JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 7, NUMBER 1 JANUARY 1966

Scattering by the Singular Potential r^{-4}

H. H. ALY†
Max-Planck Institut für Physik und Astrophysik, München, Germany
AND
H. J. W. MÜLLER
Institut für Theoretische Physik der Universität München, München, Germany
(Received 20 May 1965)

The potential $1/r^4$ is the only known potential more singular than the centrifugal term, for which the Schrödinger equation can be solved exactly. In the present investigation, we consider the potential in more detail than has been done so far. In particular the physical S-matrix is obtained, shown to be unitary, and compared with expressions of other derivations given in the literature. The eigenvalues of the Mathieu equation are finally discussed, and the behavior of the Regge trajectories is indicated.

INTRODUCTION

IN recent years a large number of papers has been published dealing with highly singular potentials in both relativistic and nonrelativistic theories. The interest in these potentials derives from many indications that the interactions responsible for high-energy elementary-particle reactions are of a more singular nature than the Yukowa potential. For example, as Sawyer[1] and Giffen and Predazzi[2] have pointed out, the weak four-fermion interaction could very well be equivalent to a potential as singular as $1/r^6$. Now a large number of methods developed specifically for the calculation of the scattering amplitude or approximations to it (e.g., the Mandelstam representation, Fredholm determinantel method, Born approximation, etc.) fail almost completely in the case of interactions as singular as the centrifugal term in the nonrelativistic Schrödinger equation. It is precisely this breakdown which led to the development of the peratization procedure, which has been discussed in several papers in connection with highly singular potentials.[3] It is therefore of considerable interest in connection with unrenormalizable field theories to find the exact type of singularity of the interaction responsible for high-energy elementary-particle reactions. In view of the difficulties encountered in quantum field theory, potential theory has always been regarded as a solvable mathematical model or prototype illustrating the general physical behavior. It is clear that exactly solvable potentials are of particular interest, since their physics can be studied more easily. Moreover, it is often possible to use these as a sort of unperturbed potentials which will only have to be modified slightly to represent a more realistic interaction and yet be solvable by standard perturbation methods (e.g. comparison of the Coulomb potential $1/r$ with the Yukawa potential $e^{-\mu r}/r$ shows that the factor $e^{-\mu r}$ may be regarded as a

† Present address: Department of Physics, American University of Beirut, Beirut, Lebanon.

[1] R. F. Sawyer, Phys. Rev. **134**, B448 (1964).

[2] M. Giffon and E. Predazzi, Nuovo Cimento **33**, 1374 (1964).

[3] N. N. Khuri and A. Pais, Rev. Mod. Phys. **36**, 590 (1964); G. Tiktopoulos and S. B. Treiman, Phys. Rev. **134**, B844 (1964); H. H. Aly, Riazuddin, and A. H. Zimerman, Phys. Rev. **136**, B1174 (1964); Nuovo Cimento **35**, 324 (1965); J. Math. Phys. **6**, 1115 (1965); T. T. Wu, Phys. Rev. **136**, B1176 (1964).

perturbation on the Coulomb potential).[4] However, this is beyond the scope of the present investigation.

It is obvious, therefore, that a highly singular potential for which the Schrödinger equation can be solved exactly is of considerable interest. Such a potential is r^{-4}. It was first discussed by Wannier and Vogt[5] for the motion of gaseous ions and electrons through a gas of medium-sized molecules. More recently Challifour and Eden[6] discussed the work of the former authors with regard to the behavior of the scattering amplitude for highly singular potentials. Such a study is indeed of great interest, since in these cases the Mandelstam representation would require an infinite number of subtractions, and this cannot be used. This has already been pointed out by Challifour and Eden, without however, indicating the specific behavior of the Regge trajectories. Neither is this attempted in the paper of Spector[7], which contains a formal expression of the S-matrix for a repulsive potential in terms of Mathieu functions defined by Meixner and Schöfke[8] (hereafter referred to as MS), whereas the work of Wannier and Vogt contains the derivation for the attractive potential in terms of functions defined earlier by Wannier.[9]

In the present investigation we discuss the problem of the potential r^{-4} in more detail. In particular, we show the connection between the work of Wannier and Vogt and that of Spector, we establish the unitarity of the S-matrix, for pure elastic scattering, and obtain the eigenvalues and thus the Regge trajectories, using more recent results on the asymptotic behavior of Mathieu functions and their characteristic numbers[10].

2. EXACT SOLUTION OF THE SCHRÖDINGER EQUATION AND CALCULATION OF THE S-MATRIX

For the singular potential

$$V(r) = -V/(\mu r)^4 \tag{2.1}$$

[$V(r)$ repulsive for $V < 0$, attractive for $V > 0$], we have the radial Schrödinger equation

$$\varphi''(r) + \left[k^2 - \frac{l(l+1)}{r^2} + \frac{V}{(\mu r)^4}\right]\varphi(r) = 0; \tag{2.2}$$

V has dimension of energy, μ dimension of reciprocal length. Setting

$$\varphi(r) = r^{\frac{1}{2}}\psi(r), \qquad \lambda = (l+\tfrac{1}{2})^2, \tag{2.3}$$

and $x = r/\gamma a = e^z$, we obtain the equation

$$\psi''(z) - [\lambda - 2h^2\cosh 2z]\psi(z) = 0, \tag{2.4}$$

where the range $0 \leq r \leq \infty$ corresponds to $-\infty \leq z \leq +\infty$, and where

$$r_a^2 k^2 = h^2, \qquad r_a^2 = V^{\frac{1}{2}}/k\mu^2. \tag{2.5}$$

Equation (2.4) is seen to be a modified Mathieu equation.

A. The Attractive Potential

We now discuss briefly the case of an attractive potential. In this case we have $V > 0$; the behavior of solutions near $r = 0$ is obtained from the behavior of the modified Mathieu functions $\psi(z)$ for $z \to -\infty$, whereas for $r \to +\infty$, we have to use corresponding solutions of (2.4) having the correct behavior for $z \to +\infty$.

Now Wannier and Vogt have discussed at length the behavior of the modified Mathieu functions for $z \to -\infty$. Applying the WKB-method to (2.2) near $r = 0$ yields immediately the behavior

$$r \to 0\colon \varphi(r) \sim r\exp\left[\pm\frac{V^{\frac{1}{2}}}{\mu^2}\,i\int\frac{dr}{r^2}\right] = r\exp\left[\mp\frac{iV^{\frac{1}{2}}}{\mu^2}\cdot\frac{1}{r}\right]. \tag{2.6}$$

We now define

$$r \to 0\colon \varphi(r) \sim r\exp\left[ikr_a^2/r\right] \tag{2.7}$$

as an ingoing wave representing the solution which is regular at the origin. Equation (2.7) together with the usual asymptotic behavior at large distances

$$r \to \infty\colon \varphi(r) \sim e^{\pm ikr} \tag{2.8}$$

completely determines the wave function. We now note the following solutions of (2.4) defined by Wannier from their asymptotic behavior for $z \to \pm\infty$:

$$\begin{aligned} he^{(1)}(z) &\sim \frac{1}{(2h\cosh z)^{\frac{1}{2}}}\exp\left[-2ih\cosh z - \frac{i\pi}{4}\right] \\ &\sim \frac{1}{(kr)^{\frac{1}{2}}}\exp\left[-ikr - \frac{i\pi}{4}\right], \\ he^{(2)}(z) &\sim \frac{1}{(2h\cosh z)^{\frac{1}{2}}}\exp\left[+2ih\cosh z + \frac{i\pi}{4}\right] \\ &\sim \frac{1}{(kr)^{\frac{1}{2}}}\exp\left[+ikr + \frac{i\pi}{4}\right], \end{aligned} \tag{2.9}$$

[4] H. J. W. Müller (to be published).
[5] G. H. Wannier and E. Vogt, Phys. Rev. 95, 1190 (1954).
[6] J. Challifour and R. J. Eden, J. Math. Phys. 4, 359 (1963).
[7] R. M. Spector, J. Math. Phys. 5, 1185 (1964).
[8] J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Späroidfunktionen* (Springer-Verlag, Berlin, 1953).
[9] G. H. Wannier, Quart. Appl. Math. 11, 33 (1953).
[10] R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. 211, 11 (1962).

for $z > 0$ and $z \to \infty$; and

$$he^{(3)}(z) \sim \frac{1}{(2h\cosh z)^{\frac{1}{2}}} \exp\left[-2ih\cosh z - \frac{i\pi}{4}\right] \sim \left(\frac{r}{kr_a^2}\right)^{\frac{1}{2}} \exp\left[-\frac{ihr_a}{r} - \frac{i\pi}{4}\right],$$
$$he^{(4)}(z) \sim \frac{1}{(2h\cosh z)^{\frac{1}{2}}} \exp\left[2ih\cosh z + \frac{i\pi}{4}\right] \sim \left(\frac{r}{kr_a^2}\right)^{\frac{1}{2}} \exp\left[+\frac{ihr_a}{r} + \frac{i\pi}{4}\right], \tag{2.10}$$

for $z < 0$ and $z \to -\infty$. Hence, all we need to obtain the phase shift is a relationship linking $he^{(1)}$ and $he^{(2)}$ with $he^{(4)}$. This may immediately be inferred from the formulas given by Wannier. However, in order to show the connection with the work of Spector, it is necessary to quote some more details of Wannier's analysis. The function $he^{(4)}(z)$ defined above for $z < 0$ may be re-expressed as a linear combination of two other linearly independent solutions $j_\nu^{\pm}(z)$, which are identical to $Me_{\mp\nu}^{(2)}$ in the notation of MS. Then we have

$$he_{(3)}^{(4)}(z) = [e^{\frac{1}{2}i\pi(\gamma\pm\beta)} je^{-}(z) - e^{-\frac{1}{2}i\pi(\gamma\pm\beta)} j_\nu^{+}(z)]/\sin\pi\beta. \tag{2.11}$$

Here β is the Floquet parameter (usually denoted by ν) and γ is another parameter introduced by Wannier to play an analogous role in regions of instability of Mathieu functions as β in regions of stability. These parameters are, of course, functions of λ and h in Eq. (2.4). The functions $je^{\pm}(z)$ may now be continued across $z = 0$ (i.e., $r = r_a$) with the help of the relations

$$je^{\pm}(z) = \tfrac{1}{2}ie^{\mp\frac{1}{2}i\pi\gamma}[e^{\mp\frac{1}{2}i\pi\beta} he^{(1)}(z) - e^{\pm\frac{1}{2}i\pi\beta} he^{(2)}(z)]. \tag{2.12}$$

Substitution of (2.12) in (2.11) yields immediately

$$he^{(4)}(z) = -\frac{\sin\pi(\gamma+\beta)}{\sin\pi\beta} he^{(1)}(z) + \frac{\sin\pi\gamma}{\sin\pi\beta} he^{(2)}(z). \tag{2.13}$$

Hence, the S-matrix is given by

$$S = -i\sin\pi\gamma/\sin\pi(\gamma+\beta). \tag{2.14}$$

γ and β are related by

$$e^{\phi} = i\sin\pi\gamma/\sin\pi\beta \tag{2.15}$$

(cf. Wannier), where $\beta \neq$ integer.

In the derivation of Spector, the following Mathieu functions defined by MS are used:

$$M_\nu^{(j)}(z, h) = \frac{1}{m_{e_\nu}(0, h^2)} \sum_{r=-\infty}^{\infty} (-1)^r C_{2r}^{\nu}(h^2) Z_{\nu+2r}^{(j)}(2h\cosh z), \tag{2.16}$$

where $Z_\sigma^{(j)}$ for $j = 1, 2, 3, 4$ represent the Bessel functions J_σ, $J_{-\sigma}$, and the Hankel functions $H_\sigma^{(1)}$, $H_\sigma^{(2)}$, respectively. We shall now assume ν to be nonintegral, since integral values of ν require special considerations and will be discussed separately. The expansions (2.16 are valid for $|\cosh z| > 1$) MS have shown that for $R(z) \to +\infty$:

$$M_\nu^{(3)}_{(4)}(z; h) \to H_{(2)}^{(1)}(2h\cosh z), \to \frac{1}{(\pi h\cosh z)^{\frac{1}{2}}} \times \exp\left[\pm i\left(2h\cosh z - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right], \to \left(\frac{2}{\pi kr}\right)^{\frac{1}{2}} \exp\left[\pm i\left(kr - \frac{\nu\pi}{2} - \frac{\pi}{4}\right)\right], \sim he_{(1)}^{(2)}(z), \qquad R(z) \to +\infty. \tag{2.17}$$

This behavior establishes the connection between the functions $M^{(j)}$ and he. In fact, a detailed comparison of Wannier's solutions with those of MS shows that the following connection formulas hold:

$$M_\nu^{(3)}(z) = \frac{1}{i}\frac{M_\nu^{(1)}(0)}{me_\nu(0)} e^{\frac{1}{2}i\pi(\gamma-\nu)} he^{(2)}(z),$$
$$M_\nu^{(4)}(z) = -\frac{1}{i}\frac{M_\nu^{(1)}(0)}{me_\nu(0)} e^{\frac{1}{2}i\pi(\gamma+\nu)} he^{(1)}(z) \tag{2.18}$$

for $R(z) > 0$, and

$$M_\nu^{(3)}(-z) = \frac{1}{i}\frac{M_\nu^{(1)}(0)}{me_\nu(0)} e^{\frac{1}{2}i\pi(\gamma-\nu)} he^{(4)}(z),$$
$$M_\nu^{(4)}(-z) = -\frac{1}{i}\frac{M_\nu^{(1)}(0)}{me_\nu(0)} e^{\frac{1}{2}i\pi(\gamma+\nu)} he^{(3)}z \tag{2.19}$$

for $R(z) < 0$. These formulas are useful in correlating Wannier's important results to the accepted standard notation introduced by MS.

Spector's method of matching the solutions is now seen to be similar to that of Wannier. In the case of $M_\nu^{(3)}$ we have instead of (2.11) and (2.12),

$$M_\nu^{(3)}(z) = \alpha Me_\nu(z) + \beta Me_{-\nu}(z), = \alpha Me_{-\nu}(-Z) + \beta Me_\nu(-z), = A'M_\nu^{(3)}(-z) + B'M_\nu^{(4)}(-z). \tag{2.20}$$

The coefficients are then given by the expressions

$$A' = \frac{W[M_\nu^{(3)}, Me_\nu]W[Me_\nu, M_\nu^{(4)}] - W[M_\nu^{(3)}, Me_{-\nu}]W[Me_{-\nu}, M_\nu^{(4)}]}{W[Me_{-\nu}, Me_\nu]W[M_\nu^{(3)}, M_\nu^{(4)}]},$$
$$B' = \frac{W[M_\nu^{(3)}, Me_\nu]W[Me_\nu, M_\nu^{(3)}] - W[M_\nu^{(3)}, Me_{-\nu}]W[Me_{-\nu}, M_\nu^{(3)}]}{W[Me_{-\nu}, Me_\nu]W[M_\nu^{(4)}, M_\nu^{(3)}]}, \tag{2.21}$$

where the W's denote Wronskians of the solutions (cf. MS p. 171). Evaluation of these expressions yields the ratio

$$\frac{A'}{B'} = \frac{R^2 - 1}{R^2 - e^{-2i\nu\pi}}, \quad \text{where} \quad R = \frac{M_{-\nu}^{(1)}(0)}{M_\nu^{(1)}(0)}. \tag{2.22}$$

The S-matrix is then given by

$$S = -i(A'/B')e^{-i\nu\pi}. \tag{2.23}$$

Alternatively we may use the matching relationship (MS p. 171)

$$W[M_\nu^{(i)}, M_\nu^{(k)}]M_\nu^{(j)}(-z) = [M_\nu^{(k)}(0)M_\nu^{(j)\prime}(0) + M_\nu^{(k)\prime}(0)M_\nu^{(j)}(0)]M_\nu^{(j)}(z) - 2M_\nu^{(j)}(0)M_\nu^{(j)\prime}(0)M_\nu^{(k)}(z), \tag{2.24}$$

which leads to the expression

$$\frac{A'}{B'} = -\frac{1}{2}\left[\frac{M_\nu^{(4)}(0)}{M_\nu^{(3)}(0)} + \frac{M_\nu^{(4)\prime}(0)}{M_\nu^{(3)\prime}(0)}\right]. \tag{2.25}$$

Using formulas re-expressing $M_\nu^{(4)}$, $M_\nu^{(3)}$ in terms of $M_{\pm\nu}^{(1)}$ (MS p. 169), we readily obtain

$$\frac{M_\nu^{(4)}(0)}{M_\nu^{(3)}(0)} = -\frac{R - e^{i\nu\pi}}{R - e^{-i\nu\pi}}. \tag{2.26}$$

Also, by the formula (MS p. 181)

$$M_{\pm\nu}^{(1)}(z) = \frac{M_{\pm\nu}^{(1)}(0)}{me_\nu(0)} Me_{\pm\nu}(z), \tag{2.27}$$

and the fact that

$$\frac{Me'_{+\nu}(0)}{Me'_{-\nu}(0)} = \frac{Ce'_\nu(0) + Se'_\nu(0)}{Ce'_\nu(0) - Se'_\nu(0)} = -1,$$

it follows that

$$\frac{M_\nu^{(4)\prime}(0)}{M_\nu^{(3)\prime}(0)} = -\frac{R + e^{i\nu\pi}}{R + e^{-i\nu\pi}}. \tag{2.28}$$

Substituting (2.26) and (2.28) into (2.25), we again obtain the expression (2.22) for the ratio A'/B'.

B. The Repulsive Potential

In the case of a repulsive potential we have $V < 0$, i.e.,

$$r_0^2 = \frac{i|V|^{\frac{1}{2}}}{k\mu^2} = ir_a^2, \quad \text{so that} \quad z = \ln\left(\frac{r}{r_a}\right) - i\frac{\pi}{4}.$$

At the point $r = r_a$, z changes from $+i\frac{1}{4}\pi$ to $-i\frac{1}{4}\pi$. The matching relationship corresponding to Eq. (2.24) now reads

$$W[M_\nu^{(i)}, M_\nu^{(k)}]M_\nu^{(j)}(-z) = \left\{M_\nu^{(i)}\left(-i\frac{\pi}{4}\right)M_\nu^{(k)\prime}\left(i\frac{\pi}{4}\right) + M_\nu^{(j)\prime}\left(-i\frac{\pi}{4}\right)M_\nu^{(k)}\left(i\frac{\pi}{4}\right)\right\}M_\nu^{(j)}(z) - \left\{M_\nu^{(j)}\left(-i\frac{\pi}{4}\right)M_\nu^{(j)\prime}\left(i\frac{\pi}{4}\right) + M_\nu^{(j)}\left(i\frac{\pi}{4}\right)M_\nu^{(j)\prime}\left(-i\frac{\pi}{4}\right)\right\}M_\nu^{(k)}(z). \tag{2.29}$$

Evaluating the ratio of the coefficients as before for $j = 3$, $k = 4$, we again obtain (2.22) (in terms of r_0), as one would expect. (Thus there appear to be some minus-sign errors in the final expression for the S-matrix given by Spector—in agreement with results obtained by Bertocchi *et al.*[11])

C. S-Matrix for Integral Values of ν

The above results are valid only for nonintegral values of ν. We shall see, however, that periodic Mathieu functions, i.e., those for integral values of ν, are of primary interest. We therefore extend our results to cover all possible values of ν. This necessitates a separate calculation of the S-matrix for integral values of the Floquet parameter.

Now the expansions (2.16) for the solutions of the modified Mathieu equation can be shown to be convergent for $|\cosh z| \geq 1$, but uniformly convergent only when $|\cosh z| > 1$ for z complex. Another set of linearly independent solutions is given by the pairs

$$Ce(z), \quad Fe(z); \qquad Se(z), \quad Ge(z). \tag{2.30}$$

These solutions (i.e. their Fournier expansions) converge uniformly for all finite values of z. They can therefore be used to join the two regions $0 \leq |x| \leq 1$, $1 \leq |x| \leq \infty$ ($x = e^z$). In order to connect these

[11] L. Bertocchi, S. Fubini, and G. Furlan, Nuovo Cimento 35, 599 (1965).

two regions for integral values of ν, we use the solutions (2.30). Introducing the notation adopted by MS (p. 200), we now write

$$M_m^{(j)}(z;h) = Mc_m^{(j)}(z;h) \quad \text{for} \quad m = 0, 1, 2, \cdots$$

and

$$(-1)^m M_{-m}^{(j)}(z;h) = Ms_m^{(j)}(z;h) \quad \text{for} \quad m = 1, 2, 3 \cdots . \tag{2.31}$$

In regions of common validity of the two types of solutions, we have then (cf. MS Sec. 2.7)

$$Mc^{(3,4)}(z,h) = \alpha^{c(3,4)} Ce(z;h^2) \pm \beta^F Fe(z;h^2), \tag{2.32}$$

and

$$Ms^{(3,4)}(z,h) = \alpha^{s(3,4)} Se(z;h^2) \pm \beta^G Ge(z;h^2),$$

where

$$\alpha_m^{c(3,4)} = \frac{Mc_m^{(3,4)}(0;h)}{Ce_m(0;h^2)}, \qquad \alpha_m^{s(3,4)} = \frac{Ms_m^{(2,4)\prime}(0;h)}{Se_m'(0;h^2)};$$

$$\beta_m^F = i\,\frac{Me_m^{(2)\prime}(0;h)}{fe_m'(0;h)}, \qquad \beta_m^G = i\,\frac{Mc_m^{(2)}(0;h)}{ge_m(0;h)}; \tag{2.33}$$

$(m = 2n, 2n+1, 2n+2$ as required).

Here and in the following Mc, Ms are understood to mean

$$Mc_{2n},\ Mc_{2n+1},\ Ms_{2n+1},\ Ms_{2n+2} \quad \text{for} \quad n = 0, 1, 2, \cdots .$$

It is clear that the calculation proceeds along the same lines as before, except that $M^{(3)}$ is now $Mc^{(3)}$, $Ms^{(3)}$, and instead of Me_ν, $Me_{-\nu}$, we have Ce, Se; Fe, Ge. Furthermore, it is necessary to remember that $Ce(z)$, $Ge(z)$ are even, whereas $Se(z)$ and $Fe(z)$ are odd functions of z. The ratio A'/B' is then found to be given by

$$\frac{A^{c\prime}}{B^{c\prime}} = \frac{\alpha^{c(3)} - \alpha^{c(4)}}{2\alpha^{c(3)}} = i\,\frac{M_c^{(2)}(0;h)}{Mc^{(3)}(0;h)}$$

for $0 < \nu \equiv m$, $m = 2n, 2n+1$; $n = 0, 1, 2, \cdots$, and (2.34)

$$\frac{A^{s\prime}}{B^{s\prime}} = \frac{\alpha^{s(3)} - \alpha^{s(4)}}{2\alpha^{s(3)}} = i\,\frac{Ms^{(2)\prime}(0;h)}{Ms^{(3)\prime}(0;h)}$$

for $0 > \nu \equiv -m$, $m = 2n+1, 2n+2$; $n = 0, 1, 2, \cdots$, since

$$M^{(3,4)} = M^{(1)} \pm iM^{(2)}.$$

The reader can easily verify these expressions by deriving them independently from the matching relationship (2.24), which is also valid for integral values of ν and $j = 3$, $k = 4$. The ratio $A^{s\prime}/B^{s\prime}$ leads to an S-matrix corresponding to negative integral values of the quantum number ν and is thus unphysical, since ν is a function of both angular momentum and energy. The expression $A^{c\prime}/B^{c\prime}$ can be evaluated with the help of series, expansions for $Mc^{(2)}$, $Mc^{(3)}$ given by MS. However, for energies not in the immediate neighborhood of the threshold, it is more expedient to use the following very simple expression for the S-matrix obtainable in terms of high-energy asymptotic expansions[12].

$$S = i \exp\left[-2(2)^{\frac{1}{2}} ig\right] \cdot \frac{\left[1 + \sum_{i=1}^{\infty} \frac{1}{(2^7 g)^i} R_i(n)\right]}{\left[1 + \sum_{i=1}^{\infty} \frac{1}{(2^7 g)^i} R_i^*(n)\right]},$$

where the R_i's are known functions of n, and $g^2 = ik\,|v|^{\frac{1}{2}}\,\mu^2$ (repulsive potential).

3. DISCUSSION OF THE PARAMETERS OF THE MATHIEU EQUATION

Comparison of Eqs. (2.14) and (2.22) readily shows that

$$R \equiv \frac{M_{-\nu}^{(1)}(0)}{M_\nu^{(1)}(0)} = e^{i\pi\gamma}. \tag{3.1}$$

This result is also in agreement with the linkage formulas (2.18) and (2.19). It follows from (3.1) that the substitution $\nu \to -\nu$ is equivalent to replacing γ by $-\gamma$, i.e., $\phi(\beta, \gamma) \equiv \phi(-\beta, -\gamma)$. Thus γ is in effect a normalization constant.

The significance of ϕ can be seen as follows. Replacing χ by $\frac{1}{2}\pi + iy$ in the Mathieu equation

$$\psi'' + \{\lambda - 2h^2 \cos 2x\}\psi = 0,$$

we obtain

$$d^2\psi/dy^2) - (\lambda + 2h^2 \cosh 2y)\psi = 0.$$

For large values of $|y|$ this equation may be approximated by

$$(d^2\psi/dy^2) - (\lambda + h^2 e^{2|y|})\psi \simeq 0; \qquad |y| \gg 0. \tag{3.2}$$

Solutions of this equation are the modified Bessel functions

$$K_{\lambda^{\frac{1}{2}}}(\pm he^{|y|}), \qquad I_{\lambda^{\frac{1}{2}}}(\pm he^{|y|}).$$

Following Wannier,[9] we call key that solution which for $y \gg 0$ has the following asymptotic behavior:

$$\text{key} \sim \frac{\exp[-he^y]}{(he^y)^{\frac{1}{2}}}, \qquad y \gg 0. \tag{3.3}$$

Further, for positive real λ and h^2, the second derivative of ψ divided by ψ is always positive, so that key (which vanishes for $y \to +\infty$) will curve up-

[12] H. H. Aly and H. J. W. Müller (to be published).

wards and never vanish for large negative y, i.e., key must have the behavior

$$\text{key} \sim \frac{1}{(he^{|y|})^{\frac{1}{2}}}\left[\pm \exp\left(-he^{|y|}\right) + \exp\left(he^{|y|} + \phi\right)\right]. \tag{3.4}$$

This follows immediately from the fact that a general solution of (3.2) may be written

$$\psi \sim [Ae^{+he^{|y|}} + Be^{-he^{|y|}}]/(he^{|y|})^{\frac{1}{2}}. \tag{3.5}$$

ϕ in Eq. (3.4) is the parameter appearing in the linkage formulas above via the expression (2.15). We now wish to establish the relationship between ϕ ϕ and the Floquet parameter ν (or β). In fact

$$\phi \equiv i\tfrac{1}{2}\pi q = i\pi(\nu + \tfrac{1}{2}), \tag{3.6}$$

where q is approximately an odd integer.

In order to prove (3.6), we use some solutions of the Mathieu equation given by Dingle and Müller[10]. We choose the general solution of the Mathieu equation written in the form

$$Ne(x) = N^A(q, h)[A(x)e^{2h\sin x} \pm \bar{A}(x)e^{-2h\sin x}], \tag{3.7}$$

where

$$\bar{A}(q, h; x) = A(q, h; -x) = A(-q, -h; x),$$

N^A being a normalization constant and $A(x)$ an expansion in powers of $1/h$. Setting $x = \frac{1}{2}\pi + iy$, the functions $A(x)$ and $\bar{A}(x)$ are easily seen to have the following property for $|y| \gg 0$:

$$\begin{aligned} A(y) &\simeq \frac{(-1)^{\frac{1}{4}q}2^{\frac{1}{2}}D(q, h)}{(-i\sinh y)^{\frac{1}{2}}}, \\ \bar{A}(y) &\simeq \frac{(-1)^{-\frac{1}{4}q}2^{\frac{1}{2}}D(-q, -h)}{(-i\sinh y)^{\frac{1}{2}}}. \end{aligned} \tag{3.8}$$

Now using the results of Dingle and Müller[10], it can be shown that the constants $D(q, h)$, $D(-q, -h)$ are identical. $Ne(y)$ is then seen to have the following asymptotic behavior for $y \ll 0$:

$$Ne(y) \simeq \frac{N^A(q, h)2^{\frac{1}{2}}D(q, h)}{(-1)^{\frac{1}{4}q}(-i\sinh y)^{\frac{1}{2}}} \times \begin{bmatrix} \pm\exp\left(-he^{|y|}\right) \\ +e^{\frac{1}{2}i\pi q}\exp\left(he^{|y|}\right) \end{bmatrix}. \tag{3.9}$$

Comparing this equation with (3.4) above, we obtain the desired result (3.6), which is also in agreement with the results of Challifour and Eden[6].

4. PROPERTIES OF THE S-MATRIX

The expressions obtained in Sec. 2 for the elastic S-matrix are, of course, expected to be unitary. We indicate the proof for two cases.

For nonintegral values of the Floquet parameter ν, we have

$$S = -i\frac{R^2 - 1}{R^2 - e^{-2i\nu\pi}}e^{-i\nu\pi}, \tag{4.1}$$

where $R = M^{(1)}_{-\nu}(0; h)/M^{(1)}_{\nu}(0; h)$. Unitarity implies

$$S^*S = 1; \tag{4.2}$$

here the asterisk on S indicates complex conjugation of the functional form of the S-matrix together with the replacement $k \to -k$, the latter being equivalent to the interchange $h \to -ih$ [cf. Eq. (2.5)]. Equation (4.2) is readily seen to be satisfied provided

$$\text{i.e.,} \quad \begin{aligned} R &= R^*e^{-i\nu\pi}, \\ R(h) &= R(-ih)e^{-i\nu\pi}. \end{aligned} \tag{4.3}$$

In order to prove Eq. (4.3), we require several properties of Mathieu functions, which are given by MS in the following form (MS pp. 170, 181, 131):

$$M^{(j)}_{\nu}(z; h) = M^{(j)}_{\nu}(z + \tfrac{1}{2}i\pi; -ih), \tag{4.4a}$$

$$M^{(1)}_{\pm\nu}(z; h)/Me_{\pm\nu}(z; h^2) = M^{(1)}_{\pm\nu}(0; h)/me_{\nu}(0; h^2), \tag{4.4b}$$

$$Me_{\nu}(z + i\pi p) = e^{i\nu p\pi}Me_{\nu}(z), \tag{4.4c}$$

$$Me_{-\nu}(z) = Me_{\nu}(-z). \tag{4.4d}$$

For a value of $z = z_0$ still to be chosen, we may write [cf. (4.4b)]

$$\begin{aligned} R &= \frac{Me_{\nu}(z_0; h^2)}{Me_{-\nu}(z_0; h^2)}\frac{M^{(1)}_{-\nu}(z_0; h)}{M^{(1)}_{\nu}(z_0; h)}, \\ &= e^{-i\pi\nu p}\frac{Me_{\nu}(z_0 + i\pi p; h^2)}{Me_{-\nu}(z; h^2)}\frac{M^{(1)}_{-\nu}(z_0; h)}{M^{(1)}_{\nu}(z_0; h)} \end{aligned} \tag{4.5}$$

by (4.4c). Setting now $p = 1$ and $z_0 = -\frac{1}{2}i\pi$, (4.5) becomes [using (4.4a, d)]

$$R = e^{-i\nu\pi}[M^{(1)}_{-\nu}(0; -ih)/M^{(1)}_{\nu}(0; -ih)] = e^{-i\nu\pi}R^*. \tag{4.6}$$

Thus Eq. (4.3) is satisfied, and hence S is unitary.

For integral values of the Floquet parameter, we have expressions (2.34), i.e.,

$$\begin{aligned} S &= e^{-im\pi}\frac{Mc^{(2)}_m(0; h)}{Mc^{(3)}_m(0; h)}, \quad m = 2n, 2n + 1 \\ &= e^{-im\pi}\frac{Ms^{(2)\prime}_m(0; h)}{Ms^{(3)\prime}_m(0; h)}, \quad m = 2n + 1, 2n + 2. \end{aligned} \tag{4.7}$$

We prove the unitarity of the first of these expressions for $m = 2n$. This may be easily surmised with the help of the relations (cf. MS p. 200)

$$M^{(3,4)} = M^{(1)} \pm iM^{(2)}, \tag{4.8}$$

$$M_m^{(2)}(z + p\pi i; h) = (-1)^{pm} M_m^{(2)}(z; h) + 2ip(-1)^m M_m^{(1)}(z; h),$$

$$Mc_{2n}^{(f)}(z; \pm ih) = Mc_{2n}^{(f)}\left(z \pm i\frac{\pi}{2}; h\right), \tag{4.9}$$

$$Ms_{2n+2}^{(f)}(z; \pm ih) = Ms_{2n+2}^{(f)}\left(z \pm i\frac{\pi}{2}; h\right).$$

Thus,

$$\begin{aligned} S_{2n}(k) &= \frac{1}{i}\frac{Mc_{2n}^{(2)}(0; h)}{Mc_{2n}^{(2)}(0; h) - iMc_{2n}^{(1)}(0; h)}, \\ &= \frac{1}{i}\frac{Mc_{2n}^{(2)}(-\frac{1}{2}\pi i; h) + iMc_{2n}^{(1)}(-\frac{1}{2}i\pi; h)}{Mc_{2n}^{(2)}(\frac{1}{2}\pi i; h)} \\ &= \frac{1}{i}\frac{Mc_{2n}^{(2)}(0; -ih) + iMc_{2n}^{(1)}(0; -ih)}{Mc_{2n}^{(2)}(0; -ih)}, \\ &= [S^*]^{-1}, \end{aligned}$$

which had to be shown. An analogous procedure for S_{2n+2} confirms that this expression is also unitary.

The poles of the S-matrix (2.14) are readily found to be given by

$$\phi = (n + \tfrac{1}{2})\pi i, \qquad n = 0, \pm 1, \pm 2, \cdots. \tag{4.10}$$

Comparison with (3.6) shows that this implies integral values of ν or q. The S-matrix given by Eqs. (2.22) and (2.23) is expected to have poles at just the same points. Thus, (4.10) appears to be incompatible with the assumption $\nu \neq$ integer used in the derivation of these expressions. However, it may be conjectured that the S-matrix derived for integral values of ν is the analytic continuation of the expression for nonintegral values of ν.

5. REGGE TRAJECTORIES

The Regge trajectories are given by the poles of the S-matrix (in the plane of complex angular momentum) or, equivalently, by the secular equation of the eigenvalue problem. Frequently it is much easier to solve the secular equation for the eigenvalues, than to extract the roots of the S-matrix. The eigenvalues of Mathieu's equation are well-known and have been studied in great detail. They are determined by the condition that the solutions be periodic functions of periods π, 2π. Only then will their eigenvalues be also eigenvalues of the radial Schrödinger equation for the potential $1/r^4$, if these periodicity conditions are also satisfied in this problem. This implies that $\psi(z)$ vanishes at the points where $\psi(iz)$ is zero, i.e., at $iz = p(h, q)$, or where

$$r = r_a e^{-ip(h,q)}, \; r_a = (V^{\frac{1}{2}}/\mu^2 k)^{\frac{1}{2}},$$

p being a real function of h for real h. This result shows that the eigenvalues of Mathieu's equation are eigenvalues of the present problem provided the cutoff r is introduced together with the hard-core boundary conditions. However, at extremely high energies or for weak coupling, this cutoff r is seen to approach zero, so that in this limit the eigenvalues of the Mathieu equation also represent eigenvalues of the present problem with the boundary conditions discussed above. The threshold behavior of the trajectories may easily be inferred from the Mathieu equation, which for

$$h^2 \to 0, \quad \text{i.e., energy} \quad k^2 \to 0,$$

is approximately given by

$$\psi'' + \lambda\psi \sim 0$$

so that, as is well-known for $h^2 \to 0$,

$$\lambda \sim n^2, \qquad n = 0, \pm 1, \pm 2, \cdots.$$

More exactly we have

$$\lambda = n^2 + O(h^2). \tag{5.1}$$

The eigenvalues corresponding to the expression of the S-matrix given in (2.34) are in the notation of MS given by

$$\begin{aligned} \lambda_{2n} &\equiv a_{2n}, & \lambda_{2n+1} &\equiv a_{2n+1}; \\ \lambda_{(-2n-1)} &\equiv b_{2n+1}, & \lambda_{(-2n-2)} &\equiv b_{2n+2}. \end{aligned} \tag{5.2}$$

For large values of m and small values of $|h^2|$ the eigenvalues a_m are given by (cf. MS, p. 120)

$$\begin{aligned} a_m(h^2) &= m^2 \\ &+ \frac{1}{2(m^2-1)}h^4 + \frac{(5m^2+7)}{32(m^2-1)^3(m^2-4)}h^8 \\ &+ \frac{9m^4 + 58m^2 + 29}{64(m^2-1)^5(m^2-4)(m^2-9)}h^{12} + \cdots. \end{aligned} \tag{5.3}$$

In the appendix we give a very brief derivation of (5.3) by a simple method recently developed for the derivation of asymptotic expansions of transcendental functions[10]. The expressions (5.3) has been studied at great length in the literature[8], and for each m many terms are known.

The high-energy asymptotic expansion of λ is

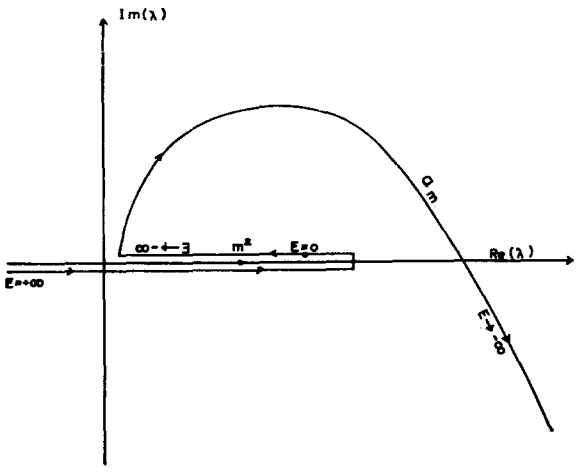

FIG. 1. Schematic trajectory of the mth Regge pole in the complex λ-plane for attractive potential.

also well known (MS p. 139; Dingle and Müller[10]):

$$a_m(h^2) = -2h^2 + 2hq - \frac{1}{2^3}(q^2+1) - \frac{1}{2^7 h}q(q^2+3) - \frac{1}{2^{12}h^2}(5q^4+34q^2+9) - \frac{1}{2^{17}h^3}q(33q^4+410q^2+405) - \frac{1}{2^{20}h^4}(63q^6+1260q^4+2943q^2+486) + \cdots, \tag{5.4}$$

where $q = 2n+1$, $n = 0, 1, 2, \cdots$.

We can now trace the trajectories of the Regge poles. Trajectories for large values of m are the only ones which account for physical bound states, including the ground state which for this singular potential would occur at $E = -\infty$. Their behavior in the vicinity of the threshold is given by the low-energy expansion, i.e., Eq. (5.3), whereas the intermediate and high-energy behavior is given by (5.4). The schematic diagram for the mth trajectory for the attractive potential is shown in Fig. 1. For simplicity the trajectory is given in the λ-plane. In order to stay on the physical sheet in the k^2-plane, we consider only regions, for which the imaginary part of $E^{\frac{1}{2}}$ is positive [$Im(E) \geq 0$], i.e., the physical sheet in the E-plane is mapped into the upper half of the k-plane, and this again is mapped into the first quadrant of the $k^{\frac{1}{2}}$-plane, and so on.

From the graph it is seen that the trajectories start at $R(\lambda) = -\infty$ for $E = +\infty$. As the energy decreases the trajectories move along the real axis and into the region of positive $R(\lambda)$. The mth trajectory is seen to move past the point $\lambda = m^2$ for some distance and reverse its direction as the energy is further reduced. Then the threshold is reached and the trajectory continues to move along the real axis. At a critical point somewhere near the origin the trajectory curves out into the upper half of the complex plane. For E approaching $-\infty$ it then travels into the lower half of the λ-plane. For very large values of m, expansion (5.4) loses its validity. However, as the energy becomes extremely large but negative the trajectory is expected to approach the ground state of the system at $\lambda = \frac{1}{4}$. It is not possible to trace the trajectories reliably in the region where they pass into the complex plane, as this is the region where both the low-energy and the high-energy expansion lose their validity. In part this follows also from a study of the radius of convergence of the low-energy expansion (cf. MS p. 121). The bound states of the mth trajectory are seen to lie at the points $R(\lambda) = \frac{1}{4}, \frac{9}{4}, \frac{25}{4}, \cdots$, within the interval $1 < R(\lambda) < m^2$.

6. CONCLUSION

For the singular potential considered in this paper the solution of the nonrelativistic Schrödinger equation does not have a simple power behavior at small distances. Moreover, near the origin the Hamilton operator is dominated by the potential term. These two characteristics are shared by the solution of the relativistic Bethe–Salpeter equation for potentials more singular than the inverse fourth power potential[13]. Recent investigations[14] indicate that an exchange of the vector meson or bubbles in the ϕ^4 field theory seem to correspond to a potential in the nonrelativistic scattering at least as singular as the centrifugal term $1/r^2$, whereas an exchange of spin-2 particles corresponds in essence to a more singular potential, for which the potential $1/r^4$ may serve as an example.

In Sec. 5 we discussed the behavior of the Regge trajectories and the occurrence of bound states for the present singular (but relatively long range) potential. This problem can be looked at in two ways. Firstly, the trajectories can be compared with those of the Coulomb potential, since both potentials have a long range. Secondly, the solution of the (nonrelativistic) radial Schrödinger equation for the potential $1/r^4$ is also the solution of the (relativistic) Klein–Gordon equation for a potential as singular as the centrifugal term[6], since

[13] A. Bastai, L. Bertocchi, G. Furlan, S. Fubini, and M. Tonin, Nuovo Cimento 30, 1512 (1963).
[14] A. P. Contogouris (to be published).

for the scalar model potential, the Klein–Gordon equation is written as

$$[(E - V)^2 - p^2 - m^2]\psi = 0.$$

It is conjectured that the mth trajectory for $m \to \infty$ would account for all bound states of the system. It is seen that the positive-energy part of the trajectories (for the attractive potential) are unphysical.

ACKNOWLEDGMENTS

We are indebted to Dr. H.-P. Dürr, Dr. M. Kretzschmar, and Dr. N. Vahedi for illuminating comments and discussions. One of us (H.H.A.) would also like to thank Dr. K. Gottstein for the hospitality extended to him at the Max–Planck Institut.

APPENDIX

We now give a very brief and simple derivation of expansion (5.3). For a discussion of the validity of the expansion and further details the reader is referred to MS.

The eigenvalues of the Mathieu equation $(-\pi \leq x \leq \pi)$

$$\psi'' + f(x)\psi = \psi'' + \{\lambda + 2h^2 \cos 2(x \pm \tfrac{1}{2}\pi)\}\psi = 0 \qquad \text{(A.1)}$$

are largely determined by the behavior of $f(x)$ in the range for which this function is positive and the solutions therefore oscillatory. Thus, for small values of h^2 and near $x = \mp\frac{1}{2}\pi$, Eq. (A.1) may be approximated by

$$\psi'' + (\lambda + 2h^2)\psi \simeq 0. \qquad \text{(A.2)}$$

Hence, [cf. (5.1)]

$$\psi = \begin{matrix}\cos \\ \sin\end{matrix} nx + o(h^2) \qquad \text{(A.3)}$$

and

$$\lambda + 2h^2 = n^2 + \Delta(h^2)\cdot h^2.$$

Writing the first approximation for

$$\psi^{(1)} = \psi_n = \cos nx,$$

we see that ψ_n leaves uncompensated terms in the Mathieu Equation amounting to

$$R_n\psi_n \equiv h^2(2 + 2\cos 2x - \Delta)\cos nx, \qquad \text{(A.4)}$$

since (A.1) may be written

$$\psi'' + n^2\psi = R_n\psi. \qquad \text{(A.5)}$$

Now it is convenient to set[10]

$$R_n\psi_n = h^2[(n, n+2)\psi_{n+2} + (n, n)\psi_n + (n, n-2)\psi_{n-2}],$$

where

$$(n, n+2) = 1 = (n, n-2), \quad (n, n) = 2 - \Delta. \qquad \text{(A.6)}$$

However, a term $\mu\psi_{n+2t}$ on the right-hand side of (A.5) may be removed by adding a further contribution

$$-[\mu/4t(t+n)]\psi_{n+2t} \quad \text{to} \quad \psi$$

(ψ_{n+2t} being a solution of

$$\psi''_{n+2t} + n^2\psi_{n+2t} = -4t(t+n)\psi_{n+2t}).$$

In this way we obtain successive contributions $\psi^{(2)}$, $\psi^{(3)}$, $\cdots$ for ψ, where

$$\psi = \psi^{(1)} + \psi^{(2)} + \psi^{(3)}, \cdots, \qquad \text{(A.7)}$$

$$\psi^{(2)} = -h^2\left[\frac{(n, n+2)}{4.1.(n+1)}\psi_{n+2} + \frac{(n, n-2)}{4(-1)(n-1)}\psi_{n-2}\right] = \frac{h^2}{4}\left[\frac{\cos(n-2)x}{(n-1)} - \frac{\cos(n+2)x}{(n+1)}\right], \qquad \text{(A.8)}$$

and so on. Then (A.7) is an eigensolution if the sum of the coefficients of all ψ_n [left uncompensative by successive approximations such as (A.8)] is set equal to zero. Thus,

$$(n, n) + h^2\left\{\frac{(n, n-2)}{4(n-1)}(n-2, n) - \frac{(n, n+2)}{4(n+1)}(n+2, n)\right\} + \cdots = 0. \qquad \text{(A.9)}$$

This equation determines the unknown function $\Delta(h^2)$. Solution of (A.9) is readily seen to yield expansion (5.3).

Local Approximation in Renormalizable and Nonrenormalizable Theories. I

K. BARDAKCI* AND B. SCHROER†‡
Institute for Advanced Study, Princeton, New Jersey
(Received 22 February 1965)

We present a framework of quantum field theory which is wide enough to incorporate renormalizable, as well as nonrenormalizable theories. A universal high-energy bound for matrix elements of fields is derived. The "string" approximation of the two-point function for various couplings is studied.

INTRODUCTION

IT is possible to formulate the content of any field theory either by specifying the Wightman functions[1] or the time-ordered (retarded) functions. The latter approach proved to be very convenient in the past in the case of renormalizable theories. However, in the case of nonrenormalizable theories, such an approach runs into difficulties.[2] In what follows, it will be shown that the Wightman approach avoids many of the difficulties connected with time ordering. Since we cannot handle a nonrenormalizable field in complete generality, we shall content ourselves mainly with the so-called ladder approximation. In this approximation we shall show that both the two-point vacuum expectation value and the two point matrix elements of fields are well defined in both renormalizable and nonrenormalizable theories, and they admit a convergent expansion in terms of the coupling constant if the latter is small. Ladder approximation to matrix elements and higher point functions are, however, not completely unique in nonrenormalizable cases. It will be shown in a subsequent paper, that there is one scaling parameter appearing, which cannot be fixed by field-theoretical principles, at least not in the considered approximation. This is contrary to the claim of recent "peratization" treatments of nonrenormalizable couplings which are based to a great deal on analogies with singular potential theory, an analogy which we consider as misleading.

The following remarks are devoted to the definition of a general framework which describes renormalizable theories and a certain class of nonrenormalizable theories. Since our concern is not so much the "axiomatic" framework itself, but rather an application of it to specific problems of "conventional" quantum field theory, the mathematical rigor of this presentation will be modest.

1. DISCUSSION OF THE FRAMEWORK

We assume that the theory is given in terms of the Wightman functions,[3] and that these functions have the full analyticity domain in coordinate space that follows from causality[4] and spectrum. There are certain nonrenormalizable interactions of a very singular nature which violate these requirements[5]; in what follows these theories are excluded.

For practical purposes, in addition to the Wightman functions, matrix elements of the form $\langle P|\ \Phi(x_1)\cdots\Phi(x_n)\ |Q\rangle$ will be considered. (Here Φ is the field and $|P\rangle$ and $|Q\rangle$ are energy momentum eigenstates.) These quantities have analyticity properties similar to those of Wightman functions.[6] Another property of Wightman functions that follows from spectral assumptions and causality are the characteristic asymptotic fall-off properties. After taking out the vacuum structure in a consistent fashion, the so-obtained "truncated" Wightman functions decrease in spacelike directions.[7] We retain this property for nonrenormalizable interactions.

In the case of renormalizable interactions, it is

* The study was supported by the Air Force Office of Scientific Research, Grant No. AF-AFOSR-42-64.
† This work supported by the National Science Foundation.
‡ Present address: University of Pittsburgh, Pittsburgh, Pennsylvania.

[1] A. S. Wightman, Phys. Rev. 101, 860 (1956).

[2] In Okubo's and Güttinger's work on nonrenormalizable models, S. Okubo, Nuovo Cimento 19, 574 (1961), W. Güttinger, Nuclear Phys. 9, 429 (1959), the basic quantities are time-ordered functions in x space.

[3] It was pointed out in B. Schroer, J. Math. Phys. 5, 1361 (1964), that the Wightman functions always can be Fourier-transformed and that all the information necessary to compute the S-matrix resides in the Wightman functions.

[4] With causality we do not necessarily mean the "operational" formulation in terms of localizable smeared-out fields (given, for example, in R. N. Streater's and A. S. Wightman's book, *PCT, Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964) but only the "formal" statement that Wightman functions are symmetric for spacelike separated points.

[5] See Ref. 3. In this paper the terminology "renormalizable of 1st and 2nd degree" was used.

[6] R. N. Streater, J. Math. Phys. 3, 256 (1962).

[7] H. Araki, K. Hepp, and D. Ruelle, Helv. Phys. Acta 35, 164 (1962).

generally assumed that Wightman functions are tempered distributions in the coordinate space, and this assumption is supported by perturbation theory. This means that the usual light cone singularity that appears when two coordinates x_1 and x_2 approach each other is not worse than $[(x_1 - x_2)^2]^{-n}$ for a certain n. If the interaction is nonrenormalizable, it is in general necessary to give up this temperedness assumption.[2] We therefore admit an arbitrary strong singularity near the light cone, and retain all the other properties of Wightman functions which are valid in ordinary renormalizable theories. These properties enable us to define a Fourier transform, no matter how bad the light-cone singularities are.

Let us give a brief discussion of this statement for the matrix element:

$$\begin{aligned} F(x; P, Q) &= \langle P|\ \Phi(-x/2)\Phi(x/2)\ |Q\rangle_{\rm tr} \\ &= \langle P|\ \Phi(-x/2)\Phi(x/2)\ |Q\rangle \\ &\quad - \langle P|\ \Phi(-x/2)\ |0\rangle\langle 0|\ \Phi(x/2)\ |Q\rangle. \end{aligned} \tag{1}$$

According to the previous remarks, this x-space distribution is the boundary value of an analytic function analytic in the holomorphic envelope of the union of the forward and backward light cone and the spacelike points. In this analyticity domain the function is symmetric under $x \to -x$. The Fourier transform is defined by

$$F(k; P, Q) = [1/(2\pi)^4] \times \lim_{\epsilon \to 0} \int d^4x\, e^{-ik_0(x_0 + i\epsilon) + i\mathbf{k}\mathbf{x}} F(x_0 + i\epsilon, \mathbf{x}; P, Q)$$

where $\epsilon \to 0$ through positive values. If F has bad singularities at $x_0 = \pm(\mathbf{x}^2)^{\frac{1}{2}}$ we cannot let $\epsilon \to 0$ in the integrand. However, since for real $\mathbf{x}$, F is analytic in the x_0 plane cut from $(\mathbf{x}^2)^{\frac{1}{2}}$ to ∞ and from $-\infty$ to $-(\mathbf{x}^2)^{\frac{1}{2}}$, we can distort the contour of integration in Eq. (3) so as to avoid the points $x_0 = \pm(\mathbf{x}^2)^{\frac{1}{2}}$. The existence of the Fourier transform is therefore assured even for very singular theories.

It is possible to derive a simple bound on the growth of $F(k; P, Q)$ for large k. As explained before, ϵ in (2) can be taken to be finite. The function F on this new path of integration is certainly bounded, and its Fourier transform will go to zero for large k by the Riemann–Lebesgue theorem. Therefore

$$F(k; P, Q)e^{-i\epsilon k_0} \to 0 \quad \text{as} \quad k_0 \to \infty. \tag{3}$$

Moreover, $\epsilon > 0$ can be taken arbitrarily small. It is also easy to see that by Lorentz invariance, ϵk_0 can be replaced by $\epsilon_\mu k^\mu$, where ϵ_μ is an arbitrary four vector in the forward cone. So the Fourier transform exists and is bounded as $k \to \infty$ along any fixed direction by $e^{(k^2)^{\frac{1}{2}}\cdot\epsilon}$ where $\epsilon > 0$ and arbitrarily small. This discussion can be generalized to higher point-expectation values. All Fourier transforms will be asymptotically of zero order growth.[8] The time-ordered boundary prescription

$$\langle P|\ T\Phi(-x/2)\Phi(x/2)\ |Q\rangle = \lim_{\epsilon \to 0} F(x_0 + i\epsilon\ \text{sign}\ x_0, \mathbf{x}; P, Q) \tag{4}$$

in general does not define a distribution (i.e., cannot be smeared with test functions nor Fourier transformed), notwithstanding that, with the exception of the light cone, (4) will be a perfectly well-behaved function in most cases.

If the Wightman function happened to be a tempered distribution, then the time-ordered function can be identified with a certain subspace of distributions, by applying dispersion theoretical techniques to the Fourier transformed Wightman function. But in general, there is no natural concept of Fourier transformation for time-ordered functions.

Finally we remark that condition (3) puts a universal high-energy bound on the total cross section. The zero-order growth for the cross section is much weaker than the polynomial growth in renormalizable theories, but by using analytic properties one can presumably improve on this bound.

2. TWO-POINT FUNCTION

The previous discussion showed that the Fourier transform $\tilde{\rho}(k^2)$ exists:

$$\langle 0|\ \Phi(0)\Phi(x)\ |0\rangle \equiv \rho(x^2) = \int d^4k\, e^{ikx}\theta(k_0)\tilde{\rho}(k^2) \tag{5}$$

and $\tilde{\rho}(k^2)$ has support in the forward cone and is bounded by $e^{\epsilon(k^2)^{\frac{1}{2}}}$ for large k^2 ($\epsilon > 0$, arbitrary). We now want to show that a violation of this bound leads to the development of a spacelike cut in the x^2 plane. Assume for example that $\tilde{\rho}$ behaves for large k^2 like

$$\tilde{\rho}(k^2) \sim e^{\alpha(k^2)^{\frac{1}{2}}} \quad \text{with} \quad \alpha > 0. \tag{6}$$

If the spectral condition is satisfied, i.e., $\tilde{\rho}(k^2) = 0$ for $k^2 < 0$, $\rho(x^2)$ in (5) converges for $\text{Im}\ x_0 > \alpha$. For other values of x^2, $\rho(x^2)$ is determined by analytic continuation. For x^2 real and <0, $\rho(x^2)$ can be written as follows:

$$\rho(\lambda) = \int d^4k\, e^{-k_0(\lambda)^{\frac{1}{2}}}\theta(k_0)\tilde{\rho}(k^2)$$

where $\lambda = -x^2$.

[8] We are using the language of J. M. Gelfand and N. Ya Vilenkin, *Generalized Functions* (Academic Press Inc., New York, 1965), Vol. 4.

According to (6) this integral diverges for $\lambda < \alpha^2$. It then follows that there must be a singularity at $\lambda = \alpha^2$, since otherwise we could prove from the inverse transformation that $\tilde{\rho}(k^2)$ is asymptotically bounded $e^{\epsilon(k^2)^{\frac{1}{2}}}$, in contradiction to the assumption of (6). This brings us back to the old bound given in (5) and the original domain of analyticity.

We now turn to a discussion of the two-point function in the so-called string approximation for three different couplings. The first coupling is the super-renormalizable Φ^3 interaction, the second one is the Φ^4 interaction and illustrates a renormalizable theory. Finally we consider the derivative coupling of the vector field to a scalar field as a model nonrenormalizable theory. The main points that emerge are the following:

(α) The equation of motion, combined with boundary conditions imposed by the spectrum, uniquely determine the solution in all cases. The solution is expandable in powers of the coupling constant for all values of this constant.

(β) The singularity near the light cone becomes worse as one proceeds from the super-renormalizable to the renormalizable case.

(γ) In the case of renormalizable theories, Fourier-transformable time-ordered functions can be defined (by subtraction procedure) up to a polynomial ambiguity whose degree depends on the coupling constant. For nonrenormalizable theories, an infinite number of arbitrary parameters enters the definition of these objects.

a. Super-renormalizable Theory

We start with the following interaction Lagrangian density:

$$L_{\text{int}} = (g/2)\phi^2(x)A(x) + \text{H.C.} \tag{7}$$

which leads to the equation of motion,

$$(\Box + m^2)\phi(x) = (g/2)[\phi(x), A(x)]_+. \tag{8}$$

The fields ϕ and A, of masses m and M, respectively, are taken to be real and scalar. Defining the Källen–Lehmann two-point function of field ϕ by

$$\rho(x^2) = \int d^4k\, e^{ik\cdot x}\tilde{\rho}(k^2) = \langle 0|\, \phi(0)\phi(x)\, |0\rangle, \tag{9}$$

we have

$$(\Box_x + m^2)(\Box_y + m^2)\rho((x-y)^2) = (g^2/4)\langle 0|\, [\phi(x), A(x)]_+[\phi(y), A(y)]_+\, |0\rangle. \tag{9'}$$

We have so far not defined the product of two field operators at the same point, hence the expressions written down so far have only formal meaning. The problem of the specification of this product is tied up with the question of the renormalization of the complete theory and in this paper we do not attempt to answer this difficult question. However, in the string approximation we are contemplating, this difficulty can be side stepped. Taking $(x-y)^2<0$, we can formally commute $A(x)$ with $\phi(y)$ in Eq. (9') and obtain

$$\begin{aligned}\tfrac{1}{4}\langle 0|\, [\phi(x)A(x)]_+[\phi(y)A(y)]_+\, |0\rangle &= \langle |\phi(x)\phi(y)A(x)A(y)|\rangle \\ &= \langle \Phi(x)\Phi(y)\rangle_0\langle A(x)A(y)\rangle_0 \\ &\quad + \sum_P \langle 0|\, \phi(x)\phi(y)\, |P\rangle\langle P|\, A(x)\phi(y)\, |0\rangle\end{aligned}$$

where we have inserted a complete set of intermediate states in the last step. The product of the two-point function, as well as the product of the matrix elements

$$M_p(x, y) = \langle 0|\, \Phi(x)\Phi(y)\, |P\rangle\langle P|\, A(x)A(y)\, |0\rangle,$$

is always well defined since in momentum space it amounts to a finite convolution. However, the summation over infinite large intermediate energies leads in general to a divergence. A proper definition of the operator product would have to take care of this divergence. The string approximation consists of only retaining the vacuum as an intermediate state and also of replacing $\langle 0|\, A(x)A(y)\, |0\rangle$ by the free propagator. This gives the following differential equation:

$$\left(4x^2\frac{d^2}{d(x^2)^2} + 8\frac{d}{dx^2} + m^2\right)^2\rho(x^2) = g^2 i\Delta^{(+)}(x^2)\rho(x^2), \tag{10}$$

where

$$i\Delta^{(+)}(x^2) = \frac{M_i}{8\pi^2(x^2)^{\frac{1}{2}}}H_1^{(2)}(M(x^2)^{\frac{1}{2}}).$$

It is clear that by analytic continuation this formula holds not only for $x^2 < 0$ but for all complex x^2. The solution will be analytic in complex x^2 plane cut for positive real x^2.

Equation (10) can be solved uniquely if one imposes the boundary condition that for large x^2, $\rho(x^2)$ goes like

$$\frac{mi}{8\pi^2(x^2)^{\frac{1}{2}}}H_1^{(2)}[m(x^2)^{\frac{1}{2}}].$$

This follows from the fact that the lowest intermediate state, with a one-particle state of mass m, dominates the large-distance behavior of $\rho(x^2)$. The integral equation that results from Eq. (10) is most

conveniently written on the half line $-\infty < x^2 < 0$. Defining $\lambda \equiv -x^2$, $\rho(-\lambda) \equiv h(\lambda)$, $D(\lambda) \equiv \Delta(-\lambda) = [M/4\pi^2(\lambda)^{\frac{1}{2}}]K_1[M(\lambda)^{\frac{1}{2}}]$, we obtain

$$h(\lambda) = \frac{m}{4\pi^2(\lambda)^{\frac{1}{2}}} K_1[m(\lambda)^{\frac{1}{2}}] + g^2 \int_\lambda^\infty d\lambda' \lambda' G(\lambda, \lambda')\, D(\lambda') h(\lambda'), \qquad (11)$$

where

$$G(\lambda, \lambda') = \frac{\pi}{4m} \frac{1}{(\lambda\lambda')^{\frac{1}{2}}} \frac{d}{dm} \{K_1[m(\lambda')^{\frac{1}{2}}] I_1[m(\lambda)^{\frac{1}{2}}] - K_1[m(\lambda)^{\frac{1}{2}}] I_0[m(\lambda')^{\frac{1}{2}}]\}.$$

Equation (11) is clearly the only integral equation that preserves the asymptotic normalization given by the inhomogeneous term. It can be solved by successive iterations in the form of a power series expansion. The equation is of Volterra type, and the kernel is of bounded norm in the space of functions square-integrable on the interval $0 \leq \lambda < \infty$. It therefore follows that the iteration always converges,[8] and the solution is an entire function in g^2. As a consequence, no bound states can emerge in this approximation.

The momentum space equation corresponding to (11) is

$$\tilde{\rho}(k^2) = \frac{1}{(2\pi)^3} \delta^+(k^2 - m^2) + \frac{g^2}{(2\pi)^3 (k^2 - m^2)^2} \times \int d^4q\, \delta^+((k-q)^2 - M^2) \tilde{\rho}(q^2). \qquad (12)$$

This equation can also be solved by a convergent iteration scheme like (11).

We want to determine the behavior of $\rho(k^2)$ for large k^2. First note that the most singular solution of Eq. (10) near $x^2 = 0$ behaves like $1/x^2$. A glance at Eq. (11) shows that such a singularity is present and comes from the free propagator contribution, but once the function $[m/4\pi^2(\lambda)^{\frac{1}{2}}]K_1[m(\lambda)^{\frac{1}{2}}]$ is subtracted from $\rho(x^2)$, the rest is only logarithmically divergent near the origin. Now the following formula can easily be established:

$$\tilde{\rho}(k^2) = -(i/4\pi^2) \oint dx^2 (x^2/k^2)^{\frac{1}{2}} J_1[(x^2k^2)^{\frac{1}{2}}] \rho(x^2). \qquad (13)$$

The contour of integration C is indicated in Fig. 1. Since $\rho(x^2)$ is L integrable on the real positive x^2 line, it follows from the Riemann–Lebesgue theorem that $\rho(k^2)$ decreases at least like $1/k^2$ for large k^2. In this approximation, one can take such a momentum space bound as the definition of a super-renormalizable interaction.

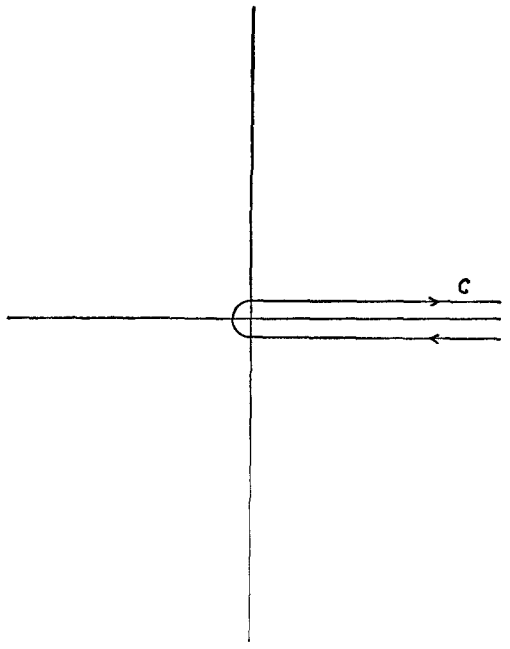

FIG. 1. The contour of integration in x^2 plane.

We finally observe that the time ordered two point function $\langle 0|\, T\phi(0)\phi(x)\, |0\rangle$ has the following momentum space expression:

$$\frac{1}{(2\pi)^4} \int d^4x\, e^{-ik\cdot x} \langle 0|\, T\phi(0)\phi(x)\, |0\rangle = \frac{i}{2\pi} \int_0^\infty dp^2 \frac{\tilde{\rho}(p^2)}{p^2 - k^2 + i\epsilon}. \qquad (14)$$

Since $\rho(k^2)$ goes to zero at infinity, the integral in question converges, and there is no trouble in uniquely defining time-ordered functions.

b. Renormalizable Theory

Consider the interaction Lagrangian density

$$L_{\text{int}} = (g/2)\phi^2(x)A^2(x) + \text{H.C.}$$

where the real scalar fields ϕ and A have masses m and M as before. The equation for the two-point function in the string approximation is

$$\left(4x^2 \frac{d^2}{d(x^2)^2} + 8 \frac{d}{dx^2} + m^2\right)^2 \rho(x^2) = g^2 i\Delta^{(+)2}(x^2) \rho(x^2). \qquad (15)$$

$i\Delta^{(+)}(x^2)$ is the free scalar propagator of mass M as before.

As in the previous case, one is led to a Volterra equation similar to (11), with $\Delta(x^2)$ replaced by $\Delta^2(x^2)$. The conclusions reached in the previous case continue to hold with the exception of large k^2 behavior of $\rho(k^2)$. To determine this behavior, we need

to solve Eq. (15) near $x^2 = 0$. Making the ansatz $\rho \sim (x^2)^\alpha$ near $x^2 = 0$, we obtain the following secular equation:

$$\alpha^2(\alpha^2 - 1) = g^2/16(4\pi^2)^2,$$

with solutions

$$\alpha_1 = \{\tfrac{1}{2} + \tfrac{1}{2}[1 + g^2/4(2\pi)^4]^{\frac{1}{2}}\}^{\frac{1}{2}},$$
$$\alpha_2 = \{\tfrac{1}{2} - \tfrac{1}{2}[1 + g^2/4(2\pi)^4]^{\frac{1}{2}}\}^{\frac{1}{2}},$$
$$\alpha_3 = -\{\tfrac{1}{2} - \tfrac{1}{2}[1 + g^2/4(2\pi)^4]^{\frac{1}{2}}\}^{\frac{1}{2}},$$
$$\alpha_4 = -\{\tfrac{1}{2} + \tfrac{1}{2}[1 + g^2/4(2\pi)^4]^{\frac{1}{2}}\}^{\frac{1}{2}}.$$

We denote the solution that behaves like $(x^2)^{\alpha_i}$ by ρ_i. Clearly, the most singular of these functions is ρ_4, which is more singular than $1/x^2$. All the other solutions are bounded near $x^2 = 0$. Since $\rho(x^2)$ must be at least as singular as $1/x^2$, it must contain ρ_4 with a nonvanishing coefficient. It is clear that ρ_4 is responsible for the leading asymptotic behavior of $\rho(k^2)$.

In Eq. (13), we make a change of variable and also split the range of integration in two parts:

$$\tilde{\rho}(k^2) = -\frac{i}{4\pi^2}\frac{1}{k^4}\left\{\int_0^\epsilon + \int_\epsilon^\infty\right\} dy^2(y^2)^{\frac{1}{2}}J_1[(y^2)^{\frac{1}{2}}] \times \Delta\rho(y^2/k^2), \quad (16)$$

where $\Delta\rho$ is the jump of ρ across the cut, and ϵ is a small number. To the integral extending from ϵ to ∞, we can apply the Riemann–Lebesgue theorem, since the integrand is bounded in this region. The integral running from 0 to ϵ may be evaluated asymptotically for large k^2 by replacing ρ by $C(x^2)^{\alpha_4}$. Neglecting insignificant multiplicative constants, we have the following asymptotic estimate:

$$\tilde{\rho}(k^2) \sim (1/k^4)\int_0^\epsilon dy^2(y^2)^{\frac{1}{2}}J_1[(y^2)^{\frac{1}{2}}](y^2/k^2)^{\alpha_4} \sim (k^2)^{-\alpha_4-2}. \quad (17)$$

Since α_4 can be arbitrarily large for large g^2, ρ can behave like an arbitrarily large power of (k^2). It is clear that the time-ordered function given by (14) may need an arbitrarily large number of subtraction constants for its definition. This situation, if it persists in an exact theory, will give rise to serious difficulties in the usual renormalization procedure, where one cannot identify a large number of ambiguities with the renormalization of mass and coupling constant.

c. Nonrenormalizable Interaction

We take a real vector meson field A of mass M, coupled to a real scalar field ϕ through derivative coupling. The interaction Lagrangian density is

$$L_{\text{int}} = (g/4)W_\mu(i\phi\partial_\mu\phi) + \text{H.C.},$$

and the equation of motion in the string approximation reads

$$\left(4x^2\frac{d^2}{d(x^2)^2} + 8\frac{d}{dx^2} + m^2\right)^2\rho(x^2) = g^2 i\Delta^{(+)}_{\mu\nu}(x^2)[\partial_\mu\partial_\nu\rho(x^2)],$$

where $\Delta_{\mu\nu}$ is the vector meson propagator given by

$$\Delta_{\mu\nu}(x^2) = \left(g_{\mu\nu} + \frac{1}{M^2}\partial_\mu\partial_\nu\right)i\Delta^{(+)}(x^2).$$

The final equation is

$$\left(4x^2\frac{d^2}{d(x^2)^2} + 8\frac{d}{dx^2} + m^2\right)^2\rho(x^2) = g^2 i\left\{6\Delta^{(+)}(x^2)\frac{d}{dx^2} - \frac{24}{M^2}x^2\Delta^{(+)\prime}(x^2)\frac{d^2}{d(x^2)^2}\right\}\rho(x^2). \quad (18)$$

The equation can be converted into a Volterra integral equation similar to Eq. (11). However, the kernel now contains the operator

$$O \equiv i\left(6\Delta^{(+)}(x^2)\frac{d}{dx^2} - \frac{24}{M^2}x^2\Delta^{(+)\prime}(x^2)\frac{d^2}{d(x^2)^2}\right)$$

instead of the function $i\Delta^{(+)}(x^2)$. We shall show that O is a bounded operator in the interval $-\infty < x^2 < -\epsilon < 0$ applied on functions analytic except for the cut on the positive real axis. (Here ϵ is a fixed positive number.) First of all, both $\Delta(x^2)$ and $\Delta'(x^2)$ are bounded functions on this interval. Secondly, given an analytic function in a region including the interval $-\infty < x^2 < -\epsilon$, if δ is the minimum distance from the boundaries of this region to a point on this interval, then the operation of differentiation increases the maximum modulus of this function at most by a factor of $1/\delta$. Hence if we norm these analytic functions by their maximum modulus, O turns our to be a bounded operator in the interval $-\infty < x^2 < -\epsilon$, whose bound, however, increases with decreasing ϵ. It then follows that the iteration solution of the Volterra equation converges for all g^2 so long as x^2 is different from zero.

Let us investigate Eq. (18) near $x^2 = 0$. To obtain the four different asymptotic solutions near the origin, we make the substitution $\rho = \exp[f(x^2)]$, and solve for the most singular term of f. The resulting asymptotic solutions are as follows:

$$\rho_1 \sim (x^2)^{\frac{1}{4}}\exp\left\{\frac{6^{\frac{1}{2}}ig}{2\pi M}\frac{1}{(x^2)^{\frac{1}{2}}}\right\},$$

$$\rho_2 \sim (x^2)^{\frac{1}{2}} \exp\left\{-\frac{6^{\frac{1}{2}} ig}{2\pi M} \frac{1}{(x^2)^{\frac{1}{2}}}\right\},$$

$$\rho_3 \sim 1 - \frac{\pi^2 M^2 m^4}{36 g^2} (x^2)^3 + \cdots,$$

$$\rho_4 \sim x^2 - \frac{M^2}{4} (x^2)^2 + \cdots.$$

From positive-definite metric and the existence of one-particle state, it follows that for x^2 negative and near zero, the correct solution must be at least as singular as $1/x^2$. Therefore, it must contain ρ_1 with a nonvanishing coefficient, since all the other ρ's are bounded near the origin. In order to determine the asymptotic behavior of $\rho(k^2)$ for large k^2, we can clearly restrict the integration region in Eq. (13) to a small loop near the origin and replace ρ by the asymptotic form of ρ_1. After a change of variable, this yields

$$\tilde{\rho}(k^2) \sim k^{-13/4} \int dy\, y^{\nu/2} J_1[(k)^{\frac{1}{2}} y] \exp\left\{\frac{iG(k)^{\frac{1}{2}}}{y}\right\} \qquad (20)$$

where $G \equiv 6g/2\pi M$. The asymptotic form of this integral can be evaluated by the saddle point method. The details are as follows.

For large k, we can replace $J_1[(k)^{\frac{1}{2}} y]$ by its asymptotic form $[(k)^{\frac{1}{2}} y]^{-\frac{1}{2}} \exp[\pm i(k)^{\frac{1}{2}} y]$, where the sign is always determined so as to get the larger of the two exponentials. Neglecting insignificant multiplicative constants, Eq. (20) can be converted into the following form:

$$\tilde{\rho}(k^2) \sim k^{-7/2} \int dy\, y^3 \exp\{i(k)^{\frac{1}{2}}[\pm y + (G/y)]\}.$$

The saddle point is at $y_0 = i(G)^{\frac{1}{2}}$, and evaluating the above integral around y_0, we obtain

$$\rho(k^2) \sim k^{-15/4} \exp\{2(Gk)^{\frac{1}{2}}\}. \qquad (21)$$

An exponential growth of this sort in momentum space, or equivalently, an exponential blowup near the origin in position space, may be taken as the characteristic feature of nonrenormalizable theories in this approximation, just as the power growth given by Eq. (17) was a feature of renormalizable interactions. It is to be noted that the bound given by (3) is never violated.

The concept of time ordering has to be given up when the momentum space behavior is as singular as given by Eq. (21), since one then needs an infinite number of subtraction constants to make Eq. (14) well defined.

One may finally ask whether the results of the previous discussion, namely expandibility in perturbation theory and essential singularities of exponential type near the light, depend crucially on the string model. It is, of course, very hard to verify these properties in any realistic situation. However, one can present an exactly soluble model discussed in Ref. 3 as a support for the above conjectures. This model contains a fermion field ψ of zero mass which satisfies the following equation:

$$i\partial\psi(x) = g^2 :\gamma_\mu A_\mu(x)\psi(x): \qquad (22)$$

where $A_\mu(x) \equiv \partial_\mu \phi_0(x)$ and $\phi_0(x)$ is a free scalar field. The two-point function has the following exact solution in this model:

$$\langle\psi(x)\bar{\psi}(y)\rangle = \langle\psi_0(x)\psi_0(y)\rangle e^{g^2 i\Delta(x-y)}, \qquad (23)$$

where ψ_0 is the free fermion field. This solution exhibits the properties of expandibility in g^2 and the singular structure near $x^2 = 0$. Of course, it is not clear whether this situation persists in a more realistic model.

ACKNOWLEDGMENT

The authors would like to thank Professor J. R. Oppenheimer for his hospitality at the Institute for Advanced Study.

Local Approximations in Renormalizable and Nonrenormalizable Theories. II

K. BARDAKCI*
The Institute for Advanced Study, Princeton, New Jersey
AND
B. SCHROER
University of Pittsburgh, Physics Department, Pittsburgh, Pennsylvania
(Received 18 March 1965)

The investigations of a previous paper are generalized to two-point matrix elements. A principle is formulated, which yields unique finite Feynman rules in the renormalizable case, i.e., permits a unique separation of counterterms. For nonrenormalizable theories this principle yields uniqueness up to a "scaling" parameter. The results are generalized to a large class of Feynman graphs. For this subset of graphs, field-theoretical principles do not determine this scaling parameter.

INTRODUCTORY REMARKS

IN a previous paper,[1] Wightman's framework[2] of quantum field theory was slightly generalized in order to allow for nonpolynomial high-energy behavior of the expectation values. It was shown that for equal fields (interaction via another field) a convectionally nonrenormalizable theory leads in the "string approximation" to a two-point function with exponential increase if the points are getting closer in spacelike direction. The physically more interesting case which contains scattering information is the case of the two-point matrix element. In the ladder approximation one can derive a fourth-order differential equation for this object. The spectrum condition leads to two boundary conditions. Normalization conditions which were available in the case of the renormalized vacuum expectation value and lead to a complete set of boundary conditions, do not exist for the matrix element. Being unable to formulate such conditions directly, we look at the perturbation theory. This leads us to the problem of determining the matrix element of the field from a given current matrix element, i.e., the inversion of the Klein–Gordon operator for a given inhomogeneous part of the differential equation. This inversion problem is studied in detail, and the form of the ambiguity of the result is determined. A principle abstracted from the fundaments of quantum field theory and the usual renormalizable perturbation theory is formulated to eliminate this ambiguity. In contradistinction to the conventionally used "minimal short distance" singularity principle (in dispersion theory known as the "minimal subtraction" principle), our principle eliminates the contact term ambiguity without making an *a priori* restriction on the high-energy behavior. Applied to nonrenormalizable theories it does, however, not lead to complete uniqueness; we only obtain uniqueness up to a scaling parameter. In every order the matrix element fulfills causality and spectrum condition. The consideration can be generalized to all Feynman graphs whose divergence in the conventional approach originates from an improper treatment of the inversion of the Klein–Gordon operator for a given but singular current. Those Feynman graphs we call the CI (current integration) graphs. The rest of the conventionally divergent graphs are infinite because of the naive handling of local operator products in the same space–time points. These CL (current limiting) graphs[3] are not treated in this paper.

1. GENERAL PROPERTIES AND INTERPRETATION OF THE TWO-POINT MATRIX ELEMENT

Spectrum conditions and causality pose a considerably more difficult problem for the matrix element

$$\langle P|\ \Phi(x)\Phi(y)\ |Q\rangle \tag{1}$$

than for the vacuum expectation value. The most general matrix element was given in form of a representation for the commutator by Jost and Lehmann[4] and Dyson[5]:

* This study was supported by the Air Force Office of Scientific Research Grant No. AF-AFOSR-42-64.

[1] K. Bardakci and B. Schroer, J. Math. Phys. 7, 10 (1966), previous paper, referred to as I.

[2] A. S. Wightman, Phys. Rev. 101, 864 (1964).

[3] This terminology arises from the allegation that a careful space–time limiting procedure is a remedy, i.e., allows a complete finite treatment (without any *ad hoc* cutoff) of these graphs. The existence of a space–time limiting procedure for the definition of the current was shown in certain soluable models. See for example, K. Johnson, Nuovo Cimento 20, 773 (1961).

[4] R. Jost and L. Lehmann, Nuovo Cimento 5, 1958 (1957).

[5] F. J. Dyson, Phys. Rev. 110, 1460 (1958).

$$\langle P|\ [\Phi(\tfrac{1}{2}x),\ \Phi(-\tfrac{1}{2}x)]\ |Q\rangle = \int_0^\infty i\Delta_\kappa^{(+)}(x)\rho(\kappa^2,\ \mathbf{x})\ d\kappa^2. \tag{2}$$

Here ρ is the Jost–Lehmann kernel (for identical fields we do not need the more general Dyson kernel) which is entire in $\mathbf{x}$. The order of this entire function reflects according to the Payley–Wiener[6] theorem the shape of the support of its Fourier transform $\tilde{\rho}(\kappa^2, \mathbf{u})$ following from the particular spectrum condition of the matrix element (1). After subtracting out the vacuum intermediate state, the support of ρ is

$$\kappa^2 \geq [M - (m^2 - u^2)^{\frac{1}{2}}]^2 \qquad |u| \leq m,$$

where M is the mass of the lowest intermediate state in (1) and $\frac{1}{4}(P + Q)^2 \equiv m^2$. This formula has been derived for renormalizable theories in which case ρ turns out to have a polynomial bound in κ^2. We observe however, that the formula

$$F_\pm = \int i\Delta_\kappa^{(\pm)}(x)\rho(\kappa^2, \mathbf{x})\ dx^2 \tag{3}$$

makes sense for all kernels which are in $(\kappa^2)^{\frac{1}{2}}$ of zero-order growth.[7] In this case $F_\pm$ is analytic for spacelike x and for Im $x < 0$ resp > 0. The Fourier transform therefore can be defined by the technique of contour shifting as explained in I. Hence the Fourier transform of (2) exists, and one easily checks that it is of zero-order growth as expected.

We argue now, that also the converse argument applies, i.e., starting from a matrix element which belongs to the nonrenormalizable class discussed in I, one can derive the Jost–Lehmann representation (3) with the kernel ρ being of zero-order growth. A formal proof is outside the scope of this paper. The asymptotic fall-off properties[8] based on the representation (3) and needed for a derivation of the asymptotic condition can be carried over to our larger class of nonrenormalizable fields.

In the next section we will study local approximations for the two-point matrix element. Using similar arguments as in I, we will obtain partial singular differential equations. The problem of solving these differential equations is tantamount to an inversion of the differential operator. As it is well known from potential theory, in order to find the "right" inversion, the differential equation has to be supplemented by an interpretation of the quantities involved. From the general framework of quantum mechanics one knows that the Schrödinger amplitude is an L^2 integrable wavefunction and that the differential operator is a self-adjoint Hamiltonian. In this case the differential equation defines a self-adjoint Hilbert space problem, and the Schrödinger boundary conditions at the origin are a consequence of this framework. A matrix element of fields is, however, a quite different object than a Schrödinger wavefunction. Consider for example the matrix element

$$\langle 0|\ \Phi(x)\Phi(y)\ |\psi\rangle. \tag{4}$$

This amplitude gives a measure for the probability of finding "field excitations," i.e., partial deviations from the vacuum around the points x and y "after" the state $|\psi\rangle$ has been prepared. Note, however, that this probability interpretation is only relative, i.e., only the ratio[9]

$$\frac{|\langle 0|\ \Phi(x)\Phi(y)\ |\psi\rangle|^2}{|\langle 0|\ \Phi(x')\Phi(y')\ |\psi\rangle|^2} \tag{5}$$

has a meaning. This relative probability interpretation does, however, not lead to a self-adjoint Hilbert space problem. The only analogy with potential theory is obtained for large spacelike separations of the field excitations.

If we take as $|\psi\rangle$ for example an incoming two-particle state, then[10]

$$\langle 0|\ \Phi(\tfrac{1}{2}\mathbf{x},\ 0)\Phi(-\tfrac{1}{2}\mathbf{x},\ 0)\ |k^{\text{in}}\rangle - \langle 0|\ \Phi^{\text{in}}(\tfrac{1}{2}\mathbf{x},\ 0)\Phi^{\text{in}}(-\tfrac{1}{2}\mathbf{x},\ 0)\ |k^{\text{in}}\rangle \xrightarrow[|x|\to\infty]{} \frac{e^{ikr}}{r} f(k,\ \cos\theta). \tag{6}$$

Here $|k^{\text{in}}\rangle$ denotes the two-particle state after taking out the center-of-mass motion, i.e., writing the two-particle Hilbert space,

$$\mathfrak{H} = \mathfrak{H}(P) \otimes \mathfrak{H}_l;$$

$|k\rangle$ is a vector out of the "little" Hilbert space. The coefficient f in front of the asymptotically least decreasing term is proportional to the elastic scattering amplitude and hence has an absolute probability interpretation.

It may be not uninteresting to point out that there is even a difference in the way the left-hand side approaches its asymptotic limit. In potential theory,

[6] For example, I. M. Gelfand and G. E. Schilow, *Verallgemeinerte Funktionen* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1962), Vol. II, Chap. III.

[7] As in I we use the terminology of I. M. Gelfand and N. Ya. Vilenkin, *Generalized Functions* (Academic Press Inc., New York, 1964), Vol. IV, p. 87.

[8] H. Araki, K. Hepp, and D. Ruelle, Helv. Phys. Acta **35**, 164 (1962).

[9] For spacelike x and y as well as x' and y', and for $|\psi\rangle$ with bounded energy, this is the quotient of two analytic functions. See R. N. Streater, J. Math. Phys. **3**, 256 (1962).

[10] Stationary scattering formulas of this type can be obtained from the Haag–Ruelle collision theory. A derivation of this and related statements will be discussed elsewhere.

the decrease is faster than any inverse power of r ("wave zone" behavior), whereas in local field theory there is a hierarchy of inverse powers. The coefficients of these inverse powers correspond to amplitudes of inelastic processes in very special final-state configurations. The analogy with potential theory breaks down completely for short distances, in which case the quantum field-theoretical framework does not give a L^2 integrability property of the matrix elements.

It is well known that these interpretational features have caused considerable confusion in the history of the Bethe–Salpeter equation,[11] which was initially thought of as a relativistic description of the quantum mechanical two-particle wavefunction[12] rather than an equation which describes the propagation of field excitation (and which only asymptotically has a particle probability interpretation). An often-cited "argument" in favor of an L^2 integrability boundary condition at the origin is the allegation that the nonrelativistic limit of the Bethe–Salpeter amplitude is the Schrödinger equation.[13] However, if one takes the nonrelativistic approximation literally (i.e., $|p_i|^2 \ll m^2$), a (ladder) approximation for a nonrelativistic quantum field theory is obtained for which the matrix element of "field excitation" does not have an absolute probability interpretation either. It is only after the subsequent step of neglecting creation and annihilation features by introducing an effective potential in the elastic region, that field-theoretical short-distance behavior is completely destroyed and any insight is lost.

Our remarks on the interpretation are presumably not new to field-theoretical experts. They are, however, helpful for an appreciation of our perturbation-theoretical treatment in the next section. From an analogy with the two-point vacuum expectation treatment one would expect an exponential increase for the equal field matrix element (1) if the spacelike distance decreases. This indeed turns out to be true for the summed up perturbation series and is in remarkable contrast to the result of Sawyer[14] and Domokes and Suranyi[15] who proposed in certain models (fermions interacting via fermion pairs) solutions which behave for short distances like

$$\exp\{f\rho/r\}, \qquad r^2 = x^2 - x_0^2 \qquad \rho = \pm i.$$

The L^2 solutions are increasingly oscillating for spacelike and Euclidean distances and exponentially increasing for timelike distances. The findings of these authors is based on the analogy of the Bethe–Salpeter equation in the unphysical Euclidean region with potential theory. However, not every solution of the differential equation in the Euclidean region fulfills the analyticity requirements following from causality and spectrum (incorporated in the Jost–Lehmann–Dyson representation) which are necessary in order to continue back into the physical points. It is an interesting question, whether these solutions solve the physical problem at all, and moreover whether there are any solutions in accordance with the principles of quantum field theory besides the one we construct in the next sections.

2. LADDER APPROXIMATION

For simplicity, consider an interacting neutral scalar theory of mass m, the interaction proceeding through a vector meson of mass M. The interaction Lagrangian is taken to be

$$L_{\text{int}} = (g/4)iW_\mu(\Phi\partial^\mu\Phi) \tag{7}$$

where Φ is the scalar field and W the vector field. Applying the Klein–Gordon operator twice to the matrix element

$$\langle 0|\ \Phi(x)\Phi(y)\ |\Psi\rangle \tag{8}$$

and using the equation of motion we formally obtain

$$(\Box_x + m^2)(\Box_y + m^2)\langle 0|\ \Phi(x)\Phi(y)\ |\Psi\rangle = g^2\langle 0|\ W^\mu(x)\partial_\mu\Phi(x)W^\nu(y)\partial_\nu\Phi(y)\ |\Psi\rangle \tag{9}$$

as in the two-point function case we take y spacelike with respect to x, commute $W^\nu(y)$ to the left, and insert a complete set of states. Retaining only the vacuum intermediate state and taking the W_ν propagator to be the free one, we obtain

$$(\Box_x + m^2)(\Box_y + m^2)\langle 0|\ \Phi(x)\Phi(y)\ |\Psi\rangle = g^2 i\Delta^{(+)}_{\mu\nu}(x-y)\partial_x^\mu\partial_y^\nu\langle 0|\ \Phi(x)\Phi(y)\ |\Psi\rangle \tag{10}$$

where

$$i\Delta^{(+)}_{\mu\nu}(x) = i[g_{\mu\nu} + (1/M^2)\partial_\mu\partial_\nu]\Delta^{(+)}(x) \tag{11}$$

with M the W mass.

If $|\psi\rangle$ does not have infinite energy components, then the product of the two distributions is perfectly well defined as in the two-point vacuum case (I).

[11] H. A. Bethe and E. E. Salpeter, Phys. Rev. **84**, 1232 (1951).

[12] A relativistic theory of a fixed number of particles is quite different from the Bethe–Salpeter equation. See F. Coester, Helv. Phys. Acta (to be published) and references given there.

[13] In certain "super-enormalizable" cases the Feymann perturbation theory indeed gives L^2 integrability at the origin. See G. C. Wick, Phys. Rev. **96**, 1124 (1954); R. E. Cutkosky, *ibid.* 1135 (1954).

[14] R. F. Sawyer, Phys. Rev. **134**, B448 (1964).

[15] G. Domokos and P. Suranyi, Acta Phys. Hung. **17**, 107 (1964).

The main problem is now to find the "right" inversion of the differential operator. *Consider the problem first in renormalizable perturbation theory.* Suppose the matrix element (8) is known in nth-order perturbation theory in g^2. Then the problem posed by (10) is the local integration of the matrix element of a local current. The solution is a particular local solution plus the most general local solution of the homogeneous equation. The latter is given in the following theorem.

Theorem: The most general local solution of the homogeneous equation

$$(\Box_x + m^2)G(x; y) = 0 \tag{12}$$

is

$$G(x; y) = \sum_{n=0}^{N} a_{\nu_1\ldots\nu_n}(y)\, D^{\nu_1\cdots\nu_n}\, i\Delta^{(+)}(x - y) \tag{13}$$

where $a_{\nu_1\ldots\nu_n}$ are tensorial functions of all variables excluding x.

Proof: Consider the commutator

$$H = G(x : y) - G(y ; x), \tag{14}$$

$H = 0$ for $(x-y)^2 < 0$. Since $(\Box_x + m^2)G(x : y) = 0$, it follows that $(\Box_x + m^2)G(y : x) = 0$ for x and y spacelike separated. By analytic continuation, this result holds for arbitrary x and y. Therefore, H vanishes for spacelike distances and is a solution of the homogeneous equation. According to a theorem of Garding and Malgrange any distribution solution[16] of the wave equation can be written in terms of distribution initial values on the hypersurface.

$$H(x ; y) = \int i\,\Delta(x - x')\, \partial_{x_0'} H(\mathbf{x}, x_0' ; y\cdot x_0')\, d^3x'. \tag{15}$$

Since the initial value has point support (according to locality), one obtains

$$H(x : y) = \sum_{n=0}^{N} a_{\nu_1\ldots\nu_n}(y)\, D_x^{\nu_1\cdots\nu_n} i\,\Delta(x - y). \tag{16}$$

Since it satisfies the homogeneous equation, $G(x : y)$ has a Fourier transform in the relative coordinate $x - y$ whose support is contained in some positive cone. Hence the commutator (16) can be uniquely disentangled and leads to Eq. (13) for the matrix element.

With the help of this theorem we obtain the most general solution for the inversion of

$$(\Box_x + m^2)\langle 0|\, \Phi(x)j(y)\, |\Psi\rangle = g(x, y; \Psi), \tag{17}$$

$$\langle 0|\, \Phi(x)j(y)\, |\Psi\rangle = \int K(x - x')g(x', y; \Psi)\, dx' + \sum_{n=0}^{N} a_{\nu_1,\ldots,\nu_n}(y)\, D_x^{\nu_1,\cdots,\nu_n} K(x - y), \tag{18}$$

with

$$K(\xi) = \frac{m}{4\pi^2(\xi^2)^{\frac{1}{4}}} K_1[m(\xi^2)^{\frac{1}{2}}]. \tag{19}$$

Here we have for convenience rotated into the Euclidean region, which is permissible if $|\psi\rangle$ is restricted in energy.[9] From now on all formulas are meant in the Euclidean region unless stated otherwise.

For plane a wave states: $|\psi\rangle = |P\rangle$ translation invariance yields

$$a_\nu(y) = e^{iP_\nu} a_\nu. \tag{20}$$

In renormalizable theories the contact terms on the right-hand side of (18) are conventionally excluded and uniqueness is obtained by insisting that only solution with mildest short-distance behavior[17] are to be taken. We reject this philosophy for the following reasons:

(a) It prohibits the understanding of short-distance behavior as an integral part of the theory itself and is tantamount to an *a priori* commitment on a feature of the theory we know least about. For nonrenormalizable theories it brings an increasing amount (with perturbation theoretical order) of arbitrariness.

(b) Even in renormalizable theories the philosophy of minimal short-distance singularity runs into difficulties once one goes outside of perturbation theory. The short-distance behavior may be worse for the summed up series depending on the size of the coupling constant.

We want to eliminate the ambiguities of the inversion problem (17) where g is the source term given in the previous perturbation iteration. [In what follows, we sometimes suppress the dependdence of $g(x\cdot y : \psi)$ on the quantities y and simply denote it by $g(x)$.] We assume that the inversion operator $(\Box_x + m^2)^{-1}$, considered as a mapping on the linear space of Wightman functions, commutes with all the operators that commute with the operator $(\Box + m^2)$ itself. There are two sets of operators that commute with the Klein–Gordon operator:

(i) The "Lorentz" generators $M_{\mu\nu} = x_\mu\, \partial/\partial x_\nu - x_\nu\, \partial/\partial x_\mu$.

(ii) Generators of translation $T_\mu = \partial/\partial x_\mu$.

[16] L. Garding and B. Malgrange, Math. Scand. 9, 5 (1961).

[17] O. Steinmann, "Perturbation Theory in the *LSZ*-Formalism," N.Y.U. preprint, and references given there.

These operators are supposed to act on a single distinguished variable x in the space of Wightman functions and should not be confused with the generators of the Poincaré group in the state space. However, the requirement that $(\Box_x + m^2)^{-1}$ commute with $M_{\mu\nu}$ and T_μ ensures the Lorentz and translation invariance of the result of the inversion, if the input has the same properties. Explicitly written, we have

$$[(\Box + m^2)^{-1} M_{\mu\nu} g](x) = M_{\mu\nu}\{[(\Box + m^2)^{-1} g](x)\}, \tag{21}$$

$$[(\Box + m^2)^{-1} T_\mu g](x) = T_\mu[(\Box + m^2)^{-1} g](x), \tag{22}$$

where $[(\Box + m^2)^{-1} g](x)$ fulfills

$$[(\Box + m^2)[(\Box + m^2)^{-1} g](x) = g(x).$$

In the case of $g(x)$ which are not too singular near $x^2 = 0$, there exists the well-known Feynman prescription which solves the inversion problem:

$$\langle 0|\, \Phi(x) J(y)\, |\psi\rangle = \int K(x - x') g(x', y, \psi)\, d^4x'. \tag{23}$$

Equation (24) agrees with Eq. (18) with the contact terms on the right hand side set equal to zero. It can explicitly be checked that the above formula yields a causal Wightman function which satisfies Eqs. (21) and (22), if these equations are already satisfied by the source function g.

There is a natural way to generalize the Feynman prescription to singular input functions coming from nonrenormalizable theories such that all the requirements we have stated are automatically fulfilled. Consider in the Euclidean region in x^2 the analytic function

$$(\Lambda/x^2)^\alpha; \qquad \Lambda > 0. \tag{24}$$

The boundary value taken from Im $x > 0$ at the Minkowski points defines a bona fide Wightman function (whose Fourier transform is contained in the forward light cone, and is positive-definite for large real α). Define the "source" function

$$g_\alpha(x) = (\Lambda/x^2)^\alpha g(x). \tag{25}$$

Clearly, for $\alpha \to 0$, g_α converges towards g in the topology of Wightman-distributions. Define a solution

$$\langle 0|\, \Phi(x) j(y)\, |\psi\rangle^{(\alpha)} = \int K(x - x') g_\alpha(x' \cdots)\, dx' \tag{26}$$

for arbitrary α as the analytic continuation starting from Re α being sufficiently large negative where the integral converges. If the singularity of g is directional-independent and behaves like $[C/(x - y)^2]^\beta$ with $\beta \neq$ integer to start with, then the limit $\alpha \to 0$ can be taken in (26) and the result is not dependent on Λ. If, however, β is an integer, then one may encounter poles in α for $\alpha \to 0$. In this case we define a solution by multiplying g with $1/\alpha$ and taking the residuum around $\alpha = 0$,

$$\langle 0|\, \Phi(x) j(y)\, |\psi\rangle \equiv \frac{1}{2\pi i} \oint \frac{d\alpha}{\alpha} \int K(x - x') g_\alpha(x' \cdots)\, dx'. \tag{27}$$

It is easy to check that this definition of a particular solution for integer β reduces to the previous one whenever $\alpha = 0$ is not a pole. In both cases we obtain a solution of the inhomogeneous differential equation which fulfills the translational property (21) for any g_α. The Lorentz-transformation property (22) allows to make a partial wave decomposition and discuss the particular solution (27) independently for any partial wave. We will use this in the next section. It is anticipated here, that this process works in any order of perturbation theory.

The prescription (27) is intimately connected with the "finite part" of Hadamard[18] and has been championed for many years by Caianiello,[19] who proposed it in the context of Dyson's renormalization procedure. The crucial step is now to find out whether our insistence in the "universality" of the inversion $(\Box + m^2)^{-1}$ fulfilling (21) and (22) leads to the Feynman prescription and to our generalization (27) for the singular cases. In the more detailed discussion in the next section it becomes evident that the Feynman prescription as well as (27) indeed fulfills our principles. The ambiguity we eliminated in (13) can be written as

$$e^{iPy} \sum_{n=0}^{N} a_{\nu_1 \cdots \nu_n}(g)\, D^{\nu_1 \cdots \nu_n} K(x - y), \tag{28}$$

where according to our interpretation as $(\Box + m^2)^{-1}$ as being universal continuous linear mapping on space of Wightman function, the $a_{\nu_1,\ldots,\nu_1}(g)$ are continuous linear functionals on the space of all (tempered) Wightman functions. Evidently a replacement $g \to T_\mu g$ does not lead to a differentiation of K, and hence

$$a_{\nu_1 \cdots \nu_n}(g) \equiv 0.$$

[18] J. Hadamard, *Lectures on Cauchy's problem in Linear Partial Differential Equations* (Dover Publications, Inc., New York, 1952).

[19] E. R. Caianiello, Nuovo Cimento, **13**, 637 (1959); **14**, 185 (1959).

3. DISCUSSION OF CONVERGENCE OF THE "LADDER" PERTURBATION SERIES IN A SPECIAL CONFIGURATION

Now consider the expectation values of the product of two field operators in a one-particle state $|\Psi\rangle$. We derive a "ladder" equation for it, which is analogous to (10),

$$(\Box_x + m^2)(\Box_y + m^2)\langle\psi|\ \Phi(x)\Phi(y)\ |\psi\rangle = g^2\langle\psi|\ W_\mu(x)\partial^\mu\Phi(x)W_\nu(y)\partial^\nu\Phi(y)\ |\psi\rangle; \tag{29}$$

commuting for spacelike distances and retaining as intermediate state only $||\psi||^{-1}\ |\psi\rangle$ we obtain

$$(\Box_x + m^2)(\Box_y + m^2)\langle\psi|\ \Phi(x)\Phi(y)\ |\psi\rangle = \frac{g^2}{||\psi||^2}\langle\psi|\ W_\mu(x)W_\nu(y)\ |\Psi\rangle\partial_x^\mu\partial_y^\nu\langle\psi|\ \Phi(x)\Phi(y)\ |\Psi\rangle. \tag{30}$$

Replacing W by the free field, and taking a sequence $|\psi\rangle$ which approaches a plane wave leads to

$$(\Box_x + m^2)(\Box_y + m^2)F(x-y;P) = g^2 i\Delta_{\mu\nu}^{(+)}\partial_x^\mu\partial_y^\nu F(x-y;P) \tag{31}$$

where $F(x-y;P) = \langle P|\ \Phi(x)\Phi(y)|P\rangle$ and $i\Delta_{\mu\nu}^{(+)}$ as previously defined.

The advantage of considering this special matrix element is that it has a simple expansion in four-dimensional spherical harmonics (in the Euclidean region),

$$F(\xi;P) = \sum_{l=0}^{\infty} C_l(\cos\theta)F_l(\xi^2), \tag{32}$$

where $\cos\theta = \xi P/(\xi^2 P^2)^{\frac{1}{2}}$, $C_l(\cos\theta) = Y_{l,0,0}(\theta) = \sin(l+1)\theta/\sin\theta$. (Note that the odd partial waves vanish in the case of identical fields.)

The expansion given above converges in the largest singularity for ellipse with foci at $\cos\theta = \pm 1$ in the complex $(\cos\theta)$ plane. This region can be determined by using the Jost–Lehmann–Dyson representation (2), which for the case when $F(x, P)$ and $F(-x,P)$ have disjoint support in momentum space can be written directly for the matrix element (after taking out the vacuum intermediate state)

$$F(x,P) = \int_0^\infty d\kappa^2 i\,\Delta^{(+)}(x,\kappa^2)\rho(\mathfrak{x},\kappa^2). \tag{33}$$

Since ρ is an entire function of x, for fixed x^2 not positive, the expansion (32) converges ($x^2 = -\xi'$, for ξ Euclidean).

Now define

$$L_l = \left(4\lambda\frac{d^2}{d\lambda^2} + 8\frac{d}{d\lambda} - m^2 - \frac{l^2+2l}{\lambda}\right), \qquad \lambda = -x^2 \quad (\lambda > 0 \text{ spacelike}), \tag{34}$$

$$N_l = -6K(\lambda)\frac{d}{d\lambda} + \frac{24}{M^2}\lambda K'\frac{d^2}{d\lambda^2} + \frac{l^2+2l}{\lambda} \times\left(K - \frac{2K'}{M^2}\right), \qquad K = i\,\Delta_M^{(+)}. \tag{35}$$

Equation (31) reduces to the following for partial waves

$$L_l^2 F_l = g^2 N_l F_l. \qquad l \text{ even}. \tag{36}$$

Two boundary conditions are readily obtained from spectral requirements. Defining as f_l the partial wave matrix element minus the vacuum intermediate-state contribution

$$f_l = F_l + \frac{2}{(2\pi)^3}\frac{l+1}{m(\lambda)^{\frac{1}{2}}}(i)^l J_{l+1}[m(\lambda)^{\frac{1}{2}}], \tag{37}$$

one obtains from spectral conditions [for example by using the Jost–Lehmann Dyson representation (33)] that f_l goes to zero faster than $\lambda^{-3/2}$ as $\lambda \to \infty$. Here we assume that $m \neq 0$ and hence the decrease is actually exponential in spacelike directions. The equation reduces asymptotically to the free one and there are four solutions,

$$\frac{1}{(\lambda)^{\frac{1}{2}}}I_{l+1}[m(\lambda)^{\frac{1}{2}}], \qquad \frac{1}{(\lambda)^{\frac{1}{2}}}K_{l+1}[m(\lambda)^{\frac{1}{2}}], \qquad \frac{\partial}{\partial m^2}\frac{1}{(\lambda)^{\frac{1}{2}}}I_{l+1}[m(\lambda)^{\frac{1}{2}}], \qquad \frac{\partial}{\partial m^2}\frac{1}{(\lambda)^{\frac{1}{2}}}K_{l+1}[m(\lambda)^{\frac{1}{2}}]. \tag{38}$$

The solution that behaves like I-functions increase exponentially for large λ and must be discarded.

The behavior near the origin is

$$F_l^{(1),(2)} \sim \lambda^{\frac{1}{4}}\exp\left[\pm\frac{(6)^{\frac{1}{2}}g}{2\pi M}\frac{1}{(\lambda)^{\frac{1}{4}}}\right], \qquad F_l^{(3),(4)} \sim (\lambda^2)^{-\frac{1}{4}\pm\frac{1}{2}[1-(l^2+2l)/3]^{\frac{1}{2}}}. \tag{39}$$

We have no direct short-distance boundary condition.

In accordance with the inversion principles established in the previous section, we invert with the α-procedure in every order of perturbation theory:

$$F_l = \sum_{n=0}^{\infty} F^{(n)}(g^2)^n, \tag{40}$$

$$F_l^{(n)} = L_l^{-2}J_l^{(n)} = \frac{1}{2\pi i}\oint\frac{d\alpha}{\alpha}L_l^{-2}J_{l,\alpha}^{(n)}, \tag{41}$$

where

$$J_{l,\alpha}^{(n)} = \left(\frac{\Delta}{x^2}\right)^\alpha N_l F_l^{(n-1)} \tag{42}$$

and

$$L_l^{-1}g_l = \int_0^\infty d\lambda'\lambda' G(\lambda,\lambda')g_l(\lambda'), \tag{43a}$$

$$L_l^{-2} g_l = \int_0^\infty d\lambda' \lambda' g_l(\lambda') \frac{\partial}{\partial m^2} G_l(\lambda\lambda'), \tag{43b}$$

$$G_l(\lambda, \lambda') = \frac{\pi}{2} \frac{1+l}{(\lambda\lambda')^{\frac{1}{2}}} \{K_{l+1}[m(\lambda')^{\frac{1}{2}}] I_{l+1}[m(\lambda)^{\frac{1}{2}}] \theta(\lambda' - \lambda) + K_{l+1}[m(\lambda)^{\frac{1}{2}}] I_{l+1}[m(\lambda')^{\frac{1}{2}}] \theta(\lambda - \lambda')\}, \tag{44}$$

where λ is the Euclidean distance and all integrations are performed in the Euclidean region. The translational property is equivalent to the following set of integral equations:

$$\frac{l+2}{2(\lambda)^{\frac{1}{2}}} (L_l^{-1} g_l) + (\lambda)^{\frac{1}{2}} \frac{d}{d\lambda} (L_l^{-1} g_l) = L_{l-1}^{-1} \left[\frac{l+2}{2(\lambda)^{\frac{1}{2}}} g_l + (\lambda)^{\frac{1}{2}} \frac{d}{d\lambda} g_l \right], \tag{45a}$$

$$-\frac{l}{2(\lambda)^{\frac{1}{2}}} (L_l^{-1} g_l) + (\lambda)^{\frac{1}{2}} \frac{d}{d\lambda} (L_l^{-1} g_l) = L_{l+1}^{-1} \left[-\frac{l}{2(\lambda)^{\frac{1}{2}}} g_l + (\lambda)^{\frac{1}{2}} \frac{d}{d\lambda} g_l \right]. \tag{45b}$$

For a g_l, which is square-integrable at the origin, these identities are fulfilled as one easily checks by partial integration. Taking g to be $J_{l,\alpha}^{(n)}$ of (42), one has identities for Re (α) sufficiently negative. Since the point $\alpha = 0$ can be reached by analytic continuation [all inversion in (45) are analytic for complex α for all expression $J^{(n)}$, occurring in perturbation theory] through complex α, this property does not get lost.

The contact terms which according to the theorem of the previous section only contributes up to the Nth partial wave, are of the form

$$a_l(g_l)\{K_{l+1}[m(\lambda)^{\frac{1}{2}}](\lambda)^{-\frac{1}{2}}\}$$

and behave for short distances like $\lambda^{-\frac{1}{2}(l-2)}$. It is just for these terms that a partial integration gives surface contributions which violate (45). An α-regularization of g_l changes only the coefficients $a_l(g_l)$ but not the integer power, and hence we have confirmed in a more detailed and direct way that the principles formulated in the previous section eliminate those ambiguities.

We now discuss the convergence of the nonrenormalizable perturbation series. It is sufficient to take the case $l = 0$, since the treatment for other values of l is quite similar. To simplify the mathematics, we shall also only consider the most singular terms in the vector boson propagator and indicate later how the more general case can be treated. With these simplifications, the differential equations under consideration reduces to

$$\left(4\lambda \frac{d^2}{d\lambda^2} + 8 \frac{d}{d\lambda} - m^2\right)^2 F = G^2 \frac{1}{\lambda} F'' \tag{46a}$$

where

$$G^2 = 24g^2/4\pi M^2$$

and the boundary condition

$$F + \frac{2}{(2\pi)^3} \frac{I_1[m(\lambda)^{\frac{1}{2}}]}{m(\lambda)^{\frac{1}{2}}} \to 0 \tag{46b}$$

for large λ.

The perturbation series (41) now simplifies to

$$F = \sum F^{(n)} (G)^n \tag{47a}$$

with

$$F^{(n+1)} = \underset{\alpha \to 0}{f.p.}\ L_0^{-2} \left(\frac{\Lambda}{\lambda}\right)^\alpha \frac{1}{\lambda} F^{(n)\prime\prime} \tag{47b}$$

where we have abbreviated the α-procedure in (41) by f.p. (finite parts).

For our immediate purpose it is convenient to forget about the boundary condition (46b) at $\lambda = \infty$. To this end we define the Green's function

$$\tilde{L}_0^{-2} g = \frac{\pi}{2} \int_0^\lambda \frac{\lambda'}{(\lambda\lambda')^{\frac{1}{2}}} \frac{\partial}{\partial m^2} \{K_1[m(\lambda)^{\frac{1}{2}}] I_1[m(\lambda')^{\frac{1}{2}}] - K_1[m(\lambda')^{\frac{1}{2}}] I_1[m(\lambda)^{\frac{1}{2}}]\} g(\lambda')\, d\lambda' \tag{48}$$

so that

$$L_0^{-2} g = \tilde{L}_0^{-2} g + (\partial/\partial m^2) I_1[m(\lambda)^{\frac{1}{2}}] c_m, \tag{49}$$

where c_m is a (m-dependent) constant.

We consider now the problem (47) with L_0^{-2} replaced by $\tilde{L}_0^{-2}$. As the zero-order input we take

$$F_1^{(0)} = I_1[m(\lambda)^{\frac{1}{2}}], \quad F_2^{(0)} = (\partial/\partial m^2) I_1[m(\lambda)^{\frac{1}{2}}]. \tag{50}$$

Assuming for the moment that the so-obtained power series F_1 and F_2 converge for small G, we obtain in this way a two-parametric manifold of solutions,

$$F = \alpha F_1 + \beta F_2, \tag{51}$$

of the differential equation (46a) which fulfill the same boundary condition at the origin as (47), but in general not the condition (46b) at $\lambda \to \infty$. The constant α and β must be so fixed that this boundary condition at $\lambda = \infty$ is satisfied. To achieve this, we define three further solutions to Eq. (46a),

$$F_3(\lambda) = \frac{K_1[m(\lambda)^{\frac{1}{2}}]}{(\lambda)^{\frac{1}{2}}} + G^2 \int_\lambda^\infty d\lambda' G_V(\lambda, \lambda') \frac{d^2 F_3(\lambda)}{d\lambda^2},$$

$$F_4(\lambda) = \frac{\partial}{\partial m^2} \frac{K_1[m(\lambda)^{\frac{1}{2}}]}{(\lambda)^{\frac{1}{2}}} + G^2 \int_\lambda^\infty d\lambda' G_V(\lambda, \lambda') \frac{d^2 F_4(\lambda)}{d\lambda^2}, \tag{52}$$

and

$$F_5(\lambda) = E(-\lambda),$$

with

$$E(\lambda) = -\frac{2}{(2\pi)^3}\frac{J_1[m(\lambda^2)^{\frac{1}{2}}]}{m(\lambda^2)^{\frac{1}{2}}} - G^2\int_\lambda^\infty d_\lambda H_r(\lambda, \lambda')\frac{d^2E(\lambda')}{d\lambda'^2}.$$

The Green's functions G_r and H_r are taken to be

$$\begin{aligned} G_r(\lambda, \lambda') &= \frac{\pi}{2}\frac{1}{(\lambda\lambda')^{\frac{1}{2}}}\frac{\partial}{\partial m^2} \\ &\times \{K_1[m(\lambda')^{\frac{1}{2}}]I_1[m(\lambda)^{\frac{1}{2}}] - (\lambda \leftrightarrow \lambda')\}, \\ H_r(\lambda, \lambda') &= \frac{\pi^2}{4}\frac{1}{(\lambda\lambda')^{\frac{1}{2}}}\frac{\partial}{\partial m^2} \\ &\times \{H_1^{(1)}[m(\lambda')^{\frac{1}{2}}]J_1[m(\lambda)^{\frac{1}{2}}] - (\lambda \leftrightarrow \lambda')\}. \end{aligned} \tag{53}$$

The integral equations for the above functions are the Volterra type, and therefore they can be iterated to yield a convergent series in G^2 for any value of G^2. Further, any arbitrary solution satisfying the boundary condition (46b) can be written in the form

$$\hat{F} = F_5(\lambda) + \gamma F_4(\lambda) + \delta F_3(\lambda), \tag{54}$$

where γ and δ are constants which can depend on G^2. To get the correct solution these constants must be adjusted to satisfy the boundary conditions at $\lambda = 0$. These are implicitly defined by F_1 and F_2. We now match the solutions defined by (51) to the solution defined by (54) (for convenience) at $\lambda = 1$.

$$\begin{aligned} F(1) &= \hat{F}(1), & F'(1) &= \hat{F}'(1), \\ F''(1) &= \hat{F}''(1), & F'''(1) &= \hat{F}'''(1). \end{aligned} \tag{55}$$

This leads to four linear equations for the four unknown parameters α, β, γ, and δ. Since the F_i's are analytic functions of G^2 around $G^2 = 0$ at the point $\lambda = 1$, then α, β, γ, and δ can be written as ratios of analytic functions of G^2. These constants will therefore be meromorphic functions of G^2, and barring the unlikely case of a pole at $G^2 = 0$, they will also have a finite region of analyticity around $G^2 = 0$. This clearly implies that the perturbation expansion for the correct solution to (46a) converges for small-enough G^2. For a discussion of convergence for the power series of F_1 and F_2 with the zero-order input (50) and the Green's function $\bar{L}_0^{-2}$ (48) we refer to the Appendix.

We could have taken a complete expansion of $i\Delta_M^{(+)}(\lambda)$ into powers of λ and products of powers times log λ. This complicates the details of the discussion, but does not effect the results.

We would like to point out that the principles of the previous section which lead us to scale dependent inversion of the Klein–Gordon operator can also be applied to the two-point function vacuum expectation values problem considered in I. In this way we would obtain a finite expression for the "unrenormalized" two-point function. For large distances, such an unrenormalized two-point function contains, in perturbation theory, besides terms proportional to $K_1[m(\lambda)^{\frac{1}{2}}]/(\lambda)^{\frac{1}{2}}$ also a term proportional to

$$(\partial/\partial m^2)K_1[m(\lambda)^{\frac{1}{2}}](\lambda)^{-\frac{1}{2}}$$

(where the constant in front is related to the "mass renormalization" δm^2). The presence of this term violates the $t^{-3/2}$ asymptotic fall-off property for the two-point function and *therefore the mass renormalization is a necessity*. The discussed inversion of the Klein–Gordon operator therefore has to be renormalized in every order of perturbation theory for those functions which warrant renormalization, i.e., the two-point and the three-point function.

It should be stressed that our perturbation-theoretical approach of the two-point matrix element does not carry over directly to the S-matrix. The elastic (forward) scattering amplitude for the particular matrix element $\langle P|\ \Phi(x)\Phi(y)\ |P\rangle$ can be obtained in an analogous fashion to (6), except that the limit has to be taken for large timelike separations. However, the obtained perturbation series for unrenormalized interactions and these limits are not generally interchangeable.

A second problem is the computation of the matrix element in the physical points from the computation of the Euclidean partial waves (32). The statement that F can be analytically continued into the physical points does not necessarily imply that the spherical decomposition can be continued. It turns out that a continuous decomposition of F with respect to basis functions which form a unitary representation of the homogeneous Lorentz group takes over the role of (32) in the physical points. The connection between the Euclidean F_l and the "Lorentz" F_λ (λ continuous) will be discussed elsewhere.

4. GENERALIZATION TO CI APPROXIMATIONS OF WIGHTMAN FUNCTIONS

The perturbation theory, which we discussed in detail for the ladder approximation of the two-point matrix element, can be carried over to certain approximations for higher point expectations values and matrix elements. Consider for example the n-point function

$$W_n(x_1 \cdots x_n) = \langle 0|\ \Phi(x_1) \cdots \Phi(x_n)\ |0\rangle. \tag{56}$$

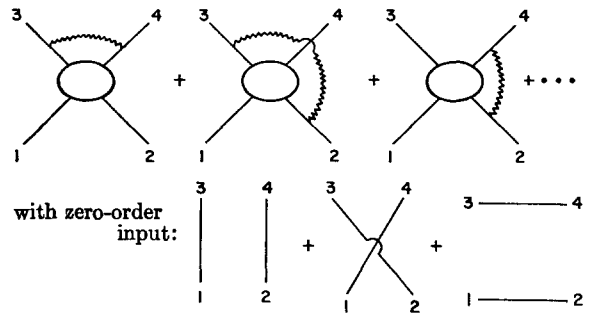

FIG. 1. The graphs of pure CI type for the four-point Wightman function.

Taking the same field equation for Φ as previously, we obtain for spacelike separated $x_1 \cdots x_n$

$$(\Box_{x_1} + m^2)(\Box_{x_2} + m^2)W_n$$
$$= g^2\langle 0|\, W_\mu(x_1)W_\nu(x_2)\partial^\mu\Phi(x_1)\partial^\nu\Phi(x_2)\Phi(x_3)\cdots\Phi(x_n)\,|0\rangle; \tag{57}$$

neglecting all intermediate states but the vacuum between the W's and the Φ's leads to the approximate equation

$$(\Box_{x_1} + m^2)(\Box_{x_2} + m^2)W_n$$
$$= g^2\langle 0|\, W_\mu(x_1)W_\nu(x_2)\,|0\rangle\partial^\mu_{x_1}\partial^\nu_{x_2}W_n. \tag{58}$$

A consideration analogous to the one in the second section shows that the same principles which eliminate the ambiguities (in the nonrenormalizable case up to a scale parameter) of the inversion in the case of the two-point matrix element also serve this purpose for (58). So for the case of a renormalizable interaction we obtain

$$W_n^{(m+1)}(x_1 \cdots x_n)$$
$$= \sum_{i<j}^{n} \iint K(x_i - x_i')K(x_j - x_j')k(x_i' - x_j')$$
$$\times W_n^{(m)}(x_1 \cdots x_i' \cdots x_j' \cdots x_n)\, dx_i'\, dx_j' \tag{59}$$

where

$$K = m[4\pi^2(\xi^2)^{\frac{1}{2}}]^{-1}K_1[m(\xi^2)^{\frac{1}{2}}],$$

and $k(\xi)$ is the two-point function of the field which is coupled to Φ. The upper index denotes the perturbation-theoretical order in g^2. Here for convenience the inversion is written down in the Euclidean region of the Wightman functions.

For the nonrenormalizable case like (58) we first reduce the inversion problem to a one-dimensional one by decomposing the Euclidean Wightman function W_n with respect to four-dimensional spherical harmonics:

$$W_n(x, \cdots x_i' \cdots x_j' \cdots x_n) = \sum \mathcal{Y}_{l,m_0,m_1}(\theta_0, \theta_1, \phi)$$
$$\times W_{n,ij}((x_i' - x_j')^2, x_1 \cdots x_n)_{l,m_0,m_1}. \tag{60}$$

One easily checks that in the $W_{n,ij}(\cdots)_{lm_0m_1}$ the problem separates. We again multiply the W_μ-boson propagator by $(\Lambda/\xi^2)^\alpha$ and invert the partial-wave differential operator as in the previous section. For the problem of continuing back into the Minkowski points the same remarks as made at the end of the previous section apply.

The expectation values which can be obtained by the iterative inversion of Klein–Gordon operators are indicated in Fig. 1 for the special case of the four-point expectation values. We call them CI (current inversion) graphs. There are other graphs which besides this inversion also involve the evaluation of two or more collapsing coordinates in the expectation value of the currents [in a way which cannot be factorized like the right-hand side of (57) in the vacuum intermediate-state approximation]. This problem is related to the problem of a careful definition of the nonlinear terms in the field equations as a "local function" of the fields involved. This important aspect of a finite (and cutoff-independent) formulation of the renormalization procedure has not even been considered in renormalizable theories in a systematic fashion. Examples of CD (current definition) graphs are given in Fig. 2. It is conceivable that the scale ambiguity in the nonrenormalizable CI graphs compensate with the scale ambiguity arising from the CD graphs. One would be inclined to conjecture that such a compensation (which is analogous to Ward's identity in quantum electrodynamics) is most likely to occur in electrodynamics of higher-spin particles. However, a discussion of the problems related with the CD graphs is very complicated and completely outside the scope of this paper.

It is worthwhile to mention that our computational procedure for the CI graphs in nonrenormalizable theories can be expressed in a very compact form by using a parameter-dependent generalized free field[20] for the W-boson field operator. The two-

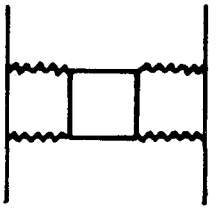

FIG. 2. Example of a CD graph.

[20] O. W. Greenberg, Ann. Phys. (N. Y.) 16, 158 (1961).

point function of this generalized free field is taken to be $(\Lambda/x^2)^\alpha$ times the free-field expression for the W-two-point function. The generalized free-field nature of W can be taken care of by writing down a (formally nonlocal) modified expression for the free part of the W-Lagrangian. In any step of the computation we have to take of course the finite part for $\alpha \to 0$.

CONCLUDING REMARKS AND ACKNOWLEDGMENTS

Although we have succeeded in finding perturbation-theoretical rules for nonrenormalizable couplings in certain approximations, the appearing of a scale ambiguity is a disconcerting feature. Only theories in which this scale parameter compensates by taking into consideration other terms (the CD graphs) can be considered as satisfactory.

It is amusing to point out that the ladder approximation of $\langle P|\ \Phi(x)\Phi(y)\ |P\rangle$ (Sec. 3) leads to an absorptive part of the forward scattering amplitude which is free of the scaling parameter Λ. This was pointed out first by M. B. Halpern.[21] It is easy to check that our perturbation-theoretical series (40) indeed leads to a power series for $\langle P|\ j(q)j(-k)\ |P\rangle$ (q, k on mass shell) which is Λ-independent and converges for arbitrary size of the coupling constant. The computation of this absorptive part is very similar to the two-point function problem discussed in our previous paper I.

After completion of this paper we received a preprint of a paper by W. Güttinger *et al.*[22] In this paper an analogous treatment of nonrenormalizable Bethe–Salpeter ladders is given. We want to emphasize, that our criticism on the analogy to singular potential theory does not apply to this paper, because the authors use a potential theoretical language only formally, but do not take it seriously as far as their computational procedure is concerned.[23]

One of the authors (K. B.) wants to thank Professor J. R. Oppenheimer for his hospitality at the Institute for Advanced Study.

APPENDIX

Consider first a second-order differential equation with $m^2 = 0$,

$$\left(8\frac{d}{d\lambda} + 4\lambda\frac{d^2}{d\lambda^2}\right)F = -\frac{g^2}{\lambda^\alpha}F. \tag{A1}$$

In this case the operator $\bar{L}_0^{-1}$ reduces to

$$Tg = \frac{\pi}{4}\int_0^\lambda \frac{\lambda'}{(\lambda\lambda')^{\frac{1}{2}}}\left\{\frac{(\lambda')^{\frac{1}{2}}}{(\lambda)^{\frac{1}{2}}} - \frac{(\lambda)^{\frac{1}{2}}}{(\lambda')^{\frac{1}{2}}}\right\}g(\lambda')\,d\lambda'. \tag{A2}$$

For $\alpha > 2$, the perturbation series corresponding to F_1 of Sec. 3 with the input $F_1^{(0)} = 1$ [corresponding to formula (50)] is

$$1 + \sum_{n=1}^{\infty}(-1)^n g^{2n}T^m\left(\frac{1}{\lambda^\alpha}\right) = 1 + \sum_{n=1}^{\infty}\frac{g^{2n}(\lambda^{-\alpha+1})^n(\pi/4)^n}{(\alpha-1)(\alpha-2)\cdots(n\alpha-n)(n\alpha-n-1)}. \tag{A3}$$

This series obviously converges. For $\alpha = 2$, the finite-part operation in $T^m(\lambda^\alpha)^{-1}$ becomes active, and we obtain the converging series

$$1 + g^2\lambda^{-1}(1-\log\lambda) + (1-\log\lambda) \times \sum_{n=2}^{\infty} g^{2n}\lambda^{-n}\left(\frac{1}{1^2}-\frac{1}{2^2}\right)\left(\frac{1}{2^2}-\frac{1}{3^2}\right)\cdots\left(\frac{1}{(n-1)^2}-\frac{1}{n^2}\right).$$

In this formula, as in the following ones, we have chosen the scale parameter $\Lambda = 1$ for convenience.

The perturbation series for the fourth-order equation

$$\left(8\frac{d}{d\lambda} + 4\lambda\frac{d^2}{d\lambda^2}\right)^2 F = \frac{g^2}{\lambda^\alpha}F, \tag{A4}$$

with the zero-order input $F_1^{(0)} = 1$ and $F_2^{(0)} = \lambda$, only involves even powers of T, and can be worked out in an analogous fashion. Therefore the functions F_1 and F_2 [Eq. (51)] which correspond to the nonrenormalizable model (A4) have a converging perturbation series.

For a discussion of the finite-mass case

$$\left(8\frac{d}{d\lambda} + 4\lambda\frac{d^2}{d\lambda^2} - m^2\right)^2 F = \frac{g^2}{\lambda^\alpha}F \tag{A5}$$

we decompose F into a double power series

$$F(\lambda) = \sum F_{i,j}(\lambda)(g^2)^i(m^2)^j. \tag{A6}$$

This gives rise to the system of differential equations,

$$\left(8\frac{d}{d\lambda} + 4\lambda\frac{d^2}{d\lambda^2}\right)^2 F_{i,j} - 2\left(8\frac{d}{d\lambda} + 4\lambda\frac{d^2}{d\lambda^2}\right)F_{i,j-1} + F_{i,j-2} = \frac{1}{\lambda^\alpha}F_{i-1,j}. \tag{A7}$$

The inversion of

$$\left(8\frac{d}{d\lambda} + 4\lambda\frac{d^2}{d\lambda^2}\right)$$

[21] M. B. Halpern, "S-Matrix Theory and Higher-Order Corrections to Weak Interaction" (Thesis, Harvard University, 1964), Part III.

[22] W. Güttinger, R. Penzl, and E. Pfaffelhuber, "Peratisation of Unrenormalizable Field Theories," University of München preprint, 1964.

[23] The treatment of singular potentials amounts to the construction of self-adjoint extensions for symmetric Hamiltonians. See K. Meetz, "Singular Potentials in Non-Relativistic Quantum Mechanics," preprint, Hamburg, 1964.

can be obtained by taking the second-order Green's function corresponding to (48),

$$\bar{L}_0^{-1} g = \frac{\pi}{2} \int_0^\lambda \frac{\lambda'}{(\lambda\lambda')^{\frac{1}{2}}} \{K_1[m(\lambda)^{\frac{1}{2}}] I_1[m(\lambda')^{\frac{1}{2}}] - (\lambda \leftrightarrow \lambda')\} g(\lambda')\, d\lambda'$$

for $m^2 = 0$. This expression is identical to the operator T in (A2). Hence we have to show that the system of equations

$$F_{i,j} - 2TF_{i,j-1} + T^2 F_{i,j-2} = T^2 \frac{1}{\lambda^\alpha} F_{i-1,j}, \tag{A8}$$

with the zero-order input [Eq. (50)]

$$\begin{aligned} F_{0j;1} &= \left(\frac{1}{4}\right)^j \frac{\lambda^j}{j!\,(j+1)!} \\ \text{and} \qquad F_{0j;2} &= j\left(\frac{1}{4}\right)^j \frac{\lambda^j}{j!\,(j+1)!}, \end{aligned} \tag{A9}$$

leads to a converging double power series (A8) F_1 and F_2.

First take Im $\alpha \neq 0$, Re arbitrary. From the structure of the system (A8) and the input (A9) follows that F_{ij} has the form

$$F_{ij} = \sum_k a_{ij}^k \lambda^{-i\alpha+k}, \tag{A10}$$

where the sum over k involves finitely-many positive values. This is evident, since the action of T on an expression of the form $\lambda^{-\beta+k}$ with Im $\beta \neq 0$ gives

$$T\lambda^{-\beta+k} = -\frac{\pi}{4} \frac{\lambda^{-\beta+k+1}}{(\beta - k - 2)(\beta - k - 1)}. \tag{A11}$$

Introducing the norm

$$N(F_{ij}) = \sum_k |a_{ij}^k| \tag{A12}$$

we see that

$$TF_{ij} = \frac{\pi}{4} \sum a_{ij}^k \frac{\lambda^{-i\alpha+l+1}}{(i\alpha - k - 2)(i\alpha - k - 1)}$$

and hence

$$\begin{aligned} N(TF_{ij}) &= \sum_k \frac{\frac{\pi}{4} |a_{ij}^k|}{|(i\alpha - k - 2)|\,|\,(i\alpha - k - 1)} \\ &\leq C \cdot \sum |a_{ij}^k| \end{aligned} \tag{A13}$$

where

$$C = \max_k \frac{\frac{1}{4}\pi}{[(\mathrm{Im}^2\,\alpha + (\ )^2)(\mathrm{Im}^2\,\alpha + (\ )^2]^{\frac{1}{2}}} \leq \frac{\frac{1}{4}\pi}{\mathrm{Im}^2\,\alpha}.$$

Therefore the mapping T is a bounded operator in the norm (A12). This leads to a convergence of the power series (A6) at the value $\lambda = 1$ for small g^2 and m^2.

In the same fashion we can show that the first three derivatives of F converge for small parameter g^2 and m^2 at $\lambda = 1$. The differential equation (A5) itself then takes care of the rest of the proof. It can be seen by a more careful discussion that the restriction to small m^2 can be removed. The restriction to small values of g^2, however, is connected with the appearance of bound states and cannot be so easily overcome.

A similar consideration holds for the differential equation (46a), if we replace the $1/\lambda$ on the right-hand side by $1/\lambda^\alpha$. In both cases the treatment of real α turns out to be much more complicated. The reason is that the finite-positive operation brings in logarithmic terms analogous to the zero-rest-mass case (A4). For $m \neq 0$ an explicit computation is, however, unfeasible. Estimates on the iteration of the system (A8) turn out to be very involved and we have not found a proof worthy of presentation.

On Series Expansion of a Determinant and Solution of the Secular Equation*

YAU W. CHAN
Brookhaven National Laboratory, Upton, New York
(Received 3 May 1965)

A determinant and its cofactors are expanded in terms of the cyclic products of its elements. With the aid of this expansion, an implicit equation for the eigenvalue and an explicit equation for amplitudes of the corresponding eigenfunction are obtained, respectively, in terms of a ratio of two simple series expansions. Comparisons with Feenberg's perturbation formula and with Sasakawa's perturbation method are also discussed.

I. INTRODUCTION

THE secular equation is known as the consequence of solving the eigenvalue problem of an operator in terms of a complete set of basic state functions. In a given space, the eigenvalue E_ω and eigenfunction Ψ_ω of an operator $\mathcal{H}$ are related by

$$(\mathcal{H} - E_\omega)\Psi_\omega = 0. \tag{1}$$

In this space, if there exists a convenient complete set of orthonormal functions u_j, the eigenfunction Ψ_ω may be expressed by the following expansion:

$$\Psi_\omega = \sum_j a_{\omega j} u_j. \tag{2}$$

To simplify our discussion, we shall use a finite set of u_j, say $j = 1, 2, \cdots, n$. There would be no loss of generality as long as we always keep in mind that n can be any number, even $n \to \infty$. Combining Eqs. (1) and (2) and taking $i = \omega$ as the chosen state, we get

$$\sum_j^n (H_{kj} - E_i\,\delta_{kj})a_{ij} = 0, \quad k = 1, 2, \cdots, n, \tag{3}$$

where $H_{kj} \equiv \langle u_k|\,\mathcal{H}\,|u_j\rangle$. This system of n homogeneous linear equations will have a nonvanishing solution if and only if the determinant formed from the coefficient of unknowns a_{ij} vanishes. This condition is known as the secular equation:

$$|H_{kj} - E_i\,\delta_{kj}| = 0. \tag{4}$$

E_i denotes one of the n solutions of the secular equation such that, for a given i,

$$\lim_{H_{ij}\to 0} E_i \to H_{ii}, \qquad \text{for all} \quad j \neq i. \tag{5}$$

That is to say, if each of the off-diagonal elements H_{ij} is allowed to decrease indefinitely to zero, the solution E_i will take H_{ii} as a limit. Thus E_i may be called the solution associated with matrix element H_{ii}.

To obtain the amplitude a_{ij} corresponding to E_i, we choose $n - 1$ equations with $k \neq i$ out of Eq. (3). By dividing each equation by a_{ii}, we have a system of $n - 1$ inhomogeneous linear equations in $n - 1$ unknowns.

$$\sum (H_{kj} - E_i\,\delta_{kj})\left(\frac{a_{ij}}{a_{ii}}\right) = -H_{ki},$$

$$k = 1, 2, \cdots, i-1, i+1, \cdots, n. \tag{6}$$

According to Cramer's rule[1] the solution of these equations is given by

$$a_{ij}/a_{ii} = -M^{ii(j)}/M^{ii}, \tag{7}$$

where M^{ii} is the complementary minor corresponding to element $(i, i) = H_{ii} - E_i$ of the determinant in Eq. (3), and $M^{ii(j)}$ is a determinant obtained from M^{ii} by replacing the j-column $(k, j) \equiv H_{kj}$ with element (k, i). However, the normalized value of a a_{ij} may be obtained using the normalizing condition $\sum_j^n |a_{ij}|^2 = 1$. We may note that

$$M^{ii(j)} = (-1)^{|i-j|-1}M^{ij} = -D^{ij}, \quad \text{for all} \quad j \neq i,$$

and

$$D^{ij} = (-1)^{i+j}M^{ij}, \qquad \text{for all} \quad i \text{ and } j, \tag{8}$$

where M^{ij} and D^{ij} are, respectively, the complementary minor and cofactor corresponding to element (i, j).

The formal solutions of E_i and a_{ij} in the form of an infinite series expansion can be derived from Eq. (3) by the method of successive order of approximation. It is generally known as the Brillouin[2]–Wigner[3]

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] H. Margenau, *The Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., New York, 1946), Chap. 10.
[2] L. Brillouin, J. Phys. Radium 3, 373 (1932).
[3] E. P. Wigner, Anz. Ungr. Akad. Wiss. Math. Naturw. 53, 475 (1935).

perturbation expansion. Feenberg[4] has transformed these solutions into a form in which no repetitive matrix elements appear in a term of given order. The Feenberg perturbation formula has also been derived by Feshbach[5] using the method of successive elimination of a_{ij} from Eq. (3). Recently, using the dispersion relation representation of a secular equation, Sasakawa[6] has proposed a new method for solving eigenvalue problems. We shall see that his Green's function expansion is closely related to the cyclic product expansion given in this note.

In this paper, we shall adopt a different method of approach. Instead of using Eq. (3), the solution of E_i and a_{ij} will be derived independently from Eqs. (4) and (7), respectively. In Sec. II, a method to expand a determinant and its minors is developed. With the aid of these expansions, solutions of E_i and a_{ij}/a_{ii} are given in Secs. III and IV, respectively. An application to the perturbation problem will be discussed in Sec. V. Finally, in Sec. VI, a brief discussion is given on the equivalence between Feenberg's formula and the result obtained by the present method. A comparison with Sasakawa's method is also briefly discussed.

I. SERIES EXPANSION OF A DETERMINANT

For convenience, an nth order determinant is simply defined by a sum of $n!$ terms, each of which is a product of n elements arranged as follows:

$$D_n = \sum (-1)^{\alpha}(1, b_1)(2, b_2) \cdots (n, b_n). \tag{9}$$

The n elements in each term are arranged in an order according to the natural order of the first indices (the row indices), whereas the arrangement of the second indices (the column indices) is one of the $n!$ permutations of integers $1, 2, \cdots, n$. In order to allow the algebraic sign of each term to be independent of particular rearrangement of the elements within the same term, the integer α is defined as the number of interchanges required for the sequence of the second indices to restore the same arrangement of the first indices which, in Eq. (9), is the natural order of n integers.

It is known that there is an intimate connection between permutations and the determinant. As in the case of permutations,[7] the product of n elements in each term of Eq. (9) may be broken up into cyclic products, each of which can be defined as follows. If we have β elements and if their indices can be arranged in a cyclic order, a cyclic product of these β elements is written as

$$C_\beta = (-1)^{\beta-1}(ijkl \cdots st), \tag{10}$$

where $(ijkl \cdots st) \equiv (i, j)(j, k)(k, l) \cdots (s, t)(t, i)$. Now $\alpha = \beta - 1$, the number of elements minus one. After a proper rearrangement of the relative order of elements, each term in D_n can be expressed as a product of C_β. Therefore,

$$D_n = \sum_{\beta_i} C_{\beta_1} C_{\beta_2} \cdots C_{\beta_r}, \tag{11}$$

with $\sum_i^r \beta_i = n$. We should note that no two C_β in the same term have a label in common. The explicit form of this expansion depends on the type of C_β chosen as principal expansion parameters. Perhaps the simplest and most interesting form is the expansion of D_n according to the product of the diagonal elements C_1^m, the product of m diagonal elements. It follows that

$$\begin{aligned} D_n &= C_1^n + \textstyle\sum_p C_1^{n-2}C_2 + \sum_p C_1^{n-3}C_3 \\ &+ \textstyle\sum_p C_1^{n-4}(C_4 + C_2^2) + \sum_p C_1^{n-5}(C_5 + C_2C_3) \\ &+ \textstyle\sum_p C_1^{n-6}(C_6 + C_4C_2 + C_3^2 + C_2^3) + \cdots \\ &+ \textstyle\sum_p C_1^{n-2r}(C_{2r} + C_{2r-2}C_2 + \cdots + C_2^r) \\ &+ \textstyle\sum_p C_1^{n-2r-1}(C_{2r+1} + C_{2r-1}C_2 + \cdots + C_3C_2^r) + \cdots \\ &+ \textstyle\sum_p (C_n + C_{n-2}C_2 + \cdots + C_3^{\frac{1}{2}[1-(-1)^n]}C_2^{\frac{1}{2}[2n-1+(-1)^n]}), \end{aligned} \tag{12}$$

where $\sum_p$ denotes the summation over all possible nonequivalent cyclic permutations.[8] For given type

$$\sum_p C_{\beta_1}^{m_1} C_{\beta_2}^{m_2} \cdots C_{\beta_r}^{m_r},$$

the total number of terms is given by[9]

$$\frac{n!}{\prod_i^r [(\beta_i!)^{m_i} m_i!]} \cdot \prod_i^r (\beta_i - 1)!^{m_i} = \frac{n!}{\prod_i^r (\beta_i^{m_i} m_i!)}. \tag{13}$$

The first factor on the left-hand side represents the number of different possible selections of n distinct objects divided into r different classes. For any class i, it consists of m_i groups, each of which has an identical number of objects β_i. The second factor represents the total number of different arrangements of objects due to the permutation of objects in each cyclic group.

[4] E. Feenberg, Phys. Rev. **74**, 206 (1948).
[5] H. Feshbach, Phys. Rev. **74**, 1548 (1948).
[6] T. Sasakawa, J. Math. Phys. **4**, 970 (1963); **5**, 379 (1964).
[7] J. S. Griffith, *The Theory of Transition-Metal Ions* (Cambridge University Press, Cambridge, England, 1961), p. 24.
[8] For example, the permutations of $(abc) = (bca) = (cab)$ are said to be equivalent, but not of $(abc) \neq (acb)$. See also Eq. (34).
[9] Consequently, $\sum_{\beta_i} [\prod_i^r (\beta_i^{m_i} m_i!)]^{-1} = 1$, since the total number of terms is equal to $n!$.

Expanding Eq. (12) to the 4th order explicitly, we get

$$D_n = \prod_i^n (i) - \sum_p^n \left[(ij) \prod_{k\neq ji}^{n-2} (k)\right] + \left[(ijk) \prod_{l\neq kji}^{n-3} (l)\right] - \sum_p^n \left\{[(ijkl) - (ij)(kl)] \prod_{m\neq lkji}^{n-4} (m)\right\} + \cdots . \quad (14)$$

By dividing the whole equation by $d_n = \prod_i^n (i)$ assuming that $d_n \neq 0$, the following form is obtained:

$$\bar{D}_n = \frac{D_n}{d_n} = \sum_p^n \left[1 - \frac{(ij)}{(i)(j)} + \frac{(ijk)}{(i)(j)(k)} - \frac{(ijkl) - (ij)(kl)}{(i)(j)(k)(l)} + \cdots\right]. \quad (15)$$

We shall call $\bar{D}_n$ the specific determinant, which will be used in the rest of this note because of its convenient form.

A similar method can be used to expand $M^{ii(j)}$ and its associated minors and cofactors. However, we may find it easier to generate the expansion of $M^{ii(j)}$ from the expansion of M^{ii}, the principal minor. The expansion of M^{ii} is simple because its labeling structure is the same as D_n except that the ith row and ith column are missing. Let $d_{ii} = \prod_{\neq i}^{n-1} (j) \neq 0$, it follows from Eq. (15) that

$$\bar{M}_{ii}^{ii} = \frac{M^{ii}}{d_{ii}} = \sum_{\substack{p \\ \neq i}}^{n-1} \left[1 - \frac{(jk)}{(j)(k)} + \frac{(jkl)}{(j)(k)(l)} - \cdots\right]. \quad (16)$$

As stated in the last section, $M^{ii(j)}$ and M^{ii} differ merely by one column. Therefore, the products of diagonal elements of both differ only by a factor $d_{ii(j)}/d_{ii} = (j, i)/(j)$. The expansion of

$$\bar{M}_{ii}^{ii(j)} (= M^{ii(j)}/d_{ii})$$

can be generated from $\bar{M}_{ii}^{ii}$ by dividing Eq. (16) into two parts. Those terms which do not have element (j) are multiplied by a factor $(j, i)/(j)$; and for the others, only those elements (x, j) are replaced by (x, i). Writing explicitly, we have

$$\bar{M}_{ii}^{ii(j)} = \sum_{\substack{p \\ \neq ij}}^{n-2} \left\{\frac{(j, i)}{(j)} \left[1 - \frac{(k, l)(l, k)}{(k)(l)} + \cdots\right] + \left[-\frac{(j, k)(k, i)}{(j)(k)} + \frac{(j, k)(k, l)(l, i)}{(j)(k)(l)} - \cdots\right]\right\}, \quad (17)$$

or

$$= \sum_{\substack{p \\ \neq ij}}^{n-2} \left\{\frac{(j, i)}{(j)} - \frac{(j, k)(k, i)}{(j)(k)} + \frac{(j, k)(k, l)(l, i) - (j, i)(k, l)(l, k)}{(j)(k)(l)} - \cdots\right\}. \quad (18)$$

With the aid of Eq. (8), the expansion of the specific minor $\bar{M}_{ii}^{ij}$ and the specific cofactor $\bar{D}_{ii}^{ij}$ can be obtained immediately.

Finally, to prepare for the discussion of the next section, we show here the result of the expansion of $\bar{D}_n$ in terms of a particular index i. To do this, Eq. (15) is divided into two parts, with the element (i) appearing in only one of them. Thus we write

$$\bar{D}_n = \sum_{\neq i}^{n-1} {}_p \left[1 - \frac{(jk)}{(j)(k)} + \frac{(jkl)}{(j)(k)(l)} - \cdots\right] + \sum_{\neq i}^{n-1} {}_p \left[-\frac{(ij)}{(i)(j)} + \frac{(ijk)}{(i)(j)(k)} - \frac{(ijkl) - (ij)(kl)}{(i)(j)(k)(l)} + \cdots\right]. \quad (19)$$

It is interesting to identify the meaning of these two sums and also to show that they agree with a well-known theorem. The first summation is identical with $\bar{M}_{ii}^{ii}$ in Eq. (16) and the second one can be shown, with Eq. (18), to be

$$-\sum_{\neq i}^{n-1} \frac{(i, j)}{(i)} \bar{M}_{ii}^{ii(i)}.$$

Therefore,

$$\bar{D}_n = \bar{M}_{ii}^{ii} - \sum_{\neq i}^{n-1} \frac{(i, j)}{(i)} \bar{M}_{ii}^{ii(i)},$$

or

$$D_n = \sum_i^n (i, j)\, D^{ij}.$$

This is the theorem, usually called Laplace development, for the expansion of a determinant in terms of its cofactors.

III. IMPLICIT SOLUTION FOR EIGENVALUE

For any eigenvalue E_i associated with matrix element H_{ii}, Eq. (4) is expanded in a form given by Eq. (19). After a simple algebraic rearrangement, the implicit solution of E_i is given by

$$E_i = H_{ii} - R_i(E_i), \quad (20)$$

where

$$R_i(E_i) = (i) = \sum_{\neq i}^{n-1} {}_p \left[\frac{(ij)}{(j)} - \frac{(ijk)}{(j)(k)} + \frac{(ijkl) - (ij)(kl)}{(j)(k)(l)} - \cdots\right] \times \left\{\sum_{\neq i}^{n-1} {}_p \left[1 - \frac{(jk)}{(j)(k)} + \frac{(jkl)}{(j)(k)(l)} - \cdots\right]\right\}^{-1}. \quad (21)$$

R_i is a function of E_i, since $(k) \equiv H_{kk} - E_i$ for all k. In principle, E_i can be solved by the iteration method[10] using Eq. (21). For example, when $H_{ii} \neq$

[10] P. M. Morse and H. Feshbach, *Method of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, Chap. 9.

H_{ii}[11] for all $j \neq i$, we may start with $R_i = 0$ as zero-iteration, i.e., ${}_{(0)}E_i = H_{ii}$. For the first iteration, ${}_{(1)}E_i$ is obtained by substituting ${}_{(0)}E_i$ in R_i of Eq. (20). Continuing with this process, one may calculate an eigenvalue with any desired degree of accuracy. This method is equivalent to the iteration method which has been used directly for a determinant without series expansion.[12] In fact, the advantage of Eq. (21) is that it can be used for a different order of approximation calculation. Higher-order terms of both the numerator and the denominator may be neglected when a complete expansion is not required. If the first-order approximation is defined by neglecting R_i, namely, $E_i^{(1)} = H_{ii}$, then Eq. (21), as written, represents the 4th-order approximation. For a pth iteration and qth-order approximation, Eq. (15) is written as

$$ {}_{(p)}E_i^{(q)} = H_{ii} - {}_{(p)}R_i^{(q)}[{}_{(p-1)}E_i^{(q)}]. \tag{22} $$

As we shall see in Sec. VI, Eq. (20) is equivalent to Feenberg's eigenvalue expansion. Hence, the condition of convergence for Feenberg's expansion which has been discussed in some detail in Ref. 10 can be applied to Eq. (20).

An alternative form of Eq. (21) may be very useful, especially if R_i is small, as in the perturbation problem. By eliminating E_i in Eq. (21) by means of Eq. (20), a solution of R_i in terms of itself is obtained:

$$ R_i(R_i) = \sum_{\neq i}^{n-1}{}_p \left[\frac{(ij)}{\{j,i\}} - \frac{(ijk)}{\{j,i\}\{k,i\}} + \cdots \right] \times \left\{ \sum_{\neq i}^{n-1}{}_p \left[1 - \frac{(jk)}{\{j,i\}\{k,i\}} + \cdots \right] \right\}^{-1}, \tag{23} $$

where $\{j,i\} \equiv H_{jj} - H_{ii} + R_i = (j) \equiv H_{jj} - E_i$. Now R_i can be calculated by iteration of itself.

At this moment, we may remind ourselves that, besides using the specific determinant expansion, R_i can be expressed also in terms of the original determinant expansion in the form of Eq. (14). Multiplying both the numerator and denominator of Eq. (21) by d_{ii}, we have

$$ R_i = \sum_{\neq i}^{n-1}{}_p \left[(ij) \prod_{\neq ij}^{n-2} (k) - (ijk) \prod_{\neq ijk}^{n-3} (l) + \cdots + \Lambda_{(0)}^{n} \right] \times \left\{ \sum_{\neq i}^{n-1}{}_p \left[\prod_{\neq i}^{n-2} (j) - (jk) \prod_{\neq ijk}^{n-3} (l) + \cdots + \Lambda_{(\neq i)}^{n-1} \right] \right\}^{-1}, \tag{24} $$

where $\Lambda_{(0)}^{n} \equiv$ the last sum of Eq. (12) and $\Lambda_{(\neq i)}^{n-1} \equiv M^{ii}$ with all $(j) = 0$. If n is finite and $(j) = 0$ for all $j \neq i$ (a hypothetical case),

$$ R_i = \Lambda_{(0)}^{n} / \Lambda_{(\neq i)}^{n-1}. \tag{25} $$

IV. EXPLICIT SOLUTION FOR AMPLITUDES

Substituting Eqs. (16) and (18) into Eq. (7) gives

$$ \frac{a_{ij}}{a_{ii}} = -\sum_{\neq ij}^{n-2}{}_p \left[\frac{(j,i)}{(j)} - \frac{(j,k)(k,i)}{(j)(k)} + \frac{(j,k)(k,l)(l,i) - (j,i)(k,l)(l,k)}{(j)(k)(l)} - \cdots \right] \times \left\{ \sum_{\neq i}^{n-1}{}_p \left[1 - \frac{(jk)}{(j)(k)} + \frac{(jkl)}{(j)(k)(l)} - \cdots \right] \right\}^{-1}. \tag{26} $$

This ratio is expressed in terms of E_i. Similarly to Eq. (23), it also can be expressed in terms of R_i by substituting $(j) = \{j,i\}$,

$$ \frac{a_{ij}}{a_{ii}} = -\sum_{\neq ij}^{n-2}{}_p \left[\frac{(j,i)}{\{j,i\}} - \frac{(j,k)(k,i)}{\{j,i\}\{k,i\}} + \cdots \right] \times \left\{ \sum_{\neq i}^{n-1}{}_p \left[1 - \frac{(jk)}{\{j,i\}\{k,i\}} + \cdots \right] \right\}^{-1}. \tag{27} $$

Therefore, the value of a_{ij}/a_{ii} can be calculated when either E_i or R_i is given. In both Eqs. (26) and (27) we note that the denominator is the same for every j value. In practice it may be set equal to a constant or simply unity to simplify the calculation, because, eventually, every value of a_{ij} will be subjected to normalization.

If both the numerator and denominator of Eq. (26) are multiplied by d_{ii}, then a_{ij}/a_{ii} is expressed in a form similar to that given in Eq. (24).

V. PERTURBATION PROBLEM

In a perturbation problem, the operator usually consists of two parts, namely,

$$ \mathcal{H} = \mathcal{H}_0 + \mathcal{H}', \tag{28} $$

where $\mathcal{H}_0$ and $\mathcal{H}'$ are called the unperturbed and the perturbed operators, respectively. For convenience, we assume that $\mathcal{H}_0 u_i = E_{0i} u_i$. Now we have $\langle u_k | \mathcal{H} | u_j \rangle = \mathcal{H}'_{kj} + (E_{0k} - E_i)\delta_{kj}$, where $H'_{kj} = \langle u_k | \mathcal{H}' | u_j \rangle$, the matrix elements of the perturbed

---

[11] When $H_{ii} = H_{j_1 j_1} = H_{j_2 j_2} = \cdots = H_{j_t j_t}$, a crude estimated value of $R_i^{(0)}$ may be obtained by means of the second-order approximation of Eq. (23). After separating the degenerate part from the nondegenerate part, one of the solutions of the quadratic equation

$$ R_i = \sum_{j_a \neq i}^{t} \frac{(ij_a)}{R_i} + \sum_{k \neq i j_a}^{n-t-1} \frac{(ik)}{H_{kk} - H_{ii} + (R_i = 0)} $$

may be used for the zero iteration. Naturally, consistency or other means should be used to justify which solution is the right one.

[12] See, for example, B. L. Cross and P. C. Crawford, J. Chem. Phys. 5, 621 (1937).

operator. Using either $(j) \equiv E_{0j} + H'_{jj} - E_i$, or $\{j, i\} \equiv E_{0j} - E_{0i} + H'_{jj} - H'_{ii} + R_i$, E_i and a_{ij} are solved with the aid of either Eqs. (21) and (26) or Eqs. (23) and (27), respectively. In order to make an interesting comparison with the Rayleigh–Schrödinger perturbation theory, we write down the results from Eqs. (23) and (27) up to third order only:

$$E_i^{(3)} = E_{0i} + H_{ii} - \left\{ \sum_{\substack{j \neq i \\ k \neq j}} \frac{H'_{ij}H'_{ji}}{E_{0j} - E_{0i} + H'_{jj} - H'_{ii} + R_i} - \sum_{\substack{j \neq i \\ k \neq j}} \frac{H'_{ij}H'_{jk}H'_{ki}}{(E_{0j} - E_{0i} + H'_{jj} - H'_{ii} + R_i)(E_{0k} - E_{0i} + H'_{kk} - H'_{ii} + R_i)} \right\} \times \left\{ 1 - \sum_{\substack{j \neq i \\ k \neq j}} \frac{H'_{jk}H'_{kj}}{(E_{0j} - E_{0i} + H'_{jj} - H'_{ii} + R_i)(E_{0k} - E_{0i} + H'_{kk} - H'_{ii} + R_i)} \right\}^{-1}, \tag{29}$$

and

$$a_{ij}^{(3)} = \left[\frac{a_{ii}}{\bar{M}_{ii}^{ii}} \right] \left[- \frac{H'_{ji}}{(E_{0j} - E_{0i} + H'_{jj} - H'_{ii} + R_i)} + \sum_{\substack{jk \neq i \\ k \neq j}} \frac{H'_{jk}H'_{ki}}{(E_{0j} - E_{0i} + H'_{jj} - H'_{ii} + R_i)(E_{0k} - E_{0i} + H'_{kk} - H_{ii} + R_i)} \right]. \tag{30}$$

No attempt will be made in this paper to discuss the difference between these results and those obtained by Rayleigh–Schrödinger's perturbation.

VI. DISCUSSION

It has been pointed out by Feenberg in Ref. 4 that his eigenvalue formula is simply a convenient way of writing the secular equation. This can be seen more instructively by following Feshbach's derivation. Therefore, there is no doubt that the Feenberg expansion and the result given by Eq. (20) are equivalent, although their forms look quite different. We shall demonstrate that Feenberg's expansion can be obtained directly from Eq. (21). We note that the product of matrix elements in each term in Feenberg's expansion is expressed only in terms of single cyclic products $(ijk \cdots)$, whereas, in Eq. (20), compound cyclic products, such as $(ijk)(lm \cdots)$ are also used. For a fraction like R_i, each compound product can be transformed into terms of single cyclic products by a method of successive dividion, dividing both the numerator and the denominator by a proper factor. This method is best illustrated by the following simple example:

$$\frac{a(1-a') + b(1-b') + c}{1-d'} = a + \frac{b(1-b') + c}{1-a'} \left(\frac{1-d'}{1-a'} \right)^{-1} = a + \frac{b + c/(1-b')}{1 - (a'-b')/(1-b')} \times \left[1 - \frac{(d'-a')/(1-b')}{1 - (a'-b')/(1-b')} \right]^{-1}. \tag{31}$$

Expanding R_i of Eq. (21) to the 5th order, we have

$$R_i = \sum_{\neq i}^{n-1} {}_p \left[\frac{(ij)}{(j)} - \cdots - \frac{(ijklm) - (ijk)(lm) - (ij)(klm)}{(j)(k)(l)(m)} + \cdots \right] \times \left\{ \sum_{\neq i}^{n-1} {}_p \left[1 - \cdots - \frac{(jklm) - (jk)(lm)}{(j)(k)(l)(m)} + \cdots \right] \right\}^{-1}. \tag{32}$$

The numerator of R_i can be transformed into

$$N_i = \left(\sum_{ij}^* \frac{(ij)}{(j)} \right) \times \left(1 - \sum_{ijkl}^* \frac{(kl)}{(k)(1)} + \sum_{ijklm}^* \frac{(klm)}{(k)(l)(m)} - \cdots \right) - \left(\sum_{ijk}^* \frac{(ijk)}{(j)(k)} \right) \left(1 - \sum_{ijklm}^* \frac{(1m)}{(1)(m)} + \cdots \right) + \sum_{ijkl}^* \frac{(ijkl)}{(j)(k)(1)} - \sum_{ijklm}^* \frac{(ijklm)}{(j)(k)(l)(m)} + \cdots, \tag{33}$$

where $\sum_{ij}^* = \sum (j \neq i)$, and

$\sum_{ijklm}^* = \sum (j \neq i, k \neq ji, l \neq kji, m \neq lkji)$, etc.

Using the method of successive division and some formal expansions, such as

$$\sum\nolimits_p (jk) = \sum_{jk}^* (jk) + \sum_{jkl}^* (kl) + \cdots, \tag{34}$$

Eq. (30) can be transformed into a repeated fractional expansion similar to that given in Eq. (31). With a further transposition of the diagonal elements, the final form is given by

$$R_i = -\sum_{ij}^{*} \cfrac{(ij) + \sum_{ijk}^{*} \cfrac{(ijk) + \sum_{ijkl}^{*} \cfrac{(ijkl) + \sum_{ijklm}^{*} \cfrac{(ijklm) + \cdots}{[-(m) - \cdots]_{ijklm}}}{\left[-(l) - \sum_{ijklm}^{*} \cfrac{(lm) + \cdots}{(m) - \cdots}\right]_{ijkl}}}{\left[-(k) - \sum_{ijkl}^{*} \cfrac{(kl) + \sum_{ijklm}^{*} \cfrac{(klm) + \cdots}{-(m) - \cdots}}{-(l) - \sum_{ijklm}^{*} \cfrac{(lm) + \cdots}{-(m) - \cdots}}\right]_{ijk}}}{\left[-(j) - \sum_{ijk}^{*} \cfrac{(jk) + \sum_{ijkl}^{*} \cfrac{(jkl) + \sum_{ijklm}^{*} \cfrac{(jklm) + \cdots}{-(m) - \cdots}}{-(l) - \sum_{ijklm}^{*} \cfrac{(lm) + \cdots}{-(m) - \cdots}}}{-(k) - \sum_{ijkl}^{*} \cfrac{(kl) + \sum_{ijklm}^{*} \cfrac{(klm) + \cdots}{-(m) - \cdots}}{-(l) - \sum_{ijklm}^{*} \cfrac{(lm) + \cdots}{-(m) - \cdots}}}\right]_{ij}}. \tag{35}$$

This can also be obtained from Eq. (3) by Feshbach' method. Let $K_{ijk\cdots} \equiv [\ ]_{ijk\cdots}$, the following simple form is obtained:

$$R_i = -\sum_{ij}^{*} \cfrac{(ij) + \sum_{ijk}^{*} \cfrac{(ijk) + \sum_{ijkl}^{*} \cfrac{(ijkl) + \cdots}{K_{ijkl}}}{K_{ijk}}}{K_{ij}},$$

or

$$R_i = -\left[\sum_{ij}^{*} \frac{(ij)}{K_{ij}} + \sum_{ijk}^{*} \frac{(ijk)}{K_{ij}K_{ijk}} + \sum_{ijkl}^{*} \frac{(ijkl)}{K_{ij}K_{ijk}K_{ijkl}} + \cdots\right]. \tag{36}$$

Similarily, K_{ij}, the denominator of Eq. (35), can be expanded in terms of other $K_{ij\cdots}$'s:

$$K_{ij} = -\left[(j) + \sum_{ijk}^{*} \frac{(jk)}{K_{ijk}} + \sum_{ijkl}^{*} \frac{(jkl)}{K_{ijk}K_{ijkl}} + \cdots\right]. \tag{37}$$

By substituting Eq. (36) into Eq. (20), the Feenberg's expansion is obtained:

$$E_i = H_{ii} + \sum_{ij}^{*} \frac{H_{ij}H_{ji}}{E_i - \epsilon_{ij}} + \sum_{ijk}^{*} \frac{H_{ij}H_{jk}H_{ki}}{(E_i - \epsilon_{ij})(E_i - \epsilon_{ijk})} + \cdots, \tag{38}$$

with $K_{ijk\cdots} = (E_i - \epsilon_{ijk\cdots})$. The expansion of ϵ_{ij} given by Feenberg can also be shown to be the result of Eq. (37):

$$\epsilon_{ij} = H_{jj} + \sum_{ijk}^{*} \frac{H_{jk}H_{kj}}{E_i - \epsilon_{ijk}} + \sum_{ijkl}^{*} \frac{H_{jk}H_{kl}H_{lj}}{(E_i - \epsilon_{ijk})(E_i - \epsilon_{ijkl})} + \cdots. \tag{39}$$

By a similar argument, it can also be demonstrated that the equation for the amplitudes given by Feenberg and Eq. (26) are equivalent.

We would like to point out that Sasakawa's Green's function expansion has generated the same expansions given by Eqs. (15)–(17). However, owing to the nature of the iteration method, his elegant work did not make extensive use of the permutation symmetry of the expansion. It is obvious that the Sasakawa dispersion relation may be obtained from either Eq. (14) or Eq. (15). To demonstrate this, we use the following identity of r matrix elements H_{ij},

$$1 \equiv \sum_{i}^{r} \left[\prod_{j\neq i}^{r-1} \frac{(j)}{H_{jj} - H_{ii}}\right] \tag{40}$$

for the r-order term, if $H_{ij} \neq H_{kk}$ for all $j \neq k$. [We note that Eq. (40) is a $(r-1)$-degree polynomial equation in E_i with at least r distinct solutions $H_{ii}, H_{jj}, \cdots$.][13] If we take the 3rd-order term as an

[13] The author thanks Dr. George Trigg for this comment.

example, we use

$$1 \equiv \frac{(j)(k)}{(H_{jj} - H_{ii})(H_{kk} - H_{ii})} + \frac{(k)(i)}{(H_{kk} - H_{jj})(H_{ii} - H_{jj})} + \frac{(i)(j)}{(H_{ii} - H_{kk})(H_{jj} - H_{kk})}. \quad (41)$$

Substituting Eq. (41) in the 3rd-order term of Eq. (15) with Sasakawa's notation $\omega_i = H_{ii}$ and $g_{ijk} = \sum_p^3 (ijk)$, we have

$$\sum_p^n \frac{(ijk)}{(i)(j)(k)} = \sum_p^n \frac{(ijk)}{(i)(j)(k)} \sum_{i\neq jk}^{3} \frac{(j)(k)}{(\omega_j - \omega_i)(\omega_k - \omega_i)} = \sum_i^n \frac{1}{(i)} \sum_{j\neq i} \sum_{\substack{k>j \\ \neq i}} \frac{g_{ijk}}{(\omega_j - \omega_i)(\omega_k - \omega_i)}. \quad (42)$$

In spite of its simplicity, the result of the present paper represents a more general form for the solutions of a secular equation and its associated eigenfunctions. Its application to practical problems should have a wider validity. The advantage of R_i and a_{ij}/a_{ii} expressed in the form of a ratio makes it possible for them to work even in some limiting cases.

VII. EXAMPLE

In this section, two simple examples selected from Sasakawa's first paper (S-I) will be used to demonstrate the practical application of the eigenvalue formulas given in this note. However, we must realize that the degree of usefulness of different methods should not be justified only by few examples. For a particular problem, one method may have certain advantages over the others, especially when approximation is used.

Example 1: The Ground-State Energy of Mathieu Equation. The matrix elements of the Mathieu equation are all zero except

$$H_{ii} = n^2 + \tfrac{1}{2}s, \qquad H_{02} = H_{20} = (8)^{-\frac{1}{2}}s, \qquad H_{ii+2} = H_{i+2i} = \tfrac{1}{4}s, \quad \text{for all } i \neq 0. \quad (43)$$

It follows that any cyclic product with more than two indices (elements) vanishes. Now Eq. (21) or (23) takes the simple form

$$R_i = \sum_{\neq i}^{\infty p} \left[\frac{(ij)}{(j)} - \frac{(ij)(kl)}{(j)(k)(l)} + \frac{(ij)(kl)(mr)}{(j)(k)(l)(m)(r)} - \cdots \right] \times \left\{ \sum_{\neq i}^{\infty p} \left[1 - \frac{(jk)}{(j)(k)} + \frac{(jk)(lm)}{(j)(k)(l)(m)} - \cdots \right] \right\}^{-1}. \quad (44)$$

For ground state $i = 0$, Eq. (44) can be written as

$$R_0 = \frac{(02)}{(2)} \sum_{\neq 0,2}^{\infty p} \left[1 - \frac{(kl)}{(k)(l)} + \frac{(kl)(mr)}{(k)(l)(m)(r)} - \cdots \right] \times \left\{ \sum_{\neq 0,2}^{\infty p} \left[1 - \frac{(kl)}{(k)(l)} + \cdots \right] - \frac{(24)}{(2)(4)} \sum_{\neq 0,2,4}^{\infty p} \left[1 - \frac{(lm)}{(l)(m)} + \cdots \right] \right\}^{-1}.$$

Dividing both the numerator and the denominator by the sum in the denominator and continuing this process of reduction, Eq. (44) can be transformed into a infinite continued fraction as

$$R_0 = \cfrac{(02)/(2)}{1 - \cfrac{(24)/[(2)(4)]}{1 - \cfrac{(46)/[(4)(6)]}{1 - \ddots}}} \quad (45)$$

It is interesting to see that the elements with odd number disappear if a finite number of fractions are used. However, we must note that it is not necessarily true when Eq. (44) is used directly. For $s = 4$, Eq. (45) becomes

$$R_0 = \cfrac{2/(2^2 + R_0)}{1 - \cfrac{1/[(2^2 + R_0)(4^2 + R_0)]}{1 - \cfrac{1/[(4^2 + R_0)(6^2 + R_0)]}{1 - 1/[(6^2 + R_0)(8^2 + R_0)]}}}, \quad (46)$$

with an approximation up to $n = 8$. The radius of convergence of R_0, ΔR_0 can be estimated by calculating how large a correction would be if the next-higher-order number were included. Therefore,

$$\Delta R_0/R_0 \simeq \pm 1/[(6^2+R_0)(8^2+R_0)(8^2+R_0)(10^2+R_0)] \simeq \pm 7 \times 10^{-8}. \quad (47)$$

Now R_0 can be solved by iteration method as described in Sec. III. The energy can be obtained by $E_0 = 2 - R_0$. Starting with ${}_{(0)}R_0 = 0$, the results of E_0 corresponding to each successive iteration are tabulated as follows, together with the results obtained previously by other methods.

Method		(1) Brillouin Wigner (Refs. 6 and 10)	(2) Feenberg (Ref. 6)	(3) Sasakawa (Ref. 6)	(4) Eq. (46)
order of iteration or approximation	0	2.00000	2.00000	2.000000	2.000000000
	1	1.00000	1.268	1.549077	1.492049464
	2	1.77778	1.55051	1.544742	1.550286204
	3	1.15407	1.54429	1.544704	1.544296736
	4		1.54487	1.544703	1.544920092
	5				1.544855297
	6				1.544862031
	7				1.544861331
	8				1.544861404
	9				1.544861396
possible error					±0.000000035
exact solution			1.54486 (Ref. 10)		

It is not surprising that the result obtained by method (4) agrees perfectly with the exact solution, because Eq. (46) is equivalent to the continued fraction formula which has been used to obtain the exact solution (see pp. 564 and 1018 of Ref. 10). Although a small discrepancy is found between the results of method (3) and (4), yet, mathematically, there is no reason to believe that the dispersion relation formula and Eq. (20) are not equivalent for the nondegenerate case.

Example 2: The Nilsson's Problem of the Degenerate Case. Now we take the example V.4 from S-I to show that Eq. (21) or (23) can also be used directly for the degenerate case. The matrix elements of this 6×6 secular equation are given by

$$H_{11} = H_{22} = H_{33} = 0, \quad H_{44} = H_{55} = H_{66} = -1$$

$$H_{14} = H_{41} = -(30)^{-\frac{1}{2}}, \quad H_{25} = H_{52} = -2(3)^{\frac{1}{2}},$$

$$H_{36} = H_{63} = -(2)^{\frac{1}{2}}, \tag{48}$$

with all others equal to zero. By direct substitution in Eq. (21) for $i = 1$, we have

$$R_1 = \frac{\frac{(14)}{(4)} - \frac{(14)(25)}{(4)(2)(5)} - \frac{(14)(36)}{(4)(3)(6)} + \frac{(14)(25)(36)}{(4)(2)(5)(3)(6)}}{1 - \frac{(25)}{(2)(5)} - \frac{(36)}{(3)(6)} + \frac{(25)(36)}{(2)(5)(3)(6)}}$$

$$= \frac{(14)}{(4)}. \tag{49}$$

Similarly we have

$$R_1 = 30/(-1 + R_1), \quad R_4 = 30/(1 + R_4),$$

$$R_2 = 12/(-1 + R_2), \quad R_5 = 12/(1 + R_5), \tag{50}$$

$$R_3 = 2/(-1 + R_3), \quad R_6 = 2/(1 + R_6).$$

Now the only problem left is to solve the six simple quadratic equations. Although there are totally twelve solutions of R_i obtainable from Eq. (50), yet, when they are substituted in the energy equation $E_i = H_{ii} - R_i$, only six value of E_i are obtained. They are

$$E_1 = 5, \quad E_2 = 3, \quad E_3 = 1,$$

$$E_4 = -6, \quad E_5 = -4, \quad E_6 = -2,$$

which agree perfectly with the results given in S-I. The identification of each solution with its original state is achieved simply by comparing the sign of each solution R_i with that obtained from Eq. (50) by setting the value of R_i on the right-hand side of each equation equal to zero. The same solutions can be obtained directly from Eq. (21), but with more labor, by iteration method as suggested in Ref. 11. Naturally, with this method, there is no need for the identification of solutions with its original state.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Joseph Chen for his very helpful discussion and suggestions and Dr. Martin Blume for his encouragement. It has also been a great benefit that both of them have called the author's attention to the works of Feenberg, Feshbach, and Sasakawa. He also wishes to acknowledge the Atomic Beam Group of the University of California at Berkeley where this work was first begun.

Distribution of Zeros of the Grand Partition Function

P. C. HEMMER AND E. HIIS HAUGE
Institutt for Teoretisk Fysikk
AND
J. O. AASEN
Institutt for Anvendt Matematikk, Norges Tekniske Høgskole, Trondheim, Norway
(Received 5 April 1965)

The distribution of zeros of the grand partition function is calculated in the thermodynamic limit for a class of one-dimensional gas models in two ways: (1) from the equation of state and (2) directly from the partition function. In this way one obtains (for these cases) a verification of the assumptions we had to make in order to associate a unique distribution of zeros with a given equation of state. In the Appendix we present some numerical evidence for the validity of these assumptions also in the case of the van der Waals gas.

1. INTRODUCTION

IN the Yang–Lee condensation theory[1] the equation of state is discussed in terms of the (in general complex) zeros of the grand partition function $Z_g(z; V, T)$, regarded as function of the fugacity z. The discussion is particularly easy if one with Yang and Lee assumes the intermolecular potential to contain a hard repulsive core, making the grand partition function a polynomial in z with positive coefficients.

The grand partition function is defined in terms of the intermolecular potential $\phi(r)$, while the equation of state is implied by the grand partition function. One may therefore say that there are two ways to explore the properties of the distribution of zeros in the complex z plane:

1. One can investigate the relation between the zeros of the polynomial Z_g and its coefficients which are complicated integrals involving the pair potential $\phi(r)$. In the general case one cannot of course hope to determine the complete distribution, but it might nevertheless be possible to say something about its location.[2]

2. One may try to infer the zero distribution from the equation of state. This problem is easier, since in the quantities involved the thermodynamical limit is already taken. It has, however, no *unique* solution.

We study below these questions on a class of interacting systems, viz. one-dimensional gas models with repulsive forces. In Sec. 3 we start with the equation of state and determine a distribution consistent with it. In view of the fact that the solution in principle is not unique it is instructive to see how one is led to one specific distribution. In Sec. 4 the more difficult problem of a direct calculation of the roots of the grand partition function is solved, and the results obtained in Sec. 3 are verified. In all cases considered the zeros are found to be distributed on the negative real axis.

By letting the lattice spacing go to zero and simultaneously extending the repulsive potential to more and more lattice sites, we obtain as a limiting case a continuum gas model with hard-core interaction.[3]

The relation between the equation of state and the zero distribution for a van der Waals gas has been studied previously,[4] and Appendix B contains some numerical evidence in support of the conclusions reached in that article.

2. MODEL

Consider a lattice gas of n atoms distributed on a one-dimensional lattice with lattice spacing δ. The interaction potential $\phi(r)$ between each pair of particles is

$$\phi(r_{ij}) = \begin{cases} +\infty & \text{if } r_{ij} < d = m\delta \\ 0 & \text{if } r_{ij} \geq d, \end{cases} \tag{1}$$

so that the hard core extends over m lattice sites. The total number of lattice sites is

$$N = L/\delta, \tag{2}$$

where L is the total length of the system. We have thus for $m = 1, 2, \cdots$, a sequence of gas models, and

[1] C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 (1952).
[2] A very good example is Yang and Lee's proof that for lattice gases with attraction the zeros are all on a circle around the origin.
[3] The hard-rod gas has been studied previously. E. Hiis Hauge and P. C. Hemmer, Physica 29, 1338 (1963).
[4] P. C. Hemmer and E. Hiis Hauge, Phys. Rev. 133, A1010 (1963).

for $m \to \infty$, $\delta \to 0$, $N \to \infty$ while L and d are fixed, a continuum gas of hard rods results.

The equation of state is obtained from the configurational partition function Q_n:

$$p/kT = (\partial/\partial L) \ln Q_n. \tag{3}$$

The partition function in this case is simply

$$Q_n = \delta^n \binom{N - mn + n}{n} = \delta^n \frac{(N - mn + n)!}{n!\,(N - mn)!}, \tag{4}$$

the binomial coefficient counting the number of possible configurations. In the thermodynamic limit $n \to \infty$, $L \to \infty$, $\rho = n/L$ = constant, one obtains from (3) and (4) the equation of state[5]

$$\begin{aligned} p &= \frac{kT}{\delta} \ln\left(1 + \frac{\rho\delta}{1 - \rho m\delta}\right) \\ &= \frac{kTm}{d} \ln\left(1 + \frac{\rho d/m}{1 - \rho d}\right). \end{aligned} \tag{5}$$

Tonks' equation of state for hard rods,

$$p = kT\rho/(1 - \rho d),$$

emerges of course in the limit $m \to \infty$.

3. THE DISTRIBUTION OF ZEROS FROM THE EQUATION OF STATE

The quantity

$$\chi(z, T) = \lim_{L\to\infty} L^{-1} \ln Z_g(z; L, T) \tag{6}$$

is related to the equation of state by the well-known Mayer equations

$$p/kT = \chi(z) \tag{7}$$

$$\rho = z\chi'(z). \tag{8}$$

Inserting for the pressure from (5), we obtain

$$\rho d = 1 - [m \exp(\chi d/m) - m + 1]^{-1}, \tag{9}$$

and by integration of (8)

$$zd/m = e^{\chi d} - e^{(1-m^{-1})\chi d}. \tag{10}$$

We *assume* that $\chi(z)$ for complex values of z is determined by an analytic continuation out from the real axis. Together with the assumption that the zeros of Z_g coalesce into lines C as $V \to \infty$, this guarantees a unique solution. By (6) $\chi(z)$ is the complex logarithmic potential of the zero distribution $g(s)$:

$$\chi(z) = \int_C g(s) \ln\left(1 - \frac{z}{z(s)}\right) ds. \tag{11}$$

In analogy with electrostatics, continuity of the real part $\Phi(z)$ of $\chi = \Phi + i\Psi$ determines the position of the charges, while the discontinuities in the imaginary part Ψ determine the value of the charge density $g(s)$.

The inverse function $\chi(z)$, as defined by (10), is multivalued with branch point z_h given by $dz/d\chi = 0$, yielding

$$z_h = -(1/d)(1 - 1/m)^{m-1}. \tag{12}$$

It is a mild convenience to assume $d = 1$ from now on. With a cut from z_h to infinity along the negative real axis [since $g(z)$ has to be symmetric about the real axis], the analytic continuation of χ is everywhere uniquely defined. On both sides of the cut we find from (10) by inserting $\chi = \Phi + i\Psi$

$$x = -\frac{m \sin(\Psi/m)}{\sin\Psi}\left[\frac{\sin[(1 - m^{-1})\Psi]}{\sin\Psi}\right]^{m-1}. \tag{13}$$

Above the cut $0 \le \Psi \le \pi$, below $-\pi \le \Psi \le 0$. The jump in Ψ across the cut determines the zero density

$$g_m(x) = -(1/\pi)(\partial\Psi/dx), \tag{14}$$

where Ψ is now restricted to $(0, \pi)$. It is seen by comparison between (12) and (13) that $\Psi = 0$ corresponds to the branch point z_h and that $x \to -\infty$ for $\Psi \to \pi$.

Hence, the resulting zero distribution (14) is always along the negative real axis and is sketched for $m = 1$, 2, and ∞ in Fig. 1. It is apparent that already nearest-neighbor repulsion yields a good approximation to the continuum result.

In general it is not possible to eliminate the parameter Ψ. Exceptions are

$$g_1(x) = \delta(x + 1) \tag{15}$$

$$g_2(x) = [\pi\,|x|\,(2\,|x| - 1)^{\frac{1}{2}}]^{-1}, \qquad x < -\tfrac{1}{2}. \tag{16}$$

The limiting distribution $g_\infty(x)$ is identical with the one previously found.[3]

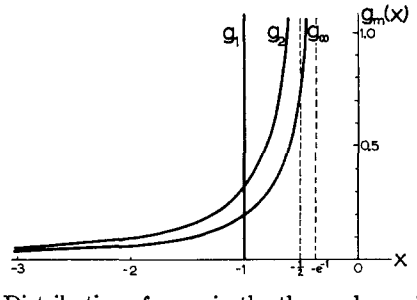

FIG. 1. Distribution of zeros in the thermodynamic limit.

4. THE ZEROS OF THE GRAND PARTITION FUNCTION

The grand partition function is known exactly in this case (with $d = 1$):

[5] T. D. Lee and C. N. Yang, Phys. Rev. 87, 410 (1952).

$$Z_g^{(m)}(z) = \sum_n Q_n z^n = \sum_{n=0}^{[N/m]} \binom{N - mn + n}{n} \left(\frac{z}{m}\right)^n, \quad (17)$$

using (4). The case $m = 1$ is trivial:

$$Z_g^{(1)}(z) = (1 + z)^N, \quad (18)$$

in accordance with (15). The case $m = 2$ can also be done exactly:

$$\begin{aligned} Z_g^{(2)}(z) &= \sum_n \binom{N - n}{n} \left(\frac{z}{2}\right)^n \\ &= \frac{1}{2\pi i} \sum_{n=0}^{\infty} \left(\frac{z}{2}\right)^n \int_{(0+)} \frac{(1 + t)^N}{[(1 + t)t]^n} \frac{dt}{t}. \end{aligned} \quad (19)$$

Assuming $|z| < |2(1 + t)t|$, summation yields

$$\begin{aligned} Z_g^{(2)}(z) &= \frac{1}{\pi i} \oint \frac{dt(1 + t)^{N+1}}{2(1 + t)t - z} \\ &= 2^{-N-1}(1 + 2z)^{-\frac{1}{2}}\{[1 + (1 + 2z)^{\frac{1}{2}}]^{N+1} \\ &\qquad - [1 - (1 + 2z)^{\frac{1}{2}}]^{N+1}\}, \end{aligned} \quad (20)$$

since the path of integration encloses both poles. The right-hand side is essentially a Chebychev polynomial of the second kind. The fraction of the zeros

$$z_k = -\{\cos [2\pi k/(N + 1)] + 1\}^{-1}, \qquad k = 1, 2, 3, \cdots [\tfrac{1}{2}N] \quad (21)$$

less than x equals (when $N \to \infty$)

$$G(x) = \begin{cases} \pi^{-1} \operatorname{Arc\,cos} (1 + x^{-1}) & x \leq -\frac{1}{2}, \\ 1 & x \geq -\frac{1}{2}, \end{cases} \quad (22)$$

which is equivalent to the previously found distribution (16).

In the general case we again write the grand partition function as a contour integral:

$$\begin{aligned} Z_g^{(m)}(z) &= \frac{1}{2\pi i} \sum_{n=0}^{\infty} \left(\frac{z}{m}\right)^n \int_{(0+)} \frac{(1 + t)^N}{[(1 + t)^{m-1}t]^n} \frac{dt}{t} \\ &= \frac{m}{2\pi i} \oint \frac{dt(1 + t)^{N+m-1}}{mt(1 + t)^{m-1} - z}, \end{aligned} \quad (23)$$

where the path of integration must satisfy

$$|z| > m\,|t(1 + t)^{m-1}|.$$

Having the results of Sec. 3 in mind, we first try to find the *real* zeros z_i less than z_h, given by (12). In this region we prefer to use Ψ as variable instead of z, see Eq. (13). In addition we change the integration variable from t to

$$v = (1 + t) \cdot \frac{\sin \Psi}{\sin [(1 - m^{-1})\Psi]}. \quad (24)$$

In this way the polynomial (23) is expressed as

$$Z_g^{(m)}(z) = I_m \cdot \sin^{N+1} [(1 - m^{-1})\Psi]/\sin^N \Psi, \quad (25)$$

where

$$I_m = \frac{1}{2\pi i} \oint \frac{dv\, v^{N+m-1}}{v^m \sin [(1 - m^{-1})\Psi] - v^{m-1} \sin \Psi + \sin (\Psi/m)}. \quad (26)$$

The contour of integration encloses all m poles. Where are these poles? Two of them,

$$v_{1,2} = e^{\pm i\Psi/m}, \quad (27)$$

are *on* the unit circle in the v plane, as is immediately verified by insertion. In Appendix A is proven that all the other poles are located *inside* the unit circle,

$$|v_i| < 1, \qquad i = 3, 4, \cdots, m. \quad (28)$$

The integral I_m equals the sum of the residues at the poles $v_1, v_2, \cdots v_m$:

$$I_m = \sum_{i=1}^{m} \operatorname{res} (v_i). \quad (29)$$

For $N \to \infty$, which is the case in which we really are interested, all residues except the two first ones vanish because of the factor v_i^N. Thus, for a large system

$$\begin{aligned} I_m &\cong \operatorname{res} (v_1) + \operatorname{res} (v_2) \\ &= M^{-1}[e^{i(N_1\Psi + \delta - \frac{1}{2}\pi)} + e^{-i(N_1\Psi + \delta - \frac{1}{2}\pi)}], \end{aligned} \quad (30)$$

with the abbreviations

N_1 = the maximum number of particles = N/m (assumed integer), (31)

$$M = |m \sin [(1 - m^{-1})\Psi] - (m - 1) \sin \Psi\, e^{-i\Psi/m}|, \quad (32)$$

and

$$\tan \delta = [\cot (\Psi/m) - m \cot \Psi]/(m - 1). \quad (33)$$

For large N, therefore, the zeros Ψ_k of I_m, and hence of Z_g, approach the solutions of

$$N_1\Psi + \delta(\Psi) = k\pi, \qquad k = 1, 2, \cdots N_1.$$

Since $0 \leq \delta(\Psi) < \frac{1}{2}\pi$ one easily sees that the zeros satisfy

$$(k - \tfrac{1}{2})(\pi/N_1) < \Psi_k \leq k(\pi/N_1), \quad k = 1, 2, \cdots N_1. \quad (34)$$

Since the degree of the original polynomial Z_g was N_1, the N_1 zeros (34) are *all* zeros.

Since z decreases when Ψ increases [see Eq. (13)], the zeros z less than a certain value z are obtained for $[N_1\Psi(z)/\pi] < k \leq N_1$. That is, the relative num-

ber of zeros $<z$ equals $1 - \pi^{-1}\Psi(z)$. By differentiation this corresponds to the following density of zeros:

$$g(z) = -\pi^{-1}[\partial\Psi(z)/\partial z]. \tag{35}$$

By comparison with Eq. (14), one sees that we now by rigorous methods have verified the distribution determined in Sec. 3. Since the method of Sec. 3 in general is easier to apply, the present verification is encouraging.

APPENDIX A

We will prove that all zeros of the polynomial

$$N(v) = v^m \sin[(1 - m^{-1})\Psi] - v^{m-1}\sin\Psi + \sin(\Psi/m) \tag{36}$$

are inside the unit circle, with the exception of the two zeros

$$v_{1,2} = \exp(\pm i\Psi/m). \tag{37}$$

The parameter Ψ may have any value in the interval

$$0 < \Psi < \pi. \tag{38}$$

Factoring out the two zeros (37) we get

$$P(v) = \frac{N(v)}{(v - v_1)(v - v_2)} = \sum_{k=1}^{m-1} \sin(k\Psi/m)v^{k-1}. \tag{39}$$

The first part of the proof consists of showing that $P(v)$ has no zeros *on* the unit circle. Assume that $v = e^{i\phi}$ is such a zero. Then

$$2ivP(v) = \sum_{k=1}^{m-1} e^{i(\phi+\Psi/m)k} - \sum_{k=1}^{m-1} e^{i(\phi-\Psi/m)k} = 0. \tag{40}$$

First note that any sum of unit vectors $\sum_{k=1}^{r} e^{ik\alpha}$ lies on a circle through the origin, centered at $[-\frac{1}{2}, \frac{1}{2}\cot(\frac{1}{2}\alpha)]$. Two such circles corresponding to different values of α have only the points 0 and -1 in common. Therefore, both sums in Eq. (40) must either vanish or else be equal to -1. In the first case ϕ satisfies

$$\phi \pm (\Psi/m) = 2k_{\pm}\pi/(m - 1) \tag{41}$$

with integers k_+, k_-. By elimination of ϕ we get

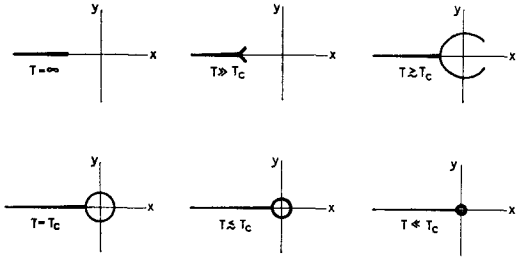

FIG. 2. Sketch of the van der Waals zero distribution at different temperatures (taken from Ref. 4).

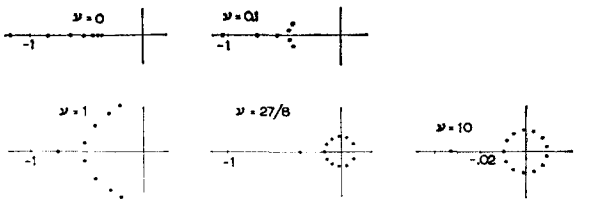

FIG. 3. Zeros of the van der Waals grand partition function, Eq. (46), for $V/d = 15$. In the figure d has been used as the unit of length. (Some negative real zeros of large modulus are not shown.)

$$\Psi = (m/(m - 1))(k_+ - k_-)\pi. \tag{42}$$

In the second case ϕ must satisfy

$$\phi_{\pm}(\Psi/m) = 2k_{\pm}\pi/m, \tag{43}$$

in which case (42) is replaced by

$$\Psi = (k_+ - k_-)\pi. \tag{44}$$

Since both (42) and (44) are in disagreement with Eq. (38) we conclude that no such pole of modulus one exists.

For the second part of the proof, we note the coefficients of $P(v)$, Eq. (39), are real, positive and increasing with m for small values of Ψ. The Enstrøm theorem[6] tells us that in this case the zeros are all inside the unit circle. Increasing Ψ, the zeros must stay inside the unit circle. This follows from continuity and from the first part of the proof.

APPENDIX B

Reference 4 contains a study of the properties of the zero distribution for a gas obeying van der Waals' equation

$$p/kT = (v - d)^{-1} - \nu v^{-2}, \qquad \nu \equiv a/kT \tag{45}$$

supplemented with the Maxwell equal area construction, based upon the same assumption as the calculation in the present Sec. 3. Qualitatively the limiting distribution was found to fall into the pattern of Fig. 2 (taken from Ref. 4), and only in limiting cases were quantitative results obtained.

We found it of interest to see how well these results agrees with a direct numerical calculation of the zeros of the following grand partition function

$$Z_g = \sum_{N=1}^{V/d} \frac{Q_N}{N!} z^N, \tag{46}$$

$$Q_N = (V - Nd)^N \exp\left[\frac{2\nu}{V}\binom{N}{2}\right], \tag{47}$$

for a finite number of particles. The canonical partition function (47) implies the van der Waals equation (45), and it is well known that the pressure from the grand canonical ensemble has a horizontal

[6] G. Enestrøm, Tôhoku Math. J. **18**, 34 (1920).

part instead of a wiggle for subcritical temperatures. Moreover, one can, by physical arguments, make the form (47) plausible as the partition function for one-dimensional molecules interacting with a hard core d plus a very weak long-range attraction, a model known to rigorously obey van der Waals' equation of state.[7]

[7] See M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963), especially the discussion in the concluding remarks.

For a maximum number of particles $V/d = 15$, the calculated zeros of the grand partition function (46) are shown in Fig. 4. The limited accuracy (8 digits) of the computer made calculations with a greater number of particles unreliable.

Qualitatively the distribution is as one would expect from the predictions of Ref. 4. The quantitative details that can be checked are in as good agreement with the calculated limited distribution as one could expect for such a small number of particles.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 1 JANUARY 1966

Approximation by Sequences of Padé Approximants in Regions of Meromorphy*

J. S. R. CHISHOLM†
University of Kent, Canterbury, England, and Physics Department, Texas A. and M., College Station, Texas

It is shown that sequences $\{P_{mn}(z)\}$ of Padé approximants with $m \to \infty$ and $n \to \infty$ approximate a function $f(z)$ inside a certain circle of meromorphy with centre at $z = 0$. Extensions of this result are discussed.

1. INTRODUCTION

THE value of continued fractions and Padé approximants in interpreting divergent series has been appreciated for a long time. The classical work of Wall[1] gives a full account of the theory of continued fractions; in a recent review paper,[2] Baker has summarized the main mathematical results to date, and has given an account of the applications of the Padé method, in particular to perturbation series. Many of the established convergence properties of Padé approximants are limited to series of Stieltjes, but Baker, Gammel, and Wills[3] examined the approximants (in particular the diagonal approximants) of many different functions, and conjectured that the convergence properties of Padé approximants were far wider than those encompassed by existing mathematical theory. Baker[4] has proved that an infinite sequence of approximants $\{P_{mn}(z)\}$, with $m + n \to \infty$, to a function $f(z)$ converges to $f(z)$ over certain regions of the complex z-plane in which the sequence is uniformly bounded. The assumption of uniform boundedness needs to be translated into more specific conditions, for example a condition ensuring that the approximants have sufficient poles and zeros to represent those of the function $f(z)$. A further question that remains to be answered is: "where are the poles of the approximants $f(z)$?" The studies of Baker, Gammel, and Wills[3] have shown that in regions of meromorphy of $f(z)$, the poles and zero of $P_{mn}(z)$ normally lie near to those of $f(z)$, but that on occasion an approximant may have other isolated "spurious poles" in the region, each accompanied by a nearby zero. It seems that there is in fact no meromorphic region within which the number of poles of $P_{mn}(z)$ is unbounded; this conjecture is an important part of the Padé problem, and is as yet unproved.

This paper shows that it is possible to establish the convergence of certain sequences of Padé approximants throughout certain sub-regions of a region of meromorphy, without assuming uniform boundedness. We have, however, to assume boundedness of the number of poles in the region. It has not been possible to justify this assumption for

* Research sponsored in part by the European Office of Aerospace Research, United States Air Force, Grant No. AF EOAR 64-61.
† National Science Foundation Senior Foreign Scientist.
[1] H. S. Wall, *Continued Fractions* (D. Van Nostrand Princeton, New Jersey, 1947).
[2] G. A. Baker, Jr., Advan. Theoret. Phys. 1, 5 (1965).
[3] G. A. Baker, Jr., J. L. Gammel, and J. G. Wills, J. Math. Anal. Appl. 2, 405 (1961).
[4] Reference 2, Theorem 4.

part instead of a wiggle for subcritical temperatures. Moreover, one can, by physical arguments, make the form (47) plausible as the partition function for one-dimensional molecules interacting with a hard core d plus a very weak long-range attraction, a model known to rigorously obey van der Waals' equation of state.[7]

[7] See M. Kac, G. E. Uhlenbeck, and P. C. Hemmer, J. Math. Phys. 4, 216 (1963), especially the discussion in the concluding remarks.

For a maximum number of particles $V/d = 15$, the calculated zeros of the grand partition function (46) are shown in Fig. 4. The limited accuracy (8 digits) of the computer made calculations with a greater number of particles unreliable.

Qualitatively the distribution is as one would expect from the predictions of Ref. 4. The quantitative details that can be checked are in as good agreement with the calculated limited distribution as one could expect for such a small number of particles.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 1 JANUARY 1966

Approximation by Sequences of Padé Approximants in Regions of Meromorphy*

J. S. R. CHISHOLM†
University of Kent, Canterbury, England, and Physics Department, Texas A. and M., College Station, Texas

It is shown that sequences $\{P_{mn}(z)\}$ of Padé approximants with $m \to \infty$ and $n \to \infty$ approximate a function $f(z)$ inside a certain circle of meromorphy with centre at $z = 0$. Extensions of this result are discussed.

1. INTRODUCTION

THE value of continued fractions and Padé approximants in interpreting divergent series has been appreciated for a long time. The classical work of Wall[1] gives a full account of the theory of continued fractions; in a recent review paper,[2] Baker has summarized the main mathematical results to date, and has given an account of the applications of the Padé method, in particular to perturbation series. Many of the established convergence properties of Padé approximants are limited to series of Stieltjes, but Baker, Gammel, and Wills[3] examined the approximants (in particular the diagonal approximants) of many different functions, and conjectured that the convergence properties of Padé approximants were far wider than those encompassed by existing mathematical theory. Baker[4] has proved that an infinite sequence of approximants $\{P_{mn}(z)\}$, with $m + n \to \infty$, to a function $f(z)$ converges to $f(z)$ over certain regions of the complex z-plane in which the sequence is uniformly bounded. The assumption of uniform boundedness needs to be translated into more specific conditions, for example a condition ensuring that the approximants have sufficient poles and zeros to represent those of the function $f(z)$. A further question that remains to be answered is: "where are the poles of the approximants $f(z)$?" The studies of Baker, Gammel, and Wills[3] have shown that in regions of meromorphy of $f(z)$, the poles and zero of $P_{mn}(z)$ normally lie near to those of $f(z)$, but that on occasion an approximant may have other isolated "spurious poles" in the region, each accompanied by a nearby zero. It seems that there is in fact no meromorphic region within which the number of poles of $P_{mn}(z)$ is unbounded; this conjecture is an important part of the Padé problem, and is as yet unproved.

This paper shows that it is possible to establish the convergence of certain sequences of Padé approximants throughout certain sub-regions of a region of meromorphy, without assuming uniform boundedness. We have, however, to assume boundedness of the number of poles in the region. It has not been possible to justify this assumption for

* Research sponsored in part by the European Office of Aerospace Research, United States Air Force, Grant No. AF EOAR 64-61.
† National Science Foundation Senior Foreign Scientist.
[1] H. S. Wall, *Continued Fractions* (D. Van Nostrand Princeton, New Jersey, 1947).
[2] G. A. Baker, Jr., Advan. Theoret. Phys. 1, 5 (1965).
[3] G. A. Baker, Jr., J. L. Gammel, and J. G. Wills, J. Math. Anal. Appl. 2, 405 (1961).
[4] Reference 2, Theorem 4.

any particular sequence, but it will be shown that spurious poles of an approximant in the sub-regions of meromorphy are compensated by a nearby zero.

2. PRELIMINARY DEFINITIONS AND RESULTS

The invariance properties of Padé approximants established by Baker, Gammel, and Wills ensure that convergence properties of sequences $\{P_{m\ m+k}\}$ (k fixed) of Padé approximants to a function $f(z)$ which are established inside a circle of meromorphy, with center at $z = 0$, can be extended over large regions of the complex z-plane. The essential result of this paper is a general convergence theorem within a sub-region inside the circle of meromorphy. Consider a function $f(z)$ which is meromorphic for $|z| < R$. Then if ρ is any number less than R, with $f(z)$ regular on $|z| = \rho$, then $f(z)$ can be expressed as

$$f(z) = \frac{\sum_0^\infty c_t z^t}{\prod_{i=1}^{l} (1 - z/\zeta_i)^{k_i}} \equiv \frac{\sum_0^\infty c_t z^t}{\sum_0^M e_u z^u}, \tag{2.1}$$

where $\sum_0^\infty c_t z^t$ is uniformly convergent at any point in the region $\mathbf{C}$ defined by $|z| \leq \rho$, and where we assume that none of the poles ζ_i coincides with the origin. We note that the total order of poles is

$$M = \sum_i k_i \tag{2.2}$$

and that $e_0 = 1$. For convenience, we normalize the function $f(z)$ by choosing

$$f(0) = c_0 = 1; \tag{2.3}$$

the results that we establish can be immediately extended to any function of form (2.1) by multiplying by a constant.

We can expand the denominator in (2.1) by the binomial theorem to give a power series expansion

$$f(z) = \sum_0^\infty d_t z^t \qquad (d_0 = 1); \tag{2.4}$$

then

$$\left[\sum_0^M e_u z^u\right]\left[\sum_0^\infty d_t z^t\right] = \sum_0^\infty c_t z^t. \tag{2.5}$$

Assuming that ζ_1 is one of the poles nearest to the origin and defining

$$\rho_i = |\zeta_i|, \tag{2.6}$$

the series (2.4) is convergent for $|z| < \rho_1$, and (2.5) is an identity between analytic functions. For $|z| > \rho_1$ the series (2.4) is divergent, but the sequence of coefficients $\{d_t\}$ is well-defined, independent of the value of z.

A Padé approximant

$$P_{mn}(z) = \left(\sum_0^n a_r z^r\right)\left(\sum_0^m b_s z^s\right)^{-1} \tag{2.7}$$

of $f(z)$ is defined by the identity

$$\left[\sum_0^m b_s z^s\right]\left[\sum_0^\infty d_t z^t\right] = \sum_0^n a_r z^r + O(z^{m+n+1}). \tag{2.8}$$

The identity has precise analytic meaning when $|z| < \rho_1$. For other values of z, (2.8) is a formal identity, the coefficients $\{a_r\}$ and $\{b_s\}$ being defined by the linear equations

$$\sum_{s=0}^{\min(r,m)} b_s d_{r-s} = a_r \qquad (r = 0, 1, \cdots, n), \tag{2.9}$$

$$\sum_{s=0}^{\min(r,m)} b_s d_{r-s} = 0 \qquad \times (r = n+1, \cdots, m+n). \tag{2.10}$$

The first of Eqs. (2.9) is $b_0 d_0 = a_0$, or $b_0 = a_0$. We choose $a_0 = b_0 = 1$. Then Eqs. (2.10) are m equations for $b_1, \cdots, b_m$ and determine them uniquely if the determinant of coefficients Δ_b is nonzero. Equations (2.9) then define $a_1, \cdots, a_n$, so that P_{mn} is uniquely determined. We note that $P_{mn}(z)$ is then the unique rational fraction of form (2.7) whose power series expansion agrees with (2.4) up to $O(z^{m+n})$.

It may happen for a particular function $f(z)$ and integer pair (m, n) that $\Delta_b = 0$. In general (2.10) will have no finite solution and we cannot define P_{mn}. When we are discussing a definite sequence of approximants, we assume that it does not contain any which cannot be uniquely defined.

It is also possible that $\Delta_b = 0$ and that Eqs. (2.10) have an infinity of solutions. We shall now discuss a function $f_0(z)$ for which this happens, when the numerator in (2.1) is a polynomial in z of degree N:

$$f_0(z) = \frac{\sum_0^N c_t z^t}{\sum_0^M e_u z^u} \equiv \frac{\sum_0^N c_t z^t}{\prod_i (1 - z/\zeta_i)^{k_i}}. \tag{2.11}$$

It is assumed that the numerator and denominator in (2.11) have no common factors. If by chance they have a common factor, the results established here are are not substantially altered.

If the expansion of (2.11) corresponding to (2.4) is

$$f_0(z) = \sum_0^\infty d_{0t} z^t, \tag{2.12}$$

then

$$\left[\sum_0^M e_u z^u\right]\left[\sum_0^\infty d_{0t} z^t\right] = \sum_0^N c_t z^t. \tag{2.13}$$

It is clear therefore that if $\pi_H(z)$ is any polynomial of degree H in z, then Eqs. (2.8) defining the Padé approximants

$$P_{mn0}(z) \equiv \Big(\sum_0^n a_{0r}z^r\Big)\Big(\sum_0^m b_{0s}z^s\Big)^{-1} \tag{2.14}$$

are satisfied if we choose

$$\begin{aligned}\sum_0^n a_{0r}z^r &\equiv \pi_H(z)\Big[\sum_0^N c_t z^t\Big] \quad (n \geq N + H),\\ \sum_0^m b_{0s}z^s &\equiv \pi_H(z)\Big[\sum_0^M e_u z^u\Big] \quad (m \geq M + H).\end{aligned} \tag{2.15}$$

So for a function of form (2.11), functions $P_{mn0}(z)$ satisfying (2.14) and (2.15) are Padé approximants of $f_0(z)$, with coefficients $\{a_{0r}\}$ and $\{b_{0s}\}$ satisfying (2.9) and (2.10). We now establish the converse, that all Padé approximants of $f_0(z)$ are of this form.

If we multiply the equation analogous to (2.8) by $\sum_0^M e_u z^u$, we obtain

$$\Big[\sum_0^m b_{0s}z^s\Big]\Big[\sum_0^N c_r z^r\Big] = \Big[\sum_0^n a_{0r}z^r\Big]\Big[\sum_0^M e_u z^u\Big] + O(z^{m+n+1}).$$

For $m \geq M$, $n \geq N$, no terms in the two products here contain terms $O(z^{m+n+1})$, and so

$$\Big[\sum_0^m b_{0s}z^s\Big]\Big[\sum_0^N c_r z^r\Big] \equiv \Big[\sum_0^n a_{0r}z^r\Big]\Big[\sum_0^M e_u z^u\Big]. \tag{2.16}$$

This equation is an identity for $|z| < \rho_1$; hence it is an identity for all z. Since $\sum^N c_r z^r$ and $\sum^M e_u z^u$ have no common factors, it follows by the remainder theorem that $\sum^n a_{0r}z^r$ and $\sum^m b_{0s}z^s$ are of the form (2.14). So we have established

Theorem 1: The class of functions defined by (2.14) and (2.15) are the Padé approximants of the function $f_0(z)$ defined by (2.11).

This theorem means that the Padé approximants of $f_0(z)$ with $m \geq M$ and $n \geq N$ are effectively identical with $f_0(z)$, and that we are justified in taking

$$P_{mn0}(z) \equiv f_0(z). \tag{2.17}$$

3. THE APPROXIMATION THEOREM

We have established (2.17) for the function $f_0(z)$, and it is not difficult to show that a function $f(z)$ of form (2.1) is approximated by $f_0(z)$. It is therefore necessary to relate the approximants $P_{mn}(z)$ and $P_{mn0}(z)$ of these two functions in order to show that $P_{mn}(z)$ approximates $f(z)$.

Since $\sum_0^\infty c_t z^t$ in (2.1) is uniformly convergent for $|z| \leq \rho$, we can find a constant L such that

$$c_t \leq L\rho^{-t} \qquad (t = 0, 1, 2, \cdots), \tag{3.1}$$

where L is independent of t.

Consider the coefficients $\{d_t\}$ in (2.4) and $\{d_{0t}\}$ in (2.12). Since the numerators in (2.1) and (2.11) agree to order z^N,

$$d_t = d_{0t} \qquad (t \leq N). \tag{3.2}$$

If we write

$$\prod_i \Big(1 - \frac{z}{\zeta_i}\Big)^{-k_i} = \sum_0^\infty g_t z^t, \tag{3.3}$$

then from (2.1), (2.11) and (3.1),

$$\begin{aligned} d_t - d_{0t} &= \sum_{u=N+1}^{t} c_u g_{t-u}\\ &\leq L\sum_{N+1}^{t} \rho^{-u} g_{t-u}, \end{aligned}$$

for $t > N$. Also for $h > N$, writing $|z| = r$,

$$\begin{aligned}\sum_{t=N+1}^{h} |(d_t - d_{0t})z^t| &\leq \sum_{t=N+1}^{h} Lr^t \sum_{u=N+1}^{t} \rho^{-u}\,|g_{t-u}|\\ &\leq L\Big[\frac{r}{\rho}\Big]^{N+1} \sum_{u=0}^{h-N-1}\Big[\frac{r}{\rho}\Big]^u \sum_{w=0}^{h-N-1} |g_w| r^w. \end{aligned} \tag{3.4}$$

Now from (3.3) and (2.6), we know that

$$\sum_{w=0}^{h-N-1} |g_w|\, r^w$$

consists of terms up to $O(r^{h-N-1})$ occurring in the expansion of

$$\prod_i \Big(1 - \frac{r}{\rho_i}\Big)^{-k_i}$$

and hence is dominated by terms in the expansion of

$$\prod_i \Big(1 - \frac{r}{\rho_1}\Big)^{-M}.$$

Therefore (3.4) gives

$$\begin{aligned}&\sum_{t=N+1}^{h} |(d_t - d_{0t})z^t|\\ &\leq L\Big(\frac{r}{\rho}\Big)^{N+1}\Big(1 - \frac{r}{\rho}\Big)^{-1} \sum_{w=0}^{h-N-1} \frac{(M + w - 1)!}{(M - 1)!\, w!}\Big(\frac{r}{\rho_1}\Big)^w\\ &\leq L\Big(\frac{r}{\rho}\Big)^{N+1}\Big(1 - \frac{r}{\rho}\Big)^{-1} \frac{(M+h-N-2)^{M-1}}{(M-1)!} \sum_0^{h-N-1}\Big(\frac{r}{\rho_1}\Big)^w\\ &= L\,\frac{(M + h - N - 2)^{M-1}}{(M - 1)!}\Big(\frac{r}{\rho}\Big)^{N+1}\\ &\quad \times \Big(1 - \frac{r}{\rho}\Big)^{-1} \frac{1 - (r/\rho_1)^{h-N-1}}{1 - (r/\rho_1)}. \end{aligned} \tag{3.5}$$

Provided that

$$r \leq \rho(1 - \epsilon) \tag{3.6}$$

and

$$|r - \rho_i| \geq \epsilon\rho_i, \tag{3.7}$$

then from (3.2) and (3.5)

$$\sum_0^h |(d_t - d_{0t})z^t| \leq L\epsilon^{-2}\frac{(M + h - N - 2)^{M-1}}{(M-1)!}\left(\frac{r}{\rho}\right)^{N+1} \times \text{Max}\left[1, \left(\frac{r}{\rho_1}\right)^{h-N-1}\right] \tag{3.8}$$

for $h > N$.

Now consider the approximants defined by (2.7), (2.8), (2.14), and (2.17). We have

$$D_{mn}(z) \equiv \left[\sum_0^n a_r z^r\right]\left[\sum_0^M b_{0s}z^s\right] - \left[\sum_0^N d_{0u}z^u\right]\left[\sum_0^m b_s z^s\right] = \left[\sum_0^m b_s z^s\right]\left[\sum_0^M b_{0s}z^s\right] \times \left[\sum_0^\infty (d_t - d_{0t})z^t\right] + O(z^{m+n+1}).$$

In order to apply Theorem 1 later, we assume

$$m \geq M, \qquad n = N;$$

then since D_{mn} contains no terms of order higher than z^{m+n},

$$|P_{mn}(z) - P_{mn0}(z)| \leq \frac{\sum_0^m |b_s z^s| \sum_0^M |b_{0s}z^s| \sum_0^{m+n} |(d_t - d_{0t})z^t|}{|\sum_0^m b_s z^s|\,|\sum_0^M b_{0s}z^s|}. \tag{3.9}$$

The last factor in the numerator is bounded by (3.8). From (2.17) we have

$$\sum_0^M b_{0s}z^s = \prod_i (1 - z/\zeta_i)^{k_i}.$$

Thus for points in C obeying (3.6) and (3.7)

$$\frac{\sum_0^M |b_{0s}z^s|}{|\sum_0^M b_{0s}z^s|} \leq \epsilon^{-M}\prod_i \left(1 + \frac{\rho}{\rho_i}\right)^{k_i}, \tag{3.10}$$

a bound which is independent of m and n.

In order to provide a suitable bound for the third factor

$$\left(\sum_0^m |b_s z^s|\right)\left(\left|\sum_0^m b_s z^s\right|\right)^{-1}$$

in (3.9), we must assume that the total order of zeros of $\sum_0^m b_s z^s$ in the region $|z| < \rho$ is bounded for all (m, n). Thus, we write

$$\sum_0^m b_s z^s = \prod_i \left(1 - \frac{z}{\xi_i}\right)^{k_i} \prod_i \left(1 - \frac{z}{\xi_i'}\right)^{k_i'}, \tag{3.11}$$

where

$$\sum_i h_i \leq \mu \tag{3.12}$$

for some finite integer μ independent of (m, n), but dependent on ρ, and where

$$\sigma_i' \equiv |\xi_i'| \geq \rho \tag{3.13}$$

for all i. If we now consider a point z satisfying (3.7) and

$$|r - \sigma_i| \geq \epsilon\sigma_i \tag{3.14}$$

then using (3.7), (3.11), (3.12), and (3.14),

$$\frac{\sum_0^m |b_s z^s|}{|\sum_0^m b_s z^s|} \leq \epsilon^{-\mu}\prod_i \left(1 + \frac{\rho}{\sigma_i}\right)^{k_i}\left[\frac{r + \rho}{r - \rho}\right]^{m-\mu}. \tag{3.15}$$

Now we have taken $N = n$, the "best" value of N. Putting $h = n + m$ in (3.8), and also using (3.9), (3.10), and (3.15), we find

$$|P_{mn}(z) - P_{mn0}(z)| \leq B(m + M - 2)^{M-1}\left(\frac{r}{\rho}\right)^{n+1}\left(\frac{\rho + r}{\rho - r}\right)^{m+1} \times \text{Max}\left[1, \left(\frac{r}{\rho_1}\right)^{m-1}\right]. \tag{3.16}$$

where B is a constant dependent upon ρ and ϵ, *but independent of m and n.*

Let us first assume that $n \geq m$; then since (M being fixed)

$$\lim_{m\to\infty} (m + M - 2)^{M/m} = 1,$$

it follows that $|P_{mn} - P_{mn0}| \to 0$ provided

$$\frac{r(\rho + r)}{\rho(\rho - r)} < 1 \qquad (r \leq \rho_1), \tag{3.17}$$

$$\frac{r^2(\rho + r)}{\rho\rho_1(\rho - r)} < 1 \qquad (r > \rho_1). \tag{3.18}$$

The condition (3.17) is

$$|z| = r < \rho(\sqrt{2} - 1) \equiv r_1. \tag{3.19}$$

say. Provided the distance ρ_1 of the nearest pole of $f(z)$ from the origin exceeds r_1, (3.19) is the condition that P_{mn} and P_{mn0} approximate each other. If $\rho_1 < r_1$, the range of r is $r < r_2$, where r_2 is the unique root of

$$r_2^2(\rho + r_2) = \rho\rho_1(\rho - r_2) \tag{3.20}$$

lying in the range $0 < r_2 < r_1$. Thus as $m, n \to \infty$ in any way such that $n \geq m$,

$$|P_{mn}(z) - P_{mn0}(z)| \to 0 \tag{3.21}$$

uniformly in z in the region defined by (3.7), (3.14) and

$$|z| \leq \text{Min}\,(r_1, r_2) - \epsilon, \tag{3.22}$$

r_1 and r_2 being the positive numbers defined by (3.19) and (3.20). Since $P_{mn}(z)$ and $P_{mn0}(z)$ are uniformly continuous in the region $\mathbf{\Delta}$ defined by (3.22) and

$$|z - \zeta_i| \geq \epsilon\rho_i, \; |z - \zeta_j| \geq \epsilon\sigma_j, \tag{3.23}$$

(3.21) holds uniformly in $\mathbf{\Delta}$.

To establish a similar result for $n \leq m$, we note that if P_{mn} is the (m, n) approximant of $f(z)$, then P_{mn}^{-1} is the (n, m) approximant to $[f(z)]^{-1}$. Now $\phi(z)$ is meromorphic where $f(z)$ is, so applying the result (3.21) to $\phi(z)$, we know that if $m, n \to \infty$ with $m \geq n$, then

$$|P_{mn}^{-1} - P_{mn0}^{-1}| \to 0$$

uniformly in the region $\mathbf{\Delta}$, except near to the zeros of $\phi(z)$ and its approximants. Therefore (3.21) holds in $\mathbf{\Delta}$ for any sequence of values of the pair (m, n) for which $m \to \infty$ and $n \to \infty$.

Now from (2.1), (2.11), (2.17), and (3.1),

$$|f(z) - P_{mn0}(z)| \leq L\epsilon^{-M-1}(r/\rho)^{n+1}$$

in the region defined by (3.7), (3.14), and (3.22). Again using uniformity of continuity, we see that

$$|f(z) - P_{mn0}(z)| \to 0$$

uniformly in the region $\mathbf{\Delta}$ as $n \to \infty$. Combining this result with (3.21), we have proved that

$$|f(z) - P_{mn}(z)| \to 0$$

uniformly in $\mathbf{\Delta}$ as $m \to \infty$ and $n \to \infty$ in any way.

Now given a number σ less than R, we can choose ρ and ϵ to satisfy

$$\rho < R, \qquad \sigma < \rho - \epsilon.$$

Then we have established

Theorem 2: A function $f(z)$ is regular and non-zero at $z = 0$, and is meromorphic in the region $|z| < R$, and σ is any positive number less than R. Let $\{P_{mn}(z)\}$ be an infinite sequence, with $m \to \infty$ and $n \to \infty$ in any way, of Padé approximants to $f(z)$, such that the numbers of poles and zeros of every approximant in the region $|z| \leq \sigma$ are each less than a number $\mu(\sigma)$ independent of m and n Then the sequence $\{P_{mn}(z)\}$ converges uniformly to $f(z)$ in the region $\mathbf{\Delta}$ defined by (3.22) and (3.23).

If $\rho_1 \geq r_1$, so that $r_1 \leq r_2$, the region (3.22) is a circle, center the origin, in which $f(z)$ is regular; inside the circle of convergence, it is easy to prove that $\{P_{mn}(z)\}$ approximates $f(z)$, so that the theorem is not very significant. But when $\rho_1 < r_1$, so that $r_2 < r_1$, (3.20) and (3.19) give

$$\frac{\rho_1}{r_2} = \frac{r_2(\rho + r_2)}{\rho(\rho - r_2)} < \frac{r_1(\rho + r_1)}{\rho(\rho - r_1)} = 1.$$

Thus at least one pole lies in the circle (3.22), for small enough ϵ. If $\rho_1 \ll \rho$, then $r_2 \ll \rho$, and (3.20) gives

$$\frac{\rho_1}{r_2} \sim \frac{r_2}{\rho}.$$

So when the nearest nonpolar singularity (ρ) is much further from the origin than the nearest pole (ρ_1), the radius (r_2) inside which $f(z)$ is approximated by $\{P_{mn}(z)\}$ much greater than ρ_1. Hence the theorem establishes convergence of $\{P_{mn}(z)\}$ well beyond the radius of convergence of $\sum d_i z^i$.

If the function $f(z)$ is meromorphic in the whole of the finite z-plane, then we can choose R and hence σ as large as we please. Thus a meromorphic function $f(z)$ is approximated by any sequence $\{P_{mn}(z)\}$ with $m \to \infty$ and $n \to \infty$ independently, provided the number of the poles and zeros of $P_{mn}(z)$ in any finite part of the z-plane is uniformly bounded.

We also note that by letting $n \to \infty$ more rapidly than m in (3.16), the restriction (3.22) on $|z|$ can be eased. For example, by choosing $n = m^2$ we obtain a sequence which approximates $f(z)$ throughout a region $|z| \leq \sigma < R$.

4. DISCUSSION

If we consider a sequence $\{P_{m\,m+k}\}$ of approximants with k fixed, the invariance theorems of Baker, Gammel, and Wills[3] allow us to replace the region of meromorphy $|z| < R$ in Theorem 2 by any finite domain $\mathbf{D}$ which is the union of regions derived from $|z| \leq R$ by transformations of the type

$$z \to w = Az/(1 + Bz).$$

Under these transformations, the region (3.22) transforms into a sub-domain $\mathbf{D}^1$ of $\mathbf{D}$. Thus sequences $\{P_{m\,m+k}\}$ will approximate $f(z)$ throughout regions of meromorphy which will often extend over large regions of the z-plane. The restriction (3.22) is

rather peculiar, and one feels that the region of uniform convergence should be the whole of **D** apart from small regions near the poles and the boundary.

We can see that the restriction (3.22) reflects the known behavior of approximants if we consider a function $f(z)$ which has an isolated essential singularity at, say, $z = R$. This singularity is "simulated" in approximants by the clustering of poles and zeros of $P_{mn}(z)$ near to $z = R$. Thus a critical factor approximated by $(\rho - r)^{m+1}$ in (3.16) is in fact expected to arise, $z = R$ being a limit point of poles. What is not reflected in Eq. (3.16) is the fact that these poles of P_{mn} are compensated by nearby zeros. However, even if the factor

$$\left(\frac{\rho + r}{\rho - r}\right)^{m+1} \tag{4.1}$$

could be eliminated from (3.16), the factor $(r/\rho_1)^{m-1}$ would still prevent us from establishing convergence of diagonal sequences throughout $|z| \leq \sigma < R$.

The problem of eliminating the factor (4.1) is part of the larger problem of locating the poles and zeros of the approximants $P_{mn}(z)$, and in particular showing that in regions of meromorphy the poles of $P_{mn}(z)$ either are near to those of $f(z)$ or are accompanied by a nearby zero. This last property can be established in any domain δ in which $\{P_{mn}(z)\}$ approximates $f(z)$ uniformly. Consider the integrals of the logarithmic derivatives of $P_{mn}(z)$ and $f(z)$ round any contour γ lying in δ which does not pass through any poles or zeros of $f(z)$ and its approximants. The difference between the integrals can be made less than 2π in magnitude by taking m and n large enough; since the integrals are integral multiples of $2\pi i$, we know that they are then equal. So if (P_1, N_1) and (P_2, N_2) are the numbers of poles and zeros of $P_{mn}(z)$ and $f(z)$ respectively,

$$P_1 - N_1 = P_2 - N_2.$$

Thus the poles and zeros of an approximant in the region δ either lie near to those of $f(z)$, or else they occur in pairs, giving nearly canceling factors in the denominator and numerator of the approximant.

Thus the two large problems in the theory of Padé approximants, the location of the poles and the specification of the region in which a sequence approximates a function $f(z)$, are closely linked.

ACKNOWLEDGMENTS

The author is indebted to Professor George Baker, Jr., for valuable discussions which resulted in a substantial improvement of this paper, and to Dr. Brian Murdoch for a critical reading of the manuscript.

Complex Space–Time and Classical Field Theory. I*

A. DAS
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received 10 May 1965)

This is the first of the series of three papers which introduces complex space–time to describe physical phenomena. The objective of this generalization is twofold: firstly, to geometrize gauge transformations and electromagnetic fields, and secondly, to quantize space–time in order to remove serious divergences from the field theory. In this paper classical fields are discussed in complex space–time with a view of subsequent generalization to quantum field theory in quantized space–time.

I. INTRODUCTION

THIS paper prepares the groundwork for the quantization of space–time which follows in the next one. If we want the quantization in such a fashion that measurements of four real coordinates should not interfere with each other, then we have to enlarge the background space at least to real eight-dimensional or complex four-dimensional space (two conjugate variables to each measurement). But this generalization immediately enriches the group of transformations under which physical laws could be covariant and thus offers the possibility of fusion of so-called iso-groups into geometrical ones. For the sake of simplicity we consider the group $L_4\ddagger \times U_1$ which is the simplest possible generalization of the proper Lorentz group to absorb the gauge transformations[1] and is of course a possible group of transformations in four-dimensional complex space–time. Moreover, the complex space–time can give a geometrical meaning to electrical properties, and we identify the electric charge as angular momenta in complex planes.

In this paper, firstly, we shall discuss the representations of the group $L_4\ddagger \times U_1$, and secondly, the Lagrangian mechanics of fields in complex space–time.

II. THE SEVEN-PARAMETER GROUP $L_4\ddagger \times U_1$

Throughout three papers, units are so chosen that $\hbar = c = l = 1$, and all physical quantities are expressed as pure numbers. Let us consider the space spanned by four complex planes coordinatized by z^k (k and other Roman indices will take values 1, 2, 3, 4) and generalize the Minkowskian metric form to

$$\Phi \overset{\text{def}}{=} -|dz^1|^2 - |dz^2|^2 - |dz^3|^2 + |dz^4|^2 = \eta_{ij}\, \overline{dz^i}\, dz^j. \tag{2.1}$$

Here we have assumed the summation convention, bar denotes complex conjugation and η_{ij} is the metric tensor.

The homogeneous linear transformation

$$z'^k = A^k_{\cdot l} z^l, \qquad \overline{z'^k} = \overline{A^k_{\cdot l}}\,\overline{z^l},$$

which leave the metric-form (2.1) invariant, satisfy

$$A^\dagger \eta A = \eta, \qquad \det A = e^{i\alpha}, \tag{2.2}$$

where matrices $A = [a^i_{\cdot j}]$, $\eta = [\eta_{ij}]$; α is a real number and the dagger denotes Hermitian conjugation.

The linear transformations in (2.2) form a 16-real-parameter continuous group.[2] This group is much richer than the ordinary homogeneous Lorentz group. It should also be noted that in this group there is no disjunct proper and improper subgroups.[3]

For the sake of simplicity we shall consider only a subgroup of the group defined in (2.2). This subgroup is denoted by $A^i_{\cdot j} \sim a^i_{\cdot j} e^{i\theta}$, where $a \equiv [a^i_{\cdot j}]$ is an element of $L_4\ddagger$, the proper, isochronous subgroup of the homogeneous Lorentz-group and $e^{i\theta}$ is an element of one-dimensional unitary group U_1. The product group[4] $L_4\ddagger \times U_1$ is a 7-real parameter[5] continuous group and suffices our prime purposes. Though we have not considered space reflections, the total reflection can be taken care of by putting $\theta = \pi$. Also, $\theta = \frac{1}{2}\pi$ generates the reciprocity transformation.

The basic space of representation[6] for $L_4\ddagger \times U_1$ can be coordinatized by z^{k+} and the conjugate space by $z^{k-} = \overline{z^{k+}}$, so that under $L_4\ddagger \times U_1$, the trans-

* This paper has been prepared under Research Grant DA-ARO(D)-124-G-602, U. S. Army Research Office, Durham.

[1] cf. J. Rzewusky, Bull. Acad. Polon. Sci. Classe (III) **6**, 26, 339 (1958); Nuovo Cimento **9**, 942 (1958).

[2] cf. A. Barut, J. Math. Phys. 5, 1652 (1964).

[3] cf. H. Fröhlich, Proc. Roy. Soc. (London) **A257**, 147, 283 (1960).

[4] The cross denotes the direct product or in matrix language the Kronecker product.

[5] It is well known that, if we want gauge transformations besides proper Lorentz transformations to be induced by transformations in spin space, then also we have to consider a seven-parameter group.

[6] This may also be considered as space of ray-representation of $L_4\ddagger$.

formations induced in these spaces are

$$\begin{aligned} z^{k'\pm'} &= a^{k'}_{\cdot k} e^{\pm i\theta} z^{k\pm}, \\ \text{mod } (z^{k'\pm'}) &\leq \text{mod } (a^{k'}_{\cdot k}) \text{ mod } (z^{k\pm}), \\ \arg (z^{k'\pm'}) &= \arg (z^{k\pm}) \pm \theta, \end{aligned} \tag{2.3}$$

where upper and lower signs should be read separately.

We shall physically interpret mod $z^{k\pm}$ as what we usually measure for positional coordinates,[7] and arg $z^{k\pm}$ are the electrical or internal coordinates. The transformations U_1 on z^{k+} form circular group[8] and will induce gauge transformations (a misnomer for rotations!) on field quantities.

The tensorial representation of $L_4\ddagger \times U_1$ would satisfy the following transformation rules:

$$T'^{\underbrace{i'j'\cdots}_{r}\underbrace{+'\cdots}_{s}\underbrace{-'\cdots}_{t}} = e^{i(s-t)\theta} \underbrace{a^{i'}_{\cdot i} a^{j'}_{\cdot j} \cdots}_{r} \times T^{ij\cdots+\cdots-\cdots}. \tag{2.4}$$

The tensor fields are defined by the following transformation properties:

$$\begin{aligned} &T'^{\underbrace{i'j'\cdots}_{r}\underbrace{+'\cdots}_{s}\underbrace{-'\cdots}_{t}}(z^{k'+'}, z^{k'-'}) \\ &\quad = e^{i(s-t)\theta} \underbrace{a^{i'}_{\cdot i} a^{j'}_{\cdot j} \cdots}_{r} T^{ij\cdots+\cdots-\cdots}(z^{k+}, z^{k-}). \end{aligned} \tag{2.4'}$$

In (2.4) the set of 4^r formal functions of complex variable z^{k+} and z^{k-} are assumed to be continuous, partially and totally differentiable in the domain of consideration. But these formal functions may not be analytic functions of four complex variables and need not satisfy Cauchy–Riemann conditions. The unimodular factor in front of right-hand side of (2.4) corresponds to gauge transformations.

The raising and lowering of indices are defined as follows:

$$\begin{aligned} A^{+}_{i} &\overset{\text{def}}{=} \eta_{ij} A^{j+}, & A^{i+} &= \eta^{ij} A^{+}_{j}, \\ A_{i-} &\overset{\text{def}}{=} A^{+}_{i}, & A^{i-} &= A^{i}_{+}, \end{aligned} \tag{2.5}$$

so that expressions like $A_{i+}B^{i+} = A^{i-}B_{i-} = \eta_{ij}A^{i-}B^{j+}$ are invariant under $L_4\ddagger \times U_1$.

Tensorial representations of $L_4\ddagger \times U_1$ in (2.4) are in general reducible, and irreducible components can be obtained by applying permutations on indices according to the Young tableux. But these representations do not exhaust all irreducible representations and leave out spinorial representations.

Therefore, to obtain all irreducible representations we should recall the homomorphism $L_4\ddagger \sim c_2$, where c_2 is a two-dimensional unimodular group.

In their basic spaces of representations, c_2, $\bar{c}_2$, U_1 induce following transformations:

$$\begin{aligned} &c_2 : V'^{\alpha} = c^{\alpha}_{\cdot\beta} V^{\beta}; \quad \bar{c}_2 : V'^{\dot{\alpha}} = c^{\dot{\alpha}}_{\cdot\dot{\beta}} V^{\dot{\beta}}; \\ &U_1 : V' = e^{i\theta} V, \quad \bar{V}' = e^{-i\theta} \bar{V}, \quad \det [c^{\alpha}_{\cdot\beta}] = 1, \end{aligned} \tag{2.6}$$

where Greek indices take value 1, 2.

The $[(2j+1)(2j'+1)]$-dimensional space of representation of $L_4\ddagger \times U_1 \sim c_2 \times \bar{c}_2 \times U_1$ is spanned by monomials

$$P_{(kk')} = \frac{(V^1)^{2j-k}(V^2)^{k}(V^{\dot{1}})^{2j'-k'}(V^{\dot{2}})^{k'} V^{t} \bar{V}^{t'}}{[(2j-k)!\, k!\, (2j'-k')!\, k'!]^{\frac{1}{2}}}, \qquad |k| \leq j, \quad |k'| \leq j'. \tag{2.7}$$

The transformation (2.6) induces the following transformations in monomials,

$$\begin{aligned} P'_{(qq')} &= \sum_{k} \sum_{k'} D_{(qq')(kk')} P_{kk'}, \\ D_{(qq')(kk')} &= e^{i(t-t')\theta} \\ &\times \sum_{k_1} \sum_{k_2} \frac{\binom{2j-q}{k_1}\binom{q}{k-k_1}\binom{2j'-q'}{k_2}\binom{q'}{k'-k_2}}{[(2j-q)!\, q!\, (2j'-q')!\, q'!]^{\frac{1}{2}}} \\ &\times (c^{1}_{\cdot 1})^{2j-q-k_1}(c^{1}_{\cdot 2})^{k_1}(c^{2}_{\cdot 1})^{q-k+k_2}(c^{2}_{\cdot 2})^{k-k_2} \cdots \\ &\times (c^{\dot{1}}_{\cdot\dot{1}})^{2j'-q'-k_2}(c^{\dot{1}}_{\dot{2}})^{k_2}(c^{\dot{2}}_{\cdot\dot{1}})^{q'-k'+k_2}(c^{\dot{2}}_{\cdot\dot{2}})^{k'-k_2}. \end{aligned} \tag{2.8}$$

The $[(2j+1)(2j'+1) \times (2j+1)(2j'+1)]$-dimensional matrices $D_{(qq')(kk')}$ are irreducible representations of the group $L_4\ddagger \times U_1$. The transformations $D_{(qq')(kk')}$ are induced by (2.6) in the space spanned by the spinors $\psi^{\alpha_1\cdots\alpha_s\dot{\beta}_1\cdots\dot{\beta}_{s'}+\cdots-\cdots}$ which is symmetric into $2j$ undotted and $2j'$ dotted indices.

Finite-dimensional representations of $L_4\ddagger \times U_1$ are not unitary because $L_4\ddagger$ is not compact, but there exist infinite-dimensional unitary representations[9] of $L_4\ddagger \times U_1$.

The representation of infinitesimal $L_4\ddagger \times U_1$ transformations is of the following form:

$$D = I + \tfrac{1}{2}\epsilon^{ab} S_{ab} + i\epsilon\tau, \tag{2.9}$$

where $|\epsilon| > |\epsilon|^2$, $|\epsilon^{ab}| > |\epsilon^{ab}|^2$; τ is an integer, and also,

$$\epsilon^{ab} = -\epsilon^{ba}, \qquad S_{ab} = -S_{ba}.$$

Lie–Cartan integrability conditions satisfied by infinitesimal operators S_{ab} are[10]

[7] The negative values may be obtained if we impose the restrictions $0 \leq \arg z^{k+} < \pi$.

[8] See S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton University Press, Princeton, New Jersey, 1948), p. 10.

[9] Scc. Harishchandra, Proc. Roy. Soc. (London) **A189**, 372 (1947).

[10] $[A, B] \overset{\text{def}}{=} AB - BA$; $\delta_{\cdot j}^{i}$ is the Kronecker delta; $c_{kp,\cdot mn}^{rs}$ are structure constants; $g_{ab,cd}$ is the metric tensor.

$$[S_{kp}, S_{mn}] = c^{rs}_{kp,mn} S_{rs},$$

$$c^{rs}_{kp,mn} \overset{\text{def}}{=} \eta_{pm}\delta^r_{\cdot k}\delta^s_{\cdot n} - \eta_{km}\delta^r_{\cdot p}\delta^s_{\cdot n} + \eta_{kn}\delta^r_{\cdot p}\delta^s_{\cdot m} - \eta_{pn}\delta^r_{\cdot k}\delta^s_{\cdot m},$$

$$g_{ab,cd} \overset{\text{def}}{=} c^{rs}_{ab,mn}c^{mn}_{cd,rs} = 4(\eta_{bc}\eta_{ad} - \eta_{bd}\eta_{ac}).$$

The fact that $\det [g_{ab,cd}] \neq 0$, and $g_{ab,cd}$ is not negative-definite shows that $L_4\ddagger$ (consequently $L_4\ddagger \times U_1$) is semisimple and noncompact.[11]

III. WAVE EQUATIONS COVARIANT UNDER $L_4\ddagger \times U_1$

We take as the basic wave equation the linear first-order form

$$(\alpha^{k+}\partial_{k+} + \alpha^{k-}\partial_{k-} - im\, I)\psi = 0, \tag{3.1}$$

where the matrices $\alpha^{k\pm}$ are required to be an irreducible representation of an abstract ring and

$$\partial_{k\pm} \overset{\text{def}}{=} \frac{\partial}{\partial z^{k\pm}}.$$

We shall derive the general commutation rules for the $\alpha^{k\pm}$ matrices from the condition of $L_4\ddagger \times U_1$ invariance which requires

$$\begin{aligned} z'^{k\pm} &= a^k_{\cdot l} z^{l\pm}, \\ \psi' &= S\psi, \\ S\alpha^{k\pm} S^{-1} &= a^k_{\cdot l}\alpha^{l\pm}. \end{aligned} \tag{3.2}$$

The quantity $P \overset{\text{def}}{=} \alpha^{k+}p_{k+} + \alpha^{k-}p_{k-}$ undergoes the following transformation:

$$\begin{aligned} P' &= \alpha^{k+}p'_{k+} + \alpha^{k-}p'_{k-} \\ &= SPS^{-1}. \end{aligned} \tag{3.3}$$

As a matrix of finite degree, P satisfies a minimal equation whose coefficients are polynomials in $p_{k\pm}$. But (3.3) shows that the minimal equation is invariant under $L_4\ddagger \times U_1$, so that it contains $p_{k\pm}$ in the invariant combination $|p|^2 \overset{\text{def}}{=} \eta^{kl}p_{k+}p_{l-}$.

The minimal equations containing even or odd powers of P can be written in the factorized forms, respectively, as

$$\begin{aligned} &(P^2 - a_1 I\,|p|^2)(P^2 - a_2 I\,|p|^2)\cdots \\ &\qquad \times (P^2 - a_{2n} I\,|p|^2) = 0, \\ \text{and}\quad &P(P^2 - a_2 I\,|p|^2)(P^2 - a_3 I\,|p|^2)\cdots \\ &\qquad \times (P^2 - a_{2n+1} I\,|p|^2) = 0. \end{aligned} \tag{3.4}$$

[11] cf. L. S. Pontrjagin, *Topological Groups* (B. G. Teubner, Leipzig, 1957–58).

We can express these equations in the following way:

$$\begin{aligned} &[\alpha^{k+}\alpha^{l+}p_{k+}p_{l+} + \alpha^{k-}\alpha^{l-}p_{k-}p_{l-} \\ &\quad + (\alpha^{k+}\alpha^{l-} + \alpha^{l-}\alpha^{k+} - a_1 I\eta^{kl})p_{k+}p_{l-}] \\ &\quad \times [\alpha^{m+}\alpha^{n+}p_{m+}p_{n+} + \alpha^{m-}\alpha^{n-}p_{m-}p_{n-} \\ &\quad + (\alpha^{m+}\alpha^{n-} + \alpha^{n-}\alpha^{m+} - a_{2n} I\eta^{mn})p_{m+}p_{n-}] = 0, \end{aligned}$$

or P times this is zero. Since these are true for all complex numbers $p_{k\pm}$, the coefficient of each term of the matrix-polynomial equation must vanish, that is

$$\begin{aligned} &\sum_{\varphi} \alpha^{k+}\alpha^{l+}\cdots\alpha^{m+}\alpha^{n+} = 0, \\ &\sum_{\varphi} \alpha^{k-}\alpha^{l-}\cdots\alpha^{m-}\alpha^{n-} = 0, \\ &\sum_{\varphi} (\alpha^{k+}\alpha^{l-} + \alpha^{l-}\alpha^{k+} - a_1 I\eta^{kl})\cdots \\ &\qquad \times (\alpha^{m+}\alpha^{n-} + \alpha^{n-}\alpha^{m+} - a_{2n} I\eta^{mn}) = 0, \end{aligned} \tag{3.5}$$

or $\alpha^{i\pm}$ times (3.5), and here $\sum_{\varphi}$ denotes summation over similar terms with all possible permutations of the indices $k\pm$, $l\pm$, $m\pm$, $n\pm$.

As an example let us study the minimal equation

$$P^2 - I\,|p|^2 = 0, \tag{3.6}$$

which corresponds to the generalization of the Dirac equation. The corresponding $\alpha^{k\pm}$, according to (3.5), satisfy

$$\begin{aligned} &\alpha^{k+}\alpha^{l+} + \alpha^{l+}\alpha^{k+} = \alpha^{k-}\alpha^{l-} + \alpha^{l-}\alpha^{k-} = 0, \\ &\alpha^{k+}\alpha^{l-} + \alpha^{l-}\alpha^{k+} = \eta^{kl} I. \end{aligned} \tag{3.7}$$

The irreducible representations of $\alpha^{k\pm}$ satisfying (3.7) are (16×16)-dimensional matrices, explicitly displayed in the following.

$$\begin{aligned} \alpha^{1+} &= \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \alpha^{1-} &= \begin{pmatrix} 0 & 0 \\ i & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \alpha^{2+} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \alpha^{2-} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 0 & 0 \\ i & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \alpha^{3+} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \alpha^{3-} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 0 & 0 \\ i & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \alpha^{4+} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ \alpha^{4-} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{3.8}$$

The generalization of Duffin–Kemmer matrices for the case of spin-zero particles has the following (9×9)-dimensional irreducible representation:

$$\alpha^{1+} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&-1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ \tfrac{1}{2}&0&0&0&0&0&0&0 \end{bmatrix}, \quad \alpha^{1-} = \begin{bmatrix} 0&0&0&0&0&0&0&-1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&\tfrac{1}{2}&0&0&0 \end{bmatrix},$$

$$\alpha^{2+} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&-1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&\tfrac{1}{2}&0&0&0&0&0&0 \end{bmatrix}, \quad \alpha^{2-} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&-1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&\tfrac{1}{2}&0&0 \end{bmatrix},$$

$$\alpha^{3+} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&-1\\ 0&0&0&0&0&0&0&0\\ 0&0&\tfrac{1}{2}&0&0&0&0&0 \end{bmatrix}, \quad \alpha^{3-} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&-1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&\tfrac{1}{2}&0&0 \end{bmatrix},$$

$$\alpha^{4+} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&\tfrac{1}{2}&0&0&0&0 \end{bmatrix}, \quad \alpha^{4-} = \begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&\tfrac{1}{2}&0 \end{bmatrix}. \tag{3.9}$$

These satisfy the relations

$$\alpha^{i\pm}\alpha^{j\pm}\alpha^{k\pm} + \alpha^{j\pm}\alpha^{i\pm}\alpha^{k\pm} + \alpha^{i\pm}\alpha^{k\pm}\alpha^{j\pm} + \alpha^{k\pm}\alpha^{j\pm}\alpha^{i\pm} + \alpha^{j\pm}\alpha^{k\pm}\alpha^{i\pm} + \alpha^{k\pm}\alpha^{i\pm}\alpha^{j\pm} = 0, \tag{3.10}$$

$$\alpha^{i\pm}\alpha^{j\mp}\alpha^{k\pm} + \alpha^{j\mp}\alpha^{i\pm}\alpha^{k\pm} + \alpha^{i\pm}\alpha^{k\pm}\alpha^{j\mp} + \alpha^{k\pm}\alpha^{j\mp}\alpha^{i\pm} + \alpha^{j\mp}\alpha^{k\pm}\alpha^{i\pm} + \alpha^{k\pm}\alpha^{i\pm}\alpha^{j\mp} = \eta^{ij}\alpha^{k\pm} + \eta^{kj}\alpha^{i\pm}.$$

To construct the energy–momentum–stress tensor and charge–current vector for a wave field ψ, we need a nonsingular matrix Λ such that

$$(\alpha^{k\pm})^{\dagger} = \Lambda\alpha^{k\mp}\Lambda^{-1}. \tag{3.11}$$

If such a matrix should exist it can be chosen Hermitian because $(\Lambda^{\dagger})^{-1}$ commutes with all the irreducible $\alpha^{k\pm}$ from (3.9), and therefore, must be a multiple of the unit matrix.

For the set of matrices in (3.8) the corresponding Λ is given by

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.12}$$

Before concluding this section we should mention that the problem of finding all wave equations covariant under $L_4\ddagger \times U_1$ hinges on the solution of the equation

$$[S_{ab}, \alpha^{k\pm}] = \delta^k_{\cdot a}\eta_{bl}\alpha^{l\pm} - \delta^k_{\cdot b}\eta_{al}\alpha^{l\pm}, \tag{3.13}$$

together with the first one of the equations (2.10).

IV. THE LAGRANGIAN FORMALISM

In this section we shall consider 15-parameter inhomogeneous linear transformations of the type

$$z^{k'\pm'} = a^{k\pm} + a^k_{\cdot m}e^{+i\theta}z^{m\pm}. \tag{4.1}$$

Tensor fields transform like (2.4) under (4.1). The infinitesimal version of (4.1) is

$$z^{k'\pm'} = \epsilon^{k\pm} + (\delta^k_{\cdot m} + \epsilon^k_{\cdot m})(1 \pm i\epsilon)z^{m\pm},$$
$$\delta z^{k\pm} \overset{\text{def}}{=} z^{k'\pm'} - z^{k\pm} = \epsilon^{k\pm} + \epsilon^k_{\cdot m}z^{m\pm} \pm i\epsilon z^{k\pm}, \tag{4.2}$$

where $|\epsilon^{k\pm}|$, $|\epsilon^k_{\cdot m}|$, $|\epsilon|$ are small positive numbers. Under (3.2), the tensor field[12] $\varphi^{\cdot\cdot}(z^+, z^-)$ suffers infinitesimal transformation and the local variation is

$$\delta\varphi^{\cdot\cdot} \overset{\text{def}}{=} \varphi'^{\cdot\cdot}(z^{+'}, z^{-'}) - \varphi^{\cdot\cdot}(z^+, z^-) = (\tfrac{1}{2}\epsilon^{ab}S_{ab} + i\epsilon\tau)\varphi^{\cdot\cdot}(z^+, z^-), \tag{4.3}$$

neglecting $O[\epsilon^2, (\epsilon^{ab})^2]$ terms. The substantial variation is defined as

$$\bar{\delta}\varphi^{\cdot\cdot} \overset{\text{def}}{=} \varphi'^{\cdot\cdot}(z^+, z^-) - \varphi^{\cdot\cdot}(z^+, z^-) = -\epsilon^{k+}\partial_{k+}\varphi^{\cdot\cdot} - \epsilon^{k-}\partial_{k-}\varphi^{\cdot\cdot} + \tfrac{1}{2}\epsilon^{ab}[S^{\cdot\cdot}_{ab\cdot\cdot} - z^+_{[b}\partial_{a]+} - z^-_{[b}\partial_{a]-}]\varphi^{\cdot\cdot} + i\epsilon[\tau - z^{k+}\partial_{k+} + z^{k-}\partial_{k-}]\varphi^{\cdot\cdot}, \tag{4.4}$$

where

$$z^+_{[b}\partial_{a]+} \overset{\text{def}}{=} z^+_b\partial_{a+} - z^+_a\partial_{b+},$$

and we have neglected higher-order terms.

We define the invariant action integral for field quantities as

$$A \overset{\text{def}}{=} \int_{\Omega} \cdots \int_{\bar{\Omega}} \cdots d^4z^+\, d^4z^- L \times [\varphi^{\cdot\cdot}, \bar{\varphi}^{\cdot\cdot}, \partial_{k+}\varphi^{\cdot\cdot}, \partial_{k-}\varphi^{\cdot\cdot}, \partial_{k+}\bar{\varphi}^{\cdot\cdot}, \partial_{k-}\bar{\varphi}^{\cdot\cdot}]. \tag{4.5}$$

In each plane the contour of integration may be chosen as $c : \arg z^+ = \frac{1}{4}\pi$, $\bar{c} : \arg z^- = -\frac{1}{4}\pi$.

Functions $\varphi^{\cdot\cdot}(z^+, z^-)$ and $\bar{\varphi}^{\cdot\cdot}(z^-, z^+)$ are assumed to be continuous and to have up to second-order partial derivatives with respect to z^{k+}, z^{k-} in the domain of consideration Ω, $\bar{\Omega}$. The domain Ω, $\bar{\Omega}$ defines a simply connected region in the eight-dimensional space spanned by Re z^{k+}, Im z^{k+}. We do not consider $\varphi^{\cdot\cdot}$'s as analytic functions of four complex variables, because in that case the variational principle $\delta A = 0$ would yield integro-differential equations instead of partial differential equations. Besides, $\varphi^{\cdot\cdot}$'s would have to satisfy the extra conditions of analyticity.

[12] $(z^{\pm}) \overset{\text{def}}{=\!=} (z^{1\pm}, z^{2\pm}, z^{3\pm}, z^{4\pm})$.

If we apply the variational principle

$$\delta A = 0, \quad \delta\varphi^{\cdot\cdot}[\sigma(\Omega)] = \delta\bar{\varphi}^{\cdot\cdot}[\bar{\sigma}(\bar{\Omega})] = \delta\varphi^{\cdot\cdot}[\bar{\sigma}(\bar{\Omega})], \\ = \delta\bar{\varphi}^{\cdot\cdot}[\sigma(\Omega)] = 0,$$

where σ, $\bar{\sigma}$ implies a regular hypersurface that encloses the eight-dimensional domain defined by Ω, $\bar{\Omega}$, then we obtain the following field equation

$$\partial_{k+}\frac{\partial L}{\partial\partial_{k+}\varphi^{\cdot\cdot}} + \partial_{k-}\frac{\partial L}{\partial\partial_{k-}\varphi^{\cdot\cdot}} - \frac{\partial L}{\partial\varphi^{\cdot\cdot}} = 0, \\ \partial_{k+}\frac{\partial L}{\partial\partial_{k+}\bar{\varphi}^{\cdot\cdot}} + \partial_{k-}\frac{\partial L}{\partial\partial_{k-}\bar{\varphi}^{\cdot\cdot}} - \frac{\partial L}{\partial\bar{\varphi}^{\cdot\cdot}} = 0. \tag{4.6}$$

The invariance of the action integral (4.5) under infinitesimal transformations (4.2) yield

$$\int_\Omega \cdots \int_{\bar{\Omega}} \cdots d^4z^+\, d^4z^-\Big[\partial_{k+}\Big(\frac{\partial L}{\partial\partial_{k+}\varphi^{\cdot\cdot}}\bar{\delta}\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k+}\bar{\varphi}^{\cdot\cdot}}\bar{\delta}\bar{\varphi}^{\cdot\cdot} + L\delta z^{k+}\Big) + \partial_{k-}\Big(\frac{\partial L}{\partial\partial_{k-}\varphi^{\cdot\cdot}}\bar{\delta}\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k-}\bar{\varphi}^{\cdot\cdot}}\bar{\delta}\bar{\varphi}^{\cdot\cdot} + L\delta z^{k-}\Big)\Big] = 0, \tag{4.7}$$

where we have neglected the second-order terms. Because (4.7) is valid for arbitrary Ω, $\bar{\Omega}$ we obtain Noether's theorems by putting the integrand to zero, i.e.,

$$\partial_{k+}\Big(\frac{\partial L}{\partial\partial_{k+}\varphi^{\cdot\cdot}}\bar{\delta}\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k+}\bar{\varphi}^{\cdot\cdot}}\bar{\delta}\bar{\varphi}^{\cdot\cdot} + L\delta z^{k+}\Big) + \partial_{k-}\Big(\frac{\partial L}{\partial\partial_{k-}\varphi^{\cdot\cdot}}\bar{\delta}\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k-}\bar{\varphi}^{\cdot\cdot}}\bar{\delta}\bar{\varphi}^{\cdot\cdot} + L\delta z^{k-}\Big) = 0. \tag{4.8}$$

If we consider the case when $\epsilon^{k\pm} \neq 0$, $\epsilon^{ab} = 0$, $\epsilon = 0$, then (4.8) yields differential energy–momentum conservations as

$$\partial_{k+}T^{k+}_{\cdot m+} + \partial_{k-}T^{k-}_{\cdot m+} = 0, \\ \partial_{k-}T^{k-}_{\cdot m-} + \partial_{k+}T^{k+}_{\cdot m-} = 0, \tag{4.9}$$

$$T^{k\pm}_{\cdot m\pm} \overset{\text{def}}{=} \frac{\partial L}{\partial\partial_{k\pm}\varphi^{\cdot\cdot}}\partial_{m\pm}\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k\pm}\bar{\varphi}^{\cdot\cdot}}\partial_{m\pm}\bar{\varphi}^{\cdot\cdot} - \delta^k_{\cdot m}L,$$

$$T^{k\pm}_{\cdot m\mp} \overset{\text{def}}{=} \frac{\partial L}{\partial\partial_{k\pm}\varphi^{\cdot\cdot}}\partial_{m\mp}\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k\pm}\bar{\varphi}^{\cdot\cdot}}\partial_{m\mp}\bar{\varphi}^{\cdot\cdot}.$$

With $\epsilon^{k\pm} = 0$, $\epsilon^{ab} \neq 0$, $\epsilon = 0$, from (4.8) we obtain differential angular momentum conservations as

$$\partial_{k+}\Big[\frac{\partial L}{\partial\partial_{k+}\varphi^{\cdot\cdot}}(S^{\cdot\cdot}_{ab\cdot\cdot} - z^+_{[b}\partial_{a]+} - z^-_{[b}\partial_{a]-})\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k+}\bar{\varphi}^{\cdot\cdot}}(S^{\cdot\cdot}_{ab\cdot\cdot} - z^+_{[b}\partial_{a]+} - z^-_{[b}\partial_{a]-})\bar{\varphi}^{\cdot\cdot} + L\delta^k_{\cdot[a}z^+_{b]}\Big] \\ + \partial_{k-}\Big[\frac{\partial L}{\partial\partial_{k-}\varphi^{\cdot\cdot}}(S^{\cdot\cdot}_{ab\cdot\cdot} - z^+_{[b}\partial_{a]+} - z^-_{[b}\partial_{a]-})\varphi^{\cdot\cdot} + \frac{\partial L}{\partial\partial_{k-}\bar{\varphi}^{\cdot\cdot}}(S^{\cdot\cdot}_{ab\cdot\cdot} - z^+_{[b}\partial_{a]+} - z^-_{[b}\partial_{a]-})\bar{\varphi}^{\cdot\cdot} + L\delta^k_{\cdot[a}z^-_{b]}\Big] = 0. \tag{4.10}$$

With $\epsilon^{k\pm} = 0$, $\epsilon^{ab} = 0$, $\epsilon \neq 0$, we obtain the differential conservation of electrical charge–current as

$$\partial_{k+}j^{k+} + \partial_{k-}j^{k-} = 0,$$

$$j^{k\pm} \overset{\text{def}}{=} \pm i\epsilon\Big[\frac{\partial L}{\partial\partial_{k\pm}\varphi^{\cdot\cdot}}(\tau - z^{m-}\partial_{m-} + z^{m+}\partial_{m+})\varphi^{\cdot\cdot} - \frac{\partial L}{\partial\partial_{k\pm}\bar{\varphi}^{\cdot\cdot}}(\tau - z^{m+}\partial_{m+} + z^{m-}\partial_{m-})\bar{\varphi}^{\cdot\cdot} + z^{k\pm}L\Big]. \tag{4.11}$$

If we assume the invariance of action integral under the infinitesimal phase transformation[13]

$$\varphi'^{\cdot\cdot} = \varphi^{\cdot\cdot}e^{i\epsilon} = \varphi^{\cdot\cdot}(1 + i\epsilon) + O(\epsilon^2), \\ \bar{\varphi}'^{\cdot\cdot} = \bar{\varphi}^{\cdot\cdot}e^{-i\epsilon} = \bar{\varphi}^{\cdot\cdot}(1 - i\epsilon) + O(\epsilon^2),$$

then we obtain the differential conservation of the density current as

$$\partial_{k+}n^{k+} + \partial_{k-}n^{k-} = 0, \\ n^{k\pm} = \pm i\Big[\frac{\partial L}{\partial\partial_{k\pm}\varphi^{\cdot\cdot}}\varphi^{\cdot\cdot} - \frac{\partial L}{\partial\partial_{k\pm}\bar{\varphi}^{\cdot\cdot}}\bar{\varphi}^{\cdot\cdot}\Big]. \tag{4.12}$$

V. THE COMPLEX SCALAR FIELD

We shall illustrate the Lagrangian formalism by the example of the complex scalar field. The Lagrangian we choose is

$$L = \eta^{km}\partial_{k+}\varphi^+(z^+, z^-)\partial_{m-}\varphi^-(z^-, z^+) - m^2\varphi^+\varphi^-. \tag{5.1}$$

Field equations (3.6) derived from (4.1) yield

$$(\eta^{kq}\partial_{k+}\partial_{q-} + m^2)\varphi^\pm = 0. \tag{5.2}$$

This is the generalization of the usual Klein–Gordon equation into the complex space–time.[14]

The canonical energy–momentum–stress tensor, the electric charge–current vector and the density current vector for this field can be obtained from

[13] This is a nongeometrical transformation!

[14] As partial differential equation, (5.2) differs from the usual Klein–Gordon equation in one respect. In Cauchy problem with the usual Klein–Gordon equation we would choose as data the functions $\varphi^\pm$ and their time derivatives on the initial hypersurface. But with reference to (5.2) we could choose as Cauchy data just the functions $\varphi^\pm$.

(4.1), (3.9), (3.11), (3.12), and they are

$$T^{k-}_{\cdot m-} = T^{k+}_{\cdot m+} = \eta^{kq}(\partial_{q-}\varphi^-\partial_{m+}\varphi^+) - \delta^k_{\cdot m}\eta^{ab}(\partial_{a+}\varphi^+\partial_{b-}\varphi^- - m^2\varphi^+\varphi^-),$$
$$j^{k\pm} = i\epsilon[\pm\eta^{kq}\partial_{q\mp}\varphi^\mp(1 - z^{p-}\partial_{p-} + z^{p+}\partial_{p+})\varphi^\pm + z^{k\pm}\eta^{ab}(\partial_{a+}\varphi^+\partial_{b-}\varphi^- - m^2\varphi^+\varphi^-)],$$
$$n^{k\pm} = i\eta^{kq}[\partial_{q\mp}\varphi^\mp\varphi^\pm]. \tag{5.3}$$

Assuming suitable restrictions, the Fourier integral theorem[15] for the functions $\varphi^\pm$ is[16]

$$\varphi^\pm(z^+, z^-) = \frac{1}{(2\pi)^4}\int\cdots\int\cdots d^4p_+\, d^4p_-a^\pm(p^+, p^-) \times \delta(\eta^{ij}p_{i+}p_{j-} - m^2)e^{\pm i(p_{i+}z^{i+} + p_{i-}z^{i-})}. \tag{5.4}$$

The Green's functions for the partial differential equation (5.2) are

$$G_{(a)}(z^+, z^-; z'^+, z'^-) = \frac{1}{(2\pi)^8}\int\cdots\int_{c(a)}\int\cdots\int_{\bar c(a)} d^4p_+\, d^4p_- \times \frac{e^{i[p_{k+}(z^{k+} - z'^{k+}) + p_{k-}(z^{k-} - z'^{k-})]}}{-\eta^{ij}p_{i+}p_{j-} + m^2}, \tag{5.5}$$

where c_a, $\bar c_a$ are different possible contours in complex p_{4+} or p_{4-} plane.

If we assume single-valuedness of $\varphi^\pm$ in 4 complex planes, and some more suitable restrictions,[17] then the Fourier–Bessel integral theorem for $\varphi^\pm$ can be written as

$$\varphi^\pm = \int_0^\infty\cdots\int_0^\infty d^4k k_1k_2k_3k_4 \times \sum_{n_1=-\infty}^{\infty}\cdots\sum_{n_4=-\infty}^{\infty} a_{n_1n_2n_3n_4}(k_1, k_2, k_3, k_4) \times J_{n_1}(2k_1r^1)J_{n_2}(2k_2r^2)J_{n_3}(2k_3r^3)J_{n_4}(2k_4r^4) \times e^{\pm i(n_1\theta^1 + n_2\theta^2 + n_3\theta^3 + n_4\theta^4)}\delta(\eta^{ij}k_ik_j - m^2), \tag{5.6}$$

where $z^{k\pm} = r^ke^{\pm i\theta^k}$ (k not to be summed).

Furthermore, if $\varphi^\pm$ does not depend on θ^k's, then (5.6) can be simplified

$$\varphi^\pm = \int_0^\infty\cdots\int_0^\infty d^4k\, k_1k_2k_3k_4a(k_1, k_2, k_3, k_4)J_0(2k_1r^1) \times J_0(2k_2r^2)J_0(2k_3r^3)J_0(2k_4r^4)\cdot\delta(\eta^{ij}k_ik_j - m^2). \tag{5.7}$$

Also, associated Green's functions can be expressed in the following form:

$$G_{(a)}(r, r') = \int_0^\infty\cdots\int_{c(a)} d^4k\, k_1k_2k_3k_4J_0(2k_1r^1) \times J_0(2k_2r^2)J_0(2k_3r^3)J_0(2k_4r^4) \times J_0(2k_1r'^1)J_0(2k_2r'^2)J_0(2k_3r'^3) \times J_0(2k_4r'^4)(-\eta^{ij}k_ik_j + m^2)^{-1}. \tag{5.8}$$

VI. THE COVARIANT WAVEFIELD

For a covariant wavefield the Lagrangian is chosen to be

$$L = \frac{1}{2i}[(\partial_{k+}\bar\psi)\alpha^{k+}\psi + (\partial_{k-}\bar\psi)\alpha^{k-}\psi - \bar\psi\alpha^{k+}\partial_{k+}\psi - \bar\psi\alpha^{k-}\partial_{k-}\psi] + m\bar\psi I\psi, \tag{6.1}$$

where $\alpha^{k\pm}$ satisfy (3.5) and $\bar\psi \overset{\text{def}}{=} \psi^\dagger\Lambda$.

The field equations (4.6) derived from (6.1) are

$$(\alpha^{k+}\partial_{k+} + \alpha^{k-}\partial_{k-} - im\, I)\psi = 0, \qquad (\partial_{k+}\bar\psi\alpha^{k+} + \partial_{k-}\bar\psi\alpha^{k-} + im\,\bar\psi I) = 0. \tag{6.2}$$

The energy–momentum–stress tensor, charge–current vector, number–current vector calculated by (4.9), (4.11), (4.12) from (6.1) are, respectively,

$$T^{k\pm}_{\cdot m\pm} = \frac{1}{2i}[(\partial_{m\pm}\bar\psi)\alpha^{k\pm}\psi - \bar\psi\alpha^{k\pm}\partial_{m\pm}\psi] - \delta^k_{\cdot m}L,$$
$$T^{k\pm}_{\cdot m\mp} = \frac{1}{2i}[(\partial_{m\mp}\bar\psi)\alpha^{k\pm}\psi - \bar\psi\alpha^{k\pm}\partial_{m\mp}\psi],$$
$$j^{k\pm} = \pm\epsilon[\bar\psi\alpha^{k\pm}(\tau - z^{l-}\partial_{l-} + z^{l+}\partial_{l+})\psi + (\tau - z^{l+}\partial_{l+} + z^{l-}\partial_{l-})\bar\psi.\alpha^{k\pm}\psi + 2iz^{k\pm}L],$$
$$n^{k\pm} = \pm\bar\psi\alpha^{k\pm}\psi. \tag{6.3}$$

The Green's functions corresponding to the matrix–differential equations (6.2) are

$$S_{(a)}(z^+, z^-; z'^+, z'^-) = \frac{1}{(2\pi)^8}\int\cdots\int_{c(a)} d^4p_+\, d^4p_- \times \frac{I}{i(P - mI)}e^{i[p_{i+}(z^{i+} - z'^{i+}) + p_{i-}(z^{i-} - z'^{i-})]}, \tag{6.4}$$

where P has been defined in (3.3).

For the set of matrices in (3.8), the Green's functions (6.4) go over

$$S_{(a)}(z^+, z^-; z'^+, z'^-) = \frac{1}{(2\pi)^8}\int\cdots \times \int_{c(a)} d^4p_+\, d^4p_- \frac{i(P + mI)}{-\eta^{ab}p_{a+}p_{b-} + m^2} \times e^{i[p_{i+}(z^{i+} - z'^{i+}) + p_{i-}(z^{i-} - z'^{i-})]}. \tag{6.4'}$$

[15] cf. R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex domain* (New York 1934).

[16] $\delta(k)$ is the Dirac delta function, and its use can be justified with distribution theory.

[17] N. Nielsen, *Handbuch der Kugel-Funktionen* (B. G. Tuebner, Leipzig, 1904), pp. 360–363.

For the case when $\alpha^{k\pm}$ are given by (3.8) we can introduce consistently the electromagnetic interaction by generalizing (6.2) to

$$\sum_{k=1}^{4}\left\{\alpha^{k+}e^{-i\theta_k}\left[\frac{\partial}{\partial r^k}-\frac{i}{r^k}\frac{\partial}{\partial\theta^k}+i\epsilon A_k(r)\right]+\alpha^{k-}e^{i\theta_k}\right.$$
$$\left.\times\left[\frac{\partial}{\partial r^k}+\frac{i}{r^k}\frac{\partial}{\partial\theta^k}+i\epsilon A_k(r)\right]-im\,I\right\}\psi=0, \quad (6.5)$$

where $2z^{k\pm}=r^k e^{\pm i\theta_k}$ (k not summed).

If we choose a special form of ψ, namely $\psi=\psi(r)$, then (6.5) goes over to

$$\left\{\underset{\sim}{\alpha}^k\left[\frac{\partial}{\partial r^k}+i\epsilon A_k(r)\right]-im\,I\right\}\psi=0, \quad (6.6)$$

where

$$\underset{\sim}{\alpha}^k \overset{\text{def}}{=} \alpha^{k+}e^{-i\theta_k}+\alpha^{k-}e^{i\theta_k} \quad (k \text{ not summed}),$$
$$\underset{\sim}{\alpha}^k\underset{\sim}{\alpha}^l+\underset{\sim}{\alpha}^l\underset{\sim}{\alpha}^k=2\eta^{kl}I. \quad (6.7)$$

The last equation shows that (6.6) is equivalent to the Dirac equation and (16 × 16)-dimensional representations of α^k's must be reducible.

Choosing another special form of ψ, namely $\psi=\chi(r)e^{iq_a\theta_a}$ (q_a's are integers), (6.5) yields

$$\sum_{k=1}^{4}\left\{\underset{\sim}{\alpha}^k\left[\frac{\partial}{\partial r^k}+i\epsilon A_k(r)\right]+\alpha_*^k\frac{q_k}{r^k}-im\,I\right\}\chi=0, \quad (6.8)$$

where

$$\alpha_*^k \overset{\text{def}}{=} \alpha^{k+}e^{-i\theta_k}-\alpha^{k-}e^{i\theta_k} \quad (k \text{ not summed}),$$
$$\alpha_*^k\alpha_*^l+\alpha_*^l\alpha_*^k=-2\eta^{kl}I,$$
$$\alpha_*^k\underset{\sim}{\alpha}^l+\underset{\sim}{\alpha}^l\alpha_*^k=0. \quad (6.9)$$

The corresponding second-order equation is

$$\left[\sum_k\sum_l\left\{\eta^{kl}\left(\frac{\partial}{\partial r^k}+i\epsilon A_k\right)\left(\frac{\partial}{\partial r^l}+i\epsilon A_l\right)-\frac{q_kq_l}{r^kr^l}\right.\right.$$
$$\left.\left.+\frac{i\epsilon}{4}\underset{\sim}{\alpha}^{[k}\underset{\sim}{\alpha}^{l]}F_{kl}\right\}-\sum_k\underset{\sim}{\alpha}^k\alpha_*^k\frac{q_k}{(r^k)^2}+m^2I\right]\chi=0, \quad (6.10)$$

where

$$F_{kl}\overset{\text{def}}{=}\frac{\partial A_l}{\partial r^k}-\frac{\partial A_k}{\partial r^l}.$$

Both (6.8) and (6.10) contains terms which reveal slightly anistropy in the physical space spanned by four r^k's.

This concludes the discussion on classical fields in complex space-time, and we shall give the quantized version of the theory in the following paper.

The Quantized Complex Space–Time and Quantum Theory of Free Fields. II*

A. DAS
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received 10 May 1965)

In this paper a covariant quantization of complex space–time is proposed. As a consequence of this quantization each of the four real coordinates can take discrete values $n^{\frac{1}{2}}l$, and furthermore, measurements of these coordinates are noninterfering with each other. Next the general theory of quantized free fields is developed in the background of quantized space–time. As an example the case of complex scalar field has been dealt with and it is found that the resulting Green's functions are nonsingular.

I. INTRODUCTION

THE divergence difficulties in the quantum field theory are as old as the theory itself. These difficulties are temporarily avoided by formal renormalization procedures. However, subtraction of infinities is not to everyone's taste, besides such approaches lack universal applicability. In recent years this problem has again been side-tracked by the dispersion theory which through ingenious efforts of many has led to the better understanding of elementary processes, especially in the field of strong interactions. But in our opinion, the complete understanding of the microworld can only come through the satisfactory solution of the classical divergence problem which unfortunately still survives. The present effort is to achieve this end. The basic idea, which is nothing new, is to introduce a fundamental length[1–3] or time into the physical theory in a covariant manner. Attempts along similar directions have been made previously either by introducing a microstructure in space–time[4,5], or by assuming nonlocal interactions[6] or nonlocal fields.[7–13]

Despite these endeavors, no fully developed quantum theory of field exists which incorporates the concept of fundamental length satisfactorily. In 1960 an attempt to formulate exhaustively the quantum field theory in the background of discrete space–time where each coordinate assumed discrete value $\pm nl$ was made.[14] There, serious divergences were eliminated in a mathematically consistent way. However, the partial difference equations which replaced usual partial differential equations did not satisfy Lorentz covariance, though the basic cellular space–time structure was integral Lorentz-invariant.[5] There seems to be no other way to formulate the convergent, covariant field theory except by the quantization of space–time with the introduction of coordinate operators. To restrict the multitude of possibilities, the following plausible conditions on the quantization are imposed: (1) Measurements of four real coordinates should be compatible or noninterfering; (2) the eigenvalue spectrum of the position operators should necessarily assume denumerable infinite number of values; (3) the quantization rules should be preferably covariant under the combined Lorentz and gauge (or, iso-) groups. Several trials[15] gave the simplest possible quantization satisfying the above criteria as $[Z^{k+}, Z^{l-}] = l^2\eta^{kl}I$ and resulted in the quantization of space–time lattices in discrete values $n^{\frac{1}{2}}l$ for each coordinate. Snyder's[4] quantization does not satisfy any of the above conditions, and Yukawa's[13] quantization, although formally similar to that presented in this paper, differs widely in its subsequent development.

Green's functions which are not vitiated by the presence of any singularity, are arrived from the study of scalar fields in this discrete space–time. Moreover, the plane wave associated with a free

* This paper has been prepared under Research Grant DA-ARO(D)-124-G-602, U. S. Army Research Office, Durham.

[1] W. Heisenberg, Z. Physik **65**, 4 (1930); Ann. Physik **32**, 20 (1938); Z. Physik **120**, 513, 673 (1942).

[2] V. Amberzumian and D. Ivanenko, Z. Physik **64**, 563 (1930).

[3] A. March, Z. Physik **104**, 93, 161 (1936); **105**, 620 (1937); **106**, 49, 291(1937).

[4] H. Snyder, Phys. Rev. **71**, 38 (1947); **72**, 68 (1948).

[5] A. Schild, Phys. Rev. **73**, 414 (1948); Can. J. Math. **1**, 29 (1949).

[6] Nonlocal interactions of local fields fail to satisfy the integrability conditions of Tomonaga–Schwinger equation.

[7] G. Wataghin, Z. Physik **88**, 92 (1934); **92**, 547 (1935).

[8] R. E. Peierls and H. McManus, Phys. Rev. **70**, 795 (1948).

[9] C. Bloch, Kgl. Danske Videnskab. Selskab, Mat.–Fys. Medd. **26**, no. 1 (1950); **27**, no. 8 (1952).

[10] P. Kristensen and C. Møller, Kgl. Danske Videnskab. Selskab, Mat.–Fys. Medd. **27**, no. 7 (1952).

[11] A. Pais and G. E. Uhlenbeck, Phys. Rev. **79**, 145 (1950).

[12] E. Arnous, W. Heitler, and Y. Takahashi, Nuovo Cimento **16**, 671 (1960).

[13] H. Yukawa, Phys. Rev. **76**, 300 (1949); **77**, 219, 849 (1950); **80**, 1047 (1950).

[14] A. Das, Nuovo Cimento **18**, 482 (1960); this paper contains extensive references of works on nonlocal fields.

[15] Under group

$$R_3 \sim SU_2\colon [\xi_\mu, \xi_\nu] = l\epsilon_{\mu\nu}, \qquad X^a = \sigma^{a\mu\nu}\xi_\mu\xi_\nu,$$

$$R_3 \times R_3'\colon [X^{aa'}, X^{bb'}] = l^2\epsilon^{abc}\epsilon^{a'b'c'}X_{cc'}, \quad \text{etc.}$$

particle in usual theory, is replaced by a plane wave damped with respect to distance, time, wavenumber, and frequency.

II. FIELD OPERATORS AND STATE VECTORS IN HILBERT SPACE

Units are so chosen that $\hbar = c = l = 1$ and all physical quantities are expressed as pure numbers.

Consider the inhomogeneous linear transformations

$$Z'^{k\pm} = a^{k\pm} + a^k_{\cdot l} Z^{l\pm}, \tag{2.1}$$

where $a^k_{\cdot l}$ belongs to $L_4\ddagger$. The group defined in (2.1) possesses infinite-dimensional unitary representations.[16] Consider infinite-dimensional-product Hilbert space $H_{I\,II} = H_I \times H_{II}$, where H_I and H_{II} are infinite-dimensional spaces[17] of representations of the group in (2.1). The transformations (2.1) induce in H_I, H_{II}, and $H_{I\,II}$ the following unitary transformations:

$$\Psi'_I = U_I \Psi_I, \quad \Psi'_{II} = U_{II}\Psi_{II}, \quad \Psi'_{I\,II} = U_{I\,II}\Psi_{I\,II},$$
$$U_{I\,II} = (U_I \times I_{II})(I_I \times U_{II}),$$
$$U_I^\dagger U_I = I_I, \quad U_{II}^\dagger U_{II} = I_{II}, \quad U_{I\,II}^\dagger U_{I\,II} = I_{I\,II} = I_I \times I_{II}, \tag{2.2}$$

where the vectors Ψ_I, Ψ_{II}, $\Psi_{I\,II}$ belong to H_I, H_{II}, $H_{I\,II}$, respectively, and I_I, I_{II}, $I_{I\,II}$ are identity operators.

A vector $\Psi_{I\,II} \in H_{I\,II}$ represents the state of a physical system to an observer in certain frame of reference. Fields are represented by linear operators[18] $\Phi^{k\cdots+\cdots}_{I\,II}(Z_I^{k+} \times I_{II}, Z_I^{k-} \times I_{II})$ which act in $H_{I\,II}$. $Z_I^{k\pm}$ $[= (Z_I^{k\mp})^\dagger]$ are linear operators which correspond to complex coordinates, and they act in H_I. Now a different observer in a new frame of reference characterized by the transformation (2.1), may either (1) ascribe to the physical system the same state vector $\Psi_{I\,II}$ and new field operators $\Phi'^{\cdots}(Z_I'^{+} \times I_{II}, Z_I'^{-} \times I_{II})$ or (2) describe the system by a new state vector $\Psi'_{I\,II} = U_{I\,II}\Psi_{I\,II}$ and unchanged operators $\Phi_I^{\cdots}{}_{II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II})$. Both methods must be equivalent and give the same expectation values for the physical quantities, i.e.

$$(\Psi_{I\,II}, \Phi'^{\cdots}_{I\,II}\Psi_{I\,II}) = (\Psi'_{I\,II}, \Phi_I^{\cdots}{}_{II}\Psi'_{I\,II}),$$

where $(\Psi_{I\,II}, \chi_{I\,II})$ denotes the scalar product between two vectors $\Psi_{I\,II}$ and $\chi_{I\,II}$ in $H_{I\,II}$. From the last equation it follows that

$$\Phi'^{\cdots}_{I\,II}(Z_I'^{+} \times I_{II}, Z_I'^{-} \times I_{II}) = U^\dagger_{I\,II}\Phi_I^{\cdots}{}_{II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II})U_{I\,II}. \tag{2.3}$$

The field operators $\Phi_I^{k\cdots+\cdots}{}_{II}(Z_I^+ \times I_{II}, Z^- \times I_{II})$ also transform like tensor fields under (2.1), so that

$$\Phi'^{k\cdots+\cdots-\cdots}_{I\,II}(Z_I'^{+} \times I_{II}, Z_I'^{-} \times I_{II}) = a^k_{\cdot m}\cdots\, \Phi^{m\cdots+\cdots-\cdots}_{I\,II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II}) = U^\dagger_{I\,II}\Phi^{k\cdots+\cdots-\cdots}_{I\,II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II})U_{I\,II}. \tag{2.4}$$

Expectation values of the field operators transform like classical tensor fields,

$$(\Psi'_{I\,II}, \Phi^{k\cdots}_{I\,II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II})\Psi'_{I\,II}) = (\Psi_{I\,II}, \Phi'^{k\cdots}_{I\,II}(Z_I'^{+} \times I_{II}, Z_I'^{-} \times I_{I\,II})\Psi_{I\,II}) = a^k_{\cdot m}\cdots(\Psi_{I\,II}, \Phi^{m\cdots}_{I\,II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II})\Psi_{I\,II}).$$

All the representations of U_I are one-dimensional and cannot be generated by similarity transformation like in (2.3). Thus it is not necessary to consider explicitly the group U_I in (2.1) and the covariance under U_I will be obvious from the notations.

Now, the infinitesimal form of (2.1) and the corresponding unitary representations are

$$Z_I'^{k\pm} = \epsilon^{k\pm} I_I + (\delta^k_{\cdot l} + \epsilon^k_{\cdot l})Z_I^{l\pm},$$
$$U_I = I_I + i(\epsilon^{k+}P_{Ik+} + \epsilon^{k-}P_{Ik-}) + \frac{i}{2}\epsilon^{ab}S_{Iab},$$
$$U_{II} = I_{II} + i(\epsilon^{k+}P_{IIk+} + \epsilon^{k-}P_{IIk-}) + \frac{i}{2}\epsilon^{ab}S_{IIab},$$
$$U_{I\,II} = I_I \times I_{II} + i[\epsilon^{k+}(P_{Ik+} \times I_{II} + I_I \times P_{IIk+}) + \epsilon^{k-}(P_{Ik-} \times I_{II} + I_I \times P_{IIk-})] + \frac{i}{2}\epsilon^{ab}(S_{Iab} \times I_{II} + I_I \times S_{IIab}), \tag{2.5}$$

where $|\epsilon^{k\pm}|$, $|\epsilon^{ab}|$ are small positive numbers and the linear operators $P_{Ik\pm}$, S_{Iab} act in H_I and so on.

From the first equation in (2.5),

$$Z_I'^{k\pm} - Z_I^{k\pm} = \epsilon^{k\pm} I_I + \tfrac{1}{2}\epsilon^{ab}\delta^k_{\cdot[a}\eta_{b]m}Z_I^{m\pm}. \tag{2.6}$$

Also from (2.3)

$$Z_I'^{k\pm} - Z_I^{k\pm} = U_I^\dagger Z_I^{k\pm}U_I - Z_I^{k\pm} = -i\epsilon^{l\pm}[P_{Il\pm}, Z_I^{k\pm}] - i\epsilon^{l\mp}[P_{Il\mp}, Z_I^{k\pm}] + \frac{i}{2}\epsilon^{ab}[S_{Iab}, Z_I^{k\pm}]. \tag{2.7}$$

---

[16] E. P. Wigner, Ann. Math. 40, 149 (1939).

[17] The cardinal numbers of H_I and H_{II} are, respectively, $\aleph_0$ and $\aleph_{c_0}$. The Hilbert space $H_{I\,II}$ goes beyond the classical limits. cf. J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955).

[18] The field operators in general are unbounded, though there exist infinite number of vectors like $\Psi_{I\,II} \in H_{I\,II}$ such that $\|\Phi_{I\,II}\cdots\Psi_{I\,II}\| < \infty$. $(Z_I^+ \times I_{II}) \equiv (Z_I^{1+} \times I_{II}, Z_I^{2+} \times I_{II}, Z_I^{3+} \times I_{II}, Z_I^{4+} \times I_{II})$.

Comparing (2.6) and (2.7) we obtain

$$\begin{aligned}[P_{II\pm}, Z_I^{k\pm}] &= i\delta^k_{\cdot I} I_I,\\ [P_{II\pm}, Z_I^{k\mp}] &= O_I,\\ [S_{Iab}, Z_I^{k\pm}] &= i\delta^k_{\cdot[a}\eta_{b]m} Z_I^{m\pm},\end{aligned} \tag{2.8}$$

where 0_I is the null operator acting on H_I.

Now under the infinitesimal unitary transformations the field operators transform as in (2.3), so that the local variation is

$$\begin{aligned}\delta_{I\,II}\Phi^{\cdot\cdot}_{I\,II} &\overset{\text{def}}{=} \Phi'^{\cdot\cdot}_{I\,II}(Z_I'^{+} \times I_{II}, Z_I'^{-} \times I_{II})\\ &\quad - \Phi^{\cdot\cdot}_{I\,II}(Z_I^{+} \times I_{II}, Z_I^{-} \times I_{II}),\\ &= i\epsilon^{k+}[(P_{Ik+} \times I_{II} + I_I \times P_{IIk+}), \Phi^{\cdot\cdot}_{I\,II}]\\ &\quad + i\epsilon^{k-}[(P_{Ik-} \times I_{II} + I_I \times P_{IIk-}), \Phi^{\cdot\cdot}_{I\,II}]\\ &\quad + \frac{i}{2}\epsilon^{ab}[(S_{Iab} \times I_{II} + I_I \times S_{IIab}), \Phi^{\cdot\cdot}_{I\,II}].\end{aligned} \tag{2.9}$$

But from tensorial transformations we have

$$\delta_{I\,II}\Phi^{\cdot\cdot}_{I\,II} = \tfrac{1}{2}\epsilon^{ab} S^{\cdot\cdot}_{ab\cdot\cdot}\Phi^{\cdot\cdot}_{I\,II}. \tag{2.10}$$

Comparing (2.9) and (2.10) we have

$$[P_{Ik\pm} \times I_{II}), \Phi^{\cdot\cdot}_{I\,II}] = -[(I_I \times P_{IIk\pm}), \Phi^{\cdot\cdot}_{I\,II}], \tag{2.11}$$

$$\begin{aligned}&[(S_{Iab} \times I_{II}), \Phi^{\cdot\cdot}_{I\,II}]\\ &\quad = -[(I_I \times S_{IIab}), \Phi^{\cdot\cdot}_{I\,II}] - iS^{\cdot\cdot}_{ab\cdot\cdot}\Phi^{\cdot\cdot}_{I\,II}.\end{aligned}$$

III. THE QUANTIZATION OF COMPLEX SPACE-TIME

To quantize space–time we shall now make the following "ansatz":

$$\begin{aligned}Z_I^{k\pm} &= \mp iP_I^{k\pm},\\ S_{Iab} &= iZ^-_{I[b}Z^+_{Ia]}.\end{aligned} \tag{3.1}$$

Inserting (3.1) in (2.8) we obtain

$$\begin{aligned}[Z_I^{k\pm}, Z_I^{l\mp}] &= \pm\eta^{kl} I_I,\\ [Z_I^{k\pm}, Z_I^{l\pm}] &= 0_I.\end{aligned} \tag{3.2}$$

Substituting (3.2) into (2.11) we get

$$\begin{aligned}&[(I_I \times P_{IIk\pm}), \Phi^{\cdot\cdot}_{I\,II}] = \pm i[(Z_{Ik\pm} \times I_{II}), \Phi^{\cdot\cdot}_{I\,II}],\\ &[(I_I \times S_{IIab}), \Phi^{\cdot\cdot}_{I\,II}]\\ &\quad = -i[(Z^-_{I[a}Z^+_{Ib]}) \times I_{II}, \Phi^{\cdot\cdot}_{I\,II}] - iS^{\cdot\cdot}_{ab\cdot\cdot}\Phi^{\cdot\cdot}_{I\,II}.\end{aligned} \tag{3.3}$$

These expressions may be considered as the defining relations for the total momentum and angular momentum operators $I_I \times P_{IIk+}$, $I_I \times S_{IIab}$ of the field $\Phi^{\cdot\cdot}_{I\,II}$.

To obtain the physical consequences of the quantum rule (3.2), we write

$$Z_I^{k+} = e^{-i\Theta_{Ik}}R_I^k, \quad Z_I^{k-} = R_I^k e^{i\Theta_{Ik}} \quad (k \text{ not summed}).$$

The results of measuring Cartesian coordinates of any physical object in space–time are given by the eigenvalues r^k of the operator R^k, and these are

$$r^k = \pm(n^k)^{\frac{1}{2}}, \qquad n^k = 0, 1, 2 \cdots \infty\ . \tag{3.4}$$

The measurements of four positional coordinates can be performed simultaneously. But there is an uncertainty principle

$$\Delta R^k \Delta\Theta^k \sim 1/2(n^k)^{\frac{1}{2}}, \quad (k \text{ not summed}), \tag{3.5}$$

where ΔR^k, $\Delta\Theta^k$ are, respectively, the uncertainties in the measurements of ordinary and electrical (or internal) coordinates.

We shall denote an eigenvector of the operator R_I^k by $|(n^k)_I^{\frac{1}{2}}$ and its conjugate vector by ${}_I(n^k)^{\frac{1}{2}}|$, and it follows that

$$\begin{aligned}R_I^k \mid (n^k)_I^{\frac{1}{2}} &= (n^k)^{\frac{1}{2}} \mid (n^k)_I^{\frac{1}{2}},\\ Z_I^{1+} \mid (n^1)_I^{\frac{1}{2}} &= (n^1 + 1)^{\frac{1}{2}} \mid (n^1 + 1)_I^{\frac{1}{2}},\\ Z_I^{1-} \mid (n^1)_I^{\frac{1}{2}} &= (n^1)^{\frac{1}{2}} \mid (n^1 - 1)_I^{\frac{1}{2}},\\ Z_I^{4+} \mid (n^4)_I^{\frac{1}{2}} &= (n^4)^{\frac{1}{2}} \mid (n^4 - 1)_I^{\frac{1}{2}},\\ Z_I^{4-} \mid (n^4)_I^{\frac{1}{2}} &= (n^4 + 1)^{\frac{1}{2}} \mid (n^4 + 1)_I^{\frac{1}{2}},\\ [{}_I(n^k)^{\frac{1}{2}} \mid (m^k)_I^{\frac{1}{2}}] &= \delta_{nm} \quad (k \text{ not summed}),\end{aligned} \tag{3.6}$$

Now the notations are elaborated. $I_I^1 \times Z_I^{2\pm} \times I_I^3 \times I_I^4$ is the full notation for $Z_I^{2\pm}$ and so on. For the simultaneous eigenstates of position operators R_I^1 R_I^2, R_I^3, R_I^4, the notation $|\,(n)_I^{\frac{1}{2}} = |\,(n^1)_I^{\frac{1}{2}} \times |\,(n^2)_I \times |\,(n^3)_I^{\frac{1}{2}} \times (n^4)_I^{\frac{1}{2}}$ is used and equations like $R_I^3 \mid (n)_I^{\frac{1}{2}} = (n^3)^{\frac{1}{2}} \mid (n)_I^{\frac{1}{2}}$ would really mean

$$\begin{aligned}&(I_I^1 \times I_I^2 \times R_I^3 \times I_I^4)\ |(n^1)_I^{\frac{1}{2}} \times |(n^2)_I^{\frac{1}{2}} \times (n^3)_I^{\frac{1}{2}} \times (n^4)_I^{\frac{1}{2}}\\ &\qquad = (n^3)_I^{\frac{1}{2}}\ |(n^1)_I^{\frac{1}{2}} \times (n^2)_I^{\frac{1}{2}} \times (n^3)_I^{\frac{1}{2}} \times (n^4)_I^{\frac{1}{2}}.\end{aligned}$$

Now from (3.6) it follows that

$$\begin{aligned}Z_I^{1-}\ |(0)_I^{\frac{1}{2}} &= 0, \quad (Z_I^{1-})^r\ |(0)_I^{\frac{1}{2}} = 0,\\ [{}_I(n^1)^{\frac{1}{2}}|\ (Z^{1-})^r\ |-(m^1)_I^{\frac{1}{2}}] &= 0,\\ [{}_I(n)^{\frac{1}{2}}\ |\Phi^{\cdot\cdot}_{I\,II}(Z_I^{+} \times I_{II}, Z_I^{-} \times I_{II})\ |-(m)_I^{\frac{1}{2}}] &= 0.\end{aligned} \tag{3.7}$$

The meaning of equation (3.7) is that physical effects in positive lattices cannot cross the origin to negative lattices. An observer measuring the lattices given by (3.4) and also verifying (3.7) might apparently conclude that the microstructure of space–time shows preferred origin, preferred axes and also preferred positive or, negative cones. He may wonder what happened to relativity! But this apparent

contradiction may be resolved in the following way. To another observer the eigenvalues of the position operator $R_I'^k = U_I^\dagger R_I^k U_I$ are the same as in (3.4), so that the measurable microstructure, after all, remains invariant to all observers.

IV. THE LAGRANGIAN FORMALISM

Let us define the Lagrangian operator $L_{I\,II}$ constructed out of the field operators $\Phi_I^{\cdot\cdot}{}_{II}$, $\Phi_I^{\dagger\cdot\cdot}{}_{II}$ and their first commutators $[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]$, $[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]$ in the following way:

$$L_{I\,II} \overset{\text{def}}{=} L\{\Phi_I^{\cdot\cdot}{}_{II}, \Phi_I^{\dagger\cdot\cdot}{}_{II}, [(Z_I^{k\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}], [Z_I^{k\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]\}. \tag{4.1}$$

We define the Action operator as[19]

$$A_{II} \overset{\text{def}}{=} \sum_n^4 \left[{}_I(n)^{\frac{1}{2}} | L_{I\,II} | (n)_I^{\frac{1}{2}} \right]. \tag{4.2}$$

The variational principle states that

$$\delta A_{II} \Psi_{II} = 0_{II}, \tag{4.3}$$

for all $\Psi_{II} \in H_{II}$ with $||\Psi|| < \infty$. (4.3) yields the following Euler–Lagrange operator equations[20]

$$\begin{aligned}
&\left[(Z_I^{k+} \times I_{II}), \frac{\partial L}{\partial [(Z_I^{k+} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} \right] \\
&\quad + \left[(Z_I^{k-} \times I_{II}), \frac{\partial L}{\partial [(Z_I^{k-} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} \right] \\
&\quad - \frac{\partial L}{\partial \Phi_I^{\cdot\cdot}{}_{II}} = 0_{I\,II}, \\
&\left[(Z_I^{k+} \times I_{II}), \frac{\partial L}{\partial [(Z_I^{k+} I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]} \right] \\
&\quad + \left[(Z_I^{k-} \times I_{II}), \frac{\partial L}{\partial [(Z_I^{k-} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]} \right] \\
&\quad - \frac{\partial L}{\partial \Phi_I^{\dagger\cdot\cdot}{}_{II}} = 0_{I\,II}.
\end{aligned} \tag{4.4}$$

We shall now derive the Noether's theorems from the invariance of Lagrangian under (2.5). For this purpose let us define all possible variations of the field operators, namely

$$\begin{aligned}
\delta_{I\,II}\Phi_I^{\cdot\cdot}{}_{II} &\overset{\text{def}}{=} U_{I\,II}^\dagger \Phi_I^{\cdot\cdot}{}_{II} U_{I\,II} - \Phi_I^{\cdot\cdot}{}_{II} \\
&= -i[\{\epsilon^{k+}(P_{Ik+} \times I_{II} + I_I \times P_{IIk+}) \\
&\quad + \epsilon^{k-}(P_{Ik-} \times I_{II} + I_I \times P_{IIk-}) \\
&\quad + \tfrac{1}{2}\epsilon^{ab}(S_{Iab} \times I_{II} + I_I \times S_{IIab})\}, \Phi_I^{\cdot\cdot}{}_{II}], \\
\delta_{II}\Phi_I^{\cdot\cdot}{}_{II} &\overset{\text{def}}{=} (I_I \times U_{II}^\dagger)\Phi_I^{\cdot\cdot}{}_{II}(I_I \times U_{II}) - \Phi_I^{\cdot\cdot}{}_{II} \\
&= i[(\epsilon^{k+}P_{Ik+} + \epsilon^{k-}P_{Ik-} + \tfrac{1}{2}\epsilon^{ab}S_{Iab}) \times I_{II}, \Phi_I^{\cdot\cdot}{}_{II}] \\
&\quad + \tfrac{1}{2}\epsilon^{ab}S_{ab}^{\cdot\cdot}{}_{\cdot\cdot}\Phi_I^{\cdot\cdot}{}_{II}, \\
&\qquad \delta_{II}(Z_I^{k\pm} \times I_{II}) = 0_{I\,II}, \\
\delta_I\Phi_I^{\cdot\cdot}{}_{II} &\overset{\text{def}}{=} (U_I^\dagger \times I_{II})\Phi_I^{\cdot\cdot}{}_{II}(U_I \times I_{II}) - \Phi_I^{\cdot\cdot}{}_{II} \\
&= -i[(\epsilon^{k+}P_{Ik+} + \epsilon^{k-}P_{Ik-} + \tfrac{1}{2}\epsilon^{ab}S_{Iab}) \times I_{II}, \Phi_I^{\cdot\cdot}{}_{II}].
\end{aligned} \tag{4.5}$$

Now the invariance of the Lagrangian operator implies

$$\begin{aligned}
0_{I\,II} &= U_{I\,II}^\dagger L_{I\,II} U_{I\,II} - L_{I\,II} = \delta_{I\,II} L_{I\,II} \\
&= \delta_I L_{I\,II} + \delta_{II} L_{I\,II} = \{-i[(\epsilon^{k+}P_{Ik+} + \epsilon^{k-}P_{Ik-} \\
&\quad + \tfrac{1}{2}\epsilon^{ab}S_{Iab}) \times I_{II}, L_{I\,II}]\} \\
&\quad + \left\{ \frac{\partial L_{I\,II}}{\partial \Phi_I^{\cdot\cdot}{}_{II}} \delta_{II}\Phi_I^{\cdot\cdot}{}_{II} + \frac{\partial L_{I\,II}}{\partial \Phi_I^{\dagger\cdot\cdot}{}_{II}} \delta_{II}\Phi_I^{\dagger\cdot\cdot}{}_{II} \right. \\
&\quad + \frac{\partial L_{I\,II}}{\partial [(Z_I^{k+} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} [(Z_I^{k+} \times I_{II}), \delta_{II}\Phi_I^{\cdot\cdot}{}_{II}] \\
&\quad + \frac{\partial L_{I\,II}}{\partial [(Z_I^{k-} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} [(Z_I^{k-} \times I_{II}), \delta_{II}\Phi_I^{\cdot\cdot}{}_{II}] \\
&\quad + \frac{\partial L_{I\,II}}{\partial [(Z_I^{k+} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]} [(Z_I^{k+} \times I_{II}), \delta_{II}\Phi_I^{\dagger\cdot\cdot}{}_{II}] \\
&\quad \left. + \frac{\partial L_{I\,II}}{\partial [(Z_I^{k-} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]} [(Z_I^{k-} \times I_{II}), \delta_{II}\Phi_I^{\dagger\cdot\cdot}{}_{II}] \right\}.
\end{aligned} \tag{4.6}$$

With (3.1), (4.4), and (4.5) the coefficients of $\epsilon^{k\pm}$ in (4.6) yield the conservation of the canonical energy–momentum–stress tensor operator in the following way

$$\begin{aligned}
&[(Z_I^{k-} \times I_{II}), T_{I\,IIk-m\pm}] \\
&\qquad + [(Z_I^{k+} \times I_{II}), T_{I\,IIk+m\pm}] = 0_{I\,II}, \\
T_{I\,IIk\mp m\pm} &\overset{\text{def}}{=} \frac{\partial L_{I\,II}}{\partial [(Z_I^{k\mp} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} \\
&\times [(Z_{Im\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}] + \frac{\partial L_{I\,II}}{\partial [(Z_I^{k\mp} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]} \\
&\times [(Z_{Im\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}] - \eta_{km} L_{I\,II}, \\
T_{I\,IIk\pm m\pm} &\overset{\text{def}}{=} \frac{\partial L_{I\,II}}{\partial [(Z_I^{k\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} [(Z_{Im\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}] \\
&+ \frac{\partial L_{I\,II}}{\partial [(Z_I^{k\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]} [(Z_{Im\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}].
\end{aligned} \tag{4.7}$$

[19] Trace is invariant under similarity transformations and replaces Lorentz-invariant space–time volume integration;

$$\sum_n^4 \overset{\text{def}}{=} \sum_{n^1} \sum_{n^2} \sum_{n^3} \sum_{n^4}.$$

[20] Cf. C. Gregory, Phys. Rev. 91, 770 (1952); 92, 1554 (1953).

Similarly, the coefficients of ϵ^{ab} in (4.6) give the conservation of angular momentum tensor operator in the following form

$$[(Z_I^{k+} \times I_{II}), J_{I\ IIabk+}] + [(Z_I^{k-} \times I_{II}), J_{I\ IIabk-}] = 0_{I\ II},$$

$$J_{I\ IIabk\pm} \overset{\text{def}}{=} \frac{\partial L_{I\ II}}{\partial[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]} \times ([(Z_{I[b}^{-} Z_{Ia]}^{+}) \times I_{II}, \Phi_I^{\cdot\cdot}{}_{II}] + S_{ab}^{\cdot\cdot}{}_{\cdot\cdot}\Phi_I^{\cdot\cdot}{}_{II}) + \frac{\partial L_{I\ II}}{\partial[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]}\{[(Z_{I[b}^{-} Z_{Ia]}^{+}) \times I_{II}, \Phi_I^{\dagger\cdot\cdot}{}_{II}] + S_{ab}^{\cdot\cdot}{}_{\cdot\cdot}\Phi_I^{\dagger\cdot\cdot}{}_{II}\} + [(\eta_{k[a} Z_{Ib]\pm}) \times I_{II}]L_{I\ II}. \quad (4.8)$$

By evoking invariance of $L_{I\ II}$ under U_I, we can derive the conservation of electrical charge–current vector operator as following

$$[(Z_I^{k+} \times I_{II}), j_{I\ IIk+}] + [(Z_I^{k-} \times I_{II}), j_{I\ IIk-}] = 0_{I\ II},$$

$$j_{I\ IIk\pm} \overset{\text{def}}{=} \pm i\epsilon\Big[\frac{\partial L_{I\ II}}{\partial[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]}\{\tau\Phi_I^{\cdot\cdot}{}_{II} - [(Z_{II-} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}](Z_I^{l-} \times I_{II}) + [(Z_{II+} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}](Z_I^{l+} \times I_{II})\} - \frac{\partial L_{I\ II}}{\partial[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]}\{\tau\Phi_I^{\dagger\cdot\cdot}{}_{II} - [(Z_{II-} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}](Z_I^{l-} \times I_{II}) + [(Z_{II+} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}] \times (Z_I^{l+} \times I_{II})\} + L_{I\ II}(Z_{Ik\pm} \times I_{II})\Big]. \quad (4.9)$$

If we demand invariance of $L_{I\ II}$ under infinitesimal phase transformation

$$\Phi'^{\cdot\cdot}{}_{II} = \Phi_I^{\cdot\cdot}{}_{II}e^{i\alpha} \simeq \Phi_I^{\cdot\cdot}{}_{II}(1 + i\alpha),$$
$$\Phi_I^{\dagger\prime\cdot\cdot}{}_{II} = \Phi_I^{\dagger\cdot\cdot}{}_{II}e^{-i\alpha} \simeq \Phi_I^{\dagger\cdot\cdot}{}_{II}(1 - i\alpha).$$

Then we obtain the conservation of number-flux vector operator as,

$$[(Z_I^{k+} \times I_{II}), N_{I\ IIk+}] + [(Z_I^{k-} \times I_{II}), N_{I\ IIk-}] = 0_{I\ II},$$

$$N_{I\ IIk\pm} \overset{\text{def}}{=} \pm i\Big[\frac{\partial L_{I\ II}}{\partial[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\cdot\cdot}{}_{II}]}\Phi_I^{\cdot\cdot}{}_{II} - \frac{\partial L_{I\ II}}{\partial[(Z_I^{k\pm} \times I_{II}), \Phi_I^{\dagger\cdot\cdot}{}_{II}]}\Phi_I^{\dagger\cdot\cdot}{}_{II}\Big] \quad (4.10)$$

From (4.7), (4.8), (4.9), and (4.10), we obtain the expressions of the total energy–momentum, angular-momentum, charge, number operators of the field as[21]

$$P_{IIk\pm} \overset{\text{def}}{=} (n^4 + 1)^{\frac{1}{2}} \sum_{\mathbf{n}} \{({}_I(n^4 + 1)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| T_{4+k\pm} |(\mathbf{n})^{\frac{1}{2}}, (n^4)_I^{\frac{1}{2}}) - ((n^4)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| T_{I\ II4-k\pm} |(\mathbf{n})^{\frac{1}{2}}, (n^4 + 1)_I^{\frac{1}{2}})_I\},$$

$$S_{IIab} \overset{\text{def}}{=} (n^4 + 1)^{\frac{1}{2}} \sum_{\mathbf{n}} \{({}_I(n^4 + 1)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| J_{I\ IIab4+} |(\mathbf{n})^{\frac{1}{2}}, (n^4)_I^{\frac{1}{2}}) - ({}_I(n^4)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| J_{I\ IIab4-} |(\mathbf{n})^{\frac{1}{2}}, (n^4 + 1)_I^{\frac{1}{2}})\},$$

$$Q_{II} \overset{\text{def}}{=} (n^4 + 1)^{\frac{1}{2}} \sum_{\mathbf{n}} \{({}_I(n^4 + 1)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| j_{I\ II4+} |(\mathbf{n})^{\frac{1}{2}}, (n^4)_I^{\frac{1}{2}}) - ({}_I(n^4)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| j_{I\ II4-} |(\mathbf{n})^{\frac{1}{2}}, (n^4 + 1)_I^{\frac{1}{2}})\},$$

$$N_{II} \overset{\text{def}}{=} (n^4 + 1)^{\frac{1}{2}} \sum_{\mathbf{n}} \{({}_I(n^4 + 1)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| N_{I\ II4+} |(\mathbf{n})^{\frac{1}{2}}, (n^4)_I^{\frac{1}{2}}) - ({}_I(n^4)^{\frac{1}{2}}, (\mathbf{n})^{\frac{1}{2}}| N_{I\ II4-} |(\mathbf{n})^{\frac{1}{2}}, (n^4 + 1)_I^{\frac{1}{2}})\}. \quad (4.11)$$

These quantities do not change with time lattice provided there is no flux of the field across spatial boundary lattices. Incidentally, $P_{IIk\pm}$, S_{IIab} in (4.11) should be identified with those occurring in (3.3).

V. THE QUANTIZATION OF COMPLEX SCALAR FIELDS

We choose the example of complex scalar fields to demonstrate the formalism developed in previous sections. The Lagrangian operator in this case, is taken to be,

$$L_{I\ II} = \eta_{kl}[(Z_I^{k+} \times I_{II}), \Phi_{I\ II}^{+}] \times [(Z_I^{l-} \times I_{II}), \Phi_{I\ II}^{-}] + m^2\Phi_{I\ II}^{+}\Phi_{I\ II}^{-}. \quad (5.1)$$

The field Eqs. (4.4) derived from (5.1) are

$$\eta_{kl}[(Z^{k+} \times I_{II}), [(Z_I^{l-} \times I_{II}), \Phi_{I\ II}^{\pm}]] - m^2\Phi_{I\ II}^{\pm} = 0_{I\ II}. \quad (5.2)$$

Taking the expectation value of (5.2) in the state $|n_I$ we obtain the following partial difference equations for the field operators $\Phi_{I\ II}^{\pm}$

$$(\blacksquare + m^2)\Phi_{II}^{\pm}(n^{\frac{1}{2}}) \overset{\text{def}}{=} (2n^1 + 1)\Phi_{II}^{\pm}(n)^{\frac{1}{2}} - (n^1 + 1)\Phi_{II}^{\pm}[(n^1 + 1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}] - n^1\Phi_{II}^{\pm}[(n^1 - 1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}] + (2n^2 + 1)\Phi_{II}^{\pm}(n^{\frac{1}{2}}) - (n^2 + 1)\Phi_{II}^{\pm}[(n^1)^{\frac{1}{2}}, (n^2 + 1)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}]$$

[21] $|(\mathbf{n})_I^{\frac{1}{2}} \equiv |(n^1)_I^{\frac{1}{2}} \times |(n^2)_I^{\frac{1}{2}} \times |(n^3)_I^{\frac{1}{2}}$.

$$
\begin{aligned}
&- n^2\Phi^{\pm}_{II}[(n^1)^{\frac{1}{2}}, (n^2 - 1)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}] \\
&+ (2n^3 + 1)\Phi^{\pm}_{II}(n^{\frac{1}{2}}) \\
&- (n^3 + 1)\Phi^{\pm}_{II}[(n^1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3 + 1)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}] \\
&- n^3\Phi^{\pm}_{II}[(n^1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3 - 1)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}] \\
&- (2n^4 + 1)\Phi^{\pm}_{II}(n^{\frac{1}{2}}) \\
&+ (n^4 + 1)\Phi^{\pm}_{II}[(n^1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4 + 1)^{\frac{1}{2}}] \\
&+ n^4\Phi^{\pm}_{II}[(n^1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4 - 1)^{\frac{1}{2}}] \\
&+ m^2\Phi^{\pm}_{II}(n^{\frac{1}{2}}) = O_{II}, \qquad (5.3)
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi^{\pm}_{II}(n^{\frac{1}{2}}) &\equiv \Phi^{\pm}_{II}[(n^1)^{\frac{1}{2}}, (n^2)^{\frac{1}{2}}, (n^3)^{\frac{1}{2}}, (n^4)^{\frac{1}{2}}] \\
&\overset{\text{def}}{=} {}_I(n)^{\frac{1}{2}}|\ \Phi^{\pm}_{I\,II}(Z^{+}_{I} \times I_{II}, Z^{-}_{I} \times I_{II})|\ (n)^{\frac{1}{2}}{}_I
\end{aligned}
$$

and physically represents the field operators at the space–time lattice $(n^1)^{\frac{1}{2}}$, $(n^2)^{\frac{1}{2}}$, $(n^3)^{\frac{1}{2}}$, $(n^4)^{\frac{1}{2}}$.

From (4.11) and (5.1) we obtain expressions for the total energy–momentum charge and number operators of the field as (calculated at $n^4 = 0$)

$$
\begin{aligned}
P_{IIk+} &= \sum_{\mathbf{n}} [-({}_I0, (\mathbf{n})^{\frac{1}{2}}|\ ([(Z^{4+}_{I} \times I_{II}), \Phi^{+}_{I\,II}] \\
&\times [(Z_{Ik+} \times I_{II}), \Phi^{-}_{I\,II}] - \eta_{4k}L_{I\,II})|\ (\mathbf{n})^{\frac{1}{2}}, 1_I) \\
&+ ({}_I1, (\mathbf{n})^{\frac{1}{2}}|\ [(Z^{4-}_{I} \times I_{II}), \Phi^{-}_{I\,II}] \\
&\times [(Z_{Ik+} \times I_{II}), \Phi^{+}_{I\,II}]|\ (\mathbf{n})^{\frac{1}{2}}, 0_I)], \\
P_{IIk-} &= \sum_{\mathbf{n}} [-({}_I0, (\mathbf{n})^{\frac{1}{2}}|\ [(Z^{4+}_{I} \times I_{II}), \Phi^{+}_{I\,II}] \\
&\times [(Z_{Ik-} \times I_{II}), \Phi^{-}_{I\,II}]|\ (\mathbf{n})^{\frac{1}{2}}, 1_I) \\
&+ ({}_I1, (\mathbf{n})^{\frac{1}{2}}|\ ([(Z^{4-} \times I_{II}), \Phi^{-}_{I\,II}] \\
&\times [(Z_{Ik-} \times I_{II}), \Phi^{+}_{I\,II}] - \eta_{4k}L_{I\,II})|\ (n)^{\frac{1}{2}}, 0_I)],
\end{aligned}
$$

$$
\begin{aligned}
S_{IIab} &= \sum_{\mathbf{n}} [-({}_I0, (\mathbf{n})^{\frac{1}{2}}|\ [(Z^{4+}_{I} \times I_{II}), \Phi^{+}_{I\,II}] \\
&\times [(Z^{-}_{I[b}Z^{+}_{Ia]}) \times I_{II}, \Phi^{-}_{I\,II}] \\
&+ (\eta_{4[a}Z_{Ib]-}) \times I_{II}L_{I\,II}|\ (\mathbf{n})^{\frac{1}{2}}, 1_I) \\
&+ ({}_I1, (\mathbf{n})^{\frac{1}{2}}|\ [(Z^{4-}_{I} \times I_{II}), \Phi^{-}_{I\,II}] \\
&\times [(Z^{-}_{I[b}Z^{+}_{Ia]}) \times I_{II}, \Phi^{+}_{II}] \\
&+ (\eta_{4[a}Z_{Ib]+}) \times I_{II}L_{I\,II}|\ (\mathbf{n})^{\frac{1}{2}}, 0_I)], \\
Q_{II} &= -i\epsilon \sum_{\mathbf{n}} [-({}_I0, (\mathbf{n})^{\frac{1}{2}}|\ ([(Z^{4+}_{I} \times I_{II}), \Phi^{+}_{I\,II}] \\
&\times [\Phi^{-}_{I\,II} - [(Z_{Ii-} \times I_{II}), \Phi^{-}_{I\,II}](Z^{i-}_{I} \times I_{II}) \\
&+ [(Z_{Ii+} \times I_{II}), \Phi^{-}_{I\,II}](Z^{i+}_{I} \times I_{II})] \\
&+ L_{I\,II}(Z_{I4-} \times I_{II}))|\ (\mathbf{n})^{\frac{1}{2}}, 1_I) \\
&+ ({}_I1, (\mathbf{n})^{\frac{1}{2}}|\ ([Z^{4-}_{I} \times I_{II}), \Phi^{-}_{I\,II}] \\
&\times [\Phi^{+}_{I\,II} - [(Z_{Ii-} \times I_{II}), \Phi^{+}_{I\,II}](Z^{i-}_{I} \times I_{II}), \\
&+ [Z_{Ii+} \times I_{II}, \Phi^{+}_{II}][((Z^{i+} \times I_{II})] \\
&+ L_{I\,II}(Z_{I4+} \times I_{II}))|\ (\mathbf{n})^{\frac{1}{2}}, 0_I)], \\
N_{II} &= -i \sum_{\mathbf{n}} [-({}_I0, (\mathbf{n})^{\frac{1}{2}}|\ [(Z^{4+}_{I} \\
&\times I_{II}), \Phi^{-}_{I\,II}]\Phi^{+}|\ (\mathbf{n})^{\frac{1}{2}}, 1_I) + ({}_I1, (\mathbf{n})^{\frac{1}{2}}|\ [(Z^{4-}_{I} \\
&\times I_{II}), \Phi^{+}_{I\,II}]\Phi^{-}_{I\,II}|\ (\mathbf{n})^{\frac{1}{2}}, 0_I)]. \qquad (5.4)
\end{aligned}
$$

Now let us make slight digression to consider the partial difference equation

$$(\blacksquare + m^2)\Phi(n^{\frac{1}{2}}) = 0, \qquad (5.5)$$

where the operation $\blacksquare$ has been defined in (5.3) and $\Phi(n^{\frac{1}{2}})$ is a bounded function of four real discrete variables. The double Green's functions corresponding to (5.5) can be defined as following

$$
\begin{aligned}
G_{(a)}[(n)^{\frac{1}{2}}, (m)^{\frac{1}{2}}] &\overset{\text{def}}{=} \iiint_0^{\infty} \int_{c_{(a)}} dk_1\, dk_2\, dk_3\, dk_4 \\
&\times \frac{e^{-(k_1+k_2+k_3+k_4)}L_{n^1}(k_1)L_{n^2}(k_2)L_{n^3}(k_3)L_{n^4}(k_4)L_{m^1}(k_1)L_{m^2}(k_2)L_{m^3}(k_3)L_{m^4}(k_4)}{k_1 + k_2 + k_3 - k_4 + m^2}, \qquad (5.6)
\end{aligned}
$$

where $L_n(k)$ is the Laguerre polynomial and $c_{(a)}$ defines different contours in the complex k_4-plane as shown in Fig. 1, one singularity being situated at $k_4 = k_1 + k_2 + k_3 + m^2$.

The Green's function $G[(n)^{\frac{1}{2}}, (m)^{\frac{1}{2}}]$ corresponding to the closed contour C is homogeneous and others corresponding to open contours are inhomogeneous and they satisfy, respectively,[22]

[22] This can easily be seen from the recurrence relation

$$(2n + 1)L_n(k) - (n + 1)L_{n+1}(k) - nL_{n-1}(k) = kL_n(k)$$

and the orthonormal property

$$\int_0^{\infty} e^{-k}L_n(k)L_m(k)\, dk = \delta_{nm};\ \delta^4_{nm} \overset{\text{def}}{=} \delta_{n^1m^1}\delta_{n^2m^2}\delta_{n^3m^3}\delta_{n^4m^4}.$$

$$(\blacksquare + m^2)G[(n)^{\frac{1}{2}}, (m)^{\frac{1}{2}}] = 0,$$
$$(\blacksquare + m^2)G_{(\omega)}[(n)^{\frac{1}{2}}, (m)^{\frac{1}{2}}] = \delta^{(4)}_{nm}. \tag{5.7}$$

It can be shown that these Green's functions are nonsingular. For subsequent use we note that the homogeneous Green's function is

$$G[(n^{\frac{1}{2}}), (m^{\frac{1}{2}})] = -2\pi i e^{-m^2} \iiint_0^\infty dk_1\, dk_2\, dk_3\, e^{-2(k_1+k_2+k_3)} L_{n^1}(k_1)L_{n^2}(k_2)L_{n^3}(k_3)$$
$$\times L_{n^4}(k_1 + k_2 + k_3 + m^2)L_{m^1}(k_1)L_{m^2}(k_2)L_{m^3}(k_3)L_{m^4}(k_1 + k_2 + k_3 + m^2),$$
$$= -\frac{\pi i}{4} e^{-m^2} \sum_{S_1=0}^{n^1}\sum_{S_2=0}^{n^2}\sum_{S_3=0}^{n^3}\sum_{S_4=0}^{n^4}\sum_{\sigma_1=0}^{m^1}\sum_{\sigma_2=0}^{m^2}\sum_{\sigma_3=0}^{m^3}\sum_{\sigma_4=0}^{m^4}\sum_{r=0}^{S_4+\sigma_4}\sum_{S=0}^{S_4+\sigma_4-r}\sum_{t=0}^{r} (-\tfrac{1}{2})(S_1 + \sigma_1 + S_2 + \sigma_2 + S_3 + \sigma_3$$
$$+ S_4 + \sigma_4)2^{(t-r)}\binom{n_1}{S_1}\binom{n_2}{S_2}\binom{n_3}{S_3}\binom{n_4}{S_4}\binom{m_1}{\sigma_1}\binom{m_2}{\sigma_2}\binom{m_3}{\sigma_3}\binom{m_4}{\sigma_4}\binom{S_4+\sigma_4}{r}\binom{S_4+\sigma_4-r}{S}\binom{r}{t}$$
$$\times (S_1 + \sigma_1 + S_4 + \sigma_4 - S - 1)!\,(S_2 + \sigma_2 + S - 1)!\,(S_3 + \sigma_3 - t - 1)!, \quad |G[(n)^{\frac{1}{2}}, (m)^{\frac{1}{2}}]| < \infty. \tag{5.8}$$

Before we pass on to the Fourier expansion of the field operators we notice

$$e^{i[p_{a+}(Z_I^{a+}\times I_{II}) + p_{a-}(Z_I^{a-}\times I_{II})]}$$
$$= e^{i[p_{a-}(Z_I^{a-}\times I_{II}) + p_{a+}(Z_I^{a+}\times I_{II})]}$$
$$= e^{-\frac{1}{2}[|p_1|^2+|p_2|^2+|p_3|^2+|p_4|^2]}$$
$$\times e^{ip_{a+}(Z_I^{a+}\times I_{II})}\cdot e^{ip_{a-}(Z_I^{a-}\times I_{II})} \tag{5.9}$$

where $p_{a\pm}$'s are ordinary c-numbers, and we have used the Baker–Hausdorff theorem.[23] Moreover from (3.6) we can obtain[24]

$$({}_I(n^{\frac{1}{2}})|\, e^{ip_{a+}(Z_I^{a+}\times I_{II})} e^{ip_{a-}(Z_I^{a-}\times I_{II})}\, |(n)^{\frac{1}{2}}_I)$$
$$= L_{n^1}(|p_1|^2)L_{n^2}(|p_2|^2)L_{n^3}(|p_3|^2)L_{n^4}(|p_4|^2)I_{II}. \tag{5.10}$$

[23] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press Inc., New York, 1963).

[24] From (3.6).

$$(Z_I^{1-})^r\, |(n^1)^{\frac{1}{2}}_I = [n^1(n^1 - 1)\cdots(n^1 - r + 1)]^{\frac{1}{2}}|\,(n^1 - r)^{\frac{1}{2}}_I,$$
$$(Z_I^{1+})^s\, |(n^1)^{\frac{1}{2}}_I = [(n^1 + 1)(n^1 + 2)\cdots(n^1 + s)]^{\frac{1}{2}}|\,(n^1 + s)^{\frac{1}{2}}_I,$$

so that

$$\left({}_I(n^1)^{\frac{1}{2}}\left|\sum_r\sum_s \frac{(ip_{1+})^s}{s!}\frac{(ip_{1-})^r}{r!}(Z_I^{1+})^s(Z_I^{1-})^r\right|(n^1)^{\frac{1}{2}}_I\right)$$
$$= \sum_r\sum_s \frac{(ip_{1+})^s}{s!}\frac{(ip_{1-})^r}{r!}[n^1\cdots(n^1 - r + 1)$$
$$\times (n^1 - r + 1)\cdots(n^1 - r + s)]^{\frac{1}{2}}\,\delta_{rs}$$
$$= n^1!\sum_r \frac{(-|p_1|^2)^r}{(r!)^2(n^1 - r)!} = L_{n^1}(|p_1|^2).$$

We assume that our field operators possess Fourier transform

$$\Phi^{\pm}_{I\,II}(Z_I^+ \times I_{II}, Z_I^- \times I_{II})$$
$$= \frac{1}{(2\pi)^4}\int\cdots\int d^4p_+\, d^4p_-[I_I \times B^{\pm}_{II}(p_+, p_-)]$$
$$\times e^{\pm i[p_{a+}(Z_I^{a+}\times I_{II}) + p_{a-}(Z_I^{a-}\times I_{II})]}. \tag{5.11}$$

Because $\Phi^{\pm}_{I\,II}$ have to satisfy (5.2), we must have

$$B^{\pm}_{II}(p_+, p_-) = 2\pi\,\delta(-\eta^{ab}p_{a+}p_{b-} + m^2)A^{\pm}_{II}(p_+, p_-). \tag{5.12}$$

The expectation values of the field operators are

$$\Phi^{\pm}_{II}(n^{\frac{1}{2}}) = ({}_I(n)^{\frac{1}{2}}|\,\Phi^{\pm}_{I\,II}\,|(n)^{\frac{1}{2}})_I$$
$$= \frac{(2\pi)^{\frac{1}{2}}}{(2\pi)^4}\int\cdots\int d^4p_+\, d^4p_-A^{\pm}_{II}(p_+, p_-)$$
$$\times \delta(-\eta^{ab}p_{a+}p_{b-} + m^2)e^{-\frac{1}{2}(|p_1|^2+|p_2|^2+|p_3|^2+|p_4|^2)}$$
$$\times L_{n^1}(|p_1|^2)L_{n^2}(|p_2|^2)L_{n^3}(|p_3|^2)L_{n^4}(|p_4|^2), \tag{5.12'}$$

by virtue of (5.9), (5.10), (5.11), and (5.12).

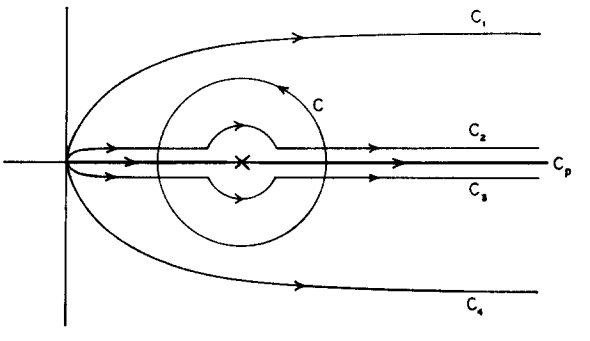

FIG. 1. Complex k_4-plane.

Denoting

$$p_{4\pm} = |p_4|\, e^{\pm i\theta_4} \text{ and } \omega^2 = |p_1|^2 + |p_2|^2 + |p_3|^2 + m^2,$$

(5.12′) can be rewritten as

$$\Phi^{\pm}_{II}(n^{\frac{1}{4}}) = \frac{(2\pi)^{\frac{1}{2}}}{(2\pi)^4} e^{-\frac{1}{2}m^2} \int \cdots \int d^3p_+\, d^3p_- \times \int d\theta_4 A^{\pm}_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4}) e^{-(|p_1|^2+|p_2|^2+|p_3|^2)} \times L_{n^1}(|p_1|^2) L_{n^2}(|p_2|^2) L_{n^3}(|p_3|^2) L_{n^4}(\omega^2). \tag{5.13}$$

Now we postulate the amplitude quantization

$$[A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4}),\ A^+_{II}(\mathbf{p}'_\pm, \omega' e^{\pm i\theta_4'}] = I_{II}\delta(|p_{1\pm} - p'_{1\pm}|^2) \times \delta(|p_{2\pm} - p'_{2\pm}|^2)\delta(|p_{3\pm} - p'_{3\pm}|^2) f(\theta_4, \theta'_4), \tag{5.14}$$

$$[A^-_{II}, A^-_{II}] = [A^+_{II}, A^+_{II}] = 0_{II},$$

$$\int_0^{2\pi}\int_0^{2\pi} f(\theta_4, \theta'_4)\, d\theta_4\, d\theta'_4 = (2\pi)^2.$$

As the consequence of this and (5.13), (5.8), one obtains

$$[\Phi^-_{II}(n^{\frac{1}{4}}), \Phi^+_{II}(n'^{\frac{1}{4}})] = iI_{II}G[(n)^{\frac{1}{4}}, (n')^{\frac{1}{4}}], \tag{5.15}$$

$$[\Phi^-_{II}(n^{\frac{1}{4}}), \Phi^-_{II}(n'^{\frac{1}{4}})] = \Phi^+_{II}(n^{\frac{1}{4}}), \Phi^+_{II}(n'^{\frac{1}{4}})] = 0_{II}.$$

The quantization (5.15) is consistent with field equations (5.2) because

$$0_{II} = [(\blacksquare + m^2)\Phi^{\pm}_{II}(n^{\frac{1}{4}}), \Phi^{\mp}_{II}(n'^{\frac{1}{4}})] = \mp iI_{II}(\blacksquare + m^2)G[(n)^{\frac{1}{4}}, (n')^{\frac{1}{4}}] = 0_{II},$$

from (5.12) and (5.7).

Straightforward calculation for energy and number operators from (5.4) and (5.13), yields

$$P_{II4+} = H_{II+} = \frac{1}{(2\pi)^4}\int\cdots\int\cdots d^3\mathbf{p}_+\, d^3\mathbf{p}_-\omega^3 e^{-\omega^2} \times \iint d\theta_4\, d\theta'_4[A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4}) \times A^+_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4'}) + A^+_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4}) \times A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4'})](e^{i\theta_4} - e^{i\theta_4'})e^{\omega^2 e^{i(\theta_4-\theta_4')}},$$

$$P_{4-} = H_{II-} = \frac{1}{(2\pi)^4}\int\cdots\int\cdots d^3\mathbf{p}_+\, d^3\mathbf{p}_-\omega^3 e^{-\omega^2} \times \iint d\theta_4\, d\theta'_4[A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4}) \times A^+_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4'})e^{i(\theta_4-\theta_4')} + A^+_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4}) \times A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4'})(2 - e^{i(\theta_4-\theta_4')})] \times (e^{-i\theta_4} - e^{-i\theta_4'})e^{\omega^2 e^{i(\theta_4-\theta_4')}},$$

$$N_{II} = \frac{-i}{(2\pi)^4}\int\cdots\int\cdots d^3\mathbf{p}_+\, d^3\mathbf{p}_-\omega^2 e^{-\omega^2} \iint d\theta_4\, d\theta'_4 \times [A^+_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4})A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4'})(1 - e^{-i(\theta_4-\theta_4')}) + A^-_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4})A^+_{II}(\mathbf{p}_\pm, \omega e^{\pm i\theta_4'})(1 - e^{i(\theta_4-\theta_4')})]. \tag{5.16}$$

The operators $H_{II\pm}$ are not 'gauge-invariant' and as such do not correspond to observables. To obtain the 'gauge-invariant' energy operator for the complex field it is noticed that the field Eqs. (5.3) are derivable from the action operator.[25]

$$A_{II} = \sum_{a=1}^{4}\sum_{b=1}^{4}\eta^{ab}n_b\Delta'_a\Phi^+_{II}(n^{\frac{1}{4}})\Delta'_b\Phi^-_{II}(n^{\frac{1}{4}}) - m^2\Phi^+_{II}\Phi^-_{II}. \tag{5.17}$$

Here and subsequently the summation convention is suspended. Following the Lagrangian formalism developed[14] for the field operators satisfying partial difference equations one arrives at the energy–momentum–stress operator

$$\begin{aligned} T^b_{II.c} = \sum_{a=1}^{4}\eta^{ab}[&(n_b\Delta'_a\Phi^+_{II})(\Delta_cE_b^{-1}\Phi^-_{II}) \\ &+ (\Delta_cE_b^{-1}\Phi^+_{II})(n_b\Delta'_a\Phi^-_{II}) \\ &+ \tfrac{1}{2}(n_b\Delta'_a\Delta_c\Phi^+_{II})(\Delta_cE_b^{-1}\Phi^-_{II}) \\ &+ \tfrac{1}{2}(\Delta_cE_b^{-1}\Phi^+_{II})(n_b\Delta'_a\Delta_c\Phi^-_{II})] - \delta^b_{\cdot c}L_{I\,II}, \\ \sum_b \Delta_bT^b_{II.c} = &-\tfrac{1}{2}(\Delta_c\Phi^+_{II})(\Delta_cE_c\Phi^-_{II}) \\ &- \tfrac{1}{2}(\Delta_cE_c\Phi^+_{II})(\Delta_c\Phi^-_{II}). \end{aligned} \tag{5.18}$$

The energy operator is

$$H_{II}(n^4) \overset{\text{def}}{=} \sum_{\mathbf{n}} T^4_{II.4}[(n)^{\frac{1}{4}}, n^4] = 2\pi\int d^3\,|\mathbf{p}|^2\, N_{II}(\mathbf{p}_\pm)[\omega^2(L^4_n(\omega^2))^2 + n_4L^{(-1)}_{n^4}L^{(-1)}_{n^4+1}], \tag{5.19}$$

where

$$N_{II}(\mathbf{p}_\pm) \overset{\text{def}}{=} A^+_{II}A^-_{II}, \quad L^{(-\alpha)}_n(x) \overset{\text{def}}{=} \sum_{m=0}^{n}(-1)^{m-n} \times \frac{\Gamma(n-\alpha+1)}{m!\,(n-m)!\,\Gamma(n-m-\alpha+1)}x^{n-m},$$

and

$$\int d^3\,|\mathbf{p}|^2\, f(\mathbf{p}_\pm) \overset{\text{def}}{=} \int_0^\infty\int_0^\infty\int_0^\infty d\,|p_1|^2\, d\,|p_2|^2\, d\,|p_3|^2\, f(\mathbf{p}_+, \mathbf{p}_-).$$

[25] $E_af(\cdots, n_a, \cdots) \overset{\text{def}}{=} f(\cdots, n_a+1, \cdots)$, $\Delta_af \overset{\text{def}}{=} (E_a - 1)f$, $\Delta_a'f \overset{\text{def}}{=} \Delta_aE_a^{-1}f$.

From (5.19) it is clear that observable energy of the field will not be exactly conserved.

Now the asymptotic approximation for Laguerre polynomials are[26]

$$e^{-|k|^2}L_n(|k|^2) \sim \frac{1}{[\pi(n+\frac{1}{4})^{\frac{1}{2}}\,|k|]^{\frac{1}{2}}} \times \cos[2(n+\tfrac{1}{4})^{\frac{1}{2}}\,|k| - \tfrac{1}{4}] + O\left[\frac{|k|^3}{2(n+\frac{1}{4})^{\frac{1}{2}}}\right], \quad (5.20)$$

$$\sim J_0[2(n+\tfrac{1}{2})^{\frac{1}{2}}\,|k|] + O(n^{-\frac{1}{2}}). \quad (5.20')$$

From (5.20) it is evident that the plane wave associated with a free particle is damped with respect to distance, time, wavenumber and frequency. The ratio of intensities of plane waves associated with a definite number of free particles in two different wavenumber states is

$$\frac{I}{I'} = \frac{|p_1'|\,|p_2'|\,|p_3'|\,(|p_1'|^2 + |p_2'|^2 + |p_3'|^2 + m^2)^{\frac{1}{2}}}{|p_1|\,|p_2|\,|p_3|\,(|p_1|^2 + |p_2|^2 + |p_3|^2 + m^2)^{\frac{1}{2}}}$$

in contrast to $I/I' = 1$ in the conventional theory.

From the asymptotic approximation of the Laguerre polynomials one can obtain the approximate energy operator H_{II} from (5.19) as

$$H_{II}(n^4) \sim \frac{2}{(n^4)^{\frac{1}{2}}} \int d^3\,|\mathbf{p}|^2\, N_{II}(\mathbf{p}_\pm)\omega \times (\{\cos[2(n^4+\tfrac{1}{4})^{\frac{1}{4}}\omega - \tfrac{1}{4}]\}^2 + \{\cos[2(n^4+\tfrac{1}{4})^{\frac{1}{4}}\omega + \tfrac{1}{4}]\}^2), \quad (5.21)$$

where we have chosen $N_{II}(\mathbf{p}_\pm) = 0_{II}$ for $\omega^2 > 2^{\frac{1}{2}}(n^4)^{\frac{1}{2}}$. Averaging the rapid fluctuation of energy with respect to time, we obtain from (5.21)

$$\overline{H_{II}(n^4)} \sim \frac{2}{(n^4)^{\frac{1}{2}}} \int d^3\,|\mathbf{p}|^2\, N_{II}(\mathbf{p}_\pm)\omega. \quad (5.22)$$

This is almost the energy–frequency relation of Planck excepting two differences. Firstly, energy depends slightly on the direction of wave-propagation revealing anisotropy of lattice structure. Secondly, energy decays slowly with time. From this result and also from the damping of wave with time as in (5.20), it seems that the natural universe will tend to disappear into nothingness in course of time!

If one writes $r = (n)^{\frac{1}{4}}$ for large n and substitutes (5.20′) into (5.8) then Eq. (5.8) of the classical fields in paper I is obtained, and this shows a correspondence between unquantized and quantized field operators.

[26] G. Sansone, *Orthogonal Functions* (Interscience Publishers, Inc., New York, 1959), pp. 348.
G. Szego, Orthogonal Polynomials (American Mathematics Society, New York, 1959), p. 197.

Complex Space–Time and Geometrization of Electromagnetism. III*

A. DAS
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received 10 May 1965)

It is shown here that if we demand the covariance under general circular transformations (rotations in complex planes), the electromagnetic potentials can be geometrized as the affine connections necessary for such covariance. The geodesic hypothesis yields the Lorentz equation of motion with an interesting correction term.

I. INTRODUCTION

THERE have been many attempts[1] to unify electromagnetism and gravitation in a single geometry. The more notable ones are due to Weyl[2] (with the nonintegrability of length), Kaluza[3] and Klein[4] (with the fifth dimension) and Einstein[5] and Schrödinger[6] (with nonsymmetric affine connections).

The basic ideas behind our approach are the following. Four straight lines passing through the origin in four complex planes can be characterized by $\arg z^{k+} = \text{const}$. Such lines replaced the usual Cartesian coordinate lines in our previous papers. Instead of straight coordinate lines let us consider four continuous, differentiable coordinate curves in complex planes which can be characterized by $\arg z^{k+} = \chi(r)$. If we now demand the covariance of the physical laws under transformations[7] from one set of such curves to another, the necessary affine connections and Riemann-type tensor could play the parts of electromagnetic potentials and fields. Moreover, we shall interpret the sum of angular momenta in complex planes as the electric charge. The resulting theory is quite simple.

In our approach one will recognize some shades of ideas occurring in previous attempts by others. For example, the unimodular factor which will represent circular transformations corresponds to the gauge factor in Weyl's theory. Electrical or internal coordinates $\arg z^{k+}$ have similarity with the fifth dimension of Kaluza's theory. Ours has the closest resemblance with the general spinor analysis as developed by Infeld and van der Waerden.[8] But in their formulation there are no electrical or internal coordinates, nor there is any explanation of the electromagnetic interaction of charged scalar fields. Our approach is free of these criticisms.

In this paper we have not exploited the completely general covariance in the complex space–time and possibilities along the same direction have been mentioned in concluding lines.

II. GENERAL CIRCULAR COVARIANCE AND ELECTROMAGNETIC FIELDS

Let us consider the complex space–time coordinatized with $z^{k\pm}$. We shall restrict the range of $\arg z^{k+}$ to $[0, \pi)$, so that $\text{mod } z^{k\pm}$ can assume both positive and negative values. Furthermore, for the sake of simplicity we shall assume $\arg z^{1+} = \arg z^{2+} = \arg z^{3+} = \arg z^{4+} = \theta$. Writing $z^{k\pm} = r^k e^{\pm i\theta}$, we shall consider the covariance of the theory under following transformations:

$$r^{i} = a^{i}_{.j} r^{j}, \tag{2.1}$$

$$\theta' = \theta + \lambda(r), \tag{2.2}$$

where $a^{i}_{.j}$ is an element of L_4‡ and $\lambda(r)$ is any continuous, differentiable, monotonic function of four real variables $(r) \overset{\text{def}}{=} (r^1, r^2, r^3, r^4)$.

In the complex space–time the obvious generalization of the Minkowskian line element is

$$ds^2 = \eta_{ij}\, dz^{i+}\, dz^{j-}$$

$$= \eta_{ij}(dr^i + ir^i\, d\theta)(dr^j - ir^j\, d\theta). \tag{2.3}$$

Though this real line element is invariant under (2.1), but it fails to be so under (2.2). Therefore, we generalize (2.3) to

$$ds^2 = \eta_{ij}[dr^i + ir^i(d\theta + A_k(r)\, dr^k)]$$

$$\times [dr^j - ir^j(d\theta + A_m(r)\, dr^m)], \tag{2.4}$$

* This paper has been prepared under Research Grant DA-ARO(D)-124-G-602, U. S. Army Research Office, Durham.

[1] V. Bargmann, Rev. Mod. Phys. 29, 169 (1957).

[2] H. Weyl, Sitzber, Preuss. Akad. D. Wiss. Phys. Math., 465 (1918).

[3] T. Kaluza, Sitzber, Preuss. Akad. D. Wiss., 966 (1921).

[4] O. Klein, Z. Physik 37, 895 (1926).

[5] A. Einstein, *Meaning of Relativity* (Methuen and Company Ltd., London, 1951).

[6] E. Schrödinger, Proc. Irish Acad. 51, 163, 205 (1948); 52, 1 (1948).

[7] These transformations are characterized by $z^{k\pm\prime} = z^{k\pm} e^{\pm i\lambda(r)}$.

[8] L. Infeld and B. L. van der Waerden, Sitzber. preuss. Akad. Wiss, Phys. Math. Kl. 380 (1933).

where the functions $A_k(r)$ of four real variables are assumed to be of class C_2.

The line element (2.4) will be invariant under both of (2.1) and (2.2) provided we ascribe the following transformation properties to $A_k(r)$

$$A'^k(r') = a^k_{\cdot i} A^i(r), \qquad A'_k(r) = A_k(r) - [\partial\lambda(r)/\partial r^k], \tag{2.5}$$

corresponding to (2.1) and (2.2).

For the equation of motion of a particle in complex space–time we shall postulate the geodesic principle

$$\delta \int ds = 0. \tag{2.6}$$

If we parametrize the world-line in the complex space–time by $z^{k\pm} = z^{k\pm}(s)$, then (2.6) is equivalent to the variational principle with a Lagrangian

$$L = (\tfrac{1}{2}) m \eta_{ij} [\dot r^i \dot r^j + r^i r^j (\dot\theta + A_k \dot r^k)^2], \tag{2.7}$$

where m is a finite constant and the dot denotes the total derivative with respect to s.

Equations of motion derivable from the Lagrangian L are

$$\begin{aligned} m\ddot r_i = m\Big[& A_i \frac{d}{ds}(\eta_{kj} r^k r^j (\dot\theta + A_n \dot r^n)) \\ & + \eta_{kj} r^k r^j (\dot\theta + A_n \dot r^n) \\ & \times \Big(\frac{\partial A_j}{\partial r^i} - \frac{\partial A_i}{\partial r^j}\Big)\dot r^j + (\dot\theta + A_k \dot r^k)^2 r^i \Big], \\ & \frac{d}{ds}[m \eta_{ij} r^i r^j (\dot\theta + A_k \dot r^k)] = 0. \end{aligned} \tag{2.8}$$

The integration of the last equation yields a constant of motion

$$q = m \eta_{ij} r^i r^j (\dot\theta + A_k \dot r^k), \tag{2.9}$$

which corresponds to the sum of angular–momenta in complex planes. Substituting (2.9) into (2.8) we have

$$\ddot r_i = \frac{q}{m}\Big(\frac{\partial A_j}{\partial r^i} - \frac{\partial A_i}{\partial r^j}\Big)\dot r^j + \Big(\frac{q}{m}\Big)^2 \frac{r_i}{(\eta_{jk} r^j r^k)^2}. \tag{2.10}$$

This equation resembles the Lorentz-equation of motion except the last correction term, provided we interpret m, q as the mass and the charge of the particle, and $A_i(r)$ as electromagnetic four potential. The physically significant fact from (2.10) is that if we reverse the sign of the charge in a given electromagnetic field, the acceleration will not be exactly reversed. Also, all charged particles will tend to move away from spatial origin even in absence of electromagnetic fields.

For the energy–momentum relation we can obtain from (2.7)

$$P_i \overset{\text{def}}{=} \frac{\partial L}{\partial \dot r^i} = m\dot r_i + qA_i, \qquad \eta^{ij}(P_i - qA_i)(P_j - qA_j) = m^2 - \frac{q^2}{\eta_{ij} r^i r^j}. \tag{2.11}$$

Under transformations (2.1) and (2.2) the field quantities transform like

$$\varphi'^{i'\cdots+'\cdots}(z'^+, z'^-) = a^{i'}_{\cdot i} \cdots e^{i\lambda(r)} \cdots \varphi^{i\cdots+\cdots}(z^+, z^-).$$

Therefore the covariant derivatives should be defined as[9]

$$\varphi^{\pm}_{;k} \overset{\text{def}}{=} \frac{\partial \varphi^{\pm}}{\partial r^k} \pm i A_k(r) \varphi^{\pm}, \tag{2.12}$$

where $A_k(r)$ transforms like in (2.5).

From (2.12) we have

$$\varphi^{\pm}_{;[ke]} = \pm i F_{ke} \varphi^{\pm}, \tag{2.13}$$

where

$$F_{ke} \overset{\text{def}}{=} \frac{\partial A_e}{\partial r^k} - \frac{\partial A_k}{\partial r^e}. \tag{2.14}$$

The quantities A_k, F_{ke} are analogous to Christoffel symbols and Riemann tensor, respectively. We can obtain Maxwell's equation from the variation of the square Lagrangian $L = (\frac{1}{4}) F_{ke} F^{ke}$ together with (2.14).

III. GENERAL COVARIANCE AND ELECTRO-GRAVITATIONAL FIELDS

In the complex space–time coordinatized by $z^{k\pm} = r^k e^{\pm i\theta}$, we shall now consider the covariance under transformations

$$r'^k = r'^k(r), \qquad \theta' = \theta + \epsilon\lambda(r), \tag{3.1}$$

where ϵ is a constant and the functions $r'(r)$, $\lambda(r)$ of four real variables are of class C_2.

Under (3.1) tensor fields transform like

$$\varphi'^{i\cdots+\cdots}(z'^+, z'^-) = \frac{\partial r'^i}{\partial r^j} \cdots e^{i\epsilon\lambda(r)} \cdots \varphi^{j\cdots+\cdots}(z^+, z^-). \tag{3.2}$$

The covariant derivative of $\varphi^{i\pm}(z^+, z^-)$ with respect to r^k can be defined as

$$\varphi^{i+}_{;k} \overset{\text{def}}{=} \frac{\partial \varphi^{i+}}{\partial r^k} + \Gamma^i_{kj} \varphi^{j+}, \quad \varphi^{i-}_{;k} = \frac{\partial \varphi^{i-}}{\partial r^k} + \overline{\Gamma^i_{kj}} \varphi^{j-}, \qquad \Gamma^i_{kj} \overset{\text{def}}{=} \left\{ \begin{matrix} i \\ kj \end{matrix} \right\} + i\epsilon \delta^i_{\cdot j} A_k, \tag{3.3}$$

[9] Needless to say, $(\eta_{ij}\varphi^{i+}\varphi^{j-})_{;k} = (\partial/\partial r^k)(\eta_{ij}\varphi^{i+}\varphi^{j-})$.

where $\{^i_{kj}\}$ are Christoffel symbols constructed out of the metric tensor[10] $g_{ij}(r)$ and $A_k(r)$ transform like in (2.5).

From (3.3) we obtain

$$\varphi_{i+;[ke]} = P^i_{.j\ ke}\varphi_{i+},$$

$$\varphi_{i-;[ke]} = \bar{P}^i_{.j\ ke}\varphi_{i-},$$

$$P^i_{.j\ ke} \overset{\text{def}}{=} \frac{\partial\Gamma^i_{je}}{\partial r^k} - \frac{\partial\Gamma^i_{jk}}{\partial r^e} + \Gamma^a_{je}\Gamma^i_{ak} - \Gamma^a_{jk}\Gamma^i_{ae} = R^i_{.j\ ke} - i\epsilon\delta^i_{.j}F_{ke}, \tag{3.4}$$

where $R^i_{.jke}$ is the usual Riemann Tensor constructed out of $g_{ij}(r)$ and F_{ke} has been defined in (2.14).

The $P^i_{.jke}$ tensor has the following properties

$$P_{ij\ ek} = -P_{ij\ ke}, \quad P_{ji\ ke} = -\bar{P}_{ij\ ke}, \tag{3.5}$$

$$P^i_{.j\ ke\ ;m} + P^i_{.j\ em\ ;k} + P^i_{.j\ mk\ ;e} = 0.$$

The possible contractions of $P^i_{.jke}$ are

$$P_{jk} \overset{\text{def}}{=} P^i_{.jki} = R_{jk} + i\epsilon F_{jk} = \bar{P}_{kj},$$

$$\pi_{ke} \overset{\text{def}}{=} P^i_{.ike} = -4i\epsilon F_{ke} = \bar{\pi}_{ek}, \tag{3.6}$$

$$P \overset{\text{def}}{=} P^i_{.i} = R = \bar{P}, \ \pi = \pi^k_{.k} = 0,$$

$$(P^k_{.m} + \bar{P}^k_{.m} - \delta^k_{.m}P)_{;k} = 0.$$

The electro–gravitational field equations should be derived from any one or linear combinations of square Lagrangians[11]

$$L' = P^{ij}\bar{P}_{ij} = R^{jk}R_{jk} + \epsilon^2F^{jk}F_{jk}, \tag{3.7}$$

$$L'' = P^{ij\ ke}\bar{P}_{ij\ ke} = R^{ij\ ke}R_{ij\ ke} + 4\epsilon^2F^{ke}F_{ke}.$$

It is already known[12] that the field equations derived from L' or L'' contain the Schwarzchild's field as one of possible solutions. However, the Nördstrom–Jefferey–Reistner solution is not obtainable from the same. These statements will be verified in the following. In case of the spherical symmetry if we choose

$$-g_{11}^{-1} = g_{44} = 1 + 2V(\rho), \quad g_{22} = -\rho^2, \quad g_{33} = -\rho^2\sin^2\theta, \quad F^{14} = q/\rho^2, \tag{3.8}$$

then field equations derivable from L'', for example, boil down to

$$2\rho^2V'''V' + 4\rho V''V' - \rho^2V''^2 + 4(1+2V)V'' - 4V'^2 - \frac{8V}{\rho^2} - \frac{12V^2}{\rho^2} - \frac{4(\epsilon q)^2}{\rho^2} = 0, \tag{3.9}$$

where the prime denotes the differentiation with respect to ρ. It is easy to verify that for $q = 0$ we can have $V(\rho) = -m/\rho$ as one of solutions of (3.9). But for $q \neq 0$, we cannot satisfy (3.9) with $V(\rho) = -(m/\rho) + (k/\rho^2)$.

Before concluding we should remark that we have not considered the most general coordinate transformations in the complex space–time (coordinatized by z^a and $z^{\dot{a}} = \bar{z}^a$). These are

$$z'^a = z'^a(z, \bar{z}), \quad z'^{\dot{a}} = z'^{\dot{a}}(\bar{z}, z).$$

Correspondingly the line element that should be considered is

$$ds^2 = g_{ab}dz^adz^b + g_{a\dot{b}}dz^adz^{\dot{b}} + g_{\dot{a}b}dz^{\dot{a}}dz^b,$$

$$g_{ab} = g_{ba}, \quad g_{a\dot{b}} = g_{\dot{b}a}, \quad g_{\dot{a}\dot{b}} = g_{\dot{b}\dot{a}},$$

which contain 36 components of the metric tensor and may consist of many other fields besides electro–gravitation. The matter is being pursued further.

ACKNOWLEDGMENT

I am grateful for the U.S. Army grant which supported this research project. It is a pleasure to thank Professor R. J. Duffin for his interest in this work. I also would like to take this opportunity to thank Professor J. L. Synge and Professor Y. Takahashi of Dublin Institute for advanced studies where this work started.

[10] $(g_{ij}\varphi^{i+}\varphi^{i-})_{;k} = (\partial/\partial r^k)(g_{ij}\varphi^{i+}\varphi^{i-})$.

[11] cf. Reference 2; W. Pauli, Z. Physik **20**, 457 (1919); C. Lanczos, Rev. Mod. Phys. **29**, 337 (1957).

[12] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, New York, 1930), p. 142.

Asymptotic Behavior of the Scattering Amplitude in the Left Half-Plane of Complex Angular Momentum

J. DILLEY
Department of Physics, Ohio University, Athens, Ohio
AND
A. O. BARUT
Department of Physics, University of Colorado, Boulder, Colorado
(Received 13 November 1964)

It is shown, for a certain class of analytic potentials, that the previously reported result

$$A(\lambda, k) \underset{|\lambda|\to\infty}{\sim} 1/k^2 V(\pm\lambda/k), \text{ Re } \lambda \geq 0,$$

where the sign $\pm$ is chosen such that $-\pi/2 \leq \arg \pm (\lambda/k) \leq \pi/2$, is valid also for Re $\lambda < 0$ except for a discrete set of points on the negative real axis. The result holds for arbitrary complex $k \neq 0$, finite or infinite.

I. GENERAL CONSIDERATIONS OF ASYMPTOTIC BEHAVIOR

IN a previous paper,[1] hereafter referred to as I, we discussed the asymptotic behavior of the scattering amplitude in the variable $\lambda = l + \frac{1}{2}$ for a wide class of analytic potentials (see Appendix). The principal result was

$$A(\lambda, k) \underset{|\lambda|\to\infty}{\sim} \frac{1}{k^2} V(\pm\lambda/k), \qquad \text{Re } \lambda \geq 0 \qquad (1)$$

where the sign $\pm$ is chosen such that $\pm\lambda/k$ is in the region where $V(\pm\lambda/k)$ is decreasing, i.e. $-\pi/2 \leq \arg \pm \lambda/k \leq \pi/2$. The purpose of this note is to show that, under certain conditions, Eq. (1) applies also for Re $\lambda < 0$.

The evaluation of the asymptotic behavior of the scattering amplitude is based on the work of Langer,[2] who considered the asymptotic solutions of the equation

$$u''(z) + [\lambda^2\phi^2(z) + x(z)]u(z) = 0 \qquad (2)$$

where $\phi^2(z)$ and $x(z)$ are arbitrary holormorphic functions of z in some region G containing the origin, and $\phi^2(z)$ has a single zero of order $n \geq 0$ at $z = 0$. It was shown in I that, by suitable transformations, the radial Schrödinger equation

$$v_\lambda''(r) + \left[k^2 - \frac{\lambda^2 - \frac{1}{4}}{r^2} - V(r)\right]v_\lambda(r) = 0 \qquad (3)$$

can be written in the form (2). In particular, from Langer's results, we can always write two linearly independent solutions $v_\lambda^\pm(r)$ of (3) whose form, for sufficiently large $|\lambda|$, is explicitly given in terms of the two functions $\phi(r)$ and $\Phi(r)$, where

$$\phi^2(r) = [k^2 - V(r)](r^2/\lambda^2) - 1, \qquad (4)$$

$$\Phi(r) = \int_{r_0}^{r} \phi(r') \frac{dr'}{r'}, \qquad (5)$$

and r_0 is a solution of the equation $\phi^2(r) = 0$, with

$$r_0 \underset{|\lambda|\to\infty}{\longrightarrow} \pm \lambda/k + O\left[\frac{1}{k^2} V(\pm\lambda/k)\right]. \qquad (6)$$

These results are identical to those reported in I, and are valid in the left as well as the right half-planes.

Since the solutions $v_\lambda^\pm(r)$ are linearly independent, the physical solution $v_\lambda(r)$, defined by the boundary condition at the origin,

$$v_\lambda(r) \underset{r\to 0}{\longrightarrow} r^{\lambda+\frac{1}{2}} \quad \text{for} \quad \text{Re } \lambda \geq 0,$$

can always be written as a linear combination of $v_\lambda^\pm(r)$,

$$v_\lambda(r) = a(\lambda)v_\lambda^+(r) + b(\lambda)v_\lambda^-(r). \qquad (7)$$

The scattering amplitude is completely determined by the two coefficients $a(\lambda)$ and $b(\lambda)$ which must be identified by using the behavior of $v_\lambda(r)$ and $v_\lambda^\pm(r)$. In I, this identification was carried out by using the known behavior of the wavefunction at $r = 0$.

$$v_\lambda(r) \underset{r\to 0}{\longrightarrow} r^{\lambda+\frac{1}{2}}. \qquad (8)$$

What we wish to note now is that although (8) does not necessarily hold for Re < 0, nevertheless, for values of λ where it does hold, all of the methods of I apply, and the asymptotic behavior of the scattering amplitude is then given by (1).

[1] A. O. Barut and J. Dilley, J. Math. Phys. 4, 1401 (1963). The reader is referred to this paper for all details concerning Langer's method and its application to the scattering amplitude.
[2] R. E. Langer, Trans. Am. Math. Soc. 34, 449 (1932).

II. THE LEFT HALF-PLANE

As is well known, the physical solutions of the Schrödinger equation can be analytically continued in λ through the equivalent integral equation[3]

$$v_\lambda = r^{\lambda+\frac{1}{2}} - \frac{1}{2\lambda}\int_0^r \left[\frac{\xi^{\lambda+\frac{1}{2}}}{r^{\lambda-\frac{1}{2}}} - \frac{r^{\lambda+\frac{1}{2}}}{\xi^{\lambda-\frac{1}{2}}}\right][V(\xi) - k^2]v_\lambda(\xi)d\xi \quad (9)$$

and, throughout the region where this solution exists, (8) will be satisfied. Now, according to Newton's work,[4] the constant term k^2 of $V(\xi) - k^2$ will generate a series of fixed poles at $\lambda = -1, -2, \cdots$, while if

$$V(r) \underset{r\to 0}{\longrightarrow} r^p, \qquad p > -2,$$

the potential $V(\xi)$ will generate another sequence of poles at $\lambda = -1 - \frac{1}{2}p, -1 - \frac{1}{2}p - \frac{1}{2}, -1 - \frac{1}{2}p - 1, \cdots$. How far the continuation can be made in this manner depends upon the precise behavior of $V(r)$ at the origin. The precise result, again from Newton's work, is that if $r^{-p}V(r)$ is m times differentiable at $r = 0$, then (9) is valid in the region Re $> -m - \frac{1}{2}$, except at the fixed points mentioned above. (Of course, for $p = 0$ or -1, the poles at $\lambda = -1, -2, \cdots$, may cancel for certain values of k, giving the well known indeterminacy points of the S-matrix.[5]) *Thus, if the behavior of the potential is such as to allow a continuation into the region where the asymptotic formulas apply, we see that* (1) *holds except in the neighborhood of a discrete set of values of* λ, *this set lying on the negative real* λ *axis for real potentials.*

We close with two final observations:

(a) For any fixed $k \neq 0$, and sufficiently large $|\lambda|$, all Regge poles must be in the neighborhood of the fixed points on the negative real λ axis. However, since $k = 0$ is excluded, and the asymptotic theory is not valid uniformly in k, these conclusions do not exclude the possibility of an infinite number of poles arriving at $\lambda = -\frac{1}{2}$ when $k = 0$.[6]

(b) The asymptotic behavior is sufficient to allow the Khuri representation in the left half-plane.[7]

APPENDIX

We give here the most precise restrictions of $V(r)$. For our present purpose, we assume that the potential can be analytically continued in the r plane to $-\pi/2 \leq \arg r \leq \pi/2$, and that, for sufficiently large $|r|$, it decreases monotonically. In addition, to derive the results in I, two further conditions must be satisfied:

(a) We must have

$$\phi^2(r) \underset{|r|\to\infty}{\longrightarrow} \phi_0^2(r)$$

for all *real* values of r, where

$$\phi^2(r) = [k^2 - V(r)](r^2/\lambda^2) - 1,$$
$$\phi_0^2(r) = (k^2r^2/\lambda^2) - 1.$$

Thus

$$\phi^2(r) - \phi_0^2(r) = -(V(r)r^2/\lambda^2) \underset{|\lambda|\to\infty}{\longrightarrow} 0$$

if $r^2V(r)$ is bounded. In particular, $V(r)$ must be less singular than $1/r^2$ at the origin and must decrease at least as $1/r^2$ as $r \to \infty$ on the real axis.

(b) In addition, according to Eq. (28) in I, one must be able to define a point $R(\lambda)$ such that

$$\left|\frac{V(R)R^2}{\lambda^2}\right| \ll \left|\frac{k^2R^2}{\lambda^2} - 1\right|.$$

If one chooses $R(\lambda) = (\lambda/k) + (1/k)\lambda^\epsilon$, then

$$\frac{V(R)R^2}{\lambda^2} \to \frac{V(\lambda/k)}{k^2}, \qquad \frac{k^2R^2}{\lambda^2} - 1 \to 2\lambda^{-1-\epsilon}$$

so that $V(\lambda)$ must decrease as $\lambda^{-1-\epsilon}$ as $|\lambda| \to \infty$. For real k, this would mean that the results are not strictly proved on arg $\lambda = \pm\pi/2$ for Yukawa potentials, although they would apply for modified Yukawa types such as $V(r) = (e - \mu r)/(1 + r^2)$. It may be, however, that the apparent difficulty exists merely in our particular method of evaluation.

[3] A. Bottino, A. M. Longoni, and T. Regge, Nuovo Cimento **23**, 954 (1962).

[4] R. G. Newton, J. Math. Phys. **3**, 867 (1962).

[5] See, for example, A. O. Barut and F. Calogero, Phys. Rev. **128**, 1383 (1962); A. Ahmadzadeh, P. Burke, and C. Tate, "Regge Trajectories for Yukawa Potentials," UCRL-10216 (1962); S. Mandelstam, Ann. Phys. **19**, 254 (1962); also Ref. 4.

[6] V. N. Gribov and I. Ya. Pomeranchuk, Phys. Letters **2**, 239 (1962); B. R. Deasi and R. G. Newton, Phys. Rev. **129**, 1445 (1963).

[7] N. N. Khuri, Phys. Rev. **130**, 429 (1963).

Mathematical Aspects of the Weyl Correspondence

JAMES C. T. POOL*
Scientific Laboratory, Ford Motor Company, Dearborn, Michigan and Department of Physics, Brandeis University, Waltham, Massachusetts
(Received 13 May 1965)

The Weyl correspondence between classical and quantum observables is rigorously formulated for a linear mechanical system with a finite number of degrees of freedom. A multiplication of functions and a *-operation are introduced to make the Hilbert space of Lebesgue square-integrable complex-valued functions on phase space into a H^*-algebra. The Weyl correspondence is realized as a *-isomorphism $f \to W(f)$ of this H^*-algebra onto the H^*-algebra of Hilbert–Schmidt operators on the Hilbert space of Lebesgue square-integrable complex-valued functions on configuration space. Moreover, the kernel of $W(f)$ is exhibited in terms of a Fourier-Plancherel transform of f. Elementary properties of the Wigner quasiprobability density function and its characteristic function are deduced and used to obtain these results.

I. INTRODUCTION

THE configuration space of a linear classical-mechanical system with one degree of freedom is the real line R. The Hilbert space utilized in the quantum theory associated with such a classical-mechanical system is the space $L^2(R)$ of Lebesgue square-integrable complex-valued functions on R. Consequently, it is asserted in the conventional Hilbert space formulation of quantum mechanics that each self-adjoint operator on $L^2(R)$ corresponds to a unique numerical-valued observable attribute of the physical system and conversely.[1] For example, the operators P and Q corresponding to momentum and position, respectively, are the following:

$$(P\psi)(q) = (\hbar/i)d\psi/dq \tag{I.1a}$$

and

$$(Q\psi)(q) = q\psi(q) \tag{I.1b}$$

for q in R and appropriate ψ in $L^2(R)$.[2] However, the physical significance of an arbitrary self-adjoint operator on $L^2(R)$ is not manifestly evident.

The Weyl correspondence provides physical significance for a large class of self-adjoint operators on $L^2(R)$. In classical mechanics, each numerical-valued observable attribute of the physical system is represented by a function on the phase space of the system[3]; in this case, the Cartesian product $R \times R$. Weyl has suggested the following method for obtaining quantum-mechanical operators corresponding to the classically observable functions on phase space.[4] Let f be a function on $R \times R$ with the Fourier representation

$$f(p, q) = (2\pi)^{-1} \iint \hat{f}(q', p') \exp i(q'p + p'q)\, dq'\, dp'. \tag{I.2}$$

According to Weyl, the corresponding quantum-mechanical operator $W(f)$ is obtained by replacing p by P and q by Q in Eq. (I.2); consequently, $W(f)$ is given by the formal expression

$$W(f) = (2\pi\hbar)^{-\frac{1}{2}} \iint \hat{f}(q', p') \exp i(q'P + p'Q)^{\sim}\, dq'\, dp' \tag{I.3}$$

where $^{\sim}$ denotes the closure of an operator.[5] [The factor $(2\pi\hbar)^{-\frac{1}{2}}$ is only a convenient normalization.] This association of the quantum-mechanical operator $W(f)$ with the classical function f on phase space is the so-called "Weyl correspondence."

The Weyl correspondence is closely related to the phase-space formulation of quantum mechanics.[6] The Wigner phase-space quasiprobability distribution function $\mathcal{R}[\psi]$ and its characteristic function $\mathcal{S}[\psi]$ corresponding to a wavefunction ψ in $L^2(R)$ replace the wavefunction in this formulation:

$$\mathcal{R}[\psi](p, q) = (2\pi\hbar)^{-\frac{1}{2}} \int \psi(q + \tfrac{1}{2}\hbar x)e^{-ipx}\psi(q - \tfrac{1}{2}\hbar x)\, dx \tag{I.4a}$$

---

* National Science Foundation Postdoctoral Fellow.

[1] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1963), pp. 75–76, 85–88.

[2] N. J. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar Publishing Company, Inc., New York, 1961), Vol. 1, pp. 103–113.

[3] Reference 1, p. 27.

[4] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publishing Company, Inc., New York, 1950), pp. 274–275.

[5] Reference 2, p. 78.

[6] For a review of the phase space formulation of quantum mechanics see, E. C. G. Sudarshan, "Structure of Dynamical Theories" in *Lectures in Theoretical Physics*, Brandeis Summer Institute 1961 (W. A. Benjamin, Inc., New York, 1962), pp. 143–199.

and

$$\mathcal{S}[\psi](q', p') = (2\pi\hbar)^{-\frac{1}{2}}[\exp i(q'P + p'Q)\tilde{}\,\psi, \psi]. \tag{I.4b}$$

The connection between the Weyl correspondence, the Wigner function and its characteristic function is indicated by the following expressions for the expectation value of $W(f)$:

$$(W(f)\psi, \psi) = \iint f(p, q)\mathcal{R}[\psi](p, q)\, dp\, dq \tag{I.5a}$$

and

$$(W(f)\psi, \psi) = \iint \hat{f}(q', p')\mathcal{S}[\psi](q', p')\, dq'\, dp'. \tag{I.5b}$$

The possibility of expressing the expectation value of the quantum-mechanical operator as phase-space integrals permits the development of a phase-space formulation of quantum mechanics. The resulting formulation of quantum statistical mechanics exhibits an analogy with classical statistical mechanics which proves to be quite useful for the investigation of some physical problems.[7]

Therefore, it is desirable to have a mathematically rigorous formulation of the Weyl correspondence, the Wigner function, and its characteristic function for both conceptual and practical reasons. The purpose of this paper is to formulate the Weyl correspondence rigorously utilizing the theory of harmonic analysis. Section III is devoted to obtaining an alternative form of Eq. (I.3). Some elementary properties of the Wigner function and its characteristic function are deduced in Sec. IV. The existence and uniqueness of the Weyl correspondence as a map $f \to W(f)$, called the Weyl transform, from the Hilbert space $L^2(R \times R)$ of Lebesgue square-integrable complex-valued functions on $R \times R$ into the algebra $\mathcal{L}_c(L^2(R))$ of continuous linear operators on $L^2(R)$ is established in Sec. V. The operator $W(f)$ is shown to be an integral operator on $L^2(R)$ and the kernel of $W(f)$ is exhibited in terms of a Fourier–Plancherel transform of f in Sec. VII. A multiplication of functions and a *-operation are introduced to make $L^2(R \times R)$ into an H^*-algebra in Sec. VIII. The Weyl transform is then shown to be a *-isomorphism of this H^*-algebraic structure of $L^2(R \times R)$ into the H^*-algebra $\mathcal{L}_{HS}(L^2(R))$ of Hilbert–Schmidt operators on $L^2(R)$. Possible generalizations of these results are indicated in Sec. IX. Necessary mathematical concepts are introduced in Secs. II and VI. The exposition assumes some acquaintance with basic concepts in measure theory[8] and the theory of Hilbert space.[9]

[7] For a review of the applications of the phase-space formulation of quantum statistical mechanics see, H. Mori, I. Oppenheim, and J. Ross, "Some Topics in Quantum Statistics: The Wigner Function and Transport Theory" in J. de Boer and G. E. Uhlenbeck, *Studies in Statistical Mechanics* (North-Holland Publishing Company, Amsterdam, 1962), Vol. 1, pp. 212–298.

II. MATHEMATICAL PRELIMINARIES

The Hilbert spaces $L^2(R)$ and $L^2(R \times R)$ of Lebesgue square-integrable complex-valued functions on R and $R \times R$, respectively, are considered throughout the following sections; consequently, the following notation for their respective inner products and norms will be adopted:

$$(\psi_1, \psi_2) = \int \psi_1(x)\psi_2(x)\, dx, \tag{II.1a}$$

$$|\psi| = (\psi, \psi)^{\frac{1}{2}} \tag{II.1b}$$

for ψ_1, ψ_2, ψ in $L^2(R)$ and

$$\langle f_1, f_2\rangle = \iint f_1(x, y)f_2(x, y)\, dx\, dy, \tag{II.2a}$$

$$\|f\| = \langle f, f\rangle^{\frac{1}{2}} \tag{II.2b}$$

for f_1, f_2, f in $L^2(R \times R)$. The Fourier–Plancherel operator F on $L^2(R)$ is defined as follows[10]:

$$(F\psi)(x) = (2\pi)^{-\frac{1}{2}} \frac{d}{dx} \int \frac{e^{-ixs} - 1}{-is} \psi(s)\, ds \tag{II.3a}$$

for x in R and ψ in $L^2(R)$. F is a unitary operator on $L^2(R)$ with the inverse

$$(F^{-1}\psi)(x) = (2\pi)^{-\frac{1}{2}} \frac{d}{dx} \int \frac{e^{ixs} - 1}{is} \psi(s)\, ds \tag{II.3b}$$

for x in R and ψ in $L^2(R)$. If ψ is in both $L^2(R)$ and $L^1(R)$, then Eqs. (II.3) reduce to the classical expressions for Fourier and inverse Fourier transforms:

$$(F\psi)(x) = (2\pi)^{-\frac{1}{2}} \int e^{-ixs}\psi(s)\, ds \tag{II.4a}$$

and

$$(F^{-1}\psi)(x) = (2\pi)^{-\frac{1}{2}} \int e^{ixs}\psi(s)\, ds \tag{II.4b}$$

for x in R. If ψ is in $L^2(R)$, then the complex conjugate of ψ is denoted by $\bar{\psi}$,

$$\bar{\psi}(x) = \overline{\psi(x)} \tag{II.5}$$

[8] See, for example, S. K. Berberian, *Measure and Integration* (The Macmillan Company, New York, 1965).

[9] See, for example, S. K. Berberian, *Introduction to Hilbert Space* (Oxford University Press, New York, 1961).

[10] Reference 2, pp. 76–77.

for x in R. The following properties of the Fourier–Plancherel operator will be utilized:

$$F\bar{\psi} = \overline{F^{-1}\psi}, \tag{II.6a}$$

$$F^{-1}\bar{\psi} = \overline{F\psi}, \tag{II.6b}$$

$$(F\psi\check{})(x) = (F^{-1}\psi)(-x) \tag{II.6c}$$

for all ψ in $L^2(R)$ and x in R.

If ψ_1 and ψ_2 are in $L^2(R)$, then the function $\psi_1 \cdot \psi_2$ on $R \times R$ is defined by

$$\psi_1 \cdot \psi_2(x, y) = \psi_1(x)\psi_2(y) \tag{II.7}$$

for (x, y) in $R \times R$; therefore,[11] $\psi_1 \cdot \psi_2$ is in $L^2(R \times R)$ and

$$\|\psi_1 \cdot \psi_2\| = |\psi_1|\,|\psi_2|. \tag{II.8}$$

Fourier–Plancherel operators F_1, F_2, and F_{12} on $L^2(R \times R)$ with respect to the first variable, the second variable and both variables, respectively, may be introduced as the unique unitary operators on $L^2(R \times R)$ such that

$$F_1\psi_1 \cdot \psi_2 = (F\psi_1) \cdot \psi_2, \tag{II.9a}$$

$$F_2\psi_1 \cdot \psi_2 = \psi_1 \cdot (F\psi_2), \tag{II.9b}$$

$$F_{12}\psi_1 \cdot \psi_2 = (F\psi_1) \cdot (F\psi_2) \tag{II.9c}$$

for all ψ_1 and ψ_2 in $L^2(R)$. F_1, F_2 and F_{12} exhibit properties analogous to Eqs. (II.6) and, furthermore,

$$F_1F_2 = F_2F_1 = F_{12}. \tag{II.10}$$

It will prove to be extremely useful to introduce a "twisting" operator T on $L^2(R \times R)$ defined by

$$(Tf)(x, y) = \hbar^{-\frac{1}{2}} f(y + \tfrac{1}{2}\hbar x, y - \tfrac{1}{2}\hbar x) \tag{II.11}$$

for (x, y) in $R \times R$. Since the matrices

$$\begin{pmatrix} \frac{1}{2}\hbar & 1 \\ -\frac{1}{2}\hbar & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \hbar^{-1} & -\hbar^{-1} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{II.12}$$

are mutual inverses and have Jacobian $\hbar$ and $\hbar^{-1}$, respectively, T is a unitary operator on $L^2(R \times R)$ with the inverse

$$(T^{-1}f)(x, y) = \hbar^{\frac{1}{2}} f(\hbar^{-1}(x - y), \tfrac{1}{2}(x + y)) \tag{II.13}$$

for (x, y) in $R \times R$ and f in $L^2(R \times R)$.

III. WEYL TRANSFORM: FORMULATION

The essential aspect of the configuration space R involved in the consideration of the Weyl correspondence is the additive group structure of R.[12] Indeed, if q' is in R and the operator $U(q')$ on $L^2(R)$ is defined by

$$(U(q')\psi)(x) = \psi(x + \hbar q') \tag{III.1}$$

for x in R and ψ in $L^2(R)$, then $q' \to U(q')$ is a weakly continuous unitary representation[13] of the additive group of reals on $L^2(R)$. It may be shown that

$$U(q') = e^{iq'P} \tag{III.2}$$

for q' in R; consequently, P is the infinitesimal generator of the representation $q' \to U(q')$. The operator Q is also the infinitesimal generator of a representation of the additive group of reals. Indeed, if p' is in R and the operator $V(p')$ on $L^2(R)$ is defined by

$$(V(p')\psi)(x) = e^{ip'x}\psi(x) \tag{III.3}$$

for x in R and ψ in $L^2(R)$, then $p' \to V(p')$ is a weakly continuous unitary representation of the additive group of reals on $L^2(R)$; furthermore,

$$V(p') = e^{ip'Q} \tag{III.4}$$

for p' in R. Since the infinitesimal generators P and Q satisfy the Heisenberg commutation relations,

$$QP - PQ \subset i\hbar I, \tag{III.5}$$

the representations $q' \to U(q')$ and $p' \to V(p')$ must be expected to possess distinguished commutation properties. Indeed, a short computation using Eqs. (III.1) and (III.3) yields the Weyl commutation relations

$$U(q')V(p') = e^{i\hbar q'p'}V(p')U(q') \tag{III.6}$$

for q' and p' in R.

The sum of the unbounded operators P and Q in the integrand of Eq. (I.3) presents an analytical inconvenience which we may eliminate by considering the representations $q' \to U(q')$ and $p' \to V(p')$. For each pair q' and p' in R, we introduce a unitary operator $W(q', p')$ on $L^2(R)$ defined by

$$W(q', p') = \exp(\tfrac{1}{2}i\hbar q'p')V(p')U(q'). \tag{III.7}$$

The commutation relations for $W(q', p')$ follow directly from Eq. (III.6):

$$W(q_1', p_1')W(q_2', p_2') = \exp[\tfrac{1}{2}i\hbar(q_1'p_2' - q_2'p_1')]W(q_1' + q_2', p_1' + p_2') \tag{III.8a}$$

[11] Reference 8, p. 131, Exercise 2.

[12] I. E. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, Rhode Island, 1963), pp. 9–18.

[13] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, Inc., New York, 1957), pp. 380–385.

for q_1', p_1', q_2', p_2' in R or

$$W(q_1', p_1')W(q_2', p_2') = \exp[\tfrac{1}{2}i\hbar B(q_1', p_1', q_2', p_2')]W(q_1' + q_2', p_1' + p_2') \tag{III.8b}$$

for q_1', p_1', q_2', p_2' in R, where B is the function from $(R \times R) \times (R \times R)$ into R defined by

$$B(x, y; x', y') = xy' - x'y \tag{III.8c}$$

for (x, y), (x', y') in $R \times R$. It follows from Eqs. (III.7) and (III.8) that $t \to W(tq', tp')$ is weakly continuous unitary representation on $L^2(R)$ of the additive group of reals for each pair q', p' in R. Consequently, by Stone's theorem, there exists a unique self-adjoint operator $R(q', p')$ on $L^2(R)$ such that

$$W(tq', tp') = e^{itR(q', p')} \tag{III.9}$$

for all t in R; indeed, it can be shown that

$$R(q', p') = (q'P + p'Q)^{\sim}. \tag{III.10}$$

Consequently, we may replace Eq. (I.3) by the following expression involving $W(q', p')$ instead of P and Q:

$$W(f) = (2\pi\hbar)^{\frac{1}{2}} \iint (F_{12}f)(q', p')W(q', p')\, dq'\, dp'. \tag{III.11}$$

The following sections are devoted to giving a well-defined meaning to this formal expression using only basic results from measure theory, the theory of Hilbert spaces and harmonic analysis.

IV. WIGNER AND CHARACTERISTIC FUNCTIONS

The goal of this section is to deduce the square-integrability of the Wigner function and its characteristic function and establish their relation as Fourier–Plancherel transforms. The following definition introduces constructs which will be shown to include the usual Wigner function and characteristic function corresponding to a wavefunction.

Definition IV.1: The maps $\mathfrak{R}$ and $\mathfrak{S}$ from the Cartesian product $L^2(R) \times L^2(R)$ into $L^2(R \times R)$ are defined by

$$\mathfrak{R}[\psi_1, \psi_2] = F_1 T \psi_1 \cdot \bar{\psi}_2 \tag{IV.1a}$$

and

$$\mathfrak{S}[\psi_1, \psi_2] = F_2^{-1} T \psi_1 \cdot \bar{\psi}_2, \tag{IV.1b}$$

respectively, for all ψ_1 and ψ_2 in $L^2(R)$.

Proposition IV.1: The maps $\mathfrak{R}$ and $\mathfrak{S}$ are sesquilinear maps,[14] that is,

$$\mathfrak{R}[\psi_1 + \psi_1', \psi_2] = \mathfrak{R}[\psi_1, \psi_2] + \mathfrak{R}[\psi_1', \psi_2], \tag{IV.2a}$$

$$\mathfrak{R}[\lambda\psi_1, \psi_2] = \lambda\mathfrak{R}[\psi_1, \psi_2], \tag{IV.2b}$$

$$\mathfrak{R}[\psi_1, \psi_2 + \psi_2'] = \mathfrak{R}[\psi_1, \psi_2] + \mathfrak{R}[\psi_1, \psi_2'], \tag{IV.2c}$$

$$\mathfrak{R}[\psi_1, \lambda\psi_2] = \bar{\lambda}\mathfrak{R}[\psi_1, \psi_2], \tag{IV.2d}$$

for all ψ_1, ψ_1', ψ_2, ψ_2' in $L^2(R)$ (and similarly for $\mathfrak{R}$ replaced by $\mathfrak{S}$) such that

$$\|\mathfrak{R}[\psi_1, \psi_2]\| = \|\mathfrak{S}[\psi_1, \psi_2]\| = |\psi_1|\,|\psi_2| \tag{IV.2e}$$

and

$$\mathfrak{R}[\psi_1, \psi_2] = F_{12}\mathfrak{S}[\psi_1, \psi_2] \tag{IV.3}$$

for all ψ_1 and ψ_2 in $L^2(R)$.

Proof: Clearly, $\psi_1, \psi_2 \to \psi_1 \cdot \bar{\psi}_2$ is a sesquilinear map of $L^2(R)$ into $L^2(R \times R)$. Since F_1, F_2^{-1}, and T are unitary operators on $L^2(R \times R)$, it is immediate that $\mathfrak{R}$ and $\mathfrak{S}$ are sesquilinear and, moreover,

$$\|\mathfrak{R}[\psi_1, \psi_2]\| = \|\psi_1 \cdot \bar{\psi}_2\|,$$

$$\|\mathfrak{S}[\psi_1, \psi_2]\| = \|\psi_1 \cdot \bar{\psi}_2\|.$$

Consequently, Eq. (IV.2e) follows from Eq. (II.8). The identity

$$F_{12}F_2^{-1} = F_1$$

follows from Eq. (II.10); therefore,

$$\begin{aligned} F_{12}\mathfrak{S}[\psi_1, \psi_2] &= F_{12}F_2^{-1}T\psi_1 \cdot \bar{\psi}_2 \\ &= F_1 T\psi_1 \cdot \bar{\psi}_2 \\ &= \mathfrak{R}[\psi_1, \psi_2]. \end{aligned}$$

The following proposition establishes explicit formulas for the functions $\mathfrak{R}[\psi_1, \psi_2]$ and $\mathfrak{S}[\psi_1, \psi_2]$ corresponding to ψ_1 and ψ_2 in $L^2(R)$.

Proposition IV.2: If ψ_1 and ψ_2 are in $L^2(R)$, then (i) for almost all (p, q) in $R \times R$,

$$\mathfrak{R}[\psi_1, \psi_2](p, q) = (2\pi\hbar)^{-\frac{1}{2}} \int \psi_1(q + \tfrac{1}{2}\hbar x)e^{-ipx}\psi_2(q - \tfrac{1}{2}\hbar x)\, dx \tag{IV.4}$$

(ii) for almost all (q', p') in $R \times R$,

$$\mathfrak{S}[\psi_1, \psi_2](q', p') = (2\pi\hbar)^{-\frac{1}{2}} \int \psi_1(x + \tfrac{1}{2}\hbar q')e^{ip'x}\psi_2(x - \tfrac{1}{2}\hbar q')\, dx \tag{IV.5a}$$

$$= (2\pi\hbar)^{-\frac{1}{2}}(W(q', p')\psi_1, \psi_2). \tag{IV.5b}$$

[14] Reference 9, pp. 123–130.

Proof: If the function h in $L^2(R \times R)$ is defined by

$$h = T\psi_1 \cdot \bar{\psi}_2,$$

then the x-section[15] h_x and the y-section h^y of h are given by

$$h_x(u) = \hbar^{-\frac{1}{2}}\psi_1(u + \tfrac{1}{2}\hbar x)\bar{\psi}_2(u - \tfrac{1}{2}\hbar x)$$

and

$$h^y(u) = \hbar^{-\frac{1}{2}}\psi_1(y + \tfrac{1}{2}\hbar u)\bar{\psi}_2(y - \tfrac{1}{2}\hbar u)$$

for u in R. For a fixed x in R, consider the functions g_1 and g_2 defined by

$$g_1(u) = \psi_1(u + \tfrac{1}{2}\hbar x)$$
$$g_2(u) = \bar{\psi}_2(u - \tfrac{1}{2}\hbar x)$$

for u in R. The translational invariance of Lebesgue measure implies that g_1 and g_2 are elements of $L^2(R)$. h_x is the product of elements of $L^2(R)$; therefore, h_x is an element of $L^1(R)$ as a function of u. Since h is in $L^2(R \times R)$,

$$\iint |h(x, y)|^2\, dx\, dy < \infty;$$

therefore,

$$\int |h_x(u)|^2\, du < \infty$$

for almost all x in R by Fubini's theorem. Consequently, h_x is an element of both $L^2(R)$ and $L^1(R)$ for almost all x. A similar argument establishes the same result for h^y.

According to Definition IV.1,

$$\mathfrak{R}[\psi_1, \psi_2] = F_1 h$$

and

$$\mathfrak{S}[\psi_1, \psi_2] = F_2^{-1} h$$

and, moreover, by the previous paragraph the classical expressions for F_1 and F_2^{-1} may be used. Therefore, for almost all (p, q) in $R \times R$,

$$\mathfrak{R}[\psi_1, \psi_2](p, q)$$
$$= (2\pi)^{-\frac{1}{2}} \int e^{-ipx} h(x, q)\, dx$$
$$= (2\pi\hbar)^{-\frac{1}{2}} \int e^{-ipx}\psi_1(q + \tfrac{1}{2}\hbar x)\bar{\psi}_2(q - \tfrac{1}{2}\hbar x)\, dx$$

and for almost all (q', p') in $R \times R$,

$$\mathfrak{S}[\psi_1, \psi_2](q', p')$$
$$= (2\pi)^{-\frac{1}{2}} \int e^{ip'x} h(q', x)\, dx$$
$$= (2\pi\hbar)^{-\frac{1}{2}} \int e^{ip'x}\psi_1(x + \tfrac{1}{2}\hbar q')\bar{\psi}_2(x - \tfrac{1}{2}\hbar q')\, dx,$$

where Eq. (II.11) has been used to evaluate $h = T\psi_1 \cdot \bar{\psi}_2$. Introducing the change of variables $x \to x' = x - \frac{1}{2}\hbar q'$, the last expression becomes

$$\mathfrak{S}[\psi_1, \psi_2](q', p')$$
$$= (2\pi\hbar)^{-\frac{1}{2}} \int \exp[ip'(x' + \tfrac{1}{2}\hbar q')]\psi_1(x' + \hbar q')\bar{\psi}_2(x')\, dx'.$$

However, it follows immediately from Eqs. (III.1), (III.3) and (III.7) that

$$(W(q', p')\psi_1)(x') = \exp[ip'(x' + \tfrac{1}{2}\hbar q')]\psi_1(x' + \hbar q').$$

Substituting this into the integrand and changing to inner-product notation yields

$$\mathfrak{S}[\psi_1, \psi_2](q', p')$$
$$= (2\pi\hbar)^{-\frac{1}{2}} \int (W(q', p')\psi_1)(x')\bar{\psi}_2(x')\, dx'$$
$$= (2\pi\hbar)^{-\frac{1}{2}}(W(q', p')\psi_1, \psi_2)$$

and completes the proof.

Consequently, if $\psi_1 = \psi_2 = \psi$ for a wavefunction ψ in $L^2(R)$, then $\mathfrak{R}[\psi, \psi]$ and $\mathfrak{S}[\psi, \psi]$ have the form of the usual Wigner quasiprobability distribution function $\mathfrak{R}[\psi]$ and its characteristic function $\mathfrak{S}[\psi]$ indicated in Eqs. (I.4). An immediate corollary is that the Wigner function and its characteristic function are square-integrable functions on phase space and

$$\|\mathfrak{R}[\psi]\| = \|\mathfrak{S}[\psi]\| = |\psi|^2.$$

This statement has been asserted in the literature on the phase-space formulation of quantum mechanics; however, only heuristic proofs involving unjustifiable interchanges of orders of integrations have been indicated.

V. WEYL TRANSFORM: EXISTENCE AND UNIQUENESS

The existence and uniqueness of the Weyl correspondence as a map from $L^2(R \times R)$ into the algebra $\mathfrak{L}_c(L^2(R))$ will now be established using only the results of the previous section and the corollary of the Riesz–Frechet theorem which gives the form of bounded sesquilinear forms on Hilbert space.[16]

Proposition V.1: There exists a unique map $W : f \to W(f)$ of $L^2(R \times R)$ into $\mathfrak{L}_c(L^2(R))$ such that

$$(W(f)\psi_1, \psi_2) = \iint f(p, q)\mathfrak{R}[\psi_1, \psi_2](p, q)\, dp\, dq$$

for all ψ_1 and ψ_2 in $L^2(R)$ and all f in $L^2(R \times R)$; moreover,

[15] Reference 8, pp. 121, 128, 142.

[16] Reference 9, pp. 130–131.

$$(W(f)\psi_1, \psi_2) = \iint (F_{12}f)(q', p')\mathcal{S}[\psi_1, \psi_2](q', p')\, dq'\, dp'$$

for all ψ_1 and ψ_2 in $L^2(R)$ and all f in $L^2(R \times R)$.

Proof: Consider a fixed f in $L^2(R \times R)$ and a fixed pair ψ_1 and ψ_2 in $L^2(R)$. $\mathcal{R}[\psi_1, \psi_2]$ and $\bar{f}$ are elements of $L^2(R \times R)$. Let $\varphi_f(\psi_1, \psi_2)$ denote their inner product:

$$\varphi_f(\psi_1, \psi_2) = \langle \mathcal{R}[\psi_1, \psi_2], \bar{f} \rangle.$$

Since

$$\mathcal{R}[\psi_1, \psi_2] = F_{12}\mathcal{S}[\psi_1, \psi_2]$$

an alternative expression for $\varphi_f(\psi_1, \psi_2)$ is the following:

$$\begin{aligned}\varphi_f(\psi_1, \psi_2) &= \langle F_{12}\mathcal{S}[\psi_1, \psi_2], \bar{f} \rangle \\ &= \langle \mathcal{S}[\psi_1, \psi_2], F_{12}^{-1}\bar{f} \rangle,\end{aligned}$$

where the unitarity of F_{12} and the property analogous to Eq. (II.6b) have been used. The definition of the inner product in $L^2(R \times R)$ yields

$$\varphi_f(\psi_1, \psi_2) = \iint f(p, q)\mathcal{R}[\psi_1, \psi_2](p, q)\, dp\, dq$$

and

$$\varphi_f(\psi_1, \psi_2) = \iint (F_{12}f)(q', p')\mathcal{S}[\psi_1, \psi_2](q', p')\, dq'\, dp'.$$

Next consider the map $\psi_1, \psi_2 \to \varphi_f(\psi_1, \psi_2)$ for a fixed f in $L^2(R \times R)$. According to Proposition IV.1, $\psi_1, \psi_2 \to \mathcal{R}[\psi_1, \psi_2]$ is a sesquilinear map of $L^2(R)$ into $L^2(R \times R)$ such that

$$\|\mathcal{R}[\psi_1, \psi_2]\| = |\psi_1|\, |\psi_2|.$$

Consequently, it follows from the linearity of the inner product in the first variable and the definition of $\varphi_f(\psi_1, \psi_2)$ that $\varphi_f(\psi_1, \psi_2)$ is a sesquilinear form on $L^2(R)$. Moreover, φ_f is bounded, since by the Cauchy–Schwartz inequality

$$\begin{aligned}|\varphi_f(\psi_1, \psi_2)| &= |\langle \mathcal{R}[\psi_1, \psi_2], \bar{f} \rangle| \\ &\leq \|\mathcal{R}[\psi_1, \psi_2]\|\, \|\bar{f}\| \\ &\leq \|f\|\, |\psi_1|\, |\psi_2|.\end{aligned}$$

Therefore, there exists a unique continuous linear operator $W(f)$ on $L^2(R)$ such that

$$\varphi_f(\psi_1, \psi_2) = (W(f)\psi_1, \psi_2)$$

for all ψ_1 and ψ_2 in $L^2(R)$ which completes the proof

If f is in $L^2(R \times R)$, then

$$\begin{aligned}(W(f)\psi_1, \psi_2) &= \iint (F_{12}f)(q', p')\mathcal{S}[\psi_1, \psi_2](q', p')\, dq'\, dp' \\ &= (2\pi\hbar)^{-\frac{1}{2}} \iint (F_{12}f)(q', p')(W(q', p')\psi_1, \psi_2)\, dq'\, dp'\end{aligned} \tag{V.2}$$

for all ψ_1 and ψ_2 according to the above proposition and Proposition IV.2. In terms of the conventional notation for an operator determined by an integral expression[17] such as Eq. (V.2), $W(f)$ should be written as follows:

$$W(f) = (2\pi\hbar)^{-\frac{1}{2}} \iint (F_{12}f)(q', p')W(q', p')\, dq'\, dp'. \tag{V.3}$$

This is precisely the formal expression of Eq. (III.11) for which a well-defined meaning was sought. It should be noted that this goal has been achieved using only basic results from the theory of measures, harmonic analysis, and Hilbert space and without recourse to more sophisticated theories of integration of vector-valued functions. Because of Eqs. (V.2) and (V.3), it is appropriate to introduce the following definition.

Definition V.1: The map W from $L^2(R \times R)$ into $\mathcal{L}_0(L^2(R))$ defined in Proposition V.1 is called the Weyl transform.

It is appropriate that a remark about uniqueness of the Weyl transform be made at this point. Proposition V.1 asserts that the Weyl transform is the only map from classically observable functions to quantum-mechanical operators which allows computation of expectation values by evaluating phase space integrals of the classically observable functions with respect to the Wigner quasiprobability density function. Evidently, other choices of density functions lead to different correspondences between classically observable functions and quantum-mechanical operators.

VI. HILBERT–SCHMIDT OPERATORS AND H^*-ALGEBRAS

The algebra $\mathcal{L}_{HS}(L^2(R))$ of Hilbert–Schmidt operators on $L^2(R)$ will play an essential role in the further discussion of the Weyl transform; therefore, some properties of Hilbert–Schmidt operators on $L^2(R)$ will be briefly noted here.[18] The Hilbert–Schmidt operators on $L^2(R)$ are precisely the in-

[17] M. A. Naimark, *Normed Rings* (P. Noordhoff, Ltd., Groningen, The Netherlands, 1960), pp. 152–154.

[18] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960), pp. 29–36.

tegral operators on $L^2(R)$ with kernels in $L^2(R \times R)$. If h is an element of $L^2(R \times R)$, then the integral operator on $L^2(R)$ with kernel h will be denoted by S_h:

$$(S_h\psi)(x) = \int h(x, y)\psi(y)\, dy \qquad \text{(VI.1)}$$

for x in R and ψ in $L^2(R)$. A *-operation may be introduced in $L^2(R \times R)$: if h is an element of $L^2(R \times R)$, then h^* is the element of $L^2(R \times R)$ defined by

$$h^*(x, y) = \bar{h}(y, x) \qquad \text{(VI.2)}$$

for (x, y) in $R \times R$. This *-operation in $L^2(R \times R)$ corresponds to the *-operation of taking the operator adjoint,

$$(S_h)^* = S_{h^*}. \qquad \text{(VI.3)}$$

A multiplication of functions may also be introduced in $L^2(R \times R)$: if g and h are elements of $L^2(R \times R)$, then $g \circ h$ is the element of $L^2(R \times R)$ defined by

$$(g \circ h)(x, y) = \int g(x, z)h(z, y)\, dz \qquad \text{(IV.4)}$$

for (x, y) in $R \times R$. This multiplication corresponds to the multiplication of operators,

$$S_g S_h = S_{g \circ h}. \qquad \text{(VI.5)}$$

In order to summarize the properties of these operations and deduce properties of the Weyl correspondence, it is convenient to introduce the concept of an H^*-algebra.[19] An H^*-algebra is a complex Hilbert space H with a *-operation $a \to a^*$ and a multiplication $a, b \to ab$ which makes H into a Banach *-algebra relative to the inner product norm $|a| = (a, a)^{\frac{1}{2}}$ of the Hilbert space and, furthermore,

$$(ab, c) = (b, a^*c) \qquad \text{(VI.6)}$$

for all a, b, c in H. Consequently, if H is a Hilbert space and operations $a \to a^*$ and $a, b \to ab$ are given, then the requirements imposed on these operations to make H into an H^*-algebra are precisely the following:

$$a(b + c) = ab + ac, \quad (b + c)a = ba + ca, \qquad \text{(VI.7a)}$$

$$\lambda(ab) = (\lambda a)b = a(\lambda b), \qquad \text{(VI.7b)}$$

$$a(bc) = (ab)c, \qquad \text{(VI.7c)}$$

$$a^{**} = a, \qquad \text{(VI.7d)}$$

$$(a + b)^* = a^* + b^*, \qquad \text{(VI.7e)}$$

$$(ab)^* = b^*a^*, \qquad \text{(VI.7f)}$$

$$(\lambda a)^* = \bar{\lambda}a^*, \qquad \text{(VI.7g)}$$

$$|a^*| = |a|, \qquad \text{(VI.7h)}$$

$$|ab| \leq |a|\,|b|, \qquad \text{(VI.7i)}$$

$$(ab, c) = (b, a^*c), \qquad \text{(VI.7j)}$$

for all a, b, c in H and all complex numbers λ.

An example of an H^*-algebra is the set $\mathfrak{L}_{HS}(L^2(R))$ of all Hilbert–Schmidt operators on $L^2(R)$ when $\mathfrak{L}_{HS}(L^2(R))$ is equipped with the Hilbert–Schmidt inner product:

$$(A, B)_{HS} = \sum_n (A\psi_n, B\psi_n)$$

for A, B in $\mathfrak{L}_{HS}(L^2(R))$ where $\{\psi_n\}$ is any orthonormal basis of $L^2(R)$. Another example of an H^*-algebra is the Hilbert space $L^2(R \times R)$ equipped with the *-operation $h \to h^*$ and the multiplication $g, h \to g \circ h$ defined by Eqs. (VI.2) and (VI.4), respectively. Moreover, this H^*-algebra is isometrically *-isomorphic to the H^*-algebra $\mathfrak{L}_{HS}(L^2(R))$ under the map $h \to S_h$.

VII. WEYL TRANSFORM: EXPLICIT FORM

The existence and uniqueness of the Weyl transform has been demonstrated; however, it would be desirable to have a constructive procedure for obtaining the operator $W(f)$ from the function f. The purpose of this section is to prove that $W(f)$ is a Hilbert–Schmidt operator and exhibit its kernel. The proof of this assertion is accomplished by exhibiting the kernel corresponding to $W(f)$ in terms of a Fourier–Plancherel transform of f. The operator K defined on $L^2(R \times R)$ by

$$(Kh)(x, y) = (T^{-1}F_1h)(y, x) \qquad \text{(VII.1)}$$

for (x, y) in $R \times R$ and h in $L^2(R \times R)$ will be used to obtain the kernel. The essential properties of K are indicated in the following proposition.

Proposition VII.1: K is a unitary operator on $L^2(R \times R)$ such that

(i) $K = T^{-1}F_1^{-1}$ (VII.2)

(ii) if h is in $L^2(R \times R)$, then

$$K\bar{h} = (Kh)^* \qquad \text{(VII.3)}$$

(iii) if ψ_1, ψ_2 are in $L^2(R)$, then

$$\mathfrak{R}[\psi_1, \psi_2] = K^{-1}\psi_1 \cdot \bar{\psi}_2. \qquad \text{(VII.4)}$$

Proof: K is clearly unitary from its definition. Since functions of the form $\psi_1 \cdot \psi_2$ with ψ_1 and ψ_2

[19] C. E. Rickart, *General Theory of Banach Algebras* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1960), p. 272.

in $L^2(R)$ are dense in $L^2(R \times R)$, it suffices to prove

$$K\psi_1 \cdot \psi_2 = T^{-1}F_1^{-1}\psi_1 \cdot \psi_2$$

for all ψ_1 and ψ_2 in $L^2(R)$. If ψ_1 and ψ_2 are in $L^2(R)$, then

$$T^{-1}F_1^{-1}\psi_1 \cdot \psi_2 = T^{-1}(F^{-1}\psi_1) \cdot \psi_2$$

by the definition of F_1. If (x, y) in $R \times R$, then

$$\begin{aligned}[T^{-1}(F^{-1}\psi_1) \cdot \psi_2](x, y) &= (F^{-1}\psi_1)(\hbar^{-1}(x - y))\psi_2(\tfrac{1}{2}(x + y)) \\ &= (F\psi_1)(\hbar^{-1}(y - x))\psi_2(\tfrac{1}{2}(x + y)),\end{aligned}$$

where Eqs. (II.6c) and (II.13) have been used. On the other hand,

$$\begin{aligned}(K\psi_1 \cdot \psi_2)(x, y) &= (T^{-1}F_1\psi_1 \cdot \psi_2)(y, x) \\ &= (T^{-1}(F\psi_1) \cdot \psi_2)(y, x) \\ &= (F\psi_1)(\hbar^{-1}(y - x))\psi_2(\tfrac{1}{2}(y + x))\end{aligned}$$

for (x, y) in $R \times R$ where Eqs. (VI.1), the definition of F_1 and Eq. (II.13) have been used. Consequently, the asserted identity

$$K = T^{-1}F_1^{-1}$$

is proven.

Let h be an element of $L^2(R \times R)$. Then for (x, y) in $R \times R$

$$\begin{aligned}(K\bar{h})(x, y) &= (T^{-1}F_1\bar{h})(y, x) \\ &= (T^{-1}\overline{F_1^{-1}h})(y, x) \\ &= \overline{(T^{-1}F_1^{-1})(y, x)},\end{aligned}$$

where the property of F_1 analogous to Eq. (II.6a) has been used. Introducing the expression of Eq. (VII.2) for K yields

$$\begin{aligned}(Kh)(x, y) &= \overline{(Kh)(y, x)} \\ &= (Kh)^*(x, y),\end{aligned}$$

where the *-operation of Eq. (VI.2) has been introduced.

If ψ_1 and ψ_2 are in $L^2(R)$, then

$$\mathfrak{R}[\psi_1, \bar{\psi}_2] = F_1T\psi_1 \cdot \bar{\psi}_2$$

by definition; however,

$$K^{-1} = (T^{-1}F_1^{-1})^{-1} = F_1T$$

which yields Eq. (VII.4) and completes the proof.

The operator K on $L^2(R \times R)$ will now be used to obtain the kernel of the operator $W(f)$ on $L^2(R)$.

Proposition VII.2: If f is an element of $L^2(R \times R)$, then $W(f)$ is the integral operator on $L^2(R)$ with the kernel Kf; consequently, $W(f)$ is a Hilbert–Schmidt operator on $L^2(R)$.

Proof: Let ψ_1 and ψ_2 be arbitrary elements of $L^2(R)$. The result

$$(W(f)\psi_1, \psi_2) = \langle \mathfrak{R}[\psi_1, \psi_2], \bar{f} \rangle$$

was established in the proof of Proposition V.1. Equation (VII.4) for $\mathfrak{R}[\psi_1, \psi_2]$ and the unitarity of K yield

$$\begin{aligned}(W(f)\psi_1, \psi_2) &= \langle K^{-1}\psi_1 \cdot \bar{\psi}_2, \bar{f} \rangle \\ &= \langle \psi_1 \cdot \bar{\psi}_2, K\bar{f} \rangle.\end{aligned}$$

Since $K\bar{f} = (Kf)^*$, it follows that

$$\begin{aligned}(W(f)\psi_1, \psi_2) &= \langle \psi_1 \cdot \bar{\psi}_2, (Kf)^* \rangle \\ &= \iint \psi_1(x)\bar{\psi}_2(y)\overline{(Kf)^*}(x, y)\, dx\, dy \\ &= \iint (Kf)(y, x)\psi_1(x)\bar{\psi}_2(y)\, dx\, dy\end{aligned}$$

where the definition of * has been employed. Fubini's theorem justifies replacing the latter integral by an iterated integral; hence,

$$\begin{aligned}(W(f)\psi_1, \psi_2) &= \int \left[\int (Kf)(y, x)\psi_1(x)\, dx \right] \psi_2(y)\, dy \\ &= \int (S_{Kf}\psi_1)(y)\bar{\psi}_2(y)\, dy \\ &= (S_{Kf}\psi_1, \psi_2).\end{aligned}$$

Since ψ_1 and ψ_2 are arbitrary, it follows that

$$W(f) = S_{Kf}$$

and $W(f)$ is the integral operator with the square-integrable kernel Kf; therefore, $W(f)$ is a Hilbert–Schmidt operator.

The following question now arises: if A is an operator on $L^2(R)$, does there exist a function f in $L^2(R \times R)$ such that $W(f) = A$? The following proposition in conjunction with the previous proposition asserts that the answer is affirmative if and only if A is a Hilbert–Schmidt operator.

Proposition VII.3: If S_h is the integral operator on $L^2(R)$ with kernel h in $L^2(R \times R)$, then $f = K^{-1}h$ is the unique element of $L^2(R \times R)$ such that

$$S_h = W(f). \tag{VII.5}$$

Proof: If f' is an element of $L^2(R \times R)$ such that

$$S_{Kf'} = S_h,$$

which implies

$$Kf' = h$$

or

$$f' = K^{-1}f,$$

which completes the proof.

It is now evident that the replacement of the operator K by another operator K' with similar properties would lead to a different correspondence between classical observable functions and quantum mechanical operators and also a "quasiprobability density function" corresponding to a wavefunction which is distinct from the Wigner quasiprobability density function.

VIII. ALGEBRAIC ASPECTS OF THE WEYL TRANSFORM

The Weyl correspondence has been realized in the previous sections as a 1–1 map $W : f \to W(f)$ called the Weyl transform from $L^2(R \times R)$ the algebra $\mathfrak{L}_{HS}(L^2(R))$ of Hilbert–Schmidt operators on $L^2(R)$. Since two Hilbert–Schmidt operators may be multiplied together to yield another Hilbert–Schmidt operator, the Weyl transform may be used to introduce a multiplication of functions in $L^2(R \times R)$. Indeed, if f and g are element of $L^2(R \times R)$, then $f \times g$ will denote the element of $L^2(R \times R)$ defined by

$$f \times g = K^{-1}(Kf \circ Kg), \qquad \text{(VIII.1)}$$

where $\circ$ denotes the multiplication of kernels introduced in Eq. (VI.4). Clearly, K maps the multiplication $\times$ into the multiplication $\circ$:

$$K(f \times g) = Kf \circ Kg, \qquad \text{(VIII.2)}$$

for all f, g in $L^2(R \times R)$. The map $f \to \bar{f}$ is a candidate for a *-operation in $L^2(R \times R)$ with the multiplication $\times$; moreover, by assertion (ii) of Proposition VII.1:

$$K\bar{f} = (Kf)^* \qquad \text{(VIII.3)}$$

for all f in $L^2(R \times R)$.

Proposition VIII.1: $L^2(R \times R)$ with the operations $f \to \bar{f}$ and $f, g \to f \times g$ is an H^*-algebra. Moreover, K is an isometric *-isomorphism of the H^*-algebra induced by $^-$ and $\times$ onto the H^*-algebra induced by * and $\circ$.

Proof: The properties of Eq. (VI.7) for $^-$ and $\times$ follow directly from the corresponding property for * and $\circ$, since K is a unitary operator with the properties of Eqs. (VIII.2) and (VIII.3). This also implies K is an isometric *-isomorphism of these two H^*-algebras.

It was noted in Sec. VI that the Hilbert–Schmidt operators form an H^*-algebra. Many of the algebraic properties of the Weyl transform are obtained by recognizing that the Weyl transform is an isomorphism of H^*-algebras as indicated in the following proposition.

Propostion VIII.2: The Weyl transform $W : f \to W(f)$ is an isometric *-isomorphism of the H^*-algebra induced on $L^2(R \times R)$ by $^-$ and $\times$ onto the H^*-algebra $\mathfrak{L}_{HS}(L^2(R))$ of Hilbert–Schmidt operators on $L^2(R)$:

$$\text{(i)} \quad W(\bar{f}) = W(f)^*, \qquad \text{(VIII.4a)}$$

$$\text{(ii)} \quad W(f \times g) = W(f)W(g), \qquad \text{(VIII.4b)}$$

$$\text{(iii)} \quad W(\lambda f) = \lambda W(f), \qquad \text{(VIII.4c)}$$

$$\text{(iv)} \quad W(f + g) = W(f) + W(g), \qquad \text{(VIII.4d)}$$

$$\text{(v)} \quad \|W(f)\| \leq \|f\| = \|W(f)\|_{HS}. \qquad \text{(VIII.4e)}$$

Proof: Since $W(f) = S_{Kf}$, it suffices to remark that $f \to Kf$ is an isometric *-isomorphism of the H^*-algebra induced in $L^2(R \times R)$ by $^-$ and $\times$ onto the H^*-algebra induced in $L^2(R \times R)$ by * and $\circ$ and that $h \to S_h$ is an isometric *-isomorphism of the latter onto the H^*-algebra $\mathfrak{L}_{HS}(L^2(R))$.

The sine and cosine brackets[20] of the phase-space formulation of quantum mechanics are related to the multiplication $\times$. Indeed, if f and g are elements of both $L^1(R \times R)$ and $L^2(R \times R)$, then a tedious, but straightforward, computation yields the following:

$$(f \times g)(x, y) = \frac{1}{4}\left(\frac{\hbar}{2\pi}\right)^4 \iiiint f(x', y')g(x'', y'') \times \exp\left[-\frac{2i}{\hbar}\det\begin{vmatrix} 1 & x & y' \\ 1 & x & y' \\ 1 & x'' & y'' \end{vmatrix}\right] dx\, dy'\, dx''\, dy'' \qquad \text{(VIII.5a)}$$

$$[f, g]_-(x, y) = \frac{1}{2i}\left(\frac{\hbar}{2\pi}\right)^4 \iiiint f(x', y')g(x'', y'') \times \sin\left[\frac{2}{\hbar}\det\begin{vmatrix} 1 & x & y \\ 1 & x' & y' \\ 1 & x'' & y'' \end{vmatrix}\right] dx'\, dy'\, dx''\, dy'' \qquad \text{(VIII.5b)}$$

$$[f, g]_+(x, y) = \frac{1}{2}\left(\frac{\hbar}{2\pi}\right)^4 \iiiint f(x', y')g(x'', y'') \times \cos\left[\frac{2}{\hbar}\det\begin{vmatrix} 1 & x & y \\ 1 & x' & y' \\ 1 & x'' & y'' \end{vmatrix}\right] dx'\, dy'\, dx''\, dy'' \qquad \text{(VIII.5c)}$$

for (x, y), in $R \times R$ where

$$[f, g]_\pm = f \times g \pm g \times f. \qquad \text{(VIII.6)}$$

It should be noted that the nonlocality of the product arises from the fact that P and Q do not commute. In fact, the determinant appearing in

[20] Reference 6; see also G. A. Baker, Jr., Phys. Rev. 109, 2198 (1957) and C. L. Mehta, J. Math. Phys. 5, 677 (1964).

Eqs. (VIII.5) is expressible in terms of the commutator function B introduced in Eq. (III.8c):

$$\det \begin{vmatrix} 1 & x & y \\ 1 & x' & y' \\ 1 & x'' & y'' \end{vmatrix}$$

$$= (x'y'' - x''y') - (xy'' - x''y) + (xy' - x'y) \qquad \text{(VIII.7a)}$$

$$= B(x, y; x', y') + B(x', y'; x'', y'') + B(x'', y''; x, y) \qquad \text{(VIII.7b)}$$

for all (x, y), (x', y'), (x'', y'') in $R \times R$.

IX. COMMENTS AND GENERALIZATIONS

The previous sections have been restricted to quantum mechanics; however, the results have extensions to quantum statistical mechanics. The Weyl transform in conjunction with the B^*-algebraic approach to quantum theory leads in a natural fashion to the introduction of density operators corresponding to the mixed states of quantum statistical mechanics.[21,22] Furthermore, the extension of the discussion of Sec. IV to Wigner quasiprobability density functions and characteristic functions induced by mixed states yields a correspondence between Wigner functions and density operators implemented by the Weyl transform.

The configuration space and phase space of a linear classical mechanical system with n degrees of freedom are R^n and $R^n \times R^n$, respectively, where R^n is the Cartesian product of n copies of R. The analysis of the previous sections may be immediately extended from the case $n = 1$ to arbitrary finite n by replacing $L^2(R)$ and $L^2(R \times R)$ by $L^2(R^n)$ and $L^2(R^n \times R^n)$, respectively, and utilizing the theory of harmonic analysis for functions of several variables.

It was remarked in Sec. III that the essential aspect of the configuration space R involved in considering Weyl correspondence was its group structure: the additive group of reals is an example of a commutative locally compact topological group. The translational invariance of Lebesgue measure and harmonic analysis on R were the essential mathematical tools relating to R used in constructing the Weyl transform. These constructs have generalizations to Haar measure and abstract harmonic analysis on an arbitrary locally compact topological group which is commutative.[23] Therefore, the question arises whether the Weyl transform admits a similar generalization and, indeed, some results have been obtained for very general commutative locally compact topological groups.[24,25] These generalizations distinguish configuration space and momentum space; consequently, some conceptual questions about the Weyl transform are clarified. The removal of the locally compact condition on the group is a generalization which would be extremely important since the replacement of the configuration space R by an infinite-dimensional real Hilbert space corresponds to the transition to a physical system with infinitely many degrees of freedom. The current developments of harmonic analysis on Hilbert space greatly encourage the possibility of considering the Weyl correspondence and Wigner quasiprobability density function in quantum field theory.

After completion of the investigations presented in the previous sections, recent papers containing some similar results were found. The approaches of Segal[24] and Kastler[26] consider the Weyl transform as a map from the Fourier–Plancherel transforms of the functions representing classical observables to operators instead of directly from the classically observable functions to quantum-mechanical operators. These approaches do not involve the explicit use of the Wigner quasiprobability density function. The essential aspects of the proof of Proposition IV.2 have appeared in a paper by McKenna and Klauder[27]; however, the relevance to the Wigner quasiprobability density function was not indicated.

The introduction of the Wigner function and its characteristic function in the study of the mathematical aspects of the Weyl correspondence permits direct proofs of results relating to the Weyl transform based on basic theorems of measure theory, Hilbert space theory, and harmonic analysis. Moreover, it has resulted in an explicit representation of the quantum-mechanical operator $W(f)$ corresponding to a classically observable function f in $L^2(R \times R)$ as a Hilbert–Schmidt integral operator on $L^2(R)$ whose kernel is Kf, where K is a "twisted" Fourier–Plancherel transform. Furthermore, the relation of the Wigner quasiprobability density function to the Weyl correspondence has been clarified by exhibiting the Wigner function $\mathfrak{R}[\psi]$ corresponding to a wavefunction ψ in $L^2(R)$ in terms of K^{-1}, the inverse

[21] J. C. T. Pool, "B^*-Algebras and the Logic of Quantum Mechanics" (unpublished).
[22] J. C. T. Pool, "Density Operators and Wigner Functions" (unpublished).
[23] L. H. Loomis, *Abstract Harmonic Analysis* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1953).
[24] I. E. Segal, Math. Scand. 13, 31 (1963).
[25] J. C. T. Pool, Lecture Notes, Seminar in Theoretical Physics, University of Iowa (1962).
[26] D. Kastler, "The C^*-algebras of a Free Boson Field: I. Discussion of Basic Facts," Commun. Math. Phys. 1, 14 (1965).
[27] J. McKenna and J. R. Klauder, J. Math. Phys. 5, 878 (1964).

of the "twisted" Fourier–Plancherel transform K: namely, $\mathfrak{R}[\psi] = K^{-1}\psi \cdot \bar{\psi}$.

Note added to proof. The approach of Kastler[26] has been utilized to obtain results similar to Proposition VIII.2 and to study the relation between density operators and quasiprobability density functions[22] in the following: G. Loupias and S. Miracle-Sole, "Sur le formalisme de la convolution gauche," Université d'Aix-Marseille Preprint (May, 1965).

ACKNOWLEDGMENTS

This research was initiated while the author was at the Scientific Laboratory, Ford Motor Company, and completed while the author was on leave of absence as a National Science Foundation Postdoctoral Fellow at Brandeis University.

The author would like to acknowledge many helpful discussions with A. B. Ramsay and express appreciation to R. W. Goodman, J. R. Klauder, and H. S. Leff for pointing out related literature. The author wishes to express his gratitude to M. Dresden for introducing him to the Weyl correspondence and also for continuing enthusiastic support and criticism. The author is especially grateful to the National Science Foundation for a Postdoctoral Fellowship, to S. S. Schweber and the members of the Department of Physics for their hospitality during the tenure of the fellowship at Brandeis University, and to the management of the Scientific Laboratory, Ford Motor Company, for granting the leave of absence which permitted acceptance of the Fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 1 JANUARY 1966

Physical Region Decoupling Procedure for Coupled Channel Scattering Amplitudes*

GORDON N. FLEMING†
Physics Department, University of Washington, Seattle, Washington
(Received 22 June 1965)

A method for expressing unitary, coupled channel, scattering amplitudes in terms of amplitudes satisfying "uncoupled" or "elastic" unitarity equations is presented. The method is given a physical interpretation by relating the equations to equations of the Heitler type. It is shown how continuous channels may be incorporated into the formalism, in some circumstances, without requiring the solution of integral equations. The influence of inelastic channels on pure elastic scattering is mentioned briefly and a pseudoelastic form of the exact unitarity equation is discussed. No applications of the method are undertaken here.

1. INTRODUCTION

IN this paper, a method for expressing scattering amplitudes satisfying coupled channel unitarity equations in terms of amplitudes satisfying "uncoupled" or "elastic" unitarity equations is presented. Many methods of expressing unitary scattering amplitudes in terms of *Hermitian* amplitudes already exist.[1-3] These methods are attractive because the Hermitian amplitude involved may be chosen arbitrarily as far as the unitarity constraint is concerned. Also, for two-particle channels, the Heitler equation, when formulated in the angular momentum representation, can be solved algebraically.[1] The other methods require the solution of linear integral equations.

Notwithstanding the existence of these methods, it seems desirable to possess a procedure for constructing fully coupled unitary amplitudes from amplitudes describing two sets of channels with no coupling between the sets. The original amplitudes might occur as the result of a calculation based on the familiar methods mentioned above. Such a procedure can probably be formulated in several

* Supported in part by the U. S. Atomic Energy Commission.
† Present address: Department of Physics, The Pennsylpania State University, University Park, Pennsylvania.
[1] W. Heitler, Proc. Cambridge Phil. Soc. 37, 291 (1941); See also M. L. Goldberger, Phys. Rev. 84, 926 (1951).
[2] M. Baker, Ann. Phys. (N. Y.) 4, 271 (1958).
[3] G. F. Chew and S. Mandelstam, Phys. Rev. 119, 467 (1960).

of the "twisted" Fourier–Plancherel transform K: namely, $\mathfrak{R}[\psi] = K^{-1}\psi \cdot \bar{\psi}$.

Note added to proof. The approach of Kastler[26] has been utilized to obtain results similar to Proposition VIII.2 and to study the relation between density operators and quasiprobability density functions[22] in the following: G. Loupias and S. Miracle-Sole, "Sur le formalisme de la convolution gauche," Université d'Aix-Marseille Preprint (May, 1965).

ACKNOWLEDGMENTS

This research was initiated while the author was at the Scientific Laboratory, Ford Motor Company, and completed while the author was on leave of absence as a National Science Foundation Postdoctoral Fellow at Brandeis University.

The author would like to acknowledge many helpful discussions with A. B. Ramsay and express appreciation to R. W. Goodman, J. R. Klauder, and H. S. Leff for pointing out related literature. The author wishes to express his gratitude to M. Dresden for introducing him to the Weyl correspondence and also for continuing enthusiastic support and criticism. The author is especially grateful to the National Science Foundation for a Postdoctoral Fellowship, to S. S. Schweber and the members of the Department of Physics for their hospitality during the tenure of the fellowship at Brandeis University, and to the management of the Scientific Laboratory, Ford Motor Company, for granting the leave of absence which permitted acceptance of the Fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 1 JANUARY 1966

Physical Region Decoupling Procedure for Coupled Channel Scattering Amplitudes*

GORDON N. FLEMING†
Physics Department, University of Washington, Seattle, Washington
(Received 22 June 1965)

A method for expressing unitary, coupled channel, scattering amplitudes in terms of amplitudes satisfying "uncoupled" or "elastic" unitarity equations is presented. The method is given a physical interpretation by relating the equations to equations of the Heitler type. It is shown how continuous channels may be incorporated into the formalism, in some circumstances, without requiring the solution of integral equations. The influence of inelastic channels on pure elastic scattering is mentioned briefly and a pseudoelastic form of the exact unitarity equation is discussed. No applications of the method are undertaken here.

1. INTRODUCTION

IN this paper, a method for expressing scattering amplitudes satisfying coupled channel unitarity equations in terms of amplitudes satisfying "uncoupled" or "elastic" unitarity equations is presented. Many methods of expressing unitary scattering amplitudes in terms of *Hermitian* amplitudes already exist.[1-3] These methods are attractive because the Hermitian amplitude involved may be chosen arbitrarily as far as the unitarity constraint is concerned. Also, for two-particle channels, the Heitler equation, when formulated in the angular momentum representation, can be solved algebraically.[1] The other methods require the solution of linear integral equations.

Notwithstanding the existence of these methods, it seems desirable to possess a procedure for constructing fully coupled unitary amplitudes from amplitudes describing two sets of channels with no coupling between the sets. The original amplitudes might occur as the result of a calculation based on the familiar methods mentioned above. Such a procedure can probably be formulated in several

* Supported in part by the U. S. Atomic Energy Commission.
† Present address: Department of Physics, The Pennsylpania State University, University Park, Pennsylvania.
[1] W. Heitler, Proc. Cambridge Phil. Soc. 37, 291 (1941); See also M. L. Goldberger, Phys. Rev. 84, 926 (1951).
[2] M. Baker, Ann. Phys. (N. Y.) 4, 271 (1958).
[3] G. F. Chew and S. Mandelstam, Phys. Rev. 119, 467 (1960).

ways[4] and the particular method described in Sec. 2 is "interpreted" in Sec. 3 by relating it to equations of the Heitler type.

It is worth noting that the method can be reversed to reduce the problem of solving a coupled many-channel problem to that of solving two such problems for a partition of the channels into two smaller subsets. Such a reduction may be desirable if, for example, the two subsets involve two-particle and three-particle channels, respectively. In Sec. 4, it is shown that if one begins with unitary amplitudes for channels described by a continuous parameter and amplitudes for discrete channels, then the coupling between the two sets of channels can be introduced and the fully coupled amplitudes can be calculated without any inversion of continuous matrices (solving of integral equations) being required.

Finally, in Sec. 5 some of the quantities introduced in Sec. 2 are related to familiar quantities in pure elastic scattering. In particular, expressions are obtained for the change in cot δ, where δ is the partial-wave phase shift, due to the coupling to inelastic channels. Also a pseudoelastic form of the exact unitarity equation is discussed and related to the process of diffraction scattering.[5]

2. THE DECOUPLING PROCEDURE

The quantities of interest are the *reduced* matrix elements of the scattering operator, T, which satisfies the unitarity equation

$$T - T^\dagger = 2iT^\dagger T. \tag{2.1}$$

Denoting all variables except the total four momentum, P_μ, by i, j, $\cdots$, etc., the reduced matrix element, $t_{ij}(s)$, where $s = P^2$, is defined by

$$\delta^4(P' - P)t_{ij}(s) \equiv \langle P', i\,|T|\,P, j\rangle. \tag{2.2}$$

It is customary to either show[6] or assume[7] that the $t_{ij}(s)$ can be defined below their respective physical thresholds in such a way that for all s greater than the lowest threshold, s_0 say, of any channel coupled to $t_{ij}(s)$ the *extended unitarity* equation,

$$t_{ij}(s) - t_{ji}(s)^* = 2i \sum_n t_{ni}(s)^* \, \delta^4(P_n - P_i) t_{nj}(s), \tag{2.3}$$

holds. The summation symbol denotes the appropriate combination of sums over discrete variables and integrals over continuous variables required for the insertion of a complete set of states between the scattering operators on the right-hand side of (2.1). For strong and electromagnetic interactions time reversal invariance[8] can be invoked to show that the variables i, j, $\cdots$, etc. can be chosen to make $t_{ij}(s)$ symmetric,[9] in which case the left-hand side of (2.3) is equal to $2i$ Im $t_{ij}(s)$. If the $t_{ij}(s)$ are gathered together in a square matrix, $f(s)$, which may be continuously infinite or partly so, and if the integrations required to eliminate the delta functions on the right-hand side of (2.3) are performed, thereby generating phase space factors which may also be gathered into a diagonal matrix, $R(s)$, then the unitarity equation reads

$$f(s) - f(s)^* = 2i\, f(s)^* R(s)\, f(s). \tag{2.4}$$

It is to be understood that the indicated matrix multiplication includes all those summations and integrations left over after the delta functions in (2.3) were eliminated. Finally, it should be remembered that each diagonal element of $R(s)$ vanishes below the physical threshold of the particular channel being described by that element.

Now suppose that of all the channels contributing to (2.4), a certain subset is more interesting than the rest. Denoting that subset by the subscript 1 and the remaining channels by the subscript 2, it is always possible to arrange $f(s)$ and $R(s)$ so that they have the form

$$f(s) \equiv \begin{bmatrix} f_{11}(s), & f_{12}(s) \\ f_{21}(s), & f_{22}(s) \end{bmatrix}; \quad R(s) \equiv \begin{bmatrix} R_1(s), & 0 \\ 0, & R_2(s) \end{bmatrix}. \tag{2.5}$$

In such a circumstance where interest is focused on the submatrix $f_{11}(s)$ the following theorem is interesting.

Theorem: If the matrix functions $h_{12}(s)$, $h_{21}(s)$, and $g_{22}(s)$ satisfy the relations

$$f_{12}(s) = f_{11}(s)h_{12}(s); \quad f_{21}(s) = h_{21}(s)f_{11}(s) \tag{2.6a}$$

[4] For example, the uncoupled phase method of M. Ross and G. Shaw, Ann. Phys. (N. Y.) 9, 391 (1960) is motivated by considerations similar to those presented here. For the extension of this method to the relativistic case and further references see P. Nath and G. Shaw, Phys. Rev. 137, B711 (1965). Another interesting approach, albeit an approximate one, is that of K. Gottfried and J. D. Jackson, Nuovo Cimento 34, 735 (1964); see also J. S. Ball and W. R. Frazer, Phys. Rev. Letters 14, 746 (1965).

[5] R. Glauber, *Lectures in Theoretical Physics* (Interscience Publishers Inc., New York, 1958), p. 315.

[6] M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964), Chap. 10, pp. 593, 609, 638, 650. An excellent set of references may be found at the end of this chapter.

[7] S. Mandelstam, Phys. Rev. 112, 1344 (1958); M. L. Goldberger, M. T. Grisaru, S. W. MacDowell, and D. Y. Wong, *ibid.* 120, 2250 (1960); M. Muraskin and K. Nishijima, *ibid.* 122, 331 (1961); E. J. Squires, *Strong Interactions and High Energy Physics*, edited by R. G. Moorhouse (Plenum Press, New York, 1964), pp. 30–32.

[8] Time-reversal invariance may not hold in weak interactions. See, for example, J. H. Christensen *et al.*, Phys. Rev. Letters 13, 138 (1964); J. A. Anderson *et al.*, *ibid.* 14, 475 (1965); W. Galbraith *et al.*, *ibid.*, 383.

[9] Reference 6, pp. 351, 891.

and

$$f_{22}(s) = g_{22}(s) + h_{21}(s)f_{11}(s)h_{12}(s), \quad (2.6b)$$

then the unitarity equations

$$f_{11}(s) - f_{11}(s)^* = 2i\, f_{11}(s)^* \times \{R_1(s) + h_{12}(s)^* R_2(s) h_{21}(s)\} f_{11}(s), \quad (2.7a)$$

$$h_{12}(s) - h_{12}(s)^* = 2i\, h_{12}(s)^* R_2(s) g_{22}(s), \quad (2.7b)$$

$$h_{21}(s) - h_{21}(s)^* = 2i\, g_{22}(s)^* R_2(s) h_{21}(s), \quad (2.7c)$$

and

$$g_{22}(s) - g_{22}(s)^* = 2i\, g_{22}(s)^* R_2(s) g_{22}(s) \quad (2.7d)$$

hold.

The proof of the theorem follows in a straightforward manner, albeit after some tedious algebra, upon substituting Eqs. (2.6) into (2.5) and then (2.5) into (2.4). Since one can always choose $h_{21}(s) = \tilde{h}_{12}(s)$, the result (2.7c) may seem redundant after (2.7b). It is stated separately to emphasize that, given (2.4)–(2.6), the symmetry properties of the h and g matrices play no role in deriving the equations (2.7).

The interest of the theorem lies in the fact that if the h matrices can be determined or estimated then $f_{11}(s)$ satisfies a closed unitarity equation with a nondiagonal effective phase-space matrix, $R_1(s) + h_{12}(s)^* R_2(s) h_{21}(s)$. At the same time the estimation of the h's is presumably facilitated by the absence of unitary coupling of the h's to the channels of primary interest. It is clear, however, that the estimation of the h's will not, in general, be a simple matter and can never be implemented until one has established a clear physical interpretation of the h's and g's. To this end, one must obtain further algebraic consequences of the theorem.

The theorem can be applied to the channels 2 and 1 in the same way it was applied to channels 1 and 2. Thus one defines $l_{12}(s)$, $l_{21}(s)$, and $g_{11}(s)$ by

$$f_{12}(s) = l_{12}(s) f_{22}(s); \qquad f_{21}(s) = f_{22}(s) l_{21}(s) \quad (2.8a)$$

and

$$f_{11}(s) = g_{11}(s) + l_{12}(s) f_{22}(s) l_{21}(s). \quad (2.8b)$$

These definitions yield the unitarity equations

$$f_{22}(s) - f_{22}(s)^* = 2i\, f_{22}(s)^* \times \{l_{21}(s)^* R_1(s) l_{12}(s) + R_2(s)\} f_{22}(s), \quad (2.9a)$$

$$l_{21}(s) - l_{21}(s)^* = 2i\, l_{21}(s)^* R_1(s) g_{11}(s), \quad (2.9b)$$

$$l_{12}(s) - l_{12}(s)^* = 2i\, g_{11}(s)^* R_1(s) l_{12}(s), \quad (2.9c)$$

$$g_{11}(s) - g_{11}(s)^* = 2i\, g_{11}(s)^* R_1(s) g_{11}(s). \quad (2.9d)$$

Now if (2.8a) and (2.6a) are substituted, in that order, into the second term on the right-hand side of (2.8b), then there follows

$$(1 - l_{12}(s)h_{21}(s)) f_{11}(s) = g_{11}(s). \quad (2.10a)$$

Similarly,

$$(1 - h_{21}(s) l_{12}(s)) f_{22}(s) = g_{22}(s). \quad (2.10b)$$

Furthermore, if (2.8b) is multiplied on the right by $h_{12}(s)$ and (2.6b) on the left by $l_{12}(s)$, then a comparison of the resulting equations with (2.6a) and (2.8a) yields

$$g_{11}(s) h_{12}(s) = l_{12}(s) g_{22}(s).$$

Assuming the g's to have inverses, one finds

$$h_{12}(s) = k_{12}(s) g_{22}(s), \quad (2.11a)$$

$$l_{12}(s) = g_{11}(s) k_{12}(s), \quad (2.11b)$$

where

$$k_{12}(s) = k_{12}(s)^*. \quad (2.12)$$

Substituting (2.11) into (2.10) there results the *integral equations*[10]

$$f_{11}(s) = g_{11}(s) + g_{11}(s) k_{12}(s) g_{22}(s) k_{21}(s) f_{11}(s) \quad (2.13a)$$

and

$$f_{22}(s) = g_{22}(s) + g_{22}(s) k_{21}(s) g_{11}(s) k_{12}(s) f_{22}(s). \quad (2.13b)$$

The solutions of these integral equations express the coupled matrix functions $f_{11}(s)$ and $f_{22}(s)$ in terms of the uncoupled matrix functions, $g_{11}(s)$ and $g_{22}(s)$, and the real matrix functions, $k_{12}(s)$ and $k_{21}(s) = \tilde{k}_{12}(s)$. The corresponding integral equations for $f_{12}(s)$ and $f_{21}(s)$ come from substituting (2.11) and (2.13a, b) into (2.6a) and (2.8a). They are

$$f_{12}(s) = g_{11}(s) k_{12}(s) g_{22}(s) + g_{11}(s) k_{12}(s) g_{22}(s) k_{21}(s) f_{12}(s), \quad (2.13c)$$

and

$$f_{21}(s) = g_{22}(s) k_{21}(s) g_{11}(s) + g_{22}(s) k_{21}(s) g_{11}(s) k_{12}(s) f_{21}(s). \quad (2.13d)$$

One can easily guess the correct form for the analogous integral equations in which the kernels on the right-hand side multiply the f's from the right.

[10] Strictly speaking, these are integral equations only if the channels require a continuous parameter for their description. Otherwise they are algebraic equations. The interesting case in which channels 1 are discrete is treated in Sec. 4.

3. PHYSICAL INTERPRETATION

The iterative "solutions" of the equations (2.13) are

$$f_{11} = g_{11} + g_{11}k_{12}g_{22}k_{21}g_{11} + \cdots, \tag{3.1a}$$

$$f_{12} = g_{11}k_{12}g_{22} + g_{11}k_{12}g_{22}k_{21}g_{11}k_{12}g_{22} + \cdots, \tag{3.1b}$$

$$f_{21} = g_{22}k_{21}g_{11} + g_{22}k_{21}g_{11}k_{12}g_{22}k_{21}g_{11} + \cdots, \tag{3.1c}$$

and

$$f_{22} = g_{22} + g_{22}k_{21}g_{11}k_{12}g_{22} + \cdots. \tag{3.1d}$$

If the nth terms in the series for f_{11} and f_{22} are denoted by $f_{11}^{(n-1)}$ and $f_{22}^{(n-1)}$, respectively, while the nth terms in the series for f_{12} and f_{21} are denoted by $f_{12}^{(n)}$ and $f_{21}^{(n)}$, respectively, then the *perturbative unitarity* relations[11]

$$f_{ij}^{(n)} - f_{ij}^{(n)*} = 2i \sum_{k=1}^{2} \sum_{r=0}^{n} f_{ik}^{(r)*} R_k f_{kj}^{(n-r)} \tag{3.2}$$

are a consequence of (2.7d), (2.9d), and (2.12), where $f_{12}^{(0)} \equiv f_{21}^{(0)} \equiv 0$.

These equations suggest that $f_{11}^{(0)}$ and $f_{22}^{(0)}$ include contributions from those processes in which channels 2 and 1, respectively, never appear as an intermediate state on the mass shell. This statement must be qualified, however, with the comment that $f_{ii}^{(0)}$ does not include *all* the virtual driving processes in f_{ii} as a glance at the higher-order terms in (3.1) a, d) shows.[12] The amplitudes $f_{12}^{(1)}$ and $f_{21}^{(1)}$ are interpreted as summing all those contributions from processes in which the transition from channels 1 to 2 or vice versa is not made more than once in an intermediate process on the mass shell. Again the qualification regarding driving forces applies.

The unitarity equations for $f_{11}^{(0)}$, $f_{12}^{(1)}$, $f_{21}^{(1)}$, $f_{22}^{(0)}$ are satisfied by the solutions of the linear equations of the Heitler form,[13]

$$f_{11}^{(0)} = b_{11} + ib_{11}R_1 f_{11}^{(0)}, \tag{3.3a}$$

$$f_{12}^{(1)} = b_{12} + ib_{11}R_1 f_{12}^{(1)} + ib_{12}R_2 f_{22}^{(0)}, \tag{3.3b}$$

$$f_{21}^{(1)} = b_{21} + ib_{21}R_1 f_{11}^{(0)} + ib_{22}R_2 f_{21}^{(1)}, \tag{3.3c}$$

$$f_{22}^{(0)} = b_{22} + ib_{22}R_2 f_{22}^{(0)}. \tag{3.3d}$$

On the other hand, from (3.1b, c),

$$f_{12}^{(1)} = f_{11}^{(0)}k_{12}f_{22}^{(0)}; \qquad f_{21}^{(1)} = f_{22}^{(0)}k_{21}f_{11}^{(0)}, \tag{3.4}$$

and upon substituting (3.4) into (3.3) and then using (3.3a, d) in (3.3b, c), one obtains

$$b_{11}k_{12}b_{22} = b_{12}. \tag{3.5}$$

The equations (3.1)–(3.5) relate the quantities g_{11}, g_{22}, k_{12}, and k_{21} to quantities occurring in more familiar formalisms for constructing unitary scattering amplitudes and may, therefore, be said to suggest the physical interpretation of the former quantities. This question, which must be considered in some detail in any particular application of the decoupling procedure, is not discussed further here.

4. DISCRETE AND CONTINUOUS CHANNELS

To this point it has been assumed that whenever matrix inversions were desired, they could be performed. In many instances of physical interest[14] this is not the case, and it is worthwhile to ask if certain of the previous results might be obtained without recourse to all the matrix inversions that have been formally employed. Thus if $f_{11}(s)$ is a discrete matrix of partial-wave amplitudes for two-particle channels, while $f_{22}(s)$ is a continuous matrix for the partial-wave amplitude of a three-particle channel, then inversion of (1, 1) matrices is routine algebra while the inversion of (2, 2) matrices can rarely be carried out in closed form if at all. It turns out that, for the preceding sections, if matrix inversion is feasible for one set of channels, the set 1 say, then it can be avoided entirely for the channels, 2.

Thus the equations (2.7) can be derived from the premises (2.6) if the channels, 1, are discrete but the equations (2.9b, c) must be modified by multiplication from the left by $f_{22}(s)^*$ and from the right by $f_{22}(s)$, respectively. Equations (2.10) and (2.11a) can still be derived, but (2.11b) and (2.12) must be multiplied from the right by $g_{22}(s)$. Finally with these modifications the integral equations, (2.13), can be derived as stated and it remains to show that they can be solved without matrix inversions of (2, 2) matrices.

The solution of (2.13a) is trivial since it involves only the inversion of a (1, 1) matrix

$$f_{11} = (1 - g_{11}k_{12}g_{22}k_{21})^{-1}g_{11}. \tag{4.1}$$

To solve for $f_{22}(s)$, multiply (2.13b) from the left by $k_{12}(s)$ to obtain

$$k_{12}f_{22} = k_{12}g_{22} + k_{12}g_{22}k_{21}g_{11}k_{12}f_{22}.$$

[11] These are not quite the same as the "unitarity" equations satisfied by the terms of conventional perturbation theory. See J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1964), Chap. 8, p. 160.

[12] In particular the "unphysical" singularity structure of f_{ij} receives contributions from every term in the series (3.1).

[13] Equations (3.3–(3.5) remain valid if the R_i on the right-hand side of (3.3) are replaced by any G_i such that Im $G_i = R_i$. Such a replacement is desirable in a discussion based on analyticity properties.

[14] For example, the coupled systems $(2\pi, 4\pi)$, etc.

This can be solved for $k_{12}f_{22}$;

$$k_{12}f_{22} = (1 - k_{12}g_{22}k_{21}g_{11})^{-1}k_{12}g_{22}.$$

Equation (2.13b) is then solved by substituting this result into the right-hand side of (2.13b), which yields

$$f_{22} = g_{22} + g_{22}k_{21}g_{11}(1 - k_{12}g_{22}k_{21}g_{11})^{-1}k_{12}g_{22}. \quad (4.2)$$

The equations (2.13c, d) may be solved in a similar way. Note that (4.2) has the form of (2.6b)

$$f_{22} = g_{22} + g_{22}k_{21}f_{11}k_{12}g_{22},$$

since it follows from (4.1) that

$$f_{11} = g_{11}(1 - k_{12}g_{22}k_{21}g_{11})^{-1}.$$

Finally since the derivation of (3.5) requires a matrix inversion in one set of channels only, it can still be carried through by looking first at k_{21} and then using the symmetry $k_{12} = k_{21}$.

5. TWO-PARTICLE ELASTIC SCATTERING

The simplest case for $f_{11}(s)$ is that of partial-wave elastic scattering in one two-particle channel. For convenience consider the common circumstance in which conservation laws allow one to work with a single function, $f_{11}(s)$, of one variable. If there exists a pure elastic interval, $s_0 \leq s \leq s_1$, in which no inelastic processes can occur, then in that interval both $g_{11}(s)$ and $f_{11}(s)$ have the form

$$g_{11}(s) = e^{i\delta^0} \sin \delta^0/R_1 = [R_1(\cot \delta^0 - i)]^{-1}, \quad (5.1)$$

$$f_{11}(s) = e^{i\delta} \sin \delta/R_1 = [R_1(\cot \delta - i)]^{-1}. \quad (5.2)$$

Upon substituting these expressions into (4.1), a little algebraic manipulation yields

$$R_1 \cot \delta = R_1 \cot \delta^0 - k_{12}g_{22}k_{21}. \quad (5.3)$$

This equation can be used to establish the influence of unitary coupling to inelastic channels on such low-energy parameters as the scattering length and effective range. It is also clear from (5.3) how resonances below the inelastic threshold can arise from strong coupling to the inelastic channel in this formalism. Such a mechanism has frequently been regarded as the explanation for the existence of some of the higher resonances in pion and nucleon interactions.[15]

The exact unitarity equation for $f_{11}(s)$ is, from (2.7a) and (2.11a),

$$f_{11} - f_{11}^* = 2i\,|f_{11}|^2\,\{R_1 + k_{12}g_{22}^*R_2g_{22}k_{21}\}. \quad (5.4)$$

On the basis of the discussion in Sec. 3, the function $b_{11}k_{12}g_{22} \equiv H_{12}$ satisfies[13]

$$H_{12} = b_{12} + ib_{12}R_2g_{22}, \quad (5.5)$$

and may be interpreted as the matrix amplitude describing the inelastic scattering that would occur if, after initial absorption, the system never returns to the elastic channel in a mass-shell intermediate state. This situation is just that which yields pure diffraction scattering in the elastic channel if no elastic driving forces are present. Consequently, on the basis of the optical theorem[16] one may expect the approximation of replacing $H_{12}R_2H_{21}$ by the appropriately normalized inelastic cross section of a diffraction scattering model to be valid at high energies. With a given b_{11}, then, the construction of a solution to (5.4) yields an amplitude incorporating corrections to the pure diffraction model. Such a method may have a bearing on recent studies of the high-energy behavior of the real parts of elastic scattering amplitudes.[17] Finally, it would be interesting to use (5.4) to investigate the influence of inelastic unitarity in the low and medium energy range through the N/D formalism since (5.4) has the algebraic form of pure elastic unitarity.[18]

[15] J. S. Ball and W. R. Frazer, Phys. Rev. Letters **7**, 204 (1961); L. F. Cook, Jr. and B. W. Lee, Phys. Rev. **127**, 183, 297 (1962); J. S. Ball, W. R. Frazer and M. Nauenberg, Phys. Rev. **128**, 478 (1962). For the general influence of closed channels on resonances see J. R. Fulco, G. L. Shaw, and D. Y. Wong, Phys. Rev. **137**, B1242 (1965).

[16] See Reference 6, p. 184.

[17] W. N. Cottingham and R. F. Peierls, Phys. Rev. **137**, B147 (1965); N. N. Khuri and F. Kinoshita, Phys. Rev. **137**, B720 (1965). A recent preprint with further references is Y. S. Jin and S. W. MacDowell, "Phase Representation and High Energy Behaviour of the Forward Scattering Amplitude," preprint, Institute for Advanced Study, Princeton, New Jersey (1965).

[18] The N/D formalism has been generalized to allow for the influence of inelastic channels in several ways. See, for example, J. D. Bjorken, Phys. Rev. Letters **4**, 473 (1960); R. Blankenbecher, Phys. Rev. **122**, 983 (1961); G. Frye and R. L. Warnock, *ibid.* **130**, 478 (1963). These various generalizations may not be consistent with one another. See M. Bander, P. W. Coulter, and G. L. Shaw, Phys. Rev. Letters **14**, 270 (1965).

Relation between the Onsager and Pfaffian Methods for Solving the Ising Problem. II. The General Lattice

C. A. HURST
University of Adelaide, Adelaide, South Australia
(Received 22 March 1965)

The considerations of a previous paper are extended to the case of a general lattice. It is shown algebraically why the Pfaffian and Onsager methods of solution of the Ising problem coincide only when the lattice is *planar*, and that the problem is then a linear one. When the lattice is nonplanar the Pfaffian method breaks down due to the appearance of unwanted negative signs, and it is shown how the Onsager method compensates for this at the expense of making the problem nonlinear.

1. INTRODUCTION

IN a previous paper,[1] it was shown that two expressions, one the partition function for the Ising problem on a rectangular lattice and the other a Pfaffian, were equal. The proof was algebraic in character, but could be interpreted as justifying the combinatorial approach which has been widely used.[2] With certain types of edge conditions the correspondence was shown not to be immediate and the source of the discrepancy was traced to the appearance of the long-range bonds associated with the imposition of cyclic or helical boundary conditions. It was claimed that the same difficulty would arise, although in a much more serious manner, when an attempt was made to identify a Pfaffian with the partition function for a lattice with a more complicated topological structure. Such an increase in topological complication arises whenever a sufficiently large number of connections hold between lattice points in a plane lattice, or when the lattice is three dimensional. It is the purpose of this paper to show in detail how the attempt to transform a general Ising model partition function into a Pfaffian will be defeated whenever the lattice is *nonplanar*. Nonplanar means that the set of lattice points and the bonds connecting them cannot be arranged in a plane in any position whatsoever without bonds crossing at points which are not lattice points. It is not the long range of the bonds in a nonplanar lattice which is significant, but the appearance of crossing points which cannot be removed by any continuous deformation of the lattice.

In Sec. 2 the algebraic relationship between the Pfaffian and Onsager methods for a general lattice will be discussed, ignoring the effect of edge conditions, which now represent only an inessential complication.

In Sec. 3 some topological aspects of this relationship will be considered. Topological aspects of the Ising model have already been extensively discussed[2-4], but I am not aware of any work in this field which directly relates the Onsager and Pfaffian methods.

2. ALGEBRAIC RELATIONSHIP BETWEEN THE ONSAGER AND THE PFAFFIAN METHODS

An Ising model on a general lattice consists of a set of N identical systems, each capable of existing in two states only, and arranged to form a regular lattice in one, two or three dimensions. These systems may interact with each other, and the energy of interaction of any pair of such systems is assumed to depend only on the states of these two systems and on nothing else. This means that we restrict ourselves to the case of two body forces only. The assumption that all the systems are identical could be relaxed, and then the lattice would be regarded as built up by the repetition of systems, each of which is itself composed of several nonidentical subsystems. Such a model has been considered, using the Pfaffian method, in an earlier paper.[5] However, in order to maintain simplicity, and because nothing essential for the purposes of this paper would be introduced by this generalization, the model described at the beginning of this paragraph will be used.

So the state of a general lattice can be specified by an N-tuple $(s_1, \cdots, s_N)$ with each s_i taking the values ± 1 only. There are thus 2^N distinct states

[1] C. A. Hurst, J. Math. Phys. (N. Y.) 5, 11 (1964), hereafter referred to as I.
[2] C. A. Hurst and H. S. Green, J. Chem. Phys. 23, 1059 (1960); P. W. Kasteleyn, J. Math. Phys. (N. Y.) 4, 287 (1963).
[3] H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964); hereafter referred to as G.
[4] S. Sherman, J. Math. Phys. (N. Y.) 1, 202 (1960); P. N. Burgoyne, 4, *ibid.* 1320 (1963).
[5] C. A. Hurst, J. Chem. Phys. 38, 2558 (1963).

of the lattice. The total energy of the lattice will be denoted by $E(s_1, \cdots, s_N)$ and the assumption of two body forces means that

$$E(s_1, \cdots, s_N) = -\sum_{i \neq j} J_{ij} s_i s_j, \tag{1}$$

apart from additional constants which do not depend on s. The expression (1) implies that the level of energy has been chosen so that the ground state has an energy $-\Sigma_{i \neq j}|J_{ij}|$ rather than 0.

All the information required to specify the type of general lattice being studied is contained in expression (1), so that the fact that we are dealing with, say, a two-dimensional square lattice with nearest-neighbour interactions, or a three-dimensional cubic lattice with long-range interactions can be obtained from the structure of the set of constants J_{ij}. This set of constants therefore expresses the topological structure of the lattice and its bond connections. Instead of there being $\frac{1}{2}N(N-1)$ distinct such constants many of them will be equal. For example the nearest-neighbour square lattice can be specified by requiring that $J_{ij} = 0$ unless $j = i \pm 1$ or $i \pm n$, $N = n^2$, and then $J_{ij} = J$. Only a single constant is required in this case. The first restriction that is usually imposed is that of translational invariance. This requires the energy of interaction of two systems to depend only on their relative positions. This means that J_{ij} depends only on the difference $i - j$ of the indices i and j and not on their separate values. The number of independent constants is thereby reduced to $2N$. The next restriction is that the lattice is symmetric under inversions so that $J_{ij} = J_{ji}$, so that these constants depend now only on $|i - j|$, the magnitude of the difference. The number of constants is thus reduced to N. The final restriction is that the longest bond present is of length n with $n \ll N$, so that $J_{ij} = 0$ unless $|i - j| \leq n$. There are then at most n independent constants J_{ij} which we now denote by $J_1, \cdots, J_n$. This restriction is equivalent to saying that the lattice is periodic with period n, and means that the lattice can be regarded as made up of a series of generalized "layers" each of length n. Without this restriction it would not be possible to regard the system as forming a lattice whatever the arrangement of the individual subsystems. Although $n \ll N$, it does not mean that n must be finite, as the example of the cubic lattice shows. However, n must be asymptotically small compared to N, for N large. Of course these restrictions may need to be modified to take edge conditions into account, but, following the remarks in the introduction, it is considered that the presence of edge effects can be neglected for large N. This is in the usual tradition of discussions on these problems, and has the consequence here that the limits on various summations that occur would be different with different edge conditions, but that these differences will be ignored.

The partition function will then be given by

$$Z = \sum_{s_1 = \pm 1} \cdots \sum_{s_N = \pm 1} \exp(-E/kT), \tag{2}$$

or, alternatively,

$$Z = \mathrm{tr}_{s_1} \cdots \mathrm{tr}_{s_N} \exp\left(\sum_{j=1}^{N} \sum_{k=1}^{n} J_k s_{j+k} s_j / kT\right), \tag{3}$$

where now the quantities s_i are regarded as commuting matrices in a 2^N-dimensional product space and tr_{s_i} means that the trace is taken over the matrix s_i wherever it occurs.

The Oguchi transformation can be employed to reduce the problem of evaluating (3) to the combinatorial problem of counting closed polygons drawn on this lattice. An alternative expression for this combinatorial problem is used as the starting point for the Pfaffian method. Instead of considering a set of N dichotomic variables $s_1, \cdots, s_N$, regarded as commuting matrices in a direct product space, we consider instead a set of nN anticommuting matrices, with a typical member $\sigma_j^{(k)}$ corresponding to a bond joining lattice point j to lattice point $j + k$.

Then we define

$$Z' = \mathrm{Tr} \prod_{j=1}^{N} \left[1 + \sum_{k>l} \left(\sigma_{j-k}^{(k)} \sigma_{j-l}^{(l)} + x_k x_l \sigma_j^{(k)} \sigma_j^{(l)}\right) + \sum_{k,l} x_k \sigma_j^{(k)} \sigma_{j-l}^{(l)} + \cdots\right], \tag{4}$$

where $x_k = \tanh(J_k/kT)$, and the notation Tr means that the trace is taken over all matrices appearing in the expansion of the product. The dots indicate terms involving fourth and higher even-order terms whose precise structure can be inferred from the later Eq. (6). Z' has been constructed with the object of defining the same combinatorial problem as does Z. The matrices $\sigma_j^{(k)}$ satisfy the anticommutation rules

$$\sigma_j^{(k)} \sigma_{j'}^{(k')} + \sigma_{j'}^{(k')} \sigma_j^{(k)} \equiv [\sigma_j^{(k)} \sigma_{j'}^{(k')}]_+ = 2\delta_{jj'} \delta_{kk'}. \tag{5}$$

It is well known[2] that Z' contains only those terms which contain even powers of each of the matrices $\sigma_j^{(k)}$, and hence, can be regarded as the generating function of closed polygons drawn on the lattice using the bonds as sides and the lattice points as

vertices, so long as all the terms appear with an explicit positive sign. This requirement may be incompatible with the relations (5), and so it is with this question that this paper is concerned. Z', if it can be evaluated, will then be proportional to Z because of the identity of the combinatorial problems which they pose. The use of Pfaffian methods to evaluate Z' has been discussed elsewhere[3], and will not be described here, except to remark that the introduction of anticommutation relations is essential and they cannot be replaced by commutation relations.

We now show how (3) and (4) can be reduced to expressions which have a very similar form so that any differences can be readily interpreted. First of all we can write (4) as

$$Z' = \mathrm{Tr} \prod_{j=1}^{N} \left[\prod_{k=2}^{n} (1 + \sigma_{j-k}^{(k)} \sigma_{j-1}^{(1)}) \times \prod_{l=1}^{n} (1 + x_l \sigma_j^{(l)} \sigma_{j-1}^{(1)}) \right], \quad (6)$$

for if the products over k and l, for fixed j, are multiplied out, and the property $\sigma^2 = 1$ used, then Eq. (4) is regained. The order of terms in the products over k and l must be opposite to that in which they occur in Eq. (4). This requirement follows from the relations (5). The structure of Eq. (4), and hence of Eq. (6), is such that pairs of bonds are not correlated so that any bond can appear with any other bond. This agrees with the assumptions made about the energy $E(s_1, \cdots, s_N)$ in Eq. (1). It is necessary for definiteness to choose a standard ordering of the anticommuting quantities, although, as will be shown later, the question of the equivalence of the two methods will be independent of this ordering. The order chosen is such that in Eq. (6) the product is written from right to left in order of decreasing l and then in order of decreasing k for fixed j, and finally from right to left for increasing j.

The factors $(1 + \sigma_{j-k}^{(k)} \sigma_{j-1}^{(1)})$ are equivalent to δ-symbols, being of 0 if $\sigma_{j-k}^{(k)} \neq \sigma_{j-1}^{(1)}$, and having the value 2 if they are identical. The argument is the same as that used in I. This means that $\sigma_j^{(k)}$ can be replaced by $\sigma_{j+k-1}^{(1)}$ in Eq. (6), and we denote this matrix by σ_{j+k}. In this reduction the order adopted above is essential. Now Eq. (6) can be replaced by

$$Z' = \mathrm{Tr} \prod_{j=1}^{N} \left[\prod_{k=1}^{n} (1 + x_k \sigma_{j+k} \sigma_j) \right] \quad (7)$$

with the product in standard order, and constant multiplying factors being dropped. The Oguchi transformation applied to Eq. (3) produces exactly the same expression for Z except that now all the matrices are the commuting quantities s_j.

The next step in the Onsager method is to replace Eq. (7) by an expression involving only $2^n \times 2^n$ matrices rather than those of dimension $2^{nN} \times 2^{nN}$. This step is effected by introducing the partial density matrices of Schultz, Mattis, and Lieb[6], following the approach used in I. We define

$$P_M(\sigma_{M+1}, \cdots, \sigma_N) = \mathrm{tr}_{\sigma_1} \cdots \mathrm{tr}_{\sigma_M} \prod_{j=1}^{M} \left[\prod_{k=1}^{n} (1 + x_k \sigma_{j+k} \sigma_k) \right] \quad (8)$$

$$P_0 = 1,$$

so that

$$P_{M+1}(\sigma_{M+2}, \cdots, \sigma_N) = \mathrm{tr}_{\sigma_{M+1}} \prod_{k=1}^{n} (1 + x_k \sigma_{M+k+1} \sigma_{M+1}) P_M. \quad (9)$$

By the same argument as used in I, only n σ-matrices can appear in any P_M so that it is sufficient to use a $2^n \times 2^n$ representation. This means that if $M = rn + m$ with $1 \leq m \leq n$ and $0 \leq r < p = [N/n]$, we may denote the matrices $\sigma_{M+1}, \cdots, \sigma_{M+n+1}$ by $\sigma'_{m+1}, \cdots, \sigma'_n, \sigma_1, \cdots, \sigma_{m-1}, \sigma_m, \sigma_{m+1}$ and adopt the representation

$$\sigma_j = \tau \times \tau \times \cdots \times \tau \times \sigma \times \delta \times \cdots \times \delta \quad \text{(n terms)} \quad (10)$$

with

$$\sigma = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \tau = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \delta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and σ appears in the jth place.

We may write P_{M+1} as

$$P_{M+1} = \mathrm{tr}_{\sigma'_{m+1}} \prod_{k=m+2}^{n} (1 + x_{k-m-1} \sigma'_k \sigma'_{m+1}) \times \prod_{l=1}^{m+1} (1 + x_{n+l-m-1} \sigma_l \sigma'_{m+1})(A_M + B_M \sigma'_{m+1}), \quad (11)$$

with the products written from right to left in order of decreasing index. The expressions A_M and B_M contains products of the matrices $\sigma'_{m+2}, \cdots, \sigma_m$ of even and odd order, respectively. On taking the trace over σ'_{m+1} and introducing the number operator $\mathbf{N}_{m+1}$ which counts the number of times

[6] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).

σ_{m+1} appears in P_M, P_{M+1} can be written as

$$P_{M+1}|0\rangle = [(1 - \mathsf{N}_m) + x_n\mathsf{N}_m] \times \prod_{k=1}^{m+1}(1 + ix_{n+k-m-1}\sigma_k\rho_{m+1}) \times \prod_{l=m+2}^{n}(1 + ix_{l-m-1}\sigma_l'\rho_{m+1})P_M|0\rangle, \quad (12)$$

where

$$\rho_j = \tau \times \tau \times \cdots \times \tau \times \rho \times \delta \times \cdots \times \delta,$$

with

$$\rho = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

in the jth place, and $|0\rangle$ is the vacuum state corresponding to all the spins being directed upwards, and satisfying

$$\mathsf{N}_j|0\rangle = 0$$

with

$$\mathsf{N}_j = \tfrac{1}{2}(1 + i\sigma_j\rho_j) = \tfrac{1}{2}(1 - \tau_j). \quad (13)$$

In Eq. (12) the factors are now written in the same order in which they appear in Eq. (4). Finally using the same arguments as in I, we have

$$Z' = \langle 0|\, V^p\, |0\rangle, \quad (14)$$

where

$$V = \prod_{m=1}^{n}\left[e^{-iK_n^*\sigma_m\rho_m}\left(\prod_{k=n-m+1}^{n-1} e^{iK_k\sigma_{k-n+m}\rho_m} \times \prod_{l=1}^{n-m} e^{iK_l\sigma_{m+l}\rho_m}\right)\right], \quad (15)$$

where $K_l = J_l/kT$ and $\tanh K^* = e^{-2k}$. (We ignore edge conditions and so suppose that p is an integer.) The corresponding expression obtained using the Onsager approach on (3) differs from Eq. (15) in that the product $i\sigma_l\rho_m$ is replaced by s_ls_m. A suitable definition of s_l is

$$s_l = \delta \times \delta \times \cdots \times \delta \times \sigma \times \delta \times \cdots \times \delta, \quad (16)$$

so that

$$\begin{aligned} s_ls_m &= i\sigma_l\tau_{l-1}\cdots\tau_{m+1}\rho_m, \quad \text{for } l > m, \\ &= i\sigma_m\tau_{m-1}\cdots\tau_{l+1}\rho_l, \quad \text{for } l < m. \end{aligned} \quad (17)$$

This means that in the Onsager method the exponents are of higher order than the second if long-range bonds are present, i.e., if $|m - l| > 1$. Such a case is called nonlinear, if, in the spirit of I we call a quadratic exponent a "linear" term. However, the presence of long-range bonds is not a sufficient condition for nonlinearity because the expression (14) can be equal to (3) even when nonlinear exponents occur. This is because the τ-matrices have eigenvalues ± 1 only, so that if Eq. (15) is expanded in powers of x_l the presence of additional τ-factors can only mean that some terms appear with a changed sign. But the difficulty with the Pfaffian method is just that the sign of certain terms is incorrect, and this suggests that the function of the τ-matrices is to compensate these incorrect signs. In the next section, the conditions under which no τ-factors are required, or equivalently, when they cancel each other, will be derived.

3. TOPOLOGICAL BASIS FOR THE EQUIVALENCE OF THE TWO METHODS

In order to compare the two expressions, we will consider a general point of the lattice whose label is $rn + m$ and consider the contribution to Eq. (14) which comes from all factors between $rn + m$ and $(r + 1)n + m$. These factors are

$$\prod_{s=1}^{m}\left\{[(1 - \mathsf{N}_s) + x_n\mathsf{N}_s]\prod_{k=1}^{s}(1 + ix_{n+k-s}\sigma_k\rho_s) \times \prod_{l=s+1}^{n}(1 + ix_{l-s}\sigma_l\rho_s)\right\} \times \prod_{s'=m+1}^{n}\left\{[(1 - \mathsf{N}_{s'}) + x_n\mathsf{N}_{s'}] \times \prod_{k'=1}^{s'}(1 + ix_{n+k'-s'}\sigma_{k'}\rho_{s'}) \times \prod_{l'=s'+1}^{n}(1 + ix_{l'-s'}\sigma_{l'}\rho_{s'})\right\}. \quad (18)$$

When Eq. (14) is expanded in powers of the x's and the vacuum expectation value taken, the only terms which survive are those which contain products of σ_m^2, ρ_m^2, and $i\rho_m\sigma_m$ for each m, and such products can then be replaced by 1. Therefore, the evaluation of Eq. (14) is completed when all the matrices have been paired in this way. Because of the anticommutation relations (5), a change of sign will occur whenever two matrices are interchanged. With each matrix factor σ_m coming from a lattice point $rn + s$, say, one may associate a bond emanating from this point, and because no bonds can be longer than n, it must terminate at the lattice point $(r + 1)n + m)$ if $s > m$, or $rn + m$ if $s < m$. Hence, for each σ_m appearing in the expansion of (18) there will be a corresponding $i\rho_m$ in the factor $s = m$ of (18). This matrix $i\rho_m$ appears either with a matrix σ_l in a term

$$(1 + ix\sigma_l\rho_m)$$

in which case it represents a bond joining $(r+1)n+m$ to $(r+1)n+1$, or it appears with the matrix σ_m in the term

$$(1 - \mathsf{N}_m) + x_n N_m,$$

when it represents a bond joining $(r+1)n+m$ to $(r+2)n+m$.

So if we write Eq. (18) in the form

$$[(1 - \mathsf{N}_m) + x_n \mathsf{N}_m](e + io\rho_m)(e' + \sigma_m o'), \qquad (19)$$

where e, e' represent terms with an even number of matrix factors, and o, o' terms with an odd number of matrix factors, the various terms can be interpreted graphically. The term e' can be regarded as corresponding to an even number of bonds terminating at $(r+1)n+m$ together with an even number of bonds which terminate elsewhere, and o' as an odd number terminating at $(r+1)n+m$ together with an odd number terminating elsewhere. As this whole expression must act on the vacuum state the only terms which survive in (19) are those contained in the expression

$$(ee' + oo') + x_n\sigma_m(oe' + eo'),$$

and then these four terms are also readily interpreted: (1) ee' represents an even number of bonds going into $(r+1)n+m$ and an even number going out; (2) oo' an odd number going in and out; (3) $x_n\sigma_m oe'$ an even number going in and an odd number going out, including one from $(r+1)n+m$ to $(r+2)n+m$, and $x_n\sigma_m eo'$ is similar.

So we can see how the expansion reproduces the bond connections, as of course it must. This classification of the bond connections given by the expansion of (14) can be represented as a linear graph of the type already used in G and which is depicted in Fig. 1. On a straight line we place N clusters of points with each cluster containing an even number of points, at most $2n$ in number. From each point in the cluster emanates a bond to terminate at a point of another cluster, and the points in the cluster are ordered in the same way as the factors in (4). This means that a bond connecting $rn+m$ with $(r+1)n+m$ must start from the left-hand point of cluster $rn+m$, and so on. The virtue of this diagram is that the number of points of intersection of these lines gives the number of sign changes resulting from interchanges of the anticommuting quantities σ and ρ. The proof of this fact is not difficult and has already been given in G. Now, as it is evident that the expansion of the Onsager form will produce no changes of sign because only commuting quantities appear,

FIG. 1. A linear graph.

the two methods are equivalent if the total parity of this graph is even, i.e., if the number of crossing points is even. The number of crossing points can be altered by interchanging points within a cluster, and such interchanges are equivalent to interchanges of factors in Eq. (6). The consequent change in the parity of the number of crossing points corresponds exactly to the changes of sign arising from an interchange of anticommuting quantities. The original graph is a nonplanar graph if the number of crossing points is nonzero, but it may be possible by suitable interchanges to remove all the crossing points so that the graph is now planar. In that case the new expression Z'' obtained from Z' by this reordering will no longer contain any explicit negative signs, and so will be identical with Z, i.e., the Pfaffian method is applicable if Z'' is used instead of Z'. Such interchanges of points within a cluster can be regarded as a topological distortion of the original linear graph. If we coalesce together all the points of the cluster belonging to a single lattice point, we can order the bonds according to the angle, reckoned in a clockwise direction, that they leave the lattice point, and the parity of the graph is the same as that of the original linear graph. Any deformation of a bond which does not change the order of its end points will not change the parity of the graph, and any deformation whatsoever can be built up from deformations of the bonds and interchanges of order within a cluster. Because the connections of this linear graph are the same as that of the original lattice, this graph is clearly a topological deformation of the original lattice graph. So, if after suitable changes of points within a cluster no further changes in crossing number are introduced on transforming back to the original lattice, it is possible to argue directly from the nature of the bonds on the lattice to the applicability of the Pfaffian method. For if the graphs on the lattice are all planar graphs, there exists an ordering of points within each cluster so that the Pfaffian method works. An explicit determination of the appropriate order and a direct proof of the absence of negative signs was given in G.

As every graph, whether planar or nonplanar, can be represented as a graph in three dimensions without crossing points, it is always possible to remove crossing points by a three-dimensional deformation, and conversely, three-dimensional graphs

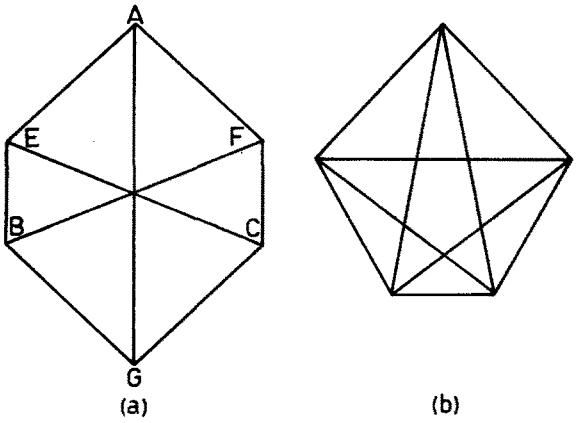

FIG. 2. Basic nonplanar graphs.

on a lattice correspond when transformed back to a linear graph to graphs with crossing points. Thus it is linear graphs which are such that they cannot be transformed to planar graphs which correspond to Ising models for which the Onsager method becomes nonlinear and for which the Pfaffian method introduces irremovable negative signs. This result goes beyond the statements made in G and by Dychne and Rumer[7] where it was stated that a uniform reordering of points in a cluster (i.e. the same reordering for every lattice point) would not help the Pfaffian method. Here it is demonstrated that even a *nonuniform* change will not do.

The necessary and sufficient condition for a graph to be nonplanar is simply expressed by a theorem of Kuratowski[8] which states that the presence, as a subgraph, of either of the two graphs depicted in Fig. 2 (or of a subgraph which is homeomorphic to either of them) is all that is required. It is evident, as shown in Fig. 3, that both the next-nearest-neighbor problem for a plane rectangular lattice

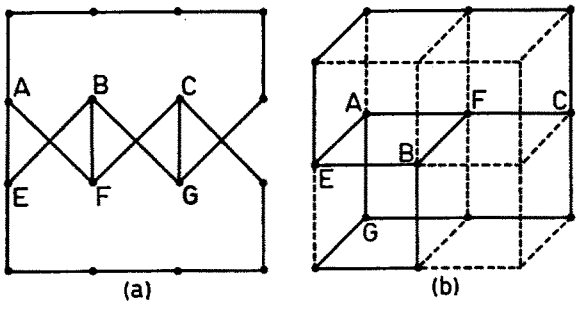

FIG. 3. Subgraph of type Fig. 2(a) for (a) next-nearest-neighbor problem, (b) cubic lattice.

and the nearest neighbour problem for a cubic lattice contain a subgraph homeomorphic to subgraph 2(a), and therefore, cannot be solved. However, this not quite what is required for the nonequivalence of the two methods. It is not sufficient that the graph be nonplanar but rather that there always exists a subgraph with an odd number of crossing points no matter how the original graph is distorted. A short proof that the two graphs of Fig. 2 always contains at least one closed polygonal subgraph with an odd number of crossing points is given in the Appendix. This means that although, for example, Fig. 2(a) cannot be an Ising model graph because its vertices have an odd number of lines incident upon them, it always contains proper Ising model graphs as subgraphs for which the sign will be incorrect.

It is interesting to see how the Onsager method enables the incorrect signs to be compensated exactly. As mentioned earlier, the factor $s_l s_n$ differs from $i\sigma_l\rho_m$ by the addition of τ-factors, and for $l > m$ there is one for each lattice point between $rn + m$ and $rn + l$. Now because

$$\tau_j\,|0\rangle = |0\rangle, \qquad \tau_j\sigma_j\,|0\rangle = -\sigma_j\,|0\rangle, \tag{20}$$

a τ-matrix in a factor associated with a particular lattice point counts the number of bonds of the jth type emanating from lattice points with lower index than the given lattice point. As every bond with index lying between m and l must terminate between m and l it must therefore cross the bond joining n and l. Hence the change in sign which this entails in the Pfaffian method is just compensated by the additional τ-factor. For $l < m$ the argument is slightly different. We can write

$$\begin{aligned} s_l s_m &= -i\rho_l\tau_{l+1}\cdots\tau_{m-1}\sigma_m \\ &= -i\rho_l\tau_{l+1}\cdots\tau_{m-1}\sigma_m U, \end{aligned} \tag{21}$$

where

$$U = \tau_1\tau_2\cdots\tau_n.$$

The addition of the factor U to this expression makes no difference because the U-matrix counts the *total* number of bonds emanating to the right of the factor $s_l s_m$ and this must be even.

But

$$\begin{aligned} &-i\rho_l\tau_{l+1}\cdots\tau_{m-1}\sigma_m U \\ &\quad = -i\sigma_l\tau_{l-1}\cdots\tau_1\tau_n\cdots\tau_{m+1}\sigma_m, \end{aligned} \tag{22}$$

so that once again the Onsager expression counts the number of bonds terminating between l and m, although in this case the bond joining m to l is

[7] A. M. Dychne and J. B. Rumer, Forschr. Physik 9, 509 (1961).

[8] C. Berge, *The Theory of Graphs and Its Applications* (Methuen and Company Ltd., London, 1962), Chap. 21.

weighted with a negative sign. For $l = m$, a similar situation arises, for we can write

$$(1 - \mathsf{N}_m) + x_n\mathsf{N}_m$$
$$= (1 - \mathsf{N}_m) + x_n\mathsf{N}_m U$$
$$= (1 - \mathsf{N}_m) - x_n\mathsf{N}_m\tau_1 \cdots \tau_{m-1}\tau_{m+1} \cdots \tau_n, \quad (23)$$

because

$$\tau_m\mathsf{N}_m = -\mathsf{N}_m.$$

Equation (23) means that a factor counting the number of bonds terminating between $rn + m$ and $(r + 1)n + m$ can be introduced to compensate for the crossing points, but the bond again has negative weight. But the combined effect of (22) and (23) is to compensate each other because all such bonds cross the lattice point with label $(r + 1)n$, and there must be an *even number of such bonds*. Hence, the minus signs cancel.

4. CONCLUSION

So we see how the nonlinearities introduced serve to correct the errors arising in the Pfaffian method from the topological structure of the model. These nonlinearities are sufficient but not always necessary. In G it was shown how a particular choice of ordering within a cluster can always be found such that any planar graph (even if it contains long-range bonds) can be solved by the Pfaffian method. Such problems are therefore *nonessentially nonlinear*, whereas those corresponding to nonplanar graphs are *essentially nonlinear* and require a more sophisticated approach.

As Kasteleyn[2] mentioned, for graphs of genus $g > 0$, the Pfaffian method can be employed, but 4^g Pfaffians are required. The insoluble problems are those with $g = \infty$. So from this point of view an essentially nonlinear problem may still be soluble in terms of a sum of linear problems, but when this sum is infinite the result is of little use. It would be interesting to investigate whether a further classification of nonplanar graphs for $g = \infty$ can be made which also characterizes the algebraic structure. Intuitively the next-nearest-neighbor problem on a rectangular lattice should be much simpler than, for example, the cubic lattice but the difference has not yet been explicitly expressed.

It appears from this discussion that the question of whether graphs form knots in three dimensions is not relevant, for the condition of bonds crossing above or below each other does not enter very naturally. This speculation does not accord with remarks made by other authors.[4]

ACKNOWLEDGMENTS

This work, which was commenced while I was a visitor at the Department of Mathematics, University of Toronto, was concluded while I was taking part in the 1965 Summer Research Institute of the Australian Mathematical Society, held in the Mathematics Department of the Australian National University. I would like to thank members of both departments for their hospitality.

APPENDIX

In the graphs of Figs. 2(a) and 2(b), there are, respectively, six and twelve closed polygonal subgraphs, each of which includes every vertex. The problem is to show that no matter how the original graph is distorted in the plane, at least one of these polygonal graphs from each set has an *odd* number of crossing points. This is not completely trivial because the parent graph may have an even total number of crossing points.

If we have a closed polygonal graph, and continuously deform one line, keeping its end points fixed, the parity of the total number of crossing points cannot change unless a crossing point passes through one of these end points. This means that the two lines meeting at a vertex must change from noncrossing to crossing, or vice versa. This follows from a theorem of Whitney's.[9]

In the case of Fig. 2(a), there are five polygonal graphs of odd parity and one of even parity and every pair of lines meeting at a vertex appears in two polygonal graphs. If any such pair is made to intersect itself, the parity of the graphs containing these lines is changed. Consequently, the number of graphs of odd parity can only change by an even number. Hence, the number of graphs of odd parity can never be made zero.

In the case of Fig. 2(b), every pair of lines meeting at a vertex is also contained in two polygonal graphs, but this time the two graphs are of opposite parity. Hence, as the parity of each one is changed by making these lines cross, there will always be at least one graph of odd parity.

[9] H. Whitney, Compositio Mathematica 4, 276 (1937).

Asymptotic Behavior of Phase Shift in λ for Strongly Singular Potentials

B. JAKŠIĆ*† AND N. LIMIĆ*
International Centre for Theoretical Physics, Trieste, Italy
(Received 19 February 1965)

The asymptotic form of the phase shift is derived for strongly singular potentials for large complex λ, $|\arg \lambda| < \frac{1}{2}\pi$, and real k. This result is valid also for usual (regular) potentials. It is shown also that the S-matrix for strongly singular potentials must have an infinite number of poles in λ-plane accumulating asymptotically in a narrow region along the imaginary axis in the first and third quadrants.

1. INTRODUCTION

THE theory of scattering on strongly singular potentials has been investigated in some generality in references.[1,2] The main results of these works solved the problem of analyticity of the S-matrix.

In this paper we intend to investigate the exact asymptotic behavior of the phase shift $\delta(\lambda, k)$ for large complex λ, arg λ fixed, and positive k for strongly singular potentials. By a strongly singular potential we understand a potential repulsive at the origin and more singular there than the centrifugal term. The method used in solving this problem is in some respect, i.e., in the problem of uniform convergence of the Jost solutions for large λ as shown in Sec. 2, close to the application of the WKB. method to the similar problem for usual potentials.[3] Of course, the problem is here more involved because of the strong singularity of the potential. Unfortunately, as is evident already in Ref. 3, it seems impossible to obtain the asymptotic form of $\delta(\lambda, k)$ using only these results. In order to solve this problem, we shall introduce a formula for the phase shift in terms of the Jost solutions, which will enable us to find out the asymptotic form of $\delta(\lambda, k)$ from the asymptotic form of the Jost solutions. The final result is given by (3.8). Of course, our method is valid also for usual potentials. In this case (3.8) is just the asymptotic form of the first Born approximation.

The only solutions of the Schrödinger equation which we shall be using are the Jost solutions $f(\lambda, \pm k, z) \equiv f_{\pm}(z)$ defined by

$$\frac{d^2}{dz^2} f_{\pm}(z) = \left[V(z) + \frac{\lambda^2 - \frac{1}{4}}{z^2} - k^2 \right] f_{\pm}(z), \qquad f_{\pm}(z) \sim \exp(\mp ikz). \tag{1.1}$$

For $V(z) \equiv 0$, the corresponding Jost solutions will be denoted by $f_0(\lambda, \pm k, z) \equiv f_{0\pm}(z)$. In (1.1) k is positive and $\lambda = |\lambda| \exp(i\omega)$, $0 \leq \omega < \frac{1}{2}\pi$. The potential $V(z)$ is assumed to be a regular analytic function in the half-plane Re $z > 0$, real on the positive real axis, satisfying at the origin the conditions

$$V(x) > 0, \qquad 0 \leq x \leq x_0, \tag{1.2}$$

$$\int_0^{x_0} \frac{1}{[V(x)]^{\frac{1}{2}}} \frac{1}{x^2}\, dx, \quad \int_0^{x_0} \frac{1}{[V(x)]^{\frac{1}{2}}} \left(\frac{V'(x)}{V(x)}\right)^2 dx,$$

$$\int_0^{x_0} \frac{1}{[V(x)]^{\frac{1}{2}}} \frac{|V''(x)|}{V(x)}\, dx \quad \text{converge}, \tag{1.3}$$

$$x^2 V(x) \quad \text{is monotonic for} \quad 0 \leq x \leq x_0. \tag{1.4}$$

A few simple examples of such potentials are potentials increasing at the origin exponentially, like $x^{-\alpha}$, $\alpha > 2$, or like $|\ln x|^{\alpha}/x^2$, $\alpha > 2$. At infinity in the half-plane Re $z > 0$, we shall assume the asymptotic behavior

$$V(z) = V_{as}(z)[1 + u(z)], \qquad u(z) \sim u_0/z, \qquad \text{Re}\, z \to \infty, \tag{1.5}$$

where

$$V_{as}(z) = V_0(e^{-\mu z}/z^{\alpha+1}), \qquad V_0 \text{ and } \alpha \text{ real}. \tag{1.6}$$

In order to derive a suitable formula for calculating the asymptotic form of the phase shift we start from the following expression[2]:

$$S(\lambda, k) = \exp\{i\pi(\lambda - \tfrac{1}{2})\} \frac{f_+(z)}{f_-(z)} \exp\left\{2ik \int_{C(0,\varepsilon)} \frac{d\xi}{f_+(\xi) f_-(\xi)}\right\}.$$

---

* On leave of absence from the University of Zagreb and the Institute "Rudjer Bošković," Zagreb.
† Present address: Institute of Physics, Torino, Italy.
[1] E. Predazzi and T. Regge, Nuovo Cimento **24**, 518 (1962).
[2] N. Limić, Nuovo Cimento **26**, 581 (1962).
[3] A. Bottino, A. Longoni, and T. Regge, Nuovo Cimento **23**, 954 (1962).

Here z is an arbitrary point in the half-plane $\mathrm{Re}\, z > 0$, and the path of integration from the origin to the point z, denoted by $C(0, z)$, is also arbitrary, provided that the integral exists. The same expression is valid for $V(z) \equiv 0$, with $S(\lambda, k) = 1$ and the free Jost solutions $f_{0\pm}(z)$. By dividing these two expressions we obtain the formula

$$S(\lambda, k) = \frac{f_+(z)/f_{0+}(z)}{f_-(z)/f_{0-}(z)} \times \exp\left\{2ik \int_{C(0,z)} \left[\frac{1}{f_+(\xi)f_-(\xi)} - \frac{1}{f_{0+}(\xi)f_{0-}(\xi)}\right] d\xi\right\}.$$

The phase shift is then given by

$$\delta(\lambda, k) = \frac{1}{2i} \ln \frac{f_+(z)/f_{0+}(z)}{f_-(z)/f_{0-}(z)} + k \int_{C(0,z)} \left[\frac{1}{f_+(\xi)f_-(\xi)} - \frac{1}{f_{0+}(\xi)f_{0-}(\xi)}\right] d\xi. \quad (1.7)$$

This is the formula which we shall be using in Sec. 3.

2. ASYMPTOTIC BEHAVIOR OF THE JOST SOLUTIONS

The first step towards the proof of the asymptotic behavior of the Jost solutions (1.1) is to choose an auxiliary differential equation as the asymptotic substitute for the Schrödinger equation. It should contain those terms of the Schrödinger equation which are dominant for $\lambda \to \infty$ and for all z. Both solutions of this auxiliary equation must be known to us in an explicit form and be simple enough to allow simple estimates. The solutions of such equation will be used for constructing the Green's function and the corresponding integral equation for the Jost solutions. We expect that the Jost solutions will approach asymptotically a solution of the auxiliary equation, if it has been chosen properly.

The general type of the differential equation

$$\frac{d^2}{dz^2} u(z) = \left[Q^2(z) + \frac{3}{4}\left(\frac{Q'(z)}{Q(z)}\right)^2 - \frac{1}{2}\frac{Q''(z)}{Q(z)}\right] u(z) \quad (2.1)$$

has two linearly independent solutions

$$u(z) = \frac{1}{[Q(z)]^{\frac{1}{2}}} \exp\left\{\mp \int Q(\xi)\, d\xi\right\}. \quad (2.2)$$

If we choose $Q^2(z)$ so that it incorporates the dominant terms of (1.1) and find out that the additional two terms in (2.1) are insignificant compared to $Q^2(z)$, then all the requirements are satisfied, and we can expect with confidence that the essentials of the asymptotic behavior of the solutions (1.1) will be reproduced by the singular solution of (2.1).

Comparing (2.1) with (1.1) we see that it is natural to choose $Q^2(z) = V(z) + \lambda^2/z^2 - k^2$. For considerations in the region $\mathrm{Re}\, z > x_0$, where the contribution of the potential is insignificant, it is simpler to take $Q^2(z) = \lambda^2/z^2 - k^2$. We shall be using both types of auxiliary differential equations, the first choice being used near the origin and the second elsewhere. Let us discuss first the solutions for the simpler second choice. They are

$$\chi_\pm(z) = \frac{1}{\binom{+}{i}[1 - (\lambda^2/k^2z^2)]^{\frac{1}{2}}} \times \exp\left\{\mp ikz + \int_{C_\pm(z,\infty)} \left[\left(\frac{\lambda^2}{\xi^2} - k^2\right)^{\frac{1}{2}} \mp ik\right] d\xi\right\},$$

$$\varphi_\pm(z) = \frac{1}{\binom{+}{i}[1 - (\lambda^2/k^2z^2)]^{\frac{1}{2}}} \times \exp\left\{\pm ikz - \int_{C_\pm(z,\infty)} \left[\left(\frac{\lambda^2}{\xi^2} - k^2\right)^{\frac{1}{2}} \mp ik\right] d\xi\right\}. \quad (2.3)$$

The cut for the double-valued square root is taken as shown in Fig. 1.

The sheet chosen is the one on which $\mathrm{Re}\,[(\lambda^2/z^2) - k^2]^{\frac{1}{2}} > 0$ on the positive real axis. Then the square root tends to $\pm ik$ as z tends to infinity along $C_\pm$, respectively. For $(1 - \lambda^2/k^2z^2)^{\frac{1}{2}}$ we chose the same cut and the sheet on which it tends to $+1$ and $(-i)$ as z tends to infinity along $C_\pm$, respectively. With these determinations of the roots we find that at the origin $\chi_\pm(z) \approx Cz^{-\lambda+\frac{1}{2}}$ and $\varphi_\pm(z) \approx Cz^{\lambda+\frac{1}{2}}$, while at infinity along $C_\pm$ we have the asymptotic behavior $\chi_\pm(z) \sim \exp(\mp ikz)$ and $\varphi_\pm(z) \sim \exp(\pm ikz)$. These functions do not vanish in $\mathrm{Re}\, z > 0$.

With the known construction of the Green's func-

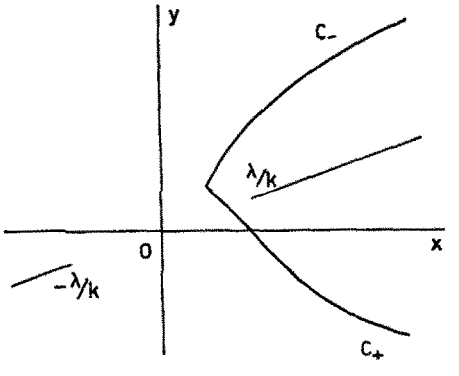

FIG. 1.

tion, we easily find the integral equation for the Jost solutions valid outside the origin

$$f_\pm(z) = \chi_\pm(z) \pm \frac{1}{2ik}\int_{C\pm(z,\infty)} [\chi_\pm(z)\varphi_\pm(z') - \chi_\pm(z')\varphi_\pm(z)][V(z') + W(z')]f_\pm(z')\,dz', \quad (2.4)$$

where

$$W(z) = -\frac{\lambda^2 k^2 + \frac{1}{4}k^4 z^2}{(\lambda^2 - k^2 z^2)^2}. \quad (2.5)$$

Introducing now the new function

$$g_\pm(z) = f_\pm(z)/\chi_\pm(z), \quad (2.6)$$

we obtain the integral equation

$$g_\pm(z) = 1 + \int_{C\pm(z,\infty)} G_\pm(z, z') \times [V(z') + W(z')]g_\pm(z')\,dz', \quad (2.7)$$

where

$$G_\pm(z, z') = \frac{1}{2[(\lambda^2/z'^2) - k^2]^{\frac{1}{4}}} \times \left(1 - \exp\left\{-2\int_{C\pm(z,z')} \left(\frac{\lambda^2}{\xi^2} - k^2\right)^{\frac{1}{2}} d\xi\right\}\right). \quad (2.8)$$

The Eq. (2.7) is the basic one for the study of the asymptotic behavior of the Jost solutions in the region outside the origin. Specially, for $V(z) \equiv 0$, this equation is valid also at the origin, and its solutions will be denoted by $g_{0\pm}(z)$.

In the neighbourhood of the origin we have to take the first choice of $Q^2(z)$, and instead of the auxiliary solutions (2.3) we have

$$\chi_{1\pm}(z) = \frac{1}{\left(\begin{matrix}+\\ i\end{matrix}\right)[1 - (\lambda^2/k^2z^2) - (V(z)/k^2)]^{\frac{1}{4}}} \exp\left(\mp ikz + \int_{C\pm(z,\infty)} \left\{\left[V(\xi) + \frac{\lambda^2}{\xi^2} - k^2\right]^{\frac{1}{2}} \mp ik\right\} d\xi\right), \quad (2.9)$$

$$\varphi_{1\pm}(z) = \frac{1}{\left(\begin{matrix}+\\ i\end{matrix}\right)[1 - (\lambda^2/k^2z^2) - (V(z)/k^2)]^{\frac{1}{4}}} \exp\left(\pm ikz - \int_{C\pm(z,\infty)} \left\{\left[V(\xi) + \frac{\lambda^2}{\xi^2} - k^2\right]^{\frac{1}{2}} \mp ik\right\} d\xi\right). \quad (2.9)$$

These solutions are defined only for $\mathrm{Re}\ z \geq 0$. The analytic structure of the square roots appearing in (2.9) is much more complicated. The new feature is that there are many more branch points. For large λ, the branch points will be located only in the small neighborhood $\epsilon(\lambda)$ of the previous branch point λ/k, and in the narrow strip $0 \leq \mathrm{Re}\ z < \eta(\lambda)$. Both $\epsilon(\lambda)$ and $\eta(\lambda)$ tend to zero as $\lambda \to \infty$. Furthermore, it follows from the repulsive character of the potential in the interval $[0, x_0]$ that there will be no branch points in this interval. All cuts starting from the branch points in the neighborhood of λ/k will be taken to go to infinity like in the previous case. The cuts from the branch points in the strip will be taken to go to infinity parallel to the imaginary axis and not intersecting the real axis. The choice of the sheet is the same as in the previous case, i.e., the real part of the square root should be positive on the positive real axis. This analysis shows that the path $C_\pm$ must approach the origin along the positive real axis.

We construct now the Green's function in terms of the functions (2.9) in the same way as before. After introducing the new function

$$g_{1\pm}(z) = f_\pm(z)/\chi_{1\pm}(z), \quad (2.10)$$

the integral equation reads as follows:

$$g_{1\pm}(z) = 1 + \int_{C\pm(z,\infty)} G_{1\pm}(z, z')W_1(z')g_{1\pm}(z')\,dz', \quad (2.11)$$

where

$$G_{1\pm}(z, z') = \frac{1}{2[V(z') + (\lambda^2/z'^2) - k^2]^{\frac{1}{4}}} \times \left(1 - \exp\left\{-2\int_{C\pm(z,z')} \left[V(\xi) + \frac{\lambda^2}{\xi^2} - k^2\right]^{\frac{1}{2}} d\xi\right\}\right), \quad (2.12)$$

and

$$W_1(z) = -\frac{1}{4z^2} - \frac{5}{16}\left[\frac{V'(z) - (2\lambda^2/z^3)}{V(z) + (\lambda^2/z^2) - k^2}\right]^2 + \frac{1}{4}\frac{V''(z) + (6\lambda^2/z^4)}{V(z) + (\lambda^2/z^2) - k^2}. \quad (2.13)$$

The obtained integral equations (2.7) and (2.11) are of the type

$$h(z) = \int_{C(z,\infty)} H_1(z, z')\,dz' + \int_{C(z,\infty)} H_2(z, z')h(z')\,dz'. \quad (2.14)$$

If the kernel $H_i(z, z')$ can be majorized by a function $K_i(z')$, uniformly with respect to z on $C(\xi, \infty)$, and if $K_i\int(z')\,|dz'| = C_i(\xi) < \infty$, then the Titchmarsh's lemma gives the following majorization for $h(z)$:

$$|h(z)| < C_1(\xi)\exp\{C_2(\xi)\}, \quad z \in C(\xi, \infty). \quad (2.15)$$

Before applying this theorem to our integral equations we have to choose the paths $C_\pm(0, \infty)$. From (1.7) we see that we need the asymptotic behavior of both Jost solutions on the same path $C(0, z_a)$. Hence, we have to choose $C_\pm(0, z_a) = C(0, z_a)$. In Sec. 3 the path $C(0, z_a)$ will be chosen so that it crosses the saddle point $z_0 = (\lambda/k)u_0$, $0 < u_0 < 1$. Having in mind all the requirements imposed by now upon $C_\pm(0, \infty)$, we choose them as shown in Fig. 2. They depend upon λ. The points z_1 and z_2 are independent upon $|\lambda|$, and they are specified in more detail in Appendix I. We are forced to introduce these points because of the previous condition that the path $C_\pm(z, \infty)$ must approach the origin along the positive real axis.

In our case $h(z) = g(z) - 1$. The uniform majorization of the kernels is done in Appendix I with the result that $C_1(\xi, \lambda) = C_2(\xi, \lambda)$ vanish for large λ, $0 \leq \omega < \frac{1}{2}\pi$, at least like $C/|\lambda|^\alpha$, $\frac{1}{2} \leq \alpha \leq 1$. Hence, using this result in the formula (2.15) applied to our three integral equations, we find that the functions $g(z)$ have the following asymptotic behavior:

$$\begin{aligned} g_\pm(z) &= 1 + \rho_\pm(z), \qquad |\rho_\pm(z)| < C/|\lambda|, \\ &\qquad\qquad z \in C_\pm(z_1, \infty), \\ g_{0\pm}(z) &= 1 + \rho_{0\pm}(z), \qquad |\rho_{0\pm}(z)| < C/|\lambda|, \\ &\qquad\qquad z \in C_\pm(0, \infty), \qquad (2.16) \\ g_{1\pm}(z) &= 1 + \rho_{1\pm}(z), \qquad |\rho_{1\pm}(z)| < C/|\lambda|^{\frac{1}{2}}, \\ &\qquad\qquad z \in C_\pm(0, \infty). \end{aligned}$$

This asymptotic convergence for large λ, $0 \leq \omega < \frac{1}{2}\pi$, is uniform with respect to z in the indicated domains. Hence, our expectations that the auxiliary functions will reproduce in some domain the essential features of the Jost functions for large λ are fulfilled.

Apart from the asymptotic behavior of the functions $g(z)$, we shall need in Sec. 3 the majorization of the asymptotic behavior of the difference $g(z) - g_0(z)$ on $C_\pm(z_1, \infty)$ and also its exact asymptotic behavior in the interval $z = (\lambda/k)u$, $a < u_p \leq u \leq u_q < 1$. From the Eq. (2.7) we obtain the following integral equation for this difference

$$g_\pm(z) - g_{0\pm}(z) = \int_{C_+(z,\infty)} G_\pm(z, z')V(z')g_\pm(z')\,dz' + \int_{C_+(z,\infty)} G_\pm(z, z')W(z')[g_\pm(z') - g_{0\pm}(z')]\,dz'. \qquad (2.17)$$

This is an integral equation of the type (2.14), and we can apply the Titchmarsh's lemma (2.15).

The estimates of the Green's functions $G_\pm(z, z')$ and of the function $W(z')$ can be found in Appendix I. It is shown there that the Green's functions are bounded by a constant uniformly over $C_\pm(z_1, \infty)$, and the integral $\int |W(z')|\,|dz'|$ tends to zero for large λ. Hence, the result is

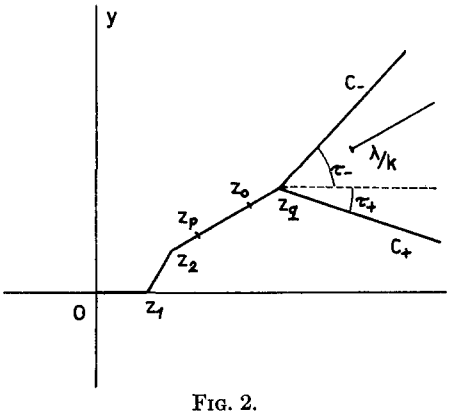

FIG. 2.

$$|g_\pm(z) - g_{0\pm}(z)| < C_1 \int_{C_+(z,\infty)} |V(z')|\,|dz'| < C\,|V_{a\pm}(z)|, \qquad (2.18)$$

where C does not depend on λ.

The second inequality follows from the analyticity and the asymptotic property (1.5) and (1.6) of the potential.

The evaluation of the exact asymptotic behavior of $g_\pm(z) - g_{0\pm}(z)$ in the region $z = (\lambda/k)u$, $0 < u < 1$ is slightly more complicated. The leading term comes from the first integral on the right-hand side of Eq. (2.17). The detailed calculations are performed in Appendix II, and the result is

$$g_\pm(z) - g_{0\pm}(z) = \frac{V_{a\pm}(z)}{\mu\{\mu + 2[(\lambda^2/z^2) - k^2]^{\frac{1}{2}}\}} \times [1 + \sigma_\pm(z)], \qquad z \in C(z_p, z_q), \qquad (2.19)$$

where

$$|\sigma_\pm(z)| < C/|\lambda|.$$

This completes our investigation of the asymptotic behavior of the Jost solutions. The formulas which we shall need in Sec. 3 are (2.16), (2.18), and (2.19).

The results of this section could have been derived easily also for complex k. But since, however, the calculations of Sec. 3 are restricted to real positive k, we kept this restriction in this section also.

3. ASYMPTOTIC BEHAVIOR OF THE PHASE SHIFT

Now we turn to the problem of the phase shift. In the formula (1.7) we shall consider first the

integral from which we expect the main contribution if the path has been properly chosen:

$$\mathfrak{I} = \mathfrak{I}(0, z_2) + \mathfrak{I}(z_2, z_q)$$
$$= k\left(\int_{C(0,z_2)} + \int_{C(z_2,z_q)}\right)\left[\frac{1}{f_+(z)f_-(z)} - \frac{1}{f_{0+}(z)f_{0-}(z)}\right]dz.$$

In terms of the functions $g(z)$, the integral $\mathfrak{I}(z_2, z_q)$ can be written as

$$\mathfrak{I}(z_2, z_q) = k\int_{C(z_2,z_q)} \frac{1}{\chi_+(z)\chi_-(z)} \times \left[\frac{1}{g_+(z)g_-(z)} - \frac{1}{g_{0+}(z)g_{0-}(z)}\right]dz.$$

The product $\chi_+(z)\chi_-(z)$ can be easily calculated from (2.3),

$$\chi_+(z)\chi_-(z) = \frac{1}{[(\lambda^2/k^2z^2) - 1]^{\frac{1}{2}}} \exp\left\{2\int_z^{\lambda/k}\left(\frac{\lambda^2}{\xi^2} - k^2\right)^{\frac{1}{2}} d\xi\right\}.$$

Inserting this expression into the integral we obtain

$$\mathfrak{I}(z_2, z_q) = \int_{C(z_2,z_q)} \left(\frac{\lambda^2}{z^2} - k^2\right)^{\frac{1}{2}} \times \exp\left\{-2\int_z^{\lambda/k}\left(\frac{\lambda^2}{\xi^2} - k^2\right)^{\frac{1}{2}} d\xi\right\} G(z)\, dz, \qquad (3.1)$$

where $G(z)$ is a function given by

$$G(z) = \frac{[g_{0+}(z) - g_+(z)]g_{0-}(z) + [g_{0-}(z) - g_-(z)]g_+(z)}{g_+(z)g_-(z)g_{0+}(z)g_{0-}(z)}. \qquad (3.2)$$

Let us first investigate the integral $\mathfrak{I}(z_p, z_q)$. In this interval we have the exact asymptotic behavior (2.19) of the difference $g_\pm(z) - g_{0\pm}(z)$. Inserting these expressions into (3.2) and using (2.16) for the functions $g(z)$, we can write down the asymptotic form of the function $G(z)$:

$$G(z) = -\frac{V_{as}(z)}{\mu\{\frac{1}{2}\mu + [(\lambda^2/z^2) - k^2]^{\frac{1}{2}}\}}[1 + \gamma(z)],$$
$$|\gamma(z)| < \frac{C}{|\lambda|}, \qquad z \in C(z_p, z_q). \qquad (3.3)$$

After inserting this expression into (3.1), introducing (1.6) and the new variable $z = (\lambda/k)u$, the integral $\mathfrak{I}(z_p, z_q)$ reads as follows:

$$\mathfrak{I}(z_p, z_q) = -\frac{k^{\alpha+1}}{\lambda^\alpha}\frac{V_0}{\mu}\int_{u_p}^{u_q} \frac{[(1/u^2) - 1]^{\frac{1}{2}}}{u^{\alpha+1}\{\frac{1}{2}\mu + k[(1/u^2) - 1]^{\frac{1}{2}}\}} \times \exp\left\{-2\lambda\int_u^1\left(\frac{1}{v^2} - 1\right)^{\frac{1}{2}} dv - \lambda\frac{\mu}{k}u\right\} \times \left[1 + \gamma\left(\frac{\lambda}{k}u\right)\right] du. \qquad (3.4)$$

Denoting the exponent by

$$h(u) = 2\int_u^1\left(\frac{1}{v^2} - 1\right)^{\frac{1}{2}} dv + \frac{\mu}{k}u,$$

we see that the equation

$$0 = h'(u) = -2[(1/u^2) - 1]^{\frac{1}{2}} + \mu/k$$

has a positive solution

$$u_0 = 1/[1 + (\mu^2/4k^2)]^{\frac{1}{2}}.$$

Since $h''(u_0) > 0$, we have a saddle point at $u = u_0$. This is the point $z_0 = (\lambda/k)u_0$ indicated in Fig. 2. We have chosen the path of integration so that it passes through this point up to the point $z_q = (\lambda/k)u_q$, $u_0 < u_q < 1$. We can use now the known asymptotic expansion which follows from the saddle point method. The reader can find it for real λ in Ref. 4 and see immediately for himself that it is valid also for complex λ with Re $\lambda > 0$. In this way we find

$$\int_{u_p}^{u_q} \frac{[(1/u^2) - 1]^{\frac{1}{2}}}{u^{\alpha+1}\{(\mu/2) + k[(1/u^2) - 1]^{\frac{1}{2}}\}} e^{-\lambda h(u)}\, du = \frac{1}{2k}\left(\frac{\pi}{2\lambda}\right)^{\frac{1}{2}}\left(\frac{\mu}{k}\right)^{\frac{1}{2}}\left(1 + \frac{\mu^2}{4k^2}\right)^{\frac{1}{2}\alpha - \frac{1}{4}} e^{-\lambda h(u_0)}\left[1 + O\left(\frac{1}{\lambda}\right)\right]. \qquad (3.5)$$

The error represented by the second integral over $\gamma(z)$ in (3.4) can be estimated absolutely. In this integral $\gamma(z)$ and λ are complex and $h(u)$ as well as the ratio in front of the exponential function are positive functions. Hence, the absolute value of the integral is smaller than the product of the majorization of $\gamma(z)$ given in (3.3) and the integral (3.5) in which λ is substituted by Re λ. Since this last integral is of the same order as (3.5), we conclude that the contribution of the integral with $\gamma(z)$ is of the order $O(1/\lambda)$ compared to the leading term (3.5). So the final result is

$$\mathfrak{I}(z_p, z_q) = -\frac{V_0}{2k}\left(\frac{\pi}{2\lambda}\right)^{\frac{1}{2}}\frac{1}{[\sinh\rho(\mu)]^{\frac{1}{2}}} \times \left[\frac{k^2}{\mu\lambda}\sinh\rho(\mu)\right]^\alpha e^{-\lambda\rho(\mu)}\left[1 + O\left(\frac{1}{\lambda}\right)\right], \qquad (3.6)$$

where $\rho(\mu) = h(u_0)$ is given by

$$\rho(\mu) = \cosh^{-1}[1 + (\mu^2/2k^2)]. \qquad (3.7)$$

Next, we have to estimate the integrals $\mathfrak{I}(0, z_2)$ and $\mathfrak{I}(z_2, z_p)$. Let us begin with the first integral $\mathfrak{I}(0, z_2)$. The functions $f_{0\pm}(z)$ behave like the func-

[4] M. Evgrafov, *Asymptotic Estimates and Entire Functions* (Gordon and Breach Science Publishers, New York, 1961), pp. 19–20.

tions $\chi_{\pm}(z)$ for large λ uniformly with respect to $z \in C(0, z_2)$. By inspecting these functions defined by (2.3) one can easily check that the uniform behavior of their product is $(2/e)(2\lambda/kze)^{2\lambda-1}$. Hence, that part of the integral $\mathfrak{z}(0, z_2)$ which contains the function $1/f_{0+}(z)f_{0-}(z)$ can be estimated by const·exp $(-\alpha\,|\lambda|\,\ln\,|\lambda|)$ which vanishes faster than the exponential function of $|\lambda|$. That part of the integral $\mathfrak{z}(0, z_2)$ which contains the function $1/f_{+}(z)f_{-}(z)$ can be estimated by the same constant depending on λ, but the estimation is more complicated.

We can transform the function $1/f_{+}(z)f_{-}(z)$ in the following way:

$$\frac{1}{f_{+}(z)f_{-}(z)} = \frac{1}{f_{0+}(z_2)f_{0-}(z_2)}\,\frac{f_{0+}(z_2)f_{0-}(z_2)}{f_{+}(z)f_{-}(z)}$$

$$= \frac{1}{f_{0+}(z_2)f_{0-}(z_2)}\left[\frac{1-(\lambda^2/k^2z^2)-V(z)/k^2}{1-(\lambda^2/k^2z^2)}\right]^{\frac{1}{2}} \times \exp\left\{-2\int_{C(z,z_2)}\left[V(\xi)+\frac{\lambda^2}{\xi^2}-k^2\right]^{\frac{1}{2}}d\xi\right\} \times \frac{g_{0+}(z_2)g_{0-}(z_2)}{g_{1+}(z)g_{1-}(z)}\left[1+O\left(\frac{1}{\lambda}\right)\right].$$

Using formulas (2.16), Appendix III, and the estimate of the function $1/f_{0+}(z)f_{0-}(z)$, we are able to obtain the simple inequality

$$\left|\frac{1}{f_{+}(z)f_{-}(z)}\right| < C_1 \exp\left[-\alpha\,|\lambda|\,\ln\,|\lambda|\right]\left[|V(z)|\right]^{\frac{1}{2}} \times \exp\left\{-p\theta(z-z_1)\int_{C(z,z_1)}[V(x)]^{\frac{1}{2}}\,dx\right\} < C_2 \exp\left[-\alpha\,|\lambda|\,\ln\,|\lambda|\right].$$

Hence, that part of the integral $\mathfrak{z}(0, z_2)$ which contains the function $1/f_{+}(z)f_{-}(z)$ under the integral sign, can be estimated by the same constant: const·exp $(-\alpha\,|\lambda|\,\ln\,|\lambda|)$, which was our assertion.

Let us look at the integral $\mathfrak{z}(z_2, z_p)$. Its estimate can be made in a simple way using the same transformation as in the case of the consideration of the integral $\mathfrak{z}(z_p, z_q)$. We shall use asymptotic forms (2.16), uniform estimate (2.18), and the notation will be as in the formula (3.5). Combining these together one obtains the following majorization:

$$|\mathfrak{z}(z_2, z_p)| < |\lambda|\,C_1 \int_0^{u_p} \exp\left\{-\operatorname{Re}\,\lambda h(u)\right\} du.$$

The minimal value of the function $h(u)$ defined on the interval (0, 1) is at the point u_0. Since $u_p < u_0$ its minimal value is at the boundary u_p of the integration. Using Theorem 3 of Ref. 4, we can obtain a simple majorization of the integral $\mathfrak{z}(z_2, z_p)$:

$$|\mathfrak{z}(z_2, z_p)| < C_2 \exp\left[-\operatorname{Re}\,\lambda h(u_p)\right].$$

Here the right-hand side tends to zero faster than the integral $\mathfrak{z}(z_p, z_q)$ expressed as in (3.6) since $h(u_p) > h(u_0) = \rho(\mu)$.

Until now we have shown that the dominant term of the integral $\mathfrak{z}(0, z_q)$ is represented by (3.6). In order to prove that the phase shift behaves for large λ just like the right-hand side of (3.6), we must still show that the logarithmic term in the expression (1.7) of the phase shift vanishes faster than the dominant term (3.6). But the function $f_{+}(z_q)f_{0-}(z_q)/f_{-}(z_q)f_{0+}(z_q)$ under the sign of the logarithm behaves like $1 + \text{const} \cdot V_{as}(z_q)$ as is easily seen by inspecting formulas (2.6), (2.10), (2.16), and (2.19). As generally $z_q = (1-\epsilon)\lambda/k$ one can always choose such a small ϵ for fixed k in order to make the function $V_{as}(z_q)$ decrease faster than the function $\exp\,[-\lambda\rho(\mu)]$ in the expression (3.6).

Thus, we have shown that the asymptotic form of the phase shift is described for large λ completely by the expression (3.6). But the asymptotic form (3.6) is valid in a larger domain than is specified during the proof: $0 \le \omega < \frac{1}{2}\pi$. Namely, if we use the unitarity condition of the S-matrix for real k: $S^*(\lambda, k) = S^{-1}(\lambda^*, k)$ and the corresponding relation between the phase shifts $\delta^*(\lambda, k) = \delta(\lambda^*, k)$, we extend immediately the validity of (3.6) to the whole half-plane $\operatorname{Re}\,\lambda > 0$:

$$\delta(\lambda, k) = -\frac{V_0}{2k}\left(\frac{\pi}{2\lambda}\right)^{\frac{1}{2}}\frac{1}{[\sinh\,\rho(\mu)]^{\frac{1}{2}}}\left[\frac{k^2}{\mu\lambda}\sinh\,\rho(\mu)\right]^{\alpha} \times e^{-\lambda\rho(\mu)}\left[1+O\left(\frac{1}{\lambda}\right)\right], \qquad |\arg\,\lambda| < \tfrac{1}{2}\pi. \tag{3.8}$$

At the end of this section let us say a few words about another possible form of the asymptotic form (3.8). The conditions on the potential permit us to represent it as the Laplace transform

$$V(z) = \int_{\mu}^{\infty}\frac{\exp\,(-tz)}{z}\,C(t)\,dt. \tag{3.9}$$

All the considerations can be repeated with the representation (3.9), and the asymptotic form of the phase shift will be in this case

$$\delta(\lambda, k) = -\frac{1}{2k}\left(\frac{\pi}{2\lambda}\right)^{\frac{1}{2}} \times \int_{\mu}^{\mu+\nu} C(t)\,\frac{e^{-\lambda\rho(t)}}{[\sinh\,\rho(t)]^{\frac{1}{2}}}\,dt\left[1+O\left(\frac{1}{\lambda}\right)\right]. \tag{3.10}$$

Here the integration cannot be taken over (μ, ∞) since the function $C(t)$ may increase faster than any power of t for large t due to the high singularity of the potential at the origin.

4. THE STRUCTURE OF THE S-MATRIX

It is known that the S-matrix is a meromorphic function in the λ plane in the case of scattering on highly singular potentials.[1,2] The S-matrix possesses a simple property of the symmetry[1]

$$S(\lambda, k) = \exp(2i\pi\lambda) S(-\lambda, k). \tag{4.1}$$

Knowing the asymptotic form of the S-matrix in the half-plane $\mathrm{Re}\,\lambda > 0$ derived in the last section, and using (4.1), one can obtain the asymptotic form of the S-matrix in the whole λ-plane except the neighborhood of the imaginary axis. We prefer to consider the quotient of the Jost functions instead of the S-matrix itself since this quotient is the function of λ^2 only

$$R(\lambda^2, k) = f(\lambda^2, k)/f(\lambda^2, -k), \tag{4.2}$$

where $f(\lambda^2, \pm k)$ are the Jost functions. Using relation

$$R(\lambda^2, k) = \exp[-i\pi(\lambda - \tfrac{1}{2})] S(\lambda, k)$$

we can establish the asymptotic form of the function $R(\lambda^2, k)$ along any ray $\lambda = |\lambda| \exp(i\omega)$, $\omega \neq \pm \tfrac{1}{2}\pi$:

$$R(\lambda^2, k) \sim \exp\{\mp i\pi(\lambda - \tfrac{1}{2})\} \quad \text{when} \quad \mathrm{Re}\,\lambda \to \pm\infty. \tag{4.3}$$

The asymptotic form (4.3) cannot be extended to the imaginary axis [for purely imaginary λ the only known property is $|R(\lambda^2, k)| = 1$] because there are an infinite number of poles and zeros accumulating at infinity in the direction of the imaginary axis. We shall prove this assertion supposing that there exists a finite number of poles only, and then we shall show the contradiction. Hence, let poles be at the points $\pm\lambda_k$ and the corresponding zeros at the points $\pm\lambda_k^*$. We suppose that all the poles are simple which does not limit the generality. Then the function

$$P(\lambda^2, k) = \prod_{k=1}^{n} \frac{(\lambda^2 - \lambda_k^2)}{(\lambda^2 - \lambda_k^{*2})} R(\lambda^2, k)$$

is the entire function without zeros with the same asymptotic form as the function $R(\lambda^2, k)$. Let us show that the order of the entire function $P(\lambda^2, k)$ is at most two. It can be easily proved that the regular solution, Jost solutions and their derivatives are the entire functions of the order two at most. Since the combination of these functions determines the Jost functions, the Jost functions may have order two at most. By our supposition the Jost functions have a finite number of zeros and can be represented in the form $f(\lambda^2, \pm k) = C_\pm \prod_\pm(\lambda^2) \exp\{Q_\pm(\lambda^2)\}$ according to Hadamard's theorem,[5] where $\prod_\pm(\lambda^2)$ is a canonical product and $Q_\pm(\lambda^2)$ is the polynomial $a_\pm + \lambda^2 b_\pm$. Hence, the quotient of the Jost functions and consequently the function $P(\lambda^2, k)$ must be of order two at most. Thus using Hadamard's theorem

$$P(\lambda^2, k) = C \exp(a + b\lambda^2).$$

One can immediately see that this form of the function $P(\lambda^2, k)$ cannot join together with its asymptotic form. Hence, the number of poles and zeros is infinite for the functions $R(\lambda^2, k)$ and the S-matrix.

At the end of this section we shall show for example the proof that the regular solution is the function of the second order at most in λ. If we introduce the function

$$\psi(x) = V(x)^{-\frac{1}{4}} \exp\left\{-\int_x^{x_0} V(y)^{\frac{1}{2}}\, dy\right\}, \qquad x \leq x_0,$$

then the regular solution $\varphi(\lambda^2, k^2, x)$ can be represented in the form $\varphi(\lambda^2, k^2, x) = \psi(x) u(\lambda^2, k^2, x)$ where the function $u(\lambda^2, k^2, x)$ is the solution of the integral equation[2]

$$u(\lambda^2, k^2, x) = 1 + \int_0^x \frac{1}{2[V(y)]^{\frac{1}{2}}} \left[1 - \exp\left\{-2\int_y^x [V(z)]^{\frac{1}{2}}\, dz\right\}\right] \times \left\{\frac{\lambda^2 - \frac{1}{4}}{y^2} - k^2 - \frac{5}{16}\left[\frac{V'(y)}{V(y)}\right]^2 + \frac{1}{4}\frac{V''(y)}{V(y)}\right\} u(\lambda^2, k^2, y)\, dy.$$

The kernel of this integral equation can be estimated as in Ref. 2, and one obtains

$$|u(\lambda^2, k^2, x)| < C_1 \exp(C_2 |\lambda|^2)$$

which proves our assertion.

ACKNOWLEDGMENTS

The first author would like to thank the Institute of Physics in Torino for kind hospitality and a grant, and to thank Professor Abdus Salam for hospitality at the International Centre for Theoretical Physics during the last stage of the work on this problem. The second author is grateful to Professor Abdus Salam and the International Atomic Energy Agency for the hospitality extended to him at the International Centre for Theoretical Physics, Trieste.

APPENDIX I

We want to show in this appendix that the kernels $H_i(z, z')$ of the integral Eqs. (2.7) and (2.11) can

[5] M. L. Cartwright, *Integral Functions* (Cambridge University Press, New York, 1962), p. 20.

be majorized by the integrable functions $K(z')$ over $C_\pm$ and that $\int K(z')\,|dz'|$ vanish for large λ. All the considerations here will be connected with the kernel $H(z, z')$ which corresponds to Eq. (2.11) since the other kernel is the specification of the first. Hence,

$$H(\xi, z) = \frac{1}{2[V(z) + (\lambda^2/z^2) - k^2]^{\frac{1}{2}}} \times \left(1 - \exp\left\{-2\int_{C_\pm(\xi,z)} \left[V(u) + \frac{\lambda^2}{u^2} - k^2\right]^{\frac{1}{2}} du\right\}\right) \times \left\{-\frac{1}{4z^2} - \frac{5}{16}\left[\frac{V'(z) - (2\lambda^2/z^3)}{V(z) + (\lambda^2/z^2) - k^2}\right]^2 + \frac{1}{4}\frac{V''(z) + (6\lambda^2/z^4)}{V(z) + (\lambda^2/z^2) - k^2}\right\}. \tag{AI.1}$$

The variable ξ appears only in the exponential function. Let us estimate first the exponential function in order to get rid of the variable ξ. We must prove that the real part of the exponent has a finite upper bound. We can rewrite the exponent in the form

$$-\int_{C_\pm(\xi,z)} \left(\frac{\lambda^2}{u^2} - k^2\right)^{\frac{1}{2}} du - \int_{C_\pm(\xi,z)} \left\{\left[V(u) + \frac{\lambda^2}{u^2} - k^2\right]^{\frac{1}{2}} - \left(\frac{\lambda^2}{u^2} - k^2\right)^{\frac{1}{2}}\right\} du. \tag{AI.2}$$

We shall divide the path $C_\pm(\xi, z)$ into four parts, $C(\xi, z_1) + C(z_1, z_2) + C(z_2, z_a) + C_\pm(z_a, \infty)$. The real part of the first integral over the first part of the path is positive since u is real. It is positive over the second part if we choose it as the part of the spiral $r = z_1 \exp(A\tau)$, $\tau \geq 0$, $A > tg\omega$, between the point $r = z_1$, $\tau = 0$, and the point $r = z_1 \exp(A\omega)$, $\tau = \omega$. Namely, if $\xi, z \in C(z_1, z_2)$ then the first integral in (AI.2) behaves like $\lambda \ln z/\xi$ for large λ. It can be easily checked that the real part of this simple function is positive if z_1 and z_2 are specified in the way mentioned. The first integral has the argument ω along the third part, hence its real part is positive.

We have $dz_\pm = dr \exp(i\tau_\pm)/\cos(\tau_\pm - \tau)$ along $C_\pm(z_a, \infty)$ where $\tau = \arg z$. For the first integral over the paths $C_\pm(z_a, \infty)$. we have the form

$$\int \left(\frac{\lambda^2}{z^2} - k^2\right)^{\frac{1}{2}} \frac{\exp(i\tau_\pm)}{\cos[\tau_\pm - \tau(r)]}\, dr. \tag{AI.3}$$

If z is on the path C_+, we can see that the real and imaginary parts of the square root are positive. Since $|\tau_+ - \tau| < \frac{1}{2}\pi$, $\tau_+ < 0$, we verify that the real part of the intergal (AI.3) is positive. If z is on the path C_-, the real part of the square root is positive, and its imaginary part is negative. But now $\tau_- > 0$, and we conclude again that the real part of (AI.3) is positive. Thus the first integral in (AI.2) has always a positive real part. For the second integral in (AI.2), it is easy to show that its real part has the finite minimal value. Namely, its absolute value on $C_\pm(z_1, \infty)$ is bounded because it behaves like $\int |du|\,|V(u)|/|[(\lambda^2/u^2) - k^2]^{\frac{1}{2}}|$ on the part $C_\pm(z_1, \infty)$, and its real part is positive on $C(\xi, z_1)$ since $V(u)$ is a positive function for $u \leq z_1$. After this analysis we can majorize the exponential function in (AI.1) by a constant.

To estimate the first factor and the brackets in (AI.1), we shall use the following rule: If $\mathbf{a}$ and $\mathbf{b}$ are vectors with an angle φ between them which is less than π, then always holds $|\mathbf{a} + \mathbf{b}| \geq \mathbf{a}\cdot\mathbf{s} + \mathbf{b}\cdot\mathbf{s} = p\,|\mathbf{a}| + p\,|\mathbf{b}|$, where $\mathbf{s}$ is the ort of the simetrale of the angle φ. In the following let us choose always λ larger than some fixed λ_0, where $|\lambda_0/k| \gg 1$. Then the first factor can be estimated by $1/p[V(x) + |\lambda|^2/x^2]^{\frac{1}{2}}$ on $C(0, z_1)$, by $|z/\lambda|$ on $C(z_1, z_a)$ and by constant on $C_\pm(z_a, \infty)$. The brackets of (AI.1) which are denoted by $W_1(z)$ have three terms. Let us show the majorization of the second term which is the most complicated.

$$\left[\frac{V'(x) - (2\lambda^2/x^3)}{V(x) + (\lambda^2/x^2) - k^2}\right]^2 = \left[\frac{V'(x) - 2(\lambda^2/x^3)}{V(x) + (\lambda^2/x^2)}\right]^2 \left[\frac{V(x) + (\lambda^2/x^2)}{V(x) + (\lambda^2/x^2) - k^2}\right]^2.$$

The second factor in this expression can be estimated by a constant depending on λ_0 only. The denominator of the first factor can be minimized by $p^2[V(x) + |\lambda|^2/x^2]^2$. Hence, the first factor can be estimated by $N[V'(x)/V(x)]^2 + M/x^2$. After estimating other terms of $W_1(x)$ in the same way, we obtain the majorization of the function (AI.1) in the interval $C(0, z_1)$ by

$$K_1(x) = \frac{C}{[V(x) + |\lambda|^2/x^2]^{\frac{1}{2}}}\left\{\frac{A_1}{x^2} + A_2\left[\frac{V'(x)}{V(x)}\right]^2 + A_3\frac{|V''(x)|}{V(x)}\right\}, \qquad x \in C(0, z_1). \tag{AI.4}$$

We shall separate the potential from the expression $W_1(z)$ in the following consideration. Thus we shall obtain $W(z)$ defined in (2.5). By pure algebraic calculations one verifies that the difference $W_1(z) - W(z)$ can be estimated by $\exp(-\nu\,|z|)$ uniformly for λ larger than λ_0 and $z \in C_\pm(z_1, \infty)$. But the estimate of the function $W(z)$ on $C_\pm(z_1, \infty)$ is simple:

$$|W(z)| < \frac{k^2\,|\lambda|^2 + \frac{1}{4}k^4\,|z|^2}{p\,|\lambda|^4 + pk^4\,|z|^4}.$$

In this way we have the following majorizations:

$$K_2(z) = (N\ |z|/|\lambda|)[\exp(-\nu\ |z|) + (P/|\lambda|^2)], \quad z \in C(z_1, z_a),$$
$$K_3(z) = M[\exp(-\nu\ |z|) + (R/|z|^2)], \quad z \in C_\pm(z_a, \infty). \tag{AI.5}$$

In this way we obtained the function $K(z)$. It still remains to be shown that the integral of the function $K(z)$ vanishes when λ tends to infinity.

$$\int_{C_\pm(0,\infty)} K(z)\ |dz| = \int_{C(0,z_1)} K_1(x)\ dx + \int_{C(z_1,z_a)} K_2(z)\ |dz| + \int_{C_\pm(z_a,\infty)} K_3(z)\ |dz|. \tag{AI.6}$$

It is easy to check that the second and third integrals on the right-hand side vanish like $1/|\lambda|$ for large λ. We must use expressions (AI.5) and the fact $|dz| < \text{const}\cdot d|z|$ on $C_\pm$. Let us show now that the first integral vanishes for large λ. We shall denote by $R(x)$ the expression in the brackets in (AI.4). We have

$$\int_0^{x_1} \frac{1}{\left[V(x) + \frac{|\lambda|^2}{x^2}\right]^{\frac{1}{2}}} R(x)\ dx = \int_0^{x_0} \left[\frac{V(x)}{V(x) + \frac{|\lambda|^2}{x^2}}\right]^{\frac{1}{2}} \frac{R(x)}{[V(x)]^{\frac{1}{2}}}\ dx + \int_{x_0}^{x_1} \left[\frac{V(x)}{V(x) + \frac{|\lambda|^2}{x^2}}\right]^{\frac{1}{2}} \frac{R(x)}{[V(x)]^{\frac{1}{2}}}\ dx,$$

where x_0 is the solution of the equation $x^2V(x) = |\lambda|$. As we postulated the monotony of the function $x^2V(x)$ at the origin, we are sure that there exists only one x_0 for large λ. Now we have

$$\int_0^{x_1} \frac{R(x)\ dx}{\left[V(x) + \frac{|\lambda|^2}{x^2}\right]^{\frac{1}{2}}} < \int_0^{x_0} \frac{R(x)}{[V(x)]^{\frac{1}{2}}}\ dx + \frac{1}{(|\lambda|)^{\frac{1}{2}}} \int_0^{x_1} \frac{R(x)}{[V(x)]^{\frac{1}{2}}}\ dx. \tag{AI.7}$$

Both integrals on the right-hand side exist because of the conditions (1.3) of the potential and both integrals vanish for large λ. This fact is evident for the second integral; for the first integral we have to remark that x_0 tends to zero. It can be shown that (AI.7) vanishes like $|\lambda|^{-\frac{1}{2}}$ in the case of the potentials which behave like $x^{-\alpha}$, $\alpha > 2$ or like $\exp(1/x)$ at the origin.

In the case of the solutions of the integral Eq. (2.7), there is no complication at the origin since these solutions are defined in the half-plane $\mathrm{Re}\, z > x_0$. Then one needs to estimate the second and third term in the brackets of (AI.1) and will conclude

$$g(z) \sim 1 + O(1/\lambda), \qquad z \in C(z_1, \infty).$$

APPENDIX II

Our intention here is to prove the asymptotic form (2.19). The starting formula will be (2.17). We shall show that the first term of the right-hand side of (2.17) where $g_\pm(z)$ is replaced by 1 is the leading term. We shall denote it by $a(z)$. Using (2.8) we can write it in the explicit form

$$a(z) = \int_{C_\pm(z,\infty)} \frac{1}{2[(\lambda^2/z'^2) - k^2]^{\frac{1}{2}}} \left(1 - \exp\left\{-2 \times \int_{C_\pm(z,z')} [(\lambda^2/\xi^2) - k^2]^{\frac{1}{2}}\ d\xi\right\}\right) V(z')\ dz'. \tag{AII.1}$$

We use the new variable $z = (\lambda/k)u$. Then the integral becomes

$$a[(\lambda/k)u] = \frac{1}{2k(\lambda/k)^\alpha} \int_u^{u_a} \frac{1 + u[(\lambda/k)x]}{x^{\alpha+1}[(1/x^2) - 1]^{\frac{1}{2}}} \times \left(\exp\{-(\lambda\mu/k)x\} - \exp\left\{-(\lambda\mu/k)x - 2\lambda \int_u^x [(1/y^2) - 1]^{\frac{1}{2}}\ dy\right\}\right) dx + O[V_{as}(z_a)]. \tag{AII.2}$$

Let us apply here Theorem 3 of Ref. 4. Although this theorem holds for real λ it can be easily generalized for complex λ from the right half of the λ-plane. As each term of the function under the sign of the integration in (AII.2) satisfies separately the conditions of the theorem mentioned, we can apply this theorem to integrals over the first and over the second part of the integrand separately. We shall obtain directly the asymptotic form in λ. Going back to the variable z we have

$$a(z) = \frac{V_{as}(z)}{\mu\{\mu + 2[(\lambda^2/z^2) - k^2]^{\frac{1}{2}}\}} [1 + \sigma(z)]. \tag{AII.3}$$

Now we have the following estimate of the Eq. (2.17):

$$|g_\pm(z) - g_{0\pm}(z) - a(z)| < A_1 \int_{C_\pm(z,\infty)} |V(z')(g_\pm(z') - 1)\ dz'| + A_2 \int_{C_\pm(z,\infty)} |W(z')a(z')\ dz'| + \int_{C_\pm(z,\infty)} |W(z')|\ |g_\pm(z') - g_{0\pm}(z') - a(z')|\ |dz'|.$$

We use the asymptotic forms (2.16) and the Titchmarsh lemma (2.15), where now

$$C_1(z) = \left(N_1/|\lambda|\right) \int_{C(z,\infty)} |V(z)\, dz|$$

and

$$C_2 = \int_{C(z_p,\infty)} |W(z)\, dz|.$$

Using Theorem 3 of Ref. 4 to the function $C_1(z)$, we can find its asymptotic behavior: $C_1(z) = (N_2/|\lambda|)\, |V_{as}(z)|\, [1 + b(z)]$. Hence,

$$|g_{\pm}(z) - g_{0\pm}(z) - a(z)| < N_3\, |V_{as}(z)|/|\lambda|. \quad \text{(AII.4)}$$

As $|a(z)|$ behaves like $|V_{as}(z)|$ for large z, we conclude that the right-hand side of (AII.4) makes the error of the order $1/|\lambda|$ to the asymptotic behavior (AII.3) of the function $a(z)$. Thus we proved the assertion (2.19).

APPENDIX III

We have to show here the validity of the inequality

$$\operatorname{Re} \int_{C(z,z_1)} \left[V(\xi) + \frac{\lambda^2}{\xi^2} - k^2 \right]^{\frac{1}{2}} d\xi > p\theta(z - z_1) \int_z^{z_1} [V(x)]^{\frac{1}{2}}\, dx. \quad \text{(AIII.1)}$$

The part of the integral over $C(z_1, z_2)$ has a positive real part as is shown in Appendix I in the estimates of the exponential function in the Green's function. Hence, we must show the validity of (AIII.1) for the integral over $C(z, z_1)$. (AIII.1) will hold if

$$\operatorname{Re}\, [V(x) + (\lambda^2/x^2) - k^2]^{\frac{1}{2}} > p[V(x)]^{\frac{1}{2}} \quad \text{for } x \in (0, z_1). \quad \text{(AIII.2)}$$

Let us denote the difference $V(x) - k^2$ by $P(x)$. If we prove the inequality (AIII.2) for the function $P(x)$ on the right-hand side instead of the function $V(x)$ it will be satisfactory because k is fixed parameter and always holds $[V(x) - k^2]^{\frac{1}{2}} > q[V(x)]^{\frac{1}{2}}$. Hence, let us denote $R = P(x) + \operatorname{Re} \lambda^2/x^2$ and $I = \operatorname{Im} \lambda^2/x^2$. We shall use

$$\operatorname{Re}\, [P(x) + \lambda^2/x^2]^{\frac{1}{2}} = (1/2^{\frac{1}{2}})[R + (R^2 + I^2)^{\frac{1}{2}}]^{\frac{1}{2}}.$$

We deduce from this formula

$$\operatorname{Re}\, [P(x) + \lambda^2/x^2]^{\frac{1}{2}} > [P(x)]^{\frac{1}{2}} \quad \text{if} \quad \operatorname{Re} \lambda^2 \geq 0$$

and

$$\operatorname{Re}\, [P(x) + \lambda^2/x^2]^{\frac{1}{2}} > [P(x)]^{\frac{1}{2}}/2^{\frac{1}{2}} \quad \text{for} \quad \operatorname{Re} \lambda^2 \geq -\operatorname{Im} \lambda^2.$$

In the range where $\operatorname{Im} \lambda^2 < -\operatorname{Re} \lambda^2$, i.e., where $\frac{3}{8}\pi < \omega < \frac{1}{2}\pi$, we shall use the formula

$$[P(x) + \lambda^2/x^2]^{\frac{1}{2}} = (R^2 + I^2)^{\frac{1}{2}} \exp\, [i/2 \operatorname{arc\,cot} (R/I)]$$

Here

$$R^2 + I^2 = P(x) + 2P(x)\, |\lambda|^2 \cos 2\omega/x^2 + |\lambda|^4/x^4$$

is larger than $P(x) \sin^2 2\omega$ and $\operatorname{arc\,ctg}\, (R/I)$ is less than 2ω. Hence, $\operatorname{Re}\, [P(x) + \lambda^2/x^2]^{\frac{1}{2}} > \sin 2\omega \cos \omega [V(x)]^{\frac{1}{2}}$. Thus, we are always able to find out the constant p in the inequality (AIII.2). It is really 1, $1/2^{\frac{1}{2}}$ or $\sin 2\omega \cos \omega$ in the case $0 \leq \omega < \frac{1}{4}\pi$, $\frac{1}{4}\pi \leq \omega < \frac{3}{8}\pi$ or $\frac{3}{8}\pi \leq \omega < \frac{1}{2}\pi$, respectively.

Rigorous Treatment of the Van Der Waals–Maxwell Theory of the Liquid–Vapor Transition*

J. L. LEBOWITZ AND O. PENROSE†
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 4 March 1965)

Rigorous upper and lower bounds are obtained for the thermodynamic free-energy density $a(\rho, \gamma)$ of a classical system of particles with two-body interaction potential $q(r) + \gamma^\nu\varphi(\gamma r)$ where ν is the number of space dimensions and ρ the density, in terms of the free-energy density $a^0(\rho)$ for the corresponding system with $\varphi(\mathbf{x}) \equiv 0$. When $\varphi(\mathbf{x})$ belongs to a class of functions, which includes those which are nonpositive and those whose ν-dimensional Fourier transforms are nonnegative, the upper and lower bounds coincide in the limit $\gamma \to 0$ and $\lim_{\gamma\to 0} a(\rho, \gamma)$ is the maximal convex function of ρ not exceeding $a^0(\rho) + \frac{1}{2}\alpha\rho^2$, where $\alpha \equiv \int \varphi(\mathbf{x})\, d\mathbf{x}$. The corresponding equation of state is given by Maxwell's equal-area rule applied to the function $p^0(\rho) + \frac{1}{2}\alpha\rho^2$ where $p^0(\rho)$ is the pressure for $\varphi(\mathbf{x}) \equiv 0$. If $a^0(\rho) + \frac{1}{2}\alpha\rho^2$ is not convex the behavior of the limiting free energy indicates a first-order phase transition. These results are easily generalized to lattice gases and thus apply also to Ising spin systems.

The two-body distribution function is found, in the limit $\gamma \to 0$, to be normally identical with that for $\varphi(\mathbf{x}) \equiv 0$, but if the system has a phase transition it has the form appropriate to a two-phase system.

Some of the upper and lower bounds on $a(\rho, \gamma)$ are simple enough to be useful for finite γ. Also, some of our results remain valid for quantum systems.

I. INTRODUCTION

PHASE transitions such as melting and boiling are familiar experiences, but their explanation from the first principles of statistical mechanics still presents a major challenge to the theoretical physicist. One of the earliest steps towards a theory of the gas–liquid phase transition was taken by van der Waals.[1] Seeing the interaction between the molecules of a classical fluid as a competition between two distinct parts of the intermolecular force, a short-range repulsive part and a long-range attractive part, he arrived at the equation of state

$$p = kT\rho/(1 - \rho b) + \tfrac{1}{2}\alpha\rho^2 \equiv p_{vdw}(\rho, T) \qquad (1.1)$$

where p is the pressure, k Boltzmann's constant, T the temperature, ρ the number density, and $-\alpha$ and b are positive constants characterizing long- and short-range parts of the potential, respectively.

When T exceeds the critical temperature $T_c \equiv -4\alpha/27bk$, the van der Waals equation of state (1) gives a good qualitative representation of the isotherms of a real fluid; for $T < T_c$, however, each isotherm includes a section where the compressibility is negative, in violation of the thermodynamic stability principle. The reason for this failure is that the argument leading to (1) assumes a single-phase system; it does not allow for the possibility of coexisting liquid and vapor phases.

Maxwell[2] showed that the coexistence region could be included in the theory by using van der Waals' equation of state for both liquid and vapor phases and using the thermodynamic equilibrium condition that the two phases must have equal pressures and chemical potentials. This leads to the following modification of (1) for $T < T_c$:

$$p = \begin{cases} p_{vdw}(\rho, T) \text{ if } \rho < \rho_v(T) \text{ or } \rho > \rho_l(T) \\ p_{sat}(T) \text{ if } \rho_v(T) < \rho < \rho_l(T) \end{cases} \qquad (1.2)$$

where $\rho_v(T)$, $\rho_l(T)$ and $p_{sat}(T)$ may be determined by the graphical construction shown in Fig. 1.

A very interesting derivation of van der Waals' equation of state with Maxwell's rule was given recently by van Kampen.[3] In this derivation the volume Ω occupied by the system is divided into a large number of cells, each small compared with the range of the long-range attractive force, but large enough to contain many particles. Avoiding the pitfall of assuming a uniform distribution of particles over cells, which leads[4] to a generalized form of the van der Waals equation of state, van Kampen obtained the distribution over cells by minimizing the free energy. His method leads to the modified equation of state (2), which implies a first-order phase transition. When $\rho < \rho_v$ or $\rho_l < \rho$, van Kampen's method indeed gives a uniform distribution over cells, but when $\rho_v < \rho < \rho_l$ it leads

* Work supported by the U. S. Air Force Office of Scientific Research at Yeshiva University, Grant No. 508-64.
† Present address: Imperial College, London, S.W. 7, England.
[1] J. van der Waals, thesis (Leiden, 1873) (cited by Kac, Uhlenbeck, and Hemmer, Ref. 6 of this paper).
[2] J. C. Maxwell, Scientific Papers (Dover Reprint, New York), p. 425.
[3] N. G. van Kampen, Phys. Rev. 135, A362 (1964).
[4] O. Ornstein, thesis (Leiden, 1908) (cited by N. G. van Kampen, Ref. 3 of this paper).

to the conclusion that this distribution is nonuniform, as it should be when two phases coexist.

Van Kampen's treatment is, however, not mathematically rigorous. In particular the conditions to be satisfied by the interaction are not specified, and various limiting processes are hinted at but not carried out explicitly. It is the purpose of the present paper to provide a rigorous treatment similar to van Kampen's and to extend it to a more general class of long-range potentials than the purely attractive potentials considered by van Kampen.

The intermolecular potential we consider has the form proposed by Kac[5] and investigated thoroughly for a one-dimensional system by Kac, Uhlenbeck, and Hemmer,[6]

$$v(\mathbf{r}) = q(\mathbf{r}) + w(\mathbf{r}, \gamma), \tag{1.3}$$

where r represents the separation of a pair of particles, $q(\mathbf{r})$ is a short-range potential, and $w(\mathbf{r}, \gamma)$ is a potential (we call it the *Kac potential*) whose range is proportional to the reciprocal of the parameter γ. Kac, Uhlenbeck, and Hemmer studied a one-dimensional system with

$$q(\mathbf{r}) \equiv \begin{cases} +\infty & \text{if } r < r_0, \\ 0 & \text{if } r > r_0, \end{cases} \tag{1.4}$$

$$w(\mathbf{r}, \gamma) \equiv \tfrac{1}{2}\alpha\gamma \exp(-\gamma r), \tag{1.5}$$

where r_0, γ, and $-\alpha$ are positive parameters. For finite γ they found no phase transition, but in the *van der Waals limit* $\gamma \to 0$ the equation of state approaches as a limit Maxwell's modification (2) of the van der Waals equation of state and thus does show a first-order phase transition. Unfortunately their method is very difficult to generalize to other potentials or to more than one dimension.

An important feature of the work of Kac, Uhlenbeck, and Hemmer was the use of the limit process $\gamma \to 0$ to provide a clear distinction between the short (finite) range of the contribution $q(\mathbf{r})$ to $v(\mathbf{r})$ and the long (infinite as $\gamma \to 0$) range of the contribution $w(\mathbf{r}, \gamma)$. It is this feature which we shall exploit here; but unlike Kac, Uhlenbeck, and Hemmer, we do not restrict the system to one dimension nor the potential to the form defined by (4) and (5). Instead of the special form (5) for $w(\mathbf{r}, \gamma)$, we use

$$w(\mathbf{r}, \gamma) = \gamma^{\nu}\varphi(\gamma\mathbf{r}), \tag{1.6}$$

where ν is the number of dimensions of the space

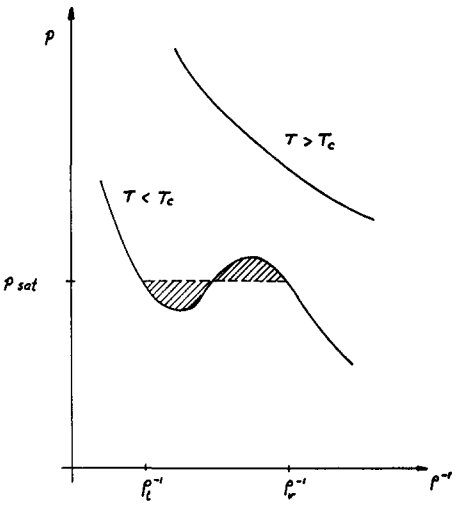

FIG. 1. Typical isotherms for the van der Waals equation of state (solid lines) and Maxwell's modification (dotted line). The shaded areas are equal.

considered; this reduces to (5) if $\nu = 1$ and $\varphi(\mathbf{x}) = \frac{1}{2}\alpha e^{-x}$. If the function $\varphi(\mathbf{x})$ is bounded in a neighborhood of the origin, say in $x < \delta$, then the Kac potential (6) has the property

$$|w(\mathbf{r}, \gamma)| < \gamma^{\nu} \operatorname*{Max}_{x<\delta} |\varphi(\mathbf{x})| \quad \text{if} \quad \gamma < \delta/r \tag{1.7}$$

so that

$$\lim_{\gamma \to 0} w(\mathbf{r}, \gamma) = 0 \quad \text{for all} \quad \mathbf{r}. \tag{1.8}$$

At the same time the integral of the Kac potential over all ν-dimensional space,

$$\int w(\mathbf{r}, \gamma)\, d\mathbf{r} = \int \varphi(\mathbf{x})\, d\mathbf{x} \equiv \alpha, \tag{1.9}$$

is independent of γ. This fact is compatible with (1.8) because the limit operation $\gamma \to 0$ does not commute with the one associated with the infinite region of integration.

The basis of our method is to obtain upper and lower bounds on the free energy

$$A(N, \Omega, \gamma) \equiv -kT \log Z(N, \Omega, \gamma), \tag{1.10}$$

where $Z(N, \Omega, \gamma)$ is the classical partition function for N particles at temperature $T \equiv 1/k\beta$ in a ν-dimensional cube Ω, defined by

$$Z(N, \Omega, \gamma) \equiv (1/N!)(mkT/2\pi\hbar^2)^{\nu N/2} \times \int_{\Omega} \cdots \int_{\Omega} e^{-\beta V}\, d\mathbf{x}_1 \cdots d\mathbf{x}_N \tag{1.11}$$

and

$$V \equiv \sum_{i \le j \le N} v(\mathbf{x}_i - \mathbf{x}_j) \tag{1.12}$$

[5] M. Kac, Phys. Fluids 2, 8 (1959).
[6] M. Kac, G. E. Uhlenbeck, and P. Hemmer, J. Math. Phys. 4, 216 (1963).

with $v(\mathbf{r})$ defined in (3). The upper and lower bounds are obtained by dividing the cube Ω into M congruent smaller cubes $\omega_1 \cdots \omega_M$ and using estimates of the interactions across cell boundaries to relate the free energy of the cube Ω to the sum of the free energies of the cubes $\omega_1 \cdots \omega_M$.

From these upper and lower bounds, the equation of state in the van der Waals limit is calculated by means of a succession of limit operations. First the thermodynamic free energy is calculated from $A(N, \Omega, \gamma)$ by taking the *thermodynamic limit*. The simplest way of taking this limit is to double the side of the cube Ω repeatedly, adjusting N at each step to the value $\rho\Omega$ where ρ, the density, is a constant, and the symbol Ω is used to represent the volume of the cube as well as the cube itself. The thermodynamic free-energy density, a function of ρ and γ, is then defined as

$$a(\rho, \gamma) \equiv \lim_{\Omega\to\infty} A(\rho\Omega, \Omega, \gamma)/\Omega \tag{1.13}$$

where $A(N, \Omega, \gamma)$ may be defined for nonintegral N by linear interpolation.[7]

The next operation is to take the van der Waals limit $\gamma \to 0$. This gives the *van der Waals free-energy density*

$$\begin{aligned} a(\rho, 0+) &\equiv \lim_{\gamma\to 0} a(\rho, \gamma) \\ &= \lim_{\gamma\to 0} \lim_{\Omega\to\infty} A(\rho\Omega, \Omega, \gamma)/\Omega. \end{aligned} \tag{1.14}$$

It is important to take these two limits in the right order. Taking the limit $\Omega \to \infty$ first, as in (14), means that the range of the Kac potential, although very large, is much less than the size of the container. If they are taken in the opposite order, then the Kac potential has a range much larger than the size of the container and in consequence of (8) its effect disappears. In fact, by applying (8) to (10) and (11) we obtain

$$\lim_{\gamma\to 0} A(N, \Omega, \gamma) = A^0(N, \Omega), \tag{1.15}$$

where A^0 is the free energy for N particles of the *reference system* in a cube Ω. By reference system we mean one for which the interaction potential is $q(\mathbf{r})$ instead of $q(\mathbf{r}) + w(\mathbf{r}, \gamma)$. Taking the thermodynamic limit of (15) we obtain, in constrast to (14), the formula

$$\lim_{\Omega\to\infty} \lim_{\gamma\to 0} A(\rho\Omega, \Omega, \gamma)/\Omega = a^0(\rho) \tag{1.16}$$

where

$$a^0(\rho) \equiv \lim_{\Omega\to\infty} A^0(\rho\Omega, \Omega)/\Omega \tag{1.17}$$

is the thermodynamic free-energy density of the reference system at particle density ρ, and is not in general equal to $a(\rho, 0+)$.

The upper and lower bounds on $A(N, \Omega, \gamma)$ which we shall derive lead to upper and lower bounds on $a(\rho, 0+)$. These bounds depend on the volume ω of the cells ω_i used. The bounds can be simplified by a third limit process $\omega \to \infty$, after which, in favorable cases, the upper and lower bounds are equal so that $a(\rho, 0+)$ can be calculated. This triple limit process corresponds to the following relationships:

$$r_0 \ll \omega^{1/\nu} \ll \gamma^{-1} \ll \Omega^{1/\nu} \tag{1.18}$$

among the four characteristic lengths of our calculation: the range of the short-range potential, the size of the cells, the range of the Kac potential, and the size of the container.

Once $a(\rho, 0+)$ has been found, the equation of state can be calculated by differentiation. The main result of our work is to show rigorously that under suitable conditions the equation of state is indeed given by Maxwell's rule (2) applied to the generalized van der Waals isotherm

$$\tilde{p}_{vdw}(\rho, T) = p^0(\rho, T) + \tfrac{1}{2}\alpha\rho^2 \tag{1.19}$$

where $p^0(\rho, T)$ is the pressure of the reference system, calculated by differentiation from $a^0(\rho)$. This result can be used to investigate the conditions under which the system will have a first-order phase transition in the van der Waals limit. A further result is to show, by studying the two-particle distribution function, that if there is such a first-order phase transition then two phases of different densities are present during the transition.

For rigorous arguments it is necessary to impose conditions on the short- and long-range potentials. We shall assume that the short-range potential satisfies

$$q(\mathbf{r}) = +\infty \quad \text{for} \quad r < r_0, \tag{1.20a}$$

$$|q(\mathbf{r})| < D_2 r^{-\nu-\epsilon} \quad \text{for} \quad r_0 < r, \tag{1.20b}$$

where r_0 (the hard-core diameter), D_2, and ϵ are positive constants; and we shall also assume that the shape function $\varphi(\mathbf{x})$ of the Kac potential satisfies

$$|\varphi(\mathbf{r})| < D_3 r^{-\nu-\epsilon} \quad \text{for all} \quad r, \tag{1.21a}$$

$$\varphi(\mathbf{r}) \quad \text{is continuous at} \quad r = 0, \tag{1.21b}$$

$$\int \varphi(\mathbf{r})\, d\mathbf{r} \text{ exists as a Riemann integral.} \tag{1.21c}$$

Further, when in Sec. V the function $\varphi(\mathbf{r})$ is ex-

[7] M. Fisher, Arch. Rat. Mech. Anals. 17, 377 (1964).

pressed as the sum of two parts satisfying the conditions (5.2), both parts are required to satisfy (1.21). The extent to which the conditions (1.20) and (1.21) can be weakened will be discussed in Sec. IX, along with the possibilities for generalizing this work to quantum systems and to lattice gases.

II. UPPER BOUND ON THE FREE ENERGY

Finding an upper bound on the free energy is equivalent to finding a lower bound on the partition function. To obtain a lower bound, we divide the cube Ω into M smaller cubical regions $\omega_1 \cdots \omega_M$, each of side $(s + t)$ where s and t are positive lengths such that $(s + t)$ is a submultiple of the side of Ω (see Fig. 2). Since the small cubes completely fill Ω its volume, which we also call Ω, is given by

$$\Omega = M(s + t)^\nu. \tag{2.1}$$

For each ω_i, let ω_i' be the cube of side s consisting of all points within ω_i whose distance from the boundary of ω_i is at least $\frac{1}{2}t$. A lower bound on $Z(N, \Omega, \gamma)$ can be obtained by selecting any set of integers $N_1, N_2, \cdots N_M$ which add up to N, and considering only the contribution to the integral in (1.11) from configurations where there are N_1 particles in the cell ω_1', N_2 in ω_2', and so on. The quantity so calculated we denote by $\tilde{Z}(N_1, N_2 \cdots N_M)$. There are $N!/(N_1!N_2! \cdots N_M!)$ ways of choosing the N_1 particles to go in the cell ω_1', N_2 to go in ω_2', etc., and since the particles are identical all these different ways give the same contribution to the integral. Multiplying a typical contribution by $N!/\prod N_i!$ we thus obtain

$$\begin{aligned} Z &\geq \tilde{Z}(N_1, N_2 \cdots N_M) \\ &= \prod_i [(1/N_i!)(m/2\pi\hbar^2\beta)^{\nu N_i/2}] \\ &\quad \times \int_{\omega'_1} \cdots \int_{\omega'_M} e^{-\beta V} \, d\mathbf{x}_1 \cdots d\mathbf{x}_N, \end{aligned} \tag{2.2}$$

where the first N_1 of the N ν-fold integrations are taken over the region ω_1', the next N_2 over ω_2', and so on.

To obtain a lower bound on the integral in (2.2) we write

$$V = V' + V'' \tag{2.3}$$

where V' is the contribution to the total potential energy from pairs of particles that are both in the same cell, and V'' is the contribution from pairs that are in different cells. If $V''_{\max}$ is some upper bound on V'', then (2) implies

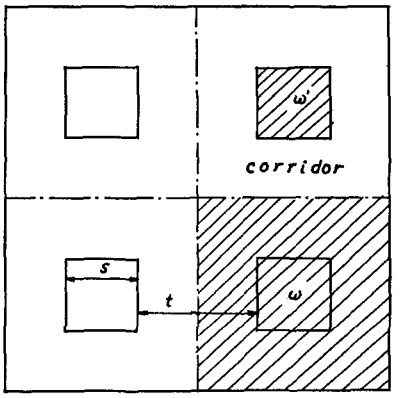

FIG. 2. Division of Ω into cells.

$$\begin{aligned} \tilde{Z} &\geq \prod_i [(1/N_i!)(m/2\pi\hbar^2\beta)^{\nu N_i/2}] \\ &\quad \times \int_{\omega'_1} \cdots \int_{\omega'_M} e^{-\beta V' - \beta V''_{\max}} \, d\mathbf{x}_1 \cdots d\mathbf{x}_N \\ &= [\prod_i Z(N_i, \omega', \gamma)] e^{-\beta V''_{\max}}, \end{aligned} \tag{2.4}$$

where $Z(N_i, \omega', \gamma)$ is the partition function for N_i particles in a cube ω', of side s.

To obtain a lower bound on the exponential factor in (4) we consider the short- and long-range contributions separately, writing

$$V'' = Q'' + W''. \tag{2.5}$$

A convenient upper bound for W'', the long-range contribution to V'', is

$$W'' \leq \sum_{i<j} N_i N_j w_{\max}(\mathbf{k}_{ij}) \tag{2.6}$$

where

$$\begin{aligned} w_{\max}(\mathbf{k}_{ij}) &\equiv \operatorname*{Max}_{\substack{\mathbf{x}\in\omega_i \\ \mathbf{y}\in\omega_j}} w(\mathbf{x} - \mathbf{y}, \gamma) \\ &= \operatorname*{Max}_{\mathbf{r}\in\omega_0} w(\mathbf{k}_{ij} + 2\mathbf{r}, \gamma) \end{aligned} \tag{2.7}$$

where ω_0 is a cube of side $(s + t)$ centered at the origin, and $\mathbf{k}_{ij}$ is the vector from the center of ω_i to that of ω_j. For an upper bound on Q'', the short-range contribution to V'', we may use the same method, in conjunction with the condition (1.20), and obtain, if $t > r_0$, the upper bound

$$\begin{aligned} Q'' &\leq D_2 \sum_{i<j} N_i N_j (r_{ij,\min})^{-\nu-\epsilon} \\ &= \tfrac{1}{2} D_2 \sum_i N_i \sum_j{}' N_j (r_{ij,\min})^{-\nu-\epsilon}, \end{aligned} \tag{2.8}$$

where $r_{ij,\min}$ is the least distance from ω_i' to ω_j' and $\sum_j'$ means a sum with the $j = i$ term omitted. To estimate this last sum we may group the cubes ω_j' into shells centered on ω_i'. The first shell contains

$3^\nu - 1$ cubes, all points of which are a distance at least t from all points of ω_i', the next shell contains $5^\nu - 3^\nu$ cubes, each one at least $2t + s$ from ω_i'; and so on. Writing also $N_{\max}$ for the maximum value of N_i, we obtain

$$Q'' \leq \tfrac{1}{2}D_2 \sum_i N_i \sum_{n=1}^{\infty} N_{\max}[(2n+1)^\nu - (2n-1)^\nu]/(nt + (n-1)s)^{\nu+\epsilon}$$
$$\leq \tfrac{1}{2}D_2 N N_{\max} J_\nu t^{-\nu-\epsilon} \tag{2.9}$$

where

$$J_\nu \equiv \sum_{n=1}^{\infty} n^{-\nu-\epsilon}[(2n+1)^\nu - (2n-1)^\nu] < \infty. \tag{2.10}$$

Substituting (9) and (6) into (5), (4), and (2), we obtain

$$A(N, \Omega, \gamma) \leq \sum_i A(N_i, \omega', \gamma) + \sum_{i<j} N_i N_j w_{\max}(\mathbf{k}_{ij}) + \tfrac{1}{2}D_2 J_\nu N N_{\max} t^{-\nu-\epsilon}, \tag{2.11}$$

where $A(N, \Omega, \gamma)$ is defined in (1.10). The upper bound (11) holds for any choice of $N_1, N_2, \cdots N_M$ whose sum is N, and in particular for the choice which minimizes the right side.

To obtain an upper bound on the thermodynamic free energy in the van der Waals limit, we apply the triple limit process described in Sec. 1 to the inequality (11). This is simplest for the case where all of $N_1, N_2 \cdots N_M$ are equal:

$$N_1 = N_2 = \cdots = N_M = \rho(s+t)^\nu. \tag{2.12}$$

Replacing $\sum_{i<j}$ in (2.11) by $\frac{1}{2}\sum_i \sum_j'$ and dividing both sides by Ω we obtain

$$A(\rho\Omega, \Omega, \gamma)/\Omega \leq [M/\Omega]A(\rho(s+t)^\nu, \omega', \gamma) + \tfrac{1}{2}[\rho^2(s+t)^{2\nu}/\Omega] \sum_i \sum_j' w_{\max}(\mathbf{k}_{ij}) + \tfrac{1}{2}D_2 J_\nu \rho^2 (s+t)^\nu t^{-\nu-\epsilon}. \tag{2.13}$$

To take the limit $\Omega \to \infty$ we require the following lemma:

$$\lim_{\Omega\to\infty} 1/M \sum_i \sum_j' w_{\max}(\mathbf{k}_{ij}) = \sum_{\mathbf{k}}' w_{\max}(\mathbf{k}) \tag{2.14}$$

where $\sum_{\mathbf{k}}'$ is an infinite sum over the complete infinite lattice of possible vectors $\mathbf{k}_{ij}$ except $\mathbf{k} = 0$. To prove (14) it is sufficient to show that the difference between its two sides vanishes. This difference can be written as the limiting value, for large M, of the expression

$$1/M \sum_i \sum_j'' w_{\max}(\mathbf{k}_{ij}) \tag{2.15}$$

with $\sum_j''$ meaning the sum over an infinite network of cells outside Ω, continuing the pattern established by the cells inside Ω, but not including the cells inside Ω. Let δ be any small positive number. By (1.6) and (1.21a), the infinite sum on the right of (14) converges absolutely, and therefore we can find a number σ such that

$$\sum_{\substack{\mathbf{k} \\ |\mathbf{k}|>\sigma}}' |w_{\max}(\mathbf{k})| < \tfrac{1}{2}\delta. \tag{2.16}$$

The sum over i in (15) may be divided into two parts, the first part including all those cells whose distance from the boundary of Ω is greater than σ, and the second, those for which it is not. For each value of i in the first part of the i-summation, the sum over j covers only a subset of the values of k covered in the sum (15), and hence this part of the i-summation contributes at most $\frac{1}{2}\delta$ to the expression (16). In the second part of the i-summation, the number of terms is at most $[\Omega - (\Omega^{1/\nu} - 2\sigma)^\nu]/(s+t)^\nu$, and this number does not exceed $2\nu\sigma\,\Omega^{1-1/\nu}(s+t)^{-\nu}$ because

$$x^\nu - y^\nu \leq \nu x^{\nu-1}(x-y) \quad \text{if} \quad x \geq y \geq 0. \tag{2.17}$$

Consequently, by (1), the second part of the i-summation contributes to the expression (15) an amount not exceeding $2\nu\sigma\,\Omega^{-1/\nu}\sum_{\mathbf{k}}' |w_{\max}(\mathbf{k})|$, which can be made $< \frac{1}{2}\delta$ by making Ω large enough. Thus the complete expression (15) is less than δ for all large Ω, and since δ is arbitrarily small, the result (14) follows from the definition of a limit, Q.E.D.

Taking the thermodynamic limit of (13) with the help of (1) and (14) we obtain

$$a(\rho, \gamma) \leq A(\rho(s+t)^\nu, \omega', \gamma)/(s+t)^\nu + \tfrac{1}{2}\rho^2(s+t)^\nu \sum_{\mathbf{k}}' w_{\max}(\mathbf{k}) + \tfrac{1}{2}D_2 J_\nu \rho^2 (s+t)^\nu / t^{\nu+\epsilon}. \tag{2.18}$$

The second limiting process is the van der Waals limit $\gamma \to 0$. To evaluate the van der Waals limit of the middle term on the right in (18) we substitute from (7) and (1.6), obtaining

$$(s+t)^\nu \sum_{\mathbf{k}}' w_{\max}(\mathbf{k}) = 2^{-\nu} \sum_{\mathbf{n}}' \Delta \operatorname*{Max}_{\mathbf{x}\in\Delta_{\mathbf{n}}} \varphi(\mathbf{x}) \tag{2.19}$$

where the sum goes over all nonzero vectors $\mathbf{n}$ with integer components. $\Delta_{\mathbf{n}}$ stands for the cube of side $2\gamma(s+t)$ centered at the point $\gamma(s+t)\mathbf{n}$ with its sides parallel to those of Ω, and $\Delta \equiv 2^\nu\gamma^\nu(s+t)^\nu$ is the volume of one of the cubes $\Delta_{\mathbf{n}}$. The network of overlapping cubes $\Delta_{\mathbf{n}}$ can be disentangled into 2^ν separate nonoverlapping networks, each of them just filling ν-dimensional space (except for one net-

work where the cube centered on the origin is missing). In the limit $\gamma \to 0$ these networks become infinitesimally fine and, by Riemann's definition of an integral,[8] the contribution of each of them to the second sum in (19) tends to $\int \varphi(\mathbf{x})\, d\mathbf{x}$; the van der Waals limit of (2.19) is therefore

$$(s+t)^\nu \lim_{\gamma\to 0} \sum_{\mathbf{k}}{}' w_{\max}(\mathbf{k}) = \int \varphi(\mathbf{x})\, d\mathbf{x} \equiv \alpha. \tag{2.20}$$

Using this result, and also (1.15), in (18) we obtain

$$a(\rho, 0+) \leq A^0(\rho(s+t)^\nu, \omega')/(s+t)^\nu + \tfrac{1}{2}\rho^2\alpha + \tfrac{1}{2}D_2 J_\nu \rho^2 (s+t)^\nu / t'^{\nu+\epsilon}. \tag{2.21}$$

The final limiting process is to make the cell size infinite by making $s \to \infty$. As Fisher has shown[7] the last term in (21), which represents the short-range interactions between cells, can be eliminated if we make t depend on s in such a way that

$$t/s \to 0 \quad \text{and} \quad s^\nu/t'^{\nu+\epsilon} \to 0 \quad \text{as} \quad s \to \infty. \tag{2.22}$$

This can be done, for example, by making $t \propto s^\eta$ where η is a constant satisfying

$$\nu/(\nu+\epsilon) < \eta < 1. \tag{2.23}$$

Applying this limit on both sides of (21) and using the continuity of $a^0(\rho)$, we obtain, since the volume of the cell ω' is s^ν,

$$a(\rho, 0+) \leq a^0(\rho) + \tfrac{1}{2}\rho^2\alpha. \tag{2.24}$$

That is, in the van der Waals limit the increase in free-energy density brought about by introducing a Kac potential into a system with short-range forces cannot exceed the increase that one would calculate by treating the particles as a smoothed-out uniform medium.

The result (2.24) can be strengthened in some cases by using the fact[7,9] that $a(\rho, \gamma)$ is a convex function[10] of ρ, so that $a(\rho, 0+)$, being the limit of a sequence of convex functions, is also convex. Because of this (2.24) implies

$$a(\rho, 0+) \leq \text{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\} \tag{2.25}$$

where CE $\{f(\rho)\}$ means, for any function $f(\rho)$, the *convex envelope* of that function, defined as the maximal convex function not exceeding f:

$$\text{CE}\,\{f(\rho)\} \equiv \text{Max}\,\phi(\rho) \quad \text{for each value of } \rho \left\{ \begin{array}{l} \phi(\cdot) \\ \phi(\cdot) \quad \text{is convex} \\ \phi(\xi) \leq f(\xi) \quad \text{for all} \quad \xi \end{array} \right\}. \tag{2.26}$$

Since the maximum of any family of convex functions is[11] itself convex, the function CE $\{f(\rho)\}$ is convex. If $f(\rho)$ is convex, then CE $\{f(\rho)\}$ and $f(\rho)$ coincide; otherwise the graph of CE $\{f(\rho)\}$ consists partly of convex segments of the graph of $f(\rho)$ and partly of segments of double tangents of this graph (Fig. 3). The construction of CE $\{f(\rho)\}$ from $f(\rho)$ is sometimes called the *double tangent construction.*

It will be shown in Sec. VI that the replacement of $a^0(\rho) + \frac{1}{2}\alpha\rho^2$ by its convex envelope is equivalent to the replacement of van der Waals' equation of state by Maxwell's modification (1.2).

III. LOWER BOUNDS ON THE FREE ENERGY: NONNEGATIVE-DEFINITE KAC POTENTIALS

A lower bound on the free energy is most easily found when the shape function $\varphi(\mathbf{x})$ defining the Kac potential (1.6) is *nonnegative definite*: that is to say, when its ν-dimensional Fourier transform

$$\Phi(\mathbf{p}) \equiv \int \varphi(\mathbf{x}) \exp(2\pi i \mathbf{p}\cdot\mathbf{x})\, d\mathbf{x} \tag{3.1}$$

is nonnegative. In this case it is possible to find a lower bound $W_{\min}$ on the long-range contribution

$$W \equiv \sum_{i<j} \gamma^\nu \varphi[\gamma(\mathbf{x}_i - \mathbf{x}_j)] = \tfrac{1}{2}\sum_i \sum_j \gamma^\nu \varphi[\gamma(\mathbf{x}_i - \mathbf{x}_j)] - \tfrac{1}{2}N\gamma^\nu\varphi(0) \tag{3.2}$$

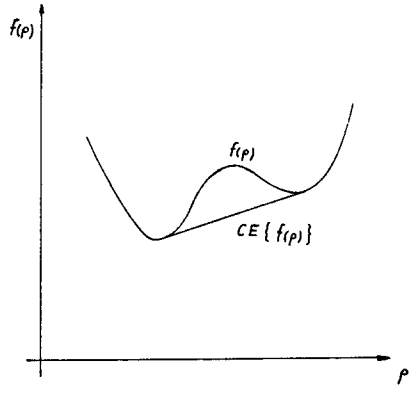

FIG. 3. A function $f(\rho)$ and its convex envelope.

[8] W. Rogosinski, *Volume and Integral* (Oliver and Boyd, London, 1952), Theorem 58.

[9] D. Ruelle, Helv. Phys. Acta **36**, 183 (1963).

[10] G. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge University Press, London and New York, 1959), Chap. 3.

[11] Proof: Since CE$\{f(\rho)\} \geq \phi(\rho)$ both ends of any chord of the graph CE$\{f(\rho)\}$ lie above or on the ends of the corresponding (same end ordinates) chord of any $\phi(\rho)$ and hence above or on the graph of $\phi(\rho)$ itself; consequently the chord of CE$\{f(\rho)\}$ also lies above or on the curve Max $\phi(\rho) \equiv$ CE$\{f(\rho)\}$.

to the total potential energy V, and by substituting such a lower bound into the basic definitions (1.10) and (1.11) we obtain

$$A(N, \Omega, \gamma) \geq A^0(N, \Omega) + W_{\min}. \tag{3.3}$$

To find a suitable $W_{\min}$ we substitute the Fourier inversion formula of (1) into (2), obtaining

$$W = -\tfrac{1}{2}N\gamma^\nu\varphi(0) + \tfrac{1}{2}\gamma^\nu \int d\mathbf{p}\, \Phi(\mathbf{p}) \left| \sum_{j=1}^{N} \exp(2\pi i\gamma\mathbf{p}\cdot\mathbf{x}_j) \right|^2 \tag{3.4}$$

for almost all configurations $\mathbf{x}_1 \cdots \mathbf{x}_N$. The "exceptional" configurations for which formulas such as (3.4) may fail will be ignored since they form a set of zero measure and therefore do not affect the configurational integral in (1.11). Let δ be any positive number. Since $\Phi(\mathbf{p})$ is[12] a continuous function of $\mathbf{p}$, and $\Phi(0) \geq 0$, we can find a positive number p_0 (depending on δ) such that

$$\Phi(\mathbf{p}) \geq (1 - \delta)\Phi(0) \tag{3.5}$$

whenever p is inside a cube of side $2p_0$ centered at the origin. Since $\Phi(\mathbf{p}) \geq 0$ when $\mathbf{p}$ is outside this cube, it follows from (4) that

$$W + \tfrac{1}{2}N\gamma^\nu\varphi(0) \geq \tfrac{1}{2}(1 - \delta)\Phi(0)\gamma^\nu \times \int d\mathbf{p} \left| \Psi(\mathbf{p}) \sum_{j=1}^{N} \exp(2\pi i\gamma\mathbf{p}\cdot\mathbf{x}) \right|^2 \tag{3.6}$$

where

$$\Psi(\mathbf{p}) \equiv \begin{cases} \prod_{n=1}^{\nu} (1 - |p_n|/p_0) & \text{if } |p_n| < p_0 \text{ for all } n, \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

and $p_1 \cdots p_\nu$ are the components of the vector $\mathbf{p}$. Applying Parseval's theorem to (6) we obtain

$$W + \tfrac{1}{2}N\gamma^\nu\varphi(0) \geq (1 - \delta)\Phi(0)\tfrac{1}{2}\gamma^\nu \times \int d(\gamma\mathbf{x}) \left\{ \sum_{j=1}^{n} \psi[\gamma(\mathbf{x} - \mathbf{x}_j)] \right\}^2 \tag{3.8}$$

where

$$\psi(\mathbf{y}) \equiv \int \Psi(\mathbf{p}) \exp(-2\pi i\mathbf{p}\cdot\mathbf{y})\, d\mathbf{p} = \prod_{n=1}^{\nu} \frac{\sin^2(\pi y_n p_0)}{\pi^2 y_n^2 p_0} \tag{3.9}$$

and $y_1 \cdots y_\nu$ are the components of the vector $\mathbf{y}$.

A lower bound on the integral in (8) can be obtained by confining the integration to a cube Ω^* concentric with Ω and similarly oriented but having a side of length $\Omega^{1/\nu} + l$ where l is an arbitrary positive quantity. Applying the Schwartz inequality to the resulting integral and writing Ω^* for the volume of the cube Ω^* we obtain

$$W + \tfrac{1}{2}N\gamma^\nu\varphi(0) \geq \tfrac{1}{2}(1 - \delta)\Phi(0) \left[\sum_{j=1}^{N} \int_{\Omega^*} \psi[\gamma(\mathbf{x} - \mathbf{x}_j)]\, d(\gamma\mathbf{x}) \right]^2 / \Omega^* \geq \tfrac{1}{2}(1 - \delta)\Phi(0) \left[N \int_{v<\gamma l} \psi(\mathbf{y})\, d\mathbf{y} \right]^2 / \Omega^* \tag{3.10}$$

since the integrand is nonnegative and every point $\mathbf{x}$ inside the sphere $|\mathbf{x} - \mathbf{x}_j| < l$ is also within Ω^*.

Let us choose l so that $l \to \infty$ and $l/\Omega^{1/\nu} \to 0$ in the thermodynamic limit; for example, we may take $l \propto \Omega^{1/2\nu}$. Then in this limit we have $\Omega^*/\Omega \to 1$ and $\int_{v<\gamma l} \psi(\mathbf{y})\, d\mathbf{y} \to 1$, and the lower bound on W/Ω implied by (10) has the thermodynamic limit $-\tfrac{1}{2}\rho\gamma^\nu\varphi(0) + \tfrac{1}{2}(1 - \delta)\Phi(0)\rho^2$.

Since this result holds for arbitrarily small δ, it is also true in the limit $\delta \to 0$; that is to say, there exists a sequence of lower bounds on W, call them $W_{\min}$, with the property[13]

$$\lim_{\Omega\to\infty} W_{\min}/\Omega = -\tfrac{1}{2}\rho\gamma^\nu\varphi(0) + \tfrac{1}{2}\Phi(0)\rho^2 = -\tfrac{1}{2}\rho\gamma^\nu\varphi(0) + \tfrac{1}{2}\alpha\rho^2 \tag{3.11}$$

in the notation of (1.8).

This result enables us to take the thermodynamic limit of (3), obtaining

$$a(\rho, \gamma) \geq a^0(\rho) - \tfrac{1}{2}\rho\gamma^\nu\varphi(0) + \tfrac{1}{2}\alpha\rho^2 \tag{3.12}$$

for nonnegative-definite potentials. In the van der Waals limit this formula reduces to

$$a(\rho, 0+) \geq a^0(\rho) + \tfrac{1}{2}\alpha\rho^2 \tag{3.13}$$

which when combined with (2.24) gives

$$a(\rho, 0+) = a^0(\rho) + \tfrac{1}{2}\alpha\rho^2 \tag{3.14}$$

for nonnegative-definite Kac potentials.

IV. LOWER BOUND ON THE FREE ENERGY: NONPOSITIVE KAC POTENTIALS

Another case where a lower bound on the free energy can be found fairly easily arises when the Kac potential is nonpositive; that is when

$$\varphi(\mathbf{x}) \leq 0 \quad \text{for all} \quad \mathbf{x}. \tag{4.1}$$

In this case it is again necessary to divide the cube Ω into cubical cells $\omega_1 \cdots \omega_M$, each of side $s + t$

[12] This follows from the existence of $\int |\varphi(\mathbf{x})|\, d\mathbf{x}$.

[13] A proof of (3.11) for systems with periodic boundary conditions was given by E. Lieb, Phys. Rev. 130, 2518 (1963). We are indebted to E. Newman and M. Austin for advice on the proof presented here.

(see Fig. 2). The definition (1.11) of the partition function may be written

$$Z(N, \Omega, \gamma) = \sum_{N_1 \cdots N_M} Z(N_1, \cdots, N_M) \qquad (4.2)$$

where the sum is over all sets of M nonnegative integers adding up to N and $Z(N_1, \cdots, N_M)$ means the contribution to (1.11) from configurations with exactly N_i particles in cell $\omega_i (i = 1, 2 \cdots M)$. Since there are $(N + M - 1)!/N!(M - 1)!$ terms in the sum it has the upper bound

$$Z(N, \Omega, \gamma) \leq [(N + M - 1)!/N!\,(M - 1)!] \times \operatorname*{Max}_{N_1 \cdots N_M} Z(N_1, \cdots, N_M), \qquad (4.3)$$

the maximum being taken over all sets of nonnegative integers $N_1 \cdots N_M$ which add up to N. The combinatorial argument which led to (2.2) gives, when applied to $Z(N_1, \cdots, N_M)$, the formula

$$Z(N_1, \cdots, N_M) = \prod_i [(1/N_i!)(m/2\pi\hbar^2\beta)^{\nu N_i/2}] \times \int_{\omega_1} \cdots \int_{\omega_M} e^{-\beta V}\, d\mathbf{x}_1 \cdots d\mathbf{x}_N \qquad (4.4)$$

where the first N_1 of the N integrations are over the cell ω_1, [not ω_i' as in (2.2)] the next N_2 over ω_2, and so on.

To obtain an upper bound on the integral in (4) we separate the potential energy V into three parts:

$$V = Q' + \tilde{Q} + W \qquad (4.5)$$

where Q' is the contribution to V from short-range interactions between particles that are in the same cell, $\tilde{Q}$ the contribution from short-range interactions between particles that are in different cells, and W is the total contribution from long-range interactions. If $\tilde{Q}_{\min}$ and $W_{\min}$ are lower bounds on $\tilde{Q}$ and W, then (4) and (5) lead to the inequality, analogous to (2.4),

$$Z(N_1, \cdots, N_M) \leq \{\prod_i Z^0(N_i, \omega)\} \times \exp[-(\tilde{Q}_{\min} + W_{\min})/kT] \qquad (4.6)$$

where $Z^0(N_i, \omega)$ is the partition function for N_i particles of the reference system in a cube of side $s + t$.

To find a suitable lower bound on $\tilde{Q}$, we split it into two parts:

$$\tilde{Q} = Q'' + Q''' \qquad (4.7)$$

where Q'' is, as in (2.5), the contribution to $\tilde{Q}$ from pairs of particles whose centers are both within the inner cells of side s denoted by ω_i' $(i = 1, 2 \cdots M)$ in Sec. 2, and Q''' is the contribution of pairs of particles, at least one of which is in the "corridor" consisting of points that are not in any of the ω_i' (see Fig. 2). The argument used to prove (2.9) proves at the same time that

$$Q'' \geq -\tfrac{1}{2} D_2 N N_{\max} J_\nu t^{-\nu-\epsilon}. \qquad (4.8)$$

The contribution Q''' has the lower bound

$$Q''' \geq -N_{\text{corr}} \Phi' \qquad (4.9)$$

where $-\Phi'$ is a lower bound on the interaction of a given particle with all its neighbors, whose existence is a consequence[14] of (1.20), and N_{corr} is the number of particles in the corridor. It can be shown[15] that the number of particles in a given region, multiplied by the reciprocal close-packing density ρ_c^{-1}, does not exceed the volume occupied by all points whose distance from the region is at most r_0. Hence, $N_{\max}$ and N_{corr} have (provided $s \geq 2r_0$,) the upper bounds

$$N_{\max}\rho_c^{-1} < (s + 2r_0)^\nu \qquad (4.10)$$

$$\begin{aligned} N_{\text{corr}}\rho_c^{-1} &\leq (\Omega^{1/\nu} + 2r_0)^\nu - M(s - 2r_0)^\nu \\ &= M\{(s + t + 2r_0 M^{-1/\nu})^\nu - (s - 2r_0)^\nu\} \\ &\leq \nu M(s + t + 2r_0 M^{-1/\nu})^{\nu-1} \\ &\quad \times [t + 2r_0(1 + M^{-1/\nu})] \end{aligned} \qquad (4.11)$$

by (2.1) and (2.17).

Substituting (10) and (11) into (7), (8), and (9), we obtain

$$\tilde{Q} \geq \tilde{Q}_{\min}, \qquad (4.12)$$

where

$$\begin{aligned} \tilde{Q}_{\min} = &-\tfrac{1}{2} D_2 N \rho_c (s + 2r_0)^\nu J_\nu t^{-\nu-\epsilon} \\ &- \rho_c \nu M(s + t + 2r_0 M^{-1/\nu})^{\nu-1} \\ &\times [t + 2r_0(1 + M^{-1/\nu})]\Phi'. \end{aligned} \qquad (4.13)$$

The details of this formula are unimportant; all that is important is its behavior under the triple limiting process described in the introduction. This behavior is

$$\begin{aligned} &\lim_{s\to\infty} \lim_{\gamma\to 0} \lim_{\Omega\to\infty} (\tilde{Q}_{\min}/\Omega) \\ &= \lim_{s\to\infty} \lim_{\gamma\to 0} \{-\tfrac{1}{2} D_2 \rho \rho_c (s + 2r_0)^\nu J_\nu t^{-\nu-\epsilon} \\ &\qquad - \nu(s + t)^{-1}(t + 2r_0)\Phi'\} \\ &= 0 \end{aligned} \qquad (4.14)$$

by (2.1) and (2.22).

[14] O. Penrose, J. Math. Phys. 4, 1312 (1963).
[15] O. Penrose, Phys. Letters 11, 224 (1964).

A simple lower bound on W, the total long-range interaction, is given by

$$W \geq \tfrac{1}{2} \sum_i \sum_j N_i N_j w_{\min}(\mathbf{k}_{ij}) \tag{4.15}$$

where

$$w_{\min}(\mathbf{k}_{ij}) \equiv \operatorname*{Min}_{\substack{\mathbf{x}\in\omega_i \\ \mathbf{y}\in\omega_j}} w(\mathbf{x}-\mathbf{y},\gamma) = \operatorname*{Min}_{\mathbf{r}\in\omega_0} w(\mathbf{k}_{ij}+2\mathbf{r},\gamma). \tag{4.16}$$

the γ-dependence of $w_{\min}(\mathbf{k})$ not being explicitly shown. These formulas are analogous to (2.6) and (2.7) but the $i = j$ terms are now included. Since $N_iN_j \leq \frac{1}{2}N_i^2 + \frac{1}{2}N_j^2$, and $w_{\min}(\mathbf{k}_{ij}) \leq 0$ by (1), we may deduce from (4.15) the inequality

$$W \geq -\tfrac{1}{2} \sum_i \sum_j (\tfrac{1}{2}N_i^2 + \tfrac{1}{2}N_j^2)\,|w_{\min}(\mathbf{k}_{ij})| = -\tfrac{1}{2} \sum_i N_i^2 \sum_j |w_{\min}(\mathbf{k}_{ij})|. \tag{4.17}$$

The sum over j may be extended, without destroying the validity of (17), to include the infinite network of cells continuing outside Ω the pattern established inside it by the cells $\omega_1 \cdots \omega_M$. In the notation of (2.14) the resulting inequality is $W \geq W_{\min}$ with

$$W_{\min} = \tfrac{1}{2} \sum_i N_i^2 \sum_{\mathbf{k}} w_{\min}(\mathbf{k}). \tag{4.18}$$

Substituting this into (6) and using (3) and (1.10), we obtain

$$A(N,\Omega,\gamma) \geq kT \log [N!\,(M-1)!/(N+M-1)!] + \operatorname*{Min}_{N_1\cdots N_M} \sum_{i=1}^{M} \{A^0(N_i,\omega) + \tfrac{1}{2}N_i^2 \sum_{\mathbf{k}} w_{\min}(\mathbf{k})\} + \tilde{Q}_{\min}. \tag{4.19}$$

The second term on the right can be simplified by means of a simple property of the convex envelope of a function, defined in (2.27). This property is

$$M^{-1} \sum_{i=1}^{M} f(N_i) \geq M^{-1} \sum_i \mathrm{CE}\,\{f(N_i)\} \geq \mathrm{CE}\,\{f(M^{-1}\sum_i N_i)\} \tag{4.20}$$

where the first inequality follows from the fact that CE $\{f\}$ is a lower bound on f, and the second from the fact that CE $\{f\}$ is convex.

Using (4.20) in (4.19) we obtain

$$A(N,\Omega,\gamma) \geq kT \log [N!\,(M-1)!/(N+M-1)!] + M\,\mathrm{CE}\,\{A^0(N/M,\omega) + \tfrac{1}{2}(N/M)^2 \sum_{\mathbf{k}} w_{\min}(\mathbf{k})\} + \tilde{Q}_{\min} \tag{4.21}$$

where CE $\{\ \}$ is the convex envelope of the quantity in braces regarded as a function of N/M.

To apply the triple limiting process described in Sec I we first divide (4.21) by Ω and take the thermodynamic limit $\Omega \to \infty$, using Stirling's approximation for the factorials. In the notation of (1.13) the result is

$$a(\rho,\gamma) \geq -kT\{(s+t)^{-\nu} \log [1+\rho(s+t)^{\nu}] + \rho \log [1+\rho^{-1}(s+t)^{-\nu}]\} + \mathrm{CE}\,\{(s+t)^{-\nu}A^0(\rho(s+t)^{\nu},\omega) + \tfrac{1}{2}\rho^2(s+t)^{\nu} \sum w_{\min}(\mathbf{k})\} + \lim_{\Omega\to\infty} \tilde{Q}_{\min}/\Omega \tag{4.22}$$

since

$$N/M = \rho\Omega/M = \rho(s+t)^{\nu}.$$

In order to carry through the other two limit processes we require the following lemma:

Let $f_n(\xi)$ be a sequence of functions converging uniformly on an interval to $f(\xi)$ as $n \to \infty$; then we have

$$\lim_{n\to\infty} \mathrm{CE}\,\{f_n(\xi)\} = \mathrm{CE}\,\{f(\xi)\}. \tag{4.23}$$

Proof: Given any $\delta > 0$ there exists an $\mathfrak{N}$ such that for $n \geq \mathfrak{N}$, $|f_n(\xi) - f(\xi)| < \delta$ for all ξ in the interval. We then have from the definition (2.26) of the convex envelope

$$\mathrm{CE}\,\{f_n(\xi)\} \geq \mathrm{CE}\,\{f(\xi)\} - \delta \quad \text{for all } \xi \text{ in the interval and } n \geq \mathfrak{N}$$

since CE $\{f(\xi)\} - \delta$ is a convex function which is $\leq f_n(\xi)$. Similarly

$$\mathrm{CE}\,\{f(\xi)\} \geq \mathrm{CE}\,\{f_n(\xi)\} - \delta \quad \text{for all } \xi \text{ in the interval and } n \geq \mathfrak{N}.$$

Hence

$$|\mathrm{CE}\,\{f_n(\xi)\} - \mathrm{CE}\,\{f(\xi)\}| \leq \delta \quad \text{for all } \xi \text{ in the interval and } n \geq \mathfrak{N}$$

which, since δ is arbitrary, proves the lemma. Q.E.D.

In our case we consider an interval, $0 \leq \rho \leq \rho_1 < \rho_c$ where ρ_1 is any density less than the close packing density ρ_c. Taking the limit $\gamma \to 0$ of (22) we obtain

$$a(\rho,0+) \geq -kT\{(s+t)^{-\nu} \log [1+\rho(s+t)^{\nu}] + \rho \log [1+\rho^{-1}(s+t)^{-\nu}]\} + \mathrm{CE}\,\{(s+t)^{-\nu}A^0[\rho(s+t)^{\nu},\omega] + \tfrac{1}{2}\alpha\rho^2\} + \lim_{\Omega\to\infty} \tilde{Q}_{\min}/\Omega \tag{4.24}$$

since, by an argument like that which led to (2.20),

$$(s + t)^\nu \lim_{\gamma \to 0} \sum_{\mathbf{k}} w_{\min}(\mathbf{k}) = \int \varphi(\mathbf{x})\, d\mathbf{x} \equiv \alpha$$

and the convergence is uniform because the difference between the sum and the integral is independent of ρ. Finally, taking the limit $s \to \infty$, where the convergence is again uniform (Sec. 7k of Ref. 7), and using (14), we obtain

$$a(\rho, 0+) \geq \mathrm{CE}\, \{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\}. \tag{4.25}$$

Combined with (2.25) this gives

$$a(\rho, 0+) = \mathrm{CE}\, \{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\} \tag{4.26}$$

for nonpositive Kac potentials. The result obtained by Kac, Uhlenbeck, and Hemmer[6] is an example of (4.26).

V. LOWER BOUND TO THE FREE ENERGY: MORE GENERAL KAC POTENTIALS

When the Kac potential is neither nonnegative-definite nor nonpositive, a lower bound can still be obtained, though the method is more complicated than before. We express $\Phi(\mathbf{p})$, the ν-dimensional Fourier transform of $\varphi(\mathbf{x})$ defined in (3.1) as the sum of two parts

$$\Phi(\mathbf{p}) = \Phi_+(\mathbf{p}) + \Phi_-(\mathbf{p}) \tag{5.1}$$

so chosen that

$$\left.\begin{aligned} &\Phi_+(\mathbf{p}) \geq 0 \quad \text{for all} \quad p, \\ &\Phi_-(0) = \min_{\mathbf{p}} \Phi_-(\mathbf{p}) = \min_{\mathbf{p}} \Phi(\mathbf{p}) \equiv \Phi_{\min}, \\ &\int \Phi_\pm(\mathbf{p})\, d\mathbf{p} < \infty, \\ &\Phi_\pm(\mathbf{p}) \quad \text{is continuous.} \end{aligned}\right\} \tag{5.2}$$

The inverse transforms of $\Phi_+(\mathbf{p})$ and $\Phi_-(\mathbf{p})$ will be denoted by $\varphi_+(\mathbf{x})$ and $\varphi_-(\mathbf{x})$ and we shall require also that $\varphi_+(\mathbf{x})$ and $\varphi_-(\mathbf{x})$ satisfy (1.21). These conditions can be satisfied (provided $d\Phi/dp_n > -\infty$ at $p = 0$) by taking

$$\Phi_+(\mathbf{p}) = \begin{cases} [\Phi(0) - \Phi_{\min}](1 - p/p')^n & \text{if } p \leq p' \\ 0 & \text{if } p \geq p' \end{cases} \tag{5.3}$$

where p' is the smallest value of p at which $\Phi(\mathbf{p})$ attains its minimum value $\Phi_{\min}$ and n is a positive integer (≥ 2) chosen sufficiently large to make $(1 - p/p')^n \leq [\Phi(\mathbf{p}) - \Phi_{\min}]/[\Phi(0) - \Phi_{\min}]$, for $p \leq p'$. When $\Phi_+(\mathbf{p})$ is given by (3) $\varphi_+(\mathbf{x})$ can be computed explicitly and shown to satisfy (1.21), and hence so will $\varphi_-(\mathbf{x})$. [Even when $d\Phi/dp = -\infty$ at $p = 0$ there is every reason to believe that the division of Φ into Φ_+ and Φ_- can be made in a way to satisfy all our conditions, provided $\varphi(\mathbf{x})$ satisfies (1.21).]

Since $\varphi_+(\mathbf{x})$ is a nonnegative-definite function its contribution to W, which may be written W_+, can be estimated by the method used in Sec. III. In analogy with (3.11), the result has the form $W_+ \geq W_{+,\min}$ where

$$\lim_{\Omega \to \infty} W_{+,\min}/\Omega = -\tfrac{1}{2}\rho\gamma^\nu\varphi_+(0) + \tfrac{1}{2}\Phi_+(0)\rho^2. \tag{5.4}$$

The other contribution W_- has a lower bound analogous to (4.15)

$$W_- \geq \tfrac{1}{2} \sum_i \sum_j N_i N_j w_{-,\min}(\mathbf{k}_{ij}) \tag{5.5}$$

where as in (4.16) $\mathbf{k}_i$ is the position vector from the center of the cell ω_i to that of ω_j, and

$$\begin{aligned} w_{-,\min}(\mathbf{k}_{ij}) &\equiv \min_{\substack{\mathbf{x} \in \omega_i \\ \mathbf{y} \in \omega_j}} w_-(\mathbf{x} - \mathbf{y}, \gamma) \\ &= \min_{\mathbf{r} \in \omega_0} w_-(\mathbf{k}_{ij} + 2\mathbf{r}, \gamma), \end{aligned} \tag{5.6}$$

where $w_-(\mathbf{x}, \gamma) = \gamma^\nu\varphi_-(\gamma\mathbf{x})$ and ω_0 is a cube of side $s + t$ centered at the origin.

The quadratic form in (5) can be diagonalized using a Fourier transformation. We define the function

$$W(\mathbf{p}) \equiv \sum_{\mathbf{k}} w_{-,\min}(\mathbf{k}) \exp(2\pi i \mathbf{p}\cdot\mathbf{k}) \tag{5.7}$$

where the sum goes over all the different values taken by the vector $\mathbf{k}_{ij}$ as both i and j range over the values 1, 2, $\cdots$ M. These values lie on a simple cubic lattice of spacing $s + t$ with a lattice point at the origin and are inside (not on the surface of) a cube of volume $2^\nu\Omega$ centered at the origin. The sum in (7) therefore covers just $[2\Omega^{1/\nu}(s + t)^{-1} - 1] = [2M^{1/\nu} - 1]^\nu \equiv M'$ lattice points. The function $W(\mathbf{p})$ is periodic in $\mathbf{p}$, the unit cell of the periodicity being a cube of side $(s + t)^{-1}$. Although the notation does not show it, $W(\mathbf{p})$ depends on $s + t$, γ, and Ω as well as on $\mathbf{p}$.

The inverse of (7) is

$$w_{-,\min}(\mathbf{k}) = (M')^{-1} \sum_{\mathbf{p}} W(\mathbf{p}) \exp(-2\pi i \mathbf{p}\cdot\mathbf{k}) \tag{5.8}$$

where the sum goes over M' values of $\mathbf{p}$ lying inside a cube of side $(s + t)^{-1}$ and on a lattice of spacing $(M')^{-1/\nu}(s + t)^{-1} = [2\Omega^{1/\nu} - (s + t)^{-1}$.] Substituting (8) into (5) and rearranging, we obtain

$$W_- \geq \tfrac{1}{2}(M')^{-1} \sum_{\mathbf{p}} W(\mathbf{p})\, |n(\mathbf{p})|^2 \tag{5.9}$$

where

$$n(\mathbf{p}) \equiv \sum_{i=1}^{M} N_i \exp(2\pi i \mathbf{p}\cdot\mathbf{k}_i) \tag{5.10}$$

and $\mathbf{k}_i$ is the position vector of the center of ω_i. For a lower bound on the sum (9) we may replace $W(\mathbf{p})$ by its minimum value, obtaining

$$\begin{aligned} W_- &\geq \tfrac{1}{2}(M')^{-1} \sum_{\mathbf{p}} |n(\mathbf{p})|^2 \operatorname*{Min}_{\mathbf{p}} W(\mathbf{p}) \\ &= \frac{1}{2} \sum_{i=1}^{M} N_i^2 \operatorname*{Min}_{\mathbf{p}} W(\mathbf{p}). \end{aligned} \tag{5.11}$$

Substituting this result into (4.6) and using (4.3) and (1.10) we obtain

$$\begin{aligned} A(N, \Omega, \gamma) &\geq kT \log[N!\,(M-1)!/(N+M-1)!] \\ &+ \operatorname*{Min}_{N_1 \cdots N_M} \sum_{i=1}^{M} \{A^0(N_i, \omega) \\ &+ \tfrac{1}{2} N_i^2 \operatorname*{Min}_{\mathbf{p}} W(\mathbf{p})\} + W_{+,\min} + \tilde{Q}_{\min} \end{aligned} \tag{5.12}$$

where $W_{+,\min}$ satisfies (4).

The inequality (12) is very similar to (4.19) and the effect on it of the triple limit process can be studied by the method used for (4.19). The result, analogous to (4.25), is

$$a(\rho, 0+) \geq \mathrm{CE}\{a^0(\rho) + \tfrac{1}{2}\alpha_- \rho^2\} + \tfrac{1}{2}\alpha_+ \rho^2, \tag{5.13}$$

where

$$\begin{aligned} \alpha_+ &\equiv (2/\rho^2) \lim_{s\to\infty} \lim_{\gamma\to 0} \lim_{\Omega\to\infty} W_{+,\min}/\Omega \\ &= \Phi_+(0) \end{aligned} \tag{5.14}$$

by (4), and

$$\alpha_- \equiv \lim_{s\to\infty} \lim_{\gamma\to 0} \lim_{\Omega\to\infty} (s+t)^{\nu} \operatorname*{Min}_{\mathbf{p}} W(\mathbf{p}). \tag{5.15}$$

The first step is to carry out the limit process $\Omega \to \infty$ in (15). It follows from the definition (7) that

$$|W(\mathbf{p}) - \lim_{\Omega\to\infty} W(\mathbf{p})| \leq \sum |w_{-,\min}(\mathbf{k})| \tag{5.16}$$

where the sum is over all lattice points $\mathbf{k}$ outside a cube of volume just less than $2^{\nu}\Omega$ centered at the origin. This sum is independent of $\mathbf{p}$, and because of (1.21) it tends to zero as $\Omega \to \infty$. Therefore $W(\mathbf{p})$ approaches its limit, as $\Omega \to \infty$ uniformly in p, so that

$$\begin{aligned} \lim_{\Omega\to\infty} \operatorname*{Min}_{\mathbf{p}} W(\mathbf{p}) &= \operatorname*{Min}_{\mathbf{p}} \lim_{\Omega\to\infty} W(\mathbf{p}) \\ &= \operatorname*{Min}_{\mathbf{p}} \sum(\mathbf{p}, \gamma) \end{aligned} \tag{5.17}$$

where

$$\sum(\mathbf{p}, \gamma) \equiv \sum_{\mathbf{k}} w_{-,\min}(\mathbf{k}) \exp(2\pi i \mathbf{p}\cdot\mathbf{k}), \tag{5.18}$$

the sum going over an infinite lattice with spacing $s + t$.

To study the effect of the next limit process, $\gamma \to 0$, we consider separately the cases where $\mathbf{p}$ is inside and outside a cube Γ of side $2p_0$ centered at the origin. The dependence of p_0 on γ will be chosen so that in the limit $\gamma \to 0$ the sum (18) may be replaced by an integral if $\mathbf{p} \in \Gamma$ and by zero otherwise.

By writing $\gamma\mathbf{q}$ for $\mathbf{p}$ and $\mathbf{x}$ for $\gamma\mathbf{k}$ in (18), and using an argument similar to that which gave (2.20), we find that

$$\begin{aligned} (s+t)^{\nu} \lim_{\gamma\to 0} \sum(\gamma\mathbf{q}, \gamma) &= \int \varphi_-(\mathbf{x}) \exp(2\pi i \mathbf{q}\cdot\mathbf{x})\, d\mathbf{x} \\ &= \Phi_-(\mathbf{q}). \end{aligned} \tag{5.19}$$

In order to use this result in (15), we must show also that the convergence to the limit is uniform provided $\gamma\mathbf{q} \in \Gamma$. To do this we use the definitions (3.1), (1.6), (18), and (6), to obtain the estimate

$$\begin{aligned} &|\Phi(\mathbf{p}/\gamma) - (s+t)^{\nu} \sum(\mathbf{p}, \gamma)| \\ &= \Big| \sum_{\mathbf{k}} \int_{\omega_0} e^{2\pi i \mathbf{p}\cdot\mathbf{k}} [e^{2\pi i \mathbf{p}\cdot\mathbf{y}} w_-(\mathbf{k}+\mathbf{y}, \gamma) \\ &\quad - \operatorname*{Min}_{\mathbf{z}\in\omega_0} w_-(\mathbf{k}+2\mathbf{z}, \gamma)]\, d\mathbf{y} \Big| \\ &\leq \sum_{\mathbf{k}} \int_{\omega_0} \{|e^{2\pi i \mathbf{p}\cdot\mathbf{y}} - 1|\, |w_-(\mathbf{k}+\mathbf{y}, \gamma)| \\ &\quad + |w_-(\mathbf{k}+\mathbf{y}, \gamma) - \operatorname*{Min}_{\mathbf{z}\in\omega_0} w_-(\mathbf{k}+2\mathbf{z}, \gamma)|\}\, d\mathbf{y} \\ &\leq \pi\nu p_0 (s+t) \int |w_-(\mathbf{x}, \gamma)|\, d\mathbf{x} \\ &\quad + \Big| \int w_-(\mathbf{x}, \gamma)\, d\mathbf{x} - (s+t)^{\nu} \sum_{k} w_{-,\min}(\mathbf{k}) \Big|. \end{aligned} \tag{5.20}$$

Since we have required that $\int |\varphi_-(\mathbf{y})|\, d\mathbf{y} < \infty$, the first part of the last member of (20) tends to zero with γ provided that

$$\lim_{\gamma\to 0} \mathbf{p}_0 = 0. \tag{5.21}$$

The second part also tends to zero with γ, as in the argument leading to (2.20), because of the Riemann integrability of φ_-. Since both parts are independent of $\mathbf{p}$, the convergence to the limit in (19) is indeed uniform for $\gamma\mathbf{q} \in \Gamma$, and it follows that

$$(s+t)^{\tau} \lim_{\gamma\to 0} \min_{\mathbf{p}\in\Gamma} \sum(\mathbf{p},\gamma) = \lim_{\gamma\to 0} \operatorname{Min}_{\mathbf{p}\in\Gamma} \Phi_-(\mathbf{p}/\gamma) = \operatorname{Min}_{\mathbf{q}} \Phi_-(\mathbf{q}) \equiv \Phi_-(0) \tag{5.22}$$

by (2), provided $\lim_{\gamma\to 0} p_0/\gamma = \infty$.

To complete the estimates of Min $\sum(\mathbf{p},\gamma)$ for (18) we must also find a lower bound for $\sum(\mathbf{p},\gamma)$ with $\mathbf{p}$ outside Γ. Whenever $\mathbf{p}$ is outside Γ, we can find a direction parallel to an edge of the cube Γ such that the component of $\mathbf{p}$ along this direction exceeds p_0. If $\mathbf{l}$ denotes a vector of length $s+t$ along this direction we therefore have

$$p_0(s+t) \le \mathbf{p}\cdot\mathbf{l} \le \tfrac{1}{2}, \tag{5.23}$$

the second inequality being a consequence of the fact [see (8)] that all the allowed values of $\mathbf{p}$ lie inside a cube of side $(s+t)^{-1}$ centered at the origin. On multiplying both sides of (18) by $1 - \exp(2\pi i\mathbf{p}\cdot\mathbf{l})$ and taking absolute values, we obtain, since $|\mathbf{l}| = s+t$,

$$\begin{aligned} &2\sin(\pi\mathbf{p}\cdot\mathbf{l})\,|\textstyle\sum(\mathbf{p}\cdot\gamma)| \\ &\quad = |\textstyle\sum_{\mathbf{k}} [w_{-,\min}(\mathbf{k}) - w_{-,\min}(\mathbf{k}-\mathbf{l})] \exp(2\pi i\mathbf{p}\cdot\mathbf{k})| \\ &\quad \le \textstyle\sum_{\mathbf{k}} |w_{-,\min}(\mathbf{k}) - w_{-,\min}(\mathbf{k}-\mathbf{l})| \\ &\quad \le \textstyle\sum_{\mathbf{k}} \operatorname{Max}_{\mathbf{x}\in\omega_0} w_-(\mathbf{k}+4\mathbf{x},\gamma) \\ &\qquad - \textstyle\sum_{\mathbf{k}} \operatorname{Min}_{\mathbf{x}\in\omega_0} w_-(\mathbf{k}+4\mathbf{x},\gamma) \end{aligned} \tag{5.24}$$

where ω_0 denotes, as in (6), a cube of side $(s+t)$ centered at the origin. Applying once more an argument of the type which led to (2.20), we find that both sums in the last member of (24) tend to the same integral in the limit $\gamma \to 0$. Thus (24) can be written in the form

$$|\textstyle\sum(\mathbf{p},\gamma)| \le \tfrac{1}{2}\,|\csc(\pi\mathbf{p}\cdot\mathbf{l})|\,\delta(\gamma) \tag{5.25}$$

where δ does not depend and on p and tends to 0 with γ. Combining (23) and (25) we obtain

$$|\textstyle\sum(\mathbf{p},\gamma)| \le \delta(\gamma)/4p_0(s+t) \quad \text{for } \mathbf{p} \text{ outside } \Gamma \tag{5.26}$$

since $\sin x \ge 2x/\pi$ for $0 \le x \le \pi/2$. Let us now choose p_0 in such a way that

$$\lim_{\gamma\to 0} \delta(\gamma)/p_0 = 0. \tag{5.27}$$

This is compatible with the previous requirements (5.21) and $\lim p_0/\gamma = \infty$; for example, the choice $p_0 \propto [\gamma + \delta(\gamma)]^{\frac{1}{2}}$ satisfies all three requirements. Combining (27) and (26) we obtain

$$\lim_{\gamma\to 0} \operatorname*{Min}_{\mathbf{p}\text{ outside }\Gamma} \sum(\mathbf{p},\gamma) = 0. \tag{5.28}$$

Combining (15) and (17) and then using (22) and (28), we obtain

$$\begin{aligned} \alpha_- &= \lim_{s\to\infty} \lim_{\gamma\to 0} (s+t)^{\tau} \operatorname*{Min}_{\mathbf{p}} \sum(\mathbf{p},\gamma) \\ &= \lim_{s\to\infty} \Phi_-(0) = \Phi_-(0) = \Phi_{\min}. \end{aligned} \tag{5.29}$$

This completes the evaluation of α_-.

Our main results (2.25) and (13) can be summarized in the formula

$$\begin{aligned} \mathrm{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha_-\rho^2\} + \tfrac{1}{2}\alpha_+\rho^2 &\le a(\rho,0+) \\ &\le \mathrm{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\}. \end{aligned} \tag{5.30}$$

From (30) we can find $a(\rho,0+)$ exactly provided the upper and lower bounds coincide. This can happen in two ways:

(i) Where $a^0(\rho) + \frac{1}{2}\alpha_-\rho^2$ coincides with its convex envelope, (e.g., if $\alpha_- = 0$), the function $a^0\rho + \frac{1}{2}\alpha\rho^2$ does so too since $\alpha = \alpha_+ + \alpha_- \ge \alpha_-$; thus (30) implies

$$\begin{aligned} &a(\rho,0+) = \mathrm{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\} = a^0(\rho) + \tfrac{1}{2}\alpha\rho^2 \\ &\text{if} \quad a^0(\rho) + \tfrac{1}{2}\alpha_-\rho^2 = \mathrm{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha_-\rho^2\}. \end{aligned} \tag{5.31}$$

In the special case $\alpha_- = 0$ this reduces to (3.14).

(ii) If $\alpha_+ = 0$ then we have $\alpha = \alpha_-$ so that (30) reduces to

$$a(\rho,0+) = \mathrm{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\} \quad \text{if} \quad \alpha_+ = 0. \tag{5.32}$$

The result (4.26) is a special case of this, since (4.1) implies $\Phi(0) \le \Phi(\mathbf{p})$ for all $\mathbf{p}$.

VI. EQUATION OF STATE

For a system with finite γ, the thermodynamic pressure $p(\rho,\gamma)$ is given[7,9] by

$$\begin{aligned} p(\rho,\gamma) &= -\partial(\rho^{-1}a(\rho,\gamma))/\partial(\rho^{-1}) \\ &= \rho\,\partial a(\rho,\gamma)/\partial\rho - a(\rho,\gamma), \end{aligned} \tag{6.1}$$

the differentiation being at constant γ and also at constant T, although the dependence on T is not explicit in the notation. Since $a(\rho,\gamma)$ is[7,9] a convex function of ρ at constant γ the derivative on the right of (1) exists[16] except on a countable set of values of ρ; it seems likely that this countable set is in fact empty but no proof is known.

Taking the limit $\gamma \to 0$ on both sides of (6.1) we obtain

$$p(\rho,0+) = \rho \lim_{\gamma\to 0} \partial a(\rho,\gamma)/\partial\rho - a(\rho,0+). \tag{6.2}$$

To evaluate the right side of (2) we must show that the order of the operations $\gamma \to 0$ and $\partial/\partial\rho$

[16] Ref. 10, p. 94.

can be reversed. To this end we use the inequality[17] for the right- and left-hand derivatives of a convex function

$$\begin{aligned}[a(\rho - \epsilon, \gamma) - a(\rho)]/(-\epsilon) &\leq D_-a(\rho, \gamma) \\ &\leq D_+a(\rho, \gamma) \\ &\leq [a(\rho + \epsilon, \gamma) - a(\rho)]/\epsilon\end{aligned} \tag{6.3}$$

where

$$D_\pm a(\rho, \gamma) \equiv \lim_{\epsilon \to 0} [a(\rho \pm \epsilon, \gamma) - a(\rho, \gamma)]/(\pm\epsilon), \tag{6.4}$$

with ϵ an arbitrary positive number.

Taking the limit $\gamma \to 0$ on both sides of (3), followed by the limit $\epsilon \to 0$, we obtain

$$\begin{aligned}D_-a(\rho, 0+) &\leq \lim_{\gamma \to 0} D_-a(\rho, \gamma) \\ &\leq \lim_{\gamma \to 0} D_+a(\rho, \gamma) \leq D_+a(\rho, 0+).\end{aligned} \tag{6.5}$$

Hence if $a(\rho, 0+)$ is differentiable all four of these expressions are equal, so that $\lim_{\gamma \to 0} D_\pm a(\rho, \gamma)$ exists and is equal to $da(\rho, 0+)/d\rho$. Substituting this result into (2) we obtain

$$p(\rho, 0+) = (\rho d/d\rho - 1)a(\rho, 0+) \tag{6.6}$$

at all points where $a(\rho, 0+)$ is differentiable. Equation (6) can be interpreted geometrically by the statement that $-p(\rho, 0+)$ is the place where the tangent at ρ to the graph of $a(\rho, 0+)$ intercepts the a-axis.

Combining (6) with (5.31) or (5.32), we obtain

$$p(\rho, 0+) = (\rho d/d\rho - 1)\ \mathrm{CE}\ \{a^0(\rho) + \tfrac{1}{2}\alpha\rho^2\} \tag{6.7}$$

for all Kac potentials to which (5.31) or (5.32) applies, including nonnegative definite and nonpositive potentials.

If the graph of $a(\rho, 0+) = \mathrm{CE}\ \{a^0(\rho) + \frac{1}{2}\alpha\rho^2\}$ has a straight segment (see Fig. 3) then the geometrical interpretation of (6) shows that $p(\rho, 0+)$ is constant along this straight segment; and the chemical potential $[a(\rho, 0+) + p(\rho, 0+)]/\rho = da(\rho, 0+)/d\rho$ is also constant. This behavior of the thermodynamic functions characterizes a first-order phase transition.

Since the straight portion of the graph $\mathrm{CE}\ \{a^0(\rho) + \frac{1}{2}\alpha\rho^2\}$ touches the graph $a^0(\rho) + \frac{1}{2}\alpha\rho^2$ at both ends, the quantity

$$\tilde{p}_{vdw}(\rho) \equiv (\rho d/d\rho - 1)(a^0(\rho) + \tfrac{1}{2}\alpha\rho^2) \tag{6.8}$$

takes the same value, call it p_{sat}, at the two ends of the phase transition region. Moreover, if ρ_l and ρ_v are the abscissas of the two points of contact, then we have

$$\begin{aligned}\int_{\rho_v}^{\rho_l} [\tilde{p}_{vdw}(\rho) - p_{sat}]\, d(\rho^{-1}) &= \{-[a^0(\rho) + \tfrac{1}{2}\alpha\rho^2 + p_{sat}]\rho^{-1}\}_{\rho_v}^{\rho_l} \\ &= \{(d/d\rho)[a^0(\rho) + \tfrac{1}{2}\alpha\rho^2]\}_{\rho_v}^{\rho_l}\end{aligned} \tag{6.9}$$

using (8) first as a definition of $\tilde{p}_{vdw}(\rho)$ and then to evaluate p_{sat}. Since the slopes of the graph $a^0(\rho) + \frac{1}{2}\alpha\rho^2$ at the two points of contact are equal, the expression (9) vanishes. This shows that ρ_l and ρ_v are related to $\tilde{p}_{vdw}(\rho)$ by Maxwell's equal-area construction (Fig. 1). The replacement of $a^0(\rho) + \frac{1}{2}\alpha\rho^2$ by its convex envelope is precisely equivalent to making Maxwell's modification (1.2) in the function $\tilde{p}_{vdw}(\rho)$ defined by (8).

For values of ρ where the upper and lower bounds of $a(\rho, 0+)$, in (5.30) do not coincide we still obtain bounds on $\lim_{\gamma \to 0} \partial a(\rho, \gamma)/\partial\rho$ and hence on $p(\rho, 0+)$ by an argument due to Fisher[18]: owing to the convexity of the graph of $a(\rho, 0+)$ its slope at $\rho = \rho_0$ must lie between the slopes of tangents to its upper bound crossing its lower bound at $\rho = \rho_0$.

A result similar to that obtained for the pressure holds also for the internal energy density

$$u(\beta, \rho, \gamma) = (\partial/\partial\beta)[\beta a(\beta, \rho, \gamma)]. \tag{6.10}$$

Since $-\beta a(\beta, \rho, \gamma)$ is convex in β we obtain, as in the derivation of (6),

$$u(\beta, \rho, 0+) = \partial/\partial\beta[\beta a(\beta, \rho, 0+)]. \tag{6.11}$$

Unfortunately we have been unable to prove any similar general statements about the specific heat and compressibility which correspond to second derivatives of the free-energy density. At sufficiently low densities, however, one can show that all the density derivatives of $p(\rho, \gamma)$ approach the corresponding derivatives of $p(\rho, 0+)$ by using Vitali's theorem together with the results of Lebowitz and Penrose[19] for the convergence of virial expansions.

VII. THE PAIR DISTRIBUTION FUNCTION

In order to understand better the effect of the Kac potential on possible phase transitions in the reference system and in the van der Waals system, it is useful to study the two-body distribution function. (In this section we do not aim at such a high standard of rigor as in the earlier sections.)

[17] Obtained by making $\delta \to 0$ in Eq. (3.18.3) of Hardy, Littlewood and Polya (Ref. 10).

[18] M. E. Fisher, "Bounds for the Derivatives of the Free Energy and the Pressure of a Hard Core System near Close Packing" (to be published).

[19] J. L. Lebowitz and O. Penrose, J. Math. Phys. 5, 841 (1964).

Instead of the usual two-body distribution function[20] $n_2(\mathbf{x}_1, \mathbf{x}_2)$, we shall study its space average

$$\bar{n}_2(\mathbf{r}) \equiv \Omega^{-1} \int n_2(\mathbf{x}, \mathbf{x} + \mathbf{r})\, d\mathbf{x}. \tag{7.1}$$

This distribution function when the system has N particles in a box Ω and a Kac parameter γ will be denoted by $\bar{n}_2(\mathbf{r}; N, \Omega, \gamma)$. This function can be determined from the formula[21]

$$\tfrac{1}{2} \int \bar{n}_2(\mathbf{r}; N, \Omega, \gamma) q'(\mathbf{r})\, d\mathbf{r} = (\partial/\partial\eta)\Omega^{-1} A[N, \Omega, \gamma, \eta q'(\mathbf{r})]_{\eta=0} \tag{7.2}$$

which holds for arbitrary bounded functions $q'(\mathbf{r})$. The integration may be taken over all space, since $\bar{n}_2(\mathbf{r}; N, \Omega, \gamma) = 0$ for large r. The symbol $A[N, \Omega, \gamma, \eta q'(\mathbf{r})]$ denotes the free energy analogous to (1.10) when the short-range interaction potential is not $q(\mathbf{r})$ but $q(\mathbf{r}) + \eta q'(\mathbf{r})$, and η is a parameter. Taking the thermodynamic limit and then the van der Waals limit in (2) we obtain, provided all the relevant limits and the derivative exist,

$$\tfrac{1}{2} \int \bar{n}_2(\mathbf{r}; \rho, 0+) q'(\mathbf{r})\, d\mathbf{r} = (\partial/\partial\eta) a[\rho, 0+, \eta q'(\mathbf{r})]_{\eta=0} \tag{7.3}$$

where

$$\bar{n}_2(\mathbf{r}; \rho, 0+) \equiv \lim_{\gamma\to 0} \lim_{\Omega\to\infty} \bar{n}_2(\mathbf{r}; \rho\Omega, \Omega, \gamma) \tag{7.4}$$

and

$$a[\rho, 0+, \eta q'(\mathbf{r})] \equiv \lim_{\gamma\to 0} \lim_{\Omega\to\infty} A[\rho\Omega, \Omega, \gamma, \eta q'(\mathbf{r})]/\Omega. \tag{7.5}$$

The permutation of the limit operations $\gamma \to 0$ and $\Omega \to \infty$ with integration and differentiation in deriving (3) is justified on the left by Lebesgue's theorem.[22] On the right it is justified [as in the argument based on (6.3) and (6.5)], by the convexity of $A[N, \Omega, \gamma, \eta q'(\mathbf{r})]$ as a function of η. This convexity can be verified by calculating the second derivative of A with respect to η. In a similar way we can obtain

$$\tfrac{1}{2} \int \bar{n}_2^0(\mathbf{r}; \rho) q'(\mathbf{r})\, d\mathbf{r} = (\partial/\partial\eta) a^0[\rho, \eta q'(\mathbf{r})]_{\eta=0} \tag{7.6}$$

where the superscript zero denotes quantities belonging to the reference system with a modified short-range potential $q(\mathbf{r}) + \eta q'(\mathbf{r})$. When the conditions of validity for our main result (5.31) or (5.32) are satisfied, (3) reduces to

$$\tfrac{1}{2} \int \bar{n}_2(\mathbf{r}; \rho, 0+) q'(\mathbf{r})\, d\mathbf{r} = (\partial/\partial\eta)\, \mathrm{CE}\, \{a^0[\rho, \eta q'(\mathbf{r})] + \tfrac{1}{2}\alpha\rho^2\}_{\eta=0}. \tag{7.7}$$

Two cases may be distinguished. In the first of these the graph of $a^0[\rho, \eta q'(\mathbf{r})] + \frac{1}{2}\alpha\rho^2$ coincides with that of its convex envelope and so (7) combines with (6) to give

$$\tfrac{1}{2} \int \bar{n}_2(\mathbf{r}; \rho, 0+) \mathbf{q}'(\mathbf{r})\, d\mathbf{r} = \tfrac{1}{2} \int \bar{n}_2^0(\mathbf{r}, \rho) q'(\mathbf{r})\, d\mathbf{r} \tag{7.8}$$

which implies

$$\bar{n}_2(\mathbf{r}; \rho, 0+) = \bar{n}_2^0(\mathbf{r}, \rho) \tag{7.9}$$

since $q'(\mathbf{r})$ is arbitrary. In this case, therefore, the Kac potential does not affect the distribution of pairs of particles, as one might expect from (1.8).

In the alternative case, where the graph of $a^0[\rho, \eta q'(\mathbf{r})] + \frac{1}{2}\alpha\rho^2$ does not coincide with that of its convex envelope, the latter is a straight line touching the former at two places, say ρ_1 and ρ_2. Both ρ_1 and ρ_2 depend, in general, on η. The equation of this straight line may be written

$$\mathrm{CE}\, \{a^0[\rho, \eta q'(\mathbf{r})] + \tfrac{1}{2}\alpha\rho^2\} = [(\rho - \rho_1)a_2 + (\rho_2 - \rho)a_1](\rho_2 - \rho_1)^{-1} (\rho_1 < \rho < \rho_2) \tag{7.10}$$

where

$$a_h \equiv a^0[\rho_h, \eta q'(\mathbf{r})] + \tfrac{1}{2}\alpha\rho_h^2 \qquad (h = 1, 2). \tag{7.11}$$

On substituting this into (7) and using (9), which applies when $\rho = \rho_1$ or $\rho = \rho_2$, we obtain

$$\tfrac{1}{2} \int \bar{n}_2(\mathbf{r}; \rho, 0+) q'(\mathbf{r})\, d\mathbf{r} = \tfrac{1}{2} \int \left[\frac{\rho - \rho_1}{\rho_2 - \rho_1} \bar{n}_2^0(\mathbf{r}; \rho_2) + \frac{\rho_2 - \rho}{\rho_2 - \rho_1} \bar{n}_2^0(\mathbf{r}, \rho_1) \right] q'(\mathbf{r})\, d\mathbf{r} \tag{7.12}$$

since

$$\frac{\partial}{\partial\rho_1} \left[\frac{(\rho - \rho_1)a_2 + (\rho_2 - \rho)a_1}{\rho_2 - \rho_1} \right] = \frac{\rho_2 - \rho}{(\rho_2 - \rho_1)^2} \left[a_1 - a_2 + (\rho_2 - \rho_1) \frac{\partial}{\partial\rho_1} a_1 \right] = 0, \text{ etc.} \tag{7.13}$$

by virtue of the double tangent construction. [In Eq. (13) we are treating η, ρ_1, ρ_2, ρ as independent

[20] T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), Sec. 29.

[21] Equation (7.2) is the classical analog of a quantum formula given by Bogolyubov and Zubarev, Zh. Eksperim. i Teor. Fiz. **28**, 129 (1955) [English transl.: Soviet Phys.—JETP **1**, 83 (1955)]. This was first used in this type of work by M. E. Fisher.

[22] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955), p. 37.

variables.] Since $q'(\mathbf{r})$ is arbitrary, it follows from (12) that

$$\bar{n}_2(\mathbf{r};\rho,0+) = \frac{\rho-\rho_1}{\rho_2-\rho_1}\bar{n}_2^0(\mathbf{r};\rho_2) + \frac{\rho_2-\rho}{\rho_2-\rho_1}\bar{n}_2^0(\mathbf{r};\rho_2), \qquad (\rho_1<\rho<\rho_2). \tag{7.14}$$

This equation can be interpreted[23] to mean that two phases coexist whenever CE $\{a^0(\rho)+\frac{1}{2}\alpha\rho^2\} < a^0(\rho)+\frac{1}{2}\alpha\rho^2$; their pair distribution functions are $\bar{n}_2^0(\mathbf{r};\rho_1)$ and $\bar{n}_2^0(\mathbf{r};\rho_2)$, and their densities ρ_1 and ρ_2, respectively. This is consistent with the interpretation of a straight portion of the graph $a(\rho,0+)$ as a first-order transition.

In interpreting the results (9) and (14) it should be remembered that their derivation involves the limit process $\gamma\to 0$; they therefore yield information only about values of r small compared with the range γ^{-1} of the Kac potential. For example, (14) implies a two-phase structure on a length scale $\ll\gamma^{-1}$ but not necessarily on a scale $\gg\gamma^{-1}$.

VIII. SIMPLE UPPER BOUND ON $a(\rho,\gamma)$

Of the various upper and lower bounds on $a(\rho,\gamma)$ obtained in Sec. II to V, the only one simple enough to be useful for finite values of γ is the lower bound (3.12) for nonnegative-definite Kac potentials. The other bounds are too unwieldy because they involve the network of cells $\omega_1\cdots\omega_M$.

A simple upper bound on $a(\rho,\gamma)$ can be obtained by a method due in essence to Gibbs.[24] We rewrite (1.11) in the form

$$Z(N,\Omega,\gamma) = Z^0(N,\Omega)\langle e^{-\beta W}\rangle^0 \tag{8.1}$$

where W is the long-range contribution to the potential energy and $\langle\ \rangle^0$ indicates a canonical average over the reference system. Since $e^{-\beta W}$ is convex we have by (1.6) and (7.1)

$$\langle e^{-\beta W}\rangle^0 \geq \exp\langle -\beta W\rangle^0 = \exp\left[-\tfrac{1}{2}\beta\iint\gamma^\nu\phi(\gamma\mathbf{r})\bar{n}_2^0(\mathbf{x},\mathbf{x}+\mathbf{r})\,d\mathbf{x}\,d\mathbf{r}\right]. \tag{8.2}$$

Combining (1) and (2) and then taking the thermodynamic limit, we obtain

$$a(\rho,\gamma)\leq a^0(\rho)+\tfrac{1}{2}\gamma^\nu\int\varphi(\gamma\mathbf{r})\bar{n}_2^0(\mathbf{r};\rho)\,d\mathbf{r} \tag{8.3}$$

where $\bar{n}_2^0$ is defined as in Sec. VII. This inequality may be strengthened in the manner used in going from (2.24) to (2.25), to give

$$a(\rho,\gamma)\leq \mathrm{CE}\left\{a^0(\rho)+\tfrac{1}{2}\gamma^\nu\int\varphi(\gamma\mathbf{r})\bar{n}_2^0(\mathbf{r};\rho)\,d\mathbf{r}\right\}. \tag{8.4}$$

A system for which the right side of (8.4) can be evaluated is the one-dimensional system considered by Kac, Uhlenbeck, and Hemmer.[6] In our notation it is defined by (1.4) and (1.5). For this system, the last term in (8.3) is essentially the Laplace transform of $\bar{n}_2^0$, and (8.4) becomes

$$a(\rho,\gamma)\leq \mathrm{CE}\,\{a^0(\rho) + \tfrac{1}{2}\alpha\rho\gamma[(1+\gamma\rho^{-1}-\gamma r_0)e^{\gamma r_0}-1]^{-1}\} \tag{8.5}$$

with

$$a^0(\rho) = \rho kT\{\log[\rho(1-\rho r_0)^{-1}(2\pi\hbar^2/mkT)^{\frac{1}{2}}]-1\}. \tag{8.6}$$

In the van der Waals limit (5) reduces to (2.25).

The argument which led to (4) also applies to lattice gases. For example, if the short-range potential is taken as

$$q(\mathbf{r}) = \begin{cases}+\infty & \text{if } \mathbf{r}=0,\\ 0 & \text{if } \mathbf{r}\neq 0,\end{cases} \tag{8.7}$$

the part of the interaction potential which prevents more than one particle occupying any site, then $\bar{n}_2^0(\mathbf{r};\rho)$ vanishes for $\mathbf{r}=0$ and takes the value ρ^2 for $\mathbf{r}\neq 0$; consequently the lattice-gas analog of (4) leads to

$$a(\rho,\gamma)\leq \mathrm{CE}\,\{a^0(\rho)+\tfrac{1}{2}\rho^2\sum{}'\,w(\mathbf{r},\gamma)\} \tag{8.8}$$

with

$$a^0(\rho) = kT[\rho\ln\rho+(1-\rho)\ln(1-\rho)]. \tag{8.9}$$

The sum $\sum'$ goes over the infinite lattice excluding $\mathbf{r}=0$.

IX. DISCUSSION

We have shown that for a class of Kac potentials, including nonnegative-definite and nonpositive potentials, the thermodynamic free-energy density is given in the van der Waals limit by

$$a(\rho,0+) = \mathrm{CE}\,\{a^0(\rho)+\tfrac{1}{2}\alpha\rho^2\} \tag{9.1}$$

and the equation of state by Maxwell's modification of the corresponding van der Waals equation of state. If $\alpha<0$, the graph of $a(\rho,0+)$ may have straight portions; these correspond to first-order phase transitions both in the thermodynamic prop-

[23] J. E. Mayer, J. Chem. Phys. 15, 187 (1947); G. E. Uhlenbeck, P. Hemmer, and M. Kac, J. Math. Phys. 4, 229 (1963).

[24] See M. Girardeau, J. Chem. Phys. 40, 899 (1964).

erties and in the behavior of the pair distribution function for $r \ll \gamma^{-1}$. If $\alpha > 0$ on the other hand, the graph of $a(\rho, 0+)$ cannot have a straight portion and thermodynamically the system can have no phase transition. A paradoxical situation arises if the reference system has a phase transition and $\alpha > 0$; then by (7.9) the pair distribution function has the form characterizing a phase transition, yet there is no phase transition in the thermodynamic sense. The explanation is that the result (7.9), which indicates the coexistence of two phases, was obtained using the limit process $\gamma \to 0$ and may therefore be relied on only when $r \ll \gamma^{-1}$. On the other hand, the term $\frac{1}{2}\alpha\rho^2$ in $a(\rho, 0+)$ indicates that on the length scale where the Kac potential operates (distances $\approx$ or $\gg \gamma^{-1}$) the system is uniform since there is no transition. It appears therefore, that the repulsive Kac potential causes the distinct liquid and gas phases of a normal first-order transition to break into droplets or froth whose chacteristic length is $\gg r_0$ but $\ll \gamma^{-1}$. This fact might possibly find a practical application.

Some of the results on which this paper is based can be generalized to quantum mechanics. The results of Sec. III generalize immediately to quantum mechanics, and so do those of Sec. II provided the boundary condition on the wavefunction is that it must vanish when the center of any particle touches the wall of the container. The main result of Sec. VIII also generalizes to quantum mechanics by virtue of Bogolyubov's inequality.[25] On the other hand Sec. IV and V do not generalize so readily. Unfortunately this means that we can at present evaluate $a(\rho, 0+)$ rigorously only for nonnegative-definite Kac potentials, which cannot produce a phase transition. Thus it remains to be shown that a van der Waals phase transition can occur in a quantum system.

The results of this paper can easily be generalized to classical lattice gases. The proofs require only minor modifications, and the main results (5.31) and (5.32) are the same. Our results apply also to Ising spin systems since these are isomorphic to lattice gases. In this way the Bragg–Williams and Weiss theories can also be dealt with in a rigorous fashion.[26]

Another direction in which our results might be generalized is to weaken the conditions (1.20) and (1.21). For example, for nonnegative-definite Kac potentials the hard-core condition (1.20a) can be replaced by the condition $q(r) > \text{const } r^{-\nu-\epsilon}$ for small r, which is sufficient[7,9] to ensure the existence of $a^0(\rho)$ and $a(\rho, \gamma)$. For more general Kac potentials the hard-core condition plays no part in the upper bound on $a(\rho, 0+)$ but is used to restrict the number of particles in a cell when lower bound on $a(\rho, 0+)$ is obtained. Possibly a more refined argument could dispense with the hard-core condition altogether.

A more interesting extension of this work would be to study Kac potentials satisfying neither of the conditions (i) and (ii) given at the end of Sec. V. For these potentials upper and lower bounds on $a(\rho, 0+)$ do not coincide, and it is possible that the behavior of the system is more complicated than in the van der Waals theory. For example the Kac potential might bring about spatial ordering with a length scale γ^{-1}. Finally from the physical point of view the most important extension of this work is to study the properties of the system when γ is finite but small. This is now being carried out along several lines.[27]

ACKNOWLEDGMENTS

We are grateful to M. Austin, M. E. Fisher, E. Lieb, D. Newman, and N. van Kampen for useful advice and suggestions.

[25] Cited in Ref. 4 of V. V. Tolmachev, Dokl. Akad. Nauk SSSR **134**, 1324 (1960) [Engl. transl.: Soviet Phys.–Doklady **5**, 984 (1961)]. See also M. Girardeau, J. Chem. Phys. **41**, 2945 (1964).

[26] In this connection see also the work of R. Brout, Phys. Rev. **115**, 824 (1959) (and later publications) who considered expansions in the inverse number of interacting spins, and G. A. Baker, Phys. Rev. **126**, 2072 (1962) who first considered the limiting process $\gamma \to 0$ in his study of spin systems.

[27] J. L. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. **6**, 1282 (1965).

Representation of Discrete Symmetry Operators

K. H. MARIWALLA*
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina
and
Department of Physics and Astronomy, University of Georgia, Athens, Georgia
(Received 22 March 1965)

Representations of discrete symmetry operators (DSO's) connected with space ($\mathcal{P}$), time (T), and generalized charge ($\mathcal{C}$) are considered. It is shown that if one writes a DSO as exp $(i\pi\Omega)$ × a phase transformation, then (under certain conditions on Ωs) to each DSO there corresponds a set of Ωs which is closed with respect a Lie algebra, which is isomorphic to the Lie algebra of generators of rotation in an n-dimensional Euclidean space; where n is the number of commuting observables that changes sign under this DSO in a given representation (e.g. linear momentum representation). In the particular case of the ($T\mathcal{C}\mathcal{P}$) operation, there are six Ωs, of which two are diagonal, viz. the generalized charge Q, and spin projection along the z axis Sz; corresponding Euclidean group is four-dimensional. For the sake of completeness, the representations are also given for the following cases: (i) nonrelativistic quantum mechanics, (ii) quantum theory of free fields, in terms of field operators.

I. INTRODUCTION

ONE of the aspects of discrete symmetries[1] that has not been studied extensively has to do with their explicit representations; the principal reason for this is obviously the lack of immediate physical application. The study of explicit expressions for the representations of *discrete symmetry operators* (DOS's) in quantum mechanics has a twofold importance. Firstly, it helps in a systematic study of the subject; secondly, it is possible that such investigations may throw some new light on the connection between the discrete symmetries and the other observables. This is then the motivation for the present paper. Such expectation is not without foundation. For we know that given a local field theory which has the symmetry of proper inhomogeneous Lorentz group (Poincaré group), one obtains into the bargain $T\mathcal{C}\mathcal{P}$ invariance if one quantizes this field theory keeping spin–statistics connection in mind. The concept of $T\mathcal{C}\mathcal{P}$ implies the existence of concept of particles–antiparticles, which did not exist in the unquantized theory. It is therefore not frivolous to suggest that $T\mathcal{C}\mathcal{P}$ may be connected with some other group of which generalized charge is one of the generators. As we shall find in Sec. IV, this is indeed the case.

To start with, in Sec. II the case of nonrelativistic quantum mechanics is considered.

In an earlier paper the author[2] has given representations for various DSO's for free fields; however, these representations were in terms of creation and annihilation operators in linear and angular momentum representations. The representations involving the direct use of field operators, rather than their positive and negative frequency parts, have not however been extensively treated.[3,4] The expressions for DSO's in terms of field operators are important for any extension to interacting fields; Sec. III is therefore devoted to finding explicit representations of DSO's for the Klein–Gordon (KG) and Dirac fields, in terms of field operators and their first derivatives.

Finally in Sec. IV an attempt is made to write down the symmetry operators in terms of (purely) algebraic relationships that exist between DSO's and the operators for observables that change sign under application of these DSO's. As the defining

* Present address: Department of Physics, Indian Institute of Technology, Bombay, P.O.: I.I.T.-Bombay, Powai, Bombay-76.

[1] The definitions of various discrete symmetry operators in terms of their effect on observables is summarized in the table. The notation has been so chosen that unitary symmetry operators are denoted by script letters and the antiunitary symmetry operators by ordinary Roman capitals.

[2] K. H. Mariwalla, Rev. Mod. Phys. **34**, 215 (1962). The complete list of references to the literature on this subject is given in this paper. This paper will be referred to as (M1).

[3] B. P. Nigam and L. L. Foldy, Phys. Rev. **102**, 1410 (1956); representations of $\mathcal{P}$ and $\mathcal{C}$ for the Dirac field are given. However, the expression obtained for $\mathcal{C}$ is not correct as it is assumed that Dirac spinors form a complete orthogonal set relative to their Hermitian conjugates, rather than their adjoints (which are distinct from Hermitian conjugates for the Dirac field), as is actually the case. See Ref. 12 and the remark following Eq. 12.

[4] P. G. Federbush and M. T. Grisaru, Nuovo Cimento **9**, 890 (1958); general formula is given valid for Klein–Gordon and Dirac fields. However, the formula requires for its validity the concept of "number ordering." Now normal ordering consists in shifting all creation operators to the left of destruction operators using commutators for Bosons and anticommutators for Fermions—i.e. for every interchange one multiplies by a factor +1 for Bosons and −1 for Fermions. In 'number ordering' one uses −1 for Bosons and +1 for Fermions. Hence, this contradicts the commutation relations. Thus, if one uses 'number ordering,' one gets $-2i\int b_{\mathbf{k}}{}^{+}b_{-\mathbf{k}}\,d^3k$ for the right side of (10), instead of zero.

table for the DSO's shows, there are just three such observables in the linear momentum representation, viz. **P**, Q, and Sz. Let n denote the number of observables that change sign under a particular DSO (the maximum is obviously three); then one finds that one can associate with this DSO $\frac{1}{2}(n+2)(n+1) - n$ "generators," which together with the n "observables" satisfy a definite Lie algebra. This Lie algebra is the same as that satisfied by the generators of the rotation group in Euclidean $(n+2)$-space. For the particular case of the operation TCP in which spin and generalized charge change sign, the corresponding Lie algebra consists of 6 operators including Sz and Q; this algebra is isomorphic to the algebra of generators of rotation in Euclidean four-space. It appears that this four-space is truly Euclidean so that the corresponding group is compact.

II. DISCRETE SYMMETRY OPERATORS IN NONRELATIVISTIC QUANTUM MECHANICS

Within the scope of nonrelativistic mechanics, since there is no concept of an antiparticle, there is no operation of particle conjugation either, though formally the equations are invariant under the change of sign of the electric charge. The nonexistence of the operator for charge conjugation is connected with the fact that there is no nonrelativistic operator for charge, as classically charge is not a canonical variable.[5] Thus within the scope of nonrelativistic quantum mechanics, there are just three symmetry operators, viz. $\mathcal{P}$, T, and their product I.

In order to construct representations for these, first note that any unitary operator may be written as $\exp(i\alpha\Omega)$, where Ω is a Hermitian operator and α is a real number; we shall refer to Ω as the "generator" of this unitary transformation. The object then is to find Ω. Now continuous symmetries are characterized by invariance of a system under certain kinds of displacements, whether linear or angular. The number α in this case is a continuous parameter, and Ω is a constant of the motion. On the other hand, discrete symmetries (if at all) give rise to multiplicative quantum numbers (e.g. ordinary parity), and α is a fixed number—i.e. only for this particular α is $\exp(i\alpha\Omega)$ a symmetry operation. Consequently, while the DSO $\exp(i\alpha\Omega)$ is itself a constant of the motion, the "generator" Ω is in general not a constant of the motion. Because of this it is difficult to find the "generators" of discrete symmetries by any general method. However, in nonrelativistic quantum mechanics, it is possible to construct representations of $\mathcal{P}$, T, and I in terms of the operators for coordinates and momenta. To do this we need the following theorem:

If A, B, are two canonically conjugate Hermitian operators, i.e. $[A, B] = i$, then the set of operators

$$\Omega_1 = \tfrac{1}{2}i(AB + BA), \qquad \Omega_2 = \tfrac{1}{2}i(A^2 - B^2), \qquad \Omega_3 = \tfrac{1}{2}(A^2 + B^2) \tag{1}$$

generate the Lie algebra

$$[\Omega_i, \Omega_j] = 2i\Omega_k; \tag{2}$$

the operators $U_i = \exp(i\pi\Omega_i)$ then have the property

$$U_iAU_i^{-1} = -A, \qquad U_iBU_i^{-1} = -B. \tag{3}$$

The proof is straightforward and will not be given here. Note that Ω_3 has the remarkable property[6] of being a generator of rotations in the space of A, B. We remark that, whereas the operators U_2, U_3 are unitary and involutary, U_1 has neither of these properties. However, U_1 has the useful property, $U_1f(A) = f(-A)$, $U_2\varphi(B) = \varphi(-B)$ for Taylor-expandable f and φ.

In order to give representations for $\mathcal{P}$, we have only to identify $A \to \mathbf{x}$, $B \to \mathbf{P}$ in Eq. (2). For time reversal, one has similarly,

$$T_i = \exp(i\pi\Omega_i') \exp(i\pi S_2)K, \tag{4}$$

where K is complex conjugation; S_2 is the y-component of spin, e.g. $\frac{1}{2}\left(\begin{smallmatrix}0 & -i\\ i & 0\end{smallmatrix}\right)$ for spin $\frac{1}{2}$ particles[7]; and Ω_i' are obtained from (1) by the substitutions[8] $A \to t$, $B \to i(\partial/\partial t)$. Finally for inversion, we have

$$I_i \sim \exp(i\pi\Omega_i') \exp(i\pi S_2)K, \tag{5}$$

where $A \to x_\mu = (x, it)$, $B \to P_\mu = (\mathbf{P}, iP_0)$. In this case I_1, I_2, I_3 are all unitary.

It is plausible that one can extend these considerations to the case of relativistic wave equations and to the free fields. The main difficulty that one encounters there is the interpretation of a position operator. Similarly one can find a representation in

[5] Recall that in nonrelativistic quantum mechanics canonical transformations are also unitary. This is of course not the case in a many particle theory, where the latter form only a small subclass of the former. In a relativistic field theory charge conjugation is a unitary operator.

[6] I am indebted to Professor E. Merzbacher for first pointing this out to me for the case $A = X$, $P = -i\partial/\partial x$; that U_3 behaves as a parity operator was earlier brought to my attention by Professor E. C. G. Sudarshan.

[7] An antilinear DSO always involves time reversal and change of sign of spin; Wigner has explicitly constructed this spin part; E. P. Wigner, *Group Theory and Its Applications to Quantum Mechanics of Atomic Spectra* (Academic Press, Inc., New York, 1959).

[8] Here the question might be raised about the Hermiticity of $-i\partial/\partial t$; however, this is ensured by the Schrödinger equation and the definition of probability.

terms of spherical polar coordinates. However these matters will not be pursued here.

III. DISCRETE SYMMETRY OPERATORS IN RELATIVISTIC QUANTUM MECHANICS: FREE FIELDS[9]

The object of this section is to find formal expressions for discrete symmetry operators which can also be used for the interacting fields in quantum field theory. As remarked in the introduction, the explicit expressions for the DSO's in terms of creation and destruction operators in linear momentum representation are given in $(M1)^2$; however, these can not be directly used for interacting fields as the frequency splitting of the (interacting) field operators, and their relationship with the corresponding frequency parts of the free fields is in general not defined. On the other hand, if symmetry operators were given in terms of the free field operators (rather than the positive and negative frequency parts), one could easily extend these expressions to interacting fields; as the required symmetry properties of free and coupled fields are same at a given instant of time, one has only to replace free fields by coupled fields to obtain the required representations. We shall therefore deal in this section with representations of DSO's in terms of free field operators. In order to construct these, one could start with expressions in terms of positive and negative frequency parts as given in (M1); however, the replacement of the frequency parts by the field operators and their derivatives involves some difficulties.

To illustrate the difficulties involved, we consider the Hermitian KG field

$$\phi(x) = \int (2\pi)^{-\frac{3}{2}} \frac{d^3k}{(2\omega)^{\frac{1}{2}}} \times (b_{\mathbf{k}} \exp(ikx) + b_{\mathbf{k}}^+ \exp(-ikx)), \quad (6)$$

where $kx = \mathbf{k}\cdot\mathbf{x} - \omega t$, $\omega = +(\mathbf{k}^2 + m^2)^{\frac{1}{2}}$; the commutation relations are

$$[b_{\mathbf{k}}, b_{\mathbf{k}'}^+]_- = \delta^3(\mathbf{k} - \mathbf{k}'), \qquad [\phi(x), \phi(x')]_- = D(x - x'), \quad (7)$$

where $D(x)$ is the Pauli–Jordan commutation function. It is obvious that the only nontrivial symmetry operators that one can construct are $\mathcal{P}$ and T, for there is only one "degree of freedom",[10] viz. momentum (i.e. only momentum changes sign for both operations, as spin and charge are zero). Furthermore, since a time-reversal operator (in a particular representation) may always be written as a unitary factor and complex conjugation, it is clear that in this case the unitary factor of T is essentially the same as $\mathcal{P}$. In terms of the $b_{\mathbf{k}}$'s a representation of $\mathcal{P}$ (see M_1) is

$$\exp\left(\tfrac{1}{2}i\pi \int d^3k b_{\mathbf{k}}^+ b_{\mathbf{k}}\right) \times \exp\left(-\tfrac{1}{2}i\pi \int d^3k b_{\mathbf{k}}^+ b_{\mathbf{k}}\right) \exp(-i\alpha_P Q), \quad (8)$$

where Q is the operator of generalized charge, which in this case is zero; the next factor involves the number operator, and the last factor on the left is responsible for the change $\mathbf{k} \to -\mathbf{k}$, and is therefore, referred to as a "discrete factor." Taken by themselves, each of the first two factors in (8) transforms positive and negative frequency parts differently by giving them different phases; in this sense they are nonlocal. In fact the number-operator densities at two different spatial points do not commute with each other, nor does it (the number-operator density) commute with local operators, such as Hamiltonian density and momentum density. With this is connected the fact that whereas the local operators such as the Hamiltonian, charge and momentum densities may always be constructed from the field operators, for the number operator one must use only the positive and negative frequency parts, because expressions of the type

$$\int (\phi\dot{\phi} - \dot{\phi}\phi)\, d^3x = i \int d^3k(-b_{\mathbf{k}} b_{\mathbf{k}}^+ + b_{\mathbf{k}}^+ b_{\mathbf{k}}) \quad (9)$$

are c-members. Similarly it does not seem possible to construct $\int d^3k b_{\mathbf{k}}^\dagger b_{-\mathbf{k}}$ from the field operators alone, as

$$\int d^3x \{\phi(\mathbf{x}, t)\dot{\phi}(-\mathbf{x}, t) - \dot{\phi}(\mathbf{x}, t)\phi(-\mathbf{x}, t)\} = i \int d^3k(b_{\mathbf{k}} b_{-\mathbf{k}}^+ - b_{\mathbf{k}}^+ b_{-\mathbf{k}}) \quad (10)$$

vanishes identically. This difficulty also arises in case of the number operator (and some DSO's) for the non-Hermitian KG and Dirac fields. Federbush and Grisaru[4] have suggested "number ordering," in

[9] The transformation properties of the field operators under the various DSO's are summarized in Table II of (M1). In what follows, the units are so chosen that $\hbar = c = 1$; m stands for mass; $(\ldots)^*$ = complex conjugate of $(\ldots)$; $(\ldots)^\sim$ = transposed of all c- and q-numbers $(\ldots)$; $(\ldots)^+$ = Hermitian conjugate of $(\ldots)$; subscript T = transposed of γ_μ matrices and Dirac spinors, and reversal of factors; superscript T = transformed operator or vector; H.C. in $[(\ldots) +$ H.C.] implies Hermitian conjugate of $(\ldots)$; dot in $\dot{\alpha}$ means $d\alpha/dt$.

[10] From the point of view of the DSO's, the pertinent observables are those that change sign under a DSO. The use of the expressions such as 'the only "degrees of freedom," ' 'the number of "observables," ' etc., should be taken to mean the number of *commuting observables that change sign* (COTCS) under a DSO.

which one uses sign convention opposite to that used in normal[11] ordering; this however violates the commutation relations.[4] Thus it appears not possible to construct a representation of $\mathcal{P}$ and T in terms of the field operators for the Hermitian KG field.

For the non-Hermitian KG field, because of the additional degree of freedom of (generalized) charge, one can construct representations of $\mathcal{P}$ and T:

$$\mathcal{P} = \exp(\tfrac{1}{2}\pi\Omega_P)\exp(\tfrac{1}{2}i\pi Q)\exp(i\alpha_P Q), \quad (11)$$

$$T = \exp(-\tfrac{1}{2}\pi\Omega_\tau)\exp(\tfrac{1}{2}i\pi Q)\exp(i\alpha_\tau Q), \quad (12)$$

where $\Omega_P = \int d^3x\{\phi^+(\mathbf{x}, t)\dot{\phi}(-\mathbf{x}, t) - H\cdot C\cdot\}$,

$$\Omega_\tau = \int d^3x\{\phi^+(\mathbf{x}, t)\dot{\phi}^*(\mathbf{x}, -t) - \dot{\phi}^*(\mathbf{x}, -t)\phi^+(\mathbf{x}, t)\};$$

the quantities $\eta_P = \exp(i\alpha_P)$, $\eta_\tau = \exp(i\alpha_\tau)$ are arbitrary phase factors of modulus unity; and $Q = i\int d^3x(\phi^+\dot{\phi} - \dot{\phi}^+\phi)$ is the operator of generalized charge. In addition there are nontrivial operators for $\mathcal{C}$, S, $\mathcal{R}$, and W. However, due to the difficulties mentioned in the last paragraph, it has not been found possible to construct these.

If one further introduces the degree freedom of spin, it turns out that the remaining DSO's can also be expressed in terms of field operators. For this we consider the Dirac field

$$\psi(x) = \int d^3k(16\pi^3\omega\Delta)^{-\frac{1}{2}}\sum_{s=1,2}\{a_{\mathbf{k}}(s, 1)u_{\mathbf{k}}(s) \times \exp(ikx) + a^+_{\mathbf{k}}(s, 2)v_{\mathbf{k}}(s)\exp(-ikx)\}, \quad (13)$$

where labels $r = 1, 2$ refer to particle, antiparticles; s refers to spin projection; Δ is an arbitrary normalization which we choose as $\frac{1}{2}\omega$; and the spinors u and v obey the relations

$$\sum_s u_{\mathbf{k}\alpha}(s)\bar{u}_{\mathbf{k}\beta}(s) = -\Delta(ik_\mu\gamma_\mu + m)_{\alpha\beta}, \qquad \sum_s v_{\mathbf{k}\alpha}(s)\bar{v}_{\mathbf{k}\beta}(s) = -\Delta(ik_\mu\gamma_\mu - m)_{\alpha\beta}, \quad (14)$$

$$\bar{u}_{\mathbf{k}}(s)u_{\mathbf{k}}(s') = -\bar{v}_{\mathbf{k}}(s)v_{\mathbf{k}}(s') = 2m\Delta\delta_{ss'}, \qquad u^+_{\mathbf{k}}(s)u_{\mathbf{k}}(s') = v^+_{\mathbf{k}}(s)v_{\mathbf{k}}(s') = 2\omega\Delta\delta_{ss'}, \quad (15)$$

$$\bar{u}_{\mathbf{k}}(s)v_{\mathbf{k}}(s') = \bar{v}_{\mathbf{k}}(s)u_{\mathbf{k}}(s') = 0, \quad \text{for all } s, s', \qquad \bar{u}\gamma_\mu u = \bar{v}\gamma_\mu v = -2i\Delta k_\mu, \quad k_4 = i\omega, \quad (16)$$

where the adjoint spinors $\bar{u}$, $\bar{v} = u^+\gamma_4$, $v^+\gamma_4$ occur in the development of the adjoint function $\bar{\psi} = \psi^+\gamma_4$.

[11] Note that for the free fields, normal ordering becomes necessary only for the Hamiltonian and amounts to renormalization. For the charge operator one symmetrizes (Bose), or antisymmetrizes (Fermi) the expressions in field operators. That, incidently, is connected with why people used to think in the olden days that spin–statistics connection had to do with particle conjugation.

It is clear from the above relations that u and v form a complete set only with respect to their adjoints and not with respect to their Hermitian conjugates.[3] In addition we have the relations

$$\bar{v}_{\mathbf{k}} = C\widetilde{u^*_{\mathbf{k}}} \equiv C_1 u^*_{\mathbf{k}}, \qquad C\gamma^*_\mu C^{-1} = -\gamma_\mu; \quad (17)$$

$$\tau u_{-\mathbf{k}}(s) = (-1)^{s'}u^*_{\mathbf{k}}(s'), \qquad \gamma_5 u_{\mathbf{k}}(s) = (-1)^{s'}v_{\mathbf{k}}(s'), \quad (18)$$

where s ($\neq s'$) refers to spin projection along k_3, and the symmetric matrix $C_1 = -\gamma_4 C$; $\tau = C^+\gamma_4$, $\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$.

We note that as in the case of KG fields the number operator cannot be constructed for the Dirac field for the same reasons; the best one can do is to have

$$\int(\bar{\psi}\psi - \psi_T\bar{\psi}_T)d^3x = \int\frac{m}{\omega}d^3k\sum_{s,r}a^+_{\mathbf{k}}(s, r)a_{\mathbf{k}}(s, r), \quad (19)$$

which is not a number operator due the factor $1/\omega$. A similar difficulty arises in case of $\mathcal{C}$ and S. Thus for $\mathcal{C}$, $\int\psi^+\psi_{\mathcal{C}}\,d^3x = 0$, where $\psi_{\mathcal{C}} \equiv C\bar{\psi}_T$; also,

$$\int(\bar{\psi}\psi_{\mathcal{C}} - \psi_{\mathcal{C}T}\bar{\psi}_T)d^3x = \int\frac{m}{\omega}\sum_{s,r\neq r'}d^3k(-1)^{r+1}a^+_{\mathbf{k}}(s, r)a_{\mathbf{k}}(s, r'), \quad (20)$$

where again the presence of $1/\omega$ spoils everything.[12] In the case of strong reflection, $\int\psi^+\psi^*_S\,d^3x = 0$, and $\int(\bar{\psi}\psi^*_S - \psi_T\bar{\psi}^*_{ST})\,d^3x = 0$, where $\psi_S \equiv \tau\gamma_4 C\bar{\psi}_T(-x_\mu)$.

However, one can still construct the operators $\mathcal{P}$, T, I, and W directly; then in terms of these one can construct $\mathcal{C}$ and S. We shall briefly summarize the results.

(i) $$\mathcal{P} \equiv \exp(\tfrac{1}{2}i\pi\Omega_P)\exp(\tfrac{1}{2}i\pi Q)\exp(-i\alpha_P Q), \quad (21)$$

where

$$\Omega_P \equiv \int[\bar{\psi}(\mathbf{x}, t)\psi(-\mathbf{x}, t) - \psi_T(-\mathbf{x}, t)\bar{\psi}_T(\mathbf{x}, t)]d^3x;$$

[12] In Ref. 4, the left-hand side of (19) is used as a number operator and the left-hand side of (20) as the 'generator' of $\mathcal{C}$. However, we have seen that these are not the proper operators due to the presence of ω. The reason they obtain a different result is that they introduce an operator w, with the properties $wu = u$, $wv = -v$, $w^2 = 1$. One example for w, given, is in our notation, γ_4. Application of these properties of w to (15) then shows that they imply $m = \omega$! It is only in this case that u and v form a complete set with respect to their Hermitian conjugates.

(ii) $$T \equiv \exp(-\tfrac{1}{2}i\pi\Omega_\tau)\exp(-i\alpha_\tau Q)K, \quad (22)$$

where

$$\Omega_\tau \equiv \tfrac{1}{2}\int d^3x(\psi^+\psi_\tau^* - \psi_{\tau T}^*\psi_T),$$

$$\psi_\tau = \tau\psi(\mathbf{x}, -t);$$

(iii) $$I = \exp(-\tfrac{1}{2}\pi\Omega_I)\exp(-i\alpha_I Q)K, \quad (23)$$

where

$$\eta_I = \exp(i\alpha_I), \qquad \Omega_I = \int \psi^+\psi_I^* d^3x$$

and

$$\psi_I = \tau\gamma_4\psi(-\mathbf{x}, -t);$$

(iv) $$\mathfrak{R} = \exp(\tfrac{1}{2}i\pi\Omega_R) \times \exp(-\tfrac{1}{2}i\pi Q)\exp(-i\alpha_R Q), \quad (24)$$

where

$$\Omega_R = \frac{1}{4}\int \{(\psi^+\psi_R - \psi_{RT}\psi_T^+) + H\cdot C\cdot\}d^3x,$$

and

$$\psi_R = \gamma_4 C\bar{\psi}_T(-\mathbf{x}, t) = [\psi^+(-\mathbf{x}, t)C]_T;$$

(iv)′ $$\mathfrak{R} = \exp(-\tfrac{1}{2}\pi\Omega_R')\exp(-i\alpha_R Q), \quad (25)$$

where

$$\Omega_R' = \frac{1}{4}\int [(\psi^+\psi_R - \psi_{RT}\psi_T^+) - H\cdot C\cdot]d^3x;$$

(v) $$W = \exp(\tfrac{1}{2}i\pi\Omega_W) \times \exp(-\tfrac{1}{2}i\pi Q)\exp(-i\alpha_W Q)K, \quad (26)$$

where

$$\Omega_W = \frac{1}{2}\int [(\psi^+\psi_W^* - \psi_{WT}^*\psi_T^+) + H\cdot C\cdot]d^3x$$

and

$$\psi_W = \tau C\bar{\psi}_T(\mathbf{x}, -t);$$

(v)′ $$W = \exp(-\tfrac{1}{2}\pi\Omega_W')\exp(-i\alpha_W Q), \quad (27)$$

where

$$\Omega_W' = \frac{1}{4}\int [(\psi^+\psi_W^* - \psi_{WT}^*\psi_T^+) - H\cdot C\cdot]d^3x;$$

(vi) Particle conjugation; using Ω_τ and Ω_W:

$$\mathfrak{C} = \exp[\tfrac{1}{2}\pi(\Omega_\tau \pm i\Omega_W)] \times \exp(\mp\tfrac{1}{2}i\pi Q)\exp(-i\alpha_c Q), \quad (28)$$

$$\mathfrak{C} = \exp[\pm\tfrac{1}{2}\pi(\Omega_\tau - \Omega_W')]\exp(-i\alpha_c Q);$$

similarly, one can construct $\mathfrak{C}$ from Ω_P and Ω_R;

(vii) Strong reflection; this can be constructed from any of the following sets of operators: $(\mathcal{P}, W)$, $(T, \mathfrak{R})$, and (I, T, W)| as an illustration we construct from $\mathcal{P}$ and W:

$$S = \exp[\tfrac{1}{2}\pi(\Omega_W' \pm i\Omega_P)] \times \exp(\mp\tfrac{1}{2}i\pi Q)\exp(-i\alpha_S Q)K, \quad (29)$$

$$S = \exp[\tfrac{1}{2}i\pi(\Omega_W + \Omega_P)]\exp(-i\alpha_S Q)K. \quad (30)$$

We remark in conclusion that it appears from the above constructions that for Bose fields $\mathfrak{C}$ is not a "local" operator; and $\mathcal{P}$, T are both "non-local" for the Hermitian KG field.

IV. DISCRETE SYMMETRY OPERATORS IN RELATIVISTIC QUANTUM MECHANICS: A FORMULATION INDEPENDENT OF A FIELD CONCEPT

In this section we seek to develop a method by which one can write down explicit expressions for symmetry operators without reference to either field operators or their positive and negative frequency parts, using mainly the general properties such as their effect on the observables. That is, the method is more or less purely algebraic. At the outset let us state that no claim to rigor is made; the approach is quite pedestrian. However, it is hoped that it would be possible to put it on a more rigorous basis at the later date.

To begin with we remark that symmetry operators commute apart from a phase factor; consequently, one can factor a symmetry operator into a factor that essentially corresponds to a phase transformation and a factor (to be referred to as a "discrete factor") that is responsible for the discrete change. These discrete factors will be unitary and may be written as $\exp(i\beta\Omega_u)$, where β is a real number (generally $\frac{1}{2}\pi$) and Ω_u is a Hermitian operator to be referred to as a "discrete generator" (DG) of the DSO U. Now, clearly the "discrete factors," and (hence?) also the "discrete generators" of different DSO's commute. This follows from the fact that the identity operator and a DSO together constitute a group. However, since we are considering all possible DSO's connected with space, time and charge, one DSO can always be written as a product of others; as a result it should be possible to relate DG's of different DSO's. To establish this connection, we assume that each DG can be expanded in terms of operators with a certain complete set of labels; thus, e.g., for a spinless, chargeless particle, we write $\Omega_u = \int \alpha^{\mathbf{k}}\, d^3k$. We then obtain two results:

(i) Integrand operators, such as, $\alpha^{\mathbf{k}}$ for different DG's satisfy a Lie algebra such as: Eq. (40);

(ii) For a particular DSO, there are more than one DG's; these DG's and the *commuting* set of *observables* that *change sign* (COTCS) under this DSO generate a Lie algebra which is same as the Lie algebra of the generators of rotation in a Euclidean $\mathfrak{N}$-space-the number of COTCS under a DSO being $(\mathfrak{N}-2)$. Next we determine the "phase" transformations which clearly bring out the distinction between integral and half integral spin fields.

We shall work throughout in the linear momentum representation; however, the developments can be carried over into any other representation *mutatis mutandis*.

We first consider the case of a particle whose field analog is a Hermitian KG field. The simplicity of this case lies in the fact that there is just one "observable" (COTCS) viz. linear momentum, that can possibly change sign under a DSO. The DSO's involved are just $\mathcal{P}$ and T. We start by writing

$$\mathcal{P} = \exp(\tfrac{1}{2}i\pi\Omega_P) \times \text{a phase transformation}, \quad (31)$$

where the first factor is a "discrete factor," and the Hermitian operator Ω_P is the "discrete generator". Now in field theory the observables can in general be written, e.g. in the coordinate representation, as integrals of certain "densities" in coordinate space; these densities are expressed in terms of fields functions (which are operator valued functions of coordinates) and their derivatives. Since one assumes that the Fourier transforms of field functions exist, one can also express the observables as integrals over certain operator valued functions in momentum space ($\mathbf{k}$-space). Following this clue from field theory, *we shall assume that one can always write the Hermitian "generators", such as Ω_P, as integrals of certain operator valued functions.* In particular, in the discrete linear momentum representation, we write $\Omega_P = \sum_{\mathbf{k}} \alpha^{\mathbf{k}}$, where the $\alpha^{\mathbf{k}}$ are, in general, non-Hermitian; to insure the Hermiticity of Ω_P, one puts the condition $\alpha^{\mathbf{k}} = (\alpha^{-\mathbf{k}})^+$. Since $k_1, k_2, \cdots$ form a complete set of labels, one can consider $\alpha^{\mathbf{k}}$ independent for different $\mathbf{k}$; then it is clear that the commutator $[\alpha^{\mathbf{k}}, \alpha^{\mathbf{k}'}]_-$ is proportional to $\delta_{\mathbf{k},-\mathbf{k}'} \equiv \hat{\delta}_{\mathbf{k},\mathbf{k}'}$. Furthermore, the commutator is Hermitian and antisymmetric with respect to $\mathbf{k}$, $\mathbf{k}'$, so that one can write

$$[\alpha^{\mathbf{k}}, \alpha^{\mathbf{k}'}]_- = \hat{\delta}_{\mathbf{k}\mathbf{k}'}(\alpha_{\mathbf{k}} - \alpha_{\mathbf{k}'}), \quad (32)$$

where $\alpha_{\mathbf{k}}$ are Hermitian; the notation has been so chosen that the upper index labels non-Hermitian operators, nondiagonal in this index, and the lower index implies Hermiticity with respect to this index. We should emphasize at this point that in obtaining (32) we have made two essential assumptions which are not altogether independent, viz. (1) one can always write a generator as an integral or a sum over certain operator densities, thus e.g., $\Omega_P = \sum_{\mathbf{k}} \alpha^{\mathbf{k}}$; (2) the $\alpha^{\mathbf{k}}$ for different $\mathbf{k}$ are independent, except for the connection $\alpha^{-\mathbf{k}} = (\alpha^{\mathbf{k}})^+$, which arises from the Hermiticity of Ω_P. Alternately one can take (32) as an *Ansatz.*[13]

In order to evaluate $[\alpha^{\mathbf{k}}, \alpha_{\mathbf{k}'}]$, take the commutator of (32) with $\alpha^{\mathbf{k}''}$ and $\alpha_{\mathbf{k}''}$; we shall not do that. Instead note that, in virtue of (32), $[\alpha^{\mathbf{k}}, \alpha_{\mathbf{k}'}]_-$ is nonzero for both $\mathbf{k} = \mathbf{k}'$ and $\mathbf{k} = -\mathbf{k}'$ and may therefore be taken as proportional to $(\hat{\delta}_{\mathbf{k}\mathbf{k}'} - \delta_{\mathbf{k}\mathbf{k}'})$. Moreover, since $\alpha_{\mathbf{k}}$ are Hermitian and $\alpha^{\mathbf{k}}$ non-Hermitian, the commutator will be non-Hermitian, so that, one can write

$$[\alpha^{\mathbf{k}}, \alpha_{\mathbf{k}'}]_- = (\hat{\delta}_{\mathbf{k}\mathbf{k}'} - \delta_{\mathbf{k}\mathbf{k}'})\alpha^{\mathbf{k}}, \quad (33)$$

which may be verified to be consistent with (32). Using relations (32) and (33), one can show that under the similarity transformation

$$\mathcal{P} \sim \exp(\tfrac{1}{2}i\pi\Omega_P), \quad (34)$$

the operators $\alpha^{\mathbf{k}}$, $\alpha_{\mathbf{k}}$ are transformed into $\alpha^{-\mathbf{k}}$, $\alpha_{-\mathbf{k}}$, respectively. Hence, for Ω_P to be a "discrete generator," the suitable definition of the momentum operator is

$$\mathbf{P} = \sum{}'_{\mathbf{k}} \mathbf{k}[\alpha^{\mathbf{k}}, \alpha^{-\mathbf{k}}]_- = \sum{}_{\mathbf{k}} \mathbf{k}\alpha_{\mathbf{k}}, \quad (35)$$

where the prime on $\sum$ denotes summation over half the $\mathbf{k}$-space only, say $k_3 > 0$. It is clear from (35) that, to obtain correspondence with field theory, one must identify $\alpha_{\mathbf{k}}$ with $n_{\mathbf{k}}$—the $\mathbf{k}$-space particle-number density. Similarly one can write for the time-reversal operator

$$T \sim \exp(\tfrac{1}{2}i\pi\Omega_\tau)K, \qquad \Omega_\tau \equiv \Omega_P, \quad (36)$$

which clearly anticommutes with (35). We remark that Ω_P is not the only discrete generator of $\mathcal{P}$. Let us define the "momentum density" by $\hat{\mathbf{p}} = \sum'_{\mathbf{k}} (\alpha_{\mathbf{k}} - \alpha_{-\mathbf{k}})$; then the operators $\Omega_1^P \equiv \frac{1}{2}\Omega_P$, $\Omega_2^P \equiv -i[\Omega_3^P, \Omega_1^P]_-$, $\Omega_3^P \equiv \frac{1}{2}\hat{\mathbf{p}}$ satisfy the angular momentum-type commutation relations[14]

[13] This point has been stressed by the referee who points out 'that in free field case the observables are not only integrals over densities "which can be expressed" in terms of field functions, but that the densities are *bilinears* in the field functions, which in turn obey the commutation relations of the form $[A, B]_- = c$-number, i.e. *which reduce their degree by two.* It is because of these two properties that a commutator in the observables becomes *linear* in the observables, as in (32), and leads to a Lie algebra.'

[14] E. C. G. Sudarshan [Proc. Indian Acad. Sci., Sect. B, 49, 66 (1959)] has discussed the case of representation of $\mathcal{P}$ for spin $\frac{1}{2}$ case from the same point of view. He remarks that the representations of discrete generators for $\mathcal{P}$ for spin $\frac{1}{2}$ case are unique in that they are *same* as the generators of rotation group in three dimensional coordinate space. In this connection, see also E. Cartan, Leçons sur la theorié de spineurs, Herman, Paris (1939). In our case "discrete generators" for $\mathcal{P}$ are generators of rotation in an abstract three-dimensional space and this holds for both integer and $\frac{1}{2}$ integer spins.

$$[\Omega_l^P, \Omega_m^P]_- = i\epsilon_{lmn}\Omega_n^P, \qquad (l, m, n = 1, 2, 3), \quad (37)$$

where ϵ_{lmn} is the Levi-civita symbol; hence, exp $(i\pi\Omega_2^P)$ is an equally good "discrete factor." This may also be verified directly.

The case of non-Hermitian KG field is analogous. Here there are two COTCS's, viz. momentum $\mathbf{P}$, and the generalized charge Q. Hence, in analogy with the Hermitian case, we expand $\Omega_P = \sum_{\mathbf{k},r} \alpha_r^{\mathbf{k}}$, $\Omega_{\mathcal{C}} = \sum_{\mathbf{k},r} \alpha_{\mathbf{k}}^r$, where $r = 1, 2$ denote particle and antiparticle, and $(\alpha_r^{\mathbf{k}})^+ = \alpha_r^{-\mathbf{k}}$, $(\alpha_{\mathbf{k}}^r)^+ = \alpha_{\mathbf{k}}^{r'}$, $r \neq r'$. Then using the arguments analogous to those employed for setting up commutation relations (32, 33), we arrive at

$$[\alpha_r^{\mathbf{k}}, \alpha_{r'}^{\mathbf{k}'}]_- = \delta_{rr'}\hat{\delta}_{\mathbf{k}\mathbf{k}'}(\alpha_{\mathbf{k}r} - \alpha_{\mathbf{k}'r'}) \quad (38)$$

$$[\alpha_{\mathbf{k}}^r, \alpha_{\mathbf{k}'}^{r'}]_- = \hat{\delta}_{rr'}\delta_{\mathbf{k}\mathbf{k}'}(\alpha_{\mathbf{k}r} - \alpha_{\mathbf{k}'r'}), \quad (39)$$

where $\hat{\delta}_{rr'} = 1$ for $r \neq r'$ and zero otherwise. Since it follows from the group property that the operator $\alpha^{\mathbf{k}r}$ in the expansion of Ω_R is proportional to the commutator of $\alpha_r^{\mathbf{k}}$ and $\alpha_{\mathbf{k}}^r$, a careful analysis leads to

$$[\alpha_r^{\mathbf{k}}, \alpha_{\mathbf{k}'}^{r'}]_- = \alpha^{\mathbf{k}r}\hat{\delta}_{\mathbf{k}\mathbf{k}'}\delta_{rr'} - \alpha^{\mathbf{k}'r'}\delta_{\mathbf{k}\mathbf{k}'}\delta_{rr'}; \quad (40)$$

$$[\alpha^{\mathbf{k}r}, \alpha^{\mathbf{k}'r'}]_- = \hat{\delta}_{\mathbf{k}\mathbf{k}'}\hat{\delta}_{rr'}(\alpha_{\mathbf{k}r} - \alpha_{\mathbf{k}'r'}). \quad (41)$$

One has further with obvious symmetry

$$[\alpha_r^{\mathbf{k}}, \alpha^{\mathbf{k}'r'}]_- = \hat{\delta}_{\mathbf{k}\mathbf{k}'}(\delta_{rr'}\alpha_{\mathbf{k}}^r - \hat{\delta}_{rr'}\alpha_{\mathbf{k}'}^{r'}), \quad (42)$$

$$[\alpha^{\mathbf{k}r}, \alpha_{\mathbf{k}'}^{r'}]_- = \hat{\delta}_{rr'}(\hat{\delta}_{\mathbf{k}\mathbf{k}'}\alpha_r^{\mathbf{k}} - \delta_{\mathbf{k}\mathbf{k}'}\alpha_{r'}^{\mathbf{k}'}). \quad (43)$$

The observables $\mathbf{P}$, Q may be defined in terms of these (as before), in more than one way; thus,

$$\mathbf{P} = \sum_{\mathbf{k}}' \mathbf{k}[\alpha_r^{\mathbf{k}}, \alpha_r^{-\mathbf{k}}]_- = \sum_{\mathbf{k}}' \sum_{r\neq r'} \mathbf{k}[\alpha^{\mathbf{k}r}, \alpha^{-\mathbf{k}r}]_-, \quad (44)$$

$$Q = \sum_{\mathbf{k}} (-1)^{r+1}[\alpha_{\mathbf{k}}^r, \alpha_{\mathbf{k}}^{r\neq r'}]_- = \sum_{\mathbf{k}} (-1)^{r+1}[\alpha^{\mathbf{k}r}, \alpha^{-\mathbf{k}r'\neq r}]_-. \quad (45)$$

Again $\alpha_{\mathbf{k}r}$ may be identified with a number operator to obtain correspondence with field theory.

As in the Hermitian case, corresponding to $\mathcal{P}$, there are two "discrete generators"

$$\Omega_1^P \equiv \tfrac{1}{2}\Omega_P, \qquad \Omega_2^P = -i[\Omega_3^P, \Omega_1^P]_-,$$

where

$$\Omega_3^P \equiv \tfrac{1}{2}\sum_{\mathbf{k}}' \sum_r [\alpha_r^{\mathbf{k}}, \alpha_r^{-\mathbf{k}}].$$

Similarly corresponding to $\mathcal{C}$, which changes only the sign of Q, the "discrete generators" are

$$\Omega_1^C \equiv \tfrac{1}{2}\Omega_C, \qquad \Omega_2^C \equiv -i[\tfrac{1}{2}Q, \Omega_1^C];$$

these together with $\Omega_3^C = \frac{1}{2}Q$ obey the commutation rules such as (37). In case of $\mathcal{R}$, there are two COTCS' viz. $\mathbf{P}$ and Q. Instead of $\mathbf{P}$ consider $\hat{\mathbf{p}}$; let

$$\Omega_1^R \equiv \tfrac{1}{2}\Omega_R, \qquad \Omega_3^R \equiv \tfrac{1}{2}Q, \qquad \Omega_6^R \equiv \tfrac{1}{2}\hat{\mathbf{P}}.$$

These operators generate a Lie algebra of six elements

$$[\Omega_l^R, \Omega_m^R]_- = if_{lmn}\Omega_n^R \qquad (l, m, n = 1, 2 \cdots 6), \quad (46)$$

where f_{lmn} is completely antisymmetric and $f_{123} = f_{561} = f_{345} = f_{642} = +1$.

It is clear from the above examples that whereas in case of $\mathcal{P}$ and $\mathcal{C}$, only one "observable" changes sign ($\mathbf{P}$ and Q, respectively), and therefore the Lie algebra of their generators is the same as that of the rotation group in a Euclidean three-space, in case of R where there are two COTCS, the Lie algebra of the generators is the same as in a Euclidean four-space. One can generalize this result by induction. Consider a DSO which changes the sign of n commuting set of observables. The first DG may be written as $\Omega_1 = \frac{1}{2}\sum \alpha^{a_1 a_2 \cdots a_n}$, where each of the a_i has two values; set up commutation relations of the type (41) and write down the expressions for the n COTCS by the method of Eqs. (44) and (45). Find $\frac{1}{2}n(n+1)$ commutators of Ω_1 with n COTCS; these together with Ω_1 furnish $\frac{1}{2}n(n+1)+1$ DG's. The $(\frac{1}{2}n+1)(n+1)$ operators form a Lie algebra isomorphic to the Lie algebra of generators of rotation in a Euclidean $(n+2)$-space.[15] The rank of the Lie group corresponding to the Lie algebra is obviously n.

To show that our conclusion regarding the algebra of the "generators" of a DSO is correct, we consider the example of the inversion operator for spin-$\frac{1}{2}$ particles. Under inversion the only "observable" that changes sign is S_3. By Wigner's construction the representation is

$$I \sim \exp(i\pi S_2)K. \quad (47)$$

Since $[S_3, S_1] = iS_2$, one can also write[16]

$$I \sim \exp(i\pi S_1)\exp(i\pi S_3)K. \quad (48)$$

As S_3 is diagonal, S_1 is as good a candidate for a discrete generator as S_2; S_1 and S_2 are therefore the only two 'discrete generators' of I, and rank of the group is one.

[15] In other words, corresponding to a DSO U, its n COTCS' and $\frac{1}{2}n(n+1)+1$ DG's together form a Lie algebra such as (46), which is isomorphic to the algebra of generators of rotation in a Euclidean $(n+2)$-space. The corresponding Lie groups are locally isomorphic in the sense that they have the same covering group.

[16] Where use has been made of the Baker-Hausdorff formula; see e.g. G. H. Weiss and A. A. Maradudin, J. Math. Phys. **3**, 771 (1962), where other references are given. As in the present case since we are not dealing with a 'free' group one must be careful in making any unrestricted use of the "polar operator" of Hausdorff.

One can analogously discuss other operators for the spin-half case. In particular we note that for strong reflection the corresponding Euclidean space is 4-dimensional and for W it is 5-dimensional. For the sake of completeness we give the necessary relations in terms of the α's for the spin half case. We postulate the operators $\alpha^{\mathbf{k}}_{sr}$, $\alpha^{r}_{\mathbf{k}s}$ and $\alpha^{sr}_{\mathbf{k}}$, where $s = 1, 2$ refers to the spin projections $+\frac{1}{2}$, $-\frac{1}{2}$ along the z axis. The self commutators for these are

$$[\alpha^{\mathbf{k}}_{sr}, \alpha^{\mathbf{k}'}_{s'r'}]_- = \hat{\delta}_{\mathbf{k}\mathbf{k}'}\delta_{ss'}\hat{\delta}_{rr'}(\alpha_{\mathbf{k}sr} - \alpha_{\mathbf{k}'s'r'}), \quad (49)$$

$$[\alpha^{r}_{\mathbf{k}s}, \alpha^{r'}_{\mathbf{k}'s'}]_- = \delta_{\mathbf{k}\mathbf{k}'}\delta_{ss'}\hat{\delta}_{rr'}(\alpha_{\mathbf{k}sr} - \alpha_{\mathbf{k}'s'r'}), \quad (50)$$

$$[\alpha^{sr}_{\mathbf{k}}, \alpha^{s'r'}_{\mathbf{k}'}]_- = \delta_{\mathbf{k}\mathbf{k}'}\hat{\delta}_{ss'}\hat{\delta}_{rr'}(\alpha_{\mathbf{k}sr} - \alpha_{\mathbf{k}'s'r'}). \quad (51)$$

One can define $\alpha^{\mathbf{k}sr}$, $\alpha^{\mathbf{k}s}$, and $\alpha^{r}_{\mathbf{k}s}$ in terms of these as for the non-Hermitian KG field. Similarly, the observables $\mathbf{P}$, Q, S_3 may be written in more than one way. Thus, e.g.

$$\mathbf{P} = \sum'_{\mathbf{k}} \sum_{sr} \mathbf{k}[\alpha^{\mathbf{k}}_{sr}, \alpha^{-\mathbf{k}}_{sr}], \text{ etc.}, \quad (52)$$

$$Q = \sum_{\mathbf{k},s} [\alpha^{1}_{\mathbf{k}s}, \alpha^{2}_{\mathbf{k}s}]_-, \quad (53)$$

$$S_3 = \tfrac{1}{2} \sum_{\mathbf{k},r} [\alpha^{2}_{\mathbf{k}r}, \alpha^{1}_{\mathbf{k}r}]_-. \quad (54)$$

Obviously then

$$S_3 = \tfrac{1}{2} \sum_{\mathbf{k},s,r} (-1)^{s+1}\alpha_{\mathbf{k}sr}, \quad (55)$$

$$S_2 = \tfrac{1}{2}i \sum_{\mathbf{k},s,r} (-1)^{s}\alpha^{s}_{\mathbf{k}r}, \quad (56)$$

$$S_1 = \tfrac{1}{2} \sum_{\mathbf{k},s,r} \alpha^{s}_{\mathbf{k}r}. \quad (57)$$

So far in writing down the representations, we have disregarded any phase factors; we saw that in the approximation that a DSO is replaced by its "discrete factors" there is no essential difference in treatment of Bose and Fermi particles. However, there is a possible effect on the structure of a DSO due to the connection between the intrinsic spin of the particle and the permutation symmetry of a many particle state. Thus, e.g., acting on a one particle state $\mathcal{R}^2$, T^2, W^2 give, respectively, $+1$, $+1$, $+\eta_W^2$ for Bosons and -1, -1, $-\eta_W^2$ for Fermions. Again one finds that

$$[W, \mathcal{P}]_- = 0 \text{ for Bosons and } [W, \mathcal{P}]_+ = 0 \text{ for Fermions.} \quad (58)$$

If one demands[17] that η_s-the arbitrary phase factor associated with strong reflection is real (± 1), then

$$[I, \mathcal{C}]_- = 0 \text{ (Bosons)}, \quad [I, \mathcal{C}]_+ = 0 \text{ (Fermions)},$$

$$[\mathcal{R}, T]_- = 0 \text{ (Bosons and Fermions)}. \quad (59)$$

[17] This is of course implied in Luders' proof of the TCP-theorem. [G. Lüders, Ann. of Phys. (N. Y.) **2**, 1 (1957)] and suggests a connection between multiplicative symmetries and the TCP invariance. Thus see, e.g., G. Feinberg and S. Weinberg, Nuovo Cimento **14**, 571 (1959).

Thus we conclude that the representations without some additional phase factors are not complete. Actually this should be obvious, for we found that for a given DSO, the DG's and the commuting observables that change sign (COTCS') under this DSO generate a Lie algebra; hence, the general representation will be of the type

$$\exp(i\alpha_1 \text{ 'DG'}) \exp(i\alpha_2 \text{ 'COTCS'}). \quad (60)$$

In addition, one can include a factor $\exp(i\beta Q)$ (β = real, arbitrary; Q = generalized charge), which gives rise to an arbitrary phase factor $\eta = \exp(i\beta)$ for each one-particle state,[18] this can always be done as generalized charge is rigorously conserved. In the following we briefly indicate the arguments for finding the phase factors.

1. Since $\mathcal{C}^2 = +1$ for both Bosons and Fermions, a factor such as $\exp(\frac{1}{2}i\pi Q)$ has no significance; hence, for an arbitrary real β,

$$\mathcal{C} \sim \exp(i\pi\Omega_1^{\mathcal{C}}) \exp(i\beta Q). \quad (61)$$

2. For time reversal both $\mathbf{P}$ and S_3 change sign; but since momentum is not an intrinsic property of a particle, the use of $\frac{1}{2}\hat{\mathbf{p}}$ as a generator of a phase transformation is irrelevant. Hence, one can write

$$T \sim \exp(i\pi\Omega_1^{T}) \exp(i\pi S_3) \exp(i\beta Q)K; \quad (62)$$

then clearly $T^2 = \exp(2i\pi S_3)$ as required. Similarly

$$W \sim \exp(i\pi\Omega_1^{W}) \exp(i\pi S_3) \exp(i\beta Q)K, \quad (63)$$

which gives $W^2 = \exp(2i\pi S_3) \exp(i\gamma Q)$.

3. In case of $\mathcal{P}$, for fermions, we know that the parity of a fermion and its antiparticle are opposite.[19] This corresponds to the well known fact that an electron–positron pair in a singlet s-state decays into two photons which have perpendicular polarization.[20] Thus

$$\mathcal{P}_{\text{Fermi}} \sim \exp(i\pi\Omega_1^{P}) \exp(\tfrac{1}{2}i\pi Q). \quad (64)$$

The factor $\exp(\frac{1}{2}i\pi Q)$ will occur in all operators involving space inversion, e.g.

$$\mathcal{R}_{\text{Fermi}} \sim \exp(i\pi\Omega_1^{R}) \exp(\tfrac{1}{2}i\pi Q), \quad (65)$$

$$I_{\text{Fermi}} \sim \exp(i\pi\Omega_1^{I}) \exp(i\pi S_3) \exp(\tfrac{1}{2}i\pi Q). \quad (66)$$

[18] For a neutral particle $\eta_a = \pm 1$, so that parity is ± 1. One generally uses $\eta^2_P = +1$ for Bosons; $+1$ or -1 for fermions; and $\eta_S = \pm 1$ for both bosons and fermions; on the other hand, the phase factors η_C, η_R, η_T, η_I are indeterminate.

[19] This is actually connected with the properties of two-valued representations of the Lorentz group. However, we will not go into this aspect.

[20] R. C. Hanna, Nature **162**, 332 (1948); C. S. Wu and I. Shaknov, Phys. Rev. **77**, 136 (1950). The "charge parity" of the pair is of course $+1$.

It is verified that $\mathcal{R}_F^2 = \exp(i\pi Q)$; $I_F^2 = \exp(2i\pi S_3)$. Similarly for strong reflection we have,

$$S_F = \exp(i\pi\Omega_1^S)\exp(i\pi S_3) \times \exp(\tfrac{1}{2}i\pi Q)\exp(i\beta Q), \quad (67)$$

which satisfies (58) and (59) for $\exp(i\beta)$ real.

Finally we note that since DG's form a Lie algebra with operators like $(\frac{1}{2}Q)$, which have half integral eigenvalues, we expect that each "discrete factor" gives rise to a factor $\pm i$ for each one-particle state. Since for Bosons the factor $\exp(\frac{1}{2}i\pi Q)$ does not occur for $\mathcal{P}$, and S_3 is integer, the net effect is to give an imaginary parity to a one-particle Bose state; this hardly makes for consistency. Furthermore, when operating on a field operator, a discrete factor will give differing signs to positive and negative frequency parts, which is not consistent, since a field operator should transform as a whole. To correct for this one can include the factor $\exp\{\frac{1}{2}i\pi \sum_{k,r}\alpha_{k,r}\}$ for Bosons. This completes the construction.

In conclusion, we remark that if one requires of representations of DSO's only to satisfy their defining relations in terms of the effect on the observables, then the use of the operator $\sum_{k,r}\alpha_{k,r}$ is redundant and the representations are completely defined in terms of the generators of rotation in an appropriate Euclidean space. However, we note that whenever linear momentum changes sign, to be consistent, one must introduce the operator $\frac{1}{2}\mathbf{P}$ in the algebra of "generators" rather than $\frac{1}{2}\hat{\mathbf{P}}$ which is an artificial entity. This difficulty is not present for strong reflection (CPT); here the algebra has six elements, including S_3 and Q, which are both related to the "intrinsic" properties of the particles; the corresponding rotation group is four-dimensional. A possible physical significance is the following. Consider the complex Lorentz group: a set of complex (infinitesimal) linear transformation which leave the complex $(z_i = x_i + iy_i)$ form

$$F = z_1^2 + z_2^2 + z_3^2 - z_4^2 \quad (68)$$

unchanged; it is clear that (68) is also the complex form (up to an isomorphism) of the real form

$$F' = x_1^2 + x_2^2 + x_3^2 + y_4^2, \quad (69)$$

which is its unique (up to an isomorphism) compact form.[21] Since (69) is the Euclidean group, one can label the representations by two numbers; in a particular representation one can take these to be S_z and $\frac{1}{2}Q$; and one can interpret the "discrete generators" of S together with S_z and $\frac{1}{2}Q$ as the generators of rotation in the space defined by the form (69). It then becomes clear, how in the complex group, S is continuously connected to the identity.[22]

TABLE I. Definitions of discrete-symmetry operators in terms of their effect on coordinates ($\mathbf{x}$, t) charge (Q), linear momentum ($\mathbf{P}$), and angular momentum ($\mathbf{J} = \mathbf{L} + \mathbf{S}$).

	U	$\mathbf{x}'$	t'	$\mathbf{P}'$	Q'	$\mathbf{J}'$
1. Identity	$\mathcal{E}$	$+\mathbf{x}$	$+t$	$+\mathbf{P}$	$+Q$	$+\mathbf{J}$
2. Space inversion	$\mathcal{P}$	$-\mathbf{x}$	$+t$	$-\mathbf{P}$	$+Q$	$+\mathbf{J}$
3. Particle conjugation	$\mathcal{C}$	$+\mathbf{x}$	$+t$	$+\mathbf{P}$	$-Q$	$+\mathbf{J}$
4. Reflection	$\mathcal{R}$	$-\mathbf{x}$	$+t$	$-\mathbf{P}$	$-Q$	$+\mathbf{J}$
5. Time reversal	T	$+\mathbf{x}$	$-t$	$-\mathbf{P}$	$+Q$	$-\mathbf{J}$
6. Inversion	I	$-\mathbf{x}$	$-t$	$+\mathbf{P}$	$+Q$	$-\mathbf{J}$
7. Weak reflection	W	$+\mathbf{x}$	$-t$	$-\mathbf{P}$	$-Q$	$-\mathbf{J}$
8. Strong reflection	S	$-\mathbf{x}$	$-t$	$+\mathbf{P}$	$-Q$	$-\mathbf{J}$

ACKNOWLEDGMENT

The author is indebted to the referee for valuable comments.

[21] L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, New Jersey, 1946), especially pp. 270–274. See also A. Grossmann, J. Math. Phys. 1, 424 (1960), Appendix.

[22] To be more precise one has to consider the inhomogeneous Lorentz group so that the particle-antiparticle concept is defined.

Interacting Fermions in One Dimension. II. Attractive Potential*

J. B. McGuire
Department of Physics, Florida Atlantic University, Boca Raton, Florida
(Received 6 May 1965)

The exact energies and wavefunctions for a system of $(N-1)$ one-dimensional fermions all of the same spin and one fermion of the opposite spin are calculated in the large volume, finite density limit, when the particles interact via an attractive delta function potential. It is found that the attractive potential gives rise to a bound state, but, in spite of the presence of this bound state, all of the physical properties which are calculated (ground-state energy, effective mass of a certain class of excitations, etc.) are analytic continuations in the coupling constant of the corresponding results in the repulsive case. In addition, it is possible to have eigenstates which do not have the bound state present. These excited states are also discussed and are found to exhibit a negative effective mass and to modify the particle density at very large distances from the different particle.

I. INTRODUCTION

In this paper we will continue our analysis of the one-dimensional fermion problem, where the fermions interact with an equal-strength delta function potential. This paper is intended as a companion to Ref. 1 (hereafter called I) and also considers the case where $(N-1)$ fermions are spin up and one fermion is spin down. This paper, however, takes the delta function potential to be attractive.

The change from a repulsive to an attractive interaction is brought about by changing the sign of the strength parameter, g, in the Hamiltonian. Making this change we obtain the Hamiltonian which describes the attractive interaction,

$$H = -\frac{\hbar^2}{2M}\sum_i \frac{d^2}{dx_i^2} - g\sum_{i>j}\sum \delta(x_i - x_j).$$

The strength g is assumed to be always positive. Henceforth, we will use units such that $\hbar = M = 1$.

As in I the first task is to obtain the finite density limit from the scattering solutions which were obtained in Ref. 2. The procedure is a two-step process:

1. Apply periodic boundary conditions to a scattering solution for a fixed number of particles.

2. Allow the size of the periodic box and the number of particle to become infinite in such a way that the density is constant.

Step 1 is carried out in exactly the way which is discussed in Secs. I, II, and III of I. The only change which is required to carry this part of the analysis over to the attractive case is the change in the sign of g. The result of step 1 is a set of rules for calculating the wavefunction. We shall restate the rules as they apply to the case of an attractive potential. The notation is the same as in I.

1. The spectrum is given by selecting N Roots of $az + ctnz = \text{const.}$, where $a = 4/gL$, $z = k_iL/2$, and $g > 0$. Subject to the additional constraint

$$\sum z_i = n\pi.$$

2. The wavefunction is given by calculating the amplitudes of $N!$ plane waves in each region of an N-dimensional configuration space. We select as the basic amplitude

$$(1\,2\,3\cdots N)_1 = 1 - e^{ik_1L} = -2ie^{ik_1L/2}\sin k_1L/2 = -2ie^{iz_1}\sin z_1$$

where region 1 is $x_1 < x_2 < x_3 \cdots < x_N$. The amplitude of any other plane wave in this region is determined by two considerations:

(a) The k associated with particle 1 determines which k appears in the numerical expression for the amplitude.

(b) The amplitude is modified by a $\pm$ sign according to whether it is an even or odd permutation of $(1, 2, 3, 4, \cdots N)$. For example,

$$(q\,2\,3\cdots 1\cdots N)_1 = -(1 - e^{ik_qL}) = +2ie^{iz_q}\sin z_q.$$

3. The amplitude of a given plane wave in some other region is calculated by first finding the amplitude of that plane wave in Region 1 and multiplying that amplitude by a factor of e^{2iz_q} for each particle that x_1 has passed. For example, in the region $x_2 < x_3 < x_4 < x_1 < x_5 \cdots < x_N$,

$$(q\,2\,3\,4\cdots 1\cdots N) = -(1 - e^{ik_qL})e^{2i(z_2+z_3+z_4)}.$$

* Work carried out, in part, at the University of California at Los Angeles and with the financial support of the National Science Foundation, and in part supported by the U. S. Atomic Energy Commission, while a Visitor at Brookhaven National Laboratory.

[1] J. B. McGuire, J. Math. Phys. 6, 432 (1965).

[2] J. B. McGuire, J. Math. Phys. 5, 622 (1964).

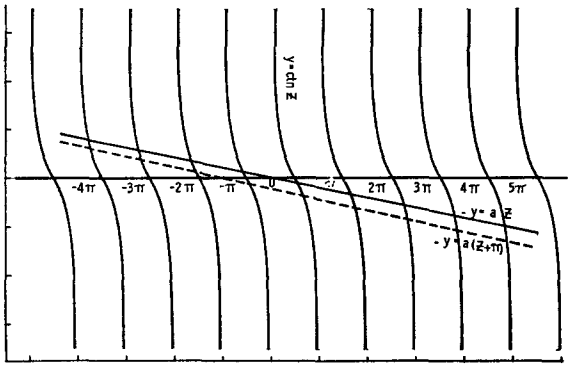

FIG. 1. Sketch of the graphical to the equation $az + \operatorname{ctn} z = 0$.

II. GROUND-STATE ENERGY OF THE SYSTEM

This set of rules allows us to calculate the ground-state energy of the system at a fixed density of particles. Fixed density, of course, implies a fixed background density of spin up particles with one particle spin down.

We must choose N roots of the equation

$$-az - c = \operatorname{ctn} z.$$

Subject to the constraint that the sum of these N roots is $n\pi$. The energy will be given by

$$E = \frac{2}{L^2} \sum_i z_i^2.$$

The constant c must be real, for as we shall later demonstrate, it is related to the total momentum of the system. The ground state of the system will occur when $c = 0$, which we will show explicitly when we analyze the excited states. Therefore we focus our attention on N roots of the equation

$$-az = \operatorname{ctn} z.$$

Figure 1 gives a graphical representation of the roots of this equation, provided the roots are real.

Since we are analyzing an attractive potential we would expect that the delta function bound state would play some role in the ground state of the system. Let us allow z to be complex:

$$z = \alpha + i\beta,$$

$$-a(\alpha + i\beta) = \frac{\tan\alpha(1 - \tanh^2\beta)}{\tan^2\alpha + \tanh^2\beta} - i\frac{\tanh\beta(1 + \tan^2\alpha)}{\tan^2\alpha + \tanh^2\beta}.$$

Separating this relation into real and imaginary parts will give us two transcendental equations with two unknowns. Let us first examine the imaginary part

$$a\beta = \frac{\tanh\beta(1 + \tan^2\alpha)}{\tan^2\alpha + \tanh^2\beta}.$$

Figure 2 is a sketch of the graphical solution of this equation. The function on the right is bounded by $f(\beta) = \tanh\beta$, $(\alpha = 0)$, and $f(\beta) = \coth\beta$, $(\alpha = \pi/2)$. We are interested only in the case where $a \ll 1$, which is the large volume (large L) limit. Thus it is seen that all of the solutions occur where

$$\pm a\beta \cong \tanh\beta \cong \coth\beta \cong 1.$$

Under these conditions the equation for the real part of ctn z is essentially independent of β and is

$$\alpha = 4 \sin\alpha \cos\alpha(e^{-2/a}/a)$$
$$= 0, \text{ in the large volume limit.}$$

Thus we see that there are but two imaginary roots of the spectral equation where $\beta = \pm(1/a)$. The remainder of the roots must be real. This would be expected because the bound state of a delta-function potential is spatially symmetric and with only one spin-down fermion there should only be one possible bound pair.

We now use the technique of I to calculate the energy, using the fact that two of the z's are imaginary,

$$z_1 = i/a = igL/4,$$
$$z_2 = -i/a = -igL/4.$$

All of the positive real roots may be approximated by

$$z_s^+ = s\pi - \operatorname{ctn}^{-1}(sa\pi)$$

to order a. Similarly, the negative real roots are

$$z_s^- = -s\pi + \operatorname{ctn}^{-1}(sa\pi).$$

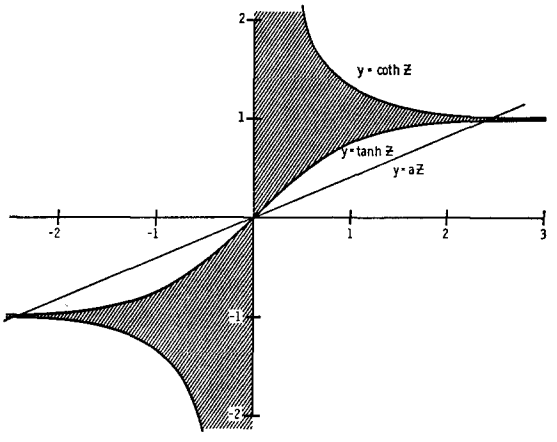

FIG. 2. Sketch of the graphical solution to the equation $a\beta = \tanh\beta(1 + \tan^2\alpha)/(\tan^2\alpha + \tanh^2\beta)^{-1}$.

The energy is

$$E = \frac{2}{L^2}(z_1^2 + z_2^2) + \frac{2}{L^2}\sum_{i>3} z_i^2$$

$$E = -\frac{g^2}{4} + \frac{4}{L^2}\sum_{s=1}^{\frac{1}{2}(N-2)} [(s\pi)^2 - 2s\pi \times \operatorname{ctn}^{-1}(sa\pi)] + O\left(\frac{1}{N}\right).$$

The second term is just the energy of the $(N-1)$-particle background Fermi gas. Let us calculate the energy shift caused by the different particle

$$E - E(N-1) = \Delta E = -\frac{g^2}{4} - \frac{8}{L^2}\sum_{s=1}^{\frac{1}{2}(N-2)} s\pi \operatorname{ctn}^{-1}(sa\pi).$$

Converting the sum to an integral we have

$$\Delta E = -\frac{g^2}{4} - \frac{g^2}{2\pi}\int_0^{2k_F/g} y \operatorname{ctn}^{-1} y\, dy.$$

Where k_F is the Fermi momentum of the background. Performing the integration and rearranging terms gives

$$\Delta E = -\frac{k_F^2}{\pi}\left[\frac{g}{2k_F} + \tan^{-1}\frac{g}{2k_F} + \left(\frac{g}{2k_F}\right)^2\left(\frac{\pi}{2} + \tan^{-1}\frac{g}{2k_F}\right)\right].$$

As $g \to 0$, $\Delta E \to 0$, as it should, for no interaction should cause no energy shift. If g becomes very large,

$$E \cong -(g^2/4) + O(g),$$

which is just the binding energy of a pair of particles in an attractive delta function potential.

This energy shift is the same as the energy shift calculated in I (repulsive potentials), with $-g$ substituted for g.

III. GROUND-STATE WAVEFUNCTION

Now let us discuss the nature of the ground-state wavefunction. In region 1 ($x_1 < x_2 < x_3 \cdots < x_N$), we may calculate the amplitude of any plane wave from the wavefunction rules of Sec. I. In particular, let us calculate $(123 \cdots N)_1$, $(123 \cdots N)_1 = 1 - e^{-2iz_1} = 1 - e^{2/a} = 1 - e^{\frac{1}{2}(gL)}$. Thus this amplitude, and all of the associated $(N-1)!$ amplitudes which come by permuting the numbers $2 \cdots N$, are exponentially large in the large volume limit. The amplitude

$$-(21 \cdots N)_1 = 1 - e^{-2iz_2} = 1 - e^{-2/a} = 1 - e^{-\frac{1}{2}(gL)}$$

is of order unity in the large volume limit. The remaining amplitudes are all calculated with real k's in Region 1, and are thus qualitatively quite similar to all of the amplitudes in the repulsive case.

If we now move on to Region 2 and calculate the amplitide $(123 \cdots N)_2$, we obtain

$$(123 \cdots N)_2 = e^{2iz_2}(1 - e^{\frac{1}{2}(gL)}) = e^{-\frac{1}{2}(gL)}(1 - e^{\frac{1}{2}(gL)}) = e^{-\frac{1}{2}(gL)} - 1,$$

which is again of order unity.

The amplitude

$$(21 \cdots N)_2 - e^{2iz_1}(21 \cdots N)_1 = e^{\frac{1}{2}(gL)}(e^{-\frac{1}{2}(gL)} - 1) = 1 - e^{\frac{1}{2}(gL)},$$

and hence is boosted from order unity to something exponentially large in the volume. In general, moving to Region 2 implies multiplication by a phase shift unless Particle 2 is part of the bound state.

If we continue moving from region to region, we can calculate the wavefunction everywhere. In every region we will find that some amplitudes are exponentially large and the remainder are of the order unity. In the large volume limit the exponential amplitudes are dominant. This dominant wavefunction may be described as follows:

1. The particle with z_1 must be to the left of the particle with z_2.

2. The spin-down particle must either have z_1 or z_2 or be between the two particles which have z_1 and z_2.

If the density were zero we would say that the different particle would have to be involved in the bound state, since it is the only particle with which a spin-up particle can bind. We find now that at finite density the different particle may have a real velocity associated with it, but it is always trapped by the bound-state wavefunction. That is, the different particle may have a real velocity only if it is "in transit" within the bound state.

IV. EXCITATION SPECTRUM OF THE SYSTEM

Now we calculate the excited states of the system and attempt to calculate the energy of excitation as a function of the total momentum of the system. As in I, we focus our attention only on excited states which arise from the collective interactions of the different particle with the background, and not on those excitations which are ordinary Fermi excitations of the background alone. In order to do this, we make use of the freedom of allowing the constant in the spectral law to have any value we choose,

provided the sum of the z_i's is an integral multiple of π.

The program is basically the same as was followed in I. First we find both the energy and total momentum as a function of the constant in the spectral law, and then eliminate the constant in the expression for the energy to obtain the energy as a function of total momentum.

First we must contend with the problems which arise from the imaginary (or complex) roots to the spectral equation. Again we restrict ourselves, *a priori*, to a constant which is real, and justify this assumption with the conclusions to which it leads.

If we assume that the z_i's have both real and imaginary parts, we obtain the relation

$$a(\alpha + i\beta) = \frac{\tan\alpha(1 - \tanh^2\beta)}{\tan^2\alpha + \tanh^2\beta} - i\frac{\tanh\beta(1 + \tan^2\alpha)}{\tan^2\alpha + \tanh^2\beta} + c.$$

If c is real, the imaginary part of this expression reads

$$a\beta = \frac{\tanh\beta(1 + \tan^2\alpha)}{\tan^2\alpha + \tanh^2\beta},$$

which is the same expression we had in the previous section. It has the property that $\beta = 1/a$ is a solution for a small essentially independent of α. This solution for β implies that $\tanh\beta \sim 1$, and thus that the real part of ctn z is essentially equal to zero (again as argued in the previous section). This implies

$$a\alpha = c.$$

Hence, with $c \neq 0$, there are two complex conjugate z's, and all of the remaining z's are real.

We now have the following array of z's:

$$z_1 = (c/a) + (i/a),$$
$$z_2 = (c/a) - (i/a),$$
$$z_s^+ = s\pi - \text{ctn}^{-1}(sa\pi - c),$$
$$z_s^- = -s\pi + \text{ctn}^{-1}(sa\pi + c).$$

Let us now calculate the total momentum of the system.

$$k = \frac{2}{L}\sum z_i = \frac{4c}{aL} + \frac{2}{4}\sum_{s=0}^{\frac{1}{2}(N-2)}[\text{ctn}(sa\pi + c) - \text{ctn}(sa\pi - c)].$$

Changing sums to integrals we obtain

$$k = \frac{4c}{aL} + \frac{2}{La\pi}\int_0^{2k_F/v}[\text{ctn}^{-1}(y + c) - \text{ctn}^{-1}(y - c)]\,dy.$$

This provides a transcendental relationship between k and c. In general, it is very difficult to manipulate, but we can get an idea of its meaning if we expand the integral for small c and do the integral. Under those circumstances we obtain

$$k = (4c/aL)[1 - (1/\pi)\tan^{-1}(2k_F/g)].$$

Thus, for small c, k is proportional to c. This justifies our choice of a real constant at the beginning of this analysis, for k itself must be real.

Now let us calculate the energy of the system as a function of c,

$$E = \frac{2}{L^2}\sum z_2^2 = \frac{2}{L^2}(z_1^2 + z_2^2) + \frac{2}{L^2}\sum_{\text{real } s} z_i^2$$
$$= \frac{4c^2}{a^2L^2} - \frac{4}{a^2L^2} + \frac{4}{L^2}\sum_{s=0}^{\frac{1}{2}(N-2)} s^2\pi^2 - \frac{4}{L^2}\sum_{s=0}^{\frac{1}{2}(N-2)} s\pi[\text{ctn}^{-1}(sa\pi - c) + \text{ctn}^{-1}(sa\pi + c)].$$

The third term on the right is again just the energy of a Fermi gas of $N - 1$ particles, so we again define an energy shift which is $E - E_F(N-1)$. Converting the sums to integrals this energy shift is

$$\Delta E(c) = \frac{4c^2}{a^2L^2} - \frac{4}{L^2a^2\pi} \times \int_0^{2k_F/v} y[\text{ctn}^{-1}(y - c) + \text{ctn}^{-1}(y + c)]\,dy - \frac{g^2}{4}.$$

Again, in full generality, this expression is difficult to interpret. The small c expansion, however, contains a great deal of information. If $c = 0$ we recover the energy shift which we calculated in the previous section. It is easy to see that $d\Delta E/dc = 0$ when $c = 0$, so the lowest order in a power series expansion is c^2. Expanding and doing the integrals we obtain

$$\Delta E_2(c) = \frac{4c^2}{a^2L^2}\left\{1 - \frac{1}{\pi}\left[\tan^{-1}\frac{2k_F}{g} + \frac{\frac{1}{2}(2k_F)}{1 + 4k_F^2/g^2}\right]\right\}.$$

The second factor of this expression is positive for all values of $2k_F/g$, and thus we see that our previous assumption that $c = 0$ was the condition for the ground state was correct.

If we now eliminate c in favor of k in the expression for the excitation energy, we find

$$\Delta E_2 = \frac{k^2}{4}\left\{1 - \frac{1}{\pi}\left[\tan^{-1}\frac{2k_F}{g} + \frac{2k_F/g}{1 + 4k_F^2/g^2}\right]\right\}\Big/\left(1 - \frac{1}{\pi}\tan^{-1}\frac{2k_F}{g}\right)^{-2}.$$

Now we can calculate the effective mass of these excitations from the expression

$$E = k^2/2m^*.$$

$$m^* = \left[2\left(1 - \frac{1}{\pi}\tan^{-1}\frac{2k_F}{g}\right)^2\right] \Big/ \left\{1 - \frac{1}{\pi}\left[\tan^{-1}\frac{2k_F}{g} + \frac{2k_F/g}{1 + 4k_F^2/g^2}\right]\right\}^{-1}.$$

As g tends to zero (weak interactions), m^* approaches 1, indicating that the excitations have the effective mass of one of the fundamental fermions. As g tends to infinity (strong interaction), m^* approaches 2, which means that the bound state is so tightly bound that the two bound particles operate as a single unit.

These results, the ground-state energy, the excitation energy, and the effective mass may be obtained from the corresponding results for the repulsive interaction by simply changing the sign of g. Perturbation theory, which relies upon complete analyticity in the coupling constant, confirms all of these results by duplicating (at least to third order) the small g power series for these quantities. It seems surprising that the rather extreme differences in the calculations (i.e., the explicit inclusion of the complex roots in the attractive case) give rise to a simple analytic continuation of a series in g through the value $g = 0$. This principle of analyticity is valid even for calculations involving the wavefunction as we shall now demonstrate by showing that the ground-state pair-correlation function also has this property.

V. GROUND-STATE PAIR CORRELATIONS

The pair-correlation function is defined to be

$$\rho(x_1, x_2) = \int_0^L \cdots \int_0^L dx_3 \cdots dx_N \times \psi^*(x_1 \cdots x_N)\psi(x_1 \cdots x_N).$$

This function gives the relative probability to find particle z at x_2, given that Particle 1 is at x_1. Of course, Particle 2 is the same as all of the other spin-down particles, and thus this function is equally well interpreted as the spin-up–spin-down pair-correlation function.

It has been pointed out to the author[3] that the derivation of the pair correlation in I is incorrect, although it will turn out that the answer is right. We will correct the derivation here, and calculate the pair-correlation function for the attractive case.

[3] E. H. Lieb and the referee of I.

The wavefunction in Region I may be written in the form of a Slater Determinant

$$\psi = \begin{vmatrix} \alpha_1 e^{ik_1x_1} & \alpha_2 e^{ik_2x_1} & \alpha_3 e^{ik_3x_1} & \cdots \\ e^{ik_1x_2} & e^{ik_2x_2} & e^{ik_3x_2} & \cdots \\ e^{ik_1x_3} & e^{ik_2x_3} & e^{ik_3x_3} & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix}.$$

It is easily checked that this is exactly the wavefunction dictated by the rules of Sec. I. We have gone to great trouble to assure that our wavefunction is periodic so let us make use of this periodicity here. We may simplify the calculation by setting $x_1 = 0$ and integrating over all coordinates from 0 to L. This covers one entire periodic cycle of the wavefunction, and has the advantage that the integrated requires knowledge of the wavefunction in Region I only.

Under these conditions

$$\psi = \begin{vmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_N \\ e^{ik_1x_1} & e^{ik_2x_2} & e^{ik_3x_2} & \cdots & e^{ik_3x_N} \\ e^{ik_1x_3} & e^{ik_2x_3} & e^{ik_3x_3} & \cdots & e^{ik_3x_N} \\ \vdots & \vdots & \vdots & & \end{vmatrix}.$$

Let ψ_m^l be the determinant of the minor formed by removing the mth row and the lth column from the preceding determinant. With this notational convenience we may expand by mirrors to factor out the dependence of any x_j. Let us factor out x_3,

$$\psi = \sum_l (-1)^{l+1} e^{ik_l x_3} \psi_3^l.$$

Thus

$$\psi^*\psi = \sum_l \sum_m (-1)^{l+m+2} \psi_3^l (\psi_3^m)^* e^{i(k_l - k_m)x_3}.$$

Only the last factor depends upon x_3, so we may perform the integration over x_3,

$$I = \int_0^L e^{i(k_l - k_m)x_3}\, dx_3 = \frac{1}{i(k_l - k_m)}\left(e^{i(k_l - k_m)L} - 1\right).$$

This integral would, of course, be $L\delta_{lm}$ if the k_i's were separated by integral multiples of $2\pi/L$ as in the noninteracting gas. As we have seen, our system does not have this property, for each of the k_i's are given by a more complicated transcendental relation. As a matter of fact, for an attractive interaction the transcendental relation is

$$k_l - k_m = (g/2)(\operatorname{ctn} k_m \tfrac{1}{2}L - \operatorname{ctn} k_l \tfrac{1}{2}L)$$
$$= \frac{g}{2}\frac{\sin (k_l - k_m)\tfrac{1}{2}L}{\sin k_l \tfrac{1}{2}L \sin k_m \tfrac{1}{2}L}$$
$$= -ig\,\frac{e^{i(k_l - k_m)L} - 1}{(1 - e^{ik_l L})(1 - e^{-ik_m L})}.$$

Using this result we find

$$I = (1/g)\alpha_l \alpha_m^*, \qquad l \neq m,$$
$$I = L, \qquad l = m.$$

The most convenient way to write this result is

$$I = (1/g)\alpha_l \alpha_m^* + [L - (1/g)\alpha_l \alpha_m^*]\delta_{lm}.$$

Having thus performed the x_3 integration we return to the determinental form,

$$\int_0^4 \psi^* \psi \, dx_3 = \frac{1}{g}\sum_l \sum_m (-1)^{l+m+2} \psi_3^l (\psi_3^m)^* \alpha_l \alpha_m^* + \sum_m \psi_3^m (\psi_3^m)^* \left(L - \frac{1}{g}\alpha_m \alpha_m^*\right).$$

The first term in this expansion is zero, because it represents the product of two determinants each of which have a repeated row. That is,

$$\frac{1}{g}\sum_l \sum_m (-1)^{l+m+2} \psi_3^l (\psi_3^m)^* \alpha_l \alpha_m^* = \frac{1}{g}\begin{vmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_N \\ e^{ik_1 x_3} & e^{ik_2 x_3} & e^{ik_3 x_3} & \cdots & e^{ik_N x_3} \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_N \\ \vdots & \vdots & \vdots & & \vdots \end{vmatrix} \begin{vmatrix} \alpha_1^* & \alpha_2^* & \alpha_3^* & \cdots & \alpha_N^* \\ e^{-ik_1 x_3} & e^{-ik_2 x_3} & & \cdots & e^{-ik_N x_3} \\ \alpha_1^* & \alpha_2^* & & \cdots & \alpha_N^* \\ \vdots & \vdots & & & \vdots \end{vmatrix} = 0.$$

Thus, we now have

$$\int_0^4 \psi^* \psi \, dx_3 = \sum_m \psi_3^m (\psi_3^m)^* \left(L - \frac{1}{g}\alpha_m \alpha_m^*\right),$$

which is very similar to the result we would have obtained had the k_i's been separated by integral multiples of $\frac{1}{4}(2\pi)$. The only difference is that the effective normalization of ψ_3^m is $[L - (1/g)\alpha_m \alpha_m^*]^{\frac{1}{2}}$ instead of $L^{\frac{1}{2}}$.

Now, as we integrate on x_4, we may repeat this process independently on each of the N determinental mirrors appearing in the above sum. The argument is exactly the same for each of these terms and the result is

$$\int_0^L \int_0^L \psi^* \psi \, dx_3 \, dx_4 = \sum_{m \neq l}\sum \psi_{3,4}^{m,l} (\psi_{3,4}^{m,l})^* \times \left(L - \frac{1}{g}\alpha_m \alpha_m^*\right)\left(L - \frac{1}{g}\alpha_l \alpha_l^*\right),$$

where $\psi_{3,4}^{m,l}$ is the determinant of the mirror formed by crossing out rows 3 and 4 and columns m and l in the original determinant.

This process continues until all of the rows greater than 3 and all of the columns but two have been crossed out. This leaves only 2×2 determinants to be evaluated. Let φ_{ij} be one of these 2×2 determinants. Of course

$$\varphi_{ij} = \psi_{3,4,5\cdots N}^{m,n,l\cdots \neq i,j} = \begin{vmatrix} \alpha_i & \alpha_j \\ e^{ik_i x_3} & e^{ik_j x_3} \end{vmatrix}.$$

In terms of these 2×2's

$$\rho(x_2) = \sum_{i>j}\sum \sum_m \sum_{l \neq m} \sum_{\substack{n \neq l \\ \neq m}} \cdots \times \left[\prod_{s \neq i,j} \left(L - \frac{1}{g}\alpha_s^* \alpha_s\right)\right] \varphi_{ij} \varphi_{ij}^*.$$

Let us convert the product into an unrestricted product and rewrite the expression

$$\rho(x_2) = \left[\prod_{s=1}^N \left(L - \frac{1}{g}\alpha_s^* \alpha_s\right)\right] \sum_m \sum_{l \neq m} \sum_{\substack{n \neq l \\ \neq m}} \cdots \sum_{i>j}\sum \frac{1}{[L - (1/g)\alpha_i \alpha_i^*]} \frac{1}{[L - (1/g)\alpha_j \alpha_j^*]} \varphi_{ij} \varphi_{ij}^*.$$

Now all of the sums but i and j are just enumerating the ways to cross off $(N - 2)$ columns such that the ith and jth columns remain, and since $\varphi_{ii} = 0$ we may write the final result

$$\rho(x_2) = \left[\prod_{s=1}^N \left(L - \frac{1}{g}\alpha_s^* \alpha_s\right)\right](N-2)!\,\frac{1}{2}\sum_i \sum_j \frac{1}{[L - (1/g)\alpha_i \alpha_i^*]} \frac{1}{[L - (1/g)\alpha_j \alpha_j^*]} \varphi_{ij} \varphi_{ij}^*.$$

This is the exact expansion for the pair-correlation function. It may be shown that this function also is periodic with period L. We wish to calculate this function only in the large L limit, thus we may approximate the factors in the denominator by L, since

$$(1/g)\alpha_s^*\alpha_s \leq 4/g$$

and hence is negligible with respect to L.

Aside from irrelevant normalization factors, we have

$$\rho(x_2) \sim \sum_i \sum_j \alpha_i^*\alpha_i + \alpha_j^*\alpha_j - \alpha_i\alpha_j^* e^{i(k_j-k_i)x_2} - \alpha_j\alpha_i^* e^{-i(k_j-k_i)x_2},$$

which is exactly the expression used to calculate the pair correlation in I.

There is one additional complication which arises in the attractive case. In the preceding derivation we have assumed that all of the k_i's are real. As we have shown previously, the attractive ground state has two complex k's, which are complex conjugates of one another. The complication comes from the integral over the coordinates. If we allow for complex k's the integral is written

$$I = \int_0^L e^{i(k_l - k_m^*)x_3}\, dx_3.$$

By exactly the same line of reasoning as before

$$I = (1/g)\alpha_l\alpha_m^*, \qquad k_l \neq k_m^*,$$
$$= L, \qquad k_l = k_m^*.$$

If we let k_1 and k_2 be the complex k's and k_3 through k_N be real, we can say

$$I = (1/g)\alpha_l\alpha_m^* \quad \text{except where} \quad l = 1, m = 2, \quad m = 2, l = 1$$
$$= L \qquad \text{in the excepted cases.}$$

Formally the complication is easily handled if we interchange the first two rows and the first two columns of the Slater Determinant for ψ^*. This, of course, does not change ψ^* at all. We now define

$$\tilde{\alpha}_i = \alpha_2^*, \qquad i = 1,$$
$$= \alpha_1^*, \qquad i = 2,$$
$$= \alpha_i^*, \qquad i \geq 3.$$

Making use of the fact that $k_1 = k_2^*$, we can now write $\psi^*\psi$ in the form

$$\psi^*\psi = \begin{vmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_N \\ e^{ik_1x_3} & e^{ik_2x_3} & e^{ik_3x_3} & \cdots & e^{ik_Nx_3} \\ e^{ik_1x_3} & e^{ik_2x_3} & e^{ik_3x_3} & \cdots & \\ \vdots & \vdots & \vdots & & \end{vmatrix} \begin{vmatrix} e^{-ik_2x_3} & e^{-ik_3x_3} & e^{-ik_3x_3} & \cdots \\ \tilde{\alpha}_1 & \tilde{\alpha}_2 & \tilde{\alpha}_3 & \cdots \\ e^{-ik_1x_3} & e^{-ik_2x_3} & e^{-ik_3x_3} & \cdots \\ \vdots & \vdots & \vdots & \end{vmatrix}.$$

This product may be expanded and integrated in exactly the same way. The only change in the result is to change α_i^* into $\tilde{\alpha}_i$ and to change the definition of the residual 2×2 determinants. If we let

$$\varphi_{ij} = \begin{vmatrix} \alpha_i & \alpha_j \\ e^{ik_ix_3} & e^{ik_jx_3} \end{vmatrix},$$
$$\tilde{\varphi}_{ij} = \begin{vmatrix} e^{-ik_ix_3} & e^{-ik_jx_3} \\ \tilde{\alpha}_i & \tilde{\alpha}_j \end{vmatrix},$$

the final result is

$$\rho(x_2) = \frac{(N-2)!}{2}\left[\prod_{s=1}^{N}\left(L - \frac{1}{g}\alpha_s\tilde{\alpha}_s\right)\right] \times \sum_i \sum_j \frac{1}{[L - (1/g)\alpha_i\tilde{\alpha}_i]} \frac{1}{[L - (1/g)\alpha_j\tilde{\alpha}_j]} \tilde{\varphi}_{ij}\tilde{\varphi}_{ij}.$$

Now let us calculate this pair correlation function in the large L limit. According to the wavefunction rules and with our usual designation of k's we have

$$\alpha_1 = (1 - e^{ik_1L}) = 1 - e^{-\frac{1}{2}gL} = \tilde{\alpha}_2 = 1$$

in the large L limit,

$$\alpha_2 = (1 - e^{ik_2L}) = 1 - e^{\frac{1}{2}(gL)} = \tilde{\alpha}_1 = -e^{\frac{1}{2}(gL)}$$

in the large L limit.

Thus the factor

$$L - (1/g)\alpha_1\tilde{\alpha}_1 = L - (1/g)\alpha_2\tilde{\alpha}_2 \cong (1/g)e^{\frac{1}{2}(gL)}$$

in the large L limit.

As before, we may neglect $(1/g)\alpha_s\tilde{\alpha}_s$ in comparison with L in the demoninator of the expression inside the sum, provided k_s is real.

Now we will compute the actual form of the pair-correlation function in the ground state. Let us neglect the normalization factors which multiply ρ and pick a convenient normalization later. We find four different kinds of terms in the summation.

(1) $i = 1, j = 2$ and $j = 1, i = 2$,

$$F_1 = g^2 e^{-igL} \begin{vmatrix} 1 & -e^{\frac{1}{2}(gL)} \\ e^{-\frac{1}{2}(gx)} & e^{\frac{1}{2}(gx)} \end{vmatrix} \begin{vmatrix} e^{\frac{1}{2}(gx)} & e^{-\frac{1}{2}(gx)} \\ -e^{\frac{1}{2}(gL)} & 1 \end{vmatrix} = g^2 e^{-gx}$$

provided $x \ll L$.

(2) $i = 1, j \geq 3; j = 1, i \geq 3$,

$$F_2 = \sum_i \frac{g}{L} e^{-\frac{1}{2}(gL)} \begin{vmatrix} 1 & \alpha_j \\ e^{-\frac{1}{2}(gx)} & e^{ik_j x} \end{vmatrix} \begin{vmatrix} e^{gx} & e^{-ik_j x} \\ -e^{\frac{1}{2}(gL)} & \alpha_j^* \end{vmatrix}$$

$$= \frac{g}{L} \sum_j e^{-ik_j x}(e^{ik_j x} - \alpha_j e^{-\frac{1}{2}(gx)})$$

$$= \frac{Ng}{L} - \frac{g e^{-\frac{1}{2}(gx)}}{L} \sum_j \alpha_j e^{-ik_j x}$$

provided $x \ll L \quad N \gg 1$.

(3) $\quad i = 2, j \geq 3; j = 2, i \geq 3,$

$$F_3 = \sum_i \frac{g}{4} e^{-\frac{1}{2}(gL)} \begin{vmatrix} -e^{-\frac{1}{2}(gL)} & \alpha_j \\ e^{gx} & e^{ik_j x} \end{vmatrix} \begin{vmatrix} e^{-\frac{1}{2}(gx)} & e^{-ik_j x} \\ 1 & \alpha_j^* \end{vmatrix}$$

$$F_3 = \frac{Ng}{4} - g e^{-\frac{1}{2}(gx)} \sum \alpha_j^* e^{ik_j x}$$

provided $x \ll L$, $N \gg 1$.

(4) $\quad i \geq 3, j \geq 3,$

$$F_4 = \sum_i \sum_j \frac{1}{4^2} \begin{vmatrix} \alpha_i & \alpha_j \\ e^{ik_i x} & e^{ik_j x} \end{vmatrix} \begin{vmatrix} e^{-ik_i x} & e^{-ik_j x} \\ \alpha_i^* & \alpha_j^* \end{vmatrix}$$

$$= -\frac{2N}{L^2} \sum_i \alpha_i \alpha_i^* + \frac{2}{4^2} |\sum_j \alpha_j e^{-ik_j x}|^2$$

$$F_4 = -\frac{2N}{L^2} \sum \alpha_i \alpha_i^* + \frac{2}{L^2} |\sum_i \alpha_i e^{-ik_i x}|^2.$$

Now the pair correlation function is proportional to

$$\rho \propto F_1 + F_2 + F_3 + \tfrac{1}{2} F_4.$$

$$\rho \propto \frac{2N_e}{L} - \frac{N}{L^2} \sum \alpha_i^* \alpha_i + g^2 e^{-gx} - \frac{g}{L} e^{-\frac{1}{2}(gx)}$$

$$\times (\sum \alpha_i e^{-ik_i x} + \sum \alpha_i^* e^{ik_i x}) + \frac{1}{L^2} |\sum \alpha_i e^{-ik_i x}|^2.$$

$$\rho \propto \frac{2N_e}{L} - \frac{N}{L^2} \sum \alpha_i \alpha_i^* + \left| \left(g e^{-\frac{1}{2}(gx)} - \frac{1}{L} \sum \alpha_i e^{-ik_i x} \right) \right|^2.$$

We now convert the sums to integrals, using the spectral law

$$\sum \alpha_i \alpha_i^* = 4 \sum_s \sin^2 k_s \tfrac{1}{2} L,$$

but

$$\operatorname{ctn} k_s \tfrac{1}{2} L = -2k_s/g$$

thus

$$\sum \alpha_i \alpha_i^* = 4 \sum_s \frac{1}{1 + 4k_s^2/g^2}.$$

In calculating this sum very little error will be introduced if we let $k_s = 2s\pi/L$. Thus, converting to an integral

$$\sum \alpha_i \alpha_i^* = \frac{L}{2\pi} \int_{-k_F}^{k_F} \frac{1}{1 + 4k^2/g^2} \, dk = \frac{2L_e}{\pi} \tan^{-1} 2k_F/g.$$

Now we take

$$\sum \alpha_s e^{-ik_s x} = 2i \sum e^{ik_s \frac{1}{2} L} \sin k_s \tfrac{1}{2} L e^{ik_s x},$$

from the spectral law

$$= 2i \sum \frac{e^{-ik_s x}}{2k_s/g + i}.$$

If $x \ll L$ we may replace k_s by $2s\pi/L$, and convert the sum to an integral. We obtain

$$\sum \alpha_s e^{-ik_s x} = -\frac{igL}{2\pi} \int_{-2k_F/g}^{2k_F/g} \frac{e^{-igxy/2}}{y + i} \, dy.$$

Putting these back into our expression for the unnormalized pair-correlation function, we obtain

$$\rho \propto \frac{2N_e}{L} - \frac{2N_e}{4\pi} \tan^{-1} 2k_F/g + g^2 \left| e^{-gx/2} - \frac{i}{2\pi} \int_{-2k_F/g}^{2k_F/g} \frac{e^{-\frac{1}{2}(igxy)}}{y + i} \, dy \right|^2.$$

If we now normalize so that the background density is unity, we have

$$\rho = 1 + \left[g\pi \Big/ 2k_F \left(1 - \frac{1}{\pi} \tan^{-1} 2k_F/g \right) \right] Q^*(x) Q(x),$$

where

$$Q(x) = e^{-\frac{1}{2}(gx)} - \frac{i}{2\pi} \int_{-2k_F/g}^{2k_F/g} \frac{e^{-igx/2y}}{y + i} \, dy,$$

or, alternatively

$$Q(x) = \frac{1}{\pi} \int_{2k_F/g}^{\infty} \frac{y \sin \frac{1}{2}(gxy) + \cos \frac{1}{2}(gxy)}{y^2 + 1} \, dy.$$

In two limits of interest:

(1) $\frac{g}{2k_F} \gg 1$ (Strong interactions, low density).

$$Q(x) = e^{-\frac{1}{2}(gx)} - \frac{2k_F}{\pi g} \frac{\sin k_F x}{k_F x},$$

$$\rho(x) = 1 + \frac{g\pi}{2k_F} \left(e^{-\frac{1}{2}(gx)} - \frac{2k_F}{g\pi} \frac{\sin k_F x}{k_F x} \right)^2;$$

(2) $\frac{g}{2k_F} \ll 1$ (Weak interaction, high density).

$$Q(x) = \frac{1}{\pi} \int_{k_F x}^{\infty} \frac{\sin z}{z} \, dz = -\frac{1}{\pi} si(k_F x)$$

$$= \frac{1}{2} - \frac{1}{\pi} si(k_F x)$$

$$\rho(x) = 1 + \frac{g}{2k_F} [si(k_F x)]^2.$$

This result is an analytic continuation in the coupling constant g of the repulsive result, therefore perturbation theory properly predicts this result. Figure 3 is a plot of this pair-correlation function for $k_F = 1$ with the strength g as a parameter.

VI. STATES WITH NO PAIR PRESENT

In addition to the states we have discussed in the previous sections, there is another class of

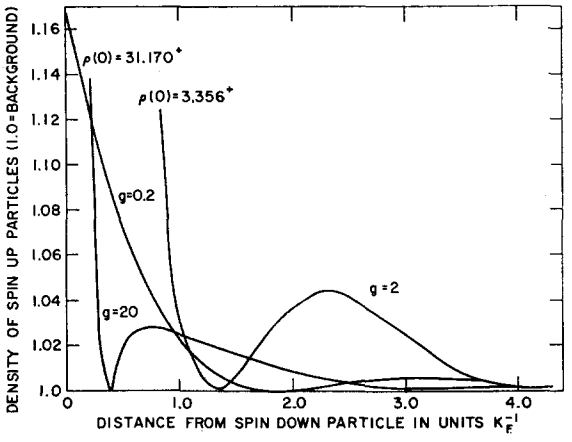

FIG. 3. Pair-correlation deviation function for the ground state.

exited states, namely those which have no bound pair present. The energy of these states may be calculated from the spectral law assuming that all of the z_i's are real.

The line of reasoning is the same as has been used throughout this paper, except that we must now include N real z's. This leads to

$$z_s^+ = s\pi - \operatorname{ctn}^{-1}(sa\pi - c), \qquad s = 1 \cdots \tfrac{1}{2}N,$$

$$z_s^- = -s\pi + \operatorname{ctn}^{-1}(sa\pi + c), \qquad s = 1 \cdots \tfrac{1}{2}N.$$

The energy is

$$E = \frac{2}{L^2}\sum z_i^2 = \frac{4}{L^2}\sum_{s=1}^{N/2} s^2\pi^2 - \frac{4}{L^2}\sum_{s=1}^{N/2} s\pi \times [\operatorname{ctn}^{-1}(sa\pi - c) + \operatorname{ctn}^{-1}(sa\pi + c)] + O\left(\frac{1}{L}\right).$$

The first term is the energy of a Fermi gas with $N + 1$ particles. We have already worked out the second term in Secs. II and IV. The result is

$$E = E(N+1) - \frac{k_F^2}{\pi}\left[\frac{g}{2k_F} + \tan^{-1}\frac{g}{2k_F} + \frac{g^2}{4k_F^2}\left(\tan^{-1}\frac{g}{2k_F} - \frac{\pi}{2}\right)\right] - \frac{g^2c^2}{4\pi}\left[\frac{\pi}{2} - \tan^{-1}\frac{g}{2k_F} + \frac{g/2k_F}{1 + (g^2/4k_F^2)}\right].$$

The connection between c and k for small c is

$$K = \frac{2}{L}\sum z_i = -\frac{2}{L}\sum_{s=1}^{\frac{1}{2}N} [\operatorname{ctn}^{-1}(sa\pi - c) - \operatorname{ctn}^{-1}(sa\pi + c)].$$

Again this sum has been worked out in Sec. IV,

$$K = -gc\left[\frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{g}{2k_F}\right].$$

Thus, for small k,

$$E = E(N+1) - \frac{k_F^2}{\pi}\left[\frac{g}{2k_F} + \tan^{-1}\frac{g}{2k_F} + \frac{g^2}{4k_F^2}\left(\tan^{-1}\frac{g}{2k_F} - \frac{\pi}{2}\right)\right] - \left\{k^2\left[\frac{\pi}{2} - \tan^{-1}\frac{g}{2k_F} + \frac{g/2k_F}{1 + g^2/4k_F^2}\right] \times 4\pi^{-1}\left[\frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{g}{2k_F}\right]^{-2}\right\}.$$

The $k = 0$ state is the most energetic states of this class, because the energy decreases as k increases. This can be seen from Fig. 1 where it is easy to see that as c is increased or decreased from zero the z's are pinched together instead of being spread apart. For g small, the energy shift is linear in g, and hence the effect of breaking a pair in the $k = 0$ state is about k_F^2, or the energy required to move two particles to the top of the Fermi sea. The different particle, however, still is most likely to be near the middle of the Fermi sea.

If g is very large the energy required to break the pair in the $k = 0$ state tends to $\frac{1}{4}g^2$, or just the binding energy of the bound state.

If we wish to break the pair in some other state of total momentum k we see that the occupation probability for the different particle maximizes at $-k$. It is possible to show that the states with a bound pair present, and those with the bound pair broken come within order g^2 of one another as k approaches k_F. If one breaks the pair and keeps $k = k_F$ it is necessary to send the different particle all the way to $-k_F$. This is reminiscent of the BCS wavefunction in the theory of superconductivity.

VII. PAIR CORRELATIONS IN THE "BROKEN PAIR" STATES

Since all of the k_i's are real, the pair correlation function is

$$\rho \propto \frac{N}{L^2}\sum \alpha_i\alpha_i^* - \left|\frac{1}{L}\sum \alpha_i e^{-ikix}\right|^2$$

for $x \ll L$ $N \gg 1$. We put this into normalized form, and use the previous calculations of the sums and obtain

$$\rho = 1 - \frac{gR(x)R^*(x)}{8k_F \tan^{-1} 2k_F/g}$$

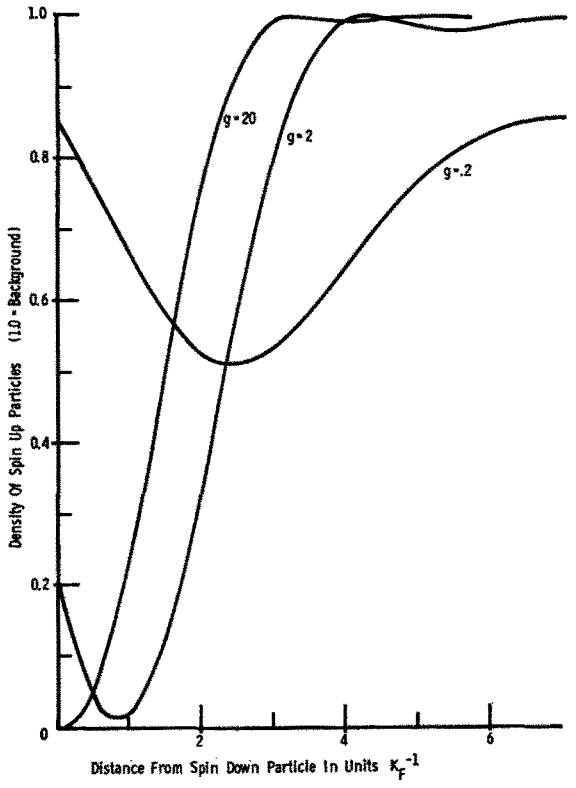

FIG. 4. Pair-correlation function for the "broken pair" $K = 0$ state.

where

$$R(x) = \int_{-2k_F/g}^{2k_F/g} \frac{e^{-\frac{1}{2}(igxy)}}{y + i}\, dy = 2\pi i(Q(x) - e^{-\frac{1}{2}(gx)}).$$

This pair-correlation function is plotted in Fig. 4 with $k_F = 1$ and g as a parameter.

The small-g behavior of this function is interesting. For small g we obtain

$$\rho(x) = 1 - (gL/N)[e^{-\frac{1}{2}gx} + si(k_F x)]^2,$$

a function which is very close to unity as g becomes small, but whose departure from unity extends over an enormous range. Define the function f as

$$f(x) = \rho(x) - 1.$$

We now claim that $f(x)$ tends to zero as g tend to zero, but the integral of $f(x)$ is independent of g. Thus, in this state, a very tiny pair correlation extends over a huge distance.

In the other extreme, when $g/k_F \gg 1$, the pair-correlation function becomes

$$\rho(x) = 1 - (\sin k_F x/k_F x)^2,$$

which is the same result one obtains for the strongly interacting repulsive case. This is just a manifestation of the fact that an infinite attractive delta-function well is as impenetrable as an infinite repulsive delta-function well, and this answer would have been anticipated from the work of Giraudeau.[4]

For intermediate values of the coupling constant the pair-correlation function tends to "pile up" probability in the immediate vicinity of the different particle, but the maximum value of this enhanced region is always below background. One striking effect of removing the bound pair is that the pair-correlation function switches from being everywhere greater than background to being everywhere less than background.

VIII. CONCLUSIONS

To a point, the results for the attractive case bear a great resemblance to those for the repulsive case. The ground-state energy shift and the effective mass of the low-lying excited states and the pair-correlation function are the same functions of the strength that they were for the repulsive case. As one would expect, the energy shift is negative so that the energy of the system is less than the energy of a Fermi gas of $N - 1$ particles.

The attractive potential also gives rise to a class of excited states which have no analog in the repulsive case. These are the "broken pair" states which have all real k's and no bound state is present. These states have an energy which is greater than the energy of a Fermi gas of $N - 1$ particles, and are characterized by a negative effective mass. These states are separated from the states with a pair present by an energy if the order of the binding energy when the total momentum is nearly equal to k_F. When no bound pair is present these states exhibit a long-range tail in the pair-correlation function.

Finally, it should be emphasized that the ground state of a system of N spin-$\frac{1}{2}$ fermions would have $\frac{1}{2}N$ spin up and $\frac{1}{2}N$ spin down.[5] It is hoped that these two papers will provide a first step toward the solution of this important problem.

ACKNOWLEDGMENT

The author wishes to acknowledge many helpful communications and discussions with Professor D. S. Saxon, Dr. B. Day, and Dr. E. H. Lieb. The hospitality of the Brookhaven National Laboratory, during the latter portions of this work, is also gratefully acknowledged.

[4] M. D. Giraudeau, J. Math. Phys. 1, 516 (1960).
[5] E. Lieb and D. Mattis, Phys. Rev. 125, 164 (1962).

Some Theorems on the Unimodular Complex Degree of Optical Coherence†

C. L. MEHTA, E. WOLF
Department of Physics and Astronomy, University of Rochester, Rochester, New York
AND
A. P. BALACHANDRAN
Department of Physics, Syracuse University, Syracuse, New York
(Received 30 March 1965)

The complex degree of coherence $\gamma(\mathbf{r}_1, \mathbf{r}_2, t_1 - t_2)$ of a stationary optical field is defined as the normalized cross-correlation function ($|\gamma| \leq 1$) of the light disturbances at two space–time points $(\mathbf{r}_1, t_1)$, $(\mathbf{r}_2, t_2)$, the disturbances being represented by means of Gabor's analytic signals.

In the present paper the general form of γ is examined under the condition that $|\gamma|$ takes on the extreme value unity for all possible time differences $\tau = t_1 - t_2$ ($-\infty < \tau < \infty$). Several cases are distinguished, depending on whether this condition is satisfied for some points or for all points in some fixed domain of space. It is pointed out, that the previously published derivations of the relevant theorems contain serious errors. The methods employed in the present paper make use of the property of nonnegative definiteness which the complex degree of coherence is shown to obey. The results have a bearing on the important but as yet unsolved "phase problem" of optical coherence theory.

1. INTRODUCTION

A BASIC quantity used in the analysis of a large number of coherence phenomena associated with stationary optical fields is the complex degree of coherence[1]

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)/\{\Gamma(\mathbf{r}_1, \mathbf{r}_1, 0)\Gamma(\mathbf{r}_2, \mathbf{r}_2, 0)\}^{\frac{1}{2}} \tag{1.1}$$

where Γ is the mutual coherence function

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle V(\mathbf{r}_1, t + \tau)V^*(\mathbf{r}_2, t)\rangle. \tag{1.2}$$

In (1.2), $V(\mathbf{r}, t)$ is a complex scalar wavefunction, $\mathbf{r}_j$ ($j = 1, 2$) are position vectors of two points in the wave field, t represents the time and angular brackets the time average. γ is an analytic signal[1,2] with respect to the variable τ, i.e., it has Fourier integral representations which contains only nonnegative frequency components,

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \int_0^\infty w(\mathbf{r}_1, \mathbf{r}_2, \nu)\, e^{-2\pi i\nu\tau}\, d\nu. \tag{1.3}$$

If one assumes that γ is square integrable, then this result implies, according to a well known theorem[3] that γ, considered as a function of complex τ, is the boundary value on the real τ-axis of a function which is analytic and regular in the lower half of the complex τ-plane. The normalization of γ ensures that[1]

$$0 \leq |\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \leq 1. \tag{1.4}$$

The extreme cases $\gamma = 0$ and $|\gamma| = 1$ are traditionally associated with "complete incoherence" and "complete coherence," respectively.

Parrent[4] (see also Refs. 5 and 6) has considered the problem of determining the general form of γ in the extreme case when $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \equiv 1$ for all τ values ($-\infty < \tau < \infty$), and for all pairs of points $\mathbf{r}_1$ and $\mathbf{r}_2$. He concluded that in this case γ must necessarily be strictly periodic in τ, and be of the form

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \exp\{i[\alpha(\mathbf{r}_1) - \alpha(\mathbf{r}_2) - 2\pi\nu_0\tau]\}. \tag{1.5}$$

Here $\alpha(\mathbf{r})$ is a real function of $\mathbf{r}$ and ν_0 is a positive constant. From (1.1) it follows that the mutual coherence function is then expressible in the form

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = U(\mathbf{r}_1)U^*(\mathbf{r}_2)\, e^{-2\pi i\nu_0\tau}, \tag{1.6}$$

where

$$U(\mathbf{r}) = \{\Gamma(\mathbf{r}, \mathbf{r}, 0)\}^{\frac{1}{2}} \exp[i\alpha(\mathbf{r})]$$

is a function of $\mathbf{r}$ only.

In the present paper we are returning to the problem treated by Parrent and consider also some related problems for two reasons. First, because as will be shown below, while the conclusions expressed by Eqs. (1.5) and (1.6) are correct, Par-

† Research supported by the Air Force Cambridge Research Laboratories and by the U. S. Atomic Energy Commission.

[1] M. Born and E. Wolf, *Principles of Optics*, (Pergamon Press, Oxford, England and Macmillan Co., New York; 1964), 2nd Ed., Chapter X.

[2] D. Gabor, J. Inst. Elec. Engrs. 93, Pt. III, 429 (1946).

[3] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, (Clarendon Press, Oxford, England, 1948), 2nd Ed., p. 128.

[4] G. B. Parrent, Jr., Opt. Acta 6, 285 (1959).

[5] G. B. Parrent, Jr., J. Opt. Soc. Am. 49, 787 (1959).

[6] M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice–Hall, Inc., Englewood Cliffs, New Jersey, 1964), p. 49.

rent's proof contains serious errors. Secondly, because this problem turns out to have a bearing on the important but as yet unsolved phase problem of coherence theory.[7] The phase problem is essentially the problem of finding conditions under which the knowledge of the modulus of $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ for all real τ-values (but fixed $\mathbf{r}_1$ and $\mathbf{r}_2$) allows a unique determination of its phase. It may be shown that (cf. Refs. 7, 8) in view of the analytic behavior of γ already mentioned, the phase ϕ of γ must necessarily be expressible in the form[9]

$$\phi(\tau) = -2\pi\nu_0\tau + \tilde{\phi}(\tau) + \sum_j \phi^{(j)}(\tau), \qquad (1.7)$$

where

$$\tilde{\phi}(\tau) = \frac{2\tau}{\pi} P \int_0^\infty \frac{\log|\gamma(\tau')|}{\tau'^2 - \tau^2}\, d\tau', \qquad (1.8)$$

and $\phi^{(j)}(\tau)$ is the argument of the Blaschke factor

$$\exp[i\phi^{(j)}] = [\tau - \tau^{(j)}]/[\tau - \tau^{(j)*}]. \qquad (1.9)$$

In (1.7), ν_0 is a real, nonnegative quantity which is independent of τ and the $\tau^{(j)}$ are zeros in the analytic continuation of $\gamma(\tau)$ in the lower half of the complex τ-plane.

It is seen from Eq. (1.7) that, apart from a linear additive term in τ, the phase $\phi(\tau)$ may in general be determined from the knowledge of the modulus of $\gamma(\tau)$ on the real τ-axis and the locations of the zeros in the analytic continuation of $\gamma(\tau)$ in the lower half of the complex τ-plane. However, the locations of the zeros cannot be quite arbitrary, since γ may be shown to obey the following condition of "nonnegative definiteness"[10]:

$$\sum_{j=1}^{n}\sum_{k=1}^{n} \gamma(\mathbf{r}_j, \mathbf{r}_k, \tau_j - \tau_k)\alpha_j\alpha_k^* \geq 0. \qquad (1.10)$$

Here n is any positive integer greater than or equal to unity, $\mathbf{r}_1, \mathbf{r}_2, \cdots \mathbf{r}_n$ are arbitrary position vectors of points in the region of space for which γ is defined, $\tau_1, \tau_2, \cdots, \tau_n$ are arbitrary real constants, and $\alpha_1, \alpha_2, \cdots, \alpha_n$ are arbitrary complex constants. The formula (1.10), which is basic for the present investigation is established in Sec. 2 below.

It has been conjectured[7] that for many cases of physical interest the analytic continuation of γ will, in fact, have no zeros at all in the lower half of the complex τ-plane, at least in the case when $\mathbf{r}_1 = \mathbf{r}_2$. From Eq. (1.7) it follows that in such cases the modulus of $\gamma(\tau)$ on the real τ-axis specifies the phase uniquely up to an additive linear term in τ, and this in turn implies that in such cases the knowledge of $|\gamma|$ for all real τ values specifies the spectral profile of the light uniquely. This conjecture is supported by explicit calculations relating to blackbody radiation.[11]

Now the result of Parrent, expressed by Eq. (1.5) above, relating to the limiting case $|\gamma(\tau)| \equiv 1$ implies that in this case there are also no zeros in the analytic continuation of $\gamma(\tau)$ in the lower half of the complex τ-plane. It is evidently of importance, for a clearer understanding of the phase problem, to find what is the reason for the absence of the zeros in this particular case. The proof of (1.5) given by Parrent[4] is based on two facts:

(1) That the most general analytic signal $f(\tau)$ whose modulus is unity on the whole real τ-axis, must necessarily be of the form

$$f(\tau) = \exp\{i(\beta + \alpha\tau)\} \prod_{n=1}^{\infty} \frac{a_n^*}{a_n}\left(\frac{a_n - \tau}{a_n^* - \tau}\right), \qquad (1.11)$$

where α and β are real constants and a_n are complex constants such that α and the imaginary part of a_n are of the same sign; [both negative or positive according as $f(\tau)$ is the boundary value on the real τ axis of a function which is analytic and regular in the lower or upper half of the complex τ-plane]; and that

(2) if $f(\tau)$ is the coherence function $\gamma(\mathbf{r}, \mathbf{r}, \tau)$, then for all real τ-values

$$\Re\{f(\tau)\} = \Re\{f(-\tau)\}, \qquad (1.12)$$

where $\Re$ denotes the real part.

The representation (1.11) was derived by Edwards

---

[7] E. Wolf, Proc. Phys. Soc. (London) **80**, 1269 (1962).
[8] J. S. Toll, Phys. Rev. **104**, 1760 (1956).
[9] In Eqs. (1.7), (1.8), (1.9) we suppress the explicit dependence of γ and ϕ on $\mathbf{r}_1$ and $\mathbf{r}_2$.
[10] In fact γ satisfies the stronger nonnegative definiteness condition

$$\int f^*(\mathbf{r}_j, \tau_j)\gamma(\mathbf{r}_j, \mathbf{r}_k, \tau_j - \tau_k)f(\mathbf{r}_k, \tau_k)\, d^3\mathbf{r}_j\, d^3\mathbf{r}_k\, d\tau_j\, d\tau_k \geq 0,$$

for all sufficiently well-behaved functions $f(\mathbf{r}, \tau)$. This condition may be shown to be equivalent to the following nonnegative-definiteness condition in the frequency domain, valid for each ν:

$$\int F^*(\mathbf{r}_j)w(\mathbf{r}_j, \mathbf{r}_k, \nu)F(\mathbf{r}_k)\, d^3\mathbf{r}_j\, d^3\mathbf{r}_k \geq 0.$$

Here $F(\mathbf{r})$ is any arbitrary, sufficiently well behaved function of $\mathbf{r}$.

This equivalence is essentially a generalization of a theorem of Bochner [S. Bochner, *Lectures on Fourier Integrals* (Princeton University Press, Princeton, New Jersey, 1959), p. 326] well known in the theory of characteristic functions. [See for example E. Lukacs, *Characteristic Functions* (Charles Griffin and Company, Ltd., London, 1960), p. 62.]

[11] Y. Kano and E. Wolf, Proc. Phys. Soc. (London) **80**, 1273 (1962).

and Parrent[12] and is essentially equivalent to a representation of the S-matrix found by van Kampen.[13] The result (1.12) is an immediate consequence of the defining Eqs. (1.1) and (1.2). Parrent asserted[4] and claimed to have proved that the condition (1.12) implies that all the product terms in (1.11) are absent or, what amounts to the same thing, that the analytic continuation of $f(\tau)$ has no zeros at all in the appropriate half of the complex τ-plane. However, that the condition (1.12) cannot, in fact, be sufficient for absence of the zeros is evident by considering the function[14]

$$f(\tau) = \frac{\tau - a}{\tau - a^*} \cdot \frac{\tau + a^*}{\tau + a} \exp(-2\pi i \nu_0 \tau), \qquad (1.13)$$

where ν_0 is a positive constant and a is a constant with a negative imaginary part. This function is an analytic signal, is unimodular on the real τ-axis and obeys the condition (1.12). It is seen to contain two zeros in the lower half of the complex τ-plane (where $f(\tau)$ is analytic and regular), namely at the points $\tau = a$ and $\tau = -a^*$.

In the present paper we will show that the real reason for the absence of the zeros in the case where $|\gamma(\tau)| \equiv 1$ is the fact that γ is a nonnegative definite function, in the sense of Eq. (1.10). In Sec. 2 we derive this nonnegative definiteness condition. In Sec. 3 we consider the case where the condition $|\gamma(\mathbf{r}, \mathbf{r}, \tau)| \equiv 1$ holds for all real values of τ and for a fixed point $\mathbf{r}$. In Sec. 4 we consider first the case where the condition $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \equiv 1$ holds for all real values of τ and for a particular pair of values of $\mathbf{r}_1$ and $\mathbf{r}_2$. The case when the condition holds for all values of $\mathbf{r}_1$ and $\mathbf{r}_2$ in some domain of space is also considered in Sec. 4.

2. NONNEGATIVE DEFINITENESS CONDITIONS SATISFIED BY $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$

In this section we derive the nonnegative definiteness condition satisfied by $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ and discuss some of its consequences which will be needed in the later parts of this paper.

Let

$$F(t) = \sum_{i=1}^{n} \beta_i V(\mathbf{r}_i, t + \tau_i), \qquad (2.1)$$

where $V(\mathbf{r}, t)$ is the complex scalar wavefunction (represented in terms of an analytic signal) of a stationary optical field; $\mathbf{r}_i$, τ_i are arbitrary values of the parameters $\mathbf{r}$ and τ, respectively; β_i are arbitrary complex parameters and n is any integer greater than or equal to unity. We then have

$$\begin{aligned} 0 &\leq \langle F(t)F^*(t)\rangle \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} \beta_j\beta_k^* \langle V(\mathbf{r}_j, t + \tau_j)V^*(\mathbf{r}_k, t + \tau_k)\rangle \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} \beta_j\beta_k^* \Gamma(\mathbf{r}_j, \mathbf{r}_k, \tau_j - \tau_k). \end{aligned} \qquad (2.2)$$

Since according to (1.1),

$$\gamma(\mathbf{r}_j, \mathbf{r}_k, \tau) = \Gamma(\mathbf{r}_j, \mathbf{r}_k, \tau)/\{\Gamma(\mathbf{r}_j, \mathbf{r}_j, 0)\Gamma(\mathbf{r}_k, \mathbf{r}_k, 0)\}^{\frac{1}{2}}, \qquad (2.3)$$

(2.2) implies that

$$\sum_{j=1}^{n}\sum_{k=1}^{n} \gamma(\mathbf{r}_j, \mathbf{r}_k, \tau_j - \tau_k)\alpha_j\alpha_k^* \geq 0 \qquad (2.4)$$

$[\alpha_j = \{\Gamma(\mathbf{r}_j, \mathbf{r}_j, 0)\}^{\frac{1}{2}}\beta_j]$. Since the parameters β_j in (2.2) are arbitrary, so are the parameters α_j in (2.4).

For $n = 1$, (2.4) expresses the obvious condition that $\gamma(\mathbf{r}, \mathbf{r}, 0) \geq 0$. For $n = 2$, it is equivalent to the statement that, in addition, one has for all values of $\mathbf{r}_1$, $\mathbf{r}_2$, and τ

$$\gamma(\mathbf{r}_1, \mathbf{r}_1, 0)\gamma(\mathbf{r}_2, \mathbf{r}_2, 0) - |\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)|^2 \geq 0, \qquad (2.5)$$

where the "Hermiticity condition"

$$\gamma(\mathbf{r}_j, \mathbf{r}_k, \tau_j - \tau_k) = \gamma^*(\mathbf{r}_k, \mathbf{r}_j, \tau_k - \tau_j), \qquad (2.6)$$

which follows from the defining Eqs. (1.1) and (1.2) for γ, has been used. This result is also obvious, since

$$\gamma(\mathbf{r}_1, \mathbf{r}_1, 0) = \gamma(\mathbf{r}_2, \mathbf{r}_2, 0) = 1 \qquad (2.7a)$$

and

$$|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \leq 1. \qquad (2.7b)$$

The consequence of (2.4) for the case $n = 3$ is, however, not quite so obvious and is of immediate interest for our problem. When $n = 3$, the condition (2.4) implies that the 3×3 *Hermitian matrix*

$$\begin{bmatrix} 1 & \gamma(\mathbf{r}_1, \mathbf{r}_2, \tau_1 - \tau_2) & \gamma(\mathbf{r}_1, \mathbf{r}_3, \tau_1 - \tau_3) \\ \gamma(\mathbf{r}_2, \mathbf{r}_1, \tau_2 - \tau_1) & 1 & \gamma(\mathbf{r}_2, \mathbf{r}_3, \tau_2 - \tau_3) \\ \gamma(\mathbf{r}_3, \mathbf{r}_1, \tau_3 - \tau_1) & \gamma(\mathbf{r}_3, \mathbf{r}_2, \tau_3 - \tau_2) & 1 \end{bmatrix} \qquad (2.8)$$

is nonnegative for all real values of τ_1, τ_2, τ_3 and for all sets of any three points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ in the space

---

[12] S. F. Edwards and G. B. Parrent, Jr., Opt. Acta 6, 367 (1959).

[13] N. G. van Kampen, Phys. Rev. 89, 1072 (1953).

[14] There are a number of algebraic errors in Parrent's proof[4]. In addition, the contribution from a large semicircle, to a contour integral leading to his Eq. (A9) is incorrectly neglected.

The theorem has also been discussed in a recent book by Beran and Parrent (Ref. 6). The derivation makes use of another theorem (their Proposition B, p. 49) which is incorrect.

domain D for which γ is defined. This condition implies, in particular, that the determinant of the matrix (2.8) is nonnegative, i.e., that

$$\begin{vmatrix} 1 & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & 1 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & 1 \end{vmatrix} \geq 0, \tag{2.9}$$

where, for the sake of convenience, we have used the abbreviation

$$\gamma_{jk} \equiv \gamma(\mathbf{r}_j, \mathbf{r}_k, \tau_j - \tau_k). \tag{2.10}$$

On evaluating the determinant (2.9) and making use of the fact that $\gamma_{jk} = \gamma_{kj}^*$, we obtain the relation

$$(1 - |\gamma_{12}|^2)(1 - |\gamma_{31}|^2) - |\gamma_{32} - \gamma_{12}\gamma_{31}|^2 \geq 0. \tag{2.11}$$

Eq. (2.11), when written in full, expresses the condition that

$$|\gamma(\mathbf{r}_3, \mathbf{r}_2, \tau_3 - \tau_2) - \gamma(\mathbf{r}_1, \mathbf{r}_2, \tau_1 - \tau_2)\gamma(\mathbf{r}_3, \mathbf{r}_1, \tau_3 - \tau_1)|^2 \leq (1 - |\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau_1 - \tau_2)|^2)(1 - |\gamma(\mathbf{r}_3, \mathbf{r}_1, \tau_3 - \tau_1)|^2), \tag{2.12}$$

for all real values of τ_1, τ_2, τ_3 and for all sets of any three points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ in the domain D.

3. THE MOST GENERAL UNIMODULAR DEGREE OF SELF COHERENCE $\gamma(\mathbf{R}, \mathbf{R}, \tau)$

Suppose that $|\gamma(\mathbf{R}, \mathbf{R}, \tau)| \equiv 1$ for all real values of τ $(-\infty < \tau < \infty)$ and for some fixed point $\mathbf{r} = \mathbf{R}$. We will show that $\gamma(\mathbf{R}, \mathbf{R}, \tau)$ must then necessarily be of the form

$$\gamma(\mathbf{R}, \mathbf{R}, \tau) = \exp(-2\pi i \nu_0 \tau), \tag{3.1}$$

where ν_0 is a positive constant. In the process of the proof, it will also be seen that the complex degree of coherence $\gamma(\mathbf{r}, \mathbf{R}, \tau)$ is also strictly periodic in τ, i.e., that it has the form

$$\gamma(\mathbf{r}, \mathbf{R}, \tau) = \gamma(\mathbf{r}, \mathbf{R}, 0) \exp(-2\pi i \nu_0 \tau). \tag{3.2}$$

Here $\mathbf{r}$ is the position vector of an arbitrary point in the space domain D for which $\gamma(\mathbf{r}, \mathbf{R}, \tau)$ is defined and $\mathbf{R}$ is, as before, the position vector of the point for which $\gamma(\mathbf{R}, \mathbf{R}, \tau)$ is unimodular for all real τ-values.

Since $|\gamma(\mathbf{R}, \mathbf{R}, \tau)| \equiv 1$ for all real values of τ, we may write

$$\gamma(\mathbf{R}, \mathbf{R}, \tau) = \exp[i\phi(\tau)], \qquad (\tau \text{ real}), \tag{3.3}$$

where ϕ is a real function of τ. Moreover, from the defining Eqs. (1.1) and (1.2), it follows that

$$\phi(-\tau) = -\phi(\tau) + 2m\pi, \tag{3.4}$$

where m is any integer.

Now as shown in the previous section, γ satisfies the inequality (2.12), viz.

$$|\gamma(\mathbf{r}_3, \mathbf{r}_2, \tau_3 - \tau_2) - \gamma(\mathbf{r}_1, \mathbf{r}_2, \tau_1 - \tau_2)\gamma(\mathbf{r}_3, \mathbf{r}_1, \tau_3 - \tau_1)|^2 \leq [1 - |\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau_1 - \tau_2)|^2][1 - |\gamma(\mathbf{r}_3, \mathbf{r}_1, \tau_3 - \tau_1)|^2] \tag{3.5}$$

for all real τ_1, τ_2, τ_3 and for all triads of points $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ in the domain D for which γ is defined. In particular if we set $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{R}$ and $\mathbf{r}_3 = \mathbf{r}$, this condition implies that

$$|\gamma(\mathbf{r}, \mathbf{R}, \tau_3 - \tau_2) - \gamma(\mathbf{R}, \mathbf{R}, \tau_1 - \tau_2)\gamma(\mathbf{r}, \mathbf{R}, \tau_3 - \tau_1)|^2 \leq 0. \tag{3.6}$$

Here we have made use of the fact that $|\gamma(\mathbf{R}, \mathbf{R}, \tau)| \equiv 1$ for all real values of τ.

Since the left-hand side of (3.6) is always *nonnegative*, only the equality sign can hold, and we obtain the relation

$$\gamma(\mathbf{r}, \mathbf{R}, \tau_3 - \tau_2) = \gamma(\mathbf{R}, \mathbf{R}, \tau_1 - \tau_2)\gamma(\mathbf{r}, \mathbf{R}, \tau_3 - \tau_1), \tag{3.7}$$

which must be valid for all real values of τ_1, τ_2, τ_3 and for any arbitrary point $\mathbf{r}$ [but with fixed $\mathbf{R}$ for which $|\gamma(\mathbf{R}, \mathbf{R}, \tau)| \equiv 1$].

If we set $\tau_1 = \tau_3 = \tau_2 + \tau$ in (3.7), it follows that

$$\gamma(\mathbf{r}, \mathbf{R}, \tau) = \gamma(\mathbf{R}, \mathbf{R}, \tau)\gamma(\mathbf{r}, \mathbf{R}, 0). \tag{3.8}$$

Next let us set $\mathbf{r} = \mathbf{R}$ in (3.7) and use also Eqs. (3.3) and (3.4). We then obtain the following functional equation for ϕ:

$$\phi(\tau_1 - \tau_2) + \phi(\tau_2 - \tau_3) + \phi(\tau_3 - \tau_1) = 2m\pi, \tag{3.9}$$

where m is an arbitrary integer. To solve the functional Eq. (3.9), we first differentiate (3.9) with respect to τ_1 and then set $\tau_1 = \tau_3 = \tau_2 + \tau$. We then obtain

$$\phi'(\tau) = -2\pi\nu_0, \tag{3.10}$$

where prime denotes differentiation with respect to the parameter τ and $\nu_0 = -(1/2\pi)\phi'(0)$ is a constant. Integration of (3.10) gives

$$\phi(\tau) = -2\pi\nu_0\tau + \beta, \tag{3.11}$$

where β is a constant, and Eq. (3.3) now gives

$$\gamma(\mathbf{R}, \mathbf{R}, \tau) = \exp[i\beta - 2\pi i\nu_0\tau]. \tag{3.12}$$

However, since $\gamma(\mathbf{R}, \mathbf{R}, 0) = 1$, β must be an integral multiple of 2π, and so we finally obtain

$$\gamma(\mathbf{R}, \mathbf{R}, \tau) = \exp(-2\pi i\nu_0\tau). \tag{3.13}$$

Now the normalized power spectrum $w(\mathbf{R}, \mathbf{R}, \nu)$

is, according to (1.3), the Fourier transform of γ, so that when γ is given by (3.13),

$$w(\mathbf{R}, \mathbf{R}, \nu) = \delta(\nu - \nu_0), \tag{3.14}$$

where δ denotes the Dirac delta function. Since w can be nonzero only for the positive values of ν, the constant ν_0 in (3.13) and (3.14) must be positive.

On substitution from (3.13) in (3.8), we obtain the relation

$$\gamma(\mathbf{r}, \mathbf{R}, \tau) = \gamma(\mathbf{r}, \mathbf{R}, 0) \exp(-2\pi i\nu_0\tau). \tag{3.15a}$$

From (3.15a) and (2.6) it follows that one also has the relation

$$\gamma(\mathbf{R}, \mathbf{r}, \tau) = \gamma(\mathbf{R}, \mathbf{r}, 0) \exp(-2\pi i\nu_0\tau). \tag{3.15b}$$

We may summarize the results which we have now established in the following:

Theorem 1:

If for some point $\mathbf{R}$ and for all real values of τ $(-\infty < \tau < \infty)$ the complex degree of self coherence $\gamma(\mathbf{R}, \mathbf{R}, \tau)$ is unimodular, then $\gamma(\mathbf{R}, \mathbf{R}, \tau)$ is necessarily of the form

$$\gamma(\mathbf{R}, \mathbf{R}, \tau) = \exp(-2\pi i\nu_0\tau), \tag{3.16}$$

where ν_0 is a positive constant.[15] Moreover, if $\mathbf{r}$ is any other point in the region D for which the complex degree of coherence is defined, then $\gamma(\mathbf{r}, \mathbf{R}, \tau)$ is also necessarily periodic in τ and is of the form

$$\gamma(\mathbf{r}, \mathbf{R}, \tau) = \gamma(\mathbf{r}, \mathbf{R}, 0) \exp(-2\pi i\nu_0\tau), \tag{3.17}$$

and similarly,

$$\gamma(\mathbf{R}, \mathbf{r}, \tau) = \gamma(\mathbf{R}, \mathbf{r}, 0) \exp(-2\pi i\nu_0\tau). \tag{3.18}$$

4. THE MOST GENERAL UNIMODULAR DEGREE OF COHERENCE $\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau)$

Suppose that $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \equiv 1$ for all real values of τ $(-\infty < \tau < \infty)$ and for some fixed space points $\mathbf{r}_1 = \mathbf{R}_1$, $\mathbf{r}_2 = \mathbf{R}_2$ in some space domain D where γ is defined. We will show that in this case $\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau)$ is also strictly periodic in τ. In the process of the proof, it will be seen that both $\gamma(\mathbf{R}_1, \mathbf{R}_1, \tau)$ and $\gamma(\mathbf{R}_2, \mathbf{R}_2, \tau)$ must necessarily be also unimodular for all real values of τ.

We begin in a similar way as before. We set $\mathbf{r}_1 = \mathbf{R}_1$, $\mathbf{r}_2 = \mathbf{R}_2$ and $\mathbf{r}_3 = \mathbf{r}$ in (2.12) and use the fact that $|\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau)| \equiv 1$. We then obtain the inequality

$$|\gamma(\mathbf{r}, \mathbf{R}_2, \tau_3 - \tau_2) - \gamma(\mathbf{R}_1, \mathbf{R}_2, \tau_1 - \tau_2)\gamma(\mathbf{r}, \mathbf{R}_1, \tau_3 - \tau_1)|^2 \leq 0, \tag{4.1}$$

which implies that the equality

$$\gamma(\mathbf{r}, \mathbf{R}_2, \tau_3 - \tau_2) = \gamma(\mathbf{R}_1, \mathbf{R}_2, \tau_1 - \tau_2)\gamma(\mathbf{r}, \mathbf{R}_1, \tau_3 - \tau_1) \tag{4.2}$$

holds for all real values of τ_1, τ_2, τ_3 and for any $\mathbf{r}$ in the domain D. On equating the moduli of both sides of (4.2), setting $\mathbf{r} = \mathbf{R}_1$ and again using the fact that $|\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau)| \equiv 1$, we obtain the relation

$$|\gamma(\mathbf{R}_1, \mathbf{R}_1, \tau)| = 1 \tag{4.3}$$

valid for all real values of τ, where $\tau = \tau_3 - \tau_1$. In a similar way it can be seen that $\gamma(\mathbf{R}_2, \mathbf{R}_2, \tau)$ is also unimodular for all real values of τ.

Having proved that the complex degree of self coherence $\gamma(\mathbf{R}_1, \mathbf{R}_1, \tau)$ is unimodular, we may now use the results of Theorem 1 established in Sec. 3. We then have the result that for all real τ

$$\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau) = \gamma(\mathbf{R}_1, \mathbf{R}_2, 0) \exp(-2\pi i\nu_0\tau), \tag{4.4}$$

where ν_0 is a positive constant. However, since $\gamma(\mathbf{R}_1, \mathbf{R}_2, 0)$ is itself unimodular, it follows that

$$\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau) = \exp[i\beta - 2\pi i\nu_0\tau], \tag{4.5}$$

where β is a real constant.

Finally setting $\tau_1 = \tau_2 = \tau_3 - \tau$ in (4.2), we obtain the following result:

$$\gamma(\mathbf{r}, \mathbf{R}_2, \tau) = \gamma(\mathbf{R}_1, \mathbf{R}_2, 0)\gamma(\mathbf{r}, \mathbf{R}_1, \tau). \tag{4.6}$$

We can summarize the results which we just obtained in the following.

Theorem 2:

If for some pair of values $\mathbf{r}_1 = \mathbf{R}_1$ and $\mathbf{r}_2 = \mathbf{R}_2$ and for all real values of τ $(-\infty < \tau < \infty)$, the complex degree of coherence $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ is unimodular, then $\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau)$ is necessarily periodic in τ. More precisely, it must have the form

$$\gamma(\mathbf{R}_1, \mathbf{R}_2, \tau) = \exp(i\beta - 2\pi i\nu_0\tau), \tag{4.7}$$

where β and ν_0 are real constants and $\nu_0 > 0$. Moreover, the following relations also hold for all real values of τ:

$$\gamma(\mathbf{R}_1, \mathbf{R}_1, \tau) = \gamma(\mathbf{R}_2, \mathbf{R}_2, \tau) = \exp(-2\pi i\nu_0\tau), \tag{4.8}$$

$$\gamma(\mathbf{r}, \mathbf{R}_2, \tau) = \gamma(\mathbf{R}_1, \mathbf{R}_2, 0)\gamma(\mathbf{r}, \mathbf{R}_1, \tau), \tag{4.9}$$

where $\mathbf{r}$ is an arbitrary point in the space domain D in which γ is defined.

Lastly let us suppose that $|\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)| \equiv 1$ for all real τ values $(-\infty < \tau < \infty)$, and that this

[15] According to a theorem on characteristic functions (cf. p. 25, Corol. 2 in the book by E. Lukacs, quoted in Ref. 10) the condition $|\gamma(R, R, \tau)| \equiv 1$ for all τ may be replaced by the weaker condition $|\gamma(R, R, \tau)| = 1$ for any two incommensurable real values τ_1 and τ_2 of τ.

condition is satisfied not only for a particular pair of points $\mathbf{r}_1$ and $\mathbf{r}_2$, but for all pairs of points in some domain D of space. We will show that under these circumstances the parameter β in (4.7) must have the form $\alpha(\mathbf{r}_1) - \alpha(\mathbf{r}_2)$, where $\alpha(\mathbf{r})$ is a real function of $\mathbf{r}$, while ν_0 remains a positive constant.

We again begin with Eq. (2.12) which implies in particular, that under the conditions just stated,

$$\gamma(\mathbf{r}_3, \mathbf{r}_2, 0) = \gamma(\mathbf{r}_1, \mathbf{r}_2, 0)\gamma(\mathbf{r}_3, \mathbf{r}_1, 0), \tag{4.10}$$

where $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ are any three points in the domain D. Writing

$$\gamma(\mathbf{r}_j, \mathbf{r}_k, 0) = \exp\{i\phi(\mathbf{r}_j, \mathbf{r}_k, 0)\},$$
$$(j, k = 1, 2, 3; \phi \text{ real}), \tag{4.11}$$

we obtain

$$\phi(\mathbf{r}_1, \mathbf{r}_2, 0) + \phi(\mathbf{r}_3, \mathbf{r}_1, 0) = \phi(\mathbf{r}_3, \mathbf{r}_2, 0) + 2m\pi, \tag{4.12}$$

where m is an arbitrary integer. Setting $\mathbf{r}_2 = 0$ in Eq. (4.12) (which is always permissible by suitable choice of the origin in D), we obtain

$$\phi(\mathbf{r}_3, \mathbf{r}_1, 0) = \alpha(\mathbf{r}_3) - \alpha(\mathbf{r}_1) + 2m\pi, \tag{4.13}$$

where $\alpha(\mathbf{r}) = \phi(\mathbf{r}, 0, 0)$ is a function of $\mathbf{r}$ only. Hence (4.11) gives

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, 0) = \exp\{i[\alpha(\mathbf{r}_1) - \alpha(\mathbf{r}_2)]\}, \tag{4.14}$$

and if (4.7) is also used, we obtain the relation

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \exp\{i[\alpha(\mathbf{r}_1) - \alpha(\mathbf{r}_2) - 2\pi\nu_0\tau]\}. \tag{4.15}$$

We may summarize the result, which we just obtained, in the following:

Theorem 3:

If for all values of $\mathbf{r}_1$ and $\mathbf{r}_2$ in a domain D of space and for all real τ values $(-\infty < \tau < \infty)$, the complex degree of coherence $\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau)$ is unimodular, then it must necessarily be of the form

$$\gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = \exp\{i[\alpha(\mathbf{r}_1) - \alpha(\mathbf{r}_2) - 2\pi\nu_0\tau]\}, \tag{4.16}$$

where $\alpha(\mathbf{r})$ is a real function of $\mathbf{r}$ and ν_0 is a positive constant.

We note that when the conditions of this theorem are satisfied, (4.16) and (1.1) imply that the mutual coherence function has the "factorized", time-periodic form

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2, \tau) = U(\mathbf{r}_1)U^*(\mathbf{r}_2)\exp(-2\pi i\nu_0\tau), \tag{4.17}$$

where

$$U(\mathbf{r}) = \{\Gamma(\mathbf{r}, \mathbf{r}, 0)\}^{\frac{1}{2}} e^{i\alpha(\mathbf{r})} = \langle V(\mathbf{r}, t)V^*(\mathbf{r}, t)\rangle^{\frac{1}{2}} e^{i\alpha(\mathbf{r})}. \tag{4.18}$$

Application of Houston's Method to the Sum of Plane Waves over the Brillouin Zone. I. Simple-Cubic and Face-Centered-Cubic Lattices*

MARIA MIASEK†
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received 22 April 1965)

In certain problems in solid-state physics, the radial functions $g_j(r)$ in the expansion $\chi(\mathbf{r}) = \sum_{j=0}^{\infty} g_j(r)K_j(\theta, \varphi)$, where $\chi(\mathbf{r})$ is a known function and the K_j's are Kubic Harmonics, are of interest. This paper deals with the functions $\chi(\mathbf{r}) \equiv N^{-1} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}$, where the sum runs over the first Brillouin Zone of a crystal. In particular, the functions $\chi(\mathbf{r})$ for simple cubic and face-centered cubic lattices are expanded into series of Kubic Harmonics and the radial functions $g_j(r)$ for several values of j are found using Houston's method, in which the expansion into series of Kubic Harmonics contains only a finite number of terms with lowest j's. $g_0(r)$ is calculated using 3, 6, and 9-term expansion, $g_2(r)$ and $g_3(r)$ using only 3 and 6-term expansion. Comparing $g_j(r)$ obtained from the formulas with different numbers of terms it is established that for r in the region $\langle 0, 2a\rangle$, where a is the lattice constant, the 6-term approximation is very good. In practice, the functions $g_j(r)$ usually occur in integrands, together with atomic orbitals, and the tabulated results are expected to be particularly useful in the study of Wannier functions in the OPW scheme.

I. INTRODUCTION: HOUSTON'S METHOD

MUCH of the present paper is based on an approximation method introduced by Houston.[1] Since this method is highly useful but does not appear to be widely appreciated today, we begin with a sketch of the method. It was given for the first time for determining the frequency distribution function of cubic crystals. Afterwards it was improved in a series of papers[2-6] and used extensively in the theory of thermodynamic properties of crystals with appropriate cubic symmetry.

In this theory integrals of the following type occur:

$$I(k) = \int_0^{2\pi} \int_0^{\pi} F(k_x, k_y, k_z) \sin\theta \, d\theta \, d\varphi, \qquad (1.1)$$

and

$$I = \iiint_{B.z.} F(k_x, k_y, k_z) \, dk_x \, dk_y \, dk_z. \qquad (1.1')$$

The integration in (1.1) is taken over the solid angle in **k**-space. For the integration, the variables k_x, k_y, k_z in the function $F(\mathbf{k})$ must be substituted by the spherical variables k, θ, φ. The integration in (1.1') is taken over the entire volume of Brillouin zone. When the function $F(\mathbf{k})$ is invariant under all symmetry operations of the complete cubic group O_h, it results from symmetry that in order to calculate (1.1) or (1.1') it is sufficient to know $F(\mathbf{k})$ in only 1/48th of the Brillouin zone. We will be concerned only with functions of this type.

In the theory of thermodynamic properties the functions $F(\mathbf{k})$ that occur are not in fact known in the whole 1/48th of the Brillouin zone, but only along some directions in this space. With functions so meagerly determined one cannot easily find good approximate values of $I(k)$ and I.

In order to pass to simple one dimensional integrals in (1.1) and (1.1'), the function $F(\mathbf{k})$ may be expanded into a series of Kubic Harmonics $K_j(\theta, \varphi)$ with the same transformation properties as $F(\mathbf{k})$[7]:

$$F(\mathbf{k}) = \sum_{j=0}^{\infty} f_j(k)K_j(\theta, \varphi). \qquad (1.2)$$

The invariance of $F(\mathbf{k})$ under all symmetry operations of cubic group means that $F(\mathbf{k})$ transforms according to irreducible representation Γ_1 of cubic group O_h.[8] Kubic Harmonics $K_j(\theta, \varphi)$ are determined for several lowest values of j.[3,7] They are orthogonal and may be normalized, so they satisfy the condition

$$\int_0^{2\pi} \int_0^{\pi} K_i^*(\theta, \varphi)K_j(\theta, \varphi) \sin\theta \, d\theta \, d\varphi = \delta_{ij}. \qquad (1.3)$$

* Supported in part by a Grant (62-145) from the U. S. Air Force Office of Scientific Research.
† Present address: Institute of Physics, Polish Academy of Science, Warsaw, Poland.
[1] W. V. Houston, Revs. Mod. Phys. 20, 161 (1948).
[2] A. B. Bhatia, Phys. Rev. 97, 363 (1955).
[3] D. D. Betts, A. B. Bhatia, and M. Wyman, Phys. Rev. 104, 37 (1956).
[4] D. D. Betts, A. B. Bhatia, and G. K. Horton, Phys. Rev. 104, 43 (1956).
[5] D. D. Betts, Can. J. Phys. 39, 233 (1961).
[6] S. Ganesan and R. Srinivasan, Can. J. Phys. 40, 1153 (1962).
[7] F. C. Von der Lage and H. A. Bethe, Phys. Rev. 71, 612 (1947).
[8] We use here the name Γ_1 for this representation. In notation used by Von der Lage and Bethe, it is called α.

The functions $f_i(k)$ are expressed by the integrals

$$f_i(k) = \int_0^{2\pi} \int_0^{\pi} K_i^*(\theta, \varphi) F(\mathbf{k}) \sin\theta \, d\theta \, d\varphi. \quad (1.4)$$

When the function $F(\mathbf{k})$ is determined only for some directions in $\mathbf{k}$-space one cannot perform the exact integration in (1.4). But it would be very simple to calculate $I(k)$ and I if we could find $f_i(k)$ in any possible way.

Houston gave a method which enables us to find $f_i(k)$ for several lowest j without using the formula (1.4). For the definite direction s the expansion (1.2) can be written

$$F_s(k_x, k_y, k_z) = F(k, \theta_s, \varphi_s) = \sum_{i=0}^{\infty} f_i(k) K_i(\theta_s, \varphi_s). \quad (1.5)$$

On the right-hand side of this expression we have in fact an infinite sum over j. Houston's idea was to use instead of the infinite sum a finite one containing a number of terms equal to the number of nonequivalent directions along which the function $F(\mathbf{k})$ was actually known. If the function $F(\mathbf{k})$ is known for n nonequivalent directions, we write

$$F(k, \theta_s, \varphi_s) = \sum_{i=0}^{n-1} f_i(k) K_i(\theta_s, \varphi_s), \qquad s = 1, 2 \cdots n. \quad (1.6)$$

The expressions (1.6) form a set of algebraic linear equations with n unknown functions $f_i(k)$. Solving this set of equations the functions $f_i(k)$ are given as linear combinations of $F(k, \theta_s, \varphi_s)$. The more directions we can use in formulas (1.6) the better an approximation for $f_i(k)$ is obtained. Near the origin, for small k, the lines used are close together and quite a small number of directions yields good approximation for $f_i(k)$ in this region; but when k increases one must use more directions in order to get a "good" $f_i(k)$. Houston used his method to consider functions of $\mathbf{k}$-vectors in the Brillouin zone, that is a region of relatively small $\mathbf{k}$ in the entire $\mathbf{k}$-space. Furthermore, a normalization factor was included assuring better accuracy of the calculated integral I.[2,6]

A similar procedure can be applied to functions $G(x, y, z)$ in r-space if we are interested in values of $G(\mathbf{r})$ only in the region of small r.

In Sec. II the algebraic details of the truncated expansion are worked out for future reference, and in Sec. III the particular function $\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}$ is studied. In Secs. IV and V we examine some more general functions $G(\mathbf{r})$ of interest for solid state problems. Using Houston's method we find the radial functions of expansion into series of Kubic Harmonics, and we discuss their usefulness in calculations.

II. SOLUTIONS OF THE SET OF EQUATIONS FOR THE RADIAL EXPANSION FUNCTIONS $g_i(r)$

We consider here a function $G(\mathbf{r})$ which transforms according to the irreducible representation Γ_1 of the group O_h. We take into account several directions along which $G(\mathbf{r})$ is known, and for every direction s we expand it according to formula (1.6):

$$G_s(r) = G(r, \theta_s, \varphi_s) = \sum_{i=0}^{n-1} g_i(r) K_i(\theta_s, \varphi_s). \quad (2.1)$$

In our considerations we use nine directions which, for cubic symmetry, are contained in an appropriate 1/48th of the total solid angle[9]:

A : [100]	D : [210]	G : [411]
B : [110]	E : [211]	H : [431]
C : [111]	F : [221]	I : [433].

K_i's are taken normalized. It is well known that K_i with $j = 1$ does not appear for the irreducible representation Γ_1.[3]

In previous papers[1,3,5] formulas for the function $g_0(r)$ were given using 3, 6, 9, and 15-term expansions in (2.1) and also for different combinations of 4 and 5 terms. We write here these formulas for $g_0(r)$ as 3, 6, and 9-term expressions:

$$g_0^{(3)}(r) = (4\pi)^{\frac{1}{2}}(1/35)[10G_A(r) + 16G_B(r) + 9G_C(r)], \quad (2.2)$$

$$g_0^{(6)}(r) = (4\pi)^{\frac{1}{2}}[0.108782G_A(r) + 0.070795G_B(r) + 0.016177G_C(r) + 0.352674G_D(r) + 0.287697G_E(r) + 0.163888G_F(r)], \quad (2.2')$$

$$g_0^{(9)}(r) = (4\pi)^{\frac{1}{2}}[0.056378G_A(r) + 0.049527G_B(r) - 0.055886G_C(r) + 0.178271G_D(r) + 0.102948G_E(r) + 0.073099G_F(r) + 0.207116G_G(r) + 0.213763G_H(r) + 0.174784G_I(r)]. \quad (2.2'')$$

The upper index indicates the number of directions used in the calculation of $g_0(r)$. The functions $G_A(r)$, $G_B(r)$, $\cdots$, $G_I(r)$ are determined for the directions A, B, $\cdots$, I. If the function $G(\mathbf{r})$ is given in an analytical form, in order to get $G_A(r)$, $\cdots$ $G_I(r)$,

[9] We use here the directions given in Refs. 3 and 5. Betts showed in his paper how these directions are situated in 1/48 of the total solid angle.

we simply pass from the Cartesian coordinates x, y, z to the spherical coordinates r, θ, φ, and subsequently put into $G(r, \theta, \varphi)$ the appropriate values of θ_s and φ_s for the given direction s.

In the present paper formulas were found also for $g_2(r)$ and $g_3(r)$—the second and the third radial functions in expansion (2.1). g_2 and g_3 are given as 3 and 6-term expressions:

$$g_2^{(3)}(r) = (4\pi)^{\frac{1}{2}} \frac{2}{5\cdot 11(3\cdot 7)^{\frac{1}{2}}} \times [35G_A(r) - 8G_B(r) - 27G_C(r)], \quad (2.3)$$

$$g_2^{(6)}(r) = (4\pi)^{\frac{1}{2}}[0.193843G_A(r) - 0.129652G_B(r) - 0.075624G_C(r) + 0.247791G_D(r) - 0.015520G_E(r) - 0.220839G_F(r)], \quad (2.3')$$

and

$$g_3^{(3)}(r) = (4\pi)^{\frac{1}{2}} \frac{3\cdot 8}{7\cdot 11(2\cdot 13)^{\frac{1}{2}}}[G_A(r) - 4G_B(r) + 3G_C(r)], \quad (2.4)$$

$$g_3^{(6)}(r) = (4\pi)^{\frac{1}{2}}[0.081883G_A(r) - 0.211995G_B(r) + 0.147168G_C(r) - 0.088535G_D(r) + 0.066324G_E(r) + 0.005155G_F(r)]. \quad (2.4')$$

One may notice that the sum of the numerical coefficients for $g_0(r)$ is equal to unity, and for $g_2(r)$ and $g_3(r)$ it is equal to zero.

III. THE APPLICATION OF HOUSTON'S METHOD TO THE SUM OF PLANE WAVES OVER THE BRILLOUIN ZONE

In some problems of solid state physics the following integrals occur[10]:

$$W = (1/N) \sum_{\mathbf{k}} \int e^{i\mathbf{k}\cdot\mathbf{r}} \Phi^*(\mathbf{r})\, d\mathbf{r}. \quad (3.1)$$

Here the sum over $\mathbf{k}$ is taken over the entire volume of the Brillouin zone. $\Phi(\mathbf{r})$ is a function which transforms according to the irreducible representation Γ_1 of the cubic group O_h. Therefore, $\Phi(\mathbf{r})$ may be expanded into series of Kubic harmonics K_i:

$$\Phi(\mathbf{r}) = \sum_i \varphi_i(r) K_i(\theta, \varphi). \quad (3.2)$$

We shall wish to consider only functions $\Phi(\mathbf{r})$ for which the radial functions $\varphi_i(r)$ are different from zero for small r and are vanishing elsewhere (e.g., atomic wavefunctions). Since the Kubic Harmonics are linear combinations of the spherical harmonics, instead of (3.2) we may write

$$\Phi(\mathbf{r}) = \sum_l \sum_m \varphi_{lm}(r) Y_{lm}(\theta, \varphi). \quad (3.3)$$

φ_{lm} is given as an appropriate $\varphi_i(r)$, where $l = 2i$, multiplied by a numerical factor c_{lm}, which is determined by symmetry considerations.

If we wanted to find the value of (3.1) in the standard way, we should use (3.3) for the function $\Phi(\mathbf{r})$ and also expand the plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ in a series of Legendre polynomials with spherical Bessel functions $j_l(kr)$ as the radial functions of this expansion. After such a substitution the following expression for W would be obtained:

$$W = \sum_l \sum_m [\sum_{\mathbf{k}} Y_{lm}(\theta_{\mathbf{k}}, \varphi_{\mathbf{k}}) G_{lm}(k)], \quad (3.4)$$

where

$$G_{lm}(k) = C \int_0^\infty j_l(kr)\varphi_{lm}^*(r) r^2\, dr. \quad (3.5)$$

Here C is a numerical factor. Calculating $G_{lm}(k)$ and afterwards summing over $\mathbf{k}$ in (3.4) is extremely laborious and ineffective. In many cases, W may be found in a very easy way using Houston's idea of expanding a function of cubic symmetry in a finite number of cubic harmonics.

Let us consider the function consisting of a sum of plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$, where the summation is taken over all $\mathbf{k}$ vectors in the Brillouin zone. We define

$$\chi(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (3.6)$$

One may usually pass here from the summation to an integration, and instead of (3.6) we write

$$\chi(\mathbf{r}) = \frac{\Omega_0}{(2\pi)^3} \iiint_{B.Z.} e^{i\mathbf{k}\cdot\mathbf{r}}\, d\mathbf{k}; \quad (3.7)$$

Ω_0 is the volume of unit cell. The integral in (3.7) can be integrated easily over the appropriate Brillouin zones for cubic crystals so that the function $\chi(\mathbf{r})$ is known in analytical form, which, as we shall see later, may be quite a complicated function of $\mathbf{r}$. $\chi(\mathbf{r})$ transforms according to the irreducible representation Γ_1 of the group O_h; therefore, the considerations of Secs. I and II apply to this function.

In order to find the value of W, we expand $\chi(\mathbf{r})$ into a series of Kubic Harmonics,

$$\chi(\mathbf{r}) = \sum_i g_i(r) K_i(\theta, \varphi). \quad (3.8)$$

Then substituting (3.2) and (3.8) into (3.1), we obtain

$$W = \sum_i \int \varphi_i^*(r) g_i(r) r^2\, dr. \quad (3.9)$$

[10] M. Miasek, Bull. Acad. Polon. Sci., Classe (III) 4, 453 (1956).

In (3.9) the sum over j for typical solid-state problems contains only a few terms. There are in fact two reasons for such a limitation of j. First, in practice the function $\Phi(\mathbf{r})$ usually involves only terms with lowest j's. [Eq. (3.2)]. Second, we shall see later that the values of functions $g_i(r)$ get smaller with increasing j especially in the region where the $\varphi_i(r)$ are determined, so that only first few terms in (3.9) decide about the value of W. In order to find W from (3.9) we need usually to calculate only a few one-dimensional integrals and in order to do that we must know the functions $g_i(r)$. It is not easy to get $g_i(r)$ from the exact formula

$$g_i(r) = \int_0^{2\pi} \int_0^{\pi} \chi^*(\mathbf{r}) K_i(\theta, \varphi) \sin\theta \, d\theta \, d\varphi, \qquad (3.10)$$

because $\chi(\mathbf{r})$ is a rather complicated function of θ and φ, but we can find without any trouble the functions $g_i(r)$ with the lowest j's using Houston's method.

In Secs. IV and V $g_i(r)$'s with $j = 0, 2, 3$ are calculated for simple cubic and face-centered cubic lattices.

IV. THE FUNCTIONS $g_i(\mathbf{r})$ FOR A SUM OF PLANE WAVES IN A SIMPLE CUBIC LATTICE

In fact, for the simple cubic lattice one can find from (3.10) the exact solutions for $g_i(r)$ in the form of a power series in r/a, where a is the lattice constant. The calculation of $g_i(r)$ using Houston's method is done here because of two reasons. The main reason is that for this lattice we can compare the results of Houston's method with the exact analytical ones. The other reason is that we can use the analytical form only when $r < a$. When r is not too large but $r \geq a$ the results of Houston's method are more useful. These $g_i(r)$ occur in calculations for CsCl-type structures.

The function $\chi(\mathbf{r})$ for simple cubic structure is obtained from (3.7) by integration over the Brillouin zone which is a cube with edges of length $2\pi/a$. We have the following expression for $\chi(\mathbf{r})$:

$$\chi^{sc}(\mathbf{r}) = \left(\frac{a}{\pi}\right)^3 \frac{\sin(\pi x/a)\sin(\pi y/a)\sin(\pi z/a)}{xyz}. \qquad (4.1)$$

In order to use $\chi(\mathbf{r})$ in formulas (2.3), (2.4), and (2.5), it must be found for the directions $A, B, \cdots I$. The appropriate functions are dependent only on the ratio $\alpha = r/a$ and have the following forms:

$$\chi_A^{sc} = \frac{\sin\varphi_A}{\varphi_A}, \qquad \varphi_A = \pi\alpha,$$

$$\chi_B^{sc} = \frac{\sin^2\varphi_B}{\varphi_B^2}, \qquad \varphi_B = \pi\alpha/2^{\frac{1}{2}},$$

$$\chi_C^{sc} = \frac{\sin^3\varphi_C}{\varphi_C^3}, \qquad \varphi_C = \pi\alpha/3^{\frac{1}{2}},$$

$$\chi_D^{sc} = \frac{\sin^2\varphi_D\cos\varphi_D}{\varphi_D^2}, \qquad \varphi_D = \pi\alpha/5^{\frac{1}{2}},$$

$$\chi_E^{sc} = \frac{\sin^3\varphi_E\cos\varphi_E}{\varphi_E^3}, \qquad \varphi_E = \pi\alpha/6^{\frac{1}{2}}, \qquad (4.2)$$

$$\chi_F^{sc} = \frac{\sin^3\varphi_F\cos^2\varphi_F}{\varphi_F^3}, \qquad \varphi_F = \pi\alpha/3,$$

$$\chi_G^{sc} = \frac{\sin 4\varphi_G \sin^2\varphi_G}{4\varphi_G^3}, \qquad \varphi_G = \pi\alpha/18^{\frac{1}{2}},$$

$$\chi_H^{sc} = \frac{\sin 4\varphi_H \sin 3\varphi_H \sin\varphi_H}{12\varphi_H^3}, \qquad \varphi_H = \pi\alpha/26^{\frac{1}{2}},$$

$$\chi_I^{sc} = \frac{\sin 4\varphi_I(\sin 3\varphi_I)^2}{36\varphi_I^3}, \qquad \varphi_I = \pi\alpha/34^{\frac{1}{2}}.$$

Using these formulas we calculated $g_0(r)$ from 3, 6, and 9-term expressions and $g_2(r)$ and $g_3(r)$ from 3 and 6-term expressions. The $g_i(r)$ are calculated in the region $0 \leq r \leq 2a$ or $0 \leq \alpha \leq 2$.

The exact solutions for $g_i(r)$ are found in the form of power series in r/a. We write down such a solution for $g_0(r)$:

$$g_0(r/a) = (4\pi)^{\frac{1}{2}} \sum_{i=0}^{\infty} b_i \left(\frac{\pi r}{a}\right)^{2i}, \qquad (4.3)$$

where

$$b_0 = a_0, \qquad b_1 = \frac{1}{3}\left(-\frac{a_0}{3!} + 2a_1\right),$$

$$b_{i>1} = \frac{(-1)^i a_0}{(2j+1)!\,(2j+1)} + \frac{2\cdot 4\cdots 2j}{3\cdot 5\cdots(2j+1)} a_i + \frac{1}{3\cdot 5\cdots(2j+1)} \times \sum_{s=1}^{i-1} \frac{(-1)^{i-s}(2\cdot 4\cdots 2s)}{(2j-2s+1)[2\cdot 4\cdots(2j-2s)]} a_s,$$

and

$$a_0 = 1, \qquad a_1 = -\frac{1}{3!},$$

$$a_{i>1} = \frac{(-1)^i}{2\cdot 4\cdots 2j}\left[\frac{2}{(2j+1)(2\cdot 4\cdots 2j)} + \sum_{s=1}^{i-1} \frac{1}{(2s+1)(2j-2s+1)(2\cdot 4\cdots 2s)[2\cdot 4\cdots(2j-2s)]}\right].$$

$g_2(r)$ and $g_3(r)$ were also found in the form (4.3) but the coefficients were still more complicated.

Calculating the numerical values of several first coefficients b_i we have

$$g_0(\alpha) = (4\pi)^{\frac{1}{2}}[1 - 1.644927\alpha^2 + 1.028210\alpha^4 - 0.353043\alpha^6 + 0.078344\alpha^8 - 0.012318\alpha^{10} + \cdots], \quad (4.4)$$

$$g_2(\alpha) = (4\pi)^{\frac{1}{2}}[-0.094473\alpha^4 + 0.074673\alpha^6 - 0.026125\alpha^8 + 0.005511\alpha^{10} - \cdots]. \quad (4.4')$$

We do not write the expression for $g_3(\alpha)$ but point out that its power series starts with an α^6 term. From the series expansion we notice that

$$g_0(\alpha = 0)/(4\pi)^{\frac{1}{2}} = 1, \qquad g_{j\neq 0}(\alpha = 0) = 0,$$

and the g_j with $j \neq 0$ behaves in the vicinity of $\alpha = 0$ as α^{2j}.

In order to compare the results (4.4) and (4.4') with results obtained by Houston's method, we expand the 3-term formulas, with χ_A, χ_B, and χ_C given by (4.2) into power series in α; we obtain then

$$g_0^{(3)}(\alpha) = (4\pi)^{\frac{1}{2}}[1 - 1.644927\alpha^2 + 1.028210\alpha^4 - 0.353043\alpha^6 + 0.079141\alpha^8 - 0.012733\alpha^{10} + \cdots], \quad (4.5)$$

$$g_2^{(3)}(\alpha) = (4\pi)^{\frac{1}{2}}[-0.094472\alpha^4 + 0.074670\alpha^6 - 0.025895\alpha^8 + 0.005428\alpha^{10} - \cdots]. \quad (4.5')$$

It was also established that $g_3^{(3)}(\alpha)$ starts with an α^6 term. Comparing the coefficients in (4.4), (4.4') and (4.5), (4.5'), we see that for lowest exponents of α the agreement is excellent and the g_i's obtained by Houston's method approximate very well the exact solutions in the region $\alpha < 1$.

In Table I, the function $g_0(\alpha)$ is tabulated from 3, 6, and 9-term formulas, and furthermore, for small α the results are given for the analytical expression (4.4).

We notice that for small α we have very good agreement between all these functions but with increasing α the functions start to diverge. The 3-term function $g_0^{(3)}$ distinctly differs from $g_0^{(6)}$ and $g_0^{(9)}$ when α increases, but $g_0^{(6)}$ and $g_0^{(9)}$ differ only very slightly for $\alpha \leq 2$.

It was mentioned in Sec. I that the applicability of Houston's method is limited. The use of the finite number of terms in the expansion does not allow to define the functions $g_i^{(n)}$ properly for all r. They have proper behavior for relatively small r but for large r they are completely wrong. The question occurs how to establish the region of r where $g_i^{(n)}(r)$ obtained from n-term formula approximate well the exact $g_i(r)$. We do not answer this question here thoroughly; we only make some conclusions comparing different results for g_0.

TABLE I. The function $g_0(\alpha)/(4\pi)^{\frac{1}{2}}$ for the simple cubic structure.

	$g_0(\alpha)/(4\pi)^{\frac{1}{2}}$ obtained by Houston's method			$g_0(\alpha)/(4\pi)^{\frac{1}{2}}$
α	3-term expression	6-term expression	9-term expression	analytical expression
0.00	1.0000	1.0000	1.0000	1.0000
0.05	0.9959	0.9959	0.9959	0.9959
0.10	0.9837	0.9837	0.9837	0.9837
0.15	0.9635	0.9635	0.9635	0.9635
0.20	0.9358	0.9358	0.9358	0.9358
0.25	0.9011	0.9011	0.9011	0.9011
0.30	0.8600	0.8600	0.8600	0.8600
0.35	0.8133	0.8133	0.8133	0.8133
0.40	0.7617	0.7617	0.7617	0.7617
0.45	0.7063	0.7062	0.7062	0.7063
0.50	0.6478	0.6478	0.6478	0.6478
0.55	0.5873	0.5874	0.5874	0.5873
0.60	0.5259	0.5259	0.5259	0.5259
0.65	0.4643	0.4643	0.4643	0.4643
0.70	0.4035	0.4035	0.4035	0.4035
0.75	0.3445	0.3444	0.3444	0.3444
0.80	0.2879	0.2878	0.2878	0.2877
0.85	0.2344	0.2342	0.2342	0.2340
0.90	0.1846	0.1844	0.1844	0.1840
0.95	0.1391	0.1387	0.1387	...
1.00	0.0981	0.0976	0.0976	...
1.05	0.0618	0.0612	0.0612	...
1.10	0.0306	0.0297	0.0297	...
1.15	0.0043	0.0030	0.0031	...
1.20	−0.0172	−0.0188	−0.0188	...
1.25	−0.0339	−0.0360	−0.0360	...
1.30	−0.0462	−0.0489	−0.0490	...
1.35	−0.0544	−0.0577	−0.0577	...
1.40	−0.0589	−0.0630	−0.0630	...
1.45	−0.0600	−0.0650	−0.0650	...
1.50	−0.0583	−0.0643	−0.0643	...
1.55	−0.0542	−0.0613	−0.0613	...
1.60	−0.0481	−0.0566	−0.0566	...
1.65	−0.0406	−0.0505	−0.0504	...
1.70	−0.0320	−0.0434	−0.0434	...
1.75	−0.0228	−0.0358	−0.0359	...
1.80	−0.0134	−0.0281	−0.0279	...
1.85	−0.0040	−0.0204	−0.0203	...
1.90	0.0050	−0.0132	−0.0130	...
1.95	0.0134	−0.0065	−0.0063	...
2.00	0.0210	−0.0006	−0.0004	...

Studying the expansion of $\chi(r)$ we took into consideration the region $0 \leq r \leq 2a$. The upper limit was chosen as $2a$ because in calculation of the integrals of the type (3.1) for typical solids the function $\Phi(r)$ is given as a linear combination of atomic orbitals centered at the origin or at the first neighbors positions and radial parts of most of these functions vanish for $r < 2a$.

Comparing $g_0^{(n)}$ with $n = 3, 6, 9$ with g_0 given analytically, we see completely proper behavior of $g_0^{(n)}$ for $\alpha < 1$. In the region $1 \leq \alpha \leq 2$, we notice that $g_0^{(3)}$ differs a little from $g_0^{(6)}$ and $g_0^{(9)}$ so it differs

certainly from exact g_0. Comparing $g_0^{(6)}$ with $g_0^{(9)}$ we may conclude that the region where the both functions have the same values can be considered as the region of a good approximation to the exact solution.

Furthermore, the comparison of the several lowest coefficients in (4.4) and (4.5) allows to suppose that for the other coefficients the agreement is also good and the small differences in coefficients can give large differences in functions only for large α.

Thus we may expect that the function $g_0^{(6)}$ is satisfactory for calculating integrals in (3.9) when the integration is over the distance not bigger than $2a$. We may of course still use $g_0^{(6)}$ beyond $0 \leq \alpha \leq 2$, but the error of the calculation increases with increasing distance of integration. We want to point out that in calculation of Debye's temperature[5] using Houston's method, the difference between θ_0^3 and θ_0^6, θ_0^9, θ_0^{15} was remarkable but the values θ_0^6, θ_0^9, and θ_0^{15} differed between themselves only very slightly.

In Table II the functions $g_2(\alpha)$ and $g_3(\alpha)$ are tabulated obtained from 3 and 6-term formulas, and also $g_2(\alpha)$ from analytical expression (4.4′) for small α. Also, here we notice very good agreement for small α and increasing discrepancies between $g_i^{(3)}$ and $g_i^{(6)}$ when α increases.

TABLE II. The functions $g_2(\alpha)/(4\pi)^{\frac{1}{2}}$ and $g_3(\alpha)/(4\pi)^{\frac{1}{2}}$ for the simple-cubic structure.

	$g_2(\alpha)/(4\pi)^{\frac{1}{2}}$ obtained by Houston's method		$g_2(\alpha)/(4\pi)^{\frac{1}{2}}$	$g_3(\alpha)/(4\pi)^{\frac{1}{2}}$ obtained by Houston's method	
α	3-term expression	6-term expression	analytical expression	3-term expression	6-term expression
0.00	0	0	0	0	0
0.05	0	0	0	0	0
0.10	0	0	0	0	0
0.15	0	0	0	0	0
0.20	−0.0001	−0.0001	−0.0001	0	0
0.25	−0.0004	−0.0004	−0.0004	0	0
0.30	−0.0007	−0.0007	−0.0007	0	0
0.35	−0.0013	−0.0013	−0.0013	0	0
0.40	−0.0021	−0.0021	−0.0021	0	0
0.45	−0.0033	−0.0033	−0.0033	0	0
0.50	−0.0048	−0.0048	−0.0048	−0.0001	−0.0001
0.55	−0.0068	−0.0068	−0.0068	−0.0002	−0.0002
0.60	−0.0092	−0.0092	−0.0092	−0.0002	−0.0002
0.65	−0.0120	−0.0120	−0.0120	−0.0004	−0.0004
0.70	−0.0152	−0.0152	−0.0152	−0.0006	−0.0006
0.75	−0.0189	−0.0189	−0.0189	−0.0008	−0.0008
0.80	−0.0229	−0.0230	−0.0229	−0.0011	−0.0011
0.85	−0.0272	−0.0273	−0.0272	−0.0015	−0.0015
0.90	−0.0318	−0.0318	−0.0316	−0.0020	−0.0020
0.95	−0.0364	−0.0365	...	−0.0026	−0.0026
1.00	−0.0410	−0.0411	...	−0.0033	−0.0032
1.05	−0.0454	−0.0457	...	−0.0040	−0.0040
1.10	−0.0496	−0.0500	...	−0.0049	−0.0049
1.15	−0.0534	−0.0539	...	−0.0059	−0.0058
1.20	−0.0567	−0.0573	...	−0.0069	−0.0068
1.25	−0.0594	−0.0601	...	−0.0079	−0.0078
1.30	−0.0612	−0.0622	...	−0.0090	−0.0089
1.35	−0.0623	−0.0635	...	−0.0101	−0.0099
1.40	−0.0624	−0.0640	...	−0.0112	−0.0109
1.45	−0.0616	−0.0636	...	−0.0122	−0.0119
1.50	−0.0599	−0.0623	...	−0.0131	−0.0127
1.55	−0.0571	−0.0600	...	−0.0139	−0.0134
1.60	−0.0535	−0.0570	...	−0.0146	−0.0139
1.65	−0.0489	−0.0531	...	−0.0150	−0.0142
1.70	−0.0436	−0.0485	...	−0.0153	−0.0143
1.75	−0.0377	−0.0433	...	−0.0153	−0.0141
1.80	−0.0311	−0.0376	...	−0.0151	−0.0138
1.85	−0.0242	−0.0316	...	−0.0146	−0.0131
1.90	−0.0170	−0.0254	...	−0.0140	−0.0122
1.95	−0.0097	−0.0191	...	−0.0130	−0.0110
2.00	−0.0025	−0.0129	...	−0.0119	−0.0096

V. THE FUNCTIONS $g_j(r)$ FOR A SUM OF PLANE WAVES IN A FACE-CENTERED-CUBIC LATTICE

The function $\chi(\mathbf{r})$ for face-centered-cubic lattice is obtained from (3.7) by integration over the Brillouin zone which is given as a truncated octahedron. We have the following exact expression for $\chi(\mathbf{r})$:

$$\chi^{fcc}(\mathbf{r}) = \frac{1}{4}\left(\frac{a}{\pi}\right)^3 \left\{ \frac{x \sin\frac{\pi x}{a}\left(\cos\frac{2\pi y}{a} + \cos\frac{2\pi z}{a}\right) + \cos\frac{\pi x}{a}\left(y \sin\frac{2\pi y}{a} + z \sin\frac{2\pi z}{a}\right)}{(x^2 - y^2)(x^2 - z^2)} \right.$$

$$+ \frac{y \sin\frac{\pi y}{a}\left(\cos\frac{2\pi x}{a} + \cos\frac{2\pi z}{a}\right) + \cos\frac{\pi y}{a}\left(x \sin\frac{2\pi x}{a} + z \sin\frac{2\pi z}{a}\right)}{(y^2 - z^2)(y^2 - x^2)}$$

$$\left. + \frac{z \sin\frac{\pi z}{a}\left(\cos\frac{2\pi x}{a} + \cos\frac{2\pi y}{a}\right) + \cos\frac{\pi z}{a}\left(x \sin\frac{2\pi x}{a} + y \sin\frac{2\pi y}{a}\right)}{(z^2 - x^2)(z^2 - y^2)} \right\}. \tag{5.1}$$

The appropriate expressions for $\chi(\mathbf{r})$ for nine directions are written below. The magnitudes $\varphi_A, \varphi_B, \cdots \varphi_I$ are the same as given for the simple-cubic structure.

$$\chi_A^{fcc} = \frac{1}{4\varphi_A^3}\sin\chi_A \times [2 - 2\cos\varphi_A + 2\varphi_A\sin\varphi_A + \varphi_A^2\cos\varphi_A],$$

$$\chi_B^{fcc} = \frac{1}{2\varphi_B^3}\sin\varphi_B\cos\varphi_B[2 - 2\cos\varphi_B + \varphi_B\sin\varphi_B],$$

$$\chi_C^{fcc} = \frac{1}{16\varphi_C^3}\sin\varphi_C[3\varphi_C^2(3 - 4\sin^2\varphi_C) + 6\varphi_C\cos\varphi_C\sin\varphi_C + \sin^2\varphi_C],$$

$$\chi_D^{fcc} = \frac{1}{12\varphi_D^3}[-10 + 32\sin^2\varphi_D - 24\sin^4\varphi_D + 10\cos\varphi_D - 15\sin^2\varphi_D\cos\varphi_D],$$

$$\chi_E^{fcc} = \frac{1}{6\varphi_E^3}[3\varphi_E\cos\varphi_E\sin^2\varphi_E(1 - 2\sin^2\varphi_E) + 2\sin\varphi_E(-1 + 4\sin^2\varphi_E) - 3\sin^4\varphi_E + \cos\varphi_E - 2\cos\varphi_E\sin^2\varphi_E], \quad (5.2)$$

$$\chi_F^{fcc} = \frac{1}{36\varphi_F^3}[3\varphi_F\sin^2\varphi_F \times (6 - 21\sin^2\varphi_F + 16\sin^4\varphi_F) + \sin\varphi_F \times (18 - 64\sin^2\varphi_F + 48\sin^4\varphi_F - 18\cos\varphi_F + 73\cos\varphi_F\sin^2\varphi_F - 64\cos\varphi_F\sin^4\varphi_F)],$$

$$\chi_G^{fcc} = \frac{1}{30\varphi_G^3}[\varphi_G\cos\varphi_G\sin^2\varphi_G \times (15 - 120\sin^2\varphi_G + 256\sin^4\varphi_G - 160\sin^6\varphi_G) + \frac{2}{15}\sin\varphi_G(-9 + 157\sin^2\varphi_G - 612\sin^4\varphi_G + 864\sin^6\varphi_G - 400\sin^8\varphi_G) + \frac{2}{15}\sin\varphi_G \times \cos\varphi_G(9 - 40\sin^2\varphi_G + 40\sin^4\varphi_G)],$$

$$\chi_H^{fcc} = \frac{1}{210\varphi_H^3}[2\sin\varphi_H\cos\varphi_H \times (13 - 160\sin^2\varphi_H + 608\sin^4\varphi_H - 896\sin^6\varphi_H + 448\sin^8\varphi_H) + \sin\varphi_H(-26 + 543\sin^2\varphi_H - 3429\sin^4\varphi_H + 8512\sin^6\varphi_H - 8960\sin^8\varphi_H + 3360\sin^{10}\varphi_H)],$$

$$\chi_I^{fcc} = \frac{1}{42\varphi_I^3}[\varphi_I\cos\varphi_I\sin^2\varphi_I \times (21 - 308\sin^2\varphi_I + 1232\sin^4\varphi_I - 1824\sin^6\varphi_I + 896\sin^8\varphi_I) + \frac{2}{7}\sin\varphi_I(-51 + 939\sin^2\varphi_I - 4744\sin^4\varphi_I + 9920\sin^6\varphi_I - 9200\sin^8\varphi_I + 3136\sin^{10}\varphi_I) + \frac{2}{7}\sin\varphi_I\cos\varphi_I \times (51 - 840\sin^2\varphi_I + 3528\sin^4\varphi_I - 5376\sin^6\varphi_I + 2688\sin^8\varphi_I)].$$

As in the case of the simple-cubic structure, the function $g_0(\alpha)$ is calculated using 3, 6, 9-term formulas and $g_2(\alpha)$ and $g_3(\alpha)$ using 3 and 6-term formulas. The functions are tabulated in Tables III and IV. As in the previous case we have very good agreement between $g_0^{(3)}$, $g_0^{(6)}$ and $g_0^{(9)}$ for small α. When α increases the functions start to diverge. The numerical values of $g_0^{(3)}$ and $g_0^{(6)}$ differ much more than the values of $g_0^{(6)}$ and $g_0^{(9)}$. For example, at $\alpha = 1.5$, we

TABLE III. The function $g_0(\alpha)/(4\pi)^{\frac{1}{2}}$ for the face-centered cubic structure.

α	$g_0(\alpha)/(4\pi)^{\frac{1}{2}}$ 3-term expression	6-term expression	9-term expression
0.00	1.0000	1.0000	1.0000
0.05	0.9903	0.9903	0.9903
0.10	0.9615	0.9615	0.9615
0.15	0.9149	0.9149	0.9149
0.20	0.8523	0.8523	0.8523
0.25	0.7765	0.7765	0.7765
0.30	0.6903	0.6903	0.6903
0.35	0.5973	0.5973	0.5973
0.40	0.5009	0.5009	0.5009
0.45	0.4046	0.4046	0.4046
0.50	0.3117	0.3117	0.3117
0.55	0.2251	0.2250	0.2250
0.60	0.1471	0.1471	0.1471
0.65	0.0797	0.0798	0.0798
0.70	0.0240	0.0241	0.0241
0.75	−0.0194	−0.0192	−0.0192
0.80	−0.0506	−0.0503	−0.0503
0.85	−0.0703	−0.0700	−0.0700
0.90	−0.0797	−0.0792	−0.0792
0.95	−0.0804	−0.0797	−0.0797
1.00	−0.0740	−0.0731	−0.0731
1.05	−0.0626	−0.0614	−0.0614
1.10	−0.0481	−0.0465	−0.0465
1.15	−0.0322	−0.0301	−0.0301
1.20	−0.0165	−0.0140	−0.0140
1.25	−0.0024	0.0006	0.0006
1.30	0.0093	0.0128	0.0128
1.35	0.0179	0.0218	0.0218
1.40	0.0231	0.0275	0.0275
1.45	0.0249	0.0298	0.0298
1.50	0.0239	0.0291	0.0291
1.55	0.0203	0.0259	0.0259
1.60	0.0150	0.0207	0.0208
1.65	0.0087	0.0144	0.0145
1.70	0.0021	0.0077	0.0077
1.75	−0.0042	0.0011	0.0012
1.80	−0.0095	−0.0047	−0.0045
1.85	−0.0135	−0.0092	−0.0091
1.90	−0.0159	−0.0124	−0.0122
1.95	−0.0167	−0.0141	−0.0138
2.00	−0.0158	−0.0142	−0.0139

TABLE IV. The functions $g_2(\alpha)/(4\pi)^{\frac{1}{2}}$ and $g_3(\alpha)/(4\pi)^{\frac{1}{2}}$ for the face-centered cubic structure.

	$g_2(\alpha)/(4\pi)^{\frac{1}{2}}$		$g_3(\alpha)/(4\pi)^{\frac{1}{2}}$	
α	3-term expression	6-term expression	3-term expression	6-term expression
0.00	0.0000	0.0000	0	0
0.05	0.0000	0.0000	0	0
0.10	0.0000	0.0000	0	0
0.15	0.0001	0.0001	0	0
0.20	0.0003	0.0003	0	0
0.25	0.0007	0.0007	0	0
0.30	0.0015	0.0015	0	0
0.35	0.0026	0.0026	0.0001	0.0001
0.40	0.0041	0.0041	0.0001	0.0001
0.45	0.0061	0.0061	0.0002	0.0002
0.50	0.0085	0.0085	0.0004	0.0004
0.55	0.0113	0.0113	0.0007	0.0007
0.60	0.0143	0.0143	0.0011	0.0011
0.65	0.0174	0.0174	0.0017	0.0017
0.70	0.0205	0.0205	0.0024	0.0024
0.75	0.0233	0.0233	0.0032	0.0032
0.80	0.0256	0.0256	0.0043	0.0043
0.85	0.0273	0.0273	0.0055	0.0055
0.90	0.0281	0.0281	0.0068	0.0068
0.95	0.0280	0.0280	0.0082	0.0083
1.00	0.0270	0.0269	0.0096	0.0097
1.05	0.0249	0.0248	0.0110	0.0111
1.10	0.0220	0.0218	0.0123	0.0124
1.15	0.0183	0.0180	0.0133	0.0136
1.20	0.0141	0.0135	0.0141	0.0144
1.25	0.0096	0.0088	0.0145	0.0149
1.30	0.0050	0.0039	0.0145	0.0151
1.35	0.0007	−0.0008	0.0140	0.0147
1.40	−0.0030	−0.0051	0.0130	0.0140
1.45	−0.0060	−0.0087	0.0116	0.0128
1.50	−0.0081	−0.0114	0.0097	0.0112
1.55	−0.0091	−0.0131	0.0075	0.0092
1.60	−0.0090	−0.0138	0.0050	0.0070
1.65	−0.0079	−0.0134	0.0024	0.0046
1.70	−0.0059	−0.0121	−0.0003	0.0022
1.75	−0.0032	−0.0101	−0.0028	−0.0002
1.80	0.0001	−0.0074	−0.0051	−0.0023
1.85	0.0035	−0.0043	−0.0071	−0.0042
1.90	0.0069	−0.0011	−0.0086	−0.0056
1.95	0.0099	0.0019	−0.0095	−0.0066
2.00	0.0124	0.0046	−0.0098	−0.0072

have $(g_0^{(6)} - g_0^{(3)}) = 0.0052$, but $(g_0^{(6)} - g_0^{(9)})$ is still equal to zero. For the functions g_2 and g_3 we have also very good agreement between $g_i^{(3)}$ and $g_i^{(6)}$ for small α.

We want to point out that the first zero for g_0 is located in the vicinity of $r \approx a$ for the simple cubic lattice and in the vicinity of $r \approx 0.7a$, which is about $a/(2)^{\frac{1}{2}}$ for the face-centered cubic lattice. The reason for this is that the first zero of $\chi(\mathbf{r})$ occurs at the positions of the first neighbors.

VI. NOTE ON APPLICATIONS TO SOLID STATE THEORY

One of the most fruitful approaches to the computation of energy band structures in solids has been the OPW (Orthogonalized Plane Wave) method.[11]

[11] C. Herring, Phys. Rev. 57, 1169 (1940); see the review by T. O. Woodruff, in Solid State Phys. 4, 367 (1957).

We write an arbitrary plane wave as

$$\psi_{ik}(\mathbf{r}) = N^{-\frac{1}{2}} e^{i(\mathbf{k}_i + \mathbf{k})\cdot\mathbf{r}}, \tag{6.1}$$

where $\mathbf{k}_i$ is 2π times a reciprocal lattice vector and $\mathbf{k}$ runs over the first Brillouin Zone of thela ttice concerned. The OPW method consists of assuming that a certain finite set of "core bands" are adequately described by tight-binding wavefunctions,

$$\phi_{ck}(\mathbf{r}) = N^{-\frac{1}{2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_c(\mathbf{r} - \mathbf{R}), \tag{6.2}$$

where $\mathbf{R}$ locates a lattice site, $\sum_{\mathbf{R}}$ runs over all of them, and $\phi_c(\mathbf{r} - \mathbf{R})$ is an atomic function centered on lattice site $\mathbf{R}$. We assume no overlapping of wavefunctions on different sites, for simplicity. The OPW method consists of orthogonalizing the plane wave (6.1) to all the core states (6.2), by the Schmidt process. Then the orthogonalized functions

$$\Psi_{ik} = \psi_{ik} - \sum_c (\phi_{ck}, \psi_{ik})\phi_{ck} \tag{6.3}$$

are used in a linear variational determination of energies for each value of $\mathbf{k}$. Associated with each $\mathbf{k}$ and each energy there are one or more linear combinations of the states Ψ_{ik}.

In some applications it is useful to construct a localized set of states known as Wannier functions.[12] If a Bloch state $u_{nk}(\mathbf{r})$ is known (such as the linear linear combination mentioned just above), the corresponding Wannier function is defined by

$$a_n(\mathbf{r} - \mathbf{R}) = N^{-\frac{1}{2}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} u_{nk}(\mathbf{r}), \tag{6.4}$$

where the sum runs over the whole first Brillouin Zone. The Wannier function associated with a single OPW (6.3) has a very simple analytical form, and may be useful[13] as a zero-order localized function. It is

$$\begin{aligned} a_i(\mathbf{r} - \mathbf{R}) &\equiv N^{-\frac{1}{2}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} \Psi_{ik} \\ &= e^{i\mathbf{k}_i\cdot\mathbf{r}} \chi(\mathbf{r} - \mathbf{R}) \\ &\quad - \sum_c \sum_{\mathbf{R}'} B_c(\mathbf{k}_i; \mathbf{R} - \mathbf{R}') \phi_c(\mathbf{r} - \mathbf{R}'), \end{aligned} \tag{6.5}$$

where B_c simply expresses the nonorthogonality of χ with atomic functions on its own site and on neighboring sites:

$$B_c(\mathbf{k}_i; \mathbf{R} - \mathbf{R}') = \int \phi_c(\mathbf{r} - \mathbf{R}' + \mathbf{R}) e^{i\mathbf{k}_i\cdot(\mathbf{r}+\mathbf{R})} \chi(\mathbf{r})\, d\mathbf{r}. \tag{6.6}$$

[12] G. H. Wannier, Phys. Rev. 52, 191 (1937).
[13] R. S. Knox (private communication).

The function $\chi(\mathbf{r})$ studied in the bulk of this paper is now seen to be part of "the Wannier function of a plane wave," and as such it enters into analytical expressions very similar to those of a tight binding theory. This explains the usefulness of having the expansion of $\chi(\mathbf{r})$ in spherical harmonics; in the special case $\mathbf{k}_i = 0$, only s-like core states contribute to the sum over c and we have, for example,

$$B_c(0;0) = \int \phi_{ns}(\mathbf{r})^* \chi(\mathbf{r})\, d\mathbf{r}, \qquad (6.7)$$
$$= (4\pi)^{\frac{1}{2}} \int_0^\infty P_{ns}(r) g_0(r) r\, dr,$$

where $\phi_{ns} = [P_{ns}(r)/r] Y_{00}(\theta, \varphi)$. The quantity (6.7) gives the overlap of $\chi(\mathbf{r})$ with an s function centered at the same lattice site. For $\mathbf{R} - \mathbf{R}' \neq 0$ in (6.6), it is necessary to perform two-center integrals, where again it is usually convenient to have spherical harmonic expansions (see, e.g., the method of Löwdin[14]). We have verified that the tables given herein are sufficient to compute most quantities of the form (3.1) arising from these considerations, with particular reference to Fe, Cu, Ar, and NaCl. Some of the results will be presented in a forthcoming publication on exciton oscillator strengths in the latter two crystals.

[14] P. O. Löwdin, Advan. Phys. 5, 1 (1956).

VII. SUMMARY

A general discussion of the "Wannier Function of a plane wave" has been presented, along with approximate tabular values of its Kubic Harmonic components $g_i(r)$ in two specific lattices, simple-cubic, and face-centered cubic. It is emphasized that the $g_i(r)$ are in fact functions of $\alpha = r/a$ so that they depend only on the structure and can be used for *any* material with simple-cubic or fcc structure.

ACKNOWLEDGMENT

The author wishes to thank Prof. R. S. Knox for for many helpful suggestions and discussions and Dr. A Kiełbasiński from the University of Warsaw for programming and performing some numerical calculations with the electronic computer GIER. It is a pleasure to thank Prof. D. L. Dexter for the generous hospitality extended to me at the University of Rochester.

On a Quantum Field Theory with Degenerate Vacuum and a Special Type of Symmetry

JAN LOPUSZANSKI* AND HELMUT REEH†‡
The Institute for Advanced Study, Princeton, New Jersey
(Received 9 June 1965)

We investigate a frame of a model of quantum field theory with a degenerate vacuum. At least one vacuum is cyclic. The field is covariant under a "gauge" transformation. We show the existence of non-gauge-invariant vacuum states. We define "observables" as gauge-invariant operators; the algebra of these does not coincide with the algebra of the field operator; the reduction of the former algebra reflects a superselection rule.

1. INTRODUCTION

THERE is a constant interest in theories with degenerate ground state in nonrelativistic (many-body problem) as well as in relativistic (quantum field theory) approach.

As far as the relativistic field theory is concerned, the degeneration of the Lorentz-invariant vacuum state has been partly elucidated by some theorems.[1-3] In particular it has been shown that there cannot be a degeneracy of the vacuum state if every vacuum state is a cyclic vector for the algebra of field operators.

In this paper we intend to investigate the case that may occur if not every, but at least one, vacuum state is cyclic. We examine a frame for models of relativistic quantum field theory which display certain symmetry properties. The basic idea is already contained in Haag's paper[4,5] concerning the Bogolubov–BCS model.[6-8] We consider field theories fulfilling the usual assumptions of the Wightman framework, i.e., the field operator is an operator valued distribution in a separable Hilbert space $\mathfrak{H}$; the smeared out field maps a dense set $\mathfrak{D}$ of $\mathfrak{H}$ into itself, transforms according to a unitary representation of the inhomogeneous Lorentz group, has a vacuum state invariant under the Lorentz transformation as a cyclic vector and is local; the spectrum of the translation operator is assumed to lie in the forward light cone. We are going to show that there may exist vacuum states noninvariant under the symmetry which is inherent to the field. This is mainly due to the reduction of the original reducible algebra of field operators possessing the mentioned symmetry property into irreducible algebras which are no longer invariant under this symmetry.

Most of the calculations are carried out in detail for a special example of gauge (which, however, is not a gauge transformation in the usual sense as, e.g., for free charged fields). Most of the results hold also for other examples of symmetry as considered, e.g., in connection with the Goldstone theorem,[9] briefly mentioned in the concluding Sec. 5.

Starting from the gauge properties of the theory we are able to give a reasonable definition of the observables. It follows that not all observables belong to the algebra of the field and vice versa. The von Neumann algebra of the field operator and that of the superselecting operator (which is contained in the algebra of observables) coincide with the set of all bounded operators. The reduction of the Hilbert space with respect to the center of the observables into the coherent sectors is in a certain way "complementary" to that which reduces the algebra of the field operator. This follows rather trivially from the assumptions made for the gauge transformation.

2. CONSTRUCTION OF THE MODEL

In this section we are going to outline the construction of the model with which we are mainly

* On leave of absence from the Institute of Theoretical Physics of the University of Wroclaw, Wroclaw, Poland.
† On leave of absence from the Max Planck Institut für Physik und Astrophysik, Munich, Germany.
‡ Work supported by the National Science Foundation.
[1] H. J. Borchers, Nuovo Cimento **24**, 214, 214 (1962).
[2] H. Reeh and S. Schlieder, Nuovo Cimento **26**, 32 (1962).
[3] H. Araki, Progr. Theoret. Phys. (Kyoto) **32**, 844 (1964).
[4] R. Haag, Nuovo Cimento **25**, 287 (1962).
[5] H. Ezawa, J. Math. Phys. **5**, 1078 (1964).
[6] J. Bardeen, L. N. Cooper, and J. R. Schriefer, Phys. Rev. **108**, 1175 (1957).
[7] N. N. Bogolubov, V. V. Tolmačov, and D. S. Širkov, A. N. SSSR, Moskva, 1958.
[8] While this paper was under preparation we learned from Dr. Guenin and Dr. Emch that they extended Haag's procedure on similar lines for the BCS model as we intend to do here for field theory. See G. Emch and M. Guenin, preprint, February 1965. After this note was finished we heard that Dr. Uhlmann developed similar ideas.
[9] See in particular J. Goldstone, A. Salam, S. Weinberg, Phys. Rev. **127**, 965 (1962); compare also J. Goldstone, Nuovo Cimento **19**, 154 (1961) and W. Gilbert, Phys. Rev. Letters **12**, 713 (1964).

concerned. To this aim let us consider a charged scalar field $A_0(x)$ and the adjoint $A_0^\dagger(x)$ in a separable Hilbert space $\mathfrak{H}_0$. Let us assume this field to satisfy the axioms listed in the introduction and to be irreducible in $\mathfrak{H}_0$, so that there is only one vacuum state Ω_0. We take then a family of duplicates of such fields, A_α, $A_\alpha^\dagger$, $\mathfrak{H}_\alpha$, Ω_α for a continuous α, $0 \leq \alpha < 2\pi$. The fields in different Hilbert spaces are linked together by a one-parameter family of isometric transformations $G_\alpha(\beta)$ [β real, $G_\alpha(\beta+2\pi) = G_\alpha(\beta)$] in the following way[10,11]:

$$\begin{aligned} G_\alpha(\beta)\mathfrak{H}_\alpha &= \mathfrak{H}_{\alpha+\beta} \\ G_\alpha(\beta)\Omega_\alpha &= \Omega_{\alpha+\beta} \end{aligned} \tag{2.1}$$

$$\begin{aligned} G_\alpha(\beta)A_\alpha(x)G_{\alpha+\beta}^\dagger(\beta) &= A_{\alpha+\beta}(x)e^{i\beta} \\ G_\alpha(\beta)A_\alpha^\dagger(x)G_{\alpha+\beta}^\dagger(\beta) &= A_{\alpha+\beta}^\dagger(x)e^{-i\beta}, \end{aligned} \tag{2.2}$$

where

$$G_{\alpha+\beta}^\dagger(\beta) \equiv (G_\alpha(\beta))^\dagger.$$

If $U_\alpha(\Lambda, a)$ denotes the representation of the Lorentz group in $\mathfrak{H}_\alpha$ [i.e., $U_\alpha(\Lambda, a)A_\alpha(x)U_\alpha^{-1}(\Lambda, a) = A_\alpha(\Lambda x + a)$] then it follows using the irreducibility of A_α in $\mathfrak{H}_\alpha$ that

$$G_\alpha(\beta)U_\alpha(\Lambda, a)G_{\alpha+\beta}^\dagger(\beta) = U_{\alpha+\beta}(\Lambda, a).$$

We get also that

$$G_\alpha(\beta)I_\alpha G_{\alpha+\beta}^\dagger(\beta) = I_{\alpha+\beta}, \tag{2.3}$$

where I_α is the identity operator in $\mathfrak{H}_\alpha$.

Let us now form the direct integral of the spaces $\mathfrak{H}_\alpha$ (for the notation see Ref. 12) as well as of operators and vectors in these spaces.

$$\begin{aligned} \mathfrak{H} &= \int^\oplus \mathfrak{H}_\alpha (d\alpha)^\natural, \\ \Omega &= \int^\oplus \Omega_\alpha, \\ A(x) &= \int^\oplus A_\alpha(x), \\ U(\Lambda, a) &= \int^\oplus U_\alpha(\Lambda, a). \end{aligned} \tag{2.4}$$

The symbol $d\alpha$ denotes here the Lebesgue measure on $0 \cdots 2\pi$; Ω is a Lorentz-invariant state in the separable[13] Hilbert space $\mathfrak{H}$, the components of which in the different spaces are just Ω_α. The meaning of the direct integral is as follows: If $\Phi, \Psi \in \mathfrak{H}$ we have $(\Phi, A\Psi) = \int_0^{2\pi} (\Phi_\alpha, A_\alpha\Psi_\alpha)\, d\alpha$. The $G_\alpha(\beta)$ give rise to an operator $G(\beta)$ in $\mathfrak{H}$

$$(\Phi, G(\beta)\Psi) = \int_0^{2\pi} (\Phi_{\alpha+\beta}, G_\alpha(\beta)\Psi_\alpha)\, d\alpha.$$

The operator $G(\beta)$ which we call hereafter a "gauge transformation" is a unitary operator in $\mathfrak{H}$, which commutes with the Lorentz group and leaves Ω invariant.

We have also

$$\begin{aligned} G(\beta)A(x)G(\beta)^{-1} &= A(x)e^{i\beta}, \\ G(\beta)A^\dagger(x)G(\beta)^{-1} &= A^\dagger(x)e^{-i\beta}. \end{aligned} \tag{2.5}$$

Let us investigate under what additional assumptions the algebra $\mathfrak{A}$ of the unbounded field operator $A(x)$ given by (2.4) is cyclic with respect to Ω in $\mathfrak{H}$.

In case Ω is not cyclic, there exists a vector $\Phi \in \mathfrak{H}$, $\Phi \neq 0$ such that $(\Phi, A\Omega) = 0$ for every $A \in \mathfrak{A} = \int^\oplus \mathfrak{A}_\alpha$. $\mathfrak{A}_\alpha$ is the algebra of the unbounded field operators in $\mathfrak{H}_\alpha$, well defined on a dense domain of $\mathfrak{H}_\alpha$. We have $(\Phi, A\Omega) = \int d\alpha(\Phi_\alpha, A_\alpha\Omega_\alpha)$. This should also vanish for the particular choice of A

$$A = B^\circ U(I, a)A^\circ = \int^\oplus B_\alpha^\circ U_\alpha(I, a)A_\alpha^\circ,$$

where B_α°, A_α° are products of the smeared-out field operator and its adjoint with test functions of compact support in Minkowski space. By virtue of (2.1) and (2.2), we get

$$\begin{aligned} 0 &= (\Phi, B^\circ U(I, a)A^\circ\Omega) \\ &= \int d\alpha(\Phi_\alpha, B_\alpha^\circ U_\alpha(I, a)A_\alpha^\circ\Omega_\alpha) \\ &= \int d\alpha\, e^{-im\alpha}(G_\alpha^\dagger(\alpha)\Phi_\alpha, B_0^\circ U_0(I, a)A_0^\circ\Omega_0), \end{aligned} \tag{2.6}$$

where m is the difference of the number of field operators and the number of its adjoints in B° and A°. Now

$$\psi_0 \equiv \int d\alpha\, e^{-im\alpha}G_\alpha^\dagger(\alpha)\Phi_\alpha$$

is a vector in $\mathfrak{H}_0$ [since (2.6) is a linear integral over vectors belonging to $\mathfrak{H}_0$ and $||\psi_0|| \leq \int d\alpha \int d\alpha'$

[10] Instead of (2.2) we could as well consider the more general looking transformation

$$G_\alpha(\beta)A_\alpha(x)G_{\alpha+\beta}{}^\dagger(\beta) = V_{\alpha+\beta}A_{\alpha+\beta}(x)V^{-1}{}_{\alpha+\beta}e^{i\beta}, \text{ etc.},$$
$$V_{\alpha+\beta}V_{\alpha+\beta}{}^\dagger = V_{\alpha+\beta}{}^\dagger V_{\alpha+\beta} = I_{\alpha+\beta},$$
$$V_{\alpha+\beta}\Omega_{\alpha+\beta} = \Omega\alpha_{+\beta},$$

but it turns out that by introducing $\tilde{G}_\alpha(\beta) = V_{\alpha+\beta}{}^\dagger G_\alpha(\beta)$ we get again (2.2) and (2.3). The transformation V_α corresponds to an irrelevant change of frame of reference in $\mathfrak{H}_\alpha$.

[11] We could start with the transformation

$$G_\alpha(\beta)A_\alpha(x)G_{\alpha+\beta}{}^\dagger(\beta) = A_{\alpha+\beta}(x)f(\beta)$$

where $f(\beta)$ is a function; it turns out, however, that this transformation reduces to (2.2) again (except for the trivial case $f = 1$).

[12] M. A. Neumark, *Normierte Algebren*, (Deutscher Verlag der Wissenschaften, Berlin, 1959), Sec. 41, p. 520.

[13] J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1957), Ch. II, Sec. 1.6.

$||\Phi_\alpha||\ ||\Phi_{\alpha'}|| < \infty$]. Thus by applying a well-known result,[14] we get

$$\lim_{a\to\infty} (\Phi, B^c U(I, a) A^c \Omega)$$

$$= \int d\alpha\, e^{-in\alpha} (G_\alpha^\dagger(\alpha)\Phi_\alpha, B_0^c\Omega_0)(\Omega_0, A_0^c\Omega_0)$$

$$= \int d\alpha\, e^{-in\alpha} (\Phi_\alpha, B_\alpha^c\Omega_\alpha)(\Omega_0, A_0^c\Omega_0), \qquad (2.7)$$

where a is spacelike and n is the difference of the number of field operators and the number of its adjoints in A_0^c. Notice that the cluster decomposition theorem proved for a dense set of states on the left-hand side of the matrix element can be easily extended to an arbitrary state out of $\mathfrak{H}$.

From (2.6) and (2.7) follows

$$\int d_\alpha\, e^{-in\alpha} (\Phi_\alpha, B_\alpha^c\Omega_\alpha)(\Omega_0, A_0^c\Omega_0) = 0 \qquad (2.8)$$

for all B_α^c, A_0^c. Should at least one Wightman function $(\Omega_0, A_0^c\Omega_0)$ for every $n = 0, \pm 1, \pm 2, \cdots$ be different from zero (this is apparently not the case for free fields!), the last equation yields

$$(\Phi_\alpha, B_\alpha^c\Omega_\alpha) = 0 \qquad (2.9)$$

for every B_α^c for (almost) all α. This, however, implies

$$\Phi_\alpha = 0, \qquad \Phi = 0, \qquad (2.10)$$

which secures the cyclicity of $\mathfrak{A}$ with respect to Ω.

The restrictions imposed above onto the Wightman functions can be weakened in the following way. Assume

$$(\Omega_0, A_0^c\Omega_0) \neq 0 \qquad (2.11)$$

for $n = 1$. Then in virtue of

$$0 \neq (\Omega_0, A_0^c\Omega_0)(\Omega_0, A_0^c\Omega_0)$$

$$= \lim_{a\to\infty} (\Omega_0, A_0^c U_0(I, a) A_0^c\Omega_0),$$

the Wightman function for $n = 2$ under the limit cannot vanish for all finite a. In a similar way we find that there are Wightman functions different from zero for every $n \geq 0$. Formula (2.11) implies also that the Wightman function for $n = -1$.

$$(\Omega_0, A_0^{c\dagger}\Omega_0) \neq 0;$$

this enables us to construct Wightman functions different from zero for every $n \leq 0$.

Should the information (2.11) be not available, we have to look for two nonvanishing Wightman functions for n_1 and n_2, different from zero and having no common integer factor (i.e., n_1 and n_2 have to be relatively prime); if such functions exist, then there are again Wightman functions different from zero for every n. This can be shown by a construction similar to the preceding one based on the fact that every integer n can be represented as follows:

$$n = \nu n_1 + \mu n_2$$

with integers ν and μ.

Thus we arrive at the conclusion that $\mathfrak{A}$ is cyclically represented in $\mathfrak{H}$ with Ω as cyclic vector provided that $(\Omega_0, A_0^c\Omega_0) \neq 0$ either for $n = 1$ or for n_1, n_2 relatively prime. The first case comprises the case when the one-point function differs from zero (as, e.g., in a theory with "tadpoles").

Assuming that the before-mentioned hypothesis holds, we can make use of Borchers' result[1] that the commutant $\mathfrak{A}'$ consisting of all bounded operators commuting with $\mathfrak{A}$ is Abelian and therefore contained in the von Neumann algebra $\mathfrak{A}''$, the double commutant of $\mathfrak{A}$[15] (a slightly different proof of Borchers' assertion is given in the Appendix I).

Thus we succeeded in constructing a reducible field A possessing the symmetry property (2.5) and being cyclic with respect to a vacuum state Ω symmetric under the gauge transformation. The subspace of vacuum states[16]

$$\mathfrak{H}(0) = \int^\oplus \mathfrak{H}_\alpha(0)$$

(where $\mathfrak{H}_\alpha(0)$ is the subspace of vacuum states in $\mathfrak{H}_\alpha$ consisting of Ω_α only), contains also vectors nonsymmetric with respect to the gauge transformation, e.g., we can choose a function $f(\alpha)$ $(0 \leq \alpha < 2\pi)$ such that

$$G(\beta) \int^\oplus f(\alpha)\Omega_\alpha = \int^\oplus f(\alpha) G_\alpha(\beta)\Omega_\alpha$$

$$= \int^\oplus f(\alpha - \beta)\Omega_\alpha \neq \int^\oplus f(\alpha)\Omega_\alpha.$$

Not all vacuum states are cyclic with respect to $\mathfrak{A}$.

The hypothesis concerning the Wightman function is a sufficient condition. There is an indication that it is also a necessary one. We shall return to this point in the next section to make it plausible.

[14] Cluster decomposition theorem, cf., e.g. H. Araki, K. Hepp, and D. Ruelle, Helv. Phys. Acta 35, 164 (1962).

[15] Cf. the remark of R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964). Sec. c.

[16] H.-J. Borchers, Ref. 1, has proved that every state invariant under translations is invariant under the whole Lorentz group, provided there exists a Lorentz-invariant state, cyclic with respect to the field.

3. REDUCTION OF THE ALGEBRA OF THE FIELD POSSESSING A GAUGE SYMMETRY

In this section we want to examine the reverse problem to that considered in Sec. 2, viz., the standard reduction of the field algebra.

We start with a reducible charged scalar field $A(x)$ in a separable Hilbert space $\mathfrak{H}$. $A(x)$ satisfies the axioms stated in the introduction. In particular, it is assumed to be cyclic with respect to a Lorentz invariant state Ω and subjected to a gauge transformation (2.5). The unitary representation of the gauge group, $G(\beta)$ (β real) is assumed to leave Ω invariant. Lorentz invariance then implies that $G(\beta)$ commutes with the representation of the Lorentz group.

Let $\mathfrak{A}$ again denote the algebra of unbounded operators generated by the field operator. We may associate with $\mathfrak{A}$ two algebras of bounded operators: a weakly closed von Neumann algebra $\mathfrak{R} \equiv \mathfrak{A}''$ and a uniformly closed symmetric algebra $\mathfrak{K}$ (cf. Appendix II) which is weakly dense in $\mathfrak{R} = \mathfrak{K}''$ and for which $\mathfrak{K}' = \mathfrak{A}' = \mathfrak{R}'$.

In virtue of a theorem of Borchers[17] we know that $\mathfrak{A}'$ is Abelian so it coincides with the center of $\mathfrak{R}$. Therefore we are able to make a decomposition of the Hilbert space $\mathfrak{H}$ into a direct integral of Hilbert spaces $\mathfrak{H}_\alpha$, with respect to $\mathfrak{A}'$

$$\mathfrak{H} = \int_D \mathfrak{H}_\alpha [d\mu(\alpha)]^{\frac{1}{2}},$$

α varies over a finite interval D (cf. Ref. 18), such that the algebra $\mathfrak{K}$ splits into irreducible algebras $\mathfrak{K}_\alpha$ on $\mathfrak{H}_\alpha$ [more precisely: almost every $\mathfrak{K}_\alpha$ in the sense of the measure $\mu(\alpha)$ is irreducible].

$$\mathfrak{K} = \int^{\oplus} \mathfrak{K}_\alpha.$$

The commutant $\mathfrak{K}'$ consists of operators of the shape

$$\int^{\oplus} f(\alpha) I_\alpha,$$

where I_α is the unity operator on $\mathfrak{H}_\alpha$ and $f(\alpha)$ a measurable function of α on D.[19] The same decomposition with respect to $\mathfrak{A}'$ yields for $\mathfrak{R} = \int^{\oplus} \mathfrak{R}_\alpha$ (Ref. 20), where again the $\mathfrak{R}_\alpha$ are irreducible in $\mathfrak{H}_\alpha$. According to Borchers[1] the unitary representation $U(\Lambda, a)$ of the Lorentz group commutes with every element out of $\mathfrak{A}'$. Therefore $U(\Lambda, a) \in \mathfrak{R}$ and

$$U(\Lambda, a) = \int^{\oplus} U_\alpha(\Lambda, a).$$

The vector Ω can be also decomposed, $\Omega = \int^{\oplus} \Omega_\alpha$. Since Ω is cyclic with respect to $\mathfrak{A}$ (almost) every Ω_α is different from zero. On the other hand the algebra $\mathfrak{R}_\alpha$ is irreducible in $\mathfrak{H}_\alpha$ and therefore Ω_α is the unique vacuum in $\mathfrak{H}_\alpha$ (and is cyclic).[1-3,21]

Since Ω is cyclic for $\mathfrak{R}$ in $\mathfrak{H}$, we get (cyclic as well as non cyclic) vacuum states different from Ω by applying an element (different from λI) out of $\mathfrak{R}'$ onto Ω.

Each space $\mathfrak{H}_\alpha$ is a Hilbert space of infinite dimensions or dimension 1 since there are no finite-dimensional nontrivial representations of the Lorentz group.

To proceed further we have to impose some additional restrictions fitting our model. In particular we need some information concerning how $G(\beta)$ is related to the decomposition of $\mathfrak{H}$. We restrict ourselves here to the case considered in Sec. 2: $d\mu(\alpha) = d\alpha$ is pure continuous, D is normed to the interval $0 \cdots 2\pi$ and $G(\beta)$ connects the different spaces in the way given by (2.1) and (2.2). This implies that the representations of the Lorentz group have the same spectrum with the same multiplicity in every space $\mathfrak{H}_\alpha$.[22] It also rules out of our consideration the trivial case of $\mathfrak{H}_\alpha$ consisting only the of vacuum state Ω_α. Due to (2.3) the operator $G(\beta)$ does not commute with $\mathfrak{A}'$ and so does not belong to $\mathfrak{R}$.

Let us now return to the remark made at the end of Sec. 2 concerning the condition imposed on the Wightman functions. Let us consider the case when in addition to (2.2) a unitary transformation $H_\alpha(\gamma)$ exists in $\mathfrak{H}_\alpha$ with the property

$$\begin{aligned} H_\alpha(\gamma) A_\alpha(x) H_\alpha^\dagger(\gamma) &= e^{i\gamma} A_\alpha(x) \\ H_\alpha(x) A_\alpha^\dagger(x) H_\alpha^\dagger(\gamma) &= e^{-i\gamma} A_\alpha(x). \end{aligned} \tag{3.1}$$

Then

$$G_\alpha(\beta) A_\alpha(x) G_{\alpha+\beta}^\dagger(\beta) = H_{\alpha+\beta}(\beta) A_{\alpha+\beta}(x) H_{\alpha+\beta}^\dagger(\beta).$$

It follows that H_α commutes with $U_\alpha(\Lambda, a)$ and that

$$H_\alpha(\gamma)\Omega_\alpha = \lambda(\gamma)\Omega_\alpha \qquad |\lambda(\gamma)| = 1.$$

For a monomial A_α of field operators (n denotes the difference of the number of field operators and the number of its adjoints) we get in virtue of (2.2)

$$(\Omega_\alpha, A_\alpha \Omega_\alpha) = (\Omega_{\alpha+\beta}, A_{\alpha+\beta}\Omega_{\alpha+\beta}) e^{in\beta}. \tag{3.2}$$

[17] H.-J. Borchers, Ref. 1, see also Appendix I.

[18] F. Riesz and B. v. Sz.-Nagy, *Vorlesungen über Funktionalanalysis* (Deutscher Verlag der Wissenschaften, Berlin, 1956), Sec. 131, p. 343.

[19] M. A. Neumark, Ref. 12, Sec. 41.4, Theorem 6, p. 525.

[20] E.g.: M. A. Neumark Ref. 12, Sec. 41.3 or I. E. Segal, Mem. Am. Math. Soc. 9, 34 (1958).

[21] Under the assumptions stated at the beginning of this section the necessary and sufficient condition for a vector to be cyclic with respect to $\mathfrak{A}$ in $\mathfrak{H}$ is to have nonvanishing components in almost every $\mathfrak{H}_\alpha$.

[22] N. I. Achieser and I. M. Glasmann, *Theorie der Linearen Operatoren im Hilbert-Raum* (Akademie Verlag, Berlin, 1954), No. 71-72.

On the other hand, by means of (3.1), we get that (3.2) is equal to

$$(\Omega_{\alpha+\beta},\ A_{\alpha+\beta}\Omega_{\alpha+\beta}).$$

Consequently, the Wightman functions vanish for $n \neq 0$ and coincide in different space for $n = 0$,

$$(\Omega_\alpha,\ A_\alpha\Omega_\alpha) = (\Omega_\beta,\ A_\beta\Omega_\beta);$$

therefore, the algebras $\mathfrak{A}_\alpha$ are equivalent [the isometric mapping is given by $H^\dagger_{\alpha+\beta}(\beta)G_\alpha(\beta)$]. In this case the reducible $\mathfrak{R}$ is a factor. Ω cannot be cyclic.[3] It is clear that $\mathfrak{A}'$ does not belong to $\mathfrak{R}$. Nevertheless, a reduction of the algebra with respect to a maximal Abelian algebra in $\mathfrak{A}'$ is still possible. It gives again the decomposition (2.4). The charged free scalar field with proper commutation relations fit into this trivial case. $H_\alpha(\gamma)$ is then the conventional gauge transformation; it belongs to the algebra $\mathfrak{A}''_\alpha$. $H(\gamma) = \int^\oplus H_\alpha(\gamma)$ belong then to $\mathfrak{R}$ and applied to $A(x)$ yields the same result as $G(\gamma)$ although both operators have not much in common.

4. THE ALGEBRA OF "OBSERVABLES" AND ITS REDUCTION

Following the example of the conventional gauge theory where the observable quantities are gauge invariant we define "observables" as all bounded self-adjoint operators on $\mathfrak{H}$ which commute with $G(\beta)$. They form a van Neumann algebra $\mathfrak{O}$. It follows immediately that:

(1) $\mathfrak{O}$ does neither coincide with $\mathfrak{R}$ nor lies inside $\mathfrak{R}$. To exhibit it let us consider the projection operator P_Ω on $\Omega = \int^\oplus \Omega_\alpha$. P_Ω commutes with $G(\beta)$ since Ω is invariant under $G(\beta)$. On the other hand, P_Ω does not belong to $\mathfrak{R}$ unless $\mathfrak{R}$ is irreducible since Ω is a cyclic vector for $\mathfrak{R}$.

(2) $\mathfrak{R}$ is not contained in $\mathfrak{O}$ either. Should that be the case, $G(\beta)$ would commute with $\mathfrak{R}$, i.e. $G(\beta) \subset \mathfrak{R}'$, and $G_\alpha(\beta)$ would not connect $\mathfrak{H}_\alpha$ with $\mathfrak{H}_{\alpha+\beta}$.

(3) The intersection $\mathfrak{R} \cap \mathfrak{O}$ is not empty, e.g., the Lorentz group is contained in $\mathfrak{R} \cap \mathfrak{O}$.

(4) $\mathfrak{R}' \cap \mathfrak{O} = \lambda I$ (λ is a number) since the elements of $\mathfrak{R}'$, different from I do not commute with $G(\beta)$. It follows that

$$(\mathfrak{R} \cup \mathfrak{O}')'' = \mathfrak{B} \tag{4.1}$$

where $\mathfrak{B}$ denotes all bounded operators on $\mathfrak{H}$.

Obviously, $G(\beta) \in \mathfrak{O}$. Let us write $G(\beta) = e^{i\beta Q}$; then all spectral projections of the "charge" $Q = Q^\dagger$ belong to $\mathfrak{O}$ (Q itself is unbounded). Since on the other hand $G(\beta) \in \mathfrak{O}'$, $G(\beta)$ belongs to the center $\mathfrak{O} \cap \mathfrak{O}'$. Again in analogy to conventional theory we call the algebra of the center the algebra of "superselecting" operators. It follows that

(5) The center $\mathfrak{O} \cap \mathfrak{O}'$ coincides with $\mathfrak{O}'$. To show it take an $\alpha \in \mathfrak{O}'$, then α commutes with $G(\beta)$ since $G(\beta) \in \mathfrak{O}$; thus $\alpha \in \mathfrak{O}$. The last statement enables us to perform a complete measurement.[23,24]

(6) $\mathfrak{O}'$ consists of all bounded functions of Q only. This results from the fact that every $o \in \mathfrak{O}'$ commutes with every bounded operator commuting with Q; thus o is a function of Q.[25]

We can reformulate (4.1): the von Neumann algebra of the field $A(x)$ and $G(\beta)$, is the algebra of all bounded operators in $\mathfrak{H}$.

One easily finds eigenvalues of $G(\beta)$ and Q. For example, for

$$\Omega^{(n)} = \int^\oplus e^{-i\alpha n}\Omega_\alpha, \tag{4.2}$$

we have

$$G(\beta)\Omega^{(n)} = e^{in\beta}\Omega^{(n)}.$$

From the periodicity of $G(\beta)$ follows that n is integer. All vectors $\mathfrak{O}\Omega^{(n)}$ belong to the same eigenvalue n. It turns out that the whole spectrum of Q is represented by the numbers $n = 0, \pm1, \pm2, \cdots$ because of the relation (2.5) and the cyclicity of $\Omega^{(n)}$. The $\Omega^{(n)}$ form an orthonormal set of vectors complete in $\mathfrak{H}(0)$.

$\Omega \equiv \Omega^{(0)}$ is distinguished by the property to be invariant under $G(\beta)$ and under a TCP-operation provided the last transformation is defined appropriately in the $\mathfrak{H}_\alpha$ (Ω_α has to be TCP-invariant for every α). Under TCP-operation $\Omega^{(n)}$ goes over into $\Omega^{(-n)}$. The projections on the $\Omega^{(n)}$ are also observables not belonging to $\mathfrak{R}$. The Wightman functions for different $\Omega^{(n)}$ coincide.

If "physically realizable states" are defined as states the projections on which are observables,[26] then the vectors $\Omega^{(n)}$ are such states. There are other vacuum states which are not physically realizable (e.g., $\Omega' = \int^\oplus \eta_\Delta\Omega_\alpha$ if η_Δ is 1 on a subinterval of $0 \cdots 2\pi$ and zero elsewhere).

Let us now examine the decomposition of $\mathfrak{O}$ with respect to $\mathfrak{O}'$. The decomposition of $\mathfrak{O}$ with respect to a maximal Abelian subalgebra of $\mathfrak{O}'$,[20] is estab-

---

[23] J. M. Jauch, Helv. Phys. Acta 33, 711 (1960).
[24] J. M. Jauch and B. Misra, Helv. Phys. Acta 34, 699 (1961).
[25] Cf., e.g., F. Riesz and B. v. Sz.-Nagy, Ref. 18, Sec. 129, p. 335.
[26] See e.g. R. Streater and S. A. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964), pp. 5–7.

lished by the spectral decomposition of Q (we could choose the Cayley transformed of Q to deal with a bounded operator), i.e.,

$$\mathfrak{H} = \sum_n \oplus \mathfrak{H}_n,$$

$$\mathfrak{O} = \sum_n \oplus \mathfrak{O}_n,$$

$$\mathfrak{O}' = \{\sum_n \oplus c_n I_n\}$$

where I_n is the unity in $\mathfrak{H}_n$, the c_n are complex numbers. The algebra $\mathfrak{O}_n$ is irreducible in $\mathfrak{H}_n$ since $\mathfrak{O}' \subset \mathfrak{O}$. The spaces $\mathfrak{H}_n$ are "coherent" sectors in the sense of "superselection" rules. Every $\mathfrak{H}_n$ contains a vacuum state Ω^n and the Lorentz group is again unitarily represented in every $\mathfrak{H}_n$. Here we see immediately that no $\mathfrak{H}_n$ can be one dimensional: Since $\Omega^{(k)}$ is cyclic we know (if $\mathfrak{H}$ does not consist of vacuum states alone) that there is a product of smeared-out field operators A so that $A\Omega^{(k)} \in \mathfrak{H}_{k+m}$ is no vacuum state. But for the same A, it holds that $A\Omega^{(n)}$ for any n is no vacuum state since $\Omega^{(n)}$ is related to $\Omega^{(k)}$ by a unitary transformation V out of $\mathfrak{A}'$ (namely $V = \int^{\oplus} e^{i(n-k)\alpha} I_\alpha$), by virtue of which $A\Omega^{(n)} = \Omega'$ would imply $A\Omega^{(k)} = V^+\Omega'$, which again would be Lorentz-invariant contrary to the assumption.

The complementary character of the two decompositions with respect to the field and the observables is exhibited by the commutation relations of the "gauge operator"

$$L = \int_0^{2\pi \oplus} \alpha I_\alpha$$

related to the field and the "charge" $Q = \Sigma \oplus nI_n$ related to the observables: $[Q, L] = i$.

5. CONCLUDING REMARKS

The results obtained for a single model in Secs. 2–4 can be generalized to physically more interesting cases, e.g., to the case of a multiplet of $n > 2$ spinless real fields $A^{(\nu)}(x)$ where the transformation of the fields under the symmetry transformation $G(\beta)$ is given by[9]

$$G(\beta)A^{(\nu)}(x)G(\beta)^{-1} = \sum_{\mu=1}^{n} (e^{i\beta T})^{\nu}{}_{\mu} A^{(\mu)}(x), \qquad \nu = 1, \cdots n, \qquad (5.1)$$

where $G(\beta)$ is unitary, $T^{\nu}_{\mu} = \bar{T}^{\mu}_{\nu}$ is a c-number tensor; β a real number. The expectation value of an infinitesimal transformation (5.1) with respect to a vacuum state Ω' noninvariant under $G(\beta)$.

$$(\Omega', G(\beta)A^{(\nu)}(x)G(\beta)^{-1}\Omega') = \beta \sum_{\mu=1}^{n} T^{\nu}_{\mu}(\Omega', A^{(\mu)}(x)\Omega')_{-} + (\Omega', {}_{-}A^{(\nu)}(x)\Omega')$$

does not need to vanish; this amounts to

$$(\Omega', A^{(\mu)}(x)\Omega') \not\equiv 0.$$

This in turn leads to the conjecture known as the Goldstone theorem.[9] We are not going here to prove or disprove the latter. We intend only to point out the following: Let us return to our model. Assume the field we are considering has asymptotic fields belonging to a certain mass. Then these free fields possess the well-known gauge symmetry (and a locally conserved current) which is completely different from the transformation $G(\beta)$ as investigated here for the interacting field (see Sec. 2). Incidentally, the free fields may possess in addition the transformation property under $G(\beta)$.

ACKNOWLEDGMENT

We are grateful to Dr. M. Guenin, Dr. K. Hepp, and Dr. D. Uhlenbrock for discussions and we are indebted to Professor J. R. Oppenheimer for his kind hospitality at the Institute for Advanced Study. We wish to thank the U.S. Educational Commission for Fulbright travel grants.

APPENDIX I

We give here a proof slightly different from that of Borchers'[1] that under the assumptions stated in the introduction, $\mathfrak{A}'$ is Abelian.

Let us consider

$$(\Omega, A^c(x_1, \cdots x_n)uB^c(y_1 \cdots y_m)\Omega), \qquad (A1)$$

where A^c and B^c is a product of A and $A^\dagger$ taken at the world points $x_1, \cdots, y_m$ and u belongs to $\mathfrak{A}'$. As u commutes with A and $A^\dagger$ (A1) is equal to

$$(\Omega, A^c(x_1 \cdots x_n)B^c(y_1, \cdots y_m)\Omega'), \qquad (A2)$$

where $\Omega' \equiv u\Omega$ is again a vacuum state invariant under the Lorentz group. Applying the standard procedure based on the spectral condition, Lorentz invariance, and locality we see that (A2) is equal to

$$(\Omega, B^c(-y_m, \cdots -y_1)A^c(-x_n, \cdots -x_1)\Omega') = (\Omega, B^c(-y_m, \cdots -y_1)uA^c(-x_n, \cdots -x_1)\Omega) = (\Omega, A^c(x_1, \cdots x_n)\theta u^\dagger \theta B^c(y_1, \cdots y_m)\Omega),$$

where θ is a TCP operator (not necessarily unique). From cyclicity of Ω follows then that

$$u = \theta u^\dagger \theta. \qquad (A3)$$

Consider now the product uv where both u and v belong to $\mathfrak{A}'$. According to (A3) we have

$$uv = \theta v^\dagger u^\dagger \theta = \theta v^\dagger \theta \theta u^\dagger \theta = vu.$$

APPENDIX II

For the construction of $\mathfrak{K}$ we proceed in the following way[27]: Since the domain of $A \in \mathfrak{A}$ as well as $A^\dagger$ is dense in $\mathfrak{K}$ we may assume that A is closed. We can write for every $A \in \mathfrak{A}$ on $\mathfrak{D}$

$$A = \tfrac{1}{4}\{(I + A^\dagger)(I + A) - (I - A^\dagger)(I - A) + i(I + iA^\dagger)(I - iA) - i(I - iA^\dagger)(I + iA)\}$$

as a linear combination of four self-adjoint operators,[28] which on the other hand are uniquely determined by A on $\mathfrak{D}$. The families of finite projections of these self-adjoint operators build up an algebra $\mathfrak{A}_s$ the uniform closure of which is $\mathfrak{K}$.

As was shown in Ref. 27 we have $\mathfrak{A}_s' = \mathfrak{A}'$. Since $\mathfrak{K}$ is the uniform closure of $\mathfrak{A}_s$ we have $\mathfrak{A}_s' = \mathfrak{K}'$ too, so that $\mathfrak{A}_s'' = \mathfrak{K}''$.

[27] H. Reeh and S. Schlieder, "Über den Zerfall der Feldoperatoralgebra im Falle einer Vakuumentartung," preprint February 1962), Appendix, Sec. 2.

[28] Cf. e.g. N. I. Achieser and I. M. Glasmann, Ref. 22, No. 46, Theorem 2.

Some Exact Solutions of the Field Equations of General Relativity*

M. MISRA
Department of Physics, Temple University, Philadelphia, Pennsylvania
(Received 10 May 1965)

A method for generating nonstatic solutions of the Einstein–Maxwell field equations in vacuo from nonstatic solutions of the empty space–time field equations is given. It is shown that under certain conditions one of the metrics under investigation admits algebraically degenerate vacuum solutions. The necessary and sufficient condition in order that the vacuum solutions for this metric be of type null is obtained. Some exact solutions are listed.

1. INTRODUCTION

In the General Theory of Relativity a method for constructing (i) a stationary external solution, and (ii) a stationary internal solution, has been given by Ehlers.[1] Subsequently Bonnor[2] has given a method for constructing solutions of the Einstein–Maxwell field equations in vacuo from static solutions of the empty space–time field equations. The method works because of the fact that the two sets of field equations for the two metrics reduce to the same set of partial differential equations if the metric potentials in the two cases have functional relationships and the electromagnetic field tensor satisfies certain conditions. Thus knowing a set of solutions of these equations one can write down solutions for the two systems of field equations. Although in a sense one is generating both types of solutions at the same time one talks of generating one system from known solutions of the other because of the functional dependence of the metrics.

The object of the present investigation is to show that nonstatic solutions of the Einstein–Maxwell field equations in vacuo can be constructed from nonstatic solutions of the empty space–time field equations. It is shown that under certain conditions one of the metrics under investigation admits vacuum solutions which are algebraically degenerate according to the Pirani–Petrov classification.[3] Also the necessary and sufficient condition in order that the vacuum solutions for this metric be of the null type according to the Pirani–Petrov classification is obtained. A solution satisfying these conditions is obtained and it is found to be a solution representing plane gravitational waves. The corresponding solution representing plane electromagnetic waves is also given.

2. PRELIMINARIES

We shall be concerned with the following two metrics:

$$ds^2 = g_{\alpha\beta}\,dx^\alpha\,dx^\beta = e^{2W}\,d^*s^2 - \bar{e}^{2W}(dx^3)^2, \tag{2.1}$$

$$d\bar{s}^2 = \bar{g}_{\alpha\beta}\,dx^\alpha\,dx^\beta = e^{2P}\,d^*s^2 - \bar{e}^{2P}(dx^3)^2, \tag{2.2}$$

together with the auxiliary metric

$$d^*s^2 = {}^*g_{ij}\,dx^i\,dx^j, \tag{2.3}$$

where W, P, and ${}^*g_{ij}$ are functions of x^i only.

Throughout the paper Greek indices take the values 1, 2, 3, 4 and Latin indices take the values 1, 2, 4. The metric (2.3) is supposed to be that of an indefinite V_3. It is worth pointing out here that the metrics (2.1) and (2.2) are related to the corresponding metrics used by Bonnor by the complex transformation $x^3 \to iX^4$, and $x^4 \to iX^3$. The importance of this transformation is well known in the study of axially symmetric fields.

The Ricci tensors corresponding to the metrics (2.1), (2.2), and (2.3) will be denoted by $R_{\alpha\beta}$, $\bar{R}_{\alpha\beta}$ and ${}^*R_{ij}$, respectively. Also a semicolon denotes covariant differentiation with respect to (2.3) and a stroke denotes covariant differentiation with respect to (2.2).

3. THE FIELD EQUATIONS

Consider an electromagnetic field tensor $F_{\alpha\beta}$. We propose to establish a correspondence between P and W with some conditions on $F_{\alpha\beta}$ so that the field equations in the following two cases:

(i) $\bar{g}_{\alpha\beta}$ and $F_{\alpha\beta}$ satisfy the Einstein–Maxwell field equations in vacuo:

$$\bar{R}_{\alpha\beta} = -8\pi E_{\alpha\beta}, \tag{3.1}$$

$$E^\beta_\alpha = -F_{\alpha\sigma}F^{\beta\sigma} + \tfrac{1}{4}\bar{g}^\beta_\alpha F_{\sigma\epsilon}F^{\sigma\epsilon}, \tag{3.2}$$

* Work supported by the Aerospace Research Laboratories of the Office of Aerospace Research, U. S. Air Force.

[1] J. Ehlers, *Gravitation—An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962), Chap. 2.

[2] W. B. Bonnor, Z. Physik 161, 439 (1961).

[3] F. A. E. Pirani, *Gravitation—An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962), Chap. 6.

$$F_{\alpha\beta|\gamma} + F_{\beta\gamma|\alpha} + F_{\gamma\alpha|\beta} = 0, \tag{3.3}$$

$$F^{\alpha\beta}{}_{|\beta} = 0, \tag{3.4}$$

and

(ii) $g_{\alpha\beta}$ satisfy the field equations for empty space–time:

$$R_{\alpha\beta} = 0, \tag{3.5}$$

reduce to the same set of partial differential equations.

A simple calculation shows[4] that the Einstein–Maxwell field equations (3.1), (3.2), (3.3), and (3.4) for the metric (2.2) reduce to

$$^*R_{ij} + 2W_{;i}W_{;j} = 0, \tag{3.6}$$

$$^*g^{ij}W_{;ij} = 0, \tag{3.7}$$

if P and W are related by

$$e^{2P} = (1/a^2)\cosh^2 W, \tag{3.8}$$

and $F_{\alpha\beta}$ is given by

$$F_{3i} = [a/(4\pi)^{\frac{1}{2}}]\,\mathrm{sech}^2\, W\cdot W_{;i}, \tag{3.9}$$

$$F_{12} = F_{14} = F_{24} = 0, \tag{3.10}$$

a being a certain positive constant.

It can be shown that (3.7) follows from the contracted Bianchi identities for $^*R_{ij}$. Hence the problem of solving the Einstein–Maxwell field equations in vacuo under the conditions stated above reduces to that of solving Eqs. (3.6).

Equations (3.6) are precisely the conditions that

$$R_{\alpha\beta} = 0.$$

We are thus led to the result: $\bar{g}_{\alpha\beta}$, $F_{\alpha\beta}$ (or equivalently P, $^*g_{ij}$, $F_{\alpha\beta}$) form a solution of Eqs. (3.1), (3.2), (3.3), and (3.4) if P and $F_{\alpha\beta}$ are given by (3.8), (3.9), and (3.10) and Eqs. (3.6) are satisfied.

Since the system of Eqs. (3.6) is equivalent to (3.5) and thus the solutions of (3.6) determine the empty space–time metric (2.1) completely, we are generating solutions of the Einstein–Maxwell field equations in vacuo under certain conditions for the metric (2.2) from the corresponding solutions of the empty space–time field equations for the metric (2.1).

The geometrical significance of Eqs. (3.6) is that $^*R_{ij}$ has one simple eigenvector W^i with the eigenvalue 2 $W_{;i}W^i_{;}$ and an eigen 2-space W = const with eigenvalue 0.

[4] M. Misra, Proc. Nat. Inst. Sci. India **29A**, 104 (1962).

4. SOME GENERAL CONSIDERATIONS

In this section we study some properties of the curvature tensor for the vacuum solutions corresponding to the metric (2.1).

The three-dimensional curvature tensor for the metric (2.3) is given by[5]

$$^*R_{ikln} = {}^*g_{in}{}^*R_{kl} - {}^*g_{kn}{}^*R_{il} + {}^*g_{kl}{}^*R_{in} - {}^*g_{il}{}^*R_{kn} - \tfrac{1}{2}({}^*g_{in}{}^*g_{kl} - {}^*g_{kn}{}^*g_{il})^*R, \tag{4.1}$$

which in view of (3.6) can be written as

$$^*R_{ikln} = R'_{ikpn}\mu^p\mu_l + R'_{iklp}\mu^p\mu_n - \tfrac{1}{2}R'_{ikln}\mu_p\mu^p, \tag{4.2}$$

where

$$^*R = {}^*g^{ij}{}^*R_{ij}, \tag{4.3}$$

$$\mu_i = (-2)^{\frac{1}{2}}W_{;i}, \tag{4.4}$$

and

$$R'_{ikln} = {}^*g_{in}{}^*g_{kl} - {}^*g_{kn}{}^*g_{il}. \tag{4.5}$$

It follows from (4.2) in view of the symmetry properties of R'_{ikln} that

$$^*R_{iklm}\mu^m = (\tfrac{1}{2}{}^*R)R'_{iklm}\mu^m. \tag{4.6}$$

Now if $^*R_{ij}$ has vanishing trace we have

$$^*R = +\mu_i\mu^i = 0, \tag{4.7}$$

so that μ_i is a null vector and we consequently have

$$^*R_{iklm}\mu^m = 0. \tag{4.8}$$

Now the nonvanishing components of the curvature tensor for the metric (2.1) are given by

$$\begin{aligned} R_{hijk} = e^{2W}[&{}^*R_{hijk} + {}^*g_{hk}(W_{;ij} - W_{;i}W_{;j}) \\ &+ {}^*g_{ij}(W_{;hk} - W_{;k}W_{;h}) \\ &- {}^*g_{hj}(W_{;ik} - W_{;i}W_{;k}) \\ &- {}^*g_{ik}(W_{;hj} - W_{;h}W_{;j}) \\ &+ ({}^*g_{hk}{}^*g_{ij} - {}^*g_{hj}{}^*g_{ik})W_{;m}W^m_{;}], \end{aligned} \tag{4.9}$$

and

$$R_{h33j} = e^{2W}[W_{;hj} - 3W_{;h}W_{;j} + {}^*g_{hj}W_{;k}W^k_{;}], \tag{4.10}$$

where $^*R_{hijk}$ is given by (4.1).

It follows from (4.9) and (4.10) in view of (4.4), (4.7), and (4.8) that

$$R_{hijk}W_{;}{}^iW_{;}{}^k = 0, \tag{4.11}$$

and

$$R_{h33j}W_{;}{}^j = 0. \tag{4.12}$$

[5] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice Hall, 1942), p. 173.

Hence we obviously have

$$W_{;[\epsilon}R_{\alpha]\beta\gamma\delta}W_{;}^{\beta}W_{;}^{\delta} = 0. \tag{4.13}$$

It follows therefore that all vacuum solutions of (2.1) are algebraically degenerate according to the Pirani–Petrov classification (see, for instance, Pirani[3] where detailed references will be found) $W_{;k}$ being a multiple Debever–Penrose vector. It also follows from the Goldberg–Sachs[6] Theorem that the null vector field $W_{;k}$ is not only geodesic but shear-free.

It is easy to see that the curvature tensor for (2.1) satisfies

$$R_{\alpha\beta\gamma\delta}W_{;}^{\delta} = 0 \tag{4.14}$$

if and only if

$$W_{;i}W_{;hj} - W_{;h}W_{;ij} = 0. \tag{4.15}$$

Thus the necessary and sufficient condition that the vacuum solutions of (2.1) be of type null according to the Pirani–Petrov classification[3] is given by (4.15). It may be noted here that the Eqs. (4.7) which can also be written as

$$W_{;i}W_{;}^{i} = 0 \tag{4.16}$$

and (4.15) are independent. Thus it should be possible to find solutions of (3.6) which satisfy (4.16) but not (4.15). Such solutions will not represent plane gravitational waves. A class of solutions which do satisfy (4.15) and (4.16) is given in Sec. 5.

The conditions (4.15) can also be put in the following equivalent forms:

$$(W_{;h}/W_{;i})_{;j} = 0, \tag{4.17}$$

and

$$^{*}R_{hk}{}^{*}R_{ik;j} - {}^{*}R_{ik}{}^{*}R_{hk;j} = 0, \tag{4.18}$$

which is possible because the vector $W_{;i}$ is supposed to be nonzero. Otherwise if $W_{;i}$ were a zero vector then the Ricci tensor for (2.3) would vanish and consequently (2.3) would be a flat metric, which is of no interest here.

The immediate interpretation of (4.17) is that the quantities $(W_{;h}/W_{;i})$ are covariant constant. These hold identically if $h = i$, which is to be expected since (4.15) is antisymmetric in h and i. We are thus left with the case when $h \neq i$. We have in general three of these quantities of which only two are independent and satisfy six equations. These equations involve $^{*}g_{ij}$ and W and therefore are rather complicated. The conditions (4.18) although quadratic in the Ricci tensor for (2.3) and therefore highly nonlinear have only one advantage over the conditions (4.17). These involve only the $^{*}g_{ij}$.

[6] J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. Suppl. 22, 13 (1962).

5. SOLUTIONS

(i) Some exact solutions of (3.6) were obtained by the author[4] by assuming the metric (2.3) to be conformally flat so that

$$^{*}g_{ij} = e^{2\sigma}\eta_{ij}, \tag{5.1}$$

where σ is a function of x^{i} and η_{ij} has its usual significance diag $(-1, -1, +1)$ and zero elsewhere.

Assuming further that σ and W are functions of u where

$$u = \alpha_{i}x^{i}, \tag{5.2}$$

α_{i} being constants which satisfy

$$\eta^{ij}\alpha_{i}\alpha_{j} = 0, \tag{5.3}$$

the solution obtained is

$$\begin{aligned} \sigma &= -\log\varphi, \\ W &= \int\left[\frac{1}{2\varphi}\frac{d^{2}\varphi}{du^{2}}\right]^{\frac{1}{2}}du + \text{const}, \end{aligned} \tag{5.4}$$

φ being an arbitrary function of u.

This solution satisfies all the conditions for plane gravitational waves as obtained by various workers. Also the corresponding solution of the Einstein–Maxwell field equations in vacuo is given by

$$\begin{gathered} e^{2P} = \frac{1}{a^{2}}\cosh^{2}W, \quad W = \int\left[\frac{1}{2\varphi}\frac{d^{2}\varphi}{du^{2}}\right]^{\frac{1}{2}}du + \text{const}, \\ F_{3i} = (4\pi)^{-\frac{1}{2}}a\,\operatorname{sech}^{2}W\cdot W_{;i}, \quad F_{12} = F_{14} = F_{24} = 0, \\ {}^{*}g_{ij} = \varphi^{-2}\eta_{ij}, \quad u = \alpha_{i}x^{i}, \quad \eta^{ij}\alpha_{i}\alpha_{j} = 0. \end{gathered} \tag{5.5}$$

This gives a class of solutions representing plane electromagnetic waves in vacuo.

(ii) If instead of assuming (2.3) to be conformally flat, we simply assume that it is conformal to another metric

$$^{*}g_{ij} = e^{2\sigma}\tilde{g}_{ij}, \tag{5.6}$$

and denote by an overhead $\sim$ all quantities defined with respect to $\tilde{g}_{ij}$, then (3.6) gives

$$\begin{aligned} &\tilde{R}_{ij} + \sigma_{,ij} - \sigma_{,i}\sigma_{,j} \\ &\quad + \tilde{g}_{ij}\tilde{g}^{km}(\sigma_{,km} + \sigma_{,k}\sigma_{,m}) = -2W_{;i}W_{;j}, \end{aligned} \tag{5.7}$$

where a comma denotes covariant derivative with respect to $\tilde{g}_{ij}$.

The substitution

$$W = \sigma \tag{5.8}$$

leads to

$$\begin{aligned} &\tilde{R}_{ij} + \sigma_{,ij} + \sigma_{,i}\sigma_{,j} \\ &\quad + \tilde{g}_{ij}\tilde{g}^{km}(\sigma_{,km} + \sigma_{,k}\sigma_{,m}) = 0, \end{aligned} \tag{5.9}$$

which on contraction gives

$$\tilde{R} + 4\tilde{g}^{ij}(\sigma_{,ij} + \sigma_{,i}\sigma_{,j}) = 0. \tag{5.10}$$

From (5.9) and (5.10) it follows that

$$\tilde{R}_{ij} + \sigma_{,ij} + \sigma_{,i}\sigma_{,j} - (\tilde{R}/4)\tilde{g}_{ij} = 0, \tag{5.11}$$

which can be written as

$$U_{,ij} + [\tilde{R}_{ij} - (\tilde{R}/4)\tilde{g}_{ij}]U = 0, \tag{5.12}$$

where

$$\sigma = \log U. \tag{5.13}$$

Equations (5.12) are a system of six partial differential equations for determining $\tilde{g}_{ij}$ and U in all four functions. Since the three contracted Bianchi identities for $\tilde{R}_{ij}$ reduce to only one independent equation, Eq. (3.7), we have five equations in four unknowns.

The important point to be noted here is that Eqs. (5.12) are linear in U—a situation which exists only in the axially symmetric case.

The simplest solution of (5.12) is

$$\tilde{g}_{ij} = \eta_{ij}, \qquad U = \beta_i x^i, \tag{5.14}$$

with β_i arbitrary constants.

While it appears that ${}^*g_{ij}$ and W for this solution are functions of three variables, that is not the case. By a suitable rotation of the axes it can be shown that ${}^*g_{ij}$ and W are functions of only one variable.

If we try to solve (5.12) by putting

$$\tilde{R}_{ij} - (\tilde{R}/4)\tilde{g}_{ij} = 0, \tag{5.15}$$

we find that

$$\tilde{R}_{hijk} = 0, \tag{5.16}$$

and consequently we are led back to the solution (5.14) already given.

(iii) If we try to solve Eq. (3.6) by assuming that ${}^*R_{ij}$ is recurrent, i.e.,

$${}^*R_{ij;k} = \theta_k {}^*R_{ij} \tag{5.17}$$

where θ_k is an arbitrary gradient vector, we have

$$W_{;ik}W_{;j} + W_{;i}W_{;jk} = \theta_k W_{;i}W_{;j}, \tag{5.18}$$

which leads to

$$W_{;ik} = W_{;i}V_{;k}, \tag{5.19}$$

where

$$V_{;k} = -\tfrac{1}{2}[\log (W_{;i}W^{i}_{;})]_{;k} + \theta_k \tag{5.20}$$

and we have assumed that $W_{;i}$ is a nonnull vector. It follows that $W_{;i}$ is itself a recurrent vector.

Thus the necessary and sufficient condition in order that ${}^*R_{ij}$ be recurrent is that $W_{;i}$ be a recurrent vector.

Now since

$$\theta_k = {}^*R_{;k} \tag{5.21}$$

and therefore

$$V_{;k} = [{}^*R - \tfrac{1}{2}\log (W_{;i}W^{i}_{;})]_{;k} \tag{5.22}$$

it follows that

$$W_{;ikl} - W_{;ilk} = W_{;il}V_{;k} + W_{;i}V_{;kl} - W_{;ik}V_{;l} - W_{;i}V_{;lk} = 0, \tag{5.23}$$

in view of (5.19). Hence

$${}^*R_{hijk} = 0. \tag{5.24}$$

Consequently from (3.6) it follows that

$$W_{;i} = 0 \tag{5.25}$$

so that W is a constant. This solution is obviously trivial. This result is true whatever the dimensions of space so long as the Ricci tensor satisfies (3.6).

ACKNOWLEDGMENT

Part of this work was carried out at Lehigh University. The author wishes to express his gratitude to Professor P. Havas for his hospitality there as well as at Temple University and for his reading of the manuscript and helpful suggestions.

Lehmann–Symanzik–Zimmermann Formalism in a Model Field Theory

LEONARD M. SCARFONE
Department of Physics, University of Vermont, Burlington, Vermont
(Received 4 June 1965)

The Lehmann–Symanzik–Zimmermann (LSZ) formalism is used as a calculational method to obtain an equation for the determination of the $V + N$ potential energy in a slightly modified Lee model. This approach to the problem is compared with the standard eigenvalue technique and the dispersion theory treatment. It is found that an exact solution to the $VN - 2N\theta$ sector can be obtained with one algebraic equation and this evidence gives additional support to the desirability of the LSZ method as a calculational tool for the model.

I. INTRODUCTION

IT has recently been shown by Maxon and Curtis[1] that the Lehmann–Symanzik–Zimmermann[2] (henceforth called LSZ) formalism can be used to calculate the quantities of interest in the first nontrivial sector of the Lee model.[3] As is well known, these are the elastic $N + \theta$ scattering amplitude, the V-particle self-energy and the wavefunction renormalization constant. The solution of the $V - N\theta$ sector is also obtainable with simple, straightforward eigenvalue equations[3] and with the methods of dispersion theory.[4,5] Using the latter technique, Amado[6] has extended the exactly solved part of the model to include both the $V + \theta \rightarrow V + \theta$ elastic scattering and the production process $V + \theta \rightarrow N + 2\theta$.

In addition to these transition amplitudes and renormalization constants, it has also been shown that dispersion methods[7] can be used to calculate two simultaneous algebraic equations for the vertex functions[8] $\Gamma = \langle V| f_N |B\rangle$ and $\Gamma' = \langle N| f_V |B\rangle$, which yield an equation for the determination of the $V + N$ potential energy in agreement with that derived by direct methods.[9] In this paper, as in Ref. 1, we are interested in demonstrating the usefulness of the LSZ formalism as a calculational technique. For this purpose, we investigate the $VN - 2N\theta$ sector of a slightly modified Lee model and show that the LSZ approach solves the $V + N$ problem and the $2N + \theta$ scattering with one algebraic equation. We see that the binding energy equation is obtained by studying the analytic structure of the $V + N$ propagator and the scattering amplitude then follows.

It will be recalled that the model field theory under consideration here describes the coupling of two heavy fermions V and N with a relativistic boson θ, such that the only virtual elementary process allowed by the selection rules is $V \rightleftarrows N + \theta$. In order to simplify the procedure in our investigation, we conveniently set the $V + N$ separation parameter equal to zero and require that all field operatiors satisfy commutation relations. It is readily verified that the latter modification does not alter the classical results for the mass and wavefunction renormalizations. In Sec. II we briefly comment on the formal structure of the model and summarize the $V + N$ problem from the point of view of dispersion theory and standard techniques. Section III will then present the LSZ approach by introducing the appropriate τ functions[10] and their Matthew–Salam equations.[11]

II. $VN - 2N\theta$ SECTOR

The basic coupling $V \rightleftarrows N + \theta$ clearly indicates that the bare and physical V-particles do not coincide and that the self-energy effects are due to $N + \theta$ "bubbles." Therefore, the renormalized Hamiltonian is

$$H = m_0 Z \psi_V^\dagger \psi_V + m \psi_N^\dagger \psi_N + \sum_k \omega a_k^\dagger a_k + g \psi_V^\dagger \psi_N A + g A^\dagger \psi_N^\dagger \psi_V, \quad (1)$$

[1] M. S. Maxon and R. B. Curtis, Phys. Rev. **137**, B996 (1965).
[2] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **1**, 205 (1955).
[3] T. D. Lee, Phys. Rev. **95**, 1329 (1954). G. Källén and W. Pauli, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **30**, No. 7 (1955).
[4] L. M. Goldberger and S. B. Treiman, Phys. Rev. **113**, 1163 (1959).
[5] P. DeCelles and G. Feldman, Nucl. Phys. **14**, 517 (1959).
[6] R. D. Amado, Phys. Rev. **122**, 696 (1961). The complete eigenvalue solution of the $V + \theta$ sector has recently been reported in the literature: A. Pagnamenta, J. Math. Phys. **6**, 955 (1965).
[7] L. M. Scarfone, Nucl. Phys. **39**, 658 (1962).
[8] $|B\rangle$, $|V\rangle$, and $|N\rangle$ are the physical $V + N$, V, and N states, respectively, and $f_V(f_N)$ are the $V(N)$ current operators at time zero.
[9] S. Weinberg, Phys. Rev. **102**, 285 (1955).
[10] The τ functions are vacuum expectation values of time-ordered products of Heisenberg operators.
[11] P. T. Matthews and A. Salam, Proc. Roy. Soc. (London) **A221**, 128 (1953).

where

$$A = \sum_k \frac{f(\omega)}{(2\omega)^{\frac{1}{2}}} a_k, \quad \omega = (k^2 + \mu^2)^{\frac{1}{2}}, \quad m_0 = m + \delta m. \tag{2}$$

The cutoff function $f(\omega)$ assures the convergence of all integrals and the renormalized coupling constant g is less than its critical value. ψ_V is the renormalized V-particle operator and Z is a renormalization constant. We are quantizing in a box of unit volume and for convenience the heavy particles have equal mass m. The nonvanishing commutators are

$$[a_k, a_{k'}^\dagger] = \delta_{kk'}, \quad [\psi_N, \psi_N^\dagger] = 1, \quad [\psi_V, \psi_V^\dagger] = Z^{-1}. \tag{3}$$

The Lee model literature now contains three independent derivations[1,3,5] of the mass and wavefunction renormalization constants and these are known to be

$$\delta m = \frac{g^2}{2Z} \sum_k \frac{f^2(\omega)}{\omega^2}, \tag{4}$$

$$Z = 1 - \frac{g^2}{2} \sum_k \frac{f^2(\omega)}{\omega^3}, \tag{5}$$

where it is required that $\langle 0| \psi_V |V\rangle = 1$. From the selection rules we know that $|B\rangle$ is a superposition of two states. One of these is the bare $V + N$ state and the other is the bare $2N + \theta_k$ multiplied by a momentum space wavefunction $\varphi(k)$ and summed over k. Calling the eigenvalue of the state $|B\rangle$ $W_B = 2m + \omega_B$ where $\omega_B < \mu$, and using (1)–(3), we obtain

$$\omega_B = \delta m + g\left(\frac{2}{Z}\right)^{\frac{1}{2}} \sum_k \frac{f(\omega)\varphi(k)}{(2\omega)^{\frac{1}{2}}}, \tag{6}$$

$$(\omega - \omega_B)\varphi(k) = -g(2/Z)^{\frac{1}{2}} f(\omega)/(2\omega)^{\frac{1}{2}}. \tag{7}$$

Eliminating $\varphi(k)$ from (6) and (7) and using (4) and (5), we get the following equation for the determination of the eigenvalue ω_B as a function of the renormalized coupling constant:

$$F(\omega_B) \equiv 1 + \beta(\omega_B) + \lambda(\omega_B) = 0, \tag{8}$$

where in general

$$\beta(\omega) = g^2\left(\frac{\omega}{2}\right) \sum_{k'} \frac{f^2(\omega')}{\omega'^3(\omega' - \omega - i\epsilon)}, \tag{9}$$

$$\lambda(\omega) = \frac{g^2}{2\omega} \sum_{k'} \frac{f^2(\omega')}{\omega'(\omega' - \omega - i\epsilon)}. \tag{10}$$

Since the fermion source has no recoil, we can interpret ω_B as the $V + N$ potential energy at zero separation. Equation (8) has been solved in Ref. 9 for the case of nonrelativistic θ-particles with no cutoff and the conclusion is that ω_B is real and single valued providing g is less than its critical value.

In Ref. 7, the vertex functions Γ and Γ' are chosen as a starting point for a dispersion theory derivation of (8). Contracting the V-particle in the former and the N-particle in the latter, we are led to consider[12]

$$\Gamma = i \int_{-\infty}^{\infty} e^{imt} \langle 0| [f_V(t), f_N] \theta(t) |B\rangle \, dt, \tag{11}$$

$$\Gamma' = i \int_{-\infty}^{\infty} e^{imt} \langle 0| [f_N(t), f_V] \theta(t) |B\rangle \, dt. \tag{12}$$

The introduction of a complete set of intermediate states into each of these expressions reveals that both Γ and Γ' are related to the vertex functions

$$K(\omega) = [(2\omega)^{\frac{1}{2}}/f(\omega)]\langle 0| f_V |N\theta_k\rangle, \tag{13a}$$

$$R(\omega) = [(2\omega)^{\frac{1}{2}}/f(\omega)]\langle N\theta_k| f_N |B\rangle. \tag{14a}$$

A dispersion theory treatment of these quantities next introduces the $N + \theta$ scattering amplitude as is clearly shown in Figs. 1 and 2. In this way,

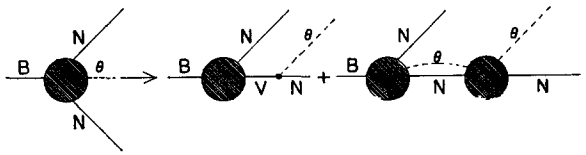

FIG. 1. Dispersion graph for the vertex $\langle N\theta|f_N|B\rangle$.

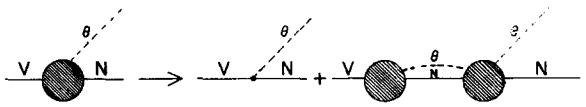

FIG. 2. Dispersion graph for the vertex $\langle 0|f_V|N\theta\rangle$.

we are forced to solve singular integral equations of the Low or Ommes types for each of these functions, and the results for K and R are[4,7]

$$K(\omega) = -g[1 + \beta(\omega)]^{-1}, \tag{13b}$$

$$R(\omega) = -Z[(\Gamma/\omega Z) + (\Gamma'/\omega_B)]K^*(\omega), \tag{14b}$$

where we have taken $|N\theta_k\rangle$ to mean "in" states. With these considerations, (11) and (12) can ultimately be written as

$$\omega_B\Gamma = \omega_B\Gamma(1 - Z^{-1}) - \delta m\Gamma', \tag{15}$$

$$\omega_B\Gamma' = \omega_B(\Gamma + Z\Gamma')\{[1 + \beta(\omega_B)]^{-1} - Z^{-1}\} - \delta m\Gamma', \tag{16}$$

respectively. When we eliminate Γ and Γ' from these equations, we are led to (6). The possibility of an exact solution in this case, as in other similar examples,[4-6] is traced to the fact that we are dealing

[12] $\theta(t) = 0$ for $t < 0$; $\theta(t) = 1$ for $t > 0$.

with a finite set of coupled integral equations brought about by the selection rules.

III. LSZ APPROACH

In this section we study the properties of the four τ functions associated with the graphs shown in Fig. 3. The $V + N$ propagator, the $VN - 2N\theta$

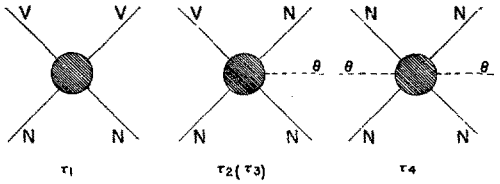

FIG. 3. Graphs corresponding to the τ functions in the $VN - 2N\theta$ sector.

vertex and the Green's function for $2N + \theta$ scattering are described by τ_1, $\tau_2(\tau_3)$, and τ_4, respectively, where these quantities are defined by the time-ordered products.

$$\tau_1(s) = \langle 0|\, T[\psi_V(s)\psi_N(s)\psi_N^\dagger(0)\psi_V^\dagger(0)]\,|0\rangle, \quad (17a)$$

$$\tau_2(s;\omega) = [(2\omega)^{\frac{1}{2}}/f(\omega)] \times \langle 0|\, T[\psi_N(s)\psi_N(s)a_k(s)\psi_N^\dagger\psi_V^\dagger]\,|0\rangle, \quad (17b)$$

$$\tau_3(s;\omega) = [(2\omega)^{\frac{1}{2}}/f(\omega)]\langle 0|\, T[\psi_V(s)\psi_N(s)\psi_N^\dagger\psi_N^\dagger a_k^\dagger]\,|0\rangle, \quad (17c)$$

$$\tau_4(s;\omega,\omega') = [(4\omega\omega')^{\frac{1}{2}}/f(\omega)f(\omega')] \times \langle 0|\, T[\psi_N(s)\psi_N(s)a_k(s)a_{k'}^\dagger\psi_N^\dagger\psi_N^\dagger]\,|0\rangle, \quad (17d)$$

where $s = t' - t$. The field equations

$$Z(i\,d/dt - m_0)\psi_V(t) = g\psi_N(t)A(t), \quad (18a)$$

$$(i\,d/dt - m)\psi_N(t) = gA^\dagger(t)\psi_V(t), \quad (18b)$$

$$(i\,d/dt - \omega)a_k(t) = g(2\omega)^{-\frac{1}{2}}f(\omega)\psi_N^\dagger(t)\psi_V(t), \quad (18c)$$

together with the equal time commutators (3) give the Matthew–Salam equations

$$Z(i\,d/ds - m_0 - m)\tau_1(s) = i\,\delta(s) + g\sum_k \frac{f^2(\omega)}{2\omega}\tau_2(s;\omega), \quad (19a)$$

$$(i\,d/ds - 2m - \omega)\tau_2(s;\omega) = 2g\tau_1(s), \quad (19b)$$

$$(i\,d/ds - 2m - \omega)\tau_3(s;\omega) = 2g\tau_1(s), \quad (19c)$$

$$(i\,d/ds - 2m - \omega)\tau_4(s;\omega,\omega') = 4i\omega\delta(s)\delta_{kk'}/f^2(\omega) + 2g\tau_3(s;\omega'). \quad (19d)$$

An equivalent equation for τ_4 is

$$(i\,d/ds - 2m - \omega')\tau_4(s;\omega,\omega') = 4i\omega\delta(s)\delta_{kk'}/f^2(\omega) + 2g\tau_2(s;\omega). \quad (19e)$$

Since $\tau_2(o;\omega) = \tau_3(o;\omega)$ and $\tau_4(o;\omega,\omega') = \tau_4(o;\omega',\omega)$, it is clear from the above equations that $\tau_2(s;\omega) = \tau_3(s;\omega)$ and that $\tau_4(s;\omega,\omega')$ is symmetric under the interchange of ω and ω'. This symmetry corresponds to an interchange of the two θ-particles in Fig. 3.

Introducing the Fourier representations

$$\tau_\alpha(s;\omega,\omega') = \frac{i}{2\pi}\int_{-\infty}^{\infty} dW\, e^{-iWs}\hat{\tau}_\alpha(W;\omega,\omega'), \quad (20a)$$

$$\hat{\tau}_\alpha(W;\omega,\omega') = \frac{1}{i}\int_{-\infty}^{\infty} ds\, e^{iWs}\tau_\alpha(s;\omega,\omega'), \quad (20b)$$

we immediately obtain

$$(W - m_0 - m)\hat{\tau}_1(W) = \frac{1}{Z} + \frac{g}{Z}\sum_k \frac{f^2(\omega)}{2\omega}\hat{\tau}_2(W;\omega), \quad (21a)$$

$$(W - 2m - \omega)\hat{\tau}_2(W;\omega) = 2g\hat{\tau}_1(W), \quad (21b)$$

$$(W - 2m - \omega)\hat{\tau}_4(W;\omega,\omega') = 4\omega\delta_{kk'}/f^2(\omega) + 2g\hat{\tau}_3(W;\omega'). \quad (21c)$$

It is obvious from the structure of these equations that the sector is solved since $\hat{\tau}_2(W,\omega)$ can be expressed in terms of $\hat{\tau}_1(W)$ by (21b) and the latter quantity can subsequently be removed from the sum in (21a) and combined with the left-hand side. In fact, we have

$$\hat{\tau}_2(W;\omega) = 2g\lim_{\epsilon\to 0+}\frac{\hat{\tau}_1(W)}{W - 2m - \omega + i\epsilon}, \quad (22)$$

$$\hat{\tau}_1(W) = (W - 2m)^{-1}F^{-1}(W - 2m). \quad (23)$$

In order to obtain (8), it is necessary to consider the analytic properties of the $V + N$ propagator, $\hat{\tau}_1(W)$. To this end, we put $\tau_1(s)$ into (20b) and insert a complete set of intermediate states which in this sector are the physical state $|B\rangle$ and the $2N + \theta$ scattering states. A simple calculation gives

$$\hat{\tau}_1(W) = \frac{|Z_B|^2/Z}{W - W_B + i\epsilon} + \sum_k \frac{|\langle 0|\,\psi_V\psi_N\,|2N\theta_k\rangle|^2}{W - 2m - \omega + i\epsilon}, \quad (24)$$

where Z_B is the normalization constant for the state $|B\rangle$. Requiring that the denominator in (23) vanish at the simple pole $W = W_B$, we thus arrive at the desired result. To determine $|Z_B|^2$ we evaluate the residue of $\hat{\tau}_1(W)$ at the pole and obtain

$$|Z_B|^2 = \left[1 + \frac{g^2}{Z}\sum_k \frac{f^2(\omega)}{\omega(\omega - \omega_B)^2}\right]^{-1}.$$

The LSZ approach solves the $V + N$ problem with one algebraic equation and this is considerably simpler than the dispersion analysis. At this point we collect the $\hat{\tau}$ functions

$$\hat{\tau}_1(W) = (W - 2m)^{-1}F^{-1}(W - 2m), \tag{25a}$$

$$\hat{\tau}_2(W;\omega) = \hat{\tau}_3(W;\omega) = 2g(W - 2m)^{-1}F^{-1}(W - 2m)(W - 2m - \omega)^{-1}, \tag{25b}$$

$$\hat{\tau}_4(W;\omega,\omega') = [4\omega\delta_{kk'}/f^2(\omega)](W - 2m - \omega)^{-1} + 4g^2(W - 2m)^{-1}F^{-1}(W - 2m) \times (W - 2m - \omega)^{-1}(W - 2m - \omega')^{-1}. \tag{25c}$$

To obtain the $2N + \theta$ scattering amplitude, we introduce the S-matrix element

$$S = \langle 2N\theta_{k'}^{(-)} \mid 2N\theta_k^{(+)}\rangle, \tag{26}$$

where the "plus" and "minus" refer to the "in" and "out" states, respectively. We proceed directly to the completely contracted form of S and write it as

$$S = \delta_{kk'} - \frac{f(\omega)f(\omega')}{2(4\omega\omega')^{\frac{1}{2}}}\int_{-\infty}^{\infty} dt' \times \int_{-\infty}^{\infty} dt\, e^{i(2m+\omega')t'}\, \bar{D}(t')\tau_4(t', t)\, \bar{D}^*(t)\, e^{-i(2m+\omega)t}, \tag{27}$$

where

$$D(t) = (i\, d/dt - 2m - \omega).$$

Replacing $\tau_4(t', t) = \tau_4(s)$ by its Fourier representation (20a) and carrying out the differentiations and integrations in (27), we get, with (25c),

$$S = \delta_{kk'} - 4\pi i g^2\delta(\omega - \omega')[f(\omega)f(\omega')/(4\omega\omega')^{\frac{1}{2}}][\omega F(\omega)]^{-1}. \tag{28}$$

It is clear that the scattering amplitude has a pole at the unphysical value of the energy $\omega = \omega_B$. In the language of dispersion theory this is, of course, a bound-state pole and it enters because the discrete state $|B\rangle$ has the same quantum numbers as the $2N + \theta$ scattering state. It is also possible to obtain (28) with the methods of dispersion theory and eigenvalue equations.

IV. CONCLUDING REMARKS

We have seen that the LSZ formalism can be used to obtain an equation for the determination of the $V + N$ potential energy and, in fact, solve the entire $VN - 2N\theta$ sector of a slightly modified Lee model. With this approach it would not be very difficult to extend these considerations to the $V + nN$ ($n > 1$) potential which has been solved by the eigenvalue technique,[13] but, on the other hand, is rather unyielding to the methods of dispersion theory.

We can also think of using the LSZ approach for the two V problem which has not been exactly solved with an eigenvalue equation[14] and seems hopelessly complicated from the dispersion-theory point of view. A zero separation of the two V-particles would be invoked as in the $V + N$ case, and the appropriate τ-functions would be those associated with the graphs shown in Fig. 4, namely, the

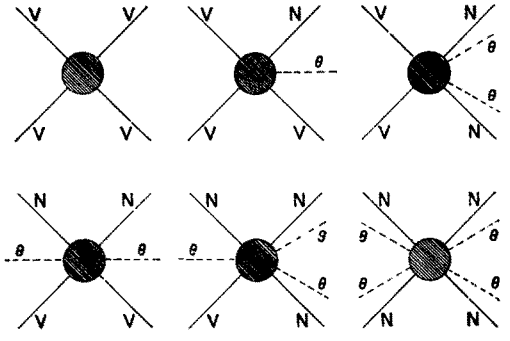

FIG. 4. Graphs corresponding to the τ functions that would be required in the $VV - VN\theta - 2N2\theta$ sector.

$V + V$ propagator, the $VV - VN\theta$ vertex, the $VV - 2N2\theta$ vertex, the Green's function for elastic $V + N\theta \to V + N\theta$ scattering, the production process $V + N + \theta \to 2N + 2\theta$ and the Green's function for $2N + 2\theta$ scattering. The LSZ method should be especially helpful here since the symmetry properties of the τ functions would be immediately obvious from the Matthew–Salam equations as is the case with the $V - N\theta$ and $VN - 2N\theta$ sectors. The two-V problem is presently under investigation.

[13] L. M. Scarfone, Nuovo Cimento 24, 480 (1962).
[14] R. Ascoli and E. Minardi, Nuovo Cimento 14, 1254 (1959).

Translations of the Discrete Bose–Einstein Operators*†

DAVID SHELUPSKY
City University of New York
(Received 12 April 1965)

The effects on a certain family of representations of the commutation relations of the operator-translations $a_k \to a_k + c_k$, $a_k^* \to a_k^* + c_k^*$, are investigated. Here a_k and a_k^* are the annihilation and creation operators of the representations, the c_k are complex numbers (scalar operators), and the representations considered are the discrete representations described by Gårding and Wightman, and Schweber and Wightman, generalizations of the usual Fock representation. Necessary and sufficient conditions on the translations are obtained for the resultant representations to remain in the family of discrete representations, and the resultant translation groups are investigated for their relevant structure.

A FUNDAMENTAL problem in the quantum theory of fields is that of understanding the representations of the canonical commutation relations and the physical situations which they describe. This has been done in various ways by Gårding and Wightman,[1] by Segal,[2] and others.

An important matter in connection with the role various representations of the canonical commutation relations may play in the quantum theory of fields may be lumped under the general name of the implementation of the pseudo-canonical transformations by canonical transformations. (The term "pseudo-canonical" here being borrowed from Segal's paper cited above.) To explain this let us give an example investigated by Haag.[3] Suppose that in some physical calculation the transformation scheme for the creation and annihilation operators

$$a_k \to b_k = a_k \cosh\theta - a_k^* \sinh\theta,$$
$$a_k^* \to b_k^* = -a_k \sinh\theta + a_k^* \cosh\theta,$$
$$k = 1, 2, 3, \cdots,$$

θ real, forces itself upon us. The operators b_k and b_k^* defined by the right-hand side of these equations satisfy the same commutation relations as do the operators a_k and a_k^*, namely,

$$[b_k, b_l] = 0, \quad [b_k^*, b_l^*] = 0, \quad [b_k, b_l^*] = \delta_{kl},$$
$$k, l = 1, 2, 3, \cdots.$$

The question now arises whether a unitary operator U exists such that $b_k = Ua_kU^{-1}$, $b_k^* = Ua_k^*U^{-1}$, that is, whether this pseudo-canonical transformation (meaning that it preserves the commutation relations) is actually a canonical transformation on the operators. Otherwise said, we must know if we remain with the same representation of the commutation relations. In the above example it turns out that for the usual Fock no-particle representation of the annihilation and creation operators, a_k and a_k^*, the derived representation has a no-particle state if and only if $\theta = 0$. Such a consideration would rule out the use of the Fock representation in the theory. These pseudo-canonical transformations, when they appear, are usually intimately involved with the covariance of the physical laws in which they occur. As a very familiar example of this we have the relation satisfied by the Pauli matrices

$$U(R)\sigma^m U^{-1}(R) = \sum_{n=1}^{3} (R^{-1})^m{}_n \sigma^n, \qquad m = 1, 2, 3,$$

where $R \to U(R)$ is the two-dimensional unitary ray representation, $D^{\frac{1}{2}}$, of the proper three-dimensional rotations. This relation is commonly used to prove that the Pauli equation for the spinning nonrelativistic electron is covariant under rotations. From our point of view the identity states that the pseudo-canonical transformations on the solutions $b^m = \sigma^m$ of the commutation relations

$$[b_r, b_s] = 2ib_t, \quad r, s, t = 1, 2, 3 \text{ in cyclic order},$$

represented by the right-hand side of the identity may be implemented by the unitary equivalence appearing on the left-hand side.

Here we shall consider those pseudo-canonical transformations which may be written formally,

$$a_k \to a_k + c_k, \quad a_k^* \to a_k^* + c_k^*, \quad k = 1, 2, 3, \cdots,$$

for a particular class of representations of the canonical commutation relations

---

* The present paper is included in a thesis submitted to Princeton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Statements made here without proof are proved in the thesis.

† The results obtained in this paper were partly obtained while the author was a National Science Foundation Fellow at Princeton University.

[1] L. Gårding and A. S. Wightman, Proc. Natl. Acad. Sci. U. S. 40 (1954).

[2] I. E. Segal, Trans. Am. Math. Soc. 88 (1958).

[3] R. Haag, Kgl. Danske Videnskab. Selskab Mat.-Fys. Medd. 29, 12 (1955).

$$[a_r, a_s] = 0, \qquad [a_r^*, a_s^*] = 0, \qquad [a_r, a_s^*] = \delta_{rs}.$$

Here the c_k are merely some sequence of complex numbers (scalar operators). The representations of the commutation relations which we consider will be what Schweber and Wightman[4] term the discrete representations of the commutation relations. For the precise definition of these representations, the simplest generalization of the usual Fock space representation of the annihilation and creation operators, and the conditions under which they arise, we refer to Refs. 1 and 4.

We use the following notation: E will be the set of all sequences $(n_1, n_2, \cdots)$ of nonnegative integers; the elements of E will be the labels for the states of our system, indicating the occupation numbers for the single-particle states of our boson field in the usual interpretation. Using the construction in Ref. 2, we denote by $[m]$ the set of all sequences $t \in E$ whose entries differ from those of m in at most a finite number of places, by $H[m]$ the corresponding Hilbert space of all complex sequences $\{c_t : t \in [m]\}$ indexed by the elements of $[m]$ such that

$$\sum_{t\in[m]} |c_t|^2 < \infty,$$

and by $R[m]$ the corresponding descrete representation of the commutation relation. Thus if $0 = (0, 0, 0, \cdots)$, then $H[0]$ and $R[0]$ are the Hilbert space and the usual Fock representation of the commutation relations. If f_t, $t \in [m]$ is the orthonormal basis for $H[m]$ corresponding to the sequence $c_p = 0$ for $p \neq t$, $c_t = 1$, then $R[m]$ is described by the equations

$$a_k f_t = t_k^{\frac{1}{2}} f_{t-\delta_k},$$
$$a_k^* f_t = (t_k + 1)^{\frac{1}{2}} f_{t+\delta_k}, \qquad k = 1, 2, 3, \cdots,$$

where t_k is the kth entry for the sequence t and $t \pm \delta_k$ is t with its kth entry increased (decreased) by unity. To complete the definition of $R[m]$, we define the domain of a_k to be those elements $\Phi \in H[m]$ for which

$$\sum_{t\in[m]} (t_k + 1)\, |(f_t, \Phi)|^2 < \infty.$$

It may then be shown that the operator a_k^* defined above actually is the adjoint of a_k, that $a_k = (a_k^*)^*$, and so that both are closed operators.

We should note that $[m]$ is a denumerable set, so that $H[m]$ has the denumerably infinite dimension cardinality and is, therefore, isomorphic to l^2.

In order to obtain a mathematically precise formulation of the problem in which complications associated with the domains of the unrounded operators are avoided, we accept the commutation relations in the form proposed by Weyl, a form from which von Neumann[5] first proved uniqueness in the case of a system having a finite number of degrees of freedom. In terms of the unitary operators,

$$U(a) = \exp\left(i \sum_{k=1}^{n} a_k q_k\right),$$
$$V(b) = \exp\left(i \sum_{k=1}^{n} b_k d_k\right), \qquad (a_i, b_i \text{ real}),$$

where $p_k = 2^{-\frac{1}{2}}(a_k + a_k^*)$ and $iq_k = 2^{-\frac{1}{2}}(a_k - a_k^*)$, the Weyl relations for a system with n degrees of freedom are

$$U(a)U(b) = U(a + b), \qquad V(a)V(b) = V(a + b),$$
$$U(a)V(b) = \exp[-i(a \cdot b)]V(b)U(a),$$
$$(a \cdot b) = \sum_{k=1}^{n} a_k b_k. \qquad (1)$$

In Ref. 1 Gårding and Wightman replaced these relations (which, by von Neumann's theorem, have as their unique continuous solutions direct sums of copies of the usual Schrödinger representation of the p and q operators) by their simplest generalizations to systems with an infinite number of degrees of freedom. Here, for these larger systems, it has long been known that the uniqueness fails. Gårding and Wightman permitted the real sequences a_i and b_i to formally be infinite sequences, but with the restriction that $a_i = b_i = 0$ for almost all i. The solution of this generalization of the Weyl relations is then not unique, and we get the class of representations of the commutation relations, both discrete and continuous, described in Ref. 1. In particular the discrete representations mentioned above satisfy these generalizations of the Weyl relations.

To get a precise form of the translations it is convenient to condense these relations even further into the form given them by Segal and von Neumann. Set

$$W(c) = W(a + ib)$$
$$= \exp[\tfrac{1}{2}i(a \cdot b)] \exp[i(b \cdot q)] \exp[i(a \cdot p)];$$

then the Weyl relations are equivalent to

$$W(c)W(d) = \exp[\tfrac{1}{2} \operatorname{Im} (\bar{c} \cdot d)]W(c + d).$$

It is easy to see that the new set $\hat{W}$ defined by

$$\hat{W}(d) = \exp\left[\sum_{k=1}^{\infty} (-d_k \bar{c}_k + \bar{d}_k c_k)\right] W(d),$$

again satisfying (1), is irreducible if the set of W is irreducible and has as infinitesimal operators

$$\hat{a}_k = a_k + c_k.$$

[4] S. S. Schweber and A. S. Wightman, Phys. Rev. **98**, 812 (1955).

[5] J. von Neumann, Math. Ann. **104**, 570 (1931).

Thus, the problem is to find when the pseudo-canonical transformation $W \to \hat{W}$ is canonical.

For this purpose it is useful to extend the relations (1) as follows: Let c be an infinite sequence of complex numbers, and, considering the representation $R[m]$, let $n \in [m]$ be a representative sequence of the class $[m]$. Suppose that

$$\sum_{k=1}^{\infty} (n_k + 1) |c_k|^2 < \infty, \tag{2}$$

and let $W^N(c)$ be the sequence of operators $W(c^N)$, where c^N is the sequence c terminated after N terms, that is,

$$(c^N)_k = c_k, \qquad 1 \leq k \leq N,$$
$$(c^N)_k = 0, \qquad k > N.$$

Then the sequence $W^N(c)$ converges strongly to a unitary operator, which will be denoted $W(c)$. If c and d are two sequences satisfying the condition (2), then the relation (1) is again satisfied. This last generalization of Weyl relations is at once pertinent to our problem, for

$$W(c)W(d)W(-c) = \hat{W}(d).$$

Thus, we have implemented the operator translation $a_k \to a_k + c_k$, the sequence c_k satisfying the condition (2), as a unitary equivalence on the operators. The condition (2) will in fact be a natural requirement on the translations acting on the discrete representation $R[n]$ for the representation to be preserved.

To study the question of the effect of the general translation on the representation $R[m]$, we may use the following method, due to Wightman. We ask if the translation $a_k \to a_k + c_k$ can possibly send $R[m]$ into $R[l]$. To answer this we consider the possibility of the existence of a vector $\Phi \in H[m]$, the space on which the operators of the representation $R[m]$ act, such that

$$(a_k^* + c_k^*)(a_k + c_k)\Phi = l_k\Phi, \qquad k = 1, 2, 3, \cdots.$$

A formal solution of these equations can be obtained, that is, a sequence Φ_t, $t \in [m]$ of complex numbers which enter as the components of Φ relative to the basis f_t of $H[m]$. On requiring that Φ_t actually form on element of the Hilbert space, that is, that

$$\sum_{t \in [m]} |\Phi_t|^2 < \infty,$$

the possibilities of the existence of the element Φ are greatly reduced. Using this method the effects of the translation on the discrete representations may easily be described. The results are the following:

(a) The translation $a_k \to a_k + c_k$, $k = 1, 2, 3, \cdots$, transforms the representation $R[m]$ into itself, and is equivalent to a unitary similarity on the representation space, if and only if the condition (2) is satisfied. This, condition incidentally, is easily seen to be unchanged if the sequence $n \in [m]$ in (2) is replaced by any other $n' \in [m]$.

(b) In all other cases the resultant representation is not a discrete representation, but one of the continuous representations distinguished by Wightman and Gårding. Thus, no discrete representation is transformed into any other by these transformations on the operators.

The condition (2) which characterizes the family $T[m]$ of translations leaving $R[m]$ invariant is consistent with the natural algebraic operations: if $a = (a_1, a_2, \cdots)$ and $b = (b_1, b_2, \cdots)$ are contained in $T[m]$ then so are $a + b = (a_1 + b_1, a_2 + b_2, \cdots)$ and $-a = (-a_1, -a_2, \cdots)$, and since $0 = (0, 0, \cdots)$, the unit under this composition, is contained in $T[m]$, $T[m]$ forms an Abelian group, a subgroup of the additive group of all vectors in the Hilbert space l^2. With the usual Hilbert space topology the translation groups of the various discrete representations are dense subgroups of the enveloping group l^2. That this is so may be seen from the fact that each of the translation groups contains the elements $c = (c_1, c_2, \cdots)$ such that $c_i = 0$ for all except possibly a finite number of values of i, and these, of course, form a dense set in the space l^2. It is not difficult to show, in fact, that this particular dense set is precisely the intersection of *all* of our translation groups in l^2.

From the point of view of the present problem the topological structure mentioned above, that of l^2 induced on the various translation groups, turns out not to be the natural one. Before we introduce other topological structure, however, it is useful to discuss more closely the correspondence between the different representations and the various translation groups. This correspondence is certainly not one-to-one. Thus if m_k is any bounded sequence of integers then the condition (2) on the complex sequence c_k means no more and no less than the condition (2) with $n_k = 0$ for all k. Therefore, all the bounded representations $R[m]$, that is, those representations for which some (and hence every) sequence $(m_1, m_2, \cdots) \in [m]$ is bounded, have the same translation group, namely all of l^2.

The situation is very easily described precisely. In the set E of all sequences of non-negative integers, we introduce the relation we call asymptotic equiv-

alence defined as follows: $r_k \sim s_k$ if there are numbers A and B such that

$$0 < A \leq (r_k + 1)/(s_k + 1) \leq B < \infty, \qquad (3)$$

for all k. This is easily seen to be an equivalence relation on E, dividing E into disjoint sets of asymptotically equivalent sequences; it is also coarser than the relation used in forming the classes $[\cdot]$ upon which the construction of the representations $R[m]$ is based, so that if $||m||$ is the sequence class containing m then $[m] \subset ||m||$, and $||m||$ may be written as the union of the classes $[t]$ contained in it.

The precise connection between the representations and the translation groups is then given by the following statement: Two representations, $R[m]$ and $R[n]$, have the same translation group, if and only if, $[m]$ and $[n]$ are contained in the same asymptotic equivalence class that is, if $||m|| = ||n||$. We denote by $T||m||$ the translation group corresponding to the family of representations $\{R[t] : t \in E, t \in ||m||\}$. Thus $T||0||$ is the translation group for the no-particle representation and also for all the bounded discrete representations. In constructing the relevant topology on these groups we will deal with $T||m||$, the single translation group corresponding to a complete family of representations, rather than with a group $T[m]$ corresponding to a single representation. Indeed there is nothing else that we can do, for the force of the above assertion is that, beginning with a translation group $T[m]$ we can only identify representations $R[m']$ with $||m'|| = ||m||$ if we ask for the representation which gave rise to it.

We now construct a topology for the group $T||m||$. To do this we select some sequence $r \in ||m||$ of integers and define the function

$$p(c) = \left[\sum_{k=1}^{\infty} (r_k + 1) \, |c_k|^2 \right]^{\frac{1}{2}}$$

on the group elements; by the definition of $T||m||$ and what we have already said, it follows that this is well defined and finite for any $c \in T||m||$. Moreover, it is easily proved in the usual way that $q(c, d) = p(c - d)$ is a metric on $T||m||$, and that, using the topology of this metric, $T||m||$ becomes a topological group and a complete metric space. By turning $T||m||$ into a Hilbert space in the natural way and then using the theorem that a Hilbert space is uniquely determined by its dimension cardinality, or by explicitly exhibiting the mapping F of $T||m||$ onto l^2 which establishes the result, we can prove that the metric topological group $T||m||$ is isomorphic, and, in fact, isometric, to the additive group of vectors of the Hilbert space l^2. Such an isometric homomorphism F is, of course, the mapping which sends $c = (c_1, c_2, \cdots) \in T||m||$ into the element $x = (x_1, x_2, \cdots) \in l^2$, where $x_k = (m_k+1)^{\frac{1}{2}}c_k$. By transitivity, then, every metric topological group $T||m||$ is isometrically isomorphic as a topological group to every other.

Having constructed these topologies on the various translation groups, let us rid ourselves of the arbitrariness involved in choosing the sequence $r \in ||m||$ used in constructing the metric on $T||m||$. To do this we choose some other sequence, $s \in ||m||$, and construct the metric $q'(c, d)$ on $T||m||$ using s just as $q(c, d)$ constructed above uses r. Then (3) holds for some A and B because r and s are asymptotically equivalent, and so we have the inequalities

$$A^{\frac{1}{2}}q'(c, d) \leq q(c, d) \leq B^{\frac{1}{2}}q'(c, d)$$

for any $c, d \in T||m||$. Thus, the metrics q and q' on $T||m||$ are equivalent, and so the topologies they determine are identical. Henceforth, we will regard $T||m||$ as being a topological group with this topology, a topology metrizable by choosing any one of the metrics q corresponding to any sequence of integers $r \in ||m||$.

The naturalness of these topologies on the translation groups is then the implication of the following assertion: The mapping $c \to W(c)$ of $T[m]$ into the group of unitary operators on the Hilbert space $H[m]$ of the representation $R[m]$ is strongly continuous, and so the mappings are strongly continuous unitary ray representations of the translation groups. The multiplication rule for the unitary operators $W(c)$ which shows that they are ray representations is, of course, the Weyl relations in their extended form mentioned earlier.

ACKNOWLEDGMENT

I thank Professor A. S. Wightman, who suggested the problem considered here and whose aid was essential in its solution. I also thank Professor V. Bargmann for his most valuable and useful criticisms, and J. McKenna and J. R. Klauder for communicating to me before publication of their work the results they arrived at relating to the subject of this paper.

Space–Time Theory of Elementary Particles. I. Born Reciprocity Principle

ERNEST E. H. SHIN
National Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts*
(Received 19 April 1965)

The use of the Born reciprocity principle as a postulate applied to "internal" Minkowski space leads to a theory of elementary particles based on a purely "space–time" point of view. The reciprocity transformations of Born, $(x_\mu, p_\mu) \rightarrow (\pm p_\mu, \mp x_\mu)$, form a discrete subgroup of the x–p rotation group $SU(2)$, which we associate with "isotopic spin" symmetry. The full dynamic symmetry of the eight-dimensional phase space with Minkowski metric is a 36-parameter group which contains the x–p rotation group $SU(2)$ and the homogeneous Lorentz group $\mathfrak{L}$ as disjoint subgroups. This gives the symmetry of "subquantic" matter which has oscillator–rotator properties. Some physical aspects of the generators and possible forms of mass formulas for both integer and odd-half integer spins are discussed. Finally, it is suggested that the nonrelativistic counterpart of the present formalism may explain the symmetry of atomic nuclei.

> "And when we have built an altar to the Invisible Light, we may set thereon
> the little lights for which our bodily vision is made."
>
> from "The Rock" by T. S. Eliot

I. INTRODUCTION

NEARLY a decade and a half ago, Born and his collaborators[1] published a series of papers in which they reported attempts to explain elementary particles using the Born reciprocity principle (BRP)[2] as a postulate. The BRP states that "the laws of nature are symmetrical with regard to space–time and momentum–energy." As several examples which are suggestive of the reciprocity property, Born cites the Hamilton principle

$$\dot{x}^k = \partial H/\partial p_k, \qquad \dot{p}_k = -\partial H/\partial x^k \qquad (k = 1, 2, 3),$$

and, in the operator formalism of quantum mechanics, the canonical commutation rules

$$x_\mu p^\nu - p^\nu x_\mu = i g^\nu_\mu \qquad (\mu, \nu = 1, 2, 3, 4),$$

and the components of angular momentum,

$$m_{\mu\nu} = (x_\mu p_\nu - x_\nu p_\mu) \text{ for } (\mu \neq \nu) \qquad (\mu, \nu = 1, 2, 3).$$

Although these examples are strongly suggestive, they merely amount to several very special cases. In general, physics seems to exhibit no obvious symmetry of the type suggested by Born. On the other hand, an interesting consequence of BRP is its implication of the existence of unit distance a and unit momentum b such that $ab = 1 (\hbar = c = 1)$. These constants are implied quite naturally by the fact that BRP suggests the existence of a reciprocal operator $(x_\mu x^\mu + p_\mu p^\mu)$ as a physically significant operator. As is well known, the relativistic equations of motion of a particle in atomic and nuclear dimensions generally do not exhibit reciprocal symmetry, e.g., Klein–Gordon equation. One possible explanation in keeping with the spirit of BRP would be that the local field theory in its usual form represents the "point-particle limit," $a \rightarrow 0$, in which case the reciprocal symmetry is broken. For example, $(x_\mu x^\mu + p_\mu p^\mu) \rightarrow p_\mu p^\mu$ in this limit. In a way, this might be viewed as a more microscopic analog of the limit $\hbar \rightarrow 0$ which represents the transition from quantum mechanics to classical mechanics of atoms. If we accept this picture, it follows then that the unit distance a must be of the order of 1 F or less thus representing the subnuclear dimensions and that BRP is most strongly realized in the most fundamental realm of nature which is found in the subnuclear dimensions. This, in turn, suggests that BRP might constitute a fundamental principle underlying the elementary particle phenomena.

It was probably this notion which led Born *et al.*[1] to their "reciprocity theory of elementary particles." Several years ago,[3] the same notion led us to propose a theory of quantization of elementary particle masses using BRP as a postulate applied to an "internal" Minkowski space such as the one proposed by Yukawa[4] in his bilocal theory of fields. Now, what do we mean by "internal" and "external"

* Supported by the U. S. Air Force Office of Scientific Research.

[1] M. Born, Nature **163**, 207 (1949); M. Born and H. S. Green, Proc. Roy. Soc. Edinburgh **A92**, 470 (1949); Nature **164**, 281 (1949); H. S. Green, Nature **163**, 208 (1949); M. Born, Rev. Mod. Phys. **21**, 463 (1949).

[2] See Ref. 1 for Born's earlier papers on his reciprocity principle.

[3] E. E. H. Shin, Phys. Rev. Letters **10**, 196 (1963).

[4] H. Yukawa, Phys. Rev. **76**, 300, 1731 (1949); **77**, 219, 849 (1950).

spaces? The "external" Minkowski space is the space–time in which the motion of a point (= center of mass) particle is mapped in the usual manner of local field theory. The "internal" Minkowski space represents the subnuclear space–time similar to the Bohm–Vigier "subquantic" medium[5] of matter in which nature manifests itself in its most fundamental form and whose space–time and momentum–energy variables supposedly describe the dynamics responsible for various elementary particle phenomena which we commonly observe from an "external" point of view. For the "internal" system of subnuclear dimensions, it is natural to expect to find the system a strongly quantized one: viz., for $a \lesssim 1$ F, the quantized matter would be characterized by an energy of the order of $b = a^{-1} \gtrsim 100$ Mev. Further, the "internal" dynamics (at least, in part) is prescribed by the postulate of BRP since physically significant operators are to be constructed from (x_μ), (p_μ) in reciprocity-invariant forms or, more precisely, to be invariant under the transformation, $(x_\mu, p_\mu) \rightarrow (\pm p_\mu, \mp x_\mu)$.

Mathematically, a reciprocal operator is one which is invariant under the reciprocity transformations of Born,[1]

$$(X'_\mu)_\alpha = \sum_{\beta=1}^{2} R_{\alpha\beta}(X_\mu)_\beta \qquad (\alpha, \beta = 1.2)$$

$$X_\mu = \begin{pmatrix} x_\mu \\ p_\mu \end{pmatrix} \tag{1}$$

$$(X_\mu)_1 = x_\mu, \qquad (X_\mu)_2 = p_\mu.$$

According to Rayski,[6] the (2×2) matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2}$$

form the common divisors for all such unimodular, unitary transformations. In all, there exist four such transformations, namely[7]

$$R_1 = I; \quad R_2 = \mathfrak{R}; \quad R_3 = \mathfrak{C}; \quad R_4 = \mathfrak{R} \times \mathfrak{C}, \tag{3}$$

where

$$I_{\alpha\beta} = \delta_{\alpha\beta} = -\mathfrak{R}_{\alpha\beta} \qquad \mathfrak{C} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}. \tag{4}$$

The simplest "nontrivial" operator which has the reciprocal symmetry and is Lorentz-invariant is $(x_\mu x^\mu + p_\mu p^\mu)$. In our previous letter, the mass operator was taken in this form, thus resulting in a linear spectrum of masses.

Clearly, the reciprocity transformations defined above are special cases of the continuous x–p rotations of $SU(2)$ defined by

$$X'_{\alpha,\mu} = X_{\alpha,\mu} + i \sum_{i=0}^{3} \sum_{\beta=1}^{2} \delta c_i (\tau_i)_{\alpha\beta} X_{\beta,\mu} \quad (\alpha, \beta = 1, 2), \tag{5}$$

where (δc_i) are arbitrary infinitesimal parameters, and

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \qquad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix};$$
$$\tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \qquad \tau_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6}$$

More precisely, they form a discrete subgroup of the complete x–p rotation group $SU(2)$: i.e., any (2×2) matrix can be represented as a linear combination of (τ_i) matrices. Note that the x–p rotations of Eq. (5) leave the four-vector indices $\mu, \nu, \cdots = 1, 2, 3, 4$ unaffected. It is commonly believed that the electromagnetic interaction breaks the isotopic spin symmetry of elementary particles, e.g., it breaks the charge independence of Yukawa interaction. In the present case, "electromagnetic" interaction may destroy the reciprocal symmetry and, therefore, the $SU(2)$ symmetry of x–p rotations if there exists "electromagnetic" vector field A_μ which depends on the "internal" variables and which would then enter into the reciprocal operator $(p_\mu p^\mu + x_\mu x^\mu)$ asymmetrically. This would in turn suggest that we might associate the isotopic spin rotations with the x–p rotations of Eq. (5). (In the usual picture of field theory, electromagnetic vector field depends only on the center-of-mass variables and the electromagnetic gauge, therefore, commutes with the present "internal" operators. In view of this, it has been brought to the author's attention that, if a vector field A_μ of the type mentioned above exists, it would probably be a "new nonelectromagnetic" field. This is certainly one of the many important questions which we wish to investigate in further detail in future publications.)

Separate from the x–p rotations of Eq. (5), there is, of course, the four-dimensional Lorentz group. The full dynamic symmetric of the eight-dimensional phase space with Minkowski metric would then be given by a large group $\mathcal{G}$ which contains the x–p rotation group $SU(2)$ and the (homogeneous) Lorentz group as subgroups. The postulate of BRP implies then that the generators of the reciprocity-invariant content of $\mathcal{G}$ represent physically meaningful quantities. More importantly, we propose that

[5] L. de Broglie, *Introduction to Vigier Theory of Elementary Particles* (Elsevier Publishing Company, New York, 1963).
[6] J. Rayski, Nuovo Cimento **2**, 255 (1955).
[7] E. E. H. Shin, J. Math. Phys. **6**, 1307 (1965).

the group $\mathcal{G}$ contains the dynamic symmetry of elementary particles.

II. RECIPROCITY AND DYNAMIC SYMMETRY

All operators of the eight-dimensional phase space which are bilinear in (x_μ) and (p_μ) are given by

$$\tilde{X}_\mu \tau_j X_\nu = \sum_{\alpha,\beta=1}^{2} \tilde{X}_{\mu,\alpha}(\tau_j)_{\alpha\beta} X_{\nu,\beta} \quad \left(\begin{matrix}\mu, \nu = 1, 2, 3, 4 \\ j = 0, 1, 2, 3\end{matrix}\right), \tag{7}$$

Of these, there are 36 "nontrivial" independent operators, which are given by

$$\begin{aligned} S^{(1)}_{\mu\nu} &= S^{(1)}_{\nu\mu} = \tfrac{1}{2}\{\tilde{X}_\mu \tau_3 X_\nu + \tilde{X}_\nu \tau_3 X_\mu\}, \\ S^{(1)}_{\mu\nu} &= S^{(2)}_{\nu\mu} = \tfrac{1}{2}\{\tilde{X}_\mu \tau_1 X_\nu + \tilde{X}_\nu \tau_1 X_\mu\}, \\ S^{(3)}_{\mu\nu} &= S^{(3)}_{\nu\mu} = \tfrac{1}{2}\{\tilde{X}_\mu \tau_0 X_\nu + \tilde{X}_\nu \tau_0 X_\mu\}, \\ m_{\mu\nu} &= \underset{(\mu\neq\nu)}{-m_{\nu\mu}} = +\tfrac{1}{2} i\{\tilde{X}_\mu \tau_2 X_\nu - \tilde{X}_\nu \tau_2 X_\mu\}. \end{aligned} \tag{8}$$

In addition to these 36 operators, there exist 10 more symmetrical operators which are "trivial" since they are the commutators,

$$g_{\mu\nu} = \tfrac{1}{2}\{\tilde{X}_\mu \tau_2 X_\nu + \tilde{X}_\nu \tau_2 X_\mu\} = -i[x_\mu, p_\nu], \tag{9}$$

which are also the flat-space metric tensors of Minkowski space having the usual properties

$$\begin{aligned} g_{\mu\nu} &= g_{\nu\mu}, \qquad g^{\mu\nu} = g^{\nu\mu}, \\ g_{k4} &= -g^4_k, \\ g_{kl} &= g^l_k \qquad (k, l = 1, 2, 3), \\ x^\mu &= g^{\mu\nu} x_\nu, \qquad x_\mu = g_{\mu\nu} x^\nu, \end{aligned}$$

and so forth. Clearly, not all operators of Eq. (8) are reciprocity-invariant. Aside from the commutators of Eq. (9), there are 16 independent operators which are reciprocity-invariant, namely

$$(S^{(3)}_{\mu\nu}), (m_{\mu\nu}) \qquad (\mu \neq \nu). \tag{10}$$

Although the remaining 20 operators $(S^{(1)}_{\mu\nu})$, $(S^{(2)}_{\mu\nu})$ are not reciprocity-invariant, they are nevertheless useful for construction of reciprocity-invariant operators to higher orders in (x_μ), (p_μ). For example,

$$(S^{(1)}_{\mu\nu} S^{(1)\mu\nu} + S^{(2)}_{\mu\nu} S^{(2)\mu\nu})$$

can be shown to be reciprocity-invariant. It is, therefore, necessary to use all 36 operators of Eq. (8) to obtain the full dynamic symmetry of the eight-dimensional phase space.

Now, it may be shown that these 36 operators together generate a 36-parameter group $\mathcal{G}$. To prove this, it will suffice to show that they generate a closed commutation algebra. Let us define

$$(\mathbf{M}_{\mu\nu}) = (M^{(1)}_{\mu\nu}, M^{(2)}_{\mu\nu}, M^{(3)}_{\mu\nu})$$

as follows:

$$\begin{aligned} M^{(1)}_{\mu\nu} &= -iS^{(1)}_{\mu\nu} = M^{(1)}_{\nu\mu}, \\ M^{(2)}_{\mu\nu} &= +iS^{(2)}_{\mu\nu} = M^{(2)}_{\nu\mu}, \\ M^{(3)}_{\mu\nu} &= S^{(3)}_{\mu\nu} = M^{(3)}_{\nu\mu}. \end{aligned} \tag{11}$$

These, together with antisymmetrical tensors $(m_{\mu\nu}) = (-m_{\nu\mu})(\mu \neq \nu)$, satisfy the following:

$$\begin{aligned} & i^{-1}[M^{(k)}_{\mu\nu}, M^{(l)}_{\lambda\kappa}] \\ &= \epsilon^{klm}\{g_{\nu\lambda} M^{(m)}_{\mu\kappa} + g_{\mu\lambda} M^{(m)}_{\nu\kappa} + g_{\nu\kappa} M^{(m)}_{\lambda\mu} + g_{\mu\kappa} M^{(m)}_{\lambda\nu}\}, \\ & i^{-1}[M^{(k)}_{\mu\nu}, m_{\lambda\kappa}] = \{g_{\nu\kappa} M^{(k)}_{\mu\lambda} + g_{\mu\kappa} M^{(k)}_{\nu\lambda} - g_{\mu\lambda} M^{(k)}_{\mu\kappa}\}, \\ & i^{-1}[m_{\mu\nu}, m_{\lambda\kappa}] \\ &= \{-g_{\mu\kappa} m_{\nu\lambda} + g_{\nu\kappa} m_{\mu\lambda} - g_{\nu\lambda} m_{\mu\kappa} + g_{\mu\lambda} m_{\nu\kappa}\} \\ &\qquad \left(\begin{matrix}k, l, m = 1, 2, 3 \\ \mu, \nu, \lambda, \kappa = 1, 2, 3, 4\end{matrix}\right), \end{aligned} \tag{12}$$

where (ϵ^{klm}) are the structure constants

$$\epsilon^{klm} = \begin{cases} +1 : k, l, m \text{ in cyclic order} \\ -1 : \text{otherwise } (k \neq l \neq m). \end{cases}$$

We may consider the 36-parameter group $\mathcal{G}$ as the eight-dimensional phase-space analog of the homogeneous Lorentz group $\mathcal{L}$ of Minkowski four-space. In fact, $\mathcal{G}$ contains the homogeneous Lorentz group $\mathcal{L}$ generated by $(m_{\mu\nu})$ as a subgroup. Now, consider the Lorentz-invariant operators

$$\begin{aligned} T_k &= \tfrac{1}{4} \operatorname{Tr}_{(4)}(M^{(k)} G) \\ &= \tfrac{1}{4} g^{\mu\nu} M^{(k)}_{\mu\nu} \qquad (k = 1, 2, 3), \end{aligned} \tag{13}$$

where $G = (-1, -1, -1, +1)$ is the Minkowski metric. According to Eq. (12), these operators satisfy the relations

$$\begin{aligned} [T_k, T_l] &= i\epsilon^{klm} T_m, \\ [T_k, m_{\lambda\kappa}] &= 0, \end{aligned} \tag{14}$$

and, therefore, generate a three-rotation group which is disjoint from $\mathcal{L}$ and isomorphic with $SU(2)$. Thus, we have

$$SU(2) \otimes \mathcal{L} \subset \mathcal{G} \tag{15}$$

as a reduced representation of $\mathcal{G}$.

We may also show that each of the three combinations of 16 operators

$$(M^{(k)}_{\mu\nu}, m_{\mu\nu}) \qquad (k = 1, 2, 3)$$

forms a 16-parameter group $\mathcal{L}'$ which is isomorphic with the 16-parameter Lie group.[8] [In particular,

[8] A. O. Barut, Phys. Rev. **135**, B839 (1964).

$(M^{(3)}_{\mu\nu}, m_{\mu\nu})$ generate a 16-parameter group $\mathfrak{L}'$ which is reciprocity-invariant.] For example, let us use $(M^{(3)}_{\mu\nu}, m_{\mu\nu})$ and define

$$A_{\mu\nu} = \tfrac{1}{2}\{M^{(3)}_{\mu\nu} - i(M_{\mu\nu} + ig_{\mu\nu})\} \quad (\mu, \nu = 1, 2, 3, 4). \tag{16}$$

According to Eq. (12), these operators satisfy

$$[A_{\mu\nu}, A_{\lambda\kappa}] = \{g_{\nu\lambda}A_{\mu\kappa} - g_{\mu\kappa}A_{\nu\lambda}\}, \tag{17}$$

which is precisely the commutation relation of the 16-parameter Lie group, e.g., the 16-parameter Lie group of Dirac γ-matrices.[8] The invariant or Casimir operator of this group is given by

$$\begin{aligned} F &= \tfrac{1}{2}\,\mathrm{Tr}_{(4)}(AG) \\ &= \tfrac{1}{2}g^{\mu\nu}A_{\mu\nu} = (T_3 + \lambda), \end{aligned} \tag{18}$$

which commutes with all operators of $\mathfrak{L}'$, where λ is an integral multiple of the identity operator. A second Casimir operator is given by

$$\begin{aligned} F^2 &= \tfrac{1}{4}\,\mathrm{Tr}_{(4)}(AGAG) \\ &= \tfrac{1}{4}g^{\mu\nu}g^{\lambda\kappa}A_{\mu\lambda}A_{\kappa\nu} \\ &= \tfrac{1}{16}\{M^{(3)}_{\mu\nu}M^{(3)\mu\nu} - m_{\mu\nu}m^{\nu\mu} + 4\lambda^2\}. \end{aligned} \tag{19}$$

The finite-dimensional representations in $\mathfrak{L}'$ of physical interest can be classified using the so-called "unitary trick" of Weyl. In the reduction $\mathfrak{L}' \to T^3 \otimes \mathfrak{L}$, the reduced representations of physical interest are given by $T^3 \otimes SU(2)_L$, where $SU(2)_L$ is the spatial three-rotation group. What has been discussed above applies also to other $\mathfrak{L}'$-contents of $\mathfrak{G}$. It must be understood, however, that $\mathfrak{L}'$ is not a *disjoint* subgroup of $\mathfrak{G}$ and consequently that the invariants (F, F^2) of $\mathfrak{L}'$ are not invariants of the whole group $\mathfrak{G}$. Obviously, it is necessary to work with the whole group $\mathfrak{G}$ to appreciate the full dynamic symmetry. A convenient way of representing the generators of group $\mathfrak{G}$ is by means of the following 8×8 representations;

$$\begin{aligned} \tilde{M}_{\mu\nu} &= \{\mathbf{M}_{\mu\nu}\cdot\boldsymbol{\tau} - i(m_{\mu\nu} + ig_{\mu\nu})\tau_0\} \\ &= \begin{bmatrix} M^{(3)}_{\mu\nu} - i(m_{\mu\nu} + ig_{\mu\nu}) & M^{(1)}_{\mu\nu} - iM^{(2)}_{\mu\nu} \\ M^{(1)}_{\mu\nu} + iM^{(2)}_{\mu\nu} & -M^{(3)}_{\mu\nu} - i(m_{\mu\nu} + ig_{\mu\nu}) \end{bmatrix}, \end{aligned} \tag{20}$$

where $(M_{\mu\nu}) = (M^{(1)}_{\mu\nu}, M^{(2)}_{\mu\nu}, M^{(3)}_{\mu\nu})$ are the symmetrical tensors defined in Eq. (11), and

$$\begin{aligned} m_{\mu\nu} + ig_{\mu\nu} &= -m_{\nu\mu} + ig_{\nu\mu} \\ &= (x_\mu p_\nu - p_\mu x_\nu). \end{aligned} \tag{21}$$

Note that the tensors belonging to the diagonal quadrants of Eq. (20) generate the $\mathfrak{L}'$-content of $\mathfrak{G}$. Let "$\mathrm{Tr}_{(2)}$" and "$\mathrm{Tr}_{(8)}$" represent the diagonal sums over the two-dimensional and the entire eight-dimensional spaces, respectively. We then have

$$\begin{aligned} M^{(k)}_{\mu\nu} &= \tfrac{1}{2}\,\mathrm{Tr}_{(2)}(\tilde{M}_{\mu\nu}\tau_k), \\ m_{\mu\nu} + ig_{\mu\nu} &= \tfrac{1}{2}i\,\mathrm{Tr}_{(2)}(\tilde{M}_{\mu\nu}\tau_0), \\ T_k &= \tfrac{1}{8}\,\mathrm{Tr}_{(8)}(\tilde{M}G\tau_k). \end{aligned} \tag{22}$$

Commutation relations among the 36 independent elements of $\tilde{M}$ are as given in Eqs. (12) and (17). For an arbitrary tensor $\mathbf{O}$ constructed from $\tilde{M}$, τ_k projects out $\mathbf{O}^{(k)}$ which transforms like $M^{(k)}$ with respect to the x–p rotations of $SU(2)$. For example, let

$$\mathbf{O}^{\lambda\kappa\cdots}_{\mu\nu\cdots} = (\tilde{M}\tilde{M}\cdots\tilde{M})^{\lambda\kappa\cdots}_{\mu\nu\cdots}$$

be an arbitrary tensor. We then have

$$\mathbf{O}^{(k)\lambda\kappa\cdots}_{\mu\nu\cdots} = \tfrac{1}{2}\,\mathrm{Tr}_{(2)}(O^{\lambda\kappa\cdots}_{\mu\nu\cdots}\tau_k). \tag{23}$$

A particular case of this is an operator $\mathbf{O}^{(k)}$ which is invariant with respect to $\mathfrak{L}$ but has the properties of $M^{(k)}$ with respect to the x–p rotations of $SU(2)$. Such an operator which is constructed from $\tilde{M}$ is defined as

$$\mathbf{O}^{(k)} = \tfrac{1}{2}\,\mathrm{Tr}_{(8)}(\tilde{M}G\tilde{M}G\cdots\tilde{M}G\tau_k). \tag{24}$$

Similarly, the invariant operators of $\mathfrak{G}$ are defined by

$$\tfrac{1}{8}\,\mathrm{Tr}_{(8)}(\tilde{M}G\tilde{M}G\cdots\tilde{M}G) \tag{25}$$

to all orders in $\tilde{M}$. For example, we have

$$\begin{aligned} F' &= \tfrac{1}{8}\,\mathrm{Tr}_{(8)}(\tilde{M}G) = N, \\ F'' &= \tfrac{1}{8}\,\mathrm{Tr}_{(8)}(\tilde{M}G\tilde{M}G) \\ &= \tfrac{1}{4}\{\mathbf{M}_{\mu\nu}\cdot\mathbf{M}^{\nu\mu} - m_{\mu\nu}m^{\nu\mu} + 4N^2\}, \end{aligned} \tag{26}$$

and similarly for others, where $N = 1, 2, 3, \cdots$ is a parameter which, physically, represents the number of four-oscillators which we build into the elementary system, i.e., for $i = 1, 2, 3, \cdots, N$,

$$\frac{1}{4i}\sum_{i=1}^{N}(x(i)_\mu p^\mu(i) - p^\mu(i)x(i)_\mu) = N. \tag{27}$$

If the mass is to be an invariant with respect to all transformations of $\mathfrak{G}$, it would then be a function of the Casimir operators. One possible (and the simplest) form of such a mass formula seems to be

$$\begin{aligned} M &= b_0 + b_1F' + b_2F'' \\ &= \{b_0 + b_1N + b_2(\mathbf{M}_{\mu\nu}\cdot\mathbf{M}^{\nu\mu} - m_{\mu\nu}m^{\nu\mu} + 4N^2)\}, \end{aligned} \tag{28}$$

where b_0, b_1, b_2 are constant parameters which can only be determined phenomenologically. Physically, such a mass formula evidently represents a system which is made up of N four-oscillators each with

mass b_1, the b_2 term representing a deviation from the simple linear sum of N masses.

Previously,[3] we proposed a linear mass formula which was basically in the form of the reciprocity- and $\mathfrak{L}$-invariant operator,

$$\text{const} \times (x_\mu x^\mu + p_\mu p^\mu).$$

Within the context of the present formalism, this is equivalent to taking the mass to have the properties of

$$T_3 = \tfrac{1}{8} \operatorname{Tr}_{(8)}(\tilde{M} G \tau_3)$$

with respect to the transformations of $\mathcal{G}$. We may consider this as a symmetry-breaking term and combine it with the $\mathcal{G}$-invariant terms, and write

$$M = M_0 + c_1 T_3 \tag{29}$$

in which M_0 and c_1 are $\mathcal{G}$-invariant parameters; e.g., for M_0, we may use Eq. (28), and take

$$M = b_0 + b_1 N + c_1 T_3 + b_2(\mathbf{M}_{\mu\nu}\cdot\mathbf{M}^{\nu\mu} - m_{\mu\nu}m^{\nu\mu} + 4N^2) \tag{30}$$

as the symmetry-breaking mass formula. To this, we may add other higher order terms which also have the properties of T_3, e.g.,

$$\begin{aligned} &\tfrac{1}{8} \operatorname{Tr}_{(8)}(\tilde{M} G \tilde{M} G \tau_3), \\ &\tfrac{1}{8} \operatorname{Tr}_{(8)}(\tilde{M} G \tilde{M} G \tilde{M} G \tau_3), \end{aligned} \tag{31}$$

and so forth, which may be interpreted as higher "perturbation" terms. [It may be pointed out that $\operatorname{Tr}_{(8)}(\tilde{M} G \tilde{M} G \tau_3)$ reduces to an expression linear in T_3 and hence does not add anything new to Eq. (30).]

In order to express the mass formulas in a more meaningful form, we must properly reduce the invariant products of tensors using physically meaningful quantities. First, let us consider the operator $m_{\mu\nu}m^{\nu\mu}$. Let $Y_K(\theta, \varphi, \omega)$ be the four-harmonics[1] for quantum number $K = 0, 1, 2, 3, \cdots$, given as a function of the polar angles, $(\theta, \varphi, \omega)$. As shown in Ref. 1,

$$Y_K = (p^2)^{\frac{1}{2}K+1} U_K \tag{32}$$

where U_K is the solution of

$$\partial^2 U_K / \partial p^\mu \partial p_\mu = 0 \tag{33}$$

and is given by

$$U_K = \frac{\partial^K}{\partial p_1^{K_1}\, \partial p_2^{K_2}\, \partial p_3^{K_3}\, \partial p_4^{K_4}} \left(\frac{1}{p^2}\right), \tag{34}$$

where $K = K_1 + K_2 + K_3 + K_4 = 0, 1, 2, 3, \cdots$. Operating on Y_K, we have

$$\begin{aligned} \tfrac{1}{8} m_{\mu\nu}m^{\mu\nu} Y_K &= \tfrac{1}{4} K(K+2) Y_K \\ &= W(W+1) Y_K, \end{aligned} \tag{35}$$

where we have defined

$$W = \tfrac{1}{2}K = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots$$

or

$$= -\tfrac{1}{2}K - 1 = -1, -\tfrac{3}{2}, -2, \cdots .$$

Secondly, reduction of $(M^{(k)}_{\mu\nu} M^{(k)\nu\mu})$ gives

$$M^{(k)}_{\mu\nu} M^{(k)\nu\mu} = \{(M^{(k)})^2 + m_{\mu\nu}m^{\nu\mu} + i[x_\mu, p^\mu]\},$$

so that

$$\begin{aligned} \mathbf{M}_{\mu\nu}\cdot\mathbf{M}^{\nu\mu} &= \{(\mathbf{M}\cdot\mathbf{M}) + 3m_{\mu\nu}m^{\mu\nu} + 3i[x_\mu, p^\mu]\} \\ &= \{16\mathbf{T}^2 + 3m_{\mu\nu}m^{\nu\mu} + 3i[x_\mu, p^\mu]\}, \end{aligned} \tag{36}$$

where $T^2 = (T_3)^2 + (T_1)^2 + (T_2)^2$. Using Eqs. (35) and (36), we may rewrite Eq. (30), for example, in the form

$$M = \{b_0 + b_1' N + c_1 T_3 + b_2'[W(W+1) - \mathbf{T}^2 - \tfrac{1}{4}N^2]\}. \tag{37}$$

Although this seems to suggest some resemblance to the Gell-Mann–Okubo mass formula, any such identification is purely arbitrary, in fact, they need not be the same.

Finally, it may be pointed out that the integer parameter N is not necessarily a positive-definite quantity since, as Finkelstein *et al.*[9] pointed out, the "ordinary" quantum mechanics is undetermined in the signature of the complex number $(-1)^{\frac{1}{2}}$ which appears in the commutation rules. That is, the "superselection" operator η is defined by

$$\begin{aligned} [x_\mu, p_\nu] &= \eta g_{\mu\nu}, \\ \eta^2 &= (-1), \end{aligned} \tag{38}$$

where the signature of η is undetermined. Associated with the two possible signatures of $\eta(= \pm i)$, we may define $N_\pm = 1, 2, 3, \cdots$ such that

$$N = (N_+ - N_-) = 0, \pm 1, \pm 2, \pm 3, \cdots , \tag{39}$$

which now includes the value zero and negative integers. This possibility[10] can be contained within

[9] D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962).

[10] A generalization of the present formalism to a "complex Minkowski space" described by canonical operators (z_μ, ω_μ) and (z_μ^*, ω_μ^*) will be the topic for Part II of this series. Such a theory contains the dual signatures in a natural manner since $[z_\mu, \omega_\nu] = -[z_\mu^*, \omega_\nu^*]$. In this case, the full dynamic symmetry of the 16-dimensional phase space is given by a 136-parameter group $\mathcal{G}'$ which contains two 36-parameter groups as disjoint subgroups. In particular, the ten $\mathfrak{L}$-invariant generators of $\mathcal{G}'$ form a 10-parameter group which is the five-dimensional real "Lorentz" group isomorphic with the 10-parameter Lie group discussed by Barut.[8]

the present formalism without going outside the group $\mathcal{G}$; i.e., $\mathcal{G}$ contains all possible representations for positive and negative N. According to a footnote in the paper by Finkelstein *et al.*,[9] Yang apparently feels that the transformation $a\eta a^{-1} = -\eta$ might be associated with the interchange of the two nucleons (p, n). In this connection, it is interesting to observe that association of the two signatures of η with (p, Ξ), for example, automatically imparts the meaning of "hypercharge" to N and that this, in turn, presents a basis for comparison between Eq. (37) with the Gell-Mann–Okubo type mass formula. A further possibility is to associate the dual signatures with the matter–antimatter dualism, the underlying idea being that some of the four-oscillators in the "internal" system are matterlike and others are anti-matterlike somewhat in the sense of Feynman "vacuum polarization." A further possibility is to interpret N literally as the "particle number" such that $N = (+1)$ for nucleons, $N = (-1)$ for leptons, and $N = 0$ for bosons. In this case, the constant b_1 would be roughly one-half of the nucleon mass (i.e., $\approx$500 MeV). This last interpretation seems to be consistent with the later generalization of the present formalism to nuclear symmetry. However, any further speculation of this type is not intended for the present paper.

III. OSCILLATOR WITH ODD-HALF SPIN

Thus far, the discussions have not touched the question on odd-half spins. More precisely, the formalism of Sec. II applies strictly to systems with integer spin angular momentum only since it gives the three components of angular momentum by $(m_{23}, m_{31}, m_{12}) = (L_1, L_2, L_3)$ such that $\mathbf{L}^2$ has eigenvalues $l(l+1)$ with $l = 0, 1, 2, 3, \cdots$. In either x- or p-representation, the eigenfunctions of Sec. II are clearly those of coupled four-oscillators. For a single four-oscillator the $\mathcal{L}$-invariant eigenfunction takes the form,[1]

$$F(p_\mu) = P^{\frac{1}{2}K} e^{-\frac{1}{2}P} L_n^{K+1}(P)\, Y_K(\theta, \varphi, \omega), \tag{40}$$

where $Y_K(\theta, \varphi, \omega)$ is the four-harmonic discussed in Sec. II and $L_n^K(P)$ is the Kth derivative of the Laguerre polynomial $L_n(P)$ of order $n \geq K+1$ given as a function of $P = -p_\mu p^\mu \geq 0$. For operators, $(x_\mu x^\mu + p_\mu p^\mu)$ and $m_{\mu\nu}m^{\mu\nu}$, for example, we have the eigenvalue equation

$$\begin{aligned} \tfrac{1}{4}\{x_\mu x^\mu + p_\mu p^\mu\} F(p_\mu) &= (n - \tfrac{1}{2}K) F(p_\mu), \\ \tfrac{1}{8} m_{\mu\nu} m^{\mu\nu} F(p_\mu) &= \tfrac{1}{4}K(K+2) F(p_\mu) \\ &= W(W+1) F(p_\mu). \end{aligned} \tag{41}$$

For four-oscillators with odd-half spin, the angular momentum operators $(m_{\mu\nu})$ are replaced by

$$\begin{aligned} J_{\mu\nu} &= (m_{\mu\nu} + \tfrac{1}{2}\sigma_{\mu\nu}) \\ &= -J_{\nu\mu}, \end{aligned} \tag{42}$$

where $(\sigma_{\mu\nu})$ are the six antisymmetrical tensors given by Dirac γ-matrices, namely

$$\sigma_{\mu\nu} = (1/2i)(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu). \tag{43}$$

According to Cheng (see Appendix in Ref. 1), the eigenfunctions of the new operator, $(J_{\mu\nu}J^{\mu\nu})$ are given by the following "generalized" four-harmonics for odd-half-spin,

$$\tilde{Y}_{K,\sigma}(\theta, \varphi, \omega)\rangle = (Z + K + 2) Y_K(\theta, \varphi, \omega)\rangle \tag{44}$$

where

$$Z = \tfrac{1}{2}\sigma_{\mu\nu}m^{\mu\nu} \tag{45}$$

and the symbol "$\cdots\rangle$" implies that Z operates to the right on a four-component spinor which is a function of the spincoordinate σ. Aside from this, Eq. (40) remains the same so that

$$F(p_\mu, \sigma)\rangle = P^{\frac{1}{2}K} e^{-\frac{1}{2}P} L_n^{K+1}(P)\, \tilde{Y}_{K,\sigma}(\theta, \varphi, \omega)\rangle \tag{46}$$

is now the eigenfunction for a four-oscillator which carries odd-half spin,

$$j = (l \pm \tfrac{1}{2}) = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots .$$

Operating on $\tilde{Y}_{K,\sigma}(\theta, \varphi, \omega)\rangle$, we have

$$\begin{aligned} \tfrac{1}{2} m_{\mu\nu} m^{\mu\nu} \tilde{Y}_{K,\sigma}\rangle &= Z(Z+2)\tilde{Y}_{K,\sigma}\rangle \\ &= K(K+2)\tilde{Y}_{K,\sigma}\rangle, \end{aligned} \tag{47}$$

which shows that $F(p_\mu, \sigma)\rangle$ is also an eigenfunction of $Z = \frac{1}{2}\sigma_{\mu\nu}m^{\mu\nu}$ for eigenvalue $Z = K$ or $-(K+2)$.

The 36 generators of the group $\mathcal{G}$ are now given by $(\mathbf{M}_{\mu\nu}, J_{\mu\nu})$. Using the commutation relations of Eq. (12), it can be shown that $Z = \frac{1}{2}\sigma_{\mu\nu}m^{\mu\nu}$ commutes with all 36 generators of $\mathcal{G}$. This implies that Z is a $\mathcal{G}$-invariant and consequently that the $\mathcal{G}$-invariant mass formula (28) may be modified to include a term linear in Z such that

$$\begin{aligned} M = \{ & b_0 + b_1 N + b' Z \\ & + b_2(\mathbf{M}_{\mu\nu}\cdot\mathbf{M}^{\mu\nu} - J_{\mu\nu}J^{\nu\mu} + 4N^2)\}. \end{aligned} \tag{48}$$

Similarly, the symmetry-breaking mass formula of Eq. (37) is now modified in the form

$$\begin{aligned} M = \{ & b_0 + b_1 N + b_1' Z + c_1 T_3 \\ & + b_2[W(W+1) - \mathbf{T}^2 - \tfrac{1}{4}N^2]\}. \end{aligned} \tag{49}$$

[Here, b_1, b_1', and b_2 are not necessarily the same as in Eq. (48).] In Eqs. (48) and (49), it is implicit

that, for more than one four-oscillator, Z represents the sum

$$Z = \sum_{i=1}^{N} Z(i) = \frac{1}{2} \sum_{i=1}^{N} \sigma_{\mu\nu} m^{\mu\nu}(i), \tag{50}$$

where $i = 1, 2, 3, \cdots, N$ for N oscillators. Clearly, Eqs. (48) and (49) are more general than their counterparts of Sec. II since these formulas can be applied to systems of both integer and odd-half integer spins in such a way that, for bosons, we simply equate the term linear in Z to zero and replace $(J_{\mu\nu}J^{\mu\nu})$ by $(m_{\mu\nu}m^{\mu\nu})$.

IV. REMARKS ON NUCLEAR SYMMETRY

While it would clearly be meaningless to talk of a "nonrelativistic" limit in the subnuclear dimensions, the nonrelativistic counterpart of the present formalism may, however, be of significance with regard to nuclear symmetry. The nonrelativistic counterpart of the 36-parameter group $\mathcal{G}$ is a 21-parameter group that is generated by

$$(\mathbf{M}_{kl}, J_{kl}) \qquad (k, l = 1, 2, 3) \tag{51}$$

which now form the following (6×6) representation:

$$\tilde{M}_{kl} = \{\mathbf{M}_{kl} \cdot \boldsymbol{\tau} - i(J_{kl} + iI_{kl})\tau_0\} = \begin{bmatrix} M_{kl}^{(3)} - i(J_{kl} + iI_{kl}) & M_{kl}^{(1)} - iM_{kl}^{(2)} \\ M_{kl}^{(1)} + iM_{kl}^{(2)} & -M_{kl}^{(3)} - i(J_{kl} - iI_{kl}) \end{bmatrix}, \tag{52}$$

where $M_{kl} = M_{lk}$, $J_{kl} = -J_{lk} = (m_{kl} + \frac{1}{2}\sigma_{kl})$, and $I_{kl} = \delta_{kl}$ is the identity operator. In this case, the identity operator (I_{kl}) replaces the Minkowski metric $G = (-1, -1, -1, +1)$ of $\mathcal{G}$. The linear and quadratic invariants F', F'' are now given by

$$\begin{aligned} F' &= \tfrac{1}{6} \operatorname{Tr} (\tilde{M}I) = N, \\ F'' &= \tfrac{1}{6} \operatorname{Tr} (\tilde{M}I\tilde{M}I) \\ &= \tfrac{1}{3}\{\mathbf{M}_{kl} \cdot \mathbf{M}^{lk} - J_{kl}J^{lk} + 4N^2\}. \end{aligned} \tag{53}$$

Here

$$N = \frac{1}{3i} \sum_{i=1}^{N} (x_k p^k - p^k x_k)_{(i)} \tag{54}$$

physically represents the number of nucleons in the nucleus and, therefore, is to be equated to the usual "baryon number," and

$$J_1 = J_{23}, \qquad J_2 = J_{31}, \qquad J_3 = J_{12} \tag{55}$$

represent three components of the total angular momentum. In addition, we also have

$$\begin{aligned} T_k &= \tfrac{1}{4} \operatorname{Tr} (M^{(k)}I) \\ &= \frac{1}{4} \sum_{l=1}^{3} M_{ll}^{(k)} \qquad (k = 1, 2, 3), \end{aligned} \tag{56}$$

which satisfy the commutation relations

$$[T_k, T_l] = i\epsilon^{klm} T_m. \tag{57}$$

The precise physical nature of (T_k) is not completely clear; i.e., aside from the fact that T_3 is one-half of the Hamiltonian of the three-oscillator. One possible interpretation would be to consider them as representing a variation of nuclear quadrupole moments in the sense of Elliot's $U(3)$ generators[11] which are quadratic in (x_k) and (p_k).

Among the 22 independent operators of the group, the 9 operators which enter into the diagonal quadrants of Eq. (52) are reciprocity-invariant operators. These nine operators generate the reciprocity-invariant $U(3)$-content of the 21-parameter group. To show this, let us define

$$B_{kl} = \tfrac{1}{2}\{M_{kl}^{(3)} - i(J_{kl} + iI_{kl})\} \quad (k, l = 1, 2, 3), \tag{58}$$

where it is understood that, in $(J_{kl} + iI_{kl})$,

$$m_{kl} + iI_{kl} = (x_k p_l - p_k x_l) \tag{59}$$

for $k = l$ as well as for $k \neq l$. We then have

$$[B_{kl}, B_{mn}] = \{\delta_{lm}B_{kn} - \delta_{kn}B_{ml}\}, \tag{60}$$

which is precisely the Lie equation satisfied by the generators of the group $U(3)$. If we consider the $U(3)$-invariant sum,

$$\begin{aligned} F &= \tfrac{1}{4} \operatorname{Tr} (BI) \\ &= (T_3 + \tfrac{3}{4}N) \end{aligned} \tag{61}$$

as a parameter, we simply have the unimodular group $SU(3)$. In short, use of the nine reciprocity-invariant operators of the diagonal quadrants of Eq. (52) alone results in the $U(3)$—or equivalently the $SU(3)$—scheme of nuclear symmetry which has been previously discussed by Elliot.[11] However, it should be interesting to see what new results can be obtained by using the whole 21-parameter group which gives the full dynamic symmetry of the six-dimensional phase space, notwithstanding, of course, the inhomogeneous translational transformations which are not contained in this group.

ACKNOWLEDGMENTS

The author wishes to thank Professor Max Born for his continued encouragement, and Dr. Benjamin Lax and Dr. Arthur J. Freeman without whose cooperation this work could not have been realized. He is also indebted to Professor L. C. Biedenharn of CERN for his constructive criticism. Finally, he is infinitely indebted to his wife Shin-Ai for her faith and devotion.

[11] J. P. Elliot, Proc. Roy. Soc. (London) **A245**, 128, 562 (1958).

Space–Time Theory of Elementary Particles. II. Complex Minkowski Space Formalism

ERNEST E. H. SHIN

National Magnetic Laboratory,† Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received 19 April 1965)

The space–time formalism of elementary particles presented previously is generalized to a complex "internal" Minkowski space described by (z_μ, ω_μ) and (z_μ^*, ω_μ^*). In this way, we are able to accommodate the dual signatures of the "superselection" operator and the odd-half spins in a natural manner. The full dynamic symmetry group of the 16-dimensional phase space is a 136-parameter group $\tilde{\mathcal{G}}$ of which the reciprocity-invariant content is a 64-parameter group isomorphic with the 64-parameter Lie group. The complete $\mathcal{L}$-invariant subgroup of $\tilde{\mathcal{G}}$ is a 10-parameter group which is identified as the five-dimensional real "Lorentz" group and which is the "unitary spin" subgroup of $\tilde{\mathcal{G}}$. Discussions on the various reduced or subgroup structures of $\tilde{\mathcal{G}}$ are given. In particular, the physical nature of the $\mathcal{L}$-invariant generators and possible forms of mass formula are discussed.

I. INTRODUCTION

PREVIOUSLY,[1] we presented a theory of elementary particles based on a purely "space–time" point of view which was developed by using the Born reciprocity principle[2] (BRP) as a postulate applied to *one-fold* 'internal" Minkowski space. (Henceforth, we shall refer to Ref. 1 as I.) In the present paper, we now generalize the formalism of I to a two-fold, complex Minkowski space. The canonical operators (x_μ), (p_μ) $(\mu = 1, 2, 3, 4)$ of the one-fold Minkowski space are now replaced by two sets of *complex* canonical operators,

$$(z_\mu), (\omega_\mu),$$
$$(z_\mu^*), (\omega_\mu^*),$$

where * denotes the complex conjugation. The motive for this generalization is explained as follows.

As Finkelstein *et al.*[3] pointed out, the ordinary quantum mechanics is undetermined in the signature of the complex number $(-1)^{\frac{1}{2}} = \eta$ which appears in the canonical commutation relations

$$[x_\mu, p_\nu] = \eta g_{\mu\nu}. \tag{1}$$

More precisely, the "superselection" operator η is capable of two possible signatures, $\pm i$. This is not to be confused with the fact that η changes its sign under a time-reversal operator since, by "two possible signatures," we mean with respect to a fixed sense of time direction. Further, we do not mean a change in signature associated with $(p_k^* = -p_k, x_k^* = x_k)$ $(k = 1, 2, 3)$, for example, since Eq. (1) remains unchanged under this operation. More properly, the two signatures may be interpreted as two possible "eigenvalues" $(= \pm 1)$ of the *real* "superselection" operator ϵ defined by

$$\eta = i\epsilon, \tag{2}$$

where ϵ changes its sign under a time-reversal operation. We feel that the most general way of accommodating the two possible signatures in a natural manner is to define two sets of canonical operators (z_μ, ω_μ) and (z_μ^*, ω_μ^*) such that

$$\begin{aligned} [z_\mu, \omega_\nu] &= -[z_\mu^*, \omega_\nu^*], \\ [z_\mu, \omega_\nu^*] &= [z_\mu^*, \omega_\nu] = 0. \end{aligned} \tag{3}$$

Associating these two sets of canonical operators with the eigenvalues $\epsilon' = (\pm 1)$ of the "superselection" operator, we shall henceforth use the notation $z_\mu^{(\pm)}$, $\omega_\mu^{(\pm)}$.

For a second reason for introducing the complex Minkowski space, it may be recalled that, in I, the description of odd-half spins required the *ad hoc* introduction of the Dirac γ-matrices or more specifically that the existence of odd-half spins was assumed, and not deduced, unlike the case of integer spins. By using complex operators, it is possible to treat odd-half spins with no less rigor than the integer spins. In this connection, it seems pertinent to mention the Vigier theory[4] in which spin is defined in terms of operators of *complex* Euclidean four-space and both integer and odd-half-integer spins are obtained in a natural manner. It is also

† Supported by the U. S. Air Force Office of Scientific Research.

[1] E. E. H. Shin, J. Math. Phys. **6**, 1307 (1965).

[2] M. Born, Nature **163**, 207 (1949); M. Born and H. S. Green, Proc. Roy. Soc. (Edinburgh) **A92**, 470 (1949); Nature **164**, 281 (1949); H. S. Green, Nature **163**, 208 (1949); M. Born, Rev. Mod. Phys. **21**, 463 (1949).

[3] D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962).

[4] L. de Broglie, *Introduction to Vigier Theory of Elementary Particles* (Elsevier Publishing Company, Inc., New York, 1963).

clear that the Vigier description of spins can be accommodated in the present formalism since the symmetry of the "rotator" motion in complex Minkowski space is naturally described by complex Euclidean four-group.

Aside from the replacement of the one-fold Minkowski space of I by a two-fold complex Minkowski space, the mathematical method adopted here is essentially the same as in I. Instead of the eight-dimensional phase space of I, we now have a 16-dimensional phase space to work on. The full dynamic symmetry of the 16-dimensional phase space is given by 136-parameter group $\tilde{\mathcal{G}}$, which is generated by 136 independent operators which are bilinear in $(z_\mu^{(\pm)})$, $(\omega_\mu^{(\pm)})$. The group $\tilde{\mathcal{G}}$ contains two 36-parameter groups $\mathcal{G}^{(+)}$ and $\mathcal{G}_0^{(-)}$ (i.e., of I) as disjoint subgroups, i.e.,

$$\mathcal{G}^{(+)} \otimes \mathcal{G}^{(-)} \subset \tilde{\mathcal{G}}, \tag{4}$$

where it is understood that each of the two 36-parameter groups has the structure

$$SU(2)^{(\pm)} \otimes \mathcal{L}^{(\pm)} \subset \mathcal{G}^{(\pm)}, \tag{5}$$

in which $\mathcal{L}$ is homogeneous Lorentz group and $SU(2)$ is the x–p rotation group containing the Born reciprocity transformations. In the reduced representation, $\mathcal{G}^{(+)} \otimes \mathcal{G}^{(-)}$ of Eq. (4), the various interpretations of I regarding the physical aspects of the 36-parameter group $\mathcal{G}$ and its representations remain unchanged here, aside from the obvious fact that we now have twice as large representations as in I.

In addition to Eq. (4), there exists another useful chain of reduction for the group $\tilde{\mathcal{G}}$,

$$\mathcal{L} \otimes G_{(10)} \subset \tilde{\mathcal{G}}, \tag{6}$$

where $\mathcal{L}$ is homogeneous Lorentz group, and $G_{(10)}$ is a 10-parameter group disjoint from $\mathcal{L}$ and is generated by ten $\mathcal{L}$-invariant operators constructed from $(z_\mu^{(\pm)})$, $(\omega_\mu^{(\pm)})$. We identify $G_{(10)}$ as a five-dimensional real "Lorentz" group which is isomorphic with the 10-parameter Lie group of Dirac γ-matrices discussed by Barut.[5] Since the essential properties of Eq. (4) are known from I, we devote much of the discussions in the present paper to the group $G_{(10)}$. Because of the $\mathcal{L}$ invariance of $G_{(10)}$, we identify it as the "unitary spin" group. Discussions regarding physical aspects of the generators, $G_{(10)}$-invariant and symmetry-breaking forms of mass formula, etc., are given.

[5] A. O. Barut, Phys. Rev. 135, B 839 (1964).

II. SYMMETRY OF 16-DIMENSIONAL PHASE SPACE

In place of the 36-bilinear operators of I, we now have 136 independent operators which are constructed by linearly combining the following bilinear operators:

$$\begin{aligned} &\bar{Z}_\mu^{(\pm)} \tau_i Z_\nu^{(\pm)}, \\ &\bar{Z}_\mu^{(\pm)} \tau_i Z_\nu^{(\mp)} \quad \begin{bmatrix} i = 0, 1, 2, 3 \\ \mu, \nu = 1, 2, 3, 4 \end{bmatrix}, \end{aligned} \tag{7}$$

where (τ_i) denote the 2×2 "isotopic spin" matrices, and

$$Z_\mu^{(\pm)} = \begin{pmatrix} z_\mu^{(\pm)} \\ \omega_\mu^{(\pm)} \end{pmatrix}, \qquad \bar{Z}_\mu^{(\pm)} = (z_\mu^{(\pm)}, \omega_\mu^{(\pm)}).$$

Excluding the bilinear operators representing the commutation rules

$$\begin{aligned} [z_\mu^{(\pm)}, \omega_\nu^{(\pm)}] &= \pm i g_{\mu\nu}, \\ [z_\mu^{(+)}, \omega_\nu^{(-)}] &= [z_\mu^{(-)}, \omega_\nu^{(+)}] = 0, \end{aligned} \tag{8}$$

where $(g_{\mu\nu})$ denote the Minkowski metric tensors as defined in I, we have the following 136 independent operators:

$$\begin{aligned} A_{\mu\nu}^{(1)} &= z_\mu^{(+)} z_\nu^{(+)} - \omega_\mu^{(+)} \omega_\nu^{(-)}, \\ A_{\mu\nu}^{(2)} &= z_\mu^{(+)} \omega_\nu^{(+)} + \omega_\mu^{(+)} z_\nu^{(+)}, \\ A_{\mu\nu}^{(3)} &= z_\mu^{(+)} z_\nu^{(+)} + \omega_\mu^{(+)} \omega_\nu^{(+)}, \\ C_{\mu\nu}^{(1)} &= z_\mu^{(+)} z_\nu^{(-)} - \omega_\mu^{(+)} \omega_\nu^{(-)}, \\ C_{\mu\nu}^{(2)} &= z_\mu^{(+)} \omega_\nu^{(-)} + \omega_\mu^{(+)} z_\nu^{(-)}, \\ C_{\mu\nu}^{(3)} &= z_\mu^{(+)} z_\nu^{(-)} + \omega_\mu^{(+)} \omega_\nu^{(-)}, \\ B_{\mu\nu}^{(1)} &= z_\mu^{(-)} z_\nu^{(-)} - \omega_\mu^{(-)} \omega_\nu^{(-)}, \\ B_{\mu\nu}^{(2)} &= z_\mu^{(-)} \omega_\nu^{(-)} + \omega_\mu^{(-)} z_\nu^{(-)}, \\ B_{\mu\nu}^{(3)} &= z_\mu^{(-)} z_\nu^{(-)} + \omega_\mu^{(-)} \omega_\nu^{(-)}, \\ D_{\mu\nu} &= z_\mu^{(+)} \omega_\nu^{(-)} - \omega_\mu^{(+)} z_\nu^{(-)}, \\ m_{\mu\nu}(+) &= z_\mu^{(+)} \omega_\nu^{(+)} - \omega_\mu^{(+)} z_\nu^{(+)}, \\ m_{\mu\nu}(-) &= z_\mu^{(-)} \omega_\nu^{(-)} - \omega_\mu^{(-)} z_\nu^{(-)}, \end{aligned}$$

(Here and henceforth, it is to be understood that the raising and lowering of the vector indices are defined in the same manner as in I. That is,

$$\begin{aligned} z^{(\pm)\mu} &= g^{\mu\nu} z_\nu^{(\pm)}, \qquad z_\mu^{(\pm)} = g_{\mu\nu} z^{(\pm)\nu}, \\ \omega^{(\pm)\mu} &= g^{\mu\nu} \omega_\nu^{(\pm)}, \qquad \omega_\mu^{(\pm)} = g_{\mu\nu} \omega^{(\pm)\nu}, \end{aligned}$$

and so forth.) To prove that the 136 operators of Eq. (9) form a group, it should suffice to show that

they generate a *closed* commutation algebra. Using the commutation rules of Eq. (8), we obtain

$$(1/i)[A_{\mu\nu}^{(k)}, A_{\lambda\kappa}^{(l)}] = \epsilon^{klm}\{g_{\nu\lambda}A_{\mu\kappa}^{(m)} + g_{\mu\lambda}A_{\nu\kappa}^{(m)} + g_{\nu\kappa}A_{\lambda\mu}^{(m)} + g_{\mu\kappa}A_{\lambda\nu}^{(m)}\},$$

$$(1/i)[B^{(k)}, B^{(l)}] = -\epsilon^{klm}\{g_{\nu\lambda}B_{\mu\kappa}^{(m)} + g_{\mu\lambda}B_{\nu\kappa}^{(m)} + g_{\nu\kappa}B_{\lambda\mu}^{(m)} + g_{\mu\kappa}B_{\lambda\nu}^{(m)}\},$$

$$(1/i)[C_{\mu\nu}^{(k)}, C_{\lambda k}^{(l)}] = -\epsilon^{klm}\{g_{\nu k}A_{\mu\lambda}^{(m)} - g_{\mu\lambda}B_{\nu\kappa}^{(m)}\},$$

$$[A_{\mu\nu}^{(k)}, B_{\lambda\kappa}^{(l)}] \equiv 0,$$

$$(1/i)[A_{\mu\nu}^{(k)}, C_{\lambda\kappa}^{(l)}] = \epsilon^{klm}\{g_{\mu\lambda}C_{\nu\kappa}^{(m)} + g_{\nu\lambda}C_{\mu\kappa}^{(m)}\},$$

$$(1/i)[B_{\mu\nu}^{(k)}, C_{\lambda\kappa}^{(l)}] = -\epsilon^{klm}\{g_{\mu\kappa}C_{\lambda\nu}^{(m)} + g_{\nu\kappa}C_{\lambda\mu}^{(m)}\},$$

$$(1/i)[D_{\lambda\kappa}, B_{\mu\nu}^{(k)}] = \{g_{\mu\kappa}C_{\lambda\nu}^{(k)} + g_{\nu\kappa}C_{\lambda\mu}^{(k)}\},$$

$$(1/i)[D_{\mu\nu}, C_{\lambda\kappa}^{(k)}] = \{g_{\nu\kappa}A_{\mu\lambda}^{(k)} + g_{\mu\lambda}B_{\nu\kappa}^{(k)}\},$$

$$(1/i)[C_{\mu\nu}^{(k)}, C_{\lambda\kappa}^{(l)}] = -\epsilon^{klm}\{g_{\nu\kappa}A_{\mu\lambda}^{(m)} - g_{\mu\lambda}B_{\nu\kappa}^{(m)}\},$$

$$(1/i)[m_{\mu\nu}(+), m_{\lambda\kappa}(+)](\mu \neq \nu, \lambda \neq \kappa)$$
$$= -(1/i)[A_{\mu\nu}^{(k)}, A_{\lambda\kappa}^{(k)}] = \{-g_{\mu\kappa}m_{\nu\lambda}(+) + g_{\nu\kappa}m_{\mu\lambda}(+) - g_{\nu\lambda}m_{\mu\kappa}(+) + g_{\mu\lambda}m_{\nu\kappa}(+)\},$$

$$(1/i)[m_{\mu\nu}(-), m_{\lambda\kappa}(-)](\mu \neq \nu, \lambda \neq \kappa)$$
$$= -(1/i)[B_{\mu\nu}^{(k)}, B_{\lambda\kappa}^{(k)}] = \{g_{\mu\kappa}m_{\nu\lambda}(-) - g_{\nu\kappa}m_{\mu\lambda}(-) + g_{\nu\lambda}m_{\mu\kappa}(-) - g_{\mu\lambda}m_{\nu\kappa}(-)\},$$

$$[m_{\mu\nu}(+), m_{\lambda\kappa}(-)] \equiv 0,$$

$$(1/i)[A_{\mu\nu}^{(k)}, m_{\lambda\kappa}(+)] = \{g_{\nu\kappa}A_{\mu\lambda}^{(k)} + g_{\mu\kappa}A_{\nu\lambda}^{(k)} - g_{\mu\lambda}A_{\nu\kappa}^{(k)} - g_{\nu\lambda}A_{\mu\kappa}^{(k)}\},$$

$$(1/i)[B_{\mu\nu}^{(k)}, m_{\lambda\kappa}(-)] = \{-g_{\nu\kappa}B_{\mu\lambda}^{(k)} + g_{\mu\kappa}B_{\nu\lambda}^{(k)} - g_{\mu\lambda}B_{\nu\kappa}^{(k)} - g_{\nu\lambda}B_{\mu\kappa}^{(k)}\},$$

$$[A_{\mu\nu}^{(k)}, m_{\lambda\kappa}(-)] = [B_{\mu\nu}^{(k)}, m_{\lambda\kappa}(+)] = 0,$$

$$(1/i)[D_{\mu\nu}, m_{\lambda\kappa}(+)] = \{g_{\mu\kappa}D_{\lambda\nu} - g_{\mu\lambda}D_{\kappa\nu}\},$$

$$(1/i)[D_{\mu\nu}, m_{\lambda\kappa}(-)] = \{g_{\nu\lambda}D_{\mu\kappa} - g_{\nu\kappa}D_{\mu\lambda}\},$$

$$(1/i)[C_{\mu\nu}^{(k)}, m_{\lambda\kappa}(+)] = \{g_{\mu\kappa}C_{\lambda\nu}^{(k)} - g_{\mu\nu}C_{\kappa\nu}^{(k)}\},$$

$$(1/i)[C_{\mu\nu}^{(k)}, m_{\lambda\kappa}(-)] = \{g_{\nu\lambda}C_{\mu\kappa}^{(k)} - g_{\nu\kappa}C_{\mu\lambda}^{(k)}\},$$

$$(1/i)[D_{\mu\nu}, D_{\lambda\kappa}] = -\{g_{\kappa\nu}m_{\mu\lambda}(+) - g_{\mu\lambda}m_{\nu\kappa}(-)\}, \quad (10)$$

where $k, l, m = 1, 2, 3$, and the structure constant $\tilde{\epsilon}^{klm}$ are defined by

$$\tilde{\epsilon}^{123} = \epsilon^{123},$$
$$\tilde{\epsilon}^{231} = -\epsilon^{231}, \quad (11)$$
$$\tilde{\epsilon}^{312} = -\epsilon^{312},$$

where

$$\epsilon^{klm} = \begin{cases} +1 : k, l, m = 1, 2, 3 \text{ in cyclic order;} \\ -1 : \text{otherwise } (k \neq l \neq m). \end{cases}$$

Equation (10) shows that the commutation algebra is closed and, therefore, that the 136 operators of Eq. (9) form a group, $\tilde{\mathcal{G}}$. To understand the mathematical and physical nature of this large group, it might be well to examine some of its subgroup or reduced structures.

III. $\mathcal{G}^{(+)} \otimes \mathcal{G}^{(-)} \subset \tilde{\mathcal{G}}$

Inspection of Eq. (10) reveals that each of the two sets of 36 operators,

$$(-iA_{\mu\nu}^{(1)}, iA_{\mu\nu}^{(2)}, A_{\mu\nu}^{(3)}, m_{\mu\nu}(+)) \quad (12)$$

and

$$(iB_{\mu\nu}^{(1)}, -iB_{\mu\nu}^{(2)}, -B_{\mu\nu}^{(3)}, m_{\mu\nu}(-)), \quad (13)$$

forms a 36-parameter group of the type discussed in I. Let us denote these by $\mathcal{G}^{(+)}$ and $\mathcal{G}^{(-)}$, respectively. Since the two sets of operators commute with each other, they form two 36-parameter groups which are disjoint from each other. In the (16 × 16) representation, the 136-parameter group contains the (8 × 8) representations of $\mathcal{G}^{(+)}$ and $\mathcal{G}^{(-)}$ in the form

$$\begin{bmatrix} \mathcal{G}^{(+)} & \text{----} \\ \text{-----} & \text{-----} \\ \text{----} & \mathcal{G}^{(-)} \end{bmatrix} \quad (14)$$

in the reduction, $\mathcal{G}^{(+)} \otimes \mathcal{G}^{(-)} \subset \tilde{\mathcal{G}}$. The crossed symmetry operations between $\mathcal{G}^{(+)}$ and $\mathcal{G}^{(-)}$ are defined by $(C_{\mu\nu}^{(k)}, D_{\mu\mu})$ (and their complex-conjugate representations) which enter in the off-diagonal quadrants of Eq. (14). In the "double $\mathcal{G}$" representation, $\mathcal{G}^{(+)} \otimes \mathcal{G}^{(-)}$, the results of I can be used without modification since, in the present case, we merely have the representations of I increased two-fold without involving any change in the nature of the generators.

IV. RECIPROCITY-INVARIANT CONTENT OF $\tilde{\mathcal{G}}$

It is seen that not all 136 operators of Eq. (9) are invariant under the reciprocity transformations[1]

$$(z_\mu^{(+)}, \omega_\mu^{(+)}) \to (\pm\omega_\mu^{(+)}, \mp z_\mu^{(+)}), (\pm z_\mu^{(+)}, \pm\omega_\mu^{(+)}),$$
$$(z_\mu^{(-)}, \omega_\mu^{(-)}) \to (\pm\omega_\mu^{(-)}, \mp z_\mu^{(-)}), (\pm z_\mu^{(-)}, \pm\omega_\mu^{(-)}). \quad (15)$$

In all, there exist 64 reciprocity-invariant operators in Eq. (9). These operators, given by

$$\begin{bmatrix} A^{(3)}_{\mu\nu}, & B^{(3)}_{\mu\nu}, & C^{(3)}_{\mu\nu} \\ D_{\mu\nu}, & m_{\mu\nu}(+), & m_{\mu\nu}(-) \end{bmatrix}, \tag{16}$$

generate a 64-parameter group $GL(8)$ which is isomorphic with 64-parameter Lie group. To show this, it is convenient to use the (8×8) representation of the generators defined as follows. Let us define 64 generators $(U_{ab}) = (U_{\alpha\beta,\mu\nu})$ $(a, b = 1, 2, 3, \cdots, 8; \alpha, \beta = 1, 2; \mu, \nu = 1, 2, 3, 4)$ such that

$$\begin{aligned} U_{11,\mu\nu} &= \tfrac{1}{2}\{A^{(3)}_{\mu\nu} - im_{\mu\nu}(+) + ig_{\mu\nu}\}, \\ U_{22,\mu\nu} &= -\tfrac{1}{2}\{B^{(3)}_{\mu\nu} - im_{\mu\nu}(-) - ig_{\mu\nu}\}, \\ U_{12,\mu\nu} &= \tfrac{1}{2}i\{C^{(3)}_{\mu\nu} - iD_{\mu\nu}\}, \\ U_{21,\mu\nu} &= \tfrac{1}{2}i\{C^{(3)}_{\mu\nu} + iD_{\mu\nu}\}, \end{aligned} \tag{17}$$

where it is understood that

$$\begin{aligned} m_{\mu\nu}(+) + ig_{\mu\nu} &= z^{(+)}_\mu \omega^{(+)}_\nu - \omega^{(+)}_\mu z^{(+)}_\nu, \\ m_{\mu\nu}(-) - ig_{\mu\nu} &= z^{(-)}_\mu \omega^{(-)}_\nu - \omega^{(-)}_\mu z^{(-)}_\nu, \end{aligned}$$

for $\mu = \nu$ as well as $\mu \neq \nu$. The (8×8) representation then takes the form

$$U_{\mu\nu} = \tfrac{1}{2}\left[\begin{array}{c|c} A^{(3)}_{\mu\nu} - i[m_{\mu\nu}(+) + ig_{\mu\nu}] & iC^{(3)}_{\mu\nu} + D_{\mu\nu} \\ \hline iC^{(3)}_{\mu\nu} - D_{\mu\nu} & -B^{(3)}_{\mu\nu} - i[m_{\mu\nu}(-) - ig_{\mu\nu}] \end{array}\right]. \tag{18}$$

According to Eq. (10), the (U_{ab}) $(a, b = 1, 2, 3, \cdots, 8)$ satisfy the following relations:

$$[U_{ab}, U_{cd}] = \{g_{bc}U_{ad} - g_{ad}U_{bc}\}, \tag{19}$$

where

$$g_{ab} = \delta_{\alpha\beta}g_{\mu\nu} - g_{ba} : \begin{matrix} \alpha, \beta = 1, 2, \\ \mu, \nu = 1, 2, 3, 4, \end{matrix} \tag{20}$$

for $(a; b) = (\alpha, \mu; \beta, \nu)$. This is precisely the commutation relation of 64-parameter Lie group; e.g., the 64-parameter Lie group formed by the (8×8) matrices which are constructed from Dirac γ-matrices and "isotopic spin" matrices (τ_i) $(i = 0, 1, 2, 3)$. As for the reduced structure of the 64-parameter group $GL(8)$, it may be pointed out that $GL(8)$ contains two 16-parameter groups $\mathfrak{L}'$ as disjoint subgroups, where each of two $\mathfrak{L}'$ is generated by 16 operators shown in the two diagonal quadrants of Eq. (18). Thus,

$$GL(8) \supset \mathfrak{L}' \otimes \mathfrak{L}'. \tag{21}$$

The properties of $\mathfrak{L}'$ were discussed in I. In particular, it may be recalled that

$$\mathfrak{L}' \supset T_3 \otimes \mathfrak{L}, \tag{22}$$

where T_3 is the z-component of "isotopic spin" T and $\mathfrak{L}$ is the homogeneous Lorentz group.

Among the 64 generators of $GL(8)$, there exist four $\mathfrak{L}$-invariant operators given by

$$\begin{aligned} A^{(3)} &= \mathrm{Tr}_{(4)}\,(A^{(3)}G) = g^{\mu\nu}A^{(3)}_{\mu\nu}, \\ B^{(3)} &= \mathrm{Tr}_{(4)}\,(B^{(3)}G) = g^{\mu\nu}B^{(3)}_{\mu\nu}, \\ C^{(3)} &= \mathrm{Tr}_{(4)}\,(C^{(3)}G) = g^{\mu\nu}C^{(3)}_{\mu\nu}, \\ D &= \mathrm{Tr}_{(4)}\,(DG) = g^{\mu\nu}D_{\mu\nu}, \end{aligned} \tag{23}$$

which commute with the generators $\{m_{\mu\nu}(+) + m_{\mu\nu}(-)\}$ of the homogeneous Lorentz group $\mathfrak{L}$. By linearly combining Eq. (23), let us define (U_i) $(j = 0, 1, 2, 3)$:

$$\begin{aligned} U_1 &= \tfrac{1}{2}iC^{(3)}, & U_2 &= \tfrac{1}{2}iD, \\ U_3 &= \tfrac{1}{4}(A^{(3)} + B^{(3)}), & U_0 &= \tfrac{1}{2}(A^{(3)} - B^{(3)}). \end{aligned} \tag{24}$$

These operators satisfy the commutation relations

$$\begin{aligned} &[U_i, U_j] = \epsilon^{ijk}U_k \qquad (i, j, k = 1, 2, 3), \\ &[U_k, U_0] = 0. \end{aligned} \tag{25}$$

This shows that the four operators (U_i) generate the group

$$U(2) = U_0 \otimes SU(2) \tag{26}$$

which gives the symmetry of the $z^{(+)} - z^{(-)}$ rotations or the conjugation operations. More specifically, the transformations of Eq. (26) induce exchange and rotations between $(z^{(+)}_\mu, \omega^{(+)}_\mu)$ and $(z^{(-)}_\mu, \omega^{(-)}_\mu)$ without affecting the four-vector indices $\mu = 1, 2, 3, 4$ and without affecting the z-ω (i.e., x-p) symmetry. Thus, the 64-parameter group $GL(8)$ has, as a reduced structure, the representation

$$U(2) \otimes \mathfrak{L} \subset GL(8), \tag{27}$$

where $\mathfrak{L}$ is the 12-parameter Lorentz group

$$\mathfrak{L}^{(+)} \otimes \mathfrak{L}^{(-)}$$

which is isomorphic with the 12-parameter complex bilateral group $R(4) \otimes R(4)$ of Vigier's theory.[4]

The invariant (Casimir) operator of $GL(8)$ is given by

$$F' = \mathrm{Tr}_{(8)}\,(UG) = (U_0 + N), \tag{28}$$

where U_0 is defined in Eq. (24) and N is an integer parameter. A second invariant of the group is given by

$$\begin{aligned} F'' &= \mathrm{Tr}_{(8)}\,(UGUG) \\ &= 2\{\mathbf{U}_{\mu\nu}\cdot\mathbf{U}^{\nu\mu} + \tfrac{1}{4}U^{(0)}_{\mu\nu}U^{(0)\mu\nu} + \tfrac{1}{2}U_0N \\ &\quad + 4N^2 - \tfrac{1}{8}(m_{\mu\nu}(+)m^{\nu\mu}(+) \\ &\quad + m_{\mu\nu}(-)m^{\nu\mu}(-))\}, \end{aligned} \tag{29}$$

where we have put

$$\begin{aligned} &\mathbf{U}_{\mu\nu} = (U^{(1)}_{\mu\nu}, U^{(2)}_{\mu\nu}, U^{(3)}_{\mu\nu}), \quad U^{(0)}_{\mu\nu} = \tfrac{1}{2}(A^{(3)}_{\mu\nu} - B^{(3)}_{\mu\nu}), \\ &U^{(1)}_{\mu\nu} = \tfrac{1}{2}iC^{(3)}_{\mu\nu}, \qquad U^{(2)}_{\mu\nu} = \tfrac{1}{2}iD_{\mu\nu}, \\ &U^{(3)}_{\mu\nu} = \tfrac{1}{4}(A^{(3)}_{\mu\nu} + B^{(3)}_{\mu\nu}). \end{aligned} \tag{30}$$

In general, the invariants of $GL(8)$ are defined by

$$\mathrm{Tr}_{(8)}\,(UGUG\cdots UG). \tag{31}$$

As in I, the tensor products such as $\mathbf{U}_{\mu\nu}\cdot\mathbf{U}^{\mu\nu}$, $U^{(0)}_{\mu\nu}U^{(0)\nu\mu}$ and $m_{\mu\nu}(\pm)m^{\nu\mu}(\pm)$ may be reduced into more meaningful forms which are given in terms of $\mathbf{U}\cdot\mathbf{U} = U_1^2 + U_2^2 + U_3^2$, U_0, $W_\pm(W_\pm + 1)$, and N, etc.

V. SPIN AND THE COMPLETE $\mathfrak{L}$-INVARIANT SUBGROUP

Among 136 generators of $\tilde{\mathfrak{G}}$, there exist 10 operators which commute with generators of the homogeneous Lorentz group $\mathfrak{L}$, given by

$$\begin{aligned} A_k &= \mathrm{Tr}_{(4)}\,(A^{(k)}G) = g^{\mu\nu}A^{(k)}_{\mu\nu}, \\ B_k &= \mathrm{Tr}_{(4)}\,(B^{(k)}G) = g^{\mu\nu}B^{(k)}_{\mu\nu}, \\ C_k &= \mathrm{Tr}_{(4)}\,(C^{(k)}G) = g^{\mu\nu}C^{(k)}_{\mu\nu}, \\ D &= \mathrm{Tr}_{(4)}\,(DG) = g^{\mu\nu}D_{\mu\nu}, \end{aligned} \tag{32}$$

where $k = 1, 2, 3$. Out of these, the four reciprocity-invariant operators (A_3, B_3, C_3, D) were shown (in Sec. IV) to generate the z–z^* rotation group $U(2)$. These ten operators commute with the antisymmetrical operators,

$$\begin{aligned} \textstyle\sum_{\mu\nu} &= m_{\mu\nu}(+) + m_{\mu\nu}(-) \\ &= -\textstyle\sum_{\nu\mu} : (\mu \neq \nu), \end{aligned} \tag{33}$$

which generate the 6-parameter homogeneous Lorentz group $\mathfrak{L}$. According to the commutation relations of Eq. (10), the ten $\mathfrak{L}$-invariant operators of Eq. (32) generate a closed commutation algebra and, therefore, form a 10-parameter group, which we denote as $G_{(10)}$. Consequently, we have another interesting reduction chain,

$$\mathfrak{L} \otimes G_{(10)} \subset \tilde{\mathfrak{G}} \tag{34}$$

in addition to $\mathfrak{G}^{(+)} \otimes \mathfrak{G}^{(-)}$ discussed in Sec. III.

Let us first discuss the rotator representations of $m_{\mu\nu}(+) \otimes m_{\mu\nu}(-)$ of which the six-parameter group $\mathfrak{L}$ generated by $(\Sigma_{\mu\nu})$ of Eq. (33) is a subgroup. The "left" and "right" rotations of the bilateral four-rotation group are defined by the 4×4 matrices $L_k^{(\pm)}$ and $R_k^{(\pm)}$ ($k = 1, 2, 3$), respectively, in the following manner:

$$\begin{aligned} \delta z_\mu^{(\pm)} &= \sum_{\nu=1}^{4}\sum_{k=1}^{3} \delta c_k^{(\pm)}(L_k^{(\pm)})_{\mu\nu}z_\nu^{(\pm)}, \\ \delta z_\mu^{(\pm)} &= \sum_{\nu=1}^{4}\sum_{k=1}^{3} \delta c_k^{(\pm)}(R_k^{(\pm)})_{\mu\nu}z_\nu^{(\pm)}, \end{aligned} \tag{35}$$

where $\delta c_k^{(\pm)}$, $\delta c_k^{(\pm)}$ ($k = 1, 2, 3$) are 12 arbitrary infinitesimal parameters. In terms of $L_k^{(\pm)}$ and $R_k^{(\pm)}$, the "left"-spins $M_k^{(\pm)}$ and the "right"-spins $N_k^{(\pm)}$ are defined by

$$\begin{aligned} M_k^{(\pm)} &= \pm\frac{1}{2i}\sum_{\mu,\nu=1}^{4} \omega_\mu^{(\pm)}(L_k^{(\pm)})_{\mu\nu}z_\nu^{(\pm)}, \\ N_k^{(\pm)} &= \pm\frac{1}{2i}\sum_{\mu,\nu=1}^{4} \omega_\mu^{(\pm)}(R_k^{(\pm)})_{\mu\nu}z_\nu^{(\pm)}. \end{aligned} \tag{36}$$

These satisfy the commutation relations

$$\begin{aligned} &[M_i^{(\pm)}, M_j^{(\pm)}] \\ &\quad = i\epsilon^{ijk}M_k^{(\pm)}, \\ &[M_i^{(\pm)}, N_j^{(\pm)}] \\ &\quad = [M_i^{(\pm)}, M_j^{(\mp)}] = [N_i^{(\pm)}, N_j^{(\mp)}] = 0, \\ &[N_i^{(\pm)}, N_j^{(\pm)}] \\ &\quad = i\epsilon^{ijk}N_k^{(\pm)} : (i, j, k = 1, 2, 3). \end{aligned} \tag{37}$$

Let us define

$$\begin{aligned} M^2(\pm) &= \sum_{i=1}^{3}(M_i^{(\pm)})^2, \\ N^2(\pm) &= \sum_{i=1}^{3}(N_i^{(\pm)})^2. \end{aligned} \tag{38}$$

Operating on the four-harmonics $Y_{K_\pm}(\theta^\pm, \varphi^\pm, \psi^\pm)$ ($K_\pm = 0, 1, 2, 3, \cdots$), we have

$$\begin{aligned} M^2(\pm)Y_{K\pm} &= N^2(\pm)Y_{K\pm}, \\ &= \tfrac{1}{4}K_\pm(K_\pm + 2)Y_{K\pm}, \\ &= W_\pm(W_\pm + 1)Y_{K\pm}, \end{aligned} \tag{39}$$

where $W_\pm = \frac{1}{2}K_\pm = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$. Thus, magnitudes of the "left" and "right" spins are identical and are equal to $W_\pm = \frac{1}{2}K_\pm$.

While $(M_i^{(+)}, N_i^{(-)})$ and $(M_i^{(-)}, N_i^{(-)})$, separately, do not commute with (C_k) and D of Eq. (32), the symmetrized "left" and "right" spins,

$$J_i' = \tfrac{1}{2}\sum_i' = (M_i^{(+)} + M_i^{(-)}),$$
$$J_i'' = \tfrac{1}{2}\sum_i'' = (N_i^{(+)} + N_i^{(-)}), \tag{40}$$

however, commute with all ten generators of the group $G_{(10)}$. The six operators of Eq. (40) are, then, the generators of the homogeneous Lorentz group $\mathfrak{L}$ in the reduction of Eq. (34). According to Eq. (37), we have

$$[J_i', J_j'] = i\epsilon^{ijk}J_k',$$
$$[J_i'', J_j''] = i\epsilon^{ijk}J_k'', \tag{41}$$
$$[J_i', J_j''] = 0.$$

Further, the "left" and "right" spins have identical eigenvalues or more precisely

$$J'(J'+1) = J''(J''+1) = J(J+1), \tag{42}$$

where $J = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$. We identify J as the spin. Unlike I, then, the odd-half spins appear as naturally as do the integer spins. The present definition of the spin is basically the same as in Vigier's theory[4] in which the odd-half spins are also defined in this manner. With this much on spin, let us now turn to the $\mathfrak{L}$-invariant group $G_{(10)}$.

VI. THE "UNITARY SPIN" SUBGROUP OF $\tilde{\mathfrak{G}}$

Since all 10 generators of $G_{(10)}$ commute with the generators of the homogeneous Lorentz group $\mathfrak{L}$, it seems reasonable to identify $G_{(10)}$ as the "unitary spin" subgroup of $\tilde{\mathfrak{G}}$. Let us represent the 10 operators of Eq. (32) as follows:

$$A_1' = -\tfrac{1}{4}iA_1, \quad B_1' = +\tfrac{1}{4}iB_1, \quad C_1' = -\tfrac{1}{2}C_1,$$
$$A_2' = +\tfrac{1}{4}iA_2, \quad B_2' = -\tfrac{1}{4}iB_2, \quad C_2' = +\tfrac{1}{2}C_2,$$
$$A_3' = \tfrac{1}{4}A_3, \quad B_3' = -\tfrac{1}{4}B_3, \quad C_3' = -\tfrac{1}{2}iC_3,$$
$$D' = \tfrac{1}{2}D. \tag{43}$$

According to Eq. (10), these operators satisfy the commutation relations

$$[A_i', A_j'] = i\epsilon^{ijk}A_k', \quad [B_i', B_j'] = i\epsilon^{ijk}B_k',$$
$$[C_i', C_j'] = i\epsilon^{ijk}(A_k' + B_k'), \tag{44}$$
$$[D', A_k'] = -\tfrac{1}{2}C_k', \; [D', B_k'] = \tfrac{1}{2}C_k',$$
$$[D', C_k'] = +(A_k' - B_k').$$

We now wish to show that the 10-parameter group $G_{(10)}$ is the five-dimensional real "Lorentz" group with the metric

$$G = (-1, +1, -1, -1, -1)$$

discussed by Barut.[5] Using Eq. (43), let us define the following antisymmetric tensors (M_{ab}) $(a, b = 1, 2, 3, 4, 5)$ of rank two:

$$M_{ab} = \begin{bmatrix} 0 & iD' & i(A_2'-B_2') & i(A_3'-B_3') & i(A_1'-B_1') \\ & 0 & C_2' & C_3' & C_1' \\ & & 0 & \frac{1}{i}(A_1'+B_1') & \frac{1}{i}(A_3'+B_3') \\ & & & 0 & \frac{1}{i}(A_2'+B_2') \\ & & & & 0 \end{bmatrix}, \tag{45}$$

where $M_{ab} = -M_{ba}$ for $b \neq a$, and $M_{aa} \equiv 0$. Using Eq. (44), it is a simple matter to show that the (M_{ab}) satisfy the commutation relation

$$[M_{ab}, M_{cd}]$$
$$= \{-g_{bc}M_{ad} - g_{ad}M_{bc} + g_{ac}M_{bd} + g_{bd}M_{ac}\}$$
$$: (g_{ab}) = (-1, +1, -1, -1, -1). \tag{46}$$

This is precisely the commutation algebra of the real "Lorentz" group with metric $(g_{ab}) = (-1, +1, -1, -1, -1)$.

For the sake of convenience, let us use the operators (T_k, R_k, L_k, Z) $(k = 1, 2, 3)$ such that

$$M_{ab} = \begin{bmatrix} 0 & iZ & iR_2 & iR_3 & iR_1 \\ & 0 & L_2 & L_3 & L_1 \\ & & 0 & T_1 & T_3 \\ & & & 0 & T_2 \\ & & & & 0 \end{bmatrix}. \tag{47}$$

In terms of these operators, the Casimir operators of Barut[5] are given by

$$F^2 = \mathrm{Tr}\,(MGMG),$$
$$= 2\{\mathbf{R}^2 + \mathbf{L}^2 - \mathbf{T}^2 - Z^2\},$$
$$G^4 = \mathrm{Tr}\,(MGMGMG),$$
$$= \{-(\mathbf{R}\cdot\mathbf{T})^2 - (\mathbf{T}\cdot\mathbf{L})^2 + [ZT - (\mathbf{R}\times\mathbf{L})]^2\}, \tag{48}$$

where

$$\mathbf{R}^2 = R_1^2 + R_2^2 + R_3^2,$$
$$\mathbf{L}^2 = L_1^2 + L_2^2 + L_3^2,$$
$$\mathbf{T} = T_1^2 + T_2^2 + T_3^2,$$

and so forth.

Since the present group is of rank two, there exist two mutually commutative operators (T_3, Z)

whose eigenvalues fully specify a state in a given representation. This choice of observables is based on the fact that $[Z, T_k] = 0$ for $k = 1, 2, 3$, while Z does not commute with (R_k) and (L_k). Thus, (Z, T_k) generate a subgroup $U(2) = Z \otimes SU(2)$ of $G_{(10)}$. Here, $T_3(= 0, \pm\frac{1}{2}, \pm 1, \cdots)$ and $2Z(= 0, \pm 1, \pm 2, \cdots)$, respectively, may be identified as the z-component of the "isotopic spin" and "hypercharge."

In the absence of a dynamic agent or mechanism (viz., interactions) which breaks the symmetry of $G_{(10)}$, mass would be an invariant of $G_{(10)}$ and given as a function of the Casimir operators defined above. One possible—and also the simplest—form of this is

$$\begin{aligned} M &= b_0 + b_1 F^2, \\ &= \{b_0 + b_1(\mathbf{R}^2 + \mathbf{L}^2 - \mathbf{T}^2 - Z^2)\}. \end{aligned} \tag{49}$$

[It may be pointed out that, in the "non-relativistic limit" of $\mathbf{L} \to 0$, Z is simply a multiple of the identity operator and $(\mathbf{R}^2 + \mathbf{L}^2 - \mathbf{T}^2) \to -(n^2 - 1)$ so that $M \sim (b_0 + b_1 n^2)$ where n is an integer.] On the other hand, if the $G_{(10)}$ symmetry is broken in the manner of

$$G_{(10)} \to Z \otimes SU(2)_T, \tag{50}$$

where $SU(2)_T$ is the "isotopic spin" group generated by (T_1, T_2, T_3) the symmetry-breaking terms may be taken to have the properties of

$$Z = (1/2i)(M_{12} - M_{21})(=\tfrac{1}{2}Y).$$

Considering the breakdown as a perturbative effect, the corresponding terms of the mass formula may be also taken as perturbation terms which are linear, bilinear, ... in (M_{ab}). Up to terms bilinear in (M_{ab}), we obtain

$$M = \{M_0 + M_1 Z + M_2(\mathbf{K}^2 - \mathbf{T}^2 - Z^2)\} \tag{51}$$

in which we have used

$$\mathbf{K} = (\mathbf{R} + \mathbf{L})$$

and

$$(MGMG)_{12} - (MGMG)_{21} = i\{\mathbf{K}^2 - \tfrac{1}{2}F^2 - \mathbf{T}^2 - Z^2\}, \tag{52}$$

and M_0, M_1, M_2 are arbitrary functions of the Casimir operators.

In order to obtain the mass formula in a more general form including the spin, it is clearly necessary to work with the full group $\tilde{\mathcal{G}}$ rather than with $G_{(10)}$ alone. For example, in the $\mathfrak{L} \otimes G_{(10)}$ representation of $\tilde{\mathcal{G}}$, Eq. (51) may be replaced by an expression which contains an additional $J(J + 1)$ term. On the other hand, in the reduction chain,

$$\begin{aligned} \tilde{\mathcal{G}} &\supset \mathcal{G}^{(+)} \otimes \mathcal{G}^{(-)}, \\ &\supset SU(2) \otimes \mathfrak{L} \otimes SU(2) \otimes \mathfrak{L}, \end{aligned} \tag{53}$$

the mass formulas discussed in I are still applicable.

VII. CONCLUDING REMARKS

Thus far, the discussions have been mostly confined to the subgroup or the reduced structure of the 136-parameter group $\tilde{\mathcal{G}}$. In order to appreciate the full physical significance of the present formalism, it is clearly necessary to work out the representations of the full group $\tilde{\mathcal{G}}$. Further, we have yet to work out the representations of even the 10-parameter group $G_{(10)}$ in order to be able to classify elementary particles according to their symmetry properties. After this comes the task of comparing the theoretical predictions with the experimental observations. Beyond this, there are other important problems such as the connection with field theory, electromagnetic structure of particles, and many others. It is, therefore, evident that what we have presented thus far is merely a small beginning of what, we believe, will be a series of papers on the same and related subject matter.

ACKNOWLEDGMENTS

The author wishes to repeat his acknowledgement of I to Prof. Max Born, Dr. Benjamin Lax, and Dr. Arthur J. Freeman, and especially to his wife, Shin-Ai. He also thanks Prof. L. C. Biedenharn of CERN for useful comments.

Spin and Isospin in S-Matrix Theory

JOHN R. TAYLOR
Palmer Physical Laboratory, Princeton, New Jersey
(Received 7 July 1965)

The Gunson–Olive technique for exploiting physical region poles and the unitarity equations to derive the analytic properties of S-matrix elements is applied to particles which belong to isospin multiplets and to particles with spin. In the former case it is shown that, in the framework of S-matrix theory, crossing symmetry leads to the requirement that particles and antiparticles transform according to complex conjugate representations; and it is possible to clarify the origin of the familiar isospin crossing matrices. In the case of particles with spin the construction of appropriate "spinor" basic states is discussed and the analytic properties of the corresponding spinor S-matrix elements or "M-functions" are derived.

I. INTRODUCTION

DURING the last two years great progress has been made towards the construction of a self-contained theory of the S-matrix based on physically realistic postulates. As a result of the work of Gunson[1] and Olive,[2] it appears that the essential ingredients of such a theory are the following five postulates:

(1) The superposition principle,
(2) The existence of a unitary S-matrix,
(3) Lorentz invariance,
(4) The decomposition law (or "connectedness structure"),
(5) Analyticity on the mass-shell.

Using these postulates the attempt has been made to prove such familiar properties as crossing symmetry, Hermitian-analyticity, the correspondence between particles and poles, and the existence of antiparticles; all of which properties appeared as somewhat unnatural postulates in previous attempts at a theory of the S-matrix.[3]

Unfortunately, the proofs of these properties are not completely satisfactory. This is because of their dependence on a knowledge of the singularities of the S-matrix, the complete enumeration of which is still an unsolved problem. At present all that is known is that the exact location of singularities should not appear explicitly in the postulate of analyticity; rather, this postulate should be formulated in such a way that all singularities are derivable from the other postulates (in particular, from unitarity and the decomposition law). Thus far no one has shown how this can be done without using the very properties it was hoped to prove (i.e., crossing symmetry, etc.).

Until these difficulties can be overcome, the S-matrix theory "proofs" of crossing symmetry, etc., can only be considered as explanations. (Given, for example, that crossing symmetry is true, there is no question but that the Gunson–Olive "proof" explains *why* it is true.) However, even to have explained such properties is a considerable achievement, since previously their only justification was their occurrence in perturbation theory. And it would seem valuable to extend the Gunson–Olive discussion of spinless particles to the case of particles which belong to isospin multiplets and to particles with spin.[4]

The principal aim of this paper, then, is to discuss the S-matrix theory of particles with isospin and—a much more complicated problem—particles with spin. Because the most recent developments in S-matrix theory are probably not very widely known and the notation is not familiar, the next section is devoted entirely to a review of the S-matrix theory of spinless particles. I first discuss the five postulates listed above and then, having defined the necessary notation, review briefly the relevant parts of the Gunson–Olive work. In Sec. III, I am then free to consider isospin invariance and its effect on the form of crossing symmetry. The last two sections are concerned with the S-matrix theory of particles with spin. More space is given to these sections because of the greater complexity of the topic. With the great number of possible labellings of spin states, one must first decide which matrix elements would be expected to have the desired analytic properties. Accordingly Sec. IV is devoted entirely to the construction of the appropriate basic spin states—the

[1] J. Gunson, J. Math. Phys. **6**, 827 (1965).
[2] D. I. Olive, Phys. Rev. **135**, B745 (1964).
[3] See, in particular, H. P. Stapp, Phys. Rev. **125**, 2139 (1962).
[4] Many of the techniques used here to include spin and isospin in the Gunson–Olive work have been known for some time. See, in particular, A. O. Barut, Phys. Rev. **130**, 436 (1963); A. O. Barut and B. C. Unal, Nuovo Cimento **28**, 112 (1963); these articles discuss crossing symmetry for particles with spin and isospin, respectively, but both take the view that crossing is a fundamental postulate.

so-called spinor states. Then in Sec. V, I discuss the analytic properties of the corresponding spinor matrix elements (or "M-functions").

II. S-MATRIX THEORY OF SPINLESS PARTICLES

The first postulate, the superposition principle, specifies the nature of the system of interest long before and long after interactions take place. The most natural statement of the principle is in terms of Hilbert space; namely, that long before and long after interactions a system can be described by "states" and that these states are found to correspond to rays[5] in a Hilbert space which has the following properties:

(i) The space is composed of orthogonal "superselection subspaces" within which any linear superposition of vectors represents a physically realizable state.

(ii) If φ and ψ are any two realizable states represented by unit vectors $|\varphi\rangle$ and $|\psi\rangle$, then $|\langle\varphi \mid \psi\rangle|^2$ is the probability that a system known to be in the state ψ be observed in the state φ.

(iii) The space is a Fock space spanned by an orthonormal basis of (improper) vectors representing states of definite numbers of particles with definite momenta. The notation used here for these states is as follows:

The states of a single (spinless) particle are labelled by their four-momentum[6] p and particle type[6] t; the two variables p and t can conveniently be denoted by a single label $k = (p, t)$ and the corresponding state vector by $|k\rangle = |p, t\rangle$. These vectors are normalized such that

$$\langle p', t' \mid p, t\rangle = p^0\delta_3(\mathbf{p}' - \mathbf{p})\delta_{t't}. \qquad (2.1)$$

The n-particle states are labelled by the n-tuple $K = (k_1, \cdots, k_n)$, where $k_i = (p_i, t_i)$, and the corresponding state vectors are just tensor products of one-particle vectors,

$$|K\rangle = |k_1, \cdots, k_n\rangle = |k_1\rangle \otimes \cdots \otimes |k_n\rangle,$$

(symmetrized according to the appropriate statistics if there are identical particles[7]). It is also convenient to introduce the n-tuples $P = (p_1, \cdots, p_n)$ and $T = (t_1, \cdots, t_n)$ and to write the state K as $K = (P, T)$. (No serious ambiguity arises from these two ways of writing the state K.)

The reason for introducing the relatively abstract and unphysical concept of Hilbert space is that it does provide the only natural expression of the superposition principle. Many recent papers on S-matrix theory[8] have tried to bypass discussion of the Hilbert space of initial and final states by defining the S-matrix elements as transition amplitudes (i.e. numbers whose squared moduli are transition probabilities). This procedure fails to settle the important question of how the *phase* of S-matrix elements can be consistently defined—a question which can only be answered using the superposition principle. It seems to be essential that this principle applies to both initial and final states [i.e. a superposition of two initial (or final) states in one superselection class *is* a realisable initial (or final) state] and that the corresponding Hilbert space enter, at least implicitly, in the theory. Indeed, if one examines the only complete attempt at formal S-matrix postulates[9] which do not include Hilbert space, one can easily see that they contain sufficient information to construct it.

The second postulate, which defines the S-matrix, does not in fact have to state that S is unitary. It states only the experimental observation that there is a one-to-one correspondence between possible initial and final states and that this correspondence is such as to preserve the superposition principle (i.e. if initial states φ and ψ lead to final states φ_s and ψ_s, then $|\langle\varphi \mid \psi\rangle| = |\langle\varphi_s \mid \psi_s\rangle|$, where $|\varphi\rangle \cdots$ are unit vectors representing $\varphi \cdots$). A famous theorem due to Wigner[10] ensures that such a correspondence between states (i.e., between rays in the Hilbert space of the first postulate) can be represented by a correspondence S between vectors in Hilbert space with S either unitary or antiunitary. That S is actually unitary follows because the relationship between

---

[5] A ray ψ corresponding to a vector $|\psi\rangle$ is the set $\psi = \{z\,|\psi\rangle\}$ where z ranges over the whole complex plane. Within any superselection subspace the rays are in one-to-one correspondence with physically realisable states, and the same label can therefore be used for both.

[6] The space component $\mathbf{p}$ ranges over the whole of three-dimensional Euclidean space; the energy component p^0 is of course redundant being defined by the mass-shell condition, $p^0 = +(\mathbf{p}^2 + m^2)^{\frac{1}{2}}$. The label t runs over all stable particles of the theory.

[7] The famous connection between statistics and spin has been discussed in the S-matrix framework by Stapp.[3] In this paper I shall for the most part avoid these complications by considering only states with all different particles. In the last section, however, the question of ordering the variables has to be discussed and the normal connection between spin and statistics will be assumed.

[8] See, for example, Ref. 3 and P. V. Landshoff, *The S-matrix without Field Theory*, Cambridge preprint, 1964 (to be published).

[9] H. P. Stapp, *Studies in the Foundations of S-matrix Theory*, California preprint UCRL-10843 (to be published). The construction mentioned here has been described in detail by J. R. Taylor, Phys. Rev. 140, B187 (1965).

[10] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959), p. 233.

coordinates and momenta is the same for initial and final states.[11]

The third postulate, Lorentz invariance, states that the probabilities of observing any definite process as measured by two different observers (i.e., Lorentz frames of reference) must be the same. Thus if two observers 0 and 0_L are related by a Lorentz transformation[12] L, the process described by 0 as a transition from the state ψ to the state φ would be described by 0_L as the transition $(\varphi_L \leftarrow \psi_L)$, and Lorentz invariance implies that

$$|\langle\varphi|\ S\ |\psi\rangle| = |\langle\varphi_L|\ S\ |\psi_L\rangle|. \tag{2.2}$$

In fact, it follows from Wigner's theorem that a Lorentz transformation L can be represented as a unitary transformation $U(L)$ on the vectors of Hilbert space. Thus Lorentz invariance implies that

$$U^\dagger(L)SU(L)\ |\psi\rangle = zS\ |\psi\rangle, \qquad (|z| = 1),$$

for any physically realizable ψ. The phase factor $z = z(L, \psi)$ may, of course, depend on ψ and L; however, it is easily seen by applying the superposition principle, that z must be the same for all ψ in the same superselection class. It then follows that the numbers $z(L, \psi)$ form a one-dimensional unitary representation of the Lorentz group, and hence, are actually unity. Thus Lorentz invariance in fact implies that

$$\langle\varphi|\ S\ |\psi\rangle = \langle\varphi|\ U^\dagger(L)SU(L)\ |\psi\rangle = \langle\varphi_L|\ S\ |\psi_L\rangle; \tag{2.3}$$

i.e., the modulus signs in Eq. (2.2) can be removed. If one defines the phases of the one-particle basis vectors such that[13]

$$|Lp, t\rangle = U(L)\ |p, t\rangle, \tag{2.4}$$

then from the superposition principle and the group property, it follows that the same equation holds for the multiparticle states; and, in terms of the basis vectors, Lorentz invariance, Eq. (2.3), becomes the following important identity:

$$\langle P', T'|\ S\ |P, T\rangle = \langle LP', T'|\ S\ |LP, T\rangle, \tag{2.5}$$

for any L, where

$$LP = (Lp_1, \cdots, Lp_n).$$

[11] In theories where it is defined as the limit when t_1 and t_2 tend to $\pm\infty$ of a time-displacement operator $U(t_1, t_2)$, S is continuously joined to the identity and so is obviously unitary. Such an argument is unfortunately not permissible in S-matrix theory.

[12] Throughout this paper the "Lorentz Group" will be taken to be the group of physically realizable Lorentz transformations; i.e. the proper orthochronous Lorentz group.

[13] The phase of $U(L)$ itself is defined by imposing the group property $U(L)U(M) = U(LM)$. That this is possible (for spinless particles) was shown by E. P. Wigner, Ann. of Math. 40, 149 (1939).

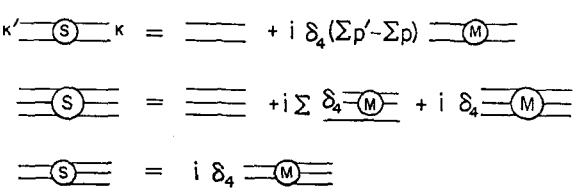

FIG. 1. The decomposition into connected parts of the three simplest S-matrix elements. The element for the process (2 particles ← 2 particles) is made up of two parts: that corresponding to no interaction and the interaction—or "connected"—part. That for (3 particles ← 3 particles) has three parts: a no-scattering part, a part corresponding to the interaction of two particles while the third is unscattered, and the genuine three-particle interaction part. That for (2 particles ← 3 particles) does not decompose at all. To conform with the time ordering of the states in the matrix element $\langle K'|\ S\ |K\rangle$, these diagrams are drawn with the initial state on the right and final state on the left; i.e. time goes from right to left.

It should be noted that, since the multiparticle state vectors are defined as products of single-particle vectors, the phase definition (2.4) leaves just one arbitrary phase—that of the rest state $|p_0, t\rangle$ say—for each type of particle in the theory.

The fourth postulate, the decomposition law (or "connectedness structure") expresses the observed fact that if two reactions are sufficiently separated in space–time then they do not affect one another.[14] This is most simply stated in terms of a diagramatic representation of S-matrix elements as shown in Fig. 1. The interpretation of these diagrams is more or less self-evident. The bubbles represent matrix elements between momentum basis vectors, the straight lines represent δ_3-functions and the δ_4-functions factored out of the connected parts M represent overall conservation of energy–momentum in the process concerned. Thus the first diagram, which represents the decomposition for elastic scattering, can be written explicitly (for distinct particles) as

$$\langle K'|\ S\ |K\rangle = p_1^0\delta_3(\mathbf{p}_1' - \mathbf{p}_1)p_2^0\delta_3(\mathbf{p}_2' - \mathbf{p}_2) + i\delta_4(p_1' + p_2' - p_1 - p_2)\langle K'|\ M\ |K\rangle.$$

The matrix element $\langle K'|\ M\ |K\rangle$ is called the connected part (without δ-functions) for the process $(K' \leftarrow K)$.

The decomposition equations of Fig. 1 do not have any physical content until some properties of the connected parts are specified. This is the task of the final postulate, analyticity. Neither the physical

[14] See J. H. Crichton and E. H. Wichmann, Phys. Rev. 132, 2788 (1963). The property is somehow connected with the short-range character of interparticle forces. An outstanding problem is to discover whether massless particles, which are generally associated with long range forces, can be incorporated into the theory. In this paper I shall, therefore, assume that all particles have nonzero mass.

FIG. 2. Unitarity for the four-point function below inelastic thresholds. The matrix elements of $M^\dagger$ are $\langle K_2|\, M^\dagger\, |K_1\rangle = \langle K_1|\, M\, |K_2\rangle^*$. The right-hand side represents an appropriate integration over "intermediate states" K of the product $\langle K_2|\, M\, |K\rangle\langle K|\, M^\dagger\, |K_1\rangle$.

origin nor the precise formulation of this postulate is yet understood. Some recent work[15] has indicated that, as was always believed to be the case, analyticity is in some way connected with causality. Whatever its origin, some features of the postulate are clear. It is certain that, with the phases fixed[16] according to Eq. (2.4), the physically measurable connected parts $\langle K'|\, M\, |K\rangle$ must be continuable as analytic functions of the momenta P and P', subject, of course, to the mass-shell and conservation constraints.[17] As already stated, it is believed that the postulate must be formulated so that the only singularities of $\langle K'|\, M\, |K\rangle$ are those which follow from unitarity and decomposition. (For this reason the postulate has been called "maximal analyticity".) Finally it is necessary to specify whether physical values of $\langle K'|\, M\, |K\rangle$ are found by passing above or below singularities lying in the physical region. This is done by the so-called[2] "$i\epsilon$-prescription", which is also probably connected with causality.[15]

Properties which, it is felt, should *not* have to appear in the analyticity postulate, but should be deducible from it, are the correspondence between particles and poles, Hermitian-analyticity, discontinuity relations, existence of antiparticles and crossing symmetry. It is the main purpose of this paper to discuss the derivation of these properties (particularly the last) for particles with spin or isospin. To facilitate this discussion, the remainder of this section is devoted to a brief sketch of the corresponding derivations for simple spinless particles.

The first step in the development of these fundamental properties is to obtain the unitarity equations for the connected parts $\langle K_2|\, M\, |K_1\rangle$. These are found by taking matrix elements of the equation $SS^\dagger = 1$ with a completeness sum inserted between S and $S^\dagger$, then substituting into the resulting equations the decomposition shown in Fig. 1. As an example, the unitarity equation for the simplest connected part—the elastic scattering amplitude or "four-point function"—is shown in Fig. 2. The exact form of the phase space integration on the right-hand side is easily derived from the normalization (2.1) and the definitions of Fig. 1.

The corresponding unitarity equation for the connected part of the process (3 particles ← 3 particles)—or "six-point function"—is shown in Fig. 3a and contains a δ-function $\delta([p_1' + p_2' - p_1]^2 - m^2)$ in the last term on the right. This implies that the six-point function itself must have a singularity, whose character can be determined as follows. If one assumes first that the singularity has the (expected) form (see Fig. 3b)

$$\langle K'|\, M\, |K\rangle \sim \langle k_1', k_2'|\, M\, |k_1, k\rangle\langle k_3', k|\, M\, |k_2, k_3\rangle D, \tag{2.6}$$

where D depends only on p^2, then substitution of this form into the unitarity equation (Fig. 3a) produces the result

$$D - D^* = 2i\delta(p^2 - m^2).$$

The $i\epsilon$-prescription (or whatever takes its place in the analyticity postulate) ensures that the correct solution of this equation is the simple pole

$$D = 1/\pi(p^2 - m^2).$$

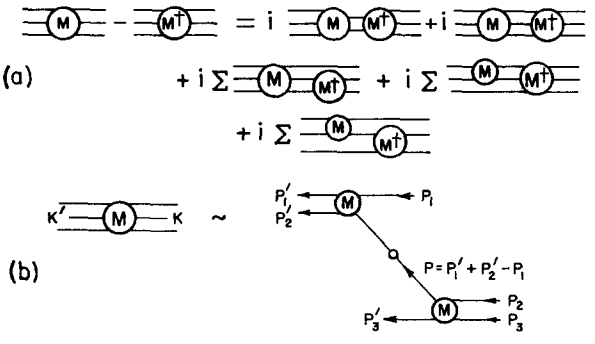

FIG. 3. (a) Unitarity for the six-point function below the four-particle threshold. The δ-function occurs in the last term on the right. (For simplicity only one type of particle is considered.) (b) The resulting singularity of the six-point function. This is, as the diagram is intended to suggest, the product of two elastic amplitudes with a simple pole term, $1/\pi(p^2 - m^2)$.

[15] D. Branson, Phys. Rev. **135**, B1255 (1964); R. J. Eden and P. V. Landshoff, Ann. Phys. (N. Y.) 31, 370 (1965).

[16] That this phase convention is significant can easily be seen; if one chooses a new basis $|p, t\rangle = z\,|p, t\rangle$ where z is any nonanalytic function of p with $|z| = 1$ and if $\langle K'|\, M\, |K\rangle$ is analytic, then obviously $(K'|\, M\, |K)$ is not.

[17] The importance of using analyticity in the momenta rather than in the invariants appears only when one considers particles with spin. For spinless particles analyticity in the momenta follows trivially from analyticity in the invariants; it was to avoid having to prove the converse result that Olive[2] proposed using analyticity in invariants as a basic postulate. However, when particles have spin it is an extremely complicated problem to construct from S-matrix elements suitable functions of the invariants. [See for example A. O. Barut, I. J. Muzinich, and D. N. Williams, Phys. Rev. **130**, 442 (1963); D. N. Williams, *The Construction of Invariant Scalar Amplitudes*, California Preprint UCRL-113 (to be published)]. Thus in general analyticity in the invariants is unnatural and inconvenient as a fundamental postulate.

To see that this solution for the singular part of the six-point function is uniquely determined by the unitarity equation (Fig. 3a), it is only necessary to suppose that the actual amplitude might have a different behavior and to consider the difference between this actual amplitude and the pole solution (2.6). This difference amplitude satisfies a unitarity equation from which the δ-function term of Fig. 3a is missing and it can be shown that such an equation implies that the difference amplitude is actually analytic[18] at $p^2 = m^2$; that is, the pole solution (2.6) for the singular part of the six-point function is unique. This argument can be reversed to show that if the six-point function has a physical region pole of this type then the unitarity equation must contain a δ-function term, which must, in turn, correspond to some particle.

Thus unitarity and the decomposition law imply that there is a one-to-one correspondence between particles (with the appropriate quantum numbers) and physical region poles of the six-point function. These poles, whose residues must factor as the product of two elastic amplitudes, correspond to the possibility that, at suitable values of the momenta, the process (3 particles ← 3 particles) can actually proceed through two successive elastic scattering (as is suggested by Fig. 3b).

The existence of physical region poles, whose residues are the products of lower amplitudes, can now be exploited[1] to deduce the properties of these lower amplitudes. For examples, Hermitian-analyticity of the four-point function (i.e. the elastic scattering amplitude) can be established[2] by considering the double pole of the eight-point function shown in Fig. 4a. This pole, which arises because the process (4 particles ← 4 particles) can proceed through three successive two-body scatterings, can be deduced from the unitarity equation in the appropriate part of the physical region; but our concern now is to substitute it into the unitarity equation [Fig. 4b] for a region where the three successive scatterings cannot occur, namely the region where $s = (p_1' + p_2' + p_3' - p_1)^2 < 4m^2$. Accordingly we first write the eight-point function as the sum of the double pole term plus a remainder (Fig. 4a). This expansion is now continued down to the region ($s < 4m^2$) where the equation of Fig. 4b is valid. Since s is just the squared total energy of the center amplitude $M(2)$, this amplitude is necessarily continued below threshold as shown in Fig. 4(c). Having made this continuation[19] we can substitute the expansion of Fig. 4(a) into the unitarity equation of Fig. 4b and continue the resulting equation back to the double pole. This process brings the outside amplitudes back to their physical values, and likewise the central amplitude $M(2)$; but the Hermitian conjugate $M^\dagger(2)$ is continued below the threshold singularity at $s = 4m^2$ [see Fig. 4(c)]. The resulting equation is shown in Fig. 4d and, since the outside amplitudes $M(1)$ and $M(3)$ can be canceled off, implies that

$$M(2_+) = M^\dagger(2_-), \tag{2.7}$$

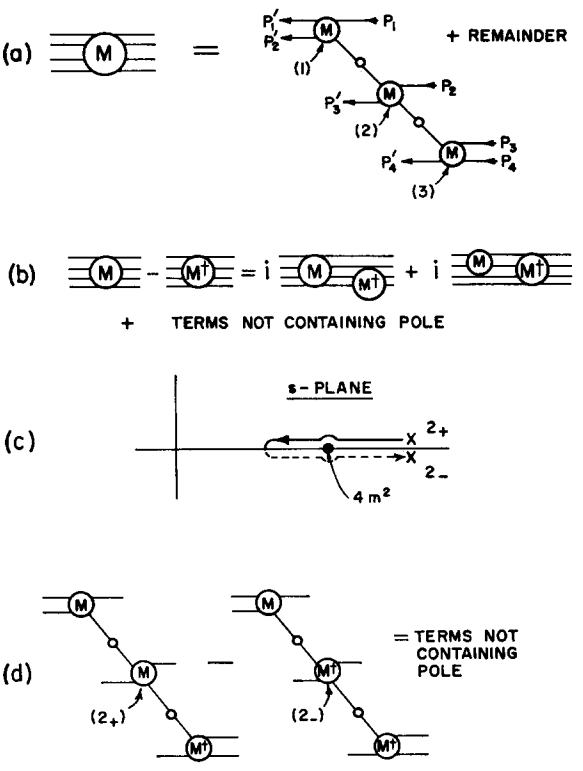

FIG. 4. (a) A double pole in the process (4 particles ← 4 particles) corresponding to three successive elastic scatterings. (b) The unitarity equation for the eight-point function in a region where the successive scatterings do *not* occur. This region is characterized by the four-momentum $(p_1' + p_2' + p_3' - p_1) = (p_2 + p_3 + p_4 - p_4')$ being inadequate to produce two particles; i.e., $(p_1' + p_2' + p_3' - p_1)^2 < 4m^2$. (c) The plane of the variable $s = (p_1' + p_2' + p_3' - p_1)^2$, the energy invariant of the elastic scattering amplitude $M(2)$. The solid arrow shows the path of continuation from the neighborhood of the triple scattering in (a) to the region of validity of the unitarity equation in (b); the broken arrow, the route by which $M^\dagger(2)$ is brought back to the double pole. (d) The equation which results from substitution of (a) into (b).

[18] The argument depends on the $i\epsilon$-prescription. Since the physical origin of this prescription is not properly understood, the argument cannot be regarded as wholly satisfactory. For further details of this and some subsequent arguments see Refs. 2, 8, and 15.

[19] One must verify that no unwanted branch points have been encountered. For this reason the argument cannot be rigorous until the complete derivation of the singularities is understood. For further details of the continuation and a discussion of anomalous thresholds see J. B. Boyling, Nuovo Cimento **33**, 1356 (1964).

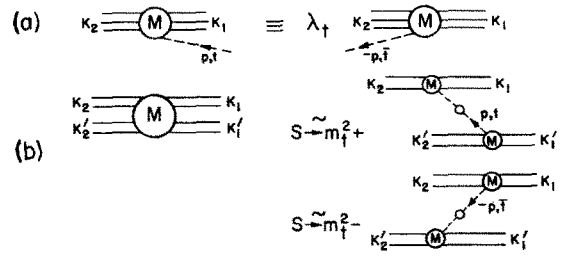

FIG. 5. (a) The crossing equation $\langle K_2|\ M\ |K_1;\ p,\ t\rangle \equiv \lambda_t\ \langle K_2;\ -p,\ \bar{t}|\ M\ |K_1\rangle$. (b) The particle ($t$) and antiparticle ($\bar{t}$) poles of the amplitude $\langle K_2,\ K_2'|\ M\ |K_1,\ K_1'\rangle$. The first occurs in that part of the physical region, where $(\sum p_2 - \sum p_1)^0 \geq m_t$; the second where $(\sum p_2 - \sum p_1)^0 \leq -m_t$.

where 2_+ and 2_- are points above and below the usual physical region cut as shown in Fig. 4(c).

The path which leads from the point 2_+ to 2_- in the space of four-momenta can, with sufficient care,[20] be seen to lead from P (on the physical sheet) back to P (on another sheet). Thus Eq. (2.7) states the identity of the functions $\langle K_2|\ M\ |K_1\rangle$ in the physical region and $\langle K_2|\ M^\dagger\ |K_1\rangle$ on another sheet.

Finally, since $\langle K_2|\ M\ |K_1\rangle$ is analytic in P_2 and P_1 whereas $\langle K_2|\ M^\dagger\ |K_1\rangle$ is analytic in P_2^* and P_1^*, Eq. (2.7) implies that, for general values of the momenta, the four-point function satisfies

$$\langle K_2|\ M\ |K_1\rangle = \langle K_2^*|\ M^\dagger\ |K_1^*\rangle; \qquad (2.8)$$

here, if

$$K = (P, T)$$

then K^* is defined as

$$K^* = (P^*, T),$$

the point P^* being reached from P by the appropriate path of continuation below threshold.

Hermitian analyticity [Eq. (2.8)] can presumably be derived for any process by considering a suitable multiple pole in some higher process. The great importance of the property is that it can be used to eliminate the Hermitian conjugate matrix elements $\langle K_2\ |M^\dagger\ |K_1\rangle$ from the unitarity equations. Thus, substitution of Eq. (2.8) into the two-particle unitarity equation of Fig. 2 gives an equation for the discontinuity of the elastic amplitude across its usual physical region cut or, more generally, the difference between its values on two different sheets.[21]

The property with which this paper is principally concerned, crossing symmetry (or the substitution rule), relates the amplitude $\langle K_2|\ M\ |K_1;\ p,\ t\rangle$ for a process (m particles $\leftarrow n+1$ particles) to the "crossed amplitude" $\langle K_2;\ -p,\ \bar{t}|\ M\ |K_1\rangle$ for the process ($m+1$ particles $\leftarrow n$ particles). [See Fig. 5(a).] This relation is proved by observing that the connected part $\langle K_2|\ M\ |K_1;\ p,\ t\rangle$ appears as a factor in the t-particle pole of any higher process ($K_2,\ K_2' \leftarrow K_1,\ K_1'$). [See Fig. 5(b).] This pole occurs at $s = (\Sigma p_2 - \Sigma p_1)^2 = m_t^2$ in that part of the physical region where $(\Sigma p_2 - \Sigma p_1)^0 \geq m_t$. Remaining close to the pole at $s = m_t^2$, one can now continue the whole amplitude to a part of the physical region where $(\Sigma p_2 - \Sigma p_1)^0 \leq -m_t$. (Such a continuation requires one to leave the physical region. It is therefore necessary to verify that it returns to the physical region, *on the physical sheet.* This cannot be rigorously done until the possibility of unwanted branch point singularities has been conclusively ruled out.) Since a pole cannot disappear, one concludes that there is a pole at $s = m_t^2$ in this new part of the physical region. Such a pole can only be interpreted as an antiparticle ($\bar{t}$) pole. [See Fig. 5(b).] In addition to proving that antiparticles must exist, this procedure establishes that the product of the two amplitudes in the particle pole is the same analytic function as the corresponding product in the antiparticle pole; that is

$$\langle K_2|\ M\ |K_1;p,\ t\rangle\langle K_2';\ p,\ t|\ M\ |K_1'\rangle$$
$$\equiv \langle K_2;\ -p,\ \bar{t}|\ M\ |K_1\rangle\langle K_2'|\ M\ |K_1';\ -p,\ \bar{t}\rangle, \qquad (2.9)$$

where, of course, the primed and unprimed variables are independent. Now, from the identity

$$A(z)B(z') \equiv \tilde{A}(z)\tilde{B}(z'),$$

where z and z' are independent, it follows that

$$A(z) \equiv \lambda\tilde{A}(z) \quad \text{and} \quad \tilde{B}(z') \equiv \lambda B(z'),$$

where λ is some constant. Thus for any fixed value of p it follows from Eq. (2.9) that

$$\langle K_2|\ M\ |K_1;\ p,\ t\rangle \equiv \lambda_t\langle K_2;\ -p,\ \bar{t}|\ M\ |K_1\rangle, \qquad (2.10)$$

and

$$\langle K_2'|\ M\ |K_1';\ -p,\ \bar{t}\rangle \equiv \lambda_t\langle K_2';\ p,\ t|\ M\ |K_1'\rangle. \qquad (2.11)$$

The factor λ_t is actually independent of p. This follows from Lorentz invariance [Eq. (2.5)] and can be shown thus: One starts with the crossing Eq. (2.10) and then applies to the right-hand side first invariance under any Lorentz transformation L, then crossing [Eq. (2.11)], and finally invariance under L^{-1}. This procedure leads to the result $\lambda_t(p) = \lambda_t(Lp)$ for any L; that is, λ_t is independent of p.

The next step is to use Hermitian analyticity,

[20] I am indebted to Dr. H. P. Stapp for help in disentangling this path.

[21] The relationship between unitarity, Hermitian-analyticity and discontinuity relations is discussed from a different point of view in J. R. Taylor, Nucl. Phys. 58, 580 (1964).

$\langle K_2| M |K_1\rangle = \langle K_2^*| M^\dagger |K_1^*\rangle$, Eq. (2.8), to prove that $|\lambda_t| = 1$. Starting with the element $\langle K_2| M |K_1; p, t\rangle$ and applying successively crossing [Eq. (2.10)], Hermitian analyticity [Eq. (2.8)], crossing [Eq. (2.11)], and finally Hermitian analyticity again, one obtains the result

$$\langle K_2| M |K_1; p, t\rangle \equiv \lambda_t \lambda_{\bar{t}}^* \langle K_2| M |K_1; p, \bar{t}\rangle. \quad (2.12)$$

One must now show that the path of continuation leading from one side of this equation to the other is such that both sides are evaluated on the same sheet. Subject to this being possible, Eq. (2.12) implies that $\lambda_t \lambda_{\bar{t}}^* = 1$ as required.

Provided the phases of the one-particle state vectors $|p, t\rangle$ have not been fixed by any convention, one can now adjust them so that the crossing phase factor λ_t in Eqs. (2.10) and (2.11) disappears. To verify this, one must distinguish two possible cases. If the particle t happens to be its own antiparticle, $t = \bar{t}$, (e.g. $t = \pi^0$) then the two Eqs. (2.10) and (2.11) are one and the same, and there is just one arbitrary phase—that of the rest-state vector $|p_0, t\rangle$—which can be adjusted. If one were to replace $|p_0, t\rangle$ by $\eta|p_0, t\rangle$, then all kets $|\ \rangle$ containing n t-particles would be replaced by $\eta^n |\ \rangle$ and all corresponding bras by $\eta^{-n} \langle\ |$. Thus in Eq. (2.10) the factor λ_t would be replaced by λ_t/η^2. By choosing $\eta = \pm\lambda_t^{\frac{1}{2}}$ one can, therefore, remove the unwanted phase factor λ_t from the crossing equation. This procedure fixes the phase of the t-particle states to within a sign.

If t is not its own antiparticle, $t \neq \bar{t}$, then there are two distinct crossing Eqs. (2.10) and (2.11), and two arbitrary phases. If one replaces $|p_0, t\rangle$ and $|p_0, \bar{t}\rangle$ by $\eta\ |p_0, t\rangle$ and $\zeta\ |p_0, \bar{t}\rangle$, respectively, it is easily seen that in *both* Eqs. (2.10) and (2.11) the phase factor λ_t is replaced by $\lambda_t/(\eta\zeta)$. This can be made equal to unity in an infinity of different ways, leaving arbitrary the relative particle–antiparticle phase η/ζ.

The final stage in the development of the basic properties of S-matrix theory is to introduce a notation which makes explicit the important fact that all processes related by crossing are described by a single analytic function, M, say. The relationship between this function and the connected parts of the processes which it describes must be fixed by some convention. A simple procedure is to regard M as the (hypothetical) amplitude obtained by crossing all particles into the final state; i.e., set

$$\langle P', T'| M |P, T\rangle = M(P', -P; T', \bar{T}). \quad (2.13)$$

This analytic function, which corresponds to Stapp's M-function[3] (or rather its connected part), has the singularities, and satisfies the unitarity and discontinuity equations, of all the channels which it describes. The fact that this one function $M(K)$ describes all processes $(P_2, T_2 \leftarrow -P_1, \bar{T}_1)$, where (K_2, K_1) is any partition of K, provides the justification for describing $M(K)$ as the n-point function, where n is the *total* number of particles involved.

III. ISOSPIN INVARIANCE IN S-MATRIX THEORY

In this section I shall consider the consequences to S-matrix theory of invariance under the isospin group[22] $SU(2, C)$. This symmetry will affect none of the considerations of the previous section until the need arises to adjust the one-particle phases in order to eliminate the phase factor λ_t from the crossing Eqs. (2.10) and (2.11). In principle such an adjustment is still perfectly possible, but in practice, in order to make use of isospin invariance, it is necessary to fix the relative phases of states within one multiplet according to some convention (for example, to ensure that the Clebsch–Gordan coefficients take their usual form). This leaves only one free phase for each multiplet and it is no longer possible to eliminate the crossing factor λ_t from all crossing equations. To see how this happens one must consider the meaning of isospin invariance.

As an essential preliminary one must introduce a suitable labelling for the particles of the theory. This means replacing the label t by a double label, $t = w_\alpha$, where w specifies the isospin multiplet (e.g. $w = \pi$, K or N) and α is the z-component of isospin T_z. The antiparticle of $t = w_\alpha$ is $\bar{t} = \bar{w}_{-\alpha}$; i.e., the particle with the opposite value of T_z in the antiparticle multiplet.

Invariance under the isospin group cannot be stated as simply as, say, invariance under rotations. Given any rotation R and any experimentally realizable initial or final state φ, one believes that there exists a realizable, rotated state $R\varphi$; rotational invariance implies simply that the probability of any process $(\psi \leftarrow \varphi)$ is the same as that of the process $(R\psi \leftarrow R\varphi)$, for any rotation R. No such statement of isospin invariance is possible. Because of the superselection rule on charge, the result of making an "isospin rotation" R_T on a physically realizable state-vector $|\varphi\rangle$ is to produce a vector $R_T |\varphi\rangle$ which in general cannot be experimentally realized. (For example, by "rotating" a proton in isospace one produces a coherent superposition of neutron and proton.) Thus it is certainly meaningless to talk of the

[22] The group of (2×2) complex unitary unimodular matrices, sometimes denoted SU_2.

transition probability for a process $(R_T\psi \leftarrow R_T\varphi)$. Instead, the statement of isospin invariance must be directly in terms of invariance properties of S-matrix elements; specifically, with any definite choice for the phases of the basis vectors, the S-matrix elements (or their connected parts) are invariant under a definite representation of the group $SU(2, C)$. If one considers an initial state $T = (c_\gamma, \cdots, d_\delta)$ and a final state $T = (a_\alpha, \cdots, b_\beta)$ then, omitting momentum variables, isospin invariance can be stated:

$$\langle a_\alpha \cdots b_\beta | M | c_\gamma \cdots d_\delta \rangle = \sum_{\alpha' \cdots \delta'} \langle a_{\alpha'} \cdots b_{\beta'} | M | c_{\gamma'} \cdots d_{\delta'} \rangle \cdot D^*_{\alpha'\alpha}(u) \cdots D^*_{\beta'\beta}(u) D_{\gamma'\gamma}(u) \cdots D_{\delta'\delta}(u), \quad (3.1)$$

where u is any element of $SU(2, C)$ and the matrices $D(u)$, one corresponding to each particle, are the well known[23] $(2T + 1)$-dimensional representatives of u, to within a phase.

Equation (3.1) states simply that the S-matrix elements are invariant under the "rotation" u in isospace. It has as an immediate consequence that isospin is conserved. This means, in particular, that the matrix elements in Eq. (3.1) are zero unless $\alpha + \cdots + \beta = \gamma + \cdots + \delta$ and that they can be decomposed, using Clebsch–Gordan coefficients, into amplitudes of definite isospin.

The phases of the elements of the matrices $D(u)$ in Eq. (3.1) are of course determined by the original choice for the phases of the vectors $|p, w_\alpha\rangle$. It is usual and convenient to readjust the latter such that the D-matrices have the usual phases.[23] However, such a readjustment means assigning definite relative phases to the particles within any multiplet and, as we shall now see, prevents one from eliminating the crossing factors λ_t.

The crossing Eq. (2.10) may be rewritten:

$$\langle K_2 | M | K_1; p, w_\alpha \rangle = \lambda_{w,\alpha} \langle K_2; -p, \bar{w}_{-\alpha} | M | K_1 \rangle, \quad (3.2)$$

where, of course, the particle types in the states K_1 and K_2 are labeled by isomultiplet and T_z, and the crossing factor $\lambda_{w,\alpha}$ depends on w and α. Now the isospin Eq. (3.1) specifies the transformations under which the matrix elements on either side of the crossing Eq. (3.2) are invariant. As far as the index α is concerned, that on the left transforms under u according to $D_{\alpha'\alpha}(u)$, (since w_α appears in the ket) and that on the right according to $D^*_{-\alpha',-\alpha}(u)$, (since $\bar{w}_{-\alpha}$ appears in the bra). This means that the whole right-hand side of Eq. (3.2) transforms according to $\lambda^{-1}_{w,\alpha'} D^*_{-\alpha',-\alpha}(u) \lambda_{w,\alpha}$. If one can prove that both sides of the crossing equation must transform in the same way under the *same element* u, it will follow that

$$D_{\alpha',\alpha}(u) = \lambda^{-1}_{w,\alpha'} D^*_{-\alpha',-\alpha}(u) \lambda_{w,\alpha}. \quad (3.3)$$

The proof that this in fact so is as follows. Starting from the matrix element $\langle K_2 | M | K_1; p, w_\alpha \rangle$ one applies successively crossing [this gives Eq. (3.2)], invariance under the isospin transformation u, [Eq. (3.1)], then crossing again, and finally invariance under u^{-1}. This last operation cancels out all D-matrices except those corresponding to the crossed particle w_α, and one obtains the result

$$\langle K_2 | M | K_1; p, w_\alpha \rangle = \sum_{\alpha', \alpha''} \langle K_2 | M | K_1; p, w_{\alpha''} \rangle \cdot D_{\alpha''\alpha'}(u^{-1}) \lambda^{-1}_{w,\alpha'} D^*_{-\alpha',-\alpha}(u) \lambda_{w,\alpha}.$$

Now conservation of T_z implies that the matrix element on the right of this equation is zero unless $\alpha'' = \alpha$. Thus the sum over α'' disappears and one finds that for any α,

$$\sum_{\alpha'} D_{\alpha\alpha'}(u^{-1}) \lambda^{-1}_{w,\alpha'} D^*_{-\alpha',-\alpha}(u) \lambda_{w,\alpha} = 1.$$

This equation implies that all diagonal elements of a certain unitary matrix are unity. Such a situation is possible only if the matrix is actually diagonal. The desired result, Eq. (3.3), follows immediately.

Equation (3.3) shows that, if the conventional assignment of isospin to particles and antiparticles is to be compatible with crossing symmetry, then the representation $D_{\alpha',\alpha}$ of $SU(2, C)$ must be similar, by the diagonal matrix of crossing-factors $\lambda_{w,\alpha}$, to the representation $D^*_{-\alpha',-\alpha}$. Fortunately this condition is satisfied, for, with the conventional choice of phases

$$D_{\alpha',\alpha}(u) = (-1)^{\alpha-\alpha'} D^*_{-\alpha',-\alpha}(u). \quad (3.4)$$

Comparing Eqs. (3.3) and (3.4) one sees that

$$\lambda_{w,\alpha} = (-1)^\alpha \lambda_w, \quad (3.5)$$

where λ_w is some phase factor independent of α.

It is important to realize that if Eq. (3.3) could not be satisfied, then crossing symmetry and isospin invariance (with the usual assignment of T_z) would be incompatible. In general, if the members of a multiplet of a conjectured symmetry group are distinguished by conserved quantum numbers (charge, in the case of the isospin group) then the result of crossing a definite particle t is to produce a second definite particle $\bar{t}$. (Thus crossing a π^+ produces a π^- and not some combination of π^+, π^0, π^-.)

[23] See, for example, M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957).

The only freedom in the crossing equations is in the phase factors λ, which, for simplicity in this discussion, we may temporarily suppose adjusted to unity. Then crossing takes the completely unambiguous form:

$$\langle \cdots | M | \cdots t \rangle = \langle \cdots \bar{t} | M | \cdots \rangle.$$

Now if one supposes that the one-particle kets $|t\rangle$ transform under some representation D and the antiparticle kets $|\bar{t}\rangle$ under some other representation $\bar{D}$, then this crossing equation implies that $D = \bar{D}^*$ or

$$\bar{D} = D^*; \tag{3.6}$$

that is, particles and antiparticles must transform under complex conjugate representations. (In general the equation need only hold within a phase.) This result is well known in field theory where, if ϕ_t is the field associated with a particle t then $\phi_t^\dagger$ is automatically the field associated with $\bar{t}$. That the result arises in S-matrix theory from a consideration of crossing symmetry is quite natural since it was through crossing that antiparticles entered in the first place.

When multiplet and antimultiplet are distinct the result (3.6) simply implies that once a representation D has been assigned to the multiplet the representation $\bar{D}$ for the antimultiplet is determined—as $\bar{D} = D^*$. Only when the particles of the antimultiplet are the same as those of the multiplet does the condition (3.6) actually place a constraint on permissible representations; in this case, $\bar{D}$ is some definite permutation of the rows and columns of D and the equation $\bar{D} = D^*$ would not be satisfied by an arbitrary representation.

In the case of isospin, $\bar{D}_{\alpha',\alpha} = D_{-\alpha',-\alpha}$ and so Eq. (3.4) shows that $\bar{D} = D^*$ *is* satisfied (within a phase factor) as required. With the usual choice of relative phases for the basis vectors $|w_\alpha\rangle$ within any multiplet, we have seen that this determines the relative phases of the corresponding crossing factors as [Eq. (3.5)]

$$\lambda_{w,\alpha} = (-1)^\alpha \lambda_w.$$

This means that one can, for example, adjust the pion phases such that crossing of a π^+ yields the π^- amplitude without any phase factor,

$$\langle \cdots | M | \cdots \pi^+ \rangle = \langle \cdots \pi^- | M | \cdots \rangle;$$

but having done so, one must accept the presence of a minus sign in the crossing equation for the π^0,

$$\langle \cdots | M | \cdots \pi^0 \rangle = -\langle \cdots \pi^0 | M | \cdots \rangle.$$

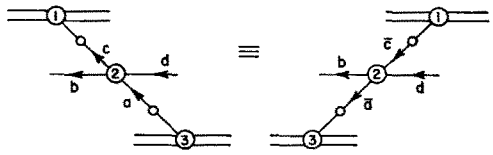

FIG. 6. The double poles of the ten-point function used to derive crossing for the four-point function. Since the two double poles are the same analytic function and the crossing properties of the five-point functions, (1) and (3), are known, those of the four-point function (2) follow in the form $\langle bc | M | da \rangle = (\lambda_a/\lambda_c)\langle b\bar{a} | M | d\bar{c} \rangle$.

The alternative procedure is of course to abandon the conventional relative phases within isospin multiplets. One could then achieve crossing symmetry without any phase factors for *all* particles, but at the expense of changing the phases of the D-matrices and Clebsch–Gordan coefficients.

An important application of crossing symmetry is to the four-point function. This needs special treatment because at least two of its lines must be crossed to reach a physically meaningful amplitude. The simplest procedure is to consider the double pole of the ten-point function shown in Fig. 6. Since the crossing properties of the two five-point functions have been established, those of the four-point function follow immediately. If one considers, for example, π-π elastic scattering, then in the operation of crossing one pion forwards and another backwards the crossing-factor λ_π of Eq. (3.5) cancels itself out and one finds that

$$\langle p_2, \pi_\beta; p_3, \pi_\gamma | M | p_1, \pi_\alpha; p_4, \pi_\delta \rangle$$
$$= (-1)^{\alpha-\gamma} \langle p_2, \pi_\beta; -p_1, \pi_{-\alpha} | M | -p_3, \pi_{-\gamma}; p_4, \pi_\delta \rangle; \tag{3.7}$$

(recall that the indices $\alpha \cdots \delta$ take the values $+1$, 0, -1 according as the pion is π^+, π^0, π^-). Using Clebsch–Gordan coefficients, the two sides of this equation can be expanded in terms of amplitudes for definite total isospin—on the left-hand side

$$\langle p_2, p_3 | M^T | p_1, p_4 \rangle = A^T, \quad \text{say},$$

and on the right

$$\langle p_2, -p_1 | M^T | -p_3, p_4 \rangle = \tilde{A}^T, \quad \text{say},$$

with $T = 0$, 1 or 2. Eq. (3.7) then yields nineteen equations for the three functions A^T in terms of the $\tilde{A}^T$. The isospin transformation properties (3.1) ensure, of course, that only three of these nineteen equations are independent. Picking out any three

independent equations one can easily solve to give

$$A^T = \sum_{T'=0}^{2} \beta_{TT'} \tilde{A}^{T'} \qquad (T = 0, 1, 2),$$

where β is the usual π-π crossing matrix.[24]

This demonstration that the isospin crossing matrices can be calculated without reference either to some postulated form of crossing symmetry or to a hypothetical isoscalar Lagrangian is, from a practical standpoint, perhaps the most important conclusion of this section.

IV. PARTICLES WITH SPIN; SPINOR BASIS VECTORS

In this and the next section I shall discuss particles with spin and use the techniques of exploiting physical region poles to prove crossing symmetry and related analytic properties. The introduction of spin presents two basic problems. The first of these arises in the following way. Corresponding to any particle, there is a pole in all appropriate connected parts. If the particle is spinless, this pole has the simple form of the product of two lower amplitudes with a pole denominator; however, if the particle has spin, then the simple product is replaced by a sum of products, one term for each spin state. This obviously makes for some difficulty in disentangling the crossing equation.

The second problem arises from the complicated transformation properties of spin states under the Lorentz group. As discussed in the previous section on isospin, the crossing phase factors (or, in the case of spin, crossing matrices) are determined by the requirement that both sides of the crossing equation transform in the same way under the appropriate group (isospin group in the last section, Lorentz group in this). Given any set of basic states, with known transformation properties, one could derive the appropriate form of crossing symmetry for the corresponding matrix elements. The procedure used here is to choose basic "spinor" states with particularly simple transformation properties such that the matrix elements on either side of the crossing equation automatically transform in the same way. (Such matrix elements are usually called M-functions.) Using these matrix elements it will be a relatively simple matter to deduce crossing symmetry.

In this section I shall deal with this latter problem—the construction of basis states with appropriate transformation properties. I shall first introduce the familiar and physically useful helicity states[25] and from these construct the required spinor states. In the next section, using the corresponding spinor matrix elements, I shall deduce crossing symmetry and some related analytic properties.

The helicity state $k = (p, \alpha, t)$ of a single particle is identified by three variables, p the four-momentum, α the helicity, which has $(2s + 1)$ distinct values, and t the particle type. Although the state (p, α, t) is uniquely determined, it is still necessary to specify the phase of the corresponding state vector $|k\rangle = |p, \alpha, t\rangle$. This is usually done by relating the general state-vector $|p, \alpha, t\rangle$ to a standard rest-state vector $|p_0, \alpha, t\rangle$ by a Lorentz transformation, chosen to map p_0 onto p while leaving the helicity unaltered. The $(2s + 1)$ rest states $|p_0, \alpha, t\rangle$ are chosen with spin quantised along the z axis and with the usual phase relationships (i.e., such that rotations are represented by the usual matrices D^s). The moving states (p, α, t) are now generated from (p_0, α, t) by an acceleration along the z axis followed by a prescribed rotation.

To create a satisfactory notation it is necessary to recall that the representations of the Lorentz group corresponding to half-integral spin are two-valued. To avoid ambiguity it is therefore necessary to label Lorentz transformations not by elements of the Lorentz group but by elements of its universal covering group $SL(2, C)$—the group of complex, unimodular 2×2 matrices. Accordingly a rotation through Euler angles (φ, θ, ψ) is labeled by 2×2 unitary matrix

$$u(\varphi, \theta, \psi) = \exp(-\tfrac{1}{2}i\varphi\sigma_3) \exp(-\tfrac{1}{2}i\theta\sigma_2) \exp(-\tfrac{1}{2}i\psi\sigma_3),$$

while a velocity transformation of magnitude and direction $\mathbf{v}$ is identified by the Hermitian matrix

$$h(\mathbf{v}) = \exp(\tfrac{1}{2}\lambda\hat{\mathbf{v}}\cdot\boldsymbol{\sigma}),$$

where $\tanh\lambda = |\mathbf{v}|$, $\hat{\mathbf{v}}$ is the unit vector in the direction of $\mathbf{v}$ and the matrices $\boldsymbol{\sigma}$ are the usual Pauli matrices. [It is convenient—though of no physical significance—that $h(\mathbf{v})$ can be written

$$h(\mathbf{v}) = (p\cdot\sigma/m)^{\frac{1}{2}} \equiv [(p^0 + \mathbf{p}\cdot\boldsymbol{\sigma})/m]^{\frac{1}{2}},$$

where p is any four-momentum corresponding to the velocity $\mathbf{v}$ and to mass m.]

Every Lorentz transformation [or rather every matrix g in $SL(2, C)$] can be expressed uniquely as the product of a velocity transformation and a rotation, $g = hu$, where $h = +(gg^\dagger)^{\frac{1}{2}}$ and $u = h^{-1}g$.

[24] G. F. Chew, *S-matrix Theory of Strong Interactions* (W. A. Benjamin, Inc., New York, 1961), p. 64.

[25] M. Jacob and G. C. Wick, Ann. Phys. (N. Y.) 7, 404 (1959).

The actual Lorentz transformation (i.e., the 4×4 real matrix) corresponding to an element g in $SL(2, C)$ can be written $L(g)$. Similarly the corresponding unitary transformation in the Hilbert space of state vectors can be written $U(g)$.

The particular transformation used to generate the basis vector $|p, \alpha, t\rangle$ from the rest-state vector $|p_0, \alpha, t\rangle$ is chosen as follows. The rest-state is first accelerated to the required speed by the element $h(v\hat{\mathbf{z}})$, where $v = |\mathbf{p}|/p^0$; it is then rotated to the required direction by $u(\varphi, \theta, -\varphi)$, φ and θ being the aximuthal and polar angles of $\mathbf{p}$. The complete transformation is denoted

$$\nu(p) = u(\varphi, \theta, -\varphi)h(v\hat{\mathbf{z}}) = h(\mathbf{v})u(\varphi, \theta, -\varphi), \qquad (4.1)$$

and the state-vector $|p, \alpha, t\rangle$ is defined as

$$|p, \alpha, t\rangle = U(\nu(p))\ |p_0, \alpha, t\rangle.$$

Having defined a set of basis vectors one must determine its transformation properties under the Lorentz group. Obviously the particle type t is invariant and, under a transformation g, the four-momentum p maps onto $L(g)p$;

$$p \xrightarrow{g} L(g)p = p_g, \quad \text{say.} \qquad (4.2)$$

To determine how the helicity changes it is only necessary to note that any g in $SL(2, C)$ can be expressed as the product of three transformations: first $\nu^{-1}(p)$ which reduces p to rest, then a rotation, $u(g, p)$, and finally $\nu(p_g)$ which produces the required final momentum p_g. Thus

$$g = \nu(p_g)u(g, p)\nu^{-1}(p). \qquad (4.3)$$

This equation actually defines $u(g, p)$, which, since it leaves the rest-momentum unaltered, is clearly a rotation. The outside two factors in this expansion of g leave the helicity unaltered while the rotation $u(g, p)$ acts on the rest state $|p_0, \alpha, t\rangle$ and is therefore represented by the usual matrices D^s. Thus under g,

$$|p, \alpha, t\rangle \xrightarrow{g} U(g)\ |p, \alpha, t\rangle = \sum_\beta |p_g, \beta, t\rangle D^s_{\beta\alpha}[u(g, p)], \qquad (4.4)$$

where p_g and $u(g, p)$ are defined by Eqs. (4.2) and (4.3).

Since the multiparticle state-vectors are constructed as products of the one-particle vectors $|p, \alpha, t\rangle$, it follows that they transform under a tensor product of the matrices in Eq. (4.4). However, since each matrix D^s depends not only on g but also on the momentum p of the corresponding particle, this transformation law is rather inconvenient. The primary motivation in constructing the so-called spinor states is to avoid the momentum dependence of the transformations of helicity states. To do this one must obviously factor out of the matrix $D^s[u(g, p)]$ in Eq. (4.4) its separate dependence on the momenta p and p_g. Now according to the definition (4.3) of $u(g, p)$

$$D^s[u(g, p)] = D^s[\nu^{-1}(p_g)g\,\nu(p)], \qquad (4.5)$$

but since the matrices D^s represent only rotations—i.e., $SU(2, C)$—it is impossible to use this factorisation without first extending the representation D^s onto the whole Lorentz group—i.e., $SL(2, C)$. It is therefore necessary to recall briefly some properties of the finite-dimensional representations of the Lorentz group.

I shall adopt as the basic $(2s + 1)$-dimensional representation of $SL(2, C)$ that usually denoted $D^{0,s}$, which is defined as follows: If

$$g = \exp(\tfrac{1}{2}\mathbf{a}\cdot\boldsymbol{\sigma}) \exp(-\tfrac{1}{2}i\mathbf{b}\cdot\boldsymbol{\sigma}) \in SL(2, C),$$

where $\mathbf{a}$ and $\mathbf{b}$ are real, then

$$D^{0,s}(g) = \exp(\mathbf{a}\cdot\mathbf{J}) \exp(-i\mathbf{b}\cdot\mathbf{J}),$$

where the three matrices $\mathbf{J}$ are the usual $(2s + 1)$-dimensional angular momentum matrices, defined with the usual conventions (J_1 and J_3 real, J_2 pure imaginary). It will be seen that, with this definition, $D^{0,\frac{1}{2}}$ coincides with $SL(2, C)$.

A second representation, usually denoted $D^{s,0}$, is defined as[26]

$$D^{s,0}(g) = D^{0,s}(g^{\dagger^{-1}}) = D^{0,s}(g)^{\dagger^{-1}}.$$

The importance of these two representations is that on the subgroup of rotations both coincide with the original representation D^s;

$$D^s(u) = D^{0,s}(u) = D^{s,0}(u), \qquad [u \in SU(2, C)].$$

Thus either can be used as an extension onto $SL(2, C)$ of the representation D^s of $SU(2, C)$.

Two further important representations of $SL(2, C)$—it is easily seen that they *are* representations—are $D^{0,s}(g^{T^{-1}})$ and $D^{s,0}(g^{T^{-1}}) \equiv D^{0,s}(g^*)$.

Any set of measurable quantities which transform under $SL(2, C)$ according to one of these four representations is called a spinor. The four different kinds of spinor, whose importance will appear shortly, are distinguished from one another by the writing of their indices. Thus, a spinor which transforms according to $D^{0,s}(g)$ is written with a raised index,

$$x^\alpha \xrightarrow{g} \sum_\beta x^\beta\ D^{0,s}_{\beta\alpha}(g), \qquad (4.6)$$

[26] It is a convenient feature of the phase conventions used that, for either $D^{0,s}$ or $D^{s,0}$, $D(g^T) = D(g)^T$, $D(g^*) = D(g)^*$, and hence, $D(g^\dagger) = D(g)^\dagger$.

while one transforming according to $D^{0,s}(g^{T^{-1}})$ is written with a lowered index,

$$y_\alpha \xrightarrow{g} \sum_\beta y_\beta D^{0,s}_{\beta\alpha}(g^{T^{-1}}). \tag{4.7}$$

The significance of the lower and upper indices is that the contraction of one lower with one upper index is an invariant—as is easily verified using Eqs. (4.6) and (4.7).

It is convenient to have a matrix to raise and lower spinor indices. This is easily constructed, for if

$$c = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in SL(2, C),$$

then for any g in $SL(2, C)$ it is easily seen that[27]

$$c^{-1}gc = g^{T^{-1}}.$$

Thus the matrix $D^{0,s}(c) = C$, say, acts as a lowering matrix,

$$\sum_\beta x^\beta C_{\beta\alpha} = x_\alpha,$$

where if x^β transforms according to $D^{0,s}(g)$ then x_α transforms under $D^{0,s}(g^{T^{-1}})$. Similarly C^{-1} acts as a raising matrix;

$$\sum_\beta y_\beta C^{\beta\alpha} = y^\alpha,$$

where the $C^{\beta\alpha}$ are elements of $C^{-1} = C^T = (-1)^{2s}C$.

If one constructs (as we shall) a ket vector which transforms according to $D^{0,s}(g)$ then the corresponding bra will transform according to $D^{0,s}(g^*)$. Spinors which transform under this complex conjugate representation are identified traditionally by upper dotted indices;

$$z^{\dot\alpha} \xrightarrow{g} \sum_{\dot\beta} z^{\dot\beta} D^{0,s}_{\dot\beta\dot\alpha}(g^*).$$

For exactly the same reasons as in the case of undotted indices, one needs to consider lower dotted indices, transforming according to $D^{0,s}(g^{\dagger^{-1}})$. The same matrix C can be used to raise and lower dotted indices.

We are now in a position to construct the required spinor basis vectors. Returning to the transformation law (4.4) for helicity basis vectors, we can now use the identity $D^s(u) = D^{0,s}(u)$ to factor the matrix $D^s[u(g, p)]$ as discussed above Eq. (4.5). Having written $D^s[u(g, p)]$ as the product of three matrices $D^{0,s}$, we can take the last of these across to the left-hand side and find that the new basis vector

$$|p, {}^\alpha, t\rangle = \sum_\beta |p, \beta, t\rangle D^{0,s}_{\beta\alpha}[\nu^{-1}(p)] \tag{4.8}$$

transforms under Lorentz transformations like a spinor;

$$|p, {}^\alpha, t\rangle \xrightarrow{g} U(g)\,|p, {}^\alpha, t\rangle = \sum_\beta |p_g, {}^\beta, t\rangle D^{0,s}_{\beta\alpha}(g).$$

It should be noted that since the matrices $D^{0,s}$ are not in general unitary the spinor basis vectors defined by Eq. (4.8) are not orthonormal. (They are of course complete since the $D^{0,s}$ are nonsingular.)

We have now achieved our first object, namely the construction of basis vectors with simple (momentum independent) transformation properties. However our principal goal is to find basis vectors such that the two sides of the crossing equation transform in the same way. Now, crossing symmetry relates a connected part of the form $\langle K_2|\ M\ |K_1; p, \alpha, t\rangle$ to another of the form $\langle K_2; -p, \alpha, \bar t|\ M\ |K_1\rangle$; thus the basis vectors must be such that the *ket* labeled (p, α, t) transforms in the same way as the *bra* labeled $(-p, \alpha, \bar t)$. This situation is impossible with only one type of basis vector, because corresponding bra and ket always transform oppositely—one according to D, the other to D^*. Thus one is forced to introduce two types of basis vector. This can be done in two ways. The conventional method has been to use different basic kets for particle and antiparticle; the ket $|t\rangle$ is chosen (as above) to transform under $D^{0,s}$, while $|\bar t\rangle$ is made to transform under D^{0,s^*}. This procedure ensures that $|t\rangle$ and $\langle\bar t|$ transform in the same way. But it is open to two objections: if the particle under consideration is its own antiparticle (and has spin greater than zero)—e.g., ρ-meson—the procedure is impossible; secondly, since neither basis is orthonormal, it introduces some irrelevant "phase-space factors" into the unitarity equation.

The alternative procedure, which I shall use, is to introduce two sorts of basis vector—denoted $|k)$ and $|k]$—for *all* particles. The first of these was defined in Eq. (4.8) so that the ket $|p, {}^\alpha, t)$ transforms according to $D^{0,s}$. The second will now be defined so that the bra $[p, {}^\alpha, t|$ transforms in the same way. The defining equation, analogous to Eq. (4.8) is

$$|p, {}_\alpha, t] = \sum_\beta |p, \beta, t)\, D^{0,s}_{\beta\alpha}[\nu^\dagger(p)]. \tag{4.9}$$

[This is arrived at in the same way as Eq. (4.8) but using the identity $D^s(u) = D^{s,0}(u)$ to factor the helicity transformation matrix.] This ket transforms according to $D^{0,s}(g^{\dagger^{-1}})$ and thus the corresponding bra $[p, {}_\alpha, t|$ transforms according to $D^{0,s}(g^{T^{-1}})$. By raising its index one obtains a bra

$$[p, {}^\alpha, t| = \sum_\beta [p, {}_\beta, t|\ C^{\beta\alpha}$$

which transforms under $D^{0,s}(g)$, as required.

[27] When $g = u$ is in $SU(2, C)$ this equation becomes $c^{-1}uc = u^*$ and is equivalent to Eq. (3.4) used in connection with isospin crossing.

It can easily be shown, using the definitions (4.8) and (4.9) of $|k)$ and $|k]$ and definition (4.1) of $\nu(p)$, that the relationship between the two types of basis vector is

$$|p, \dot{\alpha}, t] = \sum_{\beta} |p, {}^{\beta}, t)\, D^{0,s}_{\beta\alpha}(p\cdot\sigma/m).$$

It can likewise be shown that the bases are *reciprocal*; that is, their scalar product is

$$[p', \alpha', t'|\, p, {}^{\alpha}, t) = p^0\delta_3(\mathbf{p}' - \mathbf{p})\,\delta_{\alpha'\alpha}\,\delta_{t't},$$

and the completeness sum takes the usual form; i.e.,

$$\sum_{\alpha} |p, \alpha, t)(p, \alpha, t| = \sum_{\alpha} |p, {}^{\alpha}, t)[p, \alpha, t|.$$

This means, for example, that the unitarity condition for matrix elements $[P', {}_{A'}, T'|\ M\ |P, {}^{A}, T)$ takes the same simple form as for matrix elements between an ordinary orthonormal set.

The multiparticle basis vectors are formed, as usual, from tensor products of the one-particle vectors; for example,

$$|K) = |P, {}^{A}, T) = |p_1, {}^{\alpha_1}, t_1) \otimes \cdots \otimes |p_n, {}^{\alpha_n}, t_n).$$

They therefore transform under the product $\otimes D^{0,s} = \mathfrak{D}$ say, with one factor for each particle; that is

$$|P, {}^{A}, T) \xrightarrow{g} U(g)\ |P, {}^{A}, T) = \sum_{B} |P_g, {}^{B}, T)\ \mathfrak{D}_{BA}(g), \tag{4.10}$$

where

$$P_g = L(g)P = [L(g)p_1, \cdots, L(g)p_n],$$

and

$$\mathfrak{D}_{BA}(g) = D^{0,s_1}_{\beta_1\alpha_1}(g) \cdots D^{0,s_n}_{\beta_n\alpha_n}(g).$$

The whole object of studying the transformation properties of the basis vectors is to be able to use the physical condition of Lorentz invariance. This condition implies that,

$$\langle\varphi|\ S\ |\psi\rangle = \langle\varphi|\ U^{\dagger}(g)SU(g)\ |\psi\rangle$$

for any physically realizable states φ and ψ and for any Lorentz transformation g. If one considers matrix elements $[K'|\ S\ |K)$ between the spinor basis vectors and substitutes the transformation law (4.10), the condition of Lorentz invariance becomes the identity[28]

$$[P', {}^{A'}, T'|\ M\ |P, {}^{A}, T) = \sum_{B',B} [P'_g, {}^{B'}, T'|\ M\ |P_g, {}^{B}, T)\ \mathfrak{D}_{B'A'}(g)\ \mathfrak{D}_{BA}(g), \tag{4.11}$$

for all g in $SL(2, C)$.

[28] The identity is obviously true both for S-matrix elements $[K'|\ S\ |K)$ and for their connected parts $[K'|\ M\ |K)$.

V. PARTICLES WITH SPIN; CROSSING SYMMETRY

It is the spinor matrix elements $[K_2|\ M\ |K_1)$ whose analytic properties are regarded as fundamantal in S-matrix theory. Just as with spinless particles, a number of their properties can be deduced from the physical unitarity condition. As already mentioned, this condition takes a particularly simple form in terms of the basis vectors $|P, {}^{A}, T)$ and $[P, {}_{A}, T|$. In fact, apart from their being summed over intermediate spin states, the unitarity equations have exactly the same form as for spinless particles, and one can use exactly the same diagrammatic representation. Thus, for example, Fig. 2 can still be used to represent unitarity for the four-point function, with the understanding that the phase-space integral on the right is of the sum

$$\sum_{\alpha,\beta} [K_2|\ M\ |p, {}^{\alpha}, t; q, {}^{\beta}, t)[p, \alpha, t; q, \beta, t|\ M^{\dagger}\ |K_1).$$

Just as before, unitarity for the six-point function, as shown in Fig. 3a, implies the presence of a singularity. If, without loss of generality, this singularity is written as

$$\{\sum_{\alpha} [k'_1; k'_2|\ M\ |k_1; p, {}^{\alpha}, t)[p, \alpha, t; k'_3|\ M\ |k_2; k_3)\}D,$$

[c.f. Eq. (2.6) and Fig. 3b], then just as before one can show that D must be a simple pole, $D = 1/\pi(p^2 - m^2)$. Thus single particle poles can be represented in the same way as for spinless particles, with, once again, the understanding that the particle's spin states are summed.

With the proof of Hermitian analyticity, one encounters the first difficulty. Starting from the double pole of Fig. 4a one can easily derive the equation shown in Fig. 4d. However, this equation now contains two sums over spin states; written schematically it is

$$\sum_{\alpha,\beta} M_{\alpha}(1)\{M_{\alpha\beta}(2_+) - M^{\dagger}_{\alpha\beta}(2_-)\}M^{\dagger}_{\beta}(3) = 0.$$

(Only the summed indices are shown explicitly.) To be able to factor off the outside two amplitudes one must know that the $(2s + 1)$ functions $M_{\alpha}(1)$ are actually independent, and likewise, the functions $M^{\dagger}_{\beta}(3)$. We are thus forced to make the additional (very plausible) assumption that *for every particle t of spin s* (and mass greater than zero) *there is at least one process with t in the initial state, and another with t in the final state, for which the corresponding* $(2s + 1)$ *amplitudes are independent.* If we use these amplitudes for $M(1)$ and $M(3)$, these can be cancelled off to give Hermitian-analyticity in a

form analogous to Eq. (2.8). (Just as before, care is needed to follow the path connecting the points 2_+ and 2_- in the space of four-momenta; but I shall not examine these details here.) As before, substitution of Hermitian-analyticity into the unitarity equations gives discontinuity equations which relate the values of the function $[K_2|\ M\ |K_1)$ on two different sheets.

The main purpose of this section is to prove crossing symmetry for particles with spin. To do this one considers, as before, the single particle poles of the eight-point function (or higher) shown in Fig. 5b. Both poles are now sums over intermediate spin states and the conclusion to be drawn from their equality is [c.f. Eq. (2.9)]

$$\sum_\alpha [K_2|\ M\ |K_1; p, {}^\alpha, t)[p, {}_\alpha, t; K_2'|\ M\ |K_1') = \sum_\alpha [K_2; -p, {}_\alpha, \bar{t}|\ M\ |K_1)[K_2'|\ M\ |-p, {}^\alpha, \bar{t}; K_1'). \tag{5.1}$$

To exploit this identity one needs the following theorem (whose proof is quite straightforward and is omitted): If

$$\sum_\alpha A_\alpha(z)B_\alpha(z') = \sum_\alpha \tilde{A}_\alpha(z)\tilde{B}_\alpha(z'),$$

where z and z' are independent variables and both of the sets $\{A_\alpha(z)\}$ and $\{B_\alpha(z)\}$ are linearly independent sets of functions; then

$$\text{and} \qquad \begin{aligned} A_\alpha(z) &= \sum_\beta \tilde{A}_\beta(z)\Lambda_{\beta\alpha}, \\ \tilde{B}_\alpha(z') &= \sum_\beta \Lambda_{\alpha\beta}B_\beta(z'), \end{aligned} \tag{5.2}$$

where the square matrix Λ is nonsingular and independent of z and z'.

This theorem can obviously be used to separate the identity (5.1). To do so one must first apply it to the case where both sets of $(2s + 1)$ amplitudes on the left-hand side of the identity are independent. Having established the result for these, one can obtain the same result for any amplitude by coupling it to one of the original two. The result, which follows directly from Eqs. (5.1) and (5.2) is [c.f. Eqs. (2.10) and (2.11)]

$$[K_2|\ M\ |K_1; p, {}^\alpha, t) = \sum_\beta [K_2; -p, {}^\beta, \bar{t}|\ M\ |K_1)\Lambda^t_{\beta\alpha}(p), \tag{5.3}$$

and

$$[K_2'|\ M\ |-p, {}_\alpha, \bar{t}; K_1') = (-1)^{2st} \sum_\beta \Lambda^t_{\alpha\beta}(p)[p, {}_\beta, t; K_2'|\ M\ |K_1'), \tag{5.4}$$

where the matrix $\Lambda^t(p)$ depends only on the momentum p and the particle type t. In the second of these equations the factor $(-1)^{2st}$ arises because it is clearly desirable in both equations to have all indices of the same type [upper in (5.3) and lower in (5.4)]; to achieve this it was necessary to switch indices on the right-hand side of Eq. (5.1) using the identity

$$\sum_\alpha x_\alpha y^\alpha = (-1)^{2s} \sum_\alpha x^\alpha y_\alpha.$$

To bring the crossing Eqs. (5.3) and (5.4) into the same simple form as those for spinless particles [Eqs. (2.10) and (2.11)], two tasks remain to be done. Firstly, one must show that the spurious factor $(-1)^{2s}$ in Eq. (5.4) indeed is spurious, and secondly, the matrix $\Lambda^t(p)$ must be shown to be just a constant phase factor times the unit matrix.

The factor $(-1)^{2st}$ contributes an extra minus sign to Eqs. (5.4) if the crossed particle t is a Fermion, but has no effect if t is a Boson. This observation serves to remind one that up to this point the question of the order of variables has been ignored and that, in the case of Fermions, this order is significant.[29] To see that the variables in Eqs. (5.1)–(5.4) have been written in the correct order it is necessary to review the derivation of the poles which give Eq. (5.1). These poles were shown in Fig. 5b) and arise from terms in the unitarity equation for the function $[K_2, K_2'|\ M\ |K_1, K_1')$. Thus the t-particle pole, corresponding to the left-hand side of Eq. (5.1), is contributed by the intermediate state $(K_1; p, \alpha, t; K_2')$ in the unitarity equation. The relevant term in the completeness sum is

$$\sum_\alpha |K_1; p, {}^\alpha, t; K_2')[K_1; p, {}_\alpha, t; K_2'|,$$

(The variables here can obviously be written in any order so long as they are in the same order in both bra and ket.) and, with the intermediate state written in this way, it is easy to see that the t-particle pole has the form shown on the left-hand side of Eq. (5.1). Similarly the $\bar{t}$-antiparticle pole is easily seen to take the form shown on the right-hand side. Once given Eq. (5.1) in the form shown, it is clear that the crossing equations come out with the variables ordered as in Eqs. (5.3) and (5.4)—namely with the (p, α, t) variables to the *right* of K_2 and K_1 in Eq. (5.3) but to the *left* of K_2' and K_1' in Eq. (5.4). Now if t is a Boson then the (p, α, t) variable can be moved to the right in Eq. (5.4) without changing anything and, of course, $(-1)^{2s} = +1$. On the other

[29] I am indebted to Dr. H. P. Stapp for emphasising this point.

hand, if t is a Fermion there is a change of sign for each interchange with any other Fermion variable. Now, in any process, the number of Fermions *modulo two* is conserved and so the states K_2' and K_1' must differ by an odd number of Fermions. Thus, if one moves the (p, α, t) variable to the right on both sides of Eq. (5.4), one introduces an extra minus sign and cancels the unwanted $(-1)^{2s}$.

Thus by always writing the crossing equations in a standard form with the crossed variable on the right of the state in which it occurs one can dispose of the spurious factor $(-1)^{2st}$ which occurs in Eq. (5.4).

To prove that $\Lambda^t(p)$ is a multiple of the unit matrix one performs the following operations,[30] choosing a process for which the $(2s + 1)$ amplitudes are independent. Starting with the crossing Eq. (5.3) one applies to its right-hand side firstly invariance under the Lorentz transformation g [Eq. (4.11)], then crossing back again, and finally invariance under g^{-1}. The result is

$$\Lambda^t(p_g) = D^{0,s}(g)\Lambda^t(p)\, D^{0,s}(g^{-1}) \qquad (5.5)$$

for any g and p, where as usual p_g is defined as

$$p_g = L(g)p.$$

Now if p is chosen to be the rest momentum $p_0 = (m, \mathbf{0})$ and $g = u$ to be any rotation, then, since p_0 is unchanged by rotations, Eq. (5.5) becomes

$$\Lambda^t(p_0) = D^s(u)\Lambda^t(p_0)D^s(u^{-1}).$$

The only matrices which commute with all rotation matrices are multiples of the unit matrix and so $\Lambda^t(p_0) = \lambda_t I$. Next, still putting $p = p_0$ but allowing g in Eq. (5.5) to range over all possible Lorentz transformations, one finds that

$$\Lambda^t(p) = \lambda_t I, \qquad \text{(for all } p\text{)}.$$

Finally, just as with spinless particles, Hermitian analyticity ensures that $|\lambda_t| = 1$.

The crossing Eqs. (5.3) and (5.4) can now be rewritten in the simple form:

$$[K_2|\ M\ |K_1; p, {}^{\alpha}, t) = \lambda_t[K_2; -p, {}^{\alpha}, \bar{t}|\ M\ |K_1),$$

and (5.6)

$$[K_2|\ M\ |K_1; p, {}^{\alpha}, \bar{t}) = \lambda_t[K_2; -p, {}^{\alpha}, t|\ M\ |K_1),$$

where λ_t depends only on the particle type t and $|\lambda_t| = 1$. Just as in the case of spinless particles, the phase factor λ_t can, in the absence of any conventional phase requirements, be eliminated by readjusting the one particle phase.

The crossing equations (5.6) establish, for the case of particles with spin, that all processes related by crossing are described by a single analytic function. A notation for this single function can be set up exactly as in Eq. (2.13). With all crossing factors λ_t adjusted to unity, the M-function is defined by the equation

$$[K'|\ M\ |K) = M(K', \tilde{K}), \qquad (5.7)$$

where if $K = (P, {}^{A}, T)$, then

$\tilde{K} = (-P, {}^{A}, \bar{T})$ *with the variables written in the opposite order*. The reversal of the order of variables in K is necessary to conform with the requirement that variables be crossed from and to the right-hand position. Once defined by Eq. (5.7), the M-function can, however, be written with the variables in any order by the usual rules for the interchange of variables.

The definition of the "M-function" for particles with spin completes this discussion of the basic properties of the S-matrix theory of particles with spin. Before concluding, however, it is perhaps worth adding a brief remark about crossing for the four-point function. This can be easily treated as before using the double pole of Fig. 6, although care is now needed with the order of variables. The result is (omitting all but particle-type variables):

$$[bc|\ M\ |da) = \epsilon_{ac}(\lambda_a/\lambda_c)[b\bar{a}|\ M\ |d\bar{c}),$$

where ϵ_{ac} is the sign change resulting from interchanging the particles a and c; that is, $\epsilon_{ac} = -1$ if both a and c are Fermions, $+1$ otherwise. This result is easily understood heuristically as follows. The crossing Eqs. (5.6) hold subject to the convention that the crossed variable is moved out of and into the right-hand position. Now crossing of the four-point function can be thought of as made up of two successive crossings, the first taking one to a hypothetical (3-particle ← 1-particle) process. However, having made this first crossing both the particles c and $\bar{a}$ are in the final state with $\bar{a}$ on the right. Before the c-particle can be crossed into the initial state it must be moved to the right of the $\bar{a}$-particle. This interchange introduces the factor ϵ_{ac}.

VI. CONCLUSION

In this paper I have described how the Gunson–Olive technique of exploiting physical region poles can be used to establish some fundamental properties in the S-matrix theory of particles with spin or iso-

[30] The argument used here is similar to that used to determine the dependence of crossing factors on isospin. Eq. (5.5) is analogous to Eq. (3.3).

spin. In these cases this technique can contribute an understanding to the following:

(1) The constraints imposed by crossing symmetry on the assignment of symmetry groups to particles and their antiparticles;
(2) The physical origin of the isospin crossing matrices and their calculation in practice;
(3) The correct generalization to include spin of such properties as single-particle poles, Hermitian-analyticity, and discontinuity relations;
(4) The reasons for constructing the spinor basis vectors and the form of crossing symmetry for the corresponding spinor matrix elements (M functions).

I have tried to describe the subject as concisely as possible while at the same time making clear those points where the theory is still inadequate. It is clear that the most important obstacle at present is the inadequate understanding of the analyticity postulate. As discussed in the introduction and Sec. II, not only is the physical origin of this postulate obscure, but its precise formulation is also unknown. The hope is that a formulation can be found such that, from the five basic postulates, one can deduce: (i) the complete singularity structure of the S-matrix (at least in principle) and (ii) such fundamental properties as crossing symmetry, Hermitian analyticity, etc. At present (i) cannot be conclusively established without prior knowledge of (ii), nor can (ii) without prior knowledge of (i). The nature of this impasse would suggest that its solution may lie in a self-consistency argument whereby (i) and (ii) are simultaneously built up step by step.[31] It remains to be seen whether this can be done.

ACKNOWLEDGMENT

I am indebted to several people for help and encouragement with this work. I should mention Dr. P. V. Landshoff, whose lecture series at Cambridge introduced me to this subject; Dr. R. J. Eden, who encouraged me to write this paper; and Dr. A. M. Jaffe, Dr. C. Edward Jones, and Dr. H. P. Stapp for several helpful criticisms and suggestions.

[31] This possibility is discussed in Ref. 2.