JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 8, NUMBER 12 DECEMBER 1967

Some New General Relativistic Dust Metrics Possessing Isometries

DAVID L. FARNSWORTH
The University of Pittsburgh, Pittsburgh, Pennsylvania

(Received 17 February 1967)

In a four-dimensional Lorentzian manifold in which Einstein's gravitational field equations hold with the field produced by pressure free dust, a significant set of new solutions is found. Assuming that the manifold possesses a four parametric group of isometries of type 5 in Bianchi's classification, which act on a three-dimensional negative definite subspace, all metrics are found. The solutions are unusual in that the geodesics which the particles follow are not orthogonal to the three-dimensional negative-definite subspace so that the space will not appear homogeneous to these observers. The isotropic expansion is nonzero for almost all these solutions.

I. INTRODUCTION

IN the four-dimensional Lorentzian manifold M we impose Einstein's general relativistic field equations. We assume that the source of the gravitational field is incoherent matter (dust). The field equations are written

$$R_{ab} - \tfrac{1}{2}g_{ab}R = -\lambda g_{ab} - \rho u_a u_b, \qquad (1.1)$$

where g_{ab} is the metric tensor, R_{ab} is the Ricci tensor, $R = R^a_a$, λ is the cosmological constant, ρ is the density of the dust so that $\rho > 0$, and u^a is the 4-velocity of the matter so that $u^a u_a = +1$. We assume that galaxies are the dust particles and that a model of the universe can be described by the solutions to Eq. (1.1).

These equations are too difficult to solve in general. As is often done,[1] we assume that the universe has some symmetries in order to simplify the equations. We assume here that M allows a four parametric group of isometries. Each one parametric subgroup generates a vector field which is said to be a Killing-vector field of M. The algebras of these vectors possess canonical forms.

If $\mathbf{X}_a$ $(a = 1, 2, 3, 4)$ are the four Killing-vector fields, then the bracket of two of them at a point of M

$$[\mathbf{X}_a, \mathbf{X}_b] = \mathbf{X}_a\mathbf{X}_b - \mathbf{X}_b\mathbf{X}_a$$

can be written

$$[\mathbf{X}_a, \mathbf{X}_b] = C^c_{ab}\mathbf{X}_c .$$

The C^a_{bc} are 24 quantities which are constants with respect to coordinate transformations and are called the constants of structure of the group. The algebra of the $\mathbf{X}_a$ is determined by the C^a_{bc} up to isomorphisms. Here we investigate Bianchi type 5.[2] The structure constants are given by

$$[\mathbf{X}_1, \mathbf{X}_2] = 0, \quad [\mathbf{X}_2, \mathbf{X}_3] = \mathbf{X}_2, \quad [\mathbf{X}_3, \mathbf{X}_1] = -\mathbf{X}_1, \qquad (1.2)$$

$$[\mathbf{X}_1, \mathbf{X}_4] = \mathbf{X}_2, \quad [\mathbf{X}_2, \mathbf{X}_4] = -\mathbf{X}_1, \quad [\mathbf{X}_3, \mathbf{X}_4] = 0. \qquad (1.3)$$

[1] J. Ehlers and W. Kundt, in *Gravitation*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), p. 49.

[2] L. Bianchi, *Lezioni sulla teoria dei gruppi continui finiti di trasformazioni* (Spoerri, Pisa, 1918).

Since all solutions to Eq. (1.1) have been found with the four Killing vectors spanning the tangent space at each point of M,[3] we assume that the Killing vectors span a three-dimensional subspace of the tangent space at each point of M which has a negative-definite metric; that is, space is homogeneous. This implies that the physics at every point in space is the same for each time. The metric and the Killing vectors can then be given the following form[4] solely on the condition imposed by Eqs. (1.2) and (1.3):

$$ds^2 = A(t)\,dx^2 + B(t)e^{2x}(dy^2 + dz^2) + dt^2,$$

$$\mathbf{X}_1 = \frac{\partial}{\partial y}, \quad \mathbf{X}_2 = \frac{\partial}{\partial z},$$

$$\mathbf{X}_3 = -\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z},$$

$$\mathbf{X}_4 = -z\frac{\partial}{\partial y} + y\frac{\partial}{\partial z}.$$

Heckmann and Schücking[5] have given a set of solutions to the dust field equations having three Killing vectors satisfying Eq. (1.2). However, these solutions as given do not allow a fourth Killing vector satisfying Eq. (1.3).

II. CHANGE OF COORDINATES

The metric is invariant under the isometries. Since u_a can be written in terms of g_{ab} by Eq. (1.1), u_a is invariant with respect to the isometries, that is,

$$[\mathbf{X}_b, u^a] = 0.$$

This forces u^a to have the form

$$u^a = \alpha(t)\frac{\partial}{\partial x} + \beta(t)\frac{\partial}{\partial t},$$

then

$$u_a = A(t)\alpha(t)\,dx + \beta(t)\,dt.$$

The twice-contracted Bianchi identities

$$(R^{ab} - \tfrac{1}{2}g^{ab}R)_{;a} = 0$$

imply that

$$u^b_{;a}u^a = 0, \tag{2.1}$$

that is, the galaxies move along geodesics. A semicolon indicates covariant differentiation so that

$$u^b_{;a}u^a = u^b_{,a}u^a + \Gamma^b_{ac}u^a u^c,$$

where a comma denotes partial differentiation and

$$\Gamma^b_{ac} = \tfrac{1}{2}g^{bs}(g_{as,c} + g_{cs,a} - g_{ac,s})$$

is the connection. Equation (2.1) along with

$$u^a u_a = A(t)\alpha^2(t) + \beta^2(t) = 1$$

gives $A(t)\alpha(t) = C$, a constant.

Therefore, unless $\alpha(t) = 0$, the dust follows geodesics which are not orthogonal to the 3-spaces of homogeneity. Therefore, although the 3-spaces are homogeneous, they will not look homogeneous to observers traveling with the dust.

We perform the change of coordinates

$$r = \int \frac{\alpha(t)}{\beta(t)}\,dt - x,$$

$$\bar{y} = y, \quad \bar{z} = z,$$

$$\tau = \int \beta(t)\,dt + A(t)\alpha(t)x,$$

in order to orient u^a in the time direction. Then

$$u_a = \frac{\partial}{\partial \tau},$$

$$ds^2 = -X^2(r,\tau)(dr)^2 - Y^2(r,\tau)(d\bar{y}^2 + d\bar{z}^2) + (d\tau)^2,$$

$$\mathbf{X}_1 = \frac{\partial}{\partial \bar{y}}, \quad \mathbf{X}_2 = \frac{\partial}{\partial \bar{z}},$$

$$\mathbf{X}_3 = -C\frac{\partial}{\partial \tau} + \frac{\partial}{\partial r} + \bar{y}\frac{\partial}{\partial \bar{y}} + \frac{\partial}{\partial \bar{z}},$$

$$\mathbf{X}_4 = -\frac{\partial}{\partial \bar{y}} + \bar{y}\frac{\partial}{\partial \bar{z}}. \tag{2.2}$$

The metric in this form is very similar to a metric considered by Bondi.[6]

We now derive expressions for $X(r,\tau)$ and $Y(r,\tau)$ from Killing's equations and the field equations. Killing's equations

$$X_{a,b} - X_{b,a} = 2\Gamma^d_{ab}X_d$$

are the necessary and sufficient conditions that $\mathbf{X}$ is a vector field generated by an isometry.[7] The only new conditions which this gives are

$$-C\dot{X} + X' = 0,$$
$$-C\dot{Y} + Y' + Y = 0, \tag{2.3}$$

where a dot over a letter means $\partial/\partial\tau$ and a prime

[3] D. Farnsworth and R. Kerr, J. Math. Phys. 7, 1625 (1966).
[4] A. S. Petrov, *Einstein-Raüme* translated by H. Treder (Akademie-Verlag, Berlin, 1964).
[5] E. Schücking and O. Heckmann, Article in *Institut International de Physique Solvay, Onzième Conseil de Physique* (Editions Stoops, Brussels, 1959); O. Heckmann and E. Schucking, in *Gravitation*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), p. 445.
[6] H. Bondi, Monthly Notices **107**, 410 (1947).
[7] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover Publications, Inc., New York, 1963), p. 208.

means $\partial/\partial r$. These equations imply that

$$X = X(Cr + \tau); \quad Y = g(Cr + \tau)\exp(-r) \quad (2.4)$$

for all C.

III. FIELD EQUATIONS AND SOLUTIONS FOR $Y' \neq 0$

The field equations for the metric, Eq. (2.2), are

$$2\frac{\dot{X}}{X}\frac{\dot{Y}}{Y} + \frac{\dot{Y}^2}{Y^2} - \frac{1}{X^2}\left(2\frac{Y''}{Y} + \frac{Y'^2}{Y^2} - 2\frac{X'}{X}\frac{Y'}{Y}\right) = \lambda + \rho, \quad (3.1)$$

$$2\frac{\ddot{Y}}{Y} + \frac{\dot{Y}^2}{Y^2} - \frac{Y'^2}{X^2Y^2} = \lambda, \quad (3.2)$$

$$\frac{\ddot{X}}{X} + \frac{\ddot{Y}}{Y} + \frac{\dot{X}\dot{Y}}{XY} - \frac{1}{X^2}\left(\frac{Y''}{Y} - \frac{X'Y'}{XY}\right) = \lambda, \quad (3.3)$$

$$\dot{Y}' - \frac{\dot{X}Y'}{X} = 0. \quad (3.4)$$

Equation (3.4) gives immediately

$$X(r, \tau) = G(r)Y'(r, \tau) \quad (3.5)$$

if $Y' \neq 0$. We consider $Y' = 0$ separately in Sec. IV. Substituting $Y = ge^{-r}$ and $X = G(ge^{-r})'$ into Eq. (2.3) yields $G(r) = Ge^{+r}$, where G is a constant. We now have

$$X = G(g' - g),$$
$$Y = ge^{-r}, \quad g = g(Cr + \tau).$$

The field equation (3.2) yields

$$2g\ddot{g} + \dot{g}^2 - \lambda g^2 - (1/G^2) = 0. \quad (3.6)$$

The field equation (3.3) reduces to the same equation and Eq. (3.1) yields

$$\rho = \frac{-6}{g - C\dot{g}}(\ddot{g} - \tfrac{1}{3}\lambda g).$$

If $\lambda = 0$ and $g \neq 0$ (i.e., $\rho \neq 0$), the same time derivative of Eq. (3.6) and $C\dot{g} = g'$ give

$$g^2\ddot{g} = D, \quad (3.7)$$

a nonzero constant. Equations (3.6) and (3.7) imply

$$\dot{g}^2 = -2D(1/g) + (1/G^2),$$

the Friedmann equation.

IV. SOLUTIONS WITH $Y' = 0$

If $Y' = 0$, then Eqs. (2.4) become

$$X = X(Cr + \tau),$$
$$Y = E\exp(\tau/C),$$

where E is a constant. The field equation (3.2) implies $\lambda = 3/C^2$ so that there are no solutions for $\lambda \leq 0$ or $C = 0$. Equation (3.3) gives

$$X(x, \tau) = M\exp(\tau/C + r) + N\exp(-2\tau/C - 2r),$$

where M and N are constants. Finally, Eq. (3.1) implies

$$\rho = \frac{-6}{C^2}\left[\left(\frac{M}{N}\right)\exp 3\left(\frac{1}{C}\tau + r\right) + 1\right]^{-1}$$

and $N \neq 0$. If $N = 0$, then $\rho = 0$.

V. PROPERTIES OF SOLUTIONS

Three quantities are defined for the timelike tangent vector to the geodesic flow lines of the dust.[1] Since these are tensor quantities, if they vanish in one coordinate system they vanish in all coordinate systems. They are isotropic expansion ($u^a_{;a}$), shear

$$[u_{(a;b)} - \tfrac{1}{3}u^c_{;c}(g_{ab} - u_au_b)],$$

and rotation ($u_{[a;b]}$). For all our solutions rotation vanishes.

Shear vanishes if and only if $C = 0, \dot{g} = 0$ or $\dot{g} = Fg$ for F a constant, and $Y' \neq 0$. If $C = 0$, expansion is $3\dot{g}/g$ and these are Friedmann models since Friedmann models are characterized by vanishing rotation and shear. If $\dot{g} = Fg$ and $F \neq 1/C$, these are also Friedmann models with expansion $3F$ in this coordinate system. If $\dot{g} = 0$, $\rho < 0$. Shear is nonvanishing for all solutions with $Y' = 0$, and expansion vanishes if and only if $M = 0$. However, $M = 0$ implies $\rho < 0$.

In general, for $Y' \neq 0$ isotropic expansion is $(\dot{X}/X) + 2(\dot{Y}/Y)$ and vanishes if and only if

$$Cg\ddot{g} - 3g\dot{g} + 2C\dot{g}^2 = 0. \quad (5.1)$$

If $C = 0$, then $\dot{g} = 0$ and $\rho = \frac{1}{2}\lambda$. This is the Einstein cosmos. If $C \neq 0$, the solution to Eq. (5.1) does not satisfy Eq. (3.6). Therefore only one of our model universes does not possess isotropic expansion, a property which the observable universe possesses.

ACKNOWLEDGMENTS

The author would like to thank Dr. R. K. Sachs and Dr. G. F. R. Ellis for their many valuable suggestions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Invariant Imbedding as a Generalization of the Resolvent Equation

ANDRÉ MOCKEL*
Department of Nuclear Engineering Sciences, University of Florida, Gainesville, Florida

(Received 15 March 1967)

Invariant imbedding equations for the Green's function of a general linear operator are shown to derive from a generalization of the resolvent equation of the theory of operators. Two applications are given: the neutron, or photon, transport in an arbitrary finite body, and the quantum-mechanical theory of collision. In the last case, a nonlinear differential equation for the amplitude of transition $T^{\beta\alpha}$ from a channel α to a channel β with conservation of the total energy is derived from the general equation.

1. INTRODUCTION

THE problem of deriving invariant imbedding equations in a way other than "particle counting", and of generalizing them to problems other than transport problems, has been paid much attention recently. Bellman and Kalaba[1] have proposed a perturbation procedure. Devooght has proposed two approaches to the generalization of the Hadamard variational formula, one[2] which generalizes an identity of Case[3] leading to various reciprocity principles, another[4] based on a variational principle.

The aim of this paper is to show that the invariant imbedding equations and the Hadamard variational formula are nothing more than a generalization of the well-known "resolvent equation" of the theory of linear operators.[5]

In Sec. 2, we show that there is an identity satisfied by any two resolvent operators connected with problems belonging to an n-parameter family. This identity first, yields the classical resolvent equation, and second, the invariant imbedding equations with respect to any of the parameters.

In Sec. 3, we apply this to a few problems in transport theory, and show that we obtain the invariant imbedding equation (or Hadamard formula) derived by Devooght.[2]

In Sec. 4 we apply our equation to a problem of scattering theory.

2. THE GENERALIZED RESOLVENT EQUATION

Let $T(\alpha^{(i)}) = T(\alpha^{(0)}, \cdots \alpha^{(n)})$ be an operator depending on n independent parameters $\alpha^{(i)}$.

For a value $\alpha_1^{(i)}$ of the parameters, we have

$$(\lambda I - T(\alpha_1^{(i)}))R(\lambda, T(\alpha_1^{(i)})) = I \tag{1}$$

and for a value $\alpha_2^{(i)}$, we have

$$R(\mu, T(\alpha_2^{(i)}))(\mu I - T(\alpha_2^{(i)})) = I \tag{2}$$

for any

$$\lambda \in \rho(T(\alpha_1^{(i)})); \quad \mu \in \rho(T(\alpha_2^{(i)})). \tag{2a}$$

Multiplying Eq. (1) to the left by $R(\mu, T(\alpha_2^{(i)}))$ and Eq. (2) to the right by $R(\lambda, T(\alpha_1^{(i)}))$, and subtracting the second from the first, we obtain

$$\begin{aligned}(\lambda - \mu)&R(\mu, T(\alpha_2^{(i)}))R(\lambda, T(\alpha_1^{(i)})) \\ &- R(\mu, T(\alpha_2^{(i)}))[T(\alpha_1^{(i)}) - T(\alpha_2^{(i)})]R(\lambda, T(\alpha_1^{(i)})) \\ &\equiv R(\mu, T(\alpha_2^{(i)})) - R(\lambda, T(\alpha_1^{(i)})).\end{aligned} \tag{3}$$

Identity (3) generalizes the classical resolvent equation (5), which can be obtained from it by making $\alpha_1^{(i)} = \alpha_2^{(i)}$. We then have

$$(\lambda - \mu)R(\mu, T)R(\lambda, T) = R(\mu, T) - R(\lambda, T) \tag{4}$$

or, by writing

$$\mu = \lambda + d\lambda,$$

$$R(\mu, T) = R(\lambda, T) + \frac{\partial R(\lambda, T)}{\partial \lambda} d\lambda,$$

$$\frac{\partial R(\lambda, T)}{\partial \lambda} = -R^2(\lambda, T). \tag{5}$$

On the other hand, by making $\lambda = \mu$ and $\alpha_2^{(i)} = \alpha_1^{(i)} + \delta\alpha^{(i)}$, and writing

$$T(\alpha_2^{(i)}) = T(\alpha_1^{(i)}) + \sum_j \delta\alpha^{(j)} \frac{\delta T(\alpha^{(i)})}{\delta \alpha^{(j)}}\bigg|_{\alpha_1^{(i)}}, \tag{6}$$

$$\begin{aligned}R(\lambda, T(\alpha_2^{(i)})) &= R(\lambda, T(\alpha_1^{(i)})) \\ &+ \sum_j \delta\alpha^{(j)} \frac{\delta R(\lambda, T(\alpha^{(i)}))}{\delta \alpha^{(j)}}\bigg|_{\alpha_1^{(i)}}.\end{aligned} \tag{7}$$

We obtain the generalized Hadamard formula

$$\begin{aligned}\sum_j \delta\alpha^{(j)}&\left[R(\lambda, T(\alpha_1^{(i)})) \frac{\delta T(\alpha^{(i)})}{\delta \alpha^{(j)}}\bigg|_{\alpha_1^{(i)}} R(\lambda, T(\alpha_1^{(i)}))\right] \\ &= \sum_j \delta\alpha^{(j)} \frac{\delta R(\lambda, T(\alpha^{(i)}))}{\delta \alpha^{(j)}}\bigg|_{\alpha_1^{(i)}}.\end{aligned} \tag{8}$$

Because of the independence of the parameters, we

---

* On leave from Université Libre de Bruxelles, Brussels, Belgium.

[1] R. Bellman and R. Kalaba, Proc. Natl. Acad. Sci. U.S.A. **47**, 336 (1961).

[2] J. Devooght, J. Math. Anal. Appl. **13**, 216 (1966).

[3] K. M. Case, Rev. Mod. Phys. **24**, 651 (1957).

[4] J. Devooght, J. Math. Phys. **7**, 1764 (1966).

[5] N. Dunford and J. T. Schwartz, *Linear Operators, Part I: General Theory* (Interscience Publishers, Inc., New York, 1958), p. 568.

can write this variational formula as a set of differential equations,

$$\frac{\delta R(\lambda, T(\alpha^{(i)}))}{\delta \alpha^{(j)}} = R(\lambda, T(\alpha^{(i)})) \frac{\delta T(\alpha^{(i)})}{\delta \alpha^{(j)}} R(\lambda, T(\alpha^{(i)})) \quad (9)$$

which are the invariant imbedding equations for the problem. We have, in order to apply them to a specific problem, to take care that relations (2a) are satisfied and, if T is a differential operator, that the derivative $\delta T(\alpha^{(i)})/\delta \alpha^{(j)}$ takes into account the variation of the boundary condition with the parameter $\alpha^{(j)}$.

3. APPLICATION TO MONOENERGETIC NEUTRON TRANSPORT THEORY

The monoenergetic Boltzmann operator is

$$B = \bar{\Omega}\bar{\nabla} \cdot + \sigma(\bar{r}) \cdot - c(\bar{r})\sigma(\bar{r}) \int f(\bar{\Omega}, \bar{\Omega}') \cdot d\bar{\Omega}', \quad (10)$$

where $\sigma(\bar{r})$ is the total cross section, $c(\bar{r})$ is the average number of secondaries per collision, and $f(\bar{\Omega}, \bar{\Omega}')$ is the probability that a neutron moving in direction $\bar{\Omega}$ be scattered in direction $\bar{\Omega}'$. We know that $\lambda = 0$ is in the resolvent set of the Boltzmann operator, provided that $c(\bar{r}) \leq 1$ and $c(\bar{r}) \to 0$ sufficiently fast for $|r| \to \infty$.

We will take as boundary condition that $\psi(\bar{r}, \bar{\Omega}) \to 0$ for $|\bar{r}| \to \infty$.

A. Geometrical Imbedding with Fixed-Point Sources

Consider an elementary source $f_\delta = \delta(\bar{r} - \bar{r}_0)\delta(\bar{\Omega} - \bar{\Omega}_0)$; we have by definition of the Green's function

$$g_p(\bar{r}, \bar{\Omega} \mid \bar{r}_0, \bar{\Omega}_0) = -R(0, B)f_\delta, \quad (11)$$

and for any vector $\mathbf{u} = \phi(\bar{r}_0, \bar{\Omega}_0)$,

$$R(0, B)\mathbf{u} = -\iint dV_0 \, d\bar{\Omega}_0 g_p(\bar{r}, \bar{\Omega} \mid \bar{r}_0, \bar{\Omega}_0)\phi(\bar{r}_0, \bar{\Omega}_0). \quad (12)$$

Let us now consider that $c(\bar{r})$ and $\sigma(\bar{r})$ are zero outside a volume limited by the surface $S_{\lambda'}$ and choose λ as the imbedding parameter. (In what follows, λ will be an imbedding parameter and no longer the spectral parameter.) We can then write

$$\sigma(\bar{r}) = \sigma(\bar{r})[1 - U(\bar{r}, \lambda)]. \quad (13)$$

With

$$U(\bar{r}, \lambda) = \begin{cases} 0 & \bar{r} \text{ INSIDE THE SURFACE } S_\lambda \\ 1 & \bar{r} \text{ OUTSIDE THE SURFACE } S_\lambda, \end{cases}$$

we have

$$\frac{\delta U(\bar{r}, \lambda)}{\delta \lambda} = -\delta(\bar{r} - \bar{r}_{s\lambda}) \frac{\delta n(\bar{r}, \lambda)}{\delta \lambda}, \quad (14)$$

$\delta(\bar{r} - \bar{r}_{s_\lambda})$ being a surface Dirac function and $(\delta n(\bar{r}, \lambda)/\delta\lambda)\delta\lambda$ the normal displacement of the point $\bar{r}$. Then,

$$\frac{\delta B}{\delta \lambda} = \sigma(r) \frac{\delta n(\bar{r}, \lambda)}{\delta \lambda} \delta(\bar{r} - \bar{r}_{s_\lambda}) \times \left[1 - c(\bar{r}) \int f(\bar{\Omega}, \bar{\Omega}') \cdot d\bar{\Omega}'\right], \quad (15)$$

and

$$\frac{\delta g_p(\bar{r}, \bar{\Omega} \mid \bar{r}_0, \bar{\Omega}_0)}{\delta \lambda} = -\iint dV' \, d\Omega' g_p(\bar{r}, \bar{\Omega} \mid \bar{r}', \bar{\Omega}')\sigma(\bar{r}') \times \frac{\delta n(\bar{r}', \lambda)}{\delta \lambda} \delta(\bar{r}' - \bar{r}_{s\lambda}) \cdot \left[g_p(\bar{r}', \bar{\Omega}' \mid \bar{r}_0, \bar{\Omega}_0) - c(\bar{r}) \int f(\bar{\Omega}', \bar{\Omega}'') g_p(\bar{r}', \bar{\Omega}'' \mid \bar{r}_0, \bar{\Omega}_0) \, d\bar{\Omega}''\right]. \quad (16)$$

Performing the integration over $\bar{r}'$, we obtain

$$\frac{\delta g_p(\bar{r}, \bar{\Omega} \mid \bar{r}_0, \bar{\Omega}_0)}{\delta \lambda} = -\int_{s_\lambda} dS' \int d\bar{\Omega}' \frac{\delta n(\bar{r}_s')}{\delta \lambda} \sigma(\bar{r}_s') g_p(\bar{r}, \bar{\Omega} \mid \bar{r}_s', \bar{\Omega}') \cdot \left[g_p(\bar{r}_s', \bar{\Omega}' \mid \bar{r}_0, \bar{\Omega}_0) - c(\bar{r}_s') \int f(\bar{\Omega}', \bar{\Omega}'') g_p(\bar{r}_s', \bar{\Omega}'' \mid \bar{r}_0, \bar{\Omega}_0) \, d\bar{\Omega}''\right]. \quad (17)$$

Making use of the well-known reciprocity relation (3),

$$g_p(\bar{r}, \bar{\Omega} \mid \bar{r}_0, \bar{\Omega}_0) = g_p(\bar{r}_0, -\bar{\Omega}_0\bar{r}, -\bar{\Omega}) \quad (18)$$

and

$$f(\bar{\Omega}, \bar{\Omega}') = f(-\bar{\Omega}', -\bar{\Omega}); \quad (19)$$

this formula is easily shown to be identical to formula (3.23) of Devooght.[2]

B. Geometrical Imbedding with Surface Sources

In many problems, it is interesting to compute the variation of the surface Green's function. The source f_δ is then a function of the parameter λ. We have

$$g_s(\bar{r}, \bar{\Omega} \mid \bar{r}_{0S_{\lambda+d\lambda'}}, \bar{\Omega}_0, \lambda + d\lambda) = -R(0, B(\lambda + d\lambda))f_\delta(\lambda + d\lambda) = -R(0, B(\lambda + d\lambda))f_\delta(\lambda) - R(0, B(\lambda))\frac{\partial f_\delta(\lambda)}{\partial \lambda} d\lambda, \quad (20)$$

neglecting terms of the second order in $d\lambda$; finally,

$$\frac{\delta g_s(\bar{r}, \bar{\Omega} \mid \bar{r}_{0s}, \bar{\Omega}_0, \lambda)}{\delta \lambda} = \frac{-\delta R(0, B(\lambda))}{\delta \lambda} f_\delta(\lambda) - R(0, B(\lambda))\frac{\partial f_\delta(\lambda)}{\partial \lambda}, \quad (21)$$

where the point $\bar{r}_{0s}$ follows the boundary in its displacement. The derivative of the source will contain

the derivative of a Dirac function, which in turn will produce a derivative of the Green's function with respect to $\bar{r}_{0s}$, evaluated at the boundary. This can be evaluated with the use of the reciprocity principle and the Boltzmann equation. There is some ambiguity in this evaluation as the derivative of the Green's function is discontinuous at the boundary. It is clear, however, from Eq. (20) that we have to evaluate the variation of the Green's function when the source point is moved from outside the medium towards the surface, and so we have to use the Boltzmann equation with the values of $c(\bar{r}_s^+)$ and $\sigma(\bar{r}_s^+)$, that is, zero.

To remember this, we will write Eq. (21) as

$$\frac{\delta g_s(\bar{r}, \bar{\Omega} \mid \bar{r}_{0s}, \bar{\Omega}_0; \lambda)}{\delta\lambda} = \frac{-\delta R(0, B(\lambda))}{\delta\lambda} f_\delta(\lambda) - R(0, B(\lambda)) \frac{\partial f_\delta(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda^+}. \quad (22)$$

We sometimes want our observation point to follow the surface when λ goes to $\lambda + d\lambda$. The general equation for this case can be written easily if we consider $g_s(\bar{r}, \bar{\Omega} \mid \bar{r}_{0s}, \bar{\Omega}_0; \lambda)$ for fixed $\bar{r}_{0s}, \bar{\Omega}_0$ and arbitrary $\bar{r}, \bar{\Omega}$ as a vector in the vector space of fluxes. To obtain the Green's function at a specific point of space, we have to apply to this vector a functional f_0^* which will transform our vector $g_s(\bar{r}, \bar{\Omega} \mid \bar{r}_{0s}, \bar{\Omega}_0; \lambda)$ into a scalar $g_s(\bar{r}_1, \bar{\Omega}_1 \mid \bar{r}_{0s}, \bar{\Omega}_0; \lambda)$, where $\bar{r}_1$ and $\bar{\Omega}_1$ are fixed. If the observation points follow the surface when λ goes to $\lambda + d\lambda$, the functional f_δ^* will be dependent on λ. We have

$$\begin{aligned} &g_s(\bar{r}_{1S\lambda+d\lambda'}, \bar{\Omega}_1 \mid \bar{r}_{0S\lambda+d\lambda'}, \bar{\Omega}_0; \lambda + d\lambda) \\ &= -f_\delta^*(\lambda + d\lambda) R(0, B(\lambda + d\lambda)) f_\delta(\lambda + d\lambda) \\ &= -f_\delta^*(\lambda + d\lambda)\Big[R(0, B(\lambda)) f_\delta(\lambda) + \frac{\delta R(0, B(\lambda))}{\delta\lambda} f_\delta(\lambda)\, d\lambda \\ &\qquad + R(0, B(\lambda)) \frac{\partial f_\delta(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda^+} d\lambda\Big]. \end{aligned} \quad (23)$$

Again the computation of the variation of the functional involves derivatives of Dirac functions, which in turn leads us to evaluate the derivative of the Green's function at the surface. This time, again, it is clear from Eq. (23) that it is the derivative of the Green's function outside the medium which is needed, so that our formula becomes

$$\begin{aligned} &\frac{\delta g(\bar{r}_{1s}, \bar{\Omega}_1 \mid \bar{r}_{0s}, \bar{\Omega}_0; \lambda)}{\delta\lambda} \\ &\qquad = \frac{-\partial f_\delta^*(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda^+} R(0, B(\lambda)) f_\delta(\lambda) \\ &\qquad\quad - f_\delta^*(\lambda) \frac{\partial R(0, B(\lambda))}{\partial\lambda} f_\delta(\lambda) \\ &\qquad\quad - f_\delta^*(\lambda) R(0, B(\lambda)) \frac{\delta f_\delta(\lambda)}{\delta\lambda}\bigg|_{\lambda=\lambda^+}. \end{aligned} \quad (24)$$

Equation (24) is the general form of the invariant imbedding equation for the reflection. By specializing it to a particular problem, we can derive the equations established by Devooght in Sec. 3C of Ref. 2.

4. APPLICATION TO QUANTUM SCATTERING THEORY

The system is governed by the Hamiltonian H and $R(E + i\epsilon)$ is the complete Green's function G^+. We separate the Hamiltonian into a free-wave term and an interaction term. If we assume that rearrangement can take place, we can do that in several ways,

$$H = H_a + H_a' = H_b + H_b'.$$

We can then define the operators

$$T_{ba}(E) = H_b' + H_b' G^+(E) H_a' \quad (25)$$

and

$$T_{bb}(E) = H_a' + H_b' G^+(E) H_a' = T_{ba} + H_a' - H_b'. \quad (26)$$

The matrix elements $T_{\beta\alpha}$ between eigenfunctions of the free Hamiltonian H_a and H_b, taken at the same total energy of the system, give the reaction amplitude. Taking λ as a parameter of the Hamiltonian, we then have

$$\frac{\partial T_{ba}(E)}{\partial\lambda} = \frac{\partial H_b'}{\partial\lambda} + \frac{\partial H_b'}{\partial\lambda} G^+(E) H_a' + H_b' \frac{\partial G^+(E)}{\partial\lambda} H_a' + H_b' G^+(E) \frac{\partial H_a'}{\partial\lambda}.$$

Using

$$\frac{\partial G^+(E)}{\partial\lambda} = G^+(E) \frac{\partial H}{\partial\lambda} G^+(E)$$

and the definition of T_{ba}, we obtain

$$\begin{aligned} &\frac{\partial T_{ba}}{\partial\lambda} \\ &= \frac{\partial H_b'}{\partial\lambda} H_b'^{-1} T_{ba} + H_b' G^+(E) \frac{\partial H}{\partial\lambda} G^+(E) H_a' \\ &\quad + H_b' G^+(E) \frac{\partial H_a'}{\partial\lambda} \\ &= \frac{\partial H_b'}{\partial\lambda} H_b'^{-1} T_{ba} + H_b' G^+(E) H_a' H_a'^{-1} \frac{\partial H}{\partial\lambda} H_b'^{-1} H_b' G^+(E) H_a' \\ &\quad + H_b' G^+(E) H_a' H_a'^{-1} \frac{\partial H_a'}{\partial\lambda} \\ &= \frac{\partial H_b'}{\partial\lambda} H_b'^{-1} T_{ba}(E) - \frac{\partial H}{\partial\lambda} H_b'^{-1} H_b' G^+(E) H_a' \\ &\quad + T_{bb} H_a'^{-1} \frac{\partial H}{\partial\lambda} H_b'^{-1} H_b' G^+(E) H_a' \\ &\quad - \frac{\partial H_a'}{\partial\lambda} + T_{bb} H_a'^{-1} \frac{\partial H_a'}{\partial\lambda} \end{aligned}$$

$$= \frac{\partial H'_b}{\partial \lambda} H'^{-1}_b T_{ba}(E) + \frac{\partial H}{\partial \lambda} - \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba}(E)$$
$$+ T_{ba} H'^{-1}_a \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba}$$
$$- T_{ba} H'^{-1}_a \frac{\partial H}{\partial \lambda} - \frac{\partial H'_a}{\partial \lambda} + T_{ba} H'^{-1}_a \frac{\partial H'_a}{\partial \lambda},$$

and finally,

$$\frac{\partial T_{ba}(E)}{\partial \lambda} = T_{ba} H'^{-1}_a \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba} + \frac{\partial H_a}{\partial \lambda}$$
$$- \frac{\partial H_b}{\partial \lambda} H'^{-1}_b T_{ba}(E) - T_{ba} H'^{-1}_a \frac{\partial H_b}{\partial \lambda}. \quad (27)$$

If λ is a parameter of the interaction only, i.e., if

$$\frac{\partial H_a}{\partial \lambda} = \frac{\partial H_b}{\partial \lambda} = 0,$$

$$\frac{\partial T_{ba}(E)}{\partial \lambda} = T_{ba}(E) H'^{-1}_a \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba}(E), \quad (28)$$

which is a generalization of formula (7.76) of Ref. 6.

ACKNOWLEDGMENTS

The author wants to thank Professor R. B. Perez for numerous discussions of the manuscript, and Dr. J. Devooght for his interesting comments.

[6] R. G. Newton, *Scattering Theory of Waves and Particles*, (McGraw-Hill Book Co., New York, 1966), pp. 246–263.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Causality and the S Matrix

ANTONIO F. RAÑADA*
Laboratoire de Physique Théorique et Hautes Energies, 91 Orsay, France

(Received 6 May 1966)

A formalism is presented in order to impose causality conditions for finite-range interactions which do not require the existence of wave packets with a sharp front. The absence of bound states and particle production is assumed. Two different causality conditions—referred to as strong and weak conditions—are studied. It is shown that in general the convolution kernel that connects the ingoing with the outgoing wave is not zero in the past. Some analytic properties of the S matrix are deduced.

INTRODUCTION

THE purpose of this paper is the study of some analytic properties imposed by causality on the scattering matrix $S_l(w)$ of two elementary particles when the interaction has finite range. A formalism is proposed which avoids the main difficulty of the problem: The nonexistence of wave packets with sharp front when the mass is different from zero. Our conclusions are independent of the mass value.

The first causality condition was proposed in 1927 by Kramers and Krönig[1] in the form of a dispersion integral (the Kramers–Krönig relation). They proved that this relation is a necessary and sufficient condition in order that light velocity in a medium is bounded by c. Schützer and Tiomno[2] formulated a causality condition for the nonrelativistic s-wave scattering. Although the paper contained a correct analysis, the condition was formulated in an incorrect way as was pointed out by Van Kampen.[3] They corrected the formulation in Ref. 4. Giambiagi and Saavedra[5] gave a mathematical improvement of the Schützer–Tiomno condition by means of a convolution kernel. They stated that this kernel $H(\tau)$ should be zero for $\tau < 0$. In this paper we prove that this condition is valid only for s waves. In higher partial waves it is too strong.

Toll[6] studied the logical equivalence between causality and the Kramers–Krönig relation in the frame of a formalism which can be applied to the scattering of zero-mass particles. This case was also treated by Van Kampen[7] who has also formulated a condition for the case of nonrelativistic particles in s waves. Van Kampen and Toll found in this way

* Present address: Junta de Energía Nuclear, Madrid 3, Spain.

[1] H. A. Kramers, *Collected Scientific Papers* (Amsterdam, 1956); R. Krönig, J. Opt. Soc. Am. **12**, 547 (1927).

[2] W. Schützer and J. Tiomno, Phys. Rev. **83**, 249 (1951).

[3] N. G. Van Kampen, Phys. Rev. **91**, 1267 (1953); Physica **20**, 115 (1954).

[4] W. Schutzer and J. Tiomno, Symposium on New Research Techniques in Physics, Rio de Janeiro (1954).

[5] J. J. Giambiagi and I. Saavedra, Nucl. Phys. **46**, 431 (1963).

[6] J. S. Toll, Phys. Rev. **104**, 1760 (1956).

[7] N. G. Van Kampen, Phys. Rev. **89**, 1072 (1953).

$$= \frac{\partial H'_b}{\partial \lambda} H'^{-1}_b T_{ba}(E) + \frac{\partial H}{\partial \lambda} - \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba}(E)$$
$$+ T_{ba} H'^{-1}_a \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba}$$
$$- T_{ba} H'^{-1}_a \frac{\partial H}{\partial \lambda} - \frac{\partial H'_a}{\partial \lambda} + T_{ba} H'^{-1}_a \frac{\partial H'_a}{\partial \lambda},$$

and finally,

$$\frac{\partial T_{ba}(E)}{\partial \lambda} = T_{ba} H'^{-1}_a \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba} + \frac{\partial H_a}{\partial \lambda}$$
$$- \frac{\partial H_b}{\partial \lambda} H'^{-1}_b T_{ba}(E) - T_{ba} H'^{-1}_a \frac{\partial H_b}{\partial \lambda}. \quad (27)$$

If λ is a parameter of the interaction only, i.e., if

$$\frac{\partial H_a}{\partial \lambda} = \frac{\partial H_b}{\partial \lambda} = 0,$$

$$\frac{\partial T_{ba}(E)}{\partial \lambda} = T_{ba}(E) H'^{-1}_a \frac{\partial H}{\partial \lambda} H'^{-1}_b T_{ba}(E), \quad (28)$$

which is a generalization of formula (7.76) of Ref. 6.

ACKNOWLEDGMENTS

The author wants to thank Professor R. B. Perez for numerous discussions of the manuscript, and Dr. J. Devooght for his interesting comments.

[6] R. G. Newton, *Scattering Theory of Waves and Particles*, (McGraw-Hill Book Co., New York, 1966), pp. 246–263.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Causality and the S Matrix

ANTONIO F. RAÑADA*
Laboratoire de Physique Théorique et Hautes Energies, 91 Orsay, France

(Received 6 May 1966)

A formalism is presented in order to impose causality conditions for finite-range interactions which do not require the existence of wave packets with a sharp front. The absence of bound states and particle production is assumed. Two different causality conditions—referred to as strong and weak conditions—are studied. It is shown that in general the convolution kernel that connects the ingoing with the outgoing wave is not zero in the past. Some analytic properties of the S matrix are deduced.

INTRODUCTION

THE purpose of this paper is the study of some analytic properties imposed by causality on the scattering matrix $S_l(w)$ of two elementary particles when the interaction has finite range. A formalism is proposed which avoids the main difficulty of the problem: The nonexistence of wave packets with sharp front when the mass is different from zero. Our conclusions are independent of the mass value.

The first causality condition was proposed in 1927 by Kramers and Krönig[1] in the form of a dispersion integral (the Kramers–Krönig relation). They proved that this relation is a necessary and sufficient condition in order that light velocity in a medium is bounded by c. Schützer and Tiomno[2] formulated a causality condition for the nonrelativistic s-wave scattering. Although the paper contained a correct analysis, the condition was formulated in an incorrect way as was pointed out by Van Kampen.[3] They corrected the formulation in Ref. 4. Giambiagi and Saavedra[5] gave a mathematical improvement of the Schützer–Tiomno condition by means of a convolution kernel. They stated that this kernel $H(\tau)$ should be zero for $\tau < 0$. In this paper we prove that this condition is valid only for s waves. In higher partial waves it is too strong.

Toll[6] studied the logical equivalence between causality and the Kramers–Krönig relation in the frame of a formalism which can be applied to the scattering of zero-mass particles. This case was also treated by Van Kampen[7] who has also formulated a condition for the case of nonrelativistic particles in s waves. Van Kampen and Toll found in this way

* Present address: Junta de Energía Nuclear, Madrid 3, Spain.

[1] H. A. Kramers, *Collected Scientific Papers* (Amsterdam, 1956); R. Krönig, J. Opt. Soc. Am. **12**, 547 (1927).

[2] W. Schützer and J. Tiomno, Phys. Rev. **83**, 249 (1951).

[3] N. G. Van Kampen, Phys. Rev. **91**, 1267 (1953); Physica **20**, 115 (1954).

[4] W. Schutzer and J. Tiomno, Symposium on New Research Techniques in Physics, Rio de Janeiro (1954).

[5] J. J. Giambiagi and I. Saavedra, Nucl. Phys. **46**, 431 (1963).

[6] J. S. Toll, Phys. Rev. **104**, 1760 (1956).

[7] N. G. Van Kampen, Phys. Rev. **89**, 1072 (1953).

analytic expressions for the S matrix which are generalizations of the Hu[8] expression. Another important reference is Hilgevoord.[9]

The main problem is the nonexistence of sharp fronts if the mass is not zero. As a consequence the cause (ingoing wave) and the effect (outgoing wave) extend in time from $-\infty$ to $+\infty$. Therefore we cannot formulate "the effect is posterior to the cause" in an straightforward manner. However we see that it is possible to express the outgoing wave as a convolution product of the ingoing wave with a kernel which plays the role of time-delay function. The causality condition will then be imposed on this kernel. We will study two different causality conditions referred to as "strong" and "weak" causality.

In Sec. 1 we give the definition of Hardy spaces and the factorization theorem. The reader is referred to the book by Hoffman.[10]

In Sec. 2 we discuss the theory of passive linear systems defined as operators in a Hilbert space. An elementary approach is to be found in Robinson.[11]

Section 3 deals with the scattering of two particles in the frame of a formalism quite analogous to that of linear systems. The only hypothesis is that the interaction has finite range R. The ingoing and outgoing waves are considered as the input and the output of a linear system. From causality we obtain the analytic properties of S matrix in the complex-energy plane.

1. FACTORIZATION OF THE FUNCTIONS OF HARDY SPACES

Let us consider the complex variable w. H^P is the space of functions which are analytic in the upper half-plane and such that

$$\int_{-\infty}^{+\infty} |f(w_r + iw_i)|^P \, dw_r < M,$$

where M is a constant. H^∞ is the set of functions analytic and uniformly bounded in the upper half-plane. H^P $(1 \leq p \leq \infty)$ are Banach spaces.[10] Their elements have limits almost everywhere in the real axis which belong to L^P and L^∞. In this sense there is an isomorphy between H^P and a subspace of L^P of the real axis. We can write $H^P \subset L^P$.

A. Paley–Wiener Theorem[12]

$f \in H^2$ if and only if it has the form

$$f(w) = \int_0^\infty \hat{f}(t) e^{iwt} \, dw$$

for some $\hat{f}$ in $L^2(0, +\infty)$. This theorem means that H^2 and $L^2(0, +\infty)$ are Fourier-transformed spaces.

B. Factorization Theorem[10]

If $f \in H^P$ $(1 \leq p \leq \infty)$, it admits the factorization

$$f(w) = F \exp(i\alpha w) BS, \tag{1}$$

where F, called "the outer factor," is

$$F = \exp \frac{1}{\pi} \left[\int_{-\infty}^{+\infty} \log |f(t)| \frac{tw + 1}{i(t - w)} \frac{dt}{1 + t^2} \right] \tag{2}$$

with the condition

$$\int_{-\infty}^{+\infty} \log |F(t)| \frac{dt}{1 + t^2} > -\infty. \tag{2'}$$

$\alpha \geq 0$, B is called Blaschke product and is

$$B(w) = \prod_{n=1}^{\infty} \frac{w - \beta_n}{w - \beta_n^*} \tag{3}$$

with

$$\operatorname{Im} \beta_n \geq 0; \quad \sum_{n=1}^{\infty} \frac{\operatorname{Im} \beta_n}{1 + |\beta_n|^2} < \infty. \tag{3'}$$

S is called a singular function and is

$$S(w) = \exp \left[-\int \frac{tw + 1}{i(t - w)} \, d\mu(t) \right], \tag{4}$$

where μ is a positive, finite, singular measure.

F and S have no zeros in the upper half-plane so that all the zeros of f are the β_n. This factorization is unique up to a phase factor.

2. PASSIVE LINEAR SYSTEMS—STRONG AND WEAK CAUSALITY

In this section we give a short account of the theory of passive linear systems. The usual form of causality turns out to be too strong to be applied to scattering systems. Instead of it we will employ a weaker form.

There are several works in which linear system theory is used in order to impose causality conditions (Refs. 13 to 17).

[8] N. Hu, Phys. Rev. **74**, 131 (1948).

[9] J. Hilgevoord, *Dispersion Relations and Causal Description* (North-Holland Publishing Company, Amsterdam, 1960).

[10] K. Hoffman, *Banach Spaces of Analytic Functions* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962).

[11] E. A. Robinson, *Random Wavelets and Cybernetic Systems* (Griffin, London, 1962).

[12] R. Paley and N. Wiener, Amer. Math. Soc. Colloq. Pub. XIX, New York (1934).

[13] D. C. Youla, L. J. Castriota, and H. J. Carlin, IRE Trans. CT-6, 102 (1959).

[14] C. L. Dolph, Ann. Acad. Sci. Fenn. **A1**, 336 (1963).

[15] A. Zemanian, *Distribution Theory and Transform Analysis* (McGraw-Hill Book Company, Inc., New York, 1965).

[16] W. Guttinger, Fortschr. Physik **14**, 483 (1966).

[17] E. J. Beltrami and M. R. Wohlers, *Distribution and the Boundary Value of Analytic Functions* (Academic Press Inc., New York, 1966).

From the physical point of view a passive linear system (PLS) is a device which transforms an input into an output. It is passive if it cannot increase the value of a given quadratic form of the input, which may have the meaning of an energy (classical signals), of a norm (wavefunctions) etc. Mathematically the definition is the following: "A passive linear system (PLS) is a bounded linear operator in $L^2(-\infty, +\infty)$ that commutes with translations." T is then a PLS if $b = Ta$ and $b, a \in L^2(-\infty, +\infty)$ and

$$b * \delta_\tau = T(a * \delta_\tau); \quad \|a\|_2 \geq K \|b\|_2 , \tag{5}$$

where K is a positive constant which can be taken as unity without loss of generality. a is the input and b the output.

Theorem 2.1: If T is a PLS, there exists a bounded function Σ such that

$$B = \Sigma A, \tag{6}$$

where B and A are the Fourier transforms of b and a. Σ will be called a "transfer function." (See Refs. 11, 14, 7, and 18.)

Theorem 2.1: If T is a PLS, there exists a rapidly decreasing distribution[19] σ such that

$$b = \sigma * a. \tag{7}$$

σ will be called time-delay distribution and is the transfer function of Σ.

Theorem 2.1′ is clearly a consequence of Theorem 2.1. Equation (7) results by Fourier transformation of (6). Relations (6) and (7) express the action of a pair of linear operators. The first $(A \to B)$ will be called T_w and the second $(a \to b)$ T_t. The respective spaces will be referred to as frequency and time space.

Corollary: The Hilbert subspace of the function whose transfer function is zero on a certain set is invariant by a PLS.—We could thus define a passive linear subsystem as the restriction of a PLS to any such subspace. It is clear that there are infinite PLS's having the same restriction to a given subspace.

We now introduce causality in PLS theory.

Definition: A causal system T is a PLS such that

$$T_t L^2(0, +\infty) \subset L^2(0, +\infty). \tag{8}$$

In the frequency space this is written

$$T_w H^2 \subset H^2. \tag{8′}$$

In order to understand this definition, recall that

$$L^2(-\infty, +\infty) = L^2(-\infty, 0) \oplus L^2(0, +\infty)$$

and the Paley–Wiener theorem. It is clear that if $a = 0$ for $t < t_1$, then $b = 0$ for $t < t_1$.

Theorem 2.2 (Causality Condition): A PLS is a causal system if and only if one of the following statements hold:

(a) The support of the time-delay distribution σ is contained in $(0, +\infty)$;

(b) The transfer function Σ is the boundary value of a function of H^∞ of the upper half-plane.

The proof is straightforward. It turns out that this causality condition is too strong to be applied to scattering. From now on it will be referred to as strong causality. We consider now a weaker form:

Let H_N^2 be a Hilbert subspace of H^2.

Definition: A PLS is said to be "causal" in H_N^2 if

$$T_w H_N^2 \subset H^2. \tag{9}$$

Let us now choose as H_N^2 the set of functions of H^2 which vanish on a certain finite set of points in the upper half-plane $(w_1, w_2, \cdots, w_N)$ such that all w's are different.

Theorem 2.3 (Casuality Condition) in H_N^2: A PLS is causal in H_N^2 in and only if one of the following statements holds:

(a) The time-delay distribution σ has the following expression for $t < 0$:

$$\sigma = \sum_{j=1}^{N} \lambda_j e^{-iw_j t}, \tag{10}$$

the λ's being complex numbers.

(b) The transfer function Σ is the boundary value of a function analytic and bounded in the upper half-plane except perhaps in the points $w_1 \cdots w_N$. In these points Σ may have simple poles.

The proof is simple and similar to that of Theorem 2.2. Condition (b) is easily proved, and by Fourier transformation one gets condition (a).

Some remarks are to be made.

(1) If all λ_j are zero we get the strong condition and the PLS is causal in all H^2.

(2) σ decreases exponentially when $t \to -\infty$ but an arbitrary exponential decrease does not give a system causal in H_N^2.

(3) Σ may have simple poles in $w_1 \cdots w_N$ but it may also be analytic in several w_j. If Σ is analytic in w_κ, then $\lambda_\kappa = 0$.

[18] A. F. Rañada, Ph.D. thesis, University of Paris (unpublished). A very elegant proof was pointed out to me by P. Eymard (private communication).

[19] L. Schwartz, *Les distributions* (Herman & Cie., Paris, 1957).

The function

$$\Sigma'(w) = \Sigma(w) \prod_{j=1}^{N} \frac{w - w_j}{w - w_j^*} \tag{11}$$

belongs to H^∞.

3. CAUSALITY AND SCATTERING

We will now apply PLS theory to scattering systems. The Giambiagi and Saavedra[5] condition will be shown to be equivalent to strong causality and it will be proved that it is only valid in s waves. In fact Giambiagi and Saavedra gave their condition for S waves so that their paper is correct. If $l \neq 0$ however a weaker form of causality is required. [I am grateful to the referee for pointing out through some counter-examples that my original causality condition (equivalent to Giambiagi and Saavedra condition in S waves) was too restrictive.]

A scattering system is linear, commutes with time translations and does not produce particles. This last property is to be related to passivity and we have thus a strong suggestion for the consideration of scattering as a PLS. This point of view can be developed for any value of m. It can also be applied both to nonrelativistic or relativistic particles, the only difference being the energy–momentum relation. For the sake of simplicity let us take first the case $m = 0$.

We assume only the following: In the e.m. system, the interaction is contained in a sphere of radius R outside of which the Klein–Gordon equation holds.

Consequently we do not make any hypothesis about the strength or form of the interaction.

Case $m = 0$: Out of the region of interaction we can write the ingoing and outgoing partial waves in the form

$$\psi_{l,\text{in}}(\bar{r}, t) = \int_{-\infty}^{+\infty} A_l(w) h_l^-(kr) e^{-iwt}\, dw P_l(\cos\theta),$$

$$\psi_{l,\text{out}}(\bar{r}, t) = \int_{-\infty}^{+\infty} S_l(w) A_l(w) h_l^+(kr) e^{-iwt}\, dw P_l(\cos\theta), \tag{12}$$

where $k = w$ and $h_l^\pm$ are the well-known spherical Hankel functions. A_l must have a zero of order $(l+1)$ at $k = 0$ to compensate the pole of $h_l^\pm$. If $A_l \in L^2(-\infty, +\infty)$ it is clear that $\psi_{l,\text{in}}$ and $\psi_{l,\text{out}}$, considered as functions of time are also in $L^2(-\infty, +\infty)$ for any $\bar{r}$. This follows from the fact

$$|S_l(w)| \leq 1, \tag{13}$$

which states that the outgoing flux, integrated from $t = -\infty$ to $t = +\infty$, is not greater than the ingoing flux. Applying the Plancherel theorem, we get also

$$\int_{-\infty}^{+\infty} |\psi_{l,\text{in}}|^2\, dt \geq \int_{-\infty}^{+\infty} |\psi_{l,\text{out}}|^2\, dt \tag{14}$$

for any $\bar{r}$. The relation between $\psi_{l,\text{in}}$ and $\psi_{l,\text{out}}$ being obviously linear and independent of time translations, the mapping

$$\psi_{l,\text{in}}(\bar{r}, t) \to \psi_{l,\text{out}}(\bar{r}, t) \tag{15}$$

is a PLS for any $\bar{r}$. Let us call it $T(r)$. Then there exists a distribution $\sigma_l^{(r)}$ such that

$$\psi_{l,\text{out}}(\bar{r}, t) = \psi_{l,\text{in}}(\bar{r}, t) * \sigma_l^{(r)} \tag{16}$$

and a transfer function which obviously has the form

$$\sum_l^{(r)}(w) = S_l(w) \frac{h_l^+(kr)}{h_l^-(kr)}. \tag{17}$$

h_l^- has l zeros in the upper half-plane and h_l^+ l zeros in the lower one.

Now suppose we fix $r = R$ and try to apply the strong causality condition to $T(R)$:

$$\sigma_l^{(R)} = 0 \quad \text{for} \quad t < 0. \tag{18}$$

As we have shown $\sum_l^{(R)}$ must then be a function of H^∞ and from (17), we see that $S_l(w)$ must be zero at the zeros of $h_l^-(kR)$. Otherwise $\sum_l^{(R)}$ would have poles.

There may be found examples in which these zeros of S_l are absent. For instance the scattering of electromagnetic waves by metallic or dielectric spheres which can be solved exactly. Therefore strong causality is not valid. Let us now try a weaker form. If we take $A_l \in H^2$ in (12), the Fourier transform of $\psi_{l,\text{in}}$ vanishes at the zeros of $h_l^-(kR)$. We can thus restrict our causality condition to the Banach subspace of H^2 formed by the functions which are zero at these points. Let us call it H_L^2. Then our condition must be

$$T_w(R) H_L^2 \subset H^2. \tag{19}$$

This is what we called in Sec. 2 causality in H_L^2.

Because of Theorem 2.3, we state instead of (18) the "weak causality condition"

$$\sigma_l^{(R)} = \sum_{j=1}^{l} \lambda_j e^{-iw_j t} \quad \text{for} \quad t < 0, \tag{20}$$

where w_j are the l zeros of $h_l^-(kR)$. The λ_j are complex numbers.

On the other hand $\sum_l^{(R)}$ must be analytic and bounded in the upper half-plane except perhaps at the points $w_1 \cdots w_l$.

From remark 3 after Theorem 2.3 and the properties of the zeros of $h_l^\pm$, it is easy to see that

$$S_l(w) e^{2ikR} \tag{21}$$

is a function of H^∞ of the upper half-plane.

It should be emphasized that the convolution kernel is not in general zero for $t < 0$ except for s waves. This is because h_0^- has no zeros.

Case $m \neq 0$: The ingoing and outgoing waves are now (taking only positive frequencies)

$$\psi_{l,\text{in}}(\bar{r}, t) = \int_m^\infty A_l(w) h_l^-(kr) e^{-iwt}\, dw P_l(\cos\theta), \quad (22)$$

$$\psi_{l,\text{out}}(\bar{r}, t) = \int_m^\infty S_l(w) A_l(w) h_l^+(kr) e^{-iwt}\, dw P_l(\cos\theta), \quad (23)$$

with

$$w = (k^2 + m^2)^{\frac{1}{2}} \quad \text{or} \quad w = k^2/2m.$$

The only difference with the case $m = 0$ is that now the mapping (15) is not defined in $L_l^2(-\infty, +\infty)$ but in the subspace of functions whose Fourier transform vanishes in $(-\infty, +m)$:

$$L_w^2(-\infty, +\infty) = L_w^2(-\infty, m) \oplus L_w^2(m, +\infty),$$
$$L_l^2(-\infty, +\infty) = \mathcal{F}L_w^2(-\infty, m) \oplus \mathcal{F}L_w^2(m, +\infty). \quad (24)$$

Our scattering system is thus a passive linear subsystem as defined in the corollary to Theorem 2.1. It is clear that there is in $\mathcal{F}\, L_w^2(-\infty, m)$ an infinite set of linear operators which commute with translations and are of norm ≤ 1. In other words there are infinite PLS which when restricted to $w > m$ have the same effect as the physical system.

We must point out that in order to describe bound states we should use unbounded extensions to all L^2. From now on we suppose that there are no bound states.

Theorems 2.1 and 2.1′ allows us then to say that there is an infinite set of rapidly decreasing distributions[19] $\sigma_{l,\alpha}^{(R)}$ such that

$$\psi_{l,\text{out}}(R, t) = \psi_{l,\text{in}}(R, t) * \sigma_{l,\alpha}^{(R)} \quad (25)$$

and an infinite set of bounded functions $\sum_{l,\alpha}^{(R)}$ such that

$$S_l(w) A_l(w) h_l^+(kR) = \textstyle\sum_{l,\alpha}^{(R)}(w) A_l(w) h_l^-(kR), \quad (26)$$

where we have not written the factors $P_l(\cos\theta)$.

It is clear that all $\sum_{l,\alpha}^{(R)}$ must be equal in the physical region $w > m$.

If we want to impose a causality condition on the convolution kernel, we find that there is an infinity of such kernels. This problem arises also for instance in the study of the causal propagation of free Klein–Gordon waves. If we only consider positive frequencies we have an infinite set of convolution kernels that propagate $\varphi^+(x)$. For instance $\Delta_\lambda = \Delta^+ + \lambda\Delta^-$, λ being an arbitrary complex number. The natural causality condition refers then to the future cone. We know, however, that most of these distributions do not have their supports contained in the future cone. Nevertheless causality is not violated because there exists one which is zero outside the cone and all are equivalent to it. This special kernel is $\Delta = \Delta^+ + \Delta^-$ as is well known.

We will also impose as causality condition the existence of one special kernel. As we have seen the strong causality condition is in general too restrictive. Because of that we state the weak causality condition.

A. Weak Causality Condition

The infinite set of distributions $\{\sigma_{l,\alpha}^{(R)}\}$ such that relation (25) holds contains one, $\sigma_l^{(R)}$, such that

$$\sigma_l^{(R)} = \sum_{j=1}^{l} \lambda_j e^{-iw_j t} \quad \text{for} \quad t < 0, \quad (27)$$

w_j being the energies corresponding to the zeros of $h_l^-(kR)$ and λ_j complex numbers.

As a consequence the Fourier transform of $\sigma_l^{(R)}$, $\sum_l^{(R)}$ is analytic and bounded in the upper half-plane except in w_κ if $\lambda_\kappa \neq 0$. Then as in the zero-mass case $S_l(w)e^{2ikR}$ is a function of H^∞.

This causality condition could be expressed in the following way. Among all the PLS which when restricted to $w > m$ have the same effect as the scattering system, there is one which is causal in H_L^2.

A case in which all λ_j vanish is the hard core. The S matrix is then zero at w_j. On the other hand, the square well illustrates the case in which all λ_j are different from zero. Consequently S_l is not zero at w_j.

A few comments are necessary about the choice of the r value. There is no problem so far as the weak condition holds. When we increase the r value, say to $R_1 > R$, the points w_j move but we have again a weak condition. We obtain then that $S_l(w)e^{2ikR_1}$ is a function of H^∞. This is a weaker result than the one obtained for $r = R$, therefore it is not interesting. The strong condition cannot hold at two different r values in contrast to the weak condition. An argument can be given to state that it can only hold at $r = R$. Suppose $T(R_1)$ with $R_1 > R$. We can consider it as three PLS in series:

$$\psi_{l,\text{in}}(R_1) \xrightarrow{T_1} \psi_{l,\text{in}}(R) \xrightarrow{T_2} \psi_{l,\text{out}}(R) \xrightarrow{T_3} \psi_{l,\text{out}}(R_1). \quad (28)$$

Of the three systems, T_3 obeys a strong condition, T_2 may obey it, but T_1 does not. In fact T_1 is the inverse system of a PLS which obeys a "strong anticausality condition" ($\sigma = 0$ for $t > 0$). Since the total kernel is

$$\sigma = \sigma_1 * \sigma_2 * \sigma_3, \quad (29)$$

one can hope that σ is not zero for $t < 0$. Of course this is not a proof.

B. S-Matrix Factorization

We have proved that $S_l(w)e^{2ikR}$ is a function of H^∞. We can thus apply the factorization theorem of Sec. 1:

$$S_l(w)e^{2ikR} = e^{i\alpha w}S_l^E(w)S_l^B(w)S_l^S(w) \tag{30}$$

with $\alpha \geq 0$ and

$$S_l^E(w) = \exp\left[\frac{1}{\pi}\int_{-\infty}^{+\infty} \log|S_l(t)e^{2ikR}|\,\frac{wt+1}{i(t-w)}\,\frac{dt}{1+t^2}\right], \tag{31}$$

k being in (31) the momentum corresponding to an energy t.

$$S_l^B(w) = \prod_{n=1}^{\infty}\frac{w-\beta_n}{w-\beta_n^*} \quad \text{Im}\,\beta_n \geq 0, \tag{32}$$

$$S_l^S(w) = \exp\left[i\int\frac{wt+1}{w-t}\,d\mu(t)\right]. \tag{33}$$

μ has its support contained in the nonphysical region. Otherwise S_l would not be defined in some points of the physical region.

The positive constant α represents a kind of over-all retardation. It is zero for electromagnetic waves and for nonrelativistic potential scattering. From the point of view of causality, α may have any nonnegative value.

S_l^E represents the inelasticity and is unity if the scattering is elastic. It has a cut running along the points of the real axis with $|S_l(w)| < 1$. In fact from the symmetry conditions $|S_l(w)| = S\,|_l(-w)|$, there are two cuts symmetric with respect to the origin.

The Blaschke product S_l^B contains zero-pole pairs. If they are not far from the real axis they can be interpreted as resonances. They can be infinite in number.

The factor $S_l^S(w)$ does not seem to have a simple physical meaning. Toll[6] found an analogous factor and interpreted it as the effect of infinitely narrow absorption lines.

However, since the S matrix has to be well defined in the physical region, these lines must be in the nonphysical interval $(-m, +m)$ (or in $w < 0$ for nonrelativistic particles).

From (30) we conclude that

$$\delta_l(w) = \tfrac{1}{2}\alpha w - Rk + \delta_l^E(w) + \sum_{n=1}^{\infty}\arctan\frac{\Gamma_n}{2(\Omega_n - w)} + \sum_{n=1}^{\infty}\frac{A_n}{B_n - w}, \tag{34}$$

where $\beta_n = \Omega_n + \frac{1}{2}i\Gamma_n$; $A_n > 0$; $\sum|A_n| < \infty$, B_n in the nonphysical region, $\alpha \geq 0$, and δ_l^E is the contribution of the outer factor.

If the strong condition holds we would write $h_l^+(kR)/h_l^-(kR)$ instead of e^{2ikR} in (30) and $\frac{1}{2}\arg[h_l^-(kR)/h_l^+(kR)]$ instead of $(-kR)$ in (34). We should obtain

$$S_l(w)\frac{h_l^+(kR)}{h_l^-(kR)} = e^{i\alpha w}S_l^E(w)S_l^B(w)S_l^S(w), \tag{35}$$

$$\delta_l(w) = \frac{\alpha}{2}w + \tfrac{1}{2}\arg\frac{h_l^-(kR)}{h_l^+(kR)} + \delta_l^E + \sum_{n=1}^{\infty}\arctan\frac{\Gamma_n}{2(\Omega_n - w)} + \sum_{n=1}^{\infty}\frac{A_n}{B_n - w}. \tag{36}$$

4. CONCLUSIONS

We have obtained in (30) an analytic expression for the S matrix valid in the upper half-plane of the energy. If we want to extend it to the lower one we must choose the cut arising from the function $w = w(k)$ in such a way that the poles β_n^* correspond to $\text{Im}\,k < 0$. This is always possible. It must be stressed that the outer factor can have poles in this lower half-plane. We have then a analytic function in the energy plane except for an essential singularity at infinity, poles in the lower half-plane (some of them represent resonances), two cuts representing the absorption (they come from S_l^E), some essential singularities in the nonphysical region (from S_l^S) and a kinematical cut from the function $w = w(k)$. It should be emphasized that this structure is the same for the strong condition as for the weak one.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor F. Lurcat for suggesting this problem and for his constant guidance and assistance during the course of this work. I wish to express my gratitude to Professor M. Levy for his warm hospitality at the Laboratoire de Physique Théorique et Hautes Energies. The author is very indebted to Professor L. Michel for his kind help and encouragement. He is also grateful to Professor K. Chadan for some fruitful suggestions and to Professor P. Eymard for his communication of the proof of Theorem 2.1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

On the Determination of the Many-Channel Potential Matrix and S Matrix from a Single Function

JOSEPH R. COX*
Department of Physics, Florida Atlantic University, Boca Raton, Florida

(Received 4 May 1967)

It is shown that, for a certain model of many-channel potential scattering, there exists, under rather general conditions, a single function from which both the potential matrix V and the corresponding S matrix can be constructed. This function is the Fredholm determinant of the Lippmann–Schwinger equation for the physical wavefunction of a system having a potential matrix which is identical with V in the interval from the origin out to a distance r, but which vanishes identically beyond this distance. A one-channel and a two-channel example are discussed.

1. INTRODUCTION

FOR the special case of a many-channel problem, Le Couteur[1] has shown that when the various elements of the S matrix for elastic and inelastic scattering are meromorphic functions of the energy, there exists a single function from which all elements of the S matrix may be obtained in a simple way. Newton[2a] has shown that this function of Le Couteur might just as well be chosen to be the Fredholm determinant of the Lippmann–Schwinger equation, and that it is therefore, with the channel momenta considered as independent variables, a regular analytic function in the whole upper half of the complex plane of each channel momentum under the very general conditions that the potential matrix has finite first and second absolute moments.[2b] Furthermore, its zeros give directly the bound states and include the resonances. Thus the study of the S matrix is reduced, in the above sense, to the study of a single function of a single complex variable whose analytic properties are rather well understood. This fact is sometimes useful, for example, in the construction of simple models.[3–5] Other related work includes the demonstration by Blankenbecler[4] that the S matrix may also be constructed from the Fredholm determinant considered as a function of the energy only by taking functional derivatives with respect to the Green's function, and the generalization by Newton[6] of the entire procedure of expressing the elements of the S matrix in terms of the Fredholm determinant to the case of "continuous channels," which is intended to be a preliminary step toward its generalization to problems in which there are three or more particles with an associated continuum due to the possibility of dissociation of initially bound particles.

It is our purpose in this paper to show that, at least in the case of a finite number of discrete channels, a single function exists from which not only the S matrix, but also the corresponding potential matrix $V(r')$ can be constructed in a straightforward way. This function is the Fredholm determinant $f(r)$ of the Lippmann–Schwinger equation for the potential matrix which is identical with $V(r')$ from the origin out to a distance r, but which vanishes identically beyond this distance. As r, the point of truncation, approaches infinity this function evidently approaches f, the Fredholm determinant for $V(r')$ mentioned above.

It should perhaps be pointed out that if f itself is known in analytic form for a single partial wave, then the bound states and the S matrix and hence, by means of the many-channel Marchenko equations,[7] the potential matrix $V(r')$ can in principle be constructed. However, just as would be expected from an analogy with the single-channel case,[8] $V(r')$ is not in general uniquely determined from the bound states and the S matrix; instead, if there are N bound states, then an N matrix-parameter family of potential matrices is obtained. In contrast, a knowledge of $f(r)$ (rather than f) in analytic form determines $V(r')$ uniquely.

In Sec. 2, a function q of the channel wavenumbers and r is introduced and then the construction of V from a knowledge of q in analytic form is discussed (we restrict ourselves for simplicity to s waves and sufficiently well-behaved potentials). It is then noted

* Address for the 1967–68 academic year: Sterling Chemistry Laboratory, Yale University, New Haven, Conn. 06520.

[1] K. J. Le Couteur, Proc. Roy. Soc. (London) **A 256**, 115 (1960).

[2a] R. G. Newton, J. Math. Phys. **2**, 188 (1961).

[2b] For the corresponding statement which holds when the energy conservation relation between the channel momenta is imposed, see Ref. 3; see also Refs. 1, 4, and 7.

[3] J. R. Cox, J. Math. Phys. **5**, 1065 (1964).

[4] R. Blankenbecler, in *Strong Interactions and High Energy Physics*, R. G. Moorehouse, Ed. (Oliver and Boyd, Edinburgh, 1964).

[5] M. Kato, Ann. Phys. (N.Y.) **31**, 130 (1965).

[6] R. G. Newton, Bull. Am. Phys. Soc. **12**, 49 (1967).

[7] J. R. Cox, Ph.D. thesis, Indiana University (1962).

[8] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Book Company, Inc., New York, 1966).

that q is, apart from a known factor, $f(r)$. Section 3 contains illustrative examples.

2. CONSTRUCTION OF $V(r)$

We start with the n-channel s-wave radial Schrödinger equation[8]

$$\psi'' + (K^2 - V(r))\psi = 0, \tag{2.1}$$

where $V(r)$ is a real symmetric $n \times n$ matrix which depends on r only, K is the diagonal matrix of channel wavenumbers $k_1, k_2, \cdots, k_n$, and ψ is a $n \times n$ matrix which consists of n column solutions of (2.1), the solutions differing by their boundary conditions. The k's are connected by the relation

$$k_1^2 = \frac{\mu_1}{\mu_i}(k_i^2 + \Delta_i^2), \quad i = 2, \cdots, n, \tag{2.2}$$

where μ_i is the reduced mass and $\hbar^2\Delta_i^2/2\mu_i$ is the threshold energy of the ith channel.

Let us assume that all elements of $V(r)$ possess finite first absolute moments. Then there exists an $n \times n$ matrix solution ϕ of (2.1) which vanishes at $r = 0$ and whose derivative with respect to r reduces to the unit matrix there[8]:

$$\phi = 0, \phi' = 1; \quad r = 0. \tag{2.3}$$

Define q by

$$q = \det(\phi' - iK\phi). \tag{2.4}$$

We now show that if the single function q is known in analytic form for all positive r, then it determines each of the $\frac{1}{2}(n^2 + n)$ distinct elements of $V(r)$.

Equation (2.4) may be expanded as follows:

$$\begin{aligned} q = {} & (-i)^n a k_1 k_2 \cdots k_n \\ & + \sum_{j=1}^{n} (-i)^{n-1} a_j k_1 k_2 \cdots k_{j-1} k_{j+1} \cdots k_n \\ & + \sum_{\substack{j=1 \\ j<k}}^{n} \sum_{k=2}^{n} (-i)^{n-2} \\ & \times a_{j,k} k_1 k_2 \cdots k_{j-1} k_{j+1} \cdots k_{k-1} k_{k+1} \cdots k_n \\ & + \sum_{\substack{j=1 \\ j<k<l}}^{n} \sum_{k=2}^{n} \sum_{l=3}^{n} (-i)^{n-3} a_{j,k,l} k_1 k_2 \cdots k_{j-1} \\ & \times k_{j+1} \cdots k_{k-1} k_{k+1} \cdots k_{l-1} k_{l+1} \cdots k_n \\ & + \cdots + a_{1,2\ldots,n}, \end{aligned}$$

where

$$a \equiv \det \phi$$

and $a_{j,k,l,\ldots}$ denotes the determinant of the matrix obtained from ϕ by differentiating all elements in the jth, kth, lth, $\cdots$ rows with respect to r, but leaving the elements in all other rows of ϕ unchanged. This expansion of q consists of 2^n terms. Since all elements of ϕ are, according to (2.1) and (2.3), even functions of all the k's, so are the a's in each term of the expansion. Hence, if we change the sign of one or more k's in q, any particular term in the expansion will either remain unchanged or merely change sign. Thus, we can readily solve for the a's in terms of the 2^n q's obtained by choosing all possible combinations of plus and minus signs for the n k's. We obtain, for example,

$$\begin{aligned} a &= b^{-1} i^n \sum_p (-)^m q^{(p)}, \\ a_j &= b^{-1} i^{n-1} k \sum_p (-)^{m(j)} q^{(p)}, \\ a_{j,k} &= b^{-1} i^{n-2} k_j k_k \sum_p (-)^{m(j,k)} q^{(p)}, \\ a_{j,k,l} &= b^{-1} i^{n-3} k_j k_k k_l \sum_p (-)^{m(j,k,l)} q^{(p)}, \end{aligned} \tag{2.5}$$

where

$$b = 2^n \prod_{i=1}^{n} k_i,$$

$q^{(p)}$ is obtained from q by choosing some possible combination of plus and minus signs to prefix the n k's (the sum is over all 2^n possible combinations), and $m(j, k, \cdots)$ denotes the number of minus signs which remain after the signs of $k_j, k_k, \cdots$ have been excluded from the combination. (In the case of m occurring in a above, the minus signs of none of the n k's is excluded.) For example, if $n = 2$, then

$$\begin{aligned} a &= (4k_1k_2)^{-1}[-q^{++} + q^{-+} + q^{+-} - q^{--}], \\ a_1 &= (4ik_2)^{-1}[-q^{++} - q^{-+} + q^{+-} + q^{--}], \\ a_2 &= (4ik_1)^{-1}[-q^{++} + q^{-+} - q^{+-} + q^{--}], \\ a_{1,2} &= (4)^{-1}[q^{++} + q^{-+} + q^{+-} + q^{--}], \end{aligned} \tag{2.6}$$

where the superscripts on the n q's refer to the choice of signs to prefix k_1 and k_2, respectively.

Let us solve

$$-\phi'' + V\phi = K^2\phi \tag{2.7}$$

for V. If $\det \phi \neq 0$, then by Cramer's rule, we have

$$V_{ij} = \delta_{ij}k_j^2 + a(i; j'')a^{-1}, \quad i, j = 1, 2, \cdots, n, \tag{2.8}$$

where $a(i; j'')$ denotes the determinant of the matrix obtained from ϕ by replacing the ith row of ϕ by the jth row of ϕ''. Consider first, the diagonal elements of V. Using the identity

$$a(j; j'') = (a_j)' - \sum_{i \neq j}^{n} a_{i,j}, \quad j = 1, 2, \cdots, n,$$

we may write (2.8) as

$$V_{jj} = k_j^2 + \left[(a_j)' - \sum_{i \neq j}^{n} a_{i,j}\right] a^{-1}$$

which in turn may be written, according to (2.5), as

$$V_{ij} = k_j\left\{\sum_p\left[\sum_{i=1}^{n}(-)^{m(i,j)}k_i q^{(p)} - i(-)^{m(j)}q'^{(p)}\right]\right\} \times \left(\sum_p (-)^m q^{(p)}\right)^{-1}, \quad (2.9)$$

where $m(j, j) \equiv m$.

For the off-diagonal elements of V, (2.8) yields

$$V_{ij} = a(i;j'')a^{-1}, \quad i \neq j. \quad (2.10)$$

Since V is symmetric,

$$a(i;j'') = a(j;i'').$$

In addition,

$$a(i;j') = a(j;i'), \quad (2.11)$$

and

$$a_k(i;j') = a_k(j';i), \quad k \neq i,j, \quad (2.12)$$

where in accordance with the notations established above, $a_k(i;j')$ denotes the determinant of the matrix obtained from ϕ by replacing the ith and kth rows of ϕ by the jth and the kth rows of ϕ', respectively. Equation (2.11) can be established as follows. From Eq. (2.7) and $V = \tilde{V}$, where "$\sim$" denotes matrix transpose, we have

$$(\tilde{\phi}\phi' - \tilde{\phi}'\phi)' = 0$$

which, together with (2.3), implies that $\tilde{\phi}\phi'$ is symmetric:

$$\sum_{k=1}^{n}\begin{vmatrix}\phi_{kp} & \phi_{kq}\\ \phi'_{kp} & \phi'_{kq}\end{vmatrix} = 0, \quad p, q = 1, 2, \cdots, n. \quad (2.13)$$

Define A_{ij} by

$$A_{ij} = \sum_{k=1}^{n} a(i,j;k,k'), \quad i \neq j, \quad (2.14)$$

where $a(i,j;k,k')$ is the determinant of the matrix obtained from ϕ by replacing the ith and jth rows of ϕ by the kth row of ϕ and the kth row of ϕ', respectively.

Next, make a Laplace expansion of $a(i,j;k,k')$ according to its ith and jth rows. Then

$$a(i,j;k,k') = \sum_{\substack{p=1\\p<q}}^{n}\sum_{q=2}^{n}\begin{vmatrix}\phi_{kp} & \phi_{kq}\\ \phi'_{kp} & \phi'_{kq}\end{vmatrix} M_{pq}^{ij}, \quad (2.15)$$

where M_{pq}^{ij} is $(-)^{i+j+p+q}$ times the determinant of the matrix which remains after the ith and jth rows and the pth and qth columns have been suppressed. Substitution of (2.15) into (2.14) and subsequent comparison with (2.13) yields, after a change in order of summation,

$$A_{ij} = 0, \quad i \neq j. \quad (2.16)$$

However, $a(i,j;k,k') = 0$ if $k \neq i,j$, because then the kth row of ϕ appears twice. Equations (2.14) and (2.16) therefore imply (2.11). Equation (2.12) can be established in a similar fashion. Instead of (2.14), one must start with

$$A_{ijk} = \sum_{l=1}^{n} a_k(i,j;l,l'), \quad i \neq j \neq k,$$

and one then concludes that $A_{ijk} = 0$, $i \neq j \neq k$. In parallel with the above case, $a_k(i,j;l,l') = 0$ if $k \neq i,j$, because then either the lth row of ϕ appears twice ($k \neq l$) or the lth row of ϕ' appears twice ($k = l$).

Next we note that for $n = 2$, there exists the identity

$$\begin{vmatrix}\phi'_{11} & \phi'_{12}\\ \phi_{21} & \phi_{22}\end{vmatrix}\begin{vmatrix}\phi_{11} & \phi_{12}\\ \phi'_{21} & \phi'_{22}\end{vmatrix} - \begin{vmatrix}\phi'_{11} & \phi'_{12}\\ \phi'_{21} & \phi'_{22}\end{vmatrix}\begin{vmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{vmatrix} = \begin{vmatrix}\phi_{11} & \phi_{12}\\ \phi'_{11} & \phi'_{12}\end{vmatrix}\begin{vmatrix}\phi'_{21} & \phi'_{22}\\ \phi_{21} & \phi_{22}\end{vmatrix}$$

which, by making the appropriate extensionals,[9] leads to the following two identities for n arbitrary:

$$a_i a_j - a_{i,j}a = a(j;i')a(i;j'), \quad i \neq j, \quad (2.17)$$

and

$$a_{k,i}a_{k,j} - a_{k,i,j}a_k = a_k(j;i')a_k(i;j'), \quad i \neq j \neq k. \quad (2.18)$$

Equations (2.11) and (2.17) imply

$$a(i;j') = (a_i a_j - a_{ij}a)^{\frac{1}{2}}. \quad (2.19)$$

Similarly, (2.12) and (2.18) imply

$$a_k(i;j') = (a_{k,i}a_{k,j} - a_{k,i,j}a_k)^{\frac{1}{2}}. \quad (2.20)$$

With the aid of Eqs. (2.19) and (2.20), and the identity

$$a(i;j'') = a'(i;j') - \sum_{k\neq j}^{n} a_k(i;j'), \quad i \neq j,$$

Eq. (2.10) may be written in the form

$$V_{ij} = \left\{[(a_i a_j - a_{i,j}a)^{\frac{1}{2}}]' - \sum_{k\neq j}^{n}(a_{k,i}a_{k,j} - a_{k,i,j}a_k)^{\frac{1}{2}}\right\}a^{-1}, \quad i \neq j. \quad (2.21)$$

Finally, Eqs. (2.5) and (2.21) yield

$$V_{ij} = (-)^n\left\{\left[\left(\sum_{p,p'}(-)^{m+m'}N_{ij}^{pp'}\right)^{\frac{1}{2}}\right]' + i\sum_{k\neq j}^{n}k_k\left(\sum_{p,p'}(-)^{m(k)+m'(k)}N_{ij}^{pp'}\right)^{\frac{1}{2}}\right\} \times \left(\sum_p (-)^m q^{(p)}\right)^{-1}, \quad i \neq j, \quad (2.22)$$

[9] A. C. Aitken, *Determinants and Matrices* (Oliver and Boyd, Edinburgh, 1956), 9th ed., p. 103.

where

$$N_{ij}^{pp'} = 4k_ik_j[q_{ij}^{++(p')}q_{ij}^{--(p)} - q_{ij}^{+-(p')}q_{ij}^{-+(p)}], \quad (2.23)$$

and where the subscripts i and j on the q's mean that in any summation over p, the signs which prefix k_i and k_j are to be held fixed at the values indicated directly above i and j, respectively, by superscripts. For example, if $n = 3$, then the possible values of $q_{12}^{-+(p)}$ are q^{-++} and q^{-+-}. The question as to which roots to take in the square roots in (2.22) naturally arises. If the q's are known, then presumably only one choice of roots will yield V_{ij}'s which are independent of the k's, and this will be the correct choice.

Writing V out explicitly according to Eqs. (2.9), (2.22), and (2.23), we have for $n = 1$,

$$V(r) = k[k + i(q^+ + q^-)'(q^- - q^+)^{-1}], \quad (2.24)$$

and for $n = 2$,

$$\begin{aligned} V_{11}(r) &= k_1\{k_1 + [i(q^{+-} + q^{--} - q^{++} - q^{-+})' \\ &\quad + k_2(q^{++} + q^{+-} + q^{-+} + q^{--})]\, d^{-1}\}, \\ V_{22}(r) &= k_2\{k_2 + [i(q^{-+} + q^{--} - q^{++} - q^{+-})' \\ &\quad + k_1(q^{++} + q^{+-} + q^{-+} + q^{--})]\, d^{-1}\}, \\ V_{12}(r) &= V_{21}(r) = 2\{[k_1k_2(q^{++}q^{--} - q^{+-}q^{-+})]^{\frac{1}{2}}\}'\, d^{-1}, \end{aligned} \quad (2.25)$$

where

$$d = q^{++} + q^{--} - q^{-+} - q^{+-}. \quad (2.26)$$

The function q, as defined by (2.4), is related to $f(r)$, the Fredholm determinant associated with the "truncated" potential matrix $V_t(r')$, where

$$V_t(r') = \begin{cases} V(r'), & r' < r; \\ 0, & r' > r, \end{cases}$$

by[10]

$$f(r) = \epsilon^{-1}q, \quad (2.27)$$

where

$$\epsilon \equiv \exp\left[-i\left(\sum_{j=1}^{n} k_j\right)r\right]. \quad (2.28)$$

Thus, Eq. (2.27) may be used in (2.9), (2.22), and (2.23) to express all elements of V in terms of $f(r)$. No particular simplification results when this is done for the off-diagonal elements of V. However, for the diagonal elements we find

$$V_{jj} = -ik_j\left\{\sum_p (-)^{m(j)}[\epsilon f']^{(p)}\right\}\left\{\sum_p (-)^m[\epsilon f]^{(p)}\right\}^{-1}. \quad (2.29)$$

Equations (2.24) and (2.25), when written in terms of $f^{(p)}(r)$, become, respectively,

$$\begin{aligned} V(r) &= ik(e^{-ikr}f'^{+}(r) + e^{ikr}f'^{-}(r)) \\ &\quad \times (e^{ikr}f^{-}(r) - e^{-ikr}f^{+}(r))^{-1} \end{aligned} \quad (2.30)$$

and

$$\begin{aligned} V_{11}(r) &= ik_1[e^{i(k_1+k_2)r}f'^{--}(r) + e^{i(k_2-k_1)r}f'^{+-}(r) \\ &\quad - e^{-i(k_1+k_2)r}f'^{++}(r) - e^{i(k_1-k_2)r}f'^{-+}(r)]D^{-1}, \\ V_{22}(r) &= ik_2[e^{i(k_1-k_2)r}f'^{-+}(r) + e^{i(k_1+k_2)r}f'^{--}(r) \\ &\quad - e^{i(k_2-k_1)r}f'^{+-}(r) - e^{-i(k_1+k_2)r}f'^{++}(r)]D^{-1}, \\ V_{12}(r) &= V_{21}(r) \\ &= 2\{[k_1k_2(f^{++}(r)f^{--}(r) - f^{+-}(r)f^{-+}(r))]^{\frac{1}{2}}\}'D^{-1}, \end{aligned} \quad (2.31)$$

where

$$\begin{aligned} D &= e^{i(k_1+k_2)r}f^{--}(r) - e^{i(k_2-k_1)r}f^{+-}(r) \\ &\quad + e^{-i(k_1+k_2)r}f^{++}(r) - e^{i(k_1-k_2)r}f^{-+}(r). \end{aligned} \quad (2.32)$$

The Fredholm determinant f associated with $V(r')$ is related to $f(r)$ by[10]

$$f = \lim_{r\to\infty} f(r), \quad (2.33)$$

and f in turn is related to the elements S_{ij} of the S matrix by[2,8]

$$S_{ij} = f_jf^{-1}; \quad S_{ij} = S_{ii}S_{jj} - f_{ij}f^{-1}, \quad i \neq j, \quad (2.34)$$

where, if $f = f(k_1, k_2, \cdots, k_n)$, then

$$f_j \equiv f(k_1, \cdots, -k_j, \cdots, k_n)$$

and

$$f_{ij} \equiv f(k_1, \cdots, -k_i, -k_j, \cdots, k_n).$$

Consequently, a knowledge of $f(r)$ in analytic form for all r not only determines V via (2.22), (2.23), (2.27), and (2.29), but also the corresponding S matrix via (2.33) and (2.34). Again writing out the $n = 1$ and $n = 2$ cases explicitly, we have

$$S = \lim_{r\to\infty} f^{-}(r)/f^{+}(r), \quad n = 1, \quad (2.35)$$

$$S_{11} = \lim_{r\to\infty} f^{-+}(r)/f^{++}(r), \quad S_{22} = \lim_{r\to\infty} f^{+-}(r)/f^{++}(r),$$

and

$$\begin{aligned} S_{12} &= S_{21} \\ &= \lim_{r\to\infty} [(f^{++}(r)f^{--}(r) - f^{+-}(r)f^{-+}(r))^{\frac{1}{2}}/f^{++}(r)], \\ &\qquad n = 2. \end{aligned} \quad (2.36)$$

Thus far we have been assuming that $f(r)$ (or q) is known in analytic form for all r. It is then possible to change the sign of one or more k's in $f(r)$ [or $f'(r)$] and so obtain $f^{(p)}(r)$ (or $f'^{(p)}(r)$) as is needed, for example, in (2.30)–(2.32). Actually, because of the energy conservation relation (2.2), if we consider $f(r)$ as a

[10] J. R. Cox, Nuovo Cimento 37, 482 (1965).

function of k_1, then we are led to associate with $f(r)$ a k_1 Riemann surface consisting of 2^n half planes, each distinguished from the others by its particular combination of signs of the imaginary parts of all the n k's, and having branch points at $k_1 = \pm\Delta_2, \cdots, \pm\Delta_n$ which for positive (upper) sign correspond to the threshold energies of the second through nth channel, respectively.[8] If it is assumed that all elements of V possess finite first absolute moments, then ϕ as defined by (2.3) and (2.7) is an analytic function of k_1 regular on the entire k_1 Riemann surface.[8] Consequently, according to Eqs. (2.4) and (2.27), $f(r)$ has similar properties. Hence, the changing of signs of one or more k's in $f(r)$ is equivalent to making an analytic continuation along the appropriate path on its k_1 Riemann surface.[1,4,7] For example, if $n = 2$, then the analytic continuation of $f^{++}(r)$, r held fixed, along a path which circumscribes the threshold branch point $k_1 = \Delta_2$, yields $f^{-+}(r)$. Similarly, other paths yield $f^{--}(r)$ and $f^{+-}(r)$.

3. EXAMPLES

Suppose

$$f(r) = \det F(r), \tag{3.1}$$

where[11]

$$F(r) = 1 + e^{iKr}\{R'(r) - iKR(r)(K^2 + \alpha^2)^{-1} \times (\alpha K^{-1}\sin Kr \cosh \alpha r - \cos Kr \sinh \alpha r) + R(r)K^{-1} \sin Kr \sinh \alpha r\}, \tag{3.2}$$

$$R(r) = -\alpha^{-1}(\sinh \alpha r)C[1 + (4\alpha^3)^{-1} \times (\sinh 2\alpha r - 2\alpha r)C]^{-1}\alpha^{-1}, \tag{3.3}$$

C is a constant $n \times n$ real symmetric matrix, and α is a diagonal $n \times n$ matrix such that

$$\alpha_{ij} = \alpha_j\delta_{ij}, \quad \alpha_1^2 = \alpha_j^2 - \Delta_j^2, \quad i, j = 1, \cdots, n. \tag{3.4}$$

Equations (2.30)–(2.32) must hold for all energies. In computing $V(r)$, it is useful to note that, as $|k_1| \to \infty$ with Im $K > 0$, the leading terms in (3.1)–(3.3) are

$$f^+(r) \to 1, \quad f^-(r) \to U(r)(4k^2)^{-1}e^{-2ikr}, \quad n = 1, \tag{3.5}$$

and, since $k_1 \to k_2 \equiv k$,

$$\begin{aligned} f^{++}(r) &\to 1, \quad f^{--}(r) \to (16k^4)^{-1} \\ &\quad \times [U_{11}(r)U_{22}(r) - U_{12}^2(r)]e^{-4ikr}, \\ f^{-+}(r) &\to -(4k^2)^{-1}U_{11}(r)e^{-2ikr}, \\ f^{+-} &\to -(4k^2)^{-1}U_{22}(r)e^{-2ikr}, \quad n = 2, \end{aligned} \tag{3.6}$$

where

$$U(r) = 2\frac{d}{dr}[R(r) \sinh \alpha r], \quad n = 1 \text{ or } 2, \tag{3.7}$$

and

$$U(r) = \begin{pmatrix} U_{11}(r) & U_{12}(r) \\ U_{12}(r) & U_{22}(r) \end{pmatrix}, \quad n = 2. \tag{3.8}$$

Use of Eqs. (3.5) and (3.6) in Eqs. (2.30), (2.31), and (2.32) then implies that, for $n = 1$ or $n = 2$,

$$V(r) = U(r).$$

The Fredholm determinant associated with $V(r)$ is, by Eqs. (2.33), (3.1), and (3.2),

$$f = (K - i\alpha)(K + i\alpha)^{-1},$$

and consequently, the S matrix for $n = 1$ or 2 is, by Eqs. (2.35) and (2.36),

$$S = [(K + i\alpha)(K - i\alpha)^{-1}]^2.$$

In conclusion, it should be emphasized that although a single function $f(r)$ underlies the discrete many-channel problem in the sense which we have just described, the necessary and sufficient conditions which such an $f(r)$ must satisfy have not been given. Thus, a consideration of $f(r)$ should mainly be of use in the study of the relationships between and the structure of the various elements of the potential matrix and the S matrix.

[11] J. R. Cox, Ann. Phys. (N.Y.) **39**, 235 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Statistical Mechanics of Dimers on a Quadratic Lattice

ARTHUR E. FERDINAND
Department of Chemistry, Cornell University, Ithaca, New York

(Received 27 March 1967)

The partition function for the square lattice completely filled with dimers is analyzed for a finite $n \times m$ rectangular lattice with edges and for the corresponding lattice with periodic boundary conditions. The total free energy is calculated asymptotically for fixed $\xi = n/m$ up to terms $o(1/n^{2-\delta})$ for any $\delta > 0$. The bulk terms proportional to nm, the surface terms proportional to $(n + m)$ which vanish with periodic boundary conditions, and the constant terms which reveal a parity and shape dependence are expressed explicitly using dilogarithms and elliptic theta functions.

1. INTRODUCTION

IN the evaluation of the free energy of a lattice which is partially filled with dimers, one considers N_d rigid dimers, i.e., figures consisting of two linked vertices, each of which fills two nearest-neighbor sites of the lattice of N sites, while the remaining $N - 2N_d$ sites are regarded as occupied by monomers or holes. Such a model is of general interest in the study of the theory transitions and in the discussion of the thermodynamics of adsorbed films. Combinational problems of this nature are also encountered in the study of the cell theory of the liquid state, where one divides the volume of the liquid into a set of cells, the centers of which form a regular lattice. To evaluate the free energy, one then considers the number of ways of dividing a given volume into a given number of double, triple cells, etc. For these reasons, many authors[1-4] have studied the problem using truncated series expressions and other approximate methods.

The problem of evaluating the partition function of a plane square lattice *completely filled* with dimers was solved analytically by Kasteleyn,[5] Fisher,[6] and Temperley and Fisher.[7] They considered an $m \times n$ lattice filled with horizontal and vertical dimers of activities x and y, respectively, and evaluated the limiting bulk free energy per lattice site, i.e.,

$$F_{\text{bulk}} = \lim_{m,n\to\infty} \frac{1}{mn} F_{m,n}, \tag{1.1}$$

where $F_{m,n}$ is the total free energy of the finite $m \times n$ lattice. Fisher[6] also analyzed the free energy $F_{m,n}$ asymptotically to the extent of deriving the "surface or boundary" terms, that is, those proportional to $m + n$ for a lattice with edges. He gave an expression of the form

$$F_{m,n} \approx mnF_{\text{bulk}} + (2m + 2n)F_{\text{surface}} + D + O(1/n) \tag{1.2}$$

and suggested that the third significant term would be the constant D. He estimated its value numerically in the symmetric case of a lattice in which $m = n$ and $x = y$.

In this paper we extend this asymptotic analysis of the partition function for the quadratic lattice completely filled with dimers—both when the lattice has edges, and when it is a torus, i.e., when periodic boundary conditions are imposed. Our primary interest is in determining the shape dependence (the dependence on n/m) of the partition function and hence of the "so called" molecular freedom exp $[2F_{m,n}/mn]$. We show how these functions approach their thermodynamic limit as the lattice becomes infinite. This work goes beyond that of Fisher,[6] since we obtain explicit analytical expressions for the constant D in terms of elliptic theta functions, and we also consider the torus or periodic case.

To summarize our results, we define the ratios

$$\tau = x/y \tag{1.3}$$

and

$$\xi = n/m;\ \xi' = (n + 1)/(m + 1). \tag{1.4}$$

For the partition function $Z^P_{m,n}(x, y)$ of a lattice with edges, we find

$$\ln Z^P_{m,n}(x, y) = mnF_{\text{bulk}} + (2m + 2n)F_{\text{surface}} + D^P_n(\tau/\xi') + o(1/n^{2-\delta}), \tag{1.5}$$

while for the corresponding torus partition function $Z^T_{m,n}(x, y)$ we find that

$$Z^T_{m,n}(x, y) = mnF_{\text{bulk}} + D^T_n(\tau/\xi) + o(1/n^{2-\delta}) \tag{1.6}$$

with $\delta > 0$. The explicit values of the constant terms

[1] J. F. Nagle, Phys. Rev. **152**, 190 (1966).
[2] R. H. Fowler and G. S. Rushbrooke, Trans. Faraday Soc. **33**, 1272 (1937).
[3] T. S. Chang, Proc. Roy. Soc. (London) **A169**, 512 (1939).
[4] W. J. C. Orr, Trans. Faraday Soc. **40**, 306 (1944).
[5] P. W. Kasteleyn, Physica **27**, 1209 (1961).
[6] M. E. Fisher, Phys. Rev. **124**, 1664 (1961).
[7] H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, (1961).

depend on whether n is even or odd, and are given by

$$D^P_{n,\text{even}}(\tau/\xi') = \tfrac{2}{3}\ln 2 - \tfrac{1}{6}\ln(\theta_2\theta_4\theta_3^{-2}) + (1/\pi)\Lambda_2(\tau) + \tfrac{1}{4}\ln(1+\tau^2) - \tfrac{1}{2}[\ln\{\tau+(1+\tau^2)^{\frac{1}{2}}\} + \ln\{1+(1+\tau^2)^{\frac{1}{2}}\}], \quad (1.7)$$

$$D^P_{n,\text{odd}}(\tau/\xi') = \tfrac{2}{3}\ln 2 - \tfrac{1}{6}\ln(\theta_3\theta_4\theta_2^{-2}) + (1/\pi)\Lambda_2(\tau) + \tfrac{1}{4}\ln(1+\tau^2) - \tfrac{1}{2}[\ln\{\tau+(1+\tau^2)^{\frac{1}{2}}\} + \ln\{1+(1+\tau^2)^{\frac{1}{2}}\}], \quad (1.8)$$

$$D^T_{n,\text{even}}(\tau/\xi) = -\tfrac{1}{3}\ln 2 + \ln(\theta_2^2+\theta_3^2+\theta_4^2) - \tfrac{2}{3}\ln(\theta_2\theta_3\theta_4), \quad (1.9)$$

$$D^T_{n,\text{odd}}(\tau/\xi) = \tfrac{2}{3}\ln 2 + \tfrac{1}{3}\ln(\theta_2\theta_3\theta_4^{-2}), \quad (1.10)$$

with

$$\Lambda_2(\varphi) = (2i)^{-1}[L_2(i\varphi) - L_2(-i\varphi)], \quad (1.11)$$

where

$$L_2(u) = -\int_0^u d\varphi\, \varphi^{-1}\ln(1-\varphi) \quad (1.12)$$

is Euler's dilogarithm. The theta functions in these formulas are defined by

$$\theta_j = \theta_j(0, q), \quad (1.13)$$

where

$$q = \exp(-\tau\pi/\xi) \quad \text{for the torus,}$$
$$= \exp(-\tau\pi/\xi') \quad \text{for the lattice with edges.} \quad (1.14)$$

It is interesting to note that the symmetry $(n, m) \to (m, n)$, when both m and n are even, corresponds precisely to a Jacobian imaginary transformation in which the modulus q of the elliptic theta functions is transformed by $\ln q \to 1/\ln q$.

In collaboration with Fisher, the author has made a similar study for the two-dimensional Ising model,[8] and this will be published in due course.

Before commencing our detailed analysis, we state the exact results of Kasteleyn[5] and Fisher[6] on which the work is based. The partition function for an $m \times n$ plane square lattice with edges completely filled with dimers is given by the finite product

$$Z^P_{m,n}(x, y) = 2^{\frac{1}{2}m[\frac{1}{2}n]} \prod_{k=1}^{\frac{1}{2}m} \prod_{l=1}^{[\frac{1}{2}n]} \times 2\left\{x^2\cos^2\frac{l\pi}{n+1} + y^2\cos^2\frac{k\pi}{m+1}\right\} \times \begin{cases} 1 & \text{for } n \text{ even} \\ y^{\frac{1}{2}m} & \text{for } n \text{ odd} \end{cases}, \quad (1.15)$$

it being assumed that m is even. If, however, the lattice is wrapped on a torus, the partition function $Z^T_{m,n}(x, y)$ may be expressed[5] as the sum of four products in the form

$$Z^T_{m,n}(x, y) = \tfrac{1}{2}(P_1 + P_2 + P_3 + P_4), \quad (1.16)$$

where

$$P_1 = \prod_{k=1}^{\frac{1}{2}m}\prod_{l=1}^{n} 2\left[y^2\sin^2\left\{\frac{(2k-1)\pi}{m}\right\} + x^2\sin^2\left\{\frac{(2l-1)\pi}{n}\right\}\right]^{\frac{1}{2}}, \quad (1.17a)$$

$$P_2 = \prod_{k=1}^{\frac{1}{2}m}\prod_{l=1}^{n} 2\left[y^2\sin^2\left\{\frac{2k\pi}{m}\right\} + x^2\sin^2\left\{\frac{(2l-1)\pi}{n}\right\}\right]^{\frac{1}{2}}, \quad (1.17b)$$

$$P_3 = \prod_{k=1}^{\frac{1}{2}m}\prod_{l=1}^{n} 2\left[y^2\sin^2\left\{\frac{(2k-1)\pi}{m}\right\} + x^2\sin^2\left\{\frac{2l\pi}{n}\right\}\right]^{\frac{1}{2}}, \quad (1.17c)$$

$$P_4 = -\prod_{k=1}^{\frac{1}{2}m}\prod_{l=1}^{n} 2\left[y^2\sin^2\left\{\frac{2k\pi}{m}\right\} + x^2\sin^2\left\{\frac{2l\pi}{n}\right\}\right]^{\frac{1}{2}}, \quad (1.17d)$$

where again m is supposed even. Note the minus sign in (1.17d). Clearly if both m and n are odd,

$$Z^P_{m,n} \equiv Z^T_{m,n} \equiv 0.$$

In Secs. 2-5, we present a systematic analysis of the toroidal partition function (1.16) for large even n, where $\xi = n/m$ remains finite. In Sec. 6 we express the results in terms of theta functions and discuss its symmetry. In Sec. 7, Eq. (1.16) is analyzed for large odd values of n. The remainder of the paper is then devoted to a similar analysis of (1.15) for the lattice with edges.

2. REDUCTION OF DOUBLE PRODUCTS

We begin with the analysis of $Z^T_{m,n}(x, y)$ for m and n both even, and discuss the case of odd n afterwards. The product P_4 is zero for all n. The conversion of double products to single products is made possible by use of the identities[5]

$$\prod_{k=1}^{\frac{1}{2}m} 2\left[u^2 + \sin^2\left\{\frac{2k\pi}{m}\right\}\right]^{\frac{1}{2}} \equiv [|u| + (1+u^2)^{\frac{1}{2}}]^{\frac{1}{2}m} - [-|u| + (1+u^2)^{\frac{1}{2}}]^{\frac{1}{2}m}, \quad (2.1)$$

$$\prod_{k=1}^{\frac{1}{2}m} 2\left[u^2 + \sin^2\left\{\frac{(2k-1)\pi}{m}\right\}\right]^{\frac{1}{2}} \equiv [|u| + (1+u^2)^{\frac{1}{2}}]^{\frac{1}{2}m} + [-|u| + (1+u^2)^{\frac{1}{2}}]^{\frac{1}{2}m}, \quad (2.2)$$

which are valid for even m.

With the definitions

$$\tau = x/y, \quad (2.3a)$$

$$s_i(l) = \tau\sin\{(2l-1)\pi/n\} \quad (i = 1, 2), \quad (2.3b)$$

$$= \tau\sin\{2l\pi/n\} \quad (i = 3), \quad (2.3c)$$

[8] B. Kaufman, Phys. Rev. 76, 1232 (1949).

P_i can be written as

$$P_i = y^{\frac{1}{2}mn} \prod_{l=1}^{n} |\{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^{\frac{1}{2}m} + (-1)^{i+1}\{-s_i(l) + (1 + s_i^2(l))\}^{\frac{1}{2}m}| \quad (2.4)$$

for $i = 1, 2, 3$. The modulus bars in the product arise from negative values of $u = s_i(l)$ and the removal of the modulus bars in Eqs. (2.1) and (2.2). Consider the product

$$\prod_i = \prod_{l=1}^{n} \{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^{\frac{1}{2}m}. \quad (2.5)$$

Since $s_i(l)$ is antisymmetric about $l = \frac{1}{2}n$, we can express the product in (2.5) as the product of two factors over the range $1 \leq l \leq \frac{1}{2}n$ by combining the factor for l with the factor corresponding to $l - n$. In so doing the factor corresponding to $l = \frac{1}{2}n$ for $i = 3$ is squared, but this makes no difference to the argument since that factor is unity. Thus

$$\prod_i = \prod_{l=1}^{n} \{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^{\frac{1}{2}m} \times \{-s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^{\frac{1}{2}m}. \quad (2.6)$$

By combining the two factors for each value of l, we conclude immediately that

$$\prod_i = 1. \quad (2.7)$$

We can therefore rewrite Eq. (2.4) as

$$F_i = y^{\frac{1}{2}mn} \prod_{l=1}^{n} |1 + (-1)^{i+1} \times \{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^{-m}|, \quad (2.8)$$

with $i = 1, 2, 3$.

3. DECOMPOSITION OF P_i

In this section we decompose P_i into two simpler products Q_i and R_i and evaluate the former. Evidently

$$\begin{aligned} s_i(l) + (1 + s_i^2(l)) &\geq 1 \quad \text{for } 1 \leq l \leq \tfrac{1}{2}n, \\ &\leq 1 \quad \text{for } \tfrac{1}{2}n \leq l \leq n. \end{aligned} \quad (3.1)$$

We therefore rewrite (2.8) in the form

$$P_i = y^{\frac{1}{2}mn} q_i R_i^2, \quad (3.2)$$

where q_i and R_i are defined as

$$q_i = \prod_{l=1}^{[\frac{1}{2}n]} \{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^m \quad (3.3)$$

and

$$R_i = \prod_{l=1}^{[\frac{1}{2}n]} [1 + (-1)^{i+1}\{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}^{-m}]. \quad (3.4)$$

With the limit $[\frac{1}{2}n]$, these definitions remain valid for odd n.

To simplify notation, define

$$\gamma_i(l) = \ln \{s_i(l) + (1 + s_i^2(l))^{\frac{1}{2}}\}, \quad (3.5)$$

$$\gamma(\varphi) = \ln \{\tau \sin \varphi + (1 + \tau^2 \sin^2 \varphi)^{\frac{1}{2}}\}, \quad (3.6)$$

and

$$\xi = n/m. \quad (3.7)$$

We will consider only sequences of lattices in which ξ remains positive and finite as the thermodynamic limit $n, m \to \infty$ is approached.

Taylor's theorem with remainder is now used to express $\gamma_1(l)$ $[= \gamma_2(l)]$ in terms of $\gamma_3(l)$ $[= \gamma(2l\pi/n)]$. Thus we have

$$\gamma_1(l) = \gamma(\varphi_l) - (\pi/n)\gamma'(\varphi_l) + (1/2!)(\pi/n)^2\gamma''(\varphi_l) - (1/3!)(\pi/n)^3\gamma'''(\varphi_l) + (1/4!)(\pi/n)^4\gamma^{iv}(\varphi_l + \theta\pi/n) \quad (3.8)$$

where $0 \leq \theta \leq 1$, so that

$$\ln (q_1/q_3) = \frac{m\pi}{n} \sum_{l=1}^{\frac{1}{2}n} \frac{\partial}{\partial \varphi_l} \times \left\{-\gamma(\varphi_l) + \frac{1}{2!}\frac{\pi}{n}\gamma'(\varphi_l) - \frac{1}{3!}\left(\frac{\pi}{n}\right)^2\gamma''(\varphi_l)\right\} + O(1/n^2), \quad (3.9)$$

where $\varphi_l = 2l\pi/n$, and the $O(1/n^2)$ term follows from the boundedness of γ^{iv} throughout the range. In order to apply the Euler–Maclaurin theorem for summation[9]

$$\sum_{r=0}^{q} f(a + r\delta) = \frac{1}{\delta}\int_a^{a+q\delta} f(\varphi)\, d\varphi + \tfrac{1}{2}[f(a) + f(a + q\delta) + (\delta/12)[f'(a + q\delta) - f'(a)] + O(\delta^3), \quad (3.10)$$

we add and subtract the term $l = 0$ in Eq. (3.9) and finally obtain

$$\ln (q_1/q_3) = (\tau\pi/2\xi) + O(1/n^2). \quad (3.11)$$

By also introducing an $l = 0$ term in (3.3) we have

$$\ln q_3 = m \sum_{l=0}^{\frac{1}{2}n} \ln \left\{\tau \sin \frac{2l\pi}{n} + \left[1 + \tau^2 \sin^2 \left(\frac{2l\pi}{n}\right)\right]^{\frac{1}{2}}\right\}. \quad (3.12)$$

Applying the Euler–Maclaurin formula afresh gives

$$\ln q_3 = -\frac{\tau\pi}{3\xi} + \frac{mn}{2\pi}\int_0^{\pi} \gamma(\varphi)\, d\varphi + O(1/n^2). \quad (3.13)$$

This may be rewritten conveniently by introducing the function

$$\Lambda_2(\varphi) = (2i)^{-1}[L_2(i\varphi) - L_2(-i\varphi)], \quad (3.14)$$

[9] A. D. Booth, *Numerical Methods* (Butterworths Scientific Publications Ltd., London, 1957), p. 54.

where

$$L_2(u) = -\int_0^u d\varphi\ \varphi^{-1} \ln(1-\varphi), \tag{3.15}$$

is Euler's dilogarithm.[5] Finally we have

$$\ln q_3 = -\frac{\tau\pi}{3\xi} + \frac{mn}{\pi}\Lambda_2(\tau) + O\left(\frac{1}{n^2}\right) \tag{3.16}$$

and

$$\ln q_1 = \ln q_2 = \frac{\tau\pi}{6\xi} + \frac{mn}{\pi}\Lambda_2(\tau) + O\left(\frac{1}{n^2}\right). \tag{3.17}$$

For the symmetric case $x = y = \tau = 1$, this reduces to

$$\ln q_3 = -\frac{\pi}{3\xi} + \frac{mn}{\pi}G + O\left(\frac{1}{n^2}\right) \tag{3.18}$$

and

$$\ln q_1 = \ln q_2 = \frac{\pi}{6\xi} + \frac{mn}{\pi}G + O\left(\frac{1}{n_2}\right), \tag{3.19}$$

where G is Catalan's constant, namely

$$G = 1^{-2} - 3^{-2} + 5^{-2} - 7^{-2} + \cdots = 0.915\,965\,594. \tag{3.20}$$

4. ANALYSIS OF R_t

The product R_3 will now be analyzed. The products R_1 and R_2 may be discussed along quite similar lines. Using (3.4), we express R_3 in the form

$$R_3 = \prod_{l=1}^{t}[1 + \exp\{-m\gamma_3(l)\}], \\ = 2\prod_{l=1}^{t-1}[1 + \exp\{-m\gamma_3(l)\}], \tag{4.1}$$

where

$$t = \tfrac{1}{2}n. \tag{4.2}$$

The function $\exp\{\gamma_3(l)\}$, which is symmetric about $l = \frac{1}{2}t$, is strictly monotonic increasing for $1 \le l \le \frac{1}{2}t$, and is strictly monotonic decreasing for $\frac{1}{2}t \le l \le (t-1)$. We therefore write

$$R_3 = 2X^2 \quad \text{for } t \text{ even}, \\ = 2X^2\{1 + O(e^{-m})\} \quad \text{for } t \text{ odd}, \tag{4.3}$$

where

$$X = \prod_{l=1}^{\frac{1}{2}t}[1 + \exp\{-m\gamma_3(l)\}], \tag{4.4}$$

so that

$$\ln X = \sum_{p=1}^{\infty}(-1)^{p+1}p^{-1}\sum_{l=1}^{\frac{1}{2}t}\exp\{-mp\gamma_3(l)\}. \tag{4.5}$$

We analyze the sum

$$S_p = \sum_{l=1}^{\frac{1}{2}t}\exp\{-mp\gamma_3(l)\} \tag{4.6}$$

by breaking the range of summation into two regions so that

$$S_p = S_{1,p} + S_{2,p}, \tag{4.7}$$

where

$$S_{1,p} = \sum_{l=1}^{s}\exp\{-mp\gamma_3(l)\} \tag{4.8}$$

and

$$S_{2,p} = \exp\{-mp\gamma_3(s+1)\} \\ \times \sum_{l=s+1}^{\frac{1}{2}t}\exp[-mp\{\gamma_3(l) - \gamma_3(s+1)\}]. \tag{4.9}$$

Now for $(s+1) \le l \le \frac{1}{2}t$, we have $\gamma_3(l) \ge \gamma_3(s+1)$, and so

$$0 \le S_{2,p} \le (\tfrac{1}{2}t - s - 1)\exp\{-mp\gamma_3(s+1)\}. \tag{4.10}$$

To obtain an expression for $S_{1,p}$, define

$$\mu = \tau\pi/\xi = \tau\pi m/n, \tag{4.11}$$

and write

$$S_{1,p} = \sum_{l=1}^{s}\exp\{-2lp\mu\} \\ + \sum_{l=1}^{s}[\{\exp(-mp\gamma_3(l) + 2lp\mu) - 1\} \\ \times \exp(-2lp\mu)]. \tag{4.12}$$

Now expanding $\gamma_3(l)$ about $l = 0$, one sees that

$$0 \le \exp\{-mp\gamma_3(l) + 2lp\mu\} - 1 \le Ap\mu\epsilon_{t,l}, \tag{4.13}$$

where A is a suitable constant and

$$\epsilon_{t,l} = l^3t^{-2} \tag{4.14}$$

is less than, say, unity. On substituting (4.13) into (4.12) and extending the first sum to infinity, we obtain

$$0 \le S_{1,p} - \sum_{l=1}^{\infty}\exp(-2lp\mu), \\ \le \sum_{l=s+1}^{\infty}\exp(-2lp\mu) + Ap\mu t^{-2}\sum_{l=1}^{s}l^3\exp(-2lp\mu). \tag{4.15}$$

On removing the maximum value of l^3 from under the sum and extending the range of summation to infinity, we find that

$$0 \le S_{1,p} - \sum_{l=1}^{\infty}\exp(-2lp\mu), \\ \le \{1 - \exp(-2p\mu)\}^{-1} \\ \times \{Ap\mu\epsilon_{t,s}\exp(-2p\mu) + \exp(-2p\mu s)\}. \tag{4.16}$$

Finally, on summing over p to calculate $\ln X$ in (4.5), we have

$$A\mu\epsilon_{t,s}\sum_{p=1}^{\infty}(-1)^{p+1}\exp(-2p\mu)\{1 - \exp(-2p\mu)\}^{-1} \\ \le \bar{K}(\mu)\epsilon_{t,s}, \tag{4.17}$$

where $K(\mu)$ is a constant depending on μ, the series

being convergent and therefore bounded, and

$$\sum_{p=1}^{\infty}(-1)^{p+1}p^{-1}\exp(-2p\mu s)\{1-\exp(-2p\mu)\}^{-1} < \{1-\exp(-2\mu)\}^{-1}\ln\{1+\exp(-2\mu s)\}. \quad (4.18)$$

Similarly, from (4.10) the sum over $S_{2,p}$ satisfies the inequality

$$0 < \sum_{p=1}^{\infty}(-1)^{p+1}p^{-1}S_{2,p} \leq (\tfrac{1}{2}t - s - 1)\ln\{1+\exp(-m\gamma_3(s+1))\}. \quad (4.19)$$

We now choose $s(t) = t^{\frac{1}{3}\delta}$ with $0 < \delta < 2$. Then $\epsilon_{t,s} = s^3(t)t^{-2}$ vanishes as $t^{\delta-2}$ when t, m, $s(t) \to \infty$. With this choice of $s(t)$, the right-hand side of Eq. (4.18) goes to zero exponentially fast, and that of (4.19) goes to zero as $t\exp(-t^{\frac{1}{3}\delta})$. Thus we now have

$$\left|\ln X - \sum_{l=1}^{\infty}\sum_{p=1}^{\infty}(-1)^{p+1}p^{-1}\exp(-2lp\mu)\right| \leq K(\xi)/n^{2-\delta}, \quad (4.20)$$

so that finally

$$R_3 = 2\prod_{l=1}^{\infty}\{1+\exp(-2l\mu)\}^2\left\{1+o\left(\frac{1}{n^{2-\delta}}\right)\right\}. \quad (4.21)$$

By performing a similar analysis of R_1 and R_2, we obtain

$$R_1 = \prod_{l=1}^{\infty}\{1+\exp[-\mu(2l-1)]\}^2\left\{1+o\left(\frac{1}{n^{2-\delta}}\right)\right\} \quad (4.22)$$

and

$$R_2 = \prod_{l=1}^{\infty}\{1-\exp[-\mu(2l-1)]\}^2\left\{1+o\left(\frac{1}{n^{2-\delta}}\right)\right\} \quad (4.23)$$

with $\delta > 0$.

5. INTRODUCTION OF THETA FUNCTIONS

For the theta functions, we adopt the notation of Whittaker and Watson,[10] where the four theta functions are defined as

$$\theta_1(z,q) = 2G'q^{\frac{1}{4}}\sin z\prod_{l=1}^{\infty}(1-2q^{2l}\cos 2z+q^{4l}), \quad (5.1a)$$

$$\theta_2(z,q) = 2G'q^{\frac{1}{4}}\cos z\prod_{l=1}^{\infty}(1+2q^{2l}\cos 2z+q^{4l}), \quad (5.1b)$$

$$\theta_3(z,q) = G'\prod_{l=1}^{\infty}(1+2q^{2l-1}\cos 2z+q^{4l-2}), \quad (5.1c)$$

$$\theta_4(z,q) = G'\prod_{l=1}^{\infty}(1-2q^{2l-1}\cos 2z+q^{4l-2}), \quad (5.1d)$$

with

$$G' = G'\left(-\frac{1}{\pi\ln q}\right) = \prod_{l=1}^{\infty}(1-q^{2l}). \quad (5.1e)$$

We shall also adopt the convenient notation $\theta_j = \theta_j(0,q)$. The product G' is expressible in terms of theta functions by virtue of the relations

$$\theta_1' = \theta_2\theta_3\theta_4 \quad (5.2)$$

and

$$\theta_1' = 2q^{\frac{1}{4}}G'^3. \quad (5.3)$$

With the function q given by

$$q = e^{-\mu} = \exp(-xm\pi/yn), \quad (5.4)$$

we can write

$$G'(\tau/\xi) = \{2q^{\frac{1}{4}}/(\theta_2\theta_3\theta_4)\}^{-\frac{1}{3}}. \quad (5.5)$$

From Eqs. (1.16), (3.16), and (3.17), and (4.21)–(4.23), we thus obtain the final result that

$$\ln Z^T_{m,n}(x,y) = mn[(1/\pi)\Lambda_2(\tau) + \tfrac{1}{2}\ln y] + D_{n,\text{even}}(\tau/\xi) + o(1/n^{2-\delta}), \quad (5.6)$$

where

$$D_{n,\text{even}}(\tau/\xi) = -\tfrac{1}{3}\ln 2 + \ln(\theta_2^2+\theta_3^2+\theta_4^2) - \tfrac{2}{3}\ln(\theta_2\theta_3\theta_4). \quad (5.7)$$

We note the special value

$$D_{n,\text{even}}(1) = \ln(\sqrt{2}+1) = 0.881\,373\,587. \quad (5.8)$$

We expect symmetry under the interchanges $x \leftrightarrow y$, $m \leftrightarrow n$, $\tau/\xi \leftrightarrow \xi/\tau$. To verify this we use Jacobi's imaginary transformation,[10] namely,

$$\begin{aligned}(\tau/\xi)^{\frac{1}{2}}\theta_4(0 \mid i\tau/\xi) &= \theta_2(0 \mid i\xi/\tau),\\ (\tau/\xi)^{\frac{1}{2}}\theta_3(0 \mid i\tau/\xi) &= \theta_3(0 \mid i\xi/\tau),\\ (\tau/\xi)^{\frac{1}{2}}\theta_2(0 \mid i\tau/\xi) &= \theta_4(0 \mid i\xi/\tau),\end{aligned} \quad (5.9)$$

from which the result

$$D^T_{n,\text{even}}(\tau/\xi) = D^T_{n,\text{even}}(\xi/\tau) \quad (5.10)$$

is evident. Similarly one easily sees[5] that

$$\tfrac{1}{2}\ln y + (1/\pi)\Lambda_2(x/y) = \tfrac{1}{2}\ln x + (1/\pi)\Lambda_2(y/x), \quad (5.11)$$

which establishes the symmetry of the complete asymptotic partition function.

Finally, for the symmetric case $x = y = \tau = 1$, we have

$$\ln Z^T_{m,n}(1,1) = mnG/\pi + D^T_{n,\text{even}}(\xi) + o(1/n^{2-\delta}). \quad (5.12)$$

6. ODD VALUES OF n

The study of the partition function for odd values of n is interesting both mathematically [for the change in

[10] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, England, 1902), 4th ed., Chap. 21.

the explicit expression for $Z^T_{m,n}(x, y)$], and physically [because of the numerical changes which indicate the long-range correlations between dimers].

When n is odd ($= 2t + 1$), the factor in Eq. (1.17b) corresponding to $l = t + 1$ and $k = \frac{1}{2}m$ vanishes so that $P_2 = 0$. Thus, for odd n and even m, the partition function reduces to

$$Z^T_{m,n}(x, y) = \tfrac{1}{2}(P_1 + P_3). \tag{6.1}$$

To analyze P_i in Eq. (2.8), we now write $n = 2t + 1$ and split the range of the product from $l = 1$ to $2t + 1$ into two ranges which can be combined in pairs to yield

$$P_i = 2q_i R_i^2 \quad \text{for } i = 1, 3. \tag{6.2}$$

Since the graph of $\sin\{2l\pi/(2t + 1)\}$ reflects about $l = \frac{1}{2}(t + 1)$ into $\sin\{(2l - 1)\pi/(2t + 1)\}$, we have

$$P_1 = P_3. \tag{6.2a}$$

Now taking into consideration the asymmetry of $\sin\{2l\pi/(2t + 1)\}$ with respect to the two ends of the range $l = 0$ and $l = t$, the analysis of Sec. 4 yields

$$P_1 = P_3 = 2q_1 \prod_{l=1}^{\infty} \{1 + \exp(-2l\mu)\}^2 \times [1 + \exp\{-\mu(2l - 1)\}]^2\{1 + o(1/n^{2-\delta})\}. \tag{6.3}$$

The Euler–Maclaurin analysis of Sec. 3 now gives

$$\ln q_1 = (mn/\pi)\Lambda_2(\tau) - \tfrac{1}{12}\mu + O(1/n^2), \tag{6.4}$$

so that

$$\ln P_1 = (mn/\pi)\Lambda_2(\tau) - \tfrac{1}{12}\mu + \ln\left\{2\prod_{l=1}^{\infty}\{1 + \exp(-2l\mu)\}^2 \times [1 + \exp\{-\mu(2l - 1)\}]^2\right\} + o(1/n^{2-\delta}), \tag{6.5}$$

or in terms of theta functions,

$$\ln P_1 = \ln P_3 = (mn/\pi)\Lambda_2(\tau) + \tfrac{2}{3}\ln 2 + \tfrac{1}{3}\ln(\theta_2\theta_3\theta_4^{-2}) + o(1/n^{2-\delta}). \tag{6.6}$$

Finally we get the same result (5.6) as before, but with $D_{n,\text{even}}(\tau/\xi)$ replaced by

$$D^T_{n,\text{odd}}(\tau/\xi) = \tfrac{2}{3}\ln 2 + \tfrac{1}{3}\ln(\theta_2\theta_3\theta_4^{-2}). \tag{6.7}$$

We note the special value

$$D^T_{n,\text{odd}}(1) = \tfrac{2}{3}\ln 2 = 0.519\,860\,385\,5. \tag{6.8}$$

To check the symmetry of Eq. (6.6) under $n \leftrightarrow m$, we would have to retrace our steps, but starting with the initial assumption that m is odd.

From Eqs. (5.8), (5.12), and (6.8) we find for $m = n$, where n is large and even, that

$$\ln Z^T_{m,n}(1, 1) - \tfrac{1}{2}[\ln Z^T_{m,n-1}(1, 1) + \ln Z^T_{m,n+1}(1, 1)] = \ln(\sqrt{2} + 1) - \tfrac{3}{4}\ln 2 = 0.361\,513\,201\,5. \tag{6.9}$$

This indicates that the free energy is always greater for n odd than for n even (m being even in both cases). This presumably is because some types of configurations which are available for n even do not exist for n odd.

7. DIMERS ON A SQUARE LATTICE WITH EDGES

As in the previous sections, we choose m to be even so that n can be either odd or even. Then with the aid of the identity[5]

$$\prod_{k=1}^{\frac{1}{2}m} 4\left[u^2 + \cos^2\left\{\frac{k\pi}{m+1}\right\}\right] \equiv \frac{[u + (1 + u^2)^{\frac{1}{2}}]^{m+1} - [u - (1 + u^2)^{\frac{1}{2}}]^{m+1}}{2(1 + u^2)^{\frac{1}{2}}}, \tag{7.1}$$

which holds for even m, the product over k in (1.15) can be performed, with the result

$$Z^P_{m,n}(x, y) = y^{\frac{1}{2}mn}\prod_{l=1}^{[\frac{1}{2}n]} \times \frac{[\{c(l) + (1 + c^2(l))^{\frac{1}{2}}\}^{m+1} - \{c(l) - (1 + c^2(l))^{\frac{1}{2}}\}^{m+1}]}{2(1 + c^2(l))^{\frac{1}{2}}}, \tag{7.2}$$

where

$$c(l) = \tau\cos\{l\pi/(n + 1)\}. \tag{7.3}$$

With the definitions

$$q_4 = \prod_{l=1}^{[\frac{1}{2}n]}\{c(l) + (1 + c^2(l))^{\frac{1}{2}}\}^{m+1}, \tag{7.4}$$

$$q_5 = \prod_{l=1}^{[\frac{1}{2}n]}\tfrac{1}{2}\{1 + c^2(l)\}^{-\frac{1}{2}}, \tag{7.5}$$

and

$$P_0 = \prod_{l=1}^{[\frac{1}{2}n]}[1 + \{c(l) + (1 + c^2(l))^{\frac{1}{2}}\}^{-2(m+1)}], \tag{7.6}$$

the partition function (7.2) can be written in the form

$$Z^P_{m,n}(x, y) = y^{\frac{1}{2}mn}q_4q_5P_0. \tag{7.7}$$

To analyze the product P_0, we transform the cosine functions in (7.6) to sine functions by setting $n = 2t$ and writing

$$l\pi/(n + 1) = \tfrac{1}{2}\pi - \tfrac{1}{2}\pi(2t - 2l + 1)/(n + 1).$$

As l assumes the values from t to 1, the term $(2t - 2l + 1)$ assumes odd-integer values, so that

$$P_0 = \prod_{l=1}^{t}[1 + \{s(l) + (1 + s^2(l))^{\frac{1}{2}}\}^{-2(m+1)}], \tag{7.8}$$

where

$$s(l) = \tau \sin\{(2l-1)\pi/2(n+1)\}. \quad (7.9)$$

On performing an analysis quite similar to that of Sec. 4, we readily find that

$$\ln P_0 = \tfrac{1}{6}\ln\{2\theta_3^2/\theta_2\theta_4\} - \tau\pi/24\xi' + o(1/n^{2-\delta}), \quad (7.10)$$

where

$$\theta_j = \theta_j(0 \mid i\tau/\xi') \quad (7.11)$$

and

$$\xi' = (n+1)/(m+1).$$

The sum $\ln q_4$ is analyzed with the aid of the Euler–Maclaurin's theorem as before. We state the results:

$$\ln q_4 = (m+1)(n+1)(1/\pi)\Lambda_2(\tau) - \tfrac{1}{2}(m+1)\ln\{\tau + (1+\tau^2)^{\frac{1}{2}}\} + \tau\pi/24\xi' + O(1/n^2), \quad (7.12)$$

$$\ln q_5 = \tfrac{1}{2}\ln 2 + \tfrac{1}{4}\ln(1+\tau^2) - \tfrac{1}{2}(n+1)\ln\{1+(1+\tau^2)^{\frac{1}{2}}\}. \quad (7.13)$$

Substitution of (7.10), (7.12), and (7.13) into (7.7) readily yields the final result:

$$\ln Z^P_{m,n}(x,y) = mn\{\tfrac{1}{2}\ln y + (1/\pi)\Lambda_2(\tau)\} - m[\tfrac{1}{2}\ln\{\tau+(1+\tau^2)^{\frac{1}{2}}\} - (1/\pi)\Lambda_2(\tau)] - n[\tfrac{1}{2}\ln\{1+(1+\tau^2)^{\frac{1}{2}}\} - (1/\pi)\Lambda_2(\tau)] + D^P_{n,\text{even}}(\tau/\xi') + o(1/n^{2-\delta}), \quad (7.14)$$

where

$$D^P_{n,\text{even}}(\tau/\xi') = \tfrac{1}{2}\ln 2 + (1/\pi)\Lambda_2(\tau) + \tfrac{1}{4}\ln(1+\tau^2) - \tfrac{1}{2}[\ln\{\tau+(1+\tau^2)^{\frac{1}{2}}\} + \ln\{1+(1+\tau^2)^{\frac{1}{2}}\}] + \tfrac{1}{6}\ln\{2\theta_3^2/\theta_2\theta_4\}. \quad (7.15)$$

The bulk term proportional to mn is the same as for the torus. The surface terms which are proportional to $m+n$ were found by Fisher,[6] who also obtained the first four terms of the expression for $D^P_{n,\text{even}}(\tau/\xi')$.

For the symmetric case $x = y = \tau = 1$, Eq. (7.14) simplifies to

$$\ln Z^P_{m,n}(1,1) = (mn/\pi)G - (m+n)\{\tfrac{1}{2}\ln(\sqrt{2}+1) - (1/\pi)G\}^{\frac{1}{2}} + D^P_{n,\text{even}}(1/\xi') + o(1/n^{2-\delta}). \quad (7.16)$$

The symmetry of the bulk term has already been tested. For the surface terms and the term $D^P_{n,\text{even}}(\tau/\xi')$ it is again readily established by means of Jacobi's imaginary transformation (5.9) and the relation (5.11) satisfied by Euler's dilogarithm.

We note the special value for the symmetric case $x = y = \tau = 1$ and $\xi = 1$:

$$D^P_{n,\text{even}}(1) = \tfrac{3}{4}\ln 2 + (1/\pi)G - \ln(\sqrt{2}+1) + \epsilon, \quad (7.17)$$

with

$$\epsilon = \tfrac{1}{4}\ln 2 = 0.173\,286\,795, \quad (7.18)$$

which may be compared with Fisher's[6] numerical estimate for the case $m = n$ which he investigated, and whose value he gives as $\epsilon \simeq 0.170$.

8. THE ODD–EVEN EFFECT

To analyze the product P_0 for odd values of $n = 2t+1$ we transform the cosines as before by writing

$$\tfrac{1}{2}\pi l/(t+1) = \tfrac{1}{2}\pi - \tfrac{1}{2}\pi(t-l+1)/(t+1).$$

As l goes from t to 1, the term $(t-l+1)$ assumes all integer values from 1 to t. The analysis of Sec. 4 then gives

$$\ln P_0 = \tfrac{1}{6}\ln\left(\frac{\theta_2^2}{\theta_4\theta_3}\right) + \frac{\pi}{12}\left(\frac{\tau}{\xi'}\right) + o\left(\frac{1}{n^{2-\delta}}\right), \quad (8.1)$$

and the now-familiar analysis of q_4 and q_5 readily yields the results

$$\ln q_4 = (m+1)(n+1)\pi^{-1}\Lambda_2(\tau) - \tfrac{1}{2}(m+1)\ln\{\tau+(1+\tau^2)^{\frac{1}{2}}\} - \tfrac{1}{12}\pi(\tau/\xi') + O(1/n^2) \quad (8.2)$$

and

$$\ln q_5 = \ln 2 + \tfrac{1}{4}\ln(1+\tau^2) - \tfrac{1}{2}(n+1)\ln\{1+(1+\tau^2)^{\frac{1}{2}}\}. \quad (8.3)$$

On substituting (8.1), (8.2), and (8.3) into (7.7), one finds the same result as for even values of n, except that $D^P_{n,\text{even}}(\tau/\xi')$ is replaced by

$$D^P_{n,\text{odd}}(\tau/\xi') = \tfrac{1}{2}\ln 2 + (1/\pi)\Lambda_2(\tau) + \tfrac{1}{6}\ln\{2\theta_2^2/\theta_3\theta_4\} + \tfrac{1}{4}\ln(1+\tau^2) - \tfrac{1}{2}[\ln\{\tau+(1+\tau^2)^{\frac{1}{2}}\} + \ln\{1+(1+\tau^2)^{\frac{1}{2}}\}]. \quad (8.4)$$

It is interesting to note that, for large m and n,

$$\ln Z^P_{m,n}(1,1) - \tfrac{1}{2}\{\ln Z^P_{m,n-1}(1,1) + \ln Z^P_{m,n+1}(1,1)\} = \tfrac{1}{8}\ln 2 = 0.086\,643\,397\,6, \quad (8.5)$$

which again illustrates the relatively far-reaching effects of the parity of n.

9. CONCLUSIONS

The expressions (5.7), (6.7), (7.15), and (8.4) represent the final results of our analysis of the torus and the lattice with edges. It is interesting to note how

well these asymptotic expressions work even for the smallest lattices. For example, the truncated formulas (7.16), (7.17), and (8.4) give the number of dimer configurations for the plane 2×2, 2×3, 3×4, and 4×4 lattices with edges as 1.96, 2.98, 11.07, and 35.71, respectively. To the nearest whole number, these are precisely the exact values 2, 3, 11, and 36, respectively. However, for the 8×8 chessboard, the truncated expression gives to the nearest integer 12 957 925, compared with the exact value 12 988 816 given by Fisher[6] for the number of dimer configurations. Despite this numerical difference, the fractional error is only 0.24%.

ACKNOWLEDGMENTS

The author would like to thank Professor M. E. Fisher for suggesting this problem and for his interest and guidance during the course of the work. He would also like to thank the Advanced Research Projects Agency for support through the Materials Science Center at Cornell, as well as the University of the West Indies for a post-graduate award.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Solution of the Dimer Problem by the Transfer Matrix Method

ELLIOTT H. LIEB*
Physics Department, Northeastern University, Boston, Massachusetts

(Received 25 May 1967)

It is shown how the monomer–dimer problem can be formulated in terms of a transfer matrix, and hence in terms of simple spin operators as was originally done for the Ising problem. Thus, we rederive the solution to the pure dimer problem without using Pfaffians. The solution is extremely simple once one sees how to formulate the transfer matrix.

1. INTRODUCTION

SINCE Onsager's solution[1] of the two-dimensional Ising model in 1944, there has been a great deal of activity in the general area of nearest neighbor planar lattice problems. Basically, two approaches have been used.[2] One is the "algebraic" or "transfer matrix" method (used by Onsager) which focuses attention on the manner in which two neighboring rows are connected to each other. The second is the so-called "combinatorial method" whereby one studies graphs on the lattice as a whole. This was first used by Kac and Ward[3] for the Ising problem.

The most recent and concise formulation of the combinatorial method reduces the various problems to the evaluation of a Pfaffian. This method had its first "original" success in the solution to the dimer problem by Kasteleyn,[4] and for this reason is also referred to as the dimer method.

While Pfaffians have been used to rederive the solution to the Ising problem,[5] no one has yet taken the complementary step of solving the dimer problem by the transfer matrix method. The purpose of this note is to eliminate this gap. Elsewhere,[6] it has been shown how the transfer matrix method for Ising-like problems can be reduced to a few simple steps involving only fermion creation and annihilation operators. The dimer problem is likewise simple, using the transfer matrix. We also show how the more difficult and unsolved monomer–dimer problem can be formulated this way. The analogy with the problem of the Ising model in a magnetic field is very transparent, but the monomer–dimer problem is somewhat simpler and therefore there is considerable hope that this new formulation of the problem may ultimately lead to its solution.

2. FORMULATION OF THE TRANSFER MATRIX

We have a square planar lattice of M rows and N columns and hence MN vertices. A dimer is a rigid rod just long enough to cover two neighboring

* Work supported by National Science Foundation Grant GP-6851.

[1] L. Onsager, Phys. Rev. **65**, 117 (1944).

[2] We make no attempt to give a complete bibliography because an excellent one is given in H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964).

[3] M. Kac and J. C. Ward, Phys. Rev. **88**, 1332 (1952).

[4] P. W. Kasteleyn, Physica **27**, 1209 (1962); J. Math. Phys. **4**, 287 (1963); H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961); M. E. Fisher, Phys. Rev. **124**, 1664 (1961).

[5] C. A. Hurst and H. S. Green, J. Chem. Phys. **33**, 1059 (1960).

[6] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).

well these asymptotic expressions work even for the smallest lattices. For example, the truncated formulas (7.16), (7.17), and (8.4) give the number of dimer configurations for the plane 2×2, 2×3, 3×4, and 4×4 lattices with edges as 1.96, 2.98, 11.07, and 35.71, respectively. To the nearest whole number, these are precisely the exact values 2, 3, 11, and 36, respectively. However, for the 8×8 chessboard, the truncated expression gives to the nearest integer 12 957 925, compared with the exact value 12 988 816 given by Fisher[6] for the number of dimer configurations. Despite this numerical difference, the fractional error is only 0.24%.

ACKNOWLEDGMENTS

The author would like to thank Professor M. E. Fisher for suggesting this problem and for his interest and guidance during the course of the work. He would also like to thank the Advanced Research Projects Agency for support through the Materials Science Center at Cornell, as well as the University of the West Indies for a post-graduate award.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Solution of the Dimer Problem by the Transfer Matrix Method

ELLIOTT H. LIEB*
Physics Department, Northeastern University, Boston, Massachusetts

(Received 25 May 1967)

It is shown how the monomer–dimer problem can be formulated in terms of a transfer matrix, and hence in terms of simple spin operators as was originally done for the Ising problem. Thus, we rederive the solution to the pure dimer problem without using Pfaffians. The solution is extremely simple once one sees how to formulate the transfer matrix.

1. INTRODUCTION

SINCE Onsager's solution[1] of the two-dimensional Ising model in 1944, there has been a great deal of activity in the general area of nearest neighbor planar lattice problems. Basically, two approaches have been used.[2] One is the "algebraic" or "transfer matrix" method (used by Onsager) which focuses attention on the manner in which two neighboring rows are connected to each other. The second is the so-called "combinatorial method" whereby one studies graphs on the lattice as a whole. This was first used by Kac and Ward[3] for the Ising problem.

The most recent and concise formulation of the combinatorial method reduces the various problems to the evaluation of a Pfaffian. This method had its first "original" success in the solution to the dimer problem by Kasteleyn,[4] and for this reason is also referred to as the dimer method.

While Pfaffians have been used to rederive the solution to the Ising problem,[5] no one has yet taken the complementary step of solving the dimer problem by the transfer matrix method. The purpose of this note is to eliminate this gap. Elsewhere,[6] it has been shown how the transfer matrix method for Ising-like problems can be reduced to a few simple steps involving only fermion creation and annihilation operators. The dimer problem is likewise simple, using the transfer matrix. We also show how the more difficult and unsolved monomer–dimer problem can be formulated this way. The analogy with the problem of the Ising model in a magnetic field is very transparent, but the monomer–dimer problem is somewhat simpler and therefore there is considerable hope that this new formulation of the problem may ultimately lead to its solution.

2. FORMULATION OF THE TRANSFER MATRIX

We have a square planar lattice of M rows and N columns and hence MN vertices. A dimer is a rigid rod just long enough to cover two neighboring

* Work supported by National Science Foundation Grant GP-6851.

[1] L. Onsager, Phys. Rev. **65**, 117 (1944).

[2] We make no attempt to give a complete bibliography because an excellent one is given in H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964).

[3] M. Kac and J. C. Ward, Phys. Rev. **88**, 1332 (1952).

[4] P. W. Kasteleyn, Physica **27**, 1209 (1962); J. Math. Phys. **4**, 287 (1963); H. N. V. Temperley and M. E. Fisher, Phil. Mag. **6**, 1061 (1961); M. E. Fisher, Phys. Rev. **124**, 1664 (1961).

[5] C. A. Hurst and H. S. Green, J. Chem. Phys. **33**, 1059 (1960).

[6] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).

vertices (either horizontally or vertically), while a monomer covers just one vertex. An allowed configuration of the lattice is one in which every vertex is covered by a monomer or a dimer such that no vertex is covered by more than one object. If h, v, and m, are, respectively, the numbers of horizontal dimers, vertical dimers, and monomers in a configuration (with $2h + 2v + m = MN$), then the partition function Z is

$$\begin{aligned} Z &= \sum \binom{\text{ALLOWED}}{\text{CONFIGURATIONS}} x^h y^v z^m \\ &= y^{\frac{1}{2}MN} \sum \binom{\text{ALLOWED}}{\text{CONFIGURATIONS}} \alpha^h \beta^m \end{aligned} \tag{2.1}$$

where x, y, and z are the appropriate "activities" and $\alpha = x/y$, $\beta = z/(y)^{\frac{1}{2}}$.

Beginning at the "bottom" of the lattice, we have a row of N vertical bonds (V bonds) followed by a row of N horizontal bonds (H bonds), and so on alternately. There are M bond rows of each type. On each V bond of the first row, we place an arrow (or spin): an up arrow (spin +) signifies the presence of a vertical dimer on that bond, while a down arrow (spin −) signifies the absence of a dimer. The next row consists of NH bonds alternating with N vertices. If an arrow points into one of these vertices (i.e., spin + on the previous V bond) that vertex is saturated and only a down arrow must be allowed to propagate out of that vertex along the next V bond. If, however, a down arrow (spin −) comes into a vertex from below, one of three things are allowed to happen: (1) an up arrow (= dimer = spin +) can propagate along the next V bond; (2) two neighboring down arrows can cóoperate to form a horizontal dimer on the intervening H bond, which means that down arrows must propagate upwards from those two vertices; (3) a down arrow can propagate upwards from the vertex in question, signifying that a monomer has been placed on that vertex.

These rules seem complicated but they can be simply formulated if one decomposes the activity on the row of vertices into two simple steps.

Step 1: Reverse all the incoming arrows. The operator that accomplishes this is simply

$$V_1 = \prod_{i=1}^{N} \sigma_i^x, \tag{2.2}$$

where σ^x is the Pauli spin matrix $\sigma^x = \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$. The operator V_1 guarantees that an incoming up arrow (spin +) will go out as a down arrow, as required. Contrariwise, an incoming down arrow (no dimer) will propagate upwards as a V dimer (spin +). To create monomers or H dimers on the row of H bonds under consideration, it is necessary to convert some of these up arrows to down arrows and this is accomplished by the second step.

Step 2: To create a monomer at the jth vertex of the row, we multiply by σ_j^-, where $\sigma^- = \left(\begin{smallmatrix}0 & 0\\ 1 & 0\end{smallmatrix}\right)$. To create a horizontal dimer at the adjacent sites j and $j+1$, we multiply by $\sigma_j^-\sigma_{j+1}^-$. Note that after multiplication by these operators, we are insured that only down arrows (i.e., no vertical dimers) propagate onto the next row of V bonds.

The operator which generates exactly m monomers on the row is $(m!)^{-1}(\sum_{i=1}^{N} \sigma_i^-)^m$ because $(\sigma_i^-)^2 = 0$. Thus, to generate an arbitrary number of monomers with the factor β^m, we multiply by the operator

$$V_2 = \exp\left(\beta \sum_{i=1}^{N} \sigma_i^-\right). \tag{2.3}$$

Likewise, to generate H dimers, we multiply by

$$V_3 = \exp\left(\alpha \sum_{i=1}^{N} \sigma_i^-\sigma_{i+1}^-\right). \tag{2.4}$$

In Eq. (2.4), we have used cyclic boundary conditions with $\sigma_{N+1}^- \equiv \sigma_1^-$. If free ends are desired, then omit the term $\sigma_1^-\sigma_N^-$.

Thus, our transfer matrix is

$$V = V_3 V_2 V_1, \tag{2.5}$$

and

$$Z = y^{\frac{1}{2}MN} \operatorname{Tr} V^M, \tag{2.6}$$

where we have used cyclic boundary conditions in the vertical direction. We note that free edge (noncyclic) boundary conditions can be used in either or both directions with no great complication, although we shall confine ourselves here to cyclic boundary conditions.

The problem is thus reduced to computing the eigenvalues of V.

3. DIAGONALIZATION OF THE TRANSFER MATRIX

It is convenient, although not essential, to eliminate V_1. To do so, we consider $V^2 = V_3V_2V_1\ V_3V_2V_1$ and make use of the fact that σ_i^x is a unitary operator with $\sigma_i^x\sigma_i^-\sigma_i^x = \sigma_i^+ = \left(\begin{smallmatrix}0 & 0\\ 1 & 0\end{smallmatrix}\right)$. Thus,

$$V^2 = V_3V_2\ \bar{V}_3\bar{V}_2 \tag{3.1}$$

with

$$\begin{aligned} \bar{V}_2 &= \exp\left(\beta \sum_{i=1}^{N} \sigma_i^+\right), \\ \bar{V}_3 &= \exp\left(\alpha \sum_{i=1}^{N} \sigma_i^+\sigma_{i+1}^+\right). \end{aligned} \tag{3.2}$$

Thus,

$$Z = y^{\frac{1}{2}MN} \sum_j \lambda_j(N)^{\frac{1}{2}M}, \tag{3.3}$$

where $\lambda_j(N)$ (for $j = 1, 2, \cdots, 2^N$) are the eigenvalues of V^2 (we have assumed M is even).

The operator V contains two kinds of operators: quadratic forms (as in V_3 and $\bar{V}_3$) and linear forms (as in V_2 and $\bar{V}_2$). It is similar to, but simpler than, the problem of the Ising model in a magnetic field, because V_2 and V_3 commute and because V_2 is the adjoint of $\bar{V}_2$ and V_3 is the adjoint of $\bar{V}_3$. Nevertheless, we still are unable to handle the linear forms and henceforth we shall set $\beta = 0$ (i.e., we consider only the pure dimer problem).

The appropriate steps to diagonalize $V_3\bar{V}_3$ are well known. The details are in Ref. 6 and it suffices to outline the steps here. First, we transform from paulions to fermions:

$$\begin{aligned} C_j &\equiv (-1)^{j-1}\left(\prod_{i=1}^{j-1}\sigma_i^z\right)\sigma_j^-, \\ C_j^\dagger &\equiv (-1)^{j-1}\left(\prod_{i=1}^{j-1}\sigma_i^z\right)\sigma_j^+, \end{aligned} \tag{3.4}$$

with $\sigma^z = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$.

This results in

$$\begin{aligned} \sigma_j^-\sigma_{j+1}^- &= -C_jC_{j+1}, \\ \sigma_j^+\sigma_{j+1}^+ &= C_j^\dagger C_{j+1}^\dagger, \end{aligned} \tag{3.5}$$

while

$$\begin{aligned} \sigma_1^-\sigma_N^- &= (-1)^{\mathcal{N}}C_NC_1, \\ \sigma_1^+\sigma_N^+ &= -(-1)^{\mathcal{N}}C_N^\dagger C_1^\dagger, \end{aligned} \tag{3.6}$$

where $\mathcal{N} = \sum_1^N C_i^\dagger C_i$ is the number operator.

Since $(-1)^{\mathcal{N}}$ commutes with a quadratic form in fermions (or paulions), it is a constant of the motion $= +1$ for states of even $\mathcal{N}$ and -1 for states of odd $\mathcal{N}$. The latter case is cyclic while the former is anticyclic.

Next we go to running waves:

$$C_j = N^{-\frac{1}{2}}e^{-i\pi/4}\sum_q e^{iqj}\eta_q, \tag{3.7}$$

with

$$q = \pm\pi/N, \pm 3\pi/N, \cdots, \pm(N-1)\pi/N \tag{3.8a}$$

for $\mathcal{N}$ even, and

$$q = 0, \pm 2\pi/N, \pm 4\pi/N, \cdots, \pm(N-2)\pi/N, \pi \tag{3.8b}$$

for $\mathcal{N}$ odd. (We have assumed that N is even.) In terms of these running waves, we have

$$V^2 = \prod_{0\le q\le\pi} A_q \tag{3.9}$$

with

$$A_q = \exp(2\alpha \sin q\eta_q\eta_{-q})\exp(2\alpha \sin q\eta_{-q}^\dagger\eta_q^\dagger). \tag{3.10}$$

Obviously, A_q and $A_{q'}$ commute with each other. An eigenvalue of V^2 is thus the product of eigenvalues of each A_q, but we must remember to use the appropriate q values [(3.8a) or (3.8b)] according to whether the total $\mathcal{N}$ value of a state is even or odd.

To diagonalize A_q (for $q \neq 0$ and $q \neq \pi$), we first list the four basis states,

$$\begin{aligned} &\Phi_0 = |0\rangle, \quad \Phi_q = \eta_q^\dagger|0\rangle, \quad \Phi_{-q} = \eta_{-q}^\dagger|0\rangle, \\ &\Phi_{-qq} = \eta_{-q}^\dagger\eta_q^\dagger|0\rangle. \end{aligned} \tag{3.11}$$

The states Φ_q and Φ_{-q} are pure odd states and are eigenvectors of A_q with eigenvalue unity. The other two states are pure even and

$$\begin{aligned} A_q\Phi_0 &= (1 + 4\alpha^2\sin^2 q)\Phi_0 + 2\alpha \sin q\Phi_{-qq}, \\ A_q\Phi_{-qq} &= 2\alpha \sin q\Phi_0 + \Phi_{-qq}. \end{aligned} \tag{3.12}$$

The two even eigenvalues of A_q are thus

$$\lambda_{\text{even}} = [\alpha \sin q \pm (1 + \alpha^2\sin^2 q)^{\frac{1}{2}}]^2. \tag{3.13}$$

For $q = 0$ or π there are only two states: $\Phi_0 = |0\rangle$ (even) and $\Phi_q = \eta_q^\dagger|0\rangle$ (odd). Both have eigenvalue unity. These values of q are effective only for $\mathcal{N}$ odd and these pairs of states may be used to insure that $\mathcal{N}$ is always odd without affecting the total eigenvalue.

Thus, from (3.3),

$$\begin{aligned} y^{-\frac{1}{2}MN}Z = 2 &\prod_{\substack{0<q<\pi \\ q\in(3.8b)}} \{[(1 + \alpha^2\sin^2 q)^{\frac{1}{2}} + \alpha \sin q]^M \\ &+ [(1 + \alpha^2\sin^2 q)^{\frac{1}{2}} - \alpha \sin q]^M + (1 + 1)\} \\ &+ \tfrac{1}{2}\prod_{\substack{0<q<\pi \\ q\in(3.8a)}} \{[(1 + \alpha^2\sin^2 q)^{\frac{1}{2}} + \alpha \sin q]^M \\ &+ [(1 + \alpha^2\sin^2 q)^{\frac{1}{2}} - \alpha \sin q]^M + (1 + 1)\} \\ &+ \tfrac{1}{2}\prod_{\substack{0<q<\pi \\ q\in(3.8a)}} \{[(1 + \alpha^2\sin^2 q)^{\frac{1}{2}} + \alpha \sin q]^M \\ &+ [(1 + \alpha^2\sin^2 q)^{\frac{1}{2}} - \alpha \sin q]^M - (1 + 1)\}. \end{aligned} \tag{3.14}$$

The first product in (3.14) gives the contribution of the $\mathcal{N}$ odd eigenvalues while the sum of the second and third products gives the contribution of the $\mathcal{N}$ even eigenvalues. [Eq. (3.14) agrees exactly with Kasteleyn's result (Ref. 4, Eq. (25)), as may be seen using Kasteleyn's identity (Ref. 4, Eq. (26)).]

For a large lattice ($MN \to \infty$), we need consider only the largest eigenvalue of V^2, which means the product on all q of Eq. (3.13), using the plus sign. In this limit, it clearly is immaterial which set we use, (3.8a) or (3.8b). Thus,

$$\begin{aligned} &\lim_{M,N\to\infty} (MN)^{-1}\ln Z \\ &\quad = \frac{1}{2\pi}\int_0^\pi dq \ln[\alpha \sin q + (1 + \alpha^2\sin^2 q)^{\frac{1}{2}}] \\ &\qquad + \tfrac{1}{2}\ln y, \end{aligned} \tag{3.15}$$

which agrees with previous results[4] for the pure dimer problem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Functional Derivatives and Vector Meson Fields*

TUAN WU CHEN
Department of Physics, University of Toronto, Toronto, Canada

(Received 10 March 1967)

More general functional derivatives for vector mesons are introduced. As an application, a finite theory is formulated for neutral vector fields based on the "strong" Lorentz condition together with the other usual assumptions of asymptotic quantum field theory.

I. INTRODUCTION

FUNCTIONAL derivatives have been useful in treating the complicated problem of quantum field theory.[1-4] In particular, functional derivatives with respect to fields have played a part in many interesting formulations with considerable success.[2-4] Perhaps the most important feature of this functional derivative is that it enables us to treat on- and off-mass-shell properties simultaneously. For the scalar and spinor fields, the functional derivatives have been discussed in many places.[1-6] For higher-spin fields, because of the complexity caused by subsidiary conditions, there are some difficulties in defining functional derivatives.[5] Later, Rohrlich and Wilner[6] were able to overcome these difficulties for a spin-1 field by writing down the Proca equation instead of the Klein–Gordon equation.

In this work, we give a more general definition for a vector field, treating the Klein–Gordon equation and the subsidiary condition independently. The definition of Rohrlich and Wilner is a particular case of our definition. More interestingly, we can treat the Klein–Gordon equation and the subsidiary condition separately, such that the former holds only weakly while the latter holds strongly.[7] This proves to be a powerful tool for developing a neutral vector meson theory.

The well-known property of gauge invariance for a photon field also exists in a massive neutral vector field. This was explored in great detail by Feldman and Matthews.[8] The present formulation, which assumes the Lorentz condition on and off mass-shell, displays this property in a natural way.

In the next section, our definition of the functional derivative for a vector field is given. The subsidiary condition is then discussed in Sec. III. In Secs. IV and V, we have formulated a neutral vector field theory based on the functional derivatives. Finally, a brief discussion is given in Sec. VI.

II. FUNCTIONAL DERIVATIVES FOR A NEUTRAL VECTOR FIELD

As for scalar or spinor fields, we define formally the functional derivatives for a vector field as follows: Assuming that the free vector fields a^μ form a complete set so that any operator F can be expanded in terms of normal products of the free fields,[7]

$$F \stackrel{s}{=} \sum_{n=0}^{\infty} \frac{1}{n!} \int f_{\alpha_1 \cdots \alpha_n}(x_1, \cdots, x_n) : a^{\alpha_1} \cdots a^{\alpha_n} : d^4x_1 \cdots d^4x_n, \quad (2.1)$$

where x_i are four-vector coordinates, $a^{\mu_i} \equiv a^{\mu_i}(x_i)$ and $f\alpha_1 \cdots \alpha_n(x_1, \cdots, x_n)$ are c-number distributions. Then the functional derivative with respect to $a^\mu(x)$, in general, can be defined by

$$\frac{\delta F}{\delta a_\mu} \equiv \sum_{n=0}^{\infty} \frac{1}{n!} \int \Gamma^{\mu\alpha} f_{\alpha\alpha_1 \cdots \alpha_n}(x, x_1, \cdots, x_n) : a^{\alpha_1} \cdots a^{\alpha_n} : d^4x_1 \cdots d^4x_n, \quad (2.2)$$

where $\Gamma^{\mu\alpha}$ is an arbitrary second-rank c-number tensor which can be written as

$$\Gamma^{\mu\alpha} \equiv c_1 g^{\mu\alpha} + c_2 \partial^\mu \partial^\alpha,$$

or, using the more convenient notations $d^{\mu\alpha}$, $e^{\mu\alpha}$,[9] one

* Supported by the National Research Council of Canada.

[1] J. S. Schwinger, Proc. Natl. Acad. Sci. U.S. **37**, 452 (1951); K. Symanzik, Z. Naturforsch. **10**, 809 (1954); N. N. Bogoliubov and D. C. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959).

[2] R. E. Pugh, Ann. Phys. (N.Y.) **23**, 335 (1963); R. E. Pugh, J. Math. Phys. **6**, 750 (1965); **7**, 482 (1966).

[3] T. W. Chen, F. Rohrlich, and M. Wilner, J. Math. Phys. **7**, 1365 (1966).

[4] T. W. Chen, Ann. Phys. (N.Y.) **42**, 476 (1967); T. W. Chen, Ph.D. thesis, Syracuse University (1966).

[5] F. Rohrlich, J. Math. Phys. **5**, 324 (1964).

[6] F. Rohrlich and M. Wilner, J. Math. Phys. **7**, 482 (1966).

[7] For the meaning of strong and weak equalities, see the remark following (2.3). Also see Refs. 2 and 4.

[8] G. Feldman and P. T. Matthews, Phys. Rev. **130**, 1633 (1963).

[9] $d^{\mu\nu}$ and $e^{\mu\nu}$ are defined as follows (see Ref. 8): $d^{\mu\nu} \equiv g^{\mu\nu} - (p^\mu p^\nu/p^2)$, $e^{\mu\nu} \equiv p^\mu p^\nu/p^2$ in the momentum space. In x space, $e^{\mu\nu}$ is given by $e^{\mu\nu}F_\nu(x) \equiv -\partial^\mu \partial^\nu \int D_R(x-y)F_\nu(y)\, d^4y$ or simply

$$d^{\mu\nu} \equiv g^{\mu\nu} - (\partial^\mu \partial^\nu/\Box), \quad e^{\mu\nu} \equiv (\partial^\mu \partial^\nu/\Box).$$

Note the relations $e^{\mu\alpha}e^\nu_\alpha = e^{\mu\nu}$, $d^{\mu\alpha}\, d^\nu_\alpha = d^{\mu\nu}$, $e^{\mu\alpha}\, d^\nu_\alpha = 0$ (also $\partial^\alpha e_{\alpha\beta} = \partial_\beta$, $\partial^\alpha\, d_{\alpha\beta} = 0$), and $e^{\mu\nu} + d^{\mu\nu} = g^{\mu\nu}$.

can write

$$\Gamma^{\mu\alpha} = d^{\mu\alpha} + \lambda e^{\mu\alpha} \tag{2.3}$$

with an arbitrary parameter λ.

Using the same notation, $\overset{s}{=}$, as a previous work,[4] we say that two operators F and G are strongly equal

$$F \overset{s}{=} G$$

if and only if

$$\left\langle \frac{\delta^n F}{\delta a^{\mu_1} \cdots \delta a^{\mu_n}} \right\rangle_0 = \left\langle \frac{\delta^n G}{\delta a^{\mu_1} \cdots \delta a^{\mu_n}} \right\rangle_0$$

for all $n \geq 0$. We shall see many equalities which hold only weakly in that the functional derivatives of each side are unequal. This weak equality will be denoted by "$\overset{w}{=}$".

Making use of the definition (2.2), we can easily relate the coefficient functions $f_{\alpha_1 \cdots \alpha_n}(x_1, \cdots, x_n)$ to the derivatives of F. We have, obviously,

$$\left\langle \frac{\delta^n F}{\delta a^{\mu_1} \cdots \delta a^{\mu_n}} \right\rangle = \Gamma_{\mu_1\alpha_1} \cdots \Gamma_{\mu_n\alpha_n} \times f^{\alpha_1 \cdots \alpha_n}(x_1, \cdots, x_n). \tag{2.4}$$

For a finite $\lambda \neq 0$, we can also write

$$\Delta^{\mu_1 \cdots \mu_n}(x, \cdots, x_n) = \Gamma'^{\mu_1\alpha_1} \Gamma'^{\mu_n\alpha_n} \left\langle \frac{\delta^n F}{\delta a^{\alpha_1} \cdots \delta a^{\alpha_n}} \right\rangle_0, \tag{2.5}$$

where

$$\Gamma'^{\mu_i\alpha_i} \equiv d^{\mu_i\alpha_i} + (1/\lambda) e^{\mu_i\alpha_i}. \tag{2.6}$$

In particular, as $\lambda = 1$, i.e., $\Gamma^{\mu\nu} = \Gamma'^{\mu\nu} = g^{\mu\nu}$, our derivative is back to the one given by Rohrlich and Wilner.[6]

It must be noted that our definition of functional derivatives is independent of the Lorentz condition; it can be applied to any four-vector field (not necessarily a pure spin-1 field). But, for a pure spin-1 field, it is important to note that our definition is consistent with

$$[F, a^\mu(x)] \overset{s}{=} -i \int \Delta^{\mu\alpha}(x - y) \frac{\delta F}{\delta a^\alpha(y)} d^4y \tag{2.7}$$

for all λ, where $\Delta^{\mu\alpha}(x - y) \equiv d^{\mu\alpha}\Delta(x - y)$. This follows immediately from the fact that

$$\int \Delta^{\mu\alpha}(x - y) e_{\alpha\beta} \frac{\delta F}{\delta a^\beta(y)} d^4y \overset{s}{=} \int e_{\beta\alpha} \Delta^{\alpha\mu}(x - y) \frac{\delta F}{\delta a_\beta(y)} d^4y \overset{s}{=} 0,$$

since (2.7) is assumed to be a well-defined convolution and $e_{\beta\alpha} d^{\alpha\mu} = 0$. However, in the next section we see how our definition affects the Lorentz condition for all λ, including, $\lambda = 0$.

III. SUBSIDIARY CONDITION AND FUNCTIONAL DERIVATIVES

The way we define the functional derivatives for a vector field leads to a weak Klein-Gordon equation just as for a scalar field,[4] i.e.,

$$Ka^\mu(x) \overset{w}{=} 0. \tag{3.1}$$

This is because

$$\frac{\delta K a^\mu(x)}{\delta a_\nu(y)} \overset{s}{=} K \frac{\delta a^\mu(x)}{\delta a_\nu(y)} \overset{s}{=} K\Gamma^{\mu\nu}\delta(x - y) \neq 0. \tag{3.2}$$

Since (3.2) holds for all λ, Eq. (3.1) is a weak equation independent of $\Gamma^{\mu\nu}$. Now let us consider the Lorentz condition. For a spin-1 field, we know

$$\partial^\alpha a_\alpha \overset{w}{=} 0, \tag{3.3}$$

and we would like to know if this equality can hold strongly. Consider its functional derivative

$$\frac{\delta \partial^\alpha a_\alpha(x)}{\delta a^\mu(y)} \overset{s}{=} \partial^\alpha \Gamma_{\alpha\mu} \delta(x - y) \overset{s}{=} \lambda \partial_\mu \delta(x - y). \tag{3.4}$$

In (3.4), we have made use of the equality $\partial^\alpha e_{\alpha\beta} = \partial_\beta$. It is obvious that, for all $\lambda \neq 0$, (3.4) does not vanish. We conclude that the Lorentz condition can only be weak for $\lambda \neq 0$. For $\lambda = 0$, (3.4) vanishes and the Lorentz condition becomes a strong equation. In the following we shall treat this case in more detail.

When $\lambda = 0$, $\Gamma^{\mu\nu}$ becomes

$$\Gamma^{\mu\nu} = d^{\mu\nu} \equiv g^{\mu\nu} - \frac{\partial^\mu \partial^\nu}{\Box}. \tag{3.5}$$

The relation (2.4) is still valid, i.e.,

$$\left\langle \frac{\delta^n F}{\delta a_{\mu_1} \cdots \delta a_{\mu_n}} \right\rangle_0 = d^{\mu_1\alpha_1} \cdots d^{\mu_n\alpha_n} \times f_{\alpha_1 \cdots \alpha_n}(x_1, \cdots, x_n). \tag{3.6}$$

Now since $\lambda = 0$, the inverse of $\Gamma^{\mu\nu}$, i.e., (2.6), is not defined, and the expression (2.5) cannot hold. We have, rather, the following expression for the functions f:

$$f_{\mu_1 \cdots \mu_n}(x_1, \cdots, x_n) = \left\langle \frac{\delta^n F}{\delta a_1^{\mu_1} \cdots \delta a_\mu^{\mu_n}} \right\rangle_0 + \chi_{\mu_1 \cdots \mu_n}, \tag{3.7}$$

where $\chi_{\mu_1 \cdots \mu_n}$ are arbitrary functions satisfying

$$d^{\mu_i\alpha_i} \chi_{\alpha_1 \cdots \alpha_n} = 0 \quad \text{for an } i \text{ of } 1 \cdots n. \tag{3.8}$$

The expression (3.7) is obviously the most general solution of (3.6). We see that, as $\lambda = 0$, the way we define the functional derivatives can only determine the coefficient functions (e.g., the f's) of an operator (e.g., F) up to an arbitrary function $\chi_{\mu_1 \cdots \mu_n}$. Since a physical quantity is determined uniquely by the derivatives of the S operator, we can already foresee certain invariant properties from (3.7). We study

these properties in the following sections. Here we merely point out that

$$d^{\mu_1\alpha_1}\cdots d^{\mu_n\alpha_n}\frac{\delta^n F}{\delta a^{\alpha_1}\cdots\delta a^{\mu_n}}=\frac{\delta^n F}{\delta a_{\mu_1}\cdots\delta a_{\mu_n}}, \quad (3.9)$$

and so

$$\partial^{\mu_i}\frac{\delta^n F}{\delta a_1^{\mu_1}\cdots\delta a_n^{\mu_n}}\overset{s}{=}0 \quad (3.10)$$

for any i of $1\cdots n$. In particular, the current $j^\mu(x)$, defined to be

$$j^\mu(x)\equiv iS^*\frac{\delta S}{\delta a_\mu},$$

will always satisfy

$$\partial^\mu j_\mu(x)\overset{s}{=}0. \quad (3.11)$$

The strong equalities $\overset{s}{=}$ here, of course, are with respect to the functional derivatives of a^μ with $\lambda=0$.

We conclude this section with the following remarks. The functional derivative defined in Eq. (2.2) with $\lambda=0$, as we have seen above, naturally leads to the strong Lorentz condition. Conversely, if the free field is assumed to satisfy the strong Lorentz condition in addition to the weak Klein–Gordon equation, then, as it is obvious from (3.4), the functional derivative is necessarily defined as (2.2) with $\lambda=0$. The in-field in this case is of the form

$$a^\mu(x)\overset{s}{=}\int d^{\mu\alpha_1}\delta(x-x_1){:}a_{\alpha_1}{:}\,d^4x_1. \quad (3.12)$$

IV. NEUTRAL VECTOR MESON THEORY

As an application we consider a neutral vector meson with mass m. All the usual assumptions of an asymptotic quantum field theory[2–4] are assumed to be true. In addition, the free fields a^μ, which span a Hilbert space H_a, are assumed to satisfy

$$Ka^\mu(x)\overset{w}{=}0, \quad (4.1)$$

$$\partial^\alpha a_\alpha(x)\overset{s}{=}0, \quad (4.2)$$

and

$$[a^\mu(x),a^\nu(y)]=-i\,d^{\mu\nu}\Delta(x-y). \quad (4.3)$$

The functional derivative is defined by (2.2) with $\lambda=0$. In particular, we have

$$\frac{\delta a^\mu(x)}{\delta a_\nu(y)}\overset{s}{=}d^{\mu\nu}\delta(x-y) \quad (4.4)$$

and

$$[a^\mu(x),F]\overset{s}{=}-i\int d^{\mu\alpha}\Delta(x-y)\frac{\delta F}{\delta a^\alpha(y)}\,d^4y. \quad (4.5)$$

The S operator related the out-field to the in-field in the usual fashion:

$$a^\mu_{\text{out}}\equiv S^*a^\mu_{\text{in}}S,\quad S^*S\overset{s}{=}1, \quad (4.6)$$

where the S operator defined in the Hilbert space H_a is of the form ($a^\mu\equiv a^\mu_{\text{in}}$)

$$S\overset{s}{=}\sum_{n=0}^{\infty}\frac{1}{n!}\int\omega^{\alpha_1\cdots\alpha_n}(x_1,\cdots,x_n){:}a_{\alpha_1}\cdots a_{\alpha_n}{:}\,d^4x_1\cdots d^4x_n. \quad (4.7)$$

The functions ω are the coefficient functions of the S operator.

Now, because of (4.1)–(4.3), one can express a^μ in the following way:

$$a^\mu\overset{s}{=}d^{\mu\alpha}\phi_\alpha, \quad (4.8)$$

where ϕ^μ satisfies

$$Kd^{\mu\alpha}\phi_\alpha\overset{w}{=}0, \quad (4.9)$$

and

$$[\phi^\mu(x),\phi^\nu(x)]\overset{s}{=}-i\Lambda^{\mu\nu}(x-y). \quad (4.10)$$

The $\Lambda^{\mu\nu}$ are given by

$$\Lambda^{\mu\nu}(x-y)\equiv d^{\mu\nu}\Delta(x-y,m^2)+\eta^2e^{\mu\nu}\Delta(x-y,\xi^2m^2) \quad (4.11)$$

with parameters η and ξ. Our purpose here is to formulate the theory for a^μ in terms of ϕ^μ, since the renormalizability and gauge-invariance properties of our theory can then become explicit. One notes that ϕ^μ so defined is, in fact, the four-vector field discussed by Feldman and Matthews.[8] Naturally the field ϕ^μ contains a scalar field with mass μm in addition to the spin-1 field with mass m. More explicitly, ϕ^μ can be written as

$$\phi^\mu(x)\overset{s}{=}d^{\mu\alpha}a_\alpha+\frac{\eta}{\xi}\frac{\partial^\mu b}{m},$$

where the b field satisfies

$$[b(x),b(y)]\overset{s}{=}i\Delta(x-y,(\xi m)^2).$$

The b field is the scalar field and has the familiar troublesome nature as a timelike photon field. One can overcome this difficulty by various ways—for example, by introducing an indefinite metric as was done by Feldman and Matthews.[8] At any rate, here it is enough merely to point out that an Hilbert space H_ϕ can be constructed from ϕ^μ, the $\phi^\mu(x)$'s form a complete set in H_ϕ, and H_ϕ contains H_a. Therefore, our functional derivatives with respect to ϕ^μ can be defined. Since

$$[\phi^\mu(x),F]\overset{s}{=}-i\int\Lambda^{\mu\alpha}(x-y)\frac{\delta F}{\delta\phi^\alpha(y)}\,d^4y, \quad (4.12)$$

and since we have introduced the arbitrary parameter η in $\Lambda^{\mu\nu}$, it will be no loss of generality to take $\lambda=1$ for the definition (2.2). Thus, in particular, we have

$$\frac{\delta\phi^\mu(x)}{\delta\phi_\nu(y)}\overset{s}{=}g^{\mu\nu}\delta(x-y), \quad (4.13)$$

and

$$\frac{\delta a^{\mu}(x)}{\delta \phi_{\nu}(y)} \overset{s}{=} d^{\mu\nu}\delta(x-y). \tag{4.14}$$

Now let us come back to the S operator defined in (4.6), which has the expression (4.7) in $H_a \subset H_\phi$. Noting that, from (4.10),

$$\begin{aligned}[a^{\mu}(x), b(y)] &\overset{s}{=} (\eta\mu m)^{-1}[d_x^{\mu\alpha}\phi_{\alpha}(x), \partial_y^{\beta}\phi_{\beta}(y)] \\ &\overset{s}{=} (\eta\mu m)^{-1}\, d_x^{\mu\alpha}\partial_y^{\beta}[\phi_{\alpha}(x), \phi_{\beta}(y)] \\ &\overset{s}{=} 0,\end{aligned}$$

we can already see that

$$[S, b] \overset{s}{=} [S, \partial^{\alpha}\phi_{\alpha}] \overset{s}{=} 0. \tag{4.15}$$

The S operator can also be written in terms of normal products of ϕ^{μ}. In order to do so, it is most convenient to make use of functional derivatives. Since

$$\frac{\delta a^{\mu}(x)}{\delta b(y)} \overset{s}{=} 0 \quad \text{and} \quad \frac{\delta S}{\delta b(y)} \overset{s}{=} 0,$$

we immediately have

$$\begin{aligned}\frac{\delta S}{\delta \phi^{\mu_1}} &\overset{s}{=} \int \frac{\delta S}{\delta a_{\alpha}(y)} \frac{\delta a_{\alpha}(y)}{\delta \phi^{\mu_1}} d^4y \\ &\overset{s}{=} \int d_{\mu_1\alpha}\delta(x-y)\frac{\delta S}{\delta a_{\alpha}(y)} d^4y \\ &\overset{s}{=} d_{\mu_1\alpha}\frac{\delta S}{\delta a_{\alpha}} \overset{s}{=} \frac{\delta S}{\delta a^{\mu_1}}.\end{aligned} \tag{4.16}$$

Similarly, we have in general

$$\begin{aligned}\frac{\delta^n S}{\delta \phi^{\mu_1}\cdots\delta\phi^{\mu_n}} &\overset{s}{=} d^{\mu_1\alpha_1}\cdots d^{\mu_n\alpha_n}\frac{\delta^n S}{\delta a^{\alpha_1}\cdots\delta a^{\alpha_n}} \\ &\overset{s}{=} \frac{\delta^n S}{\delta a^{\mu_1}\cdots\delta a^{\mu_n}}\end{aligned} \tag{4.17}$$

for all $n > 0$.

The above results can be summarized as follows: Using the S operator of (4.6), we can now write

$$\phi^{\mu}_{\text{out}} \overset{s}{=} S^{*}\phi^{\mu}_{\text{in}}S, \tag{4.18}$$

and the S operator is of the form

$$\begin{aligned}S &\overset{s}{=} \sum_{n=0}^{\infty}\frac{1}{n!}\int\left\langle\frac{\delta^n S}{\delta\phi^{\mu_1}\cdots\delta\phi^{\mu_n}}\right\rangle_0 : \phi^{\mu_1}\cdots\phi^{\mu_n} \\ &\quad : d^4x_1\cdots d^4x_n \overset{s}{=} \sum_{n=0}^{\infty}\frac{1}{n!}\int d^{\beta_1\alpha_1}\cdots d^{\mu_n\alpha_n} \\ &\quad \times \omega_{\alpha_1\cdots\alpha_n}(x_1\cdots x_n): \phi_{\mu_1}\cdots\phi_{\mu_n}: d^4x_1\cdots d^4x_n,\end{aligned} \tag{4.19}$$

where $\omega_{\alpha_1\cdots\alpha_n}(x_1\cdots x_n)$ are the coefficient functions given in (4.7).

Therefore, we can now formulate the theory for a^{μ} in terms of ϕ^{μ} since the matrix elements between a^{μ}'s are the same as those between ϕ^{μ}'s. The fact that the matrix elements are independent of the parameters η, ξ, as was also thoroughly discussed by Feldman and Matthews,[8] can be seen to be the gauge-invariance property of our theory

In the next section, we confine ourselves to the most convenient gauge, i.e., $\eta = 1$, $\xi = 1$, and formulate our theory in the framework of the previous approach.[2-4]

V. FINITE THEORY OF NEUTRAL VECTOR MESON

The current $j^{\mu}(x)$ and the interpolating field are defined by

$$j^{\mu}(x) \equiv iS^{*}\frac{\delta S}{\delta\phi_{\mu}} \overset{s}{=} iS^{*}\frac{\delta S}{\delta a_{\mu}} \tag{5.1}$$

and the field equation

$$K\Phi^{\mu} \overset{s}{=} j^{\mu}(x) + K\phi^{\mu}(x). \tag{5.2}$$

Since we choose $\eta = \xi = 1$, i.e., Fermi–Stuckelberg gauge, the in-field $\phi^{\mu}(x)$ satisfies

$$K\phi^{\mu}(x) \overset{w}{=} 0, \tag{5.3}$$

$$[\phi^{\mu}(x), \phi^{\nu}(y)] \overset{s}{=} -ig^{\mu\nu}\Delta(x-y). \tag{5.4}$$

Making use of the asymptotic condition, we can rewrite Eq. (5.2):

$$\Phi^{\mu}(x) \overset{s}{=} \phi^{\mu}(x) - \int \Delta_R(x-y)j^{\mu}(y)\, d^4y. \tag{5.5}$$

Exactly analogous to a previous work,[3,4] the following equation can be derived from the assumption of Bogoliubov causality:

$$i\frac{\delta j^{\mu}(x)}{\delta\phi_{\nu}(y)} \overset{s}{=} K_xK_y\theta_{xy}[\Phi^{\mu}(x), \Phi^{\nu}(y)] - ig^{\mu\nu}K_x\delta(x-y). \tag{5.6}$$

One must recall that the S operator in the present formulation is always restricted by the condition (4.17). It is quite obvious that the current satisfies

$$\frac{\delta^n j^{\mu}(x)}{\delta\phi_{\mu_1}\cdots\delta\phi_{\mu_n}} \overset{s}{=} d^{\mu_1\alpha_1}\cdots d^{\mu_n\alpha_n}\frac{\delta^n j^{\mu}(x)}{\delta\phi^{\alpha_1}\cdots\delta\phi^{\alpha_n}}. \tag{5.7}$$

A more useful equation can be obtained by eliminating $\Phi^{\mu}(x)$ from (5.1), (5.4), and (5.5). In analogy to the scalar case,[2-4] we have

$$\begin{aligned}(1 - B_{12})S^{*}\frac{\delta^2 S}{\delta\phi_{\mu_1}\phi_{\mu_2}} &\overset{s}{=} -P_{12}\left(\frac{\delta S^{*}}{\delta\phi_{\mu_1}}\frac{\delta S}{\delta\phi_{\mu_2}}\right) \\ &\quad - P_{21}\left(\frac{\delta S^{*}}{\delta\phi_{\mu_2}}\frac{\delta S}{\delta\phi_{\mu_1}}\right),\end{aligned} \tag{5.8}$$

where B_{12}, P_{12} are projection operators and are given

by

$$P_{12}(F(x_1x_2)) \equiv (K_1K_2)\theta_{12}\int\Delta_A(x_1 - y_1) \times \Delta_B(x_2 - y_2)F(y_1y_2)\,dy_1\,dy_2 \quad (5.9)$$

and $1 - B_{12} \equiv P_{12} + P_{21}$. Again because of the conditions (4.17), (5.8) can also be rewritten as

$$(1 - B_{12})S^*\frac{\delta^2 S}{\delta\phi_{\mu_1}\phi_{\mu_2}} \overset{s}{=} -d^{\mu_1\alpha_1}d^{\mu_2\alpha_2}P_{12}\left(\frac{\delta S^*}{\delta\phi^{\alpha_1}}\frac{\delta S}{\delta\phi^{\alpha_2}}\right) - d^{\mu_1\alpha_1}d^{\mu_2\alpha_2}P_{12}\left(\frac{\delta S^*}{\delta\phi^{\alpha_2}}\frac{\delta S^*}{\delta\phi^{\alpha_1}}\right). \quad (5.10)$$

The solutions to (5.10) are, in general, of the form

$$S^*\frac{\delta^2 S}{\delta\phi_{\mu_1}\delta\phi_{\mu_2}} \overset{s}{=} d^{\mu_1\alpha_1}d^{\mu_2\alpha_2}\beta_{\alpha_1\alpha_2}(x_1 - x_2) - d^{\mu_1\alpha_1}d^{\mu_2\alpha_2} \times \left\{P_{12}\left(\frac{\delta S^*}{\delta\phi^{\alpha_1}}\frac{\delta S}{\delta\phi^{\alpha_2}}\right) + P_{21}\left(\frac{\delta S^*}{\delta\phi^{\alpha_2}}\frac{\delta S}{\delta\phi^{\alpha_1}}\right)\right\}, \quad (5.11)$$

where $\beta^{\mu\nu}(12)$ is an homogeneous solution to (5.10). The same arguments as given in a previous work[2-4] follow from Eq. (5.11). Since the integral equation (5.11) is the same as the one for the photon, we shall not go into the details of this argument. We merely point out here that, from Eq. (5.11), it is very convenient to obtain perturbative solutions and that, because of $\langle\phi^{\mu_1}\phi^{\mu_2}\rangle_0 = -ig^{\mu_1\mu_2}\Delta_+(x_1 - x_2)$, Eq. (5.11) gives the finite results, as is the case for scalar or photon fields.

VI. DISCUSSION

We have formulated a vector meson theory using a new definition of the functional derivative. It is seen to be applicable to interactions with currents satisfying $\partial^\mu j_\mu(x) \overset{s}{=} 0$. In the present work, we have only treated a neutral vector meson, but extensions to a charged vector meson or to interactions of a vector field and other fields are quite straightforward. It should be noted that all the notations of strong and weak equalities in previous sections are only with respect to vector fields. Consequently, the present formulation is applicable to, for example, an interaction like $\bar\psi\gamma^\mu\psi\phi_\mu$.

The present formulation reduces to a photon theory as $m = 0$. In particular, if we introduce $j^\mu(x) \equiv i{:}\,\bar\psi(x)\gamma^\mu\psi(x)$, it becomes quantum electrodynamics. This follows immediately from the fact that our basic equation (5.11) is valid for both $m \neq 0$ and $m = 0$. For $m = 0$, Eq. (5.11) is the same as Pugh's equation for quantum electrodynamics[2] except that we have $d^{\mu\nu}$'s explicitly in our equation. As an example, for the two-point functions, Eq. (5.11) becomes

$$\left\langle\frac{\delta^2 S}{\delta a_{\mu_1}\delta a_{\mu_2}}\right\rangle_0 = \left\langle\frac{\delta^2 S}{\delta\phi_{\mu_1}\delta\phi_{\mu_2}}\right\rangle_0 = -d^{\mu_1\alpha_1}d^{\mu_2\alpha_2}\{P_{12}\langle j_{\alpha_1}j_{\alpha_2}\rangle_0 + P_{21}\langle j_{\alpha_2}j_{\alpha_1}\rangle_0\}.$$

Because this is a function of $x_1 - x_2$ and only the transversal part of $\{P_{12}\langle j_{\alpha_1}j_{\alpha_2}\rangle_0 + P_{21}\langle j_{\alpha_2}j_{\alpha_1}\rangle_0\}$ will contribute, we have

$$\left\langle\frac{\delta^2 S}{\delta a_{\mu_1}\delta a_{\mu_2}}\right\rangle_0 = -\tfrac{1}{3}d^{\mu_1\mu_2}\{P_{12}\langle j_1^\alpha j_{2_\alpha}\rangle_0 + P_{21}\langle j_2^\alpha j_{1_\alpha}\rangle_0\}.$$

Substituting $j_\mu = i{:}\,\bar\psi_{\text{in}}\gamma_\mu\psi_{\text{in}}{:}$ for quantum electrodynamics (the photon with mass m and the electron with mass M), we have the second-order two-point function

$$\begin{aligned}
&\left\langle\frac{\delta^2 S}{\delta a_{\mu_1}\delta a_{\mu_2}}\right\rangle_0^{(2)} \\
&= -\tfrac{1}{3}d^{\mu_1\mu_2}\left\{3P_{12}\left[\int_{4M^2}^\infty d\kappa^2(\kappa^2 + 2M^2) \times \left(\frac{\kappa^2 - 4M^2}{\kappa^2}\right)^{\frac{1}{2}}\Delta_+(x_1 - x_2, \kappa^2)\right] + (1 \leftrightarrow 2)\right\} \\
&= d^{\mu_1\mu_2}K_1K_2\left\{\frac{i\theta(x_1 - x_2)}{3(2\pi)^2}\int_{4M^2}^\infty\frac{d\kappa^2}{(\kappa^2 - m^2)^2} \times \left(\frac{\kappa^2 - 4M^2}{\kappa^2}\right)^{\frac{1}{2}}(\kappa^2 + 2M^2)\Delta_c(x_1 - x_2, \kappa^2)\right\} \\
&= (\Box_1 - m^2)^2\left\{\frac{i}{3(2\pi)^6}\int d^4p e^{ip(x_1 - x_2)}\left(\frac{p^{\mu_1}p^{\mu_2}}{p^2} - g^{\mu_1\mu_2}\right) \times \int_{4M^2}^\infty\frac{d\kappa^2}{\kappa^2 - m^2}\left(\frac{\kappa^2 - 4M^2}{\kappa^2}\right)^{\frac{1}{2}} \times \frac{\kappa^2 + 2M^2}{\kappa^2 - m^2}\frac{1}{\kappa^2 + p^2 - i\epsilon}\right\}.
\end{aligned}$$

We see that this is the renormalized photon propagator[2] when $m = 0$.

In conclusion, we have introduced a new definition of functional derivatives for vector fields. As an application, a finite, gauge-invariant formulation for neutral vector fields is developed based on free fields satisfying the Klein–Gordon equation *weakly* but the Lorentz condition *strongly*. It produces the renormalized results of the conventional neutral vector meson theory and photon theory. The most important feature of the strong Lorentz condition is that it leads to gauge-invariant matrix elements.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor R. E. Pugh and Professor F. Rohrlich for many stimulating discussions. He also wishes to thank Professor Pugh for reading this manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Generalization of the Determinantal Method to Continuous Channels*†

ROGER G. NEWTON
Indiana University, Bloomington, Indiana

(Received 21 November 1966)

The procedure by which all elements of the S matrix in multichannel scattering problems are expressed in terms of the Fredholm determinant of the Lippmann–Schwinger equation, is generalized to the case of "continuous channels." This is a preliminary step toward its generalization to three (or more)-particle problems, in which the possibility of "ionizing" an initial bound state always introduces such a continuum. It is found that, whereas in the case of discrete channels the necessary functions can be obtained from the Fredholm determinant either by analytic continuation in the total energy, or by a "substitution rule," for continuous channels only the latter procedure works, and it is not equivalent to an analytic continuation.

IN the many-channel problem the various elements of the S matrix for elastic and inelastic scattering can be obtained directly from the Fredholm determinant $D(E)$ of the Lippmann–Schwinger equation. Let $D^-_{\alpha\beta\ldots}(E)$ be the analytic continuation of $D(E)$ along a contour that circumscribes the thresholds of the α, β, $\cdots$ channels, and no others. Then the S matrix element for elastic scattering in the α channel is given by[1]

$$S_{\alpha\alpha}(E) = D^-_\alpha(E)/D(E) \tag{1}$$

and the element for inelastic scattering from the α to the β channel, by[1]

$$S^2_{\alpha\beta} = (D^-_\alpha D^-_\beta/D^2) - (D^-_{\alpha\beta}/D). \tag{2}$$

It has been a tantalizing problem for some time to generalize these relations to the case of "continuum channels" which arise whenever more than two particles are possible as final products of a reaction. Only after such a generalization can one proceed to transfer these determinantal techniques to realistic problems involving more than two particles. In the present paper we give such a generalization.

An alternative way of defining the functions $D^-_{\alpha\beta}\ldots$ is to consider the Fredholm determinant D as a function of all the channel momenta separately:

$$D = D(k_1, k_2, \cdots).$$

We then say

$$D_{\alpha\beta}\ldots(k_1, k_2, \cdots) = D(\cdots, -k_\alpha, \cdots, -k_\beta, \cdots), \tag{3}$$

that is, the momenta $-k_\alpha$, $-k_\beta$, $\cdots$ are substituted for $k_\alpha, k_\beta, \cdots$. If Eqs. (1) and (2) were to be used for numerical calculations, that would in fact be the practical manner of calculating $D_{\alpha\beta}\ldots$. It would be quite unnecessary to perform an actual analytical continuation. The kernel K of the Lippmann–Schwinger equation

$$K = GV \tag{4}$$

contains the channel momenta explicitly. Each element of the diagonal matrix Green's function G depends on its own channel momentum only. Replacing k_α by $-k_\alpha$ means simply replacing G^+_α by G^-_α, but leaving all other elements of G as G^+. We may call this definition of $D_{\alpha\beta}\ldots$ the *substitution rule*.

The two alternative ways of defining $D^-_{\alpha\beta}\ldots$ are, of course, identical. Each k_α may be regarded as a function of the total energy E,

$$k_\alpha = (E - E_\alpha)^{\frac{1}{2}},$$

where E_α is the threshold of the α channel. The substitutions are then accomplished by analytic continuation in E around the branch points at the relevant thresholds.

Now even in the conventional many-channel problem, there are occasions when the substitution method can be used, yet it is not an analytic continuation. That happens when there is coupling between different angular momenta. Neutron–proton scattering in the triplet state of parity $(-)^{j+1}$ is the simplest such example. We may treat it as a two-channel problem in which the two-channel momenta are equal. Equations (1) to (3) are then applicable, with $k_1^2 = k_2^2$. But if we consider $D(k_1, k_2)$ as a function of $E = k_1^2 = k_2^2$, then there is no way of getting from $D(k_1, k_2)$ to $D(k_1, -k_2)$ by analytic continuation in E.

Consider now a problem with a "continuity of channels." That is, some of the matrix summations in the many-channel problem are replaced by integrals. (This is intended purely as a mathematical model and is not meant to imply that any physical reaction problem should in fact be treated that way.) In that case one cannot meaningfully speak of separate-channel momenta, and Eq. (3) becomes meaningless. There is then also no sensible way of making an analytic continuation in E in order to get

* Supported in part by the National Science Foundation and the Army Research Office (Durham)

† A condensed version of this paper was presented at the Annual Meeting of the American Physical Society, New York, January 1967.

[1] R. G. Newton, J. Math. Phys. **2**, 188 (1961).

from $D(E)$ to an analog of $D_{\alpha\beta}\ldots(E)$. But it is nevertheless meaningful to generalize the substitution rule in the following way.

When the channel index contains a continuum, then K in Eq. (4) is an integral operator in the additional sense of channel integration, with the structure

$$K(\alpha, \beta) = G(E - E_\alpha)V(\alpha, \beta).$$

We may now specify that in a certain given region, say $E_\alpha^{(1)} < E_\alpha < E_\alpha^{(2)}$, the outgoing-wave Green's function is to be replaced by an incoming-wave Green's function. This will lead to a new operator K^- and a new Fredholm determinant D^- (which, of course, depends on $E_\alpha^{(1)}$ and $E_\alpha^{(2)}$, as well as on E). The idea then is to make the interval $E_\alpha^{(2)}$–$E_\alpha^{(1)}$ very small. We see that in the limit in which it shrinks to zero we get a sensible analog of Eqs. (1) and (2).

Before proceeding to the proof of these statements, let us discuss their implications. In the conventional multichannel problem in which each channel has a different threshold, it was sufficient to consider the Fredholm determinant D simply as a function of the total energy E. In that limited context the substitution rule is equivalent to the result of an analytic continuation in E. When we go beyond this region we find that the substitution rule has a relatively simple generalization. It is then, however, not equivalent to an analytic continuation in the energy. In fact, there is no way of getting the result of the substitution rule by considering D as a function of the energy alone (and possibly, of other conserved quantities). The manipulations that lead from D to D^- must be performed in the course of its construction as a Fredholm determinant of $\mathbf{1} - GV$; they cannot be performed on the resulting function once it has been constructed. We may therefore conclude that while the equivalence of substitution rule and analytic continuation in the "two-particle sector" is an interesting and useful fact, since it has no analog in the much larger more-than-two-particle region, it should not be regarded as very fundamental. In a certain sense it may be said to be *accidental*.[2]

[2] The reason why I am stressing this point so explicitly is that there may possibly be a lesson to be learned here whose importance transcends the many-channel problem and nonrelativistic quantum mechanics. With all due caution necessary when drawing analogies among different parts of physics, it is, I believe, worthwhile nevertheless to point to a possible parallel in field theory. There too we have a substitution rule that connects collisions or reaction processes of different particles or channels. In the realm of two-particle to two-particle collisions, this substitution rule has been recognized as being equivalent to an analytic continuation. This recognition was then exploited for many fruitful dispersion theoretical techniques and finally attempts were, and are, made to put the entire analytic continuation apparatus at the very basis of the field theoretical dynamics. It is not inconceivable that here too we are victims of an "accident." Some of the analytic continuation procedures may have no meaningful generalization to the much richer area of particle production. One may then have to conclude that analyticity considerations, though useful and interesting, are not as fundamental as we thought.

Another way of stating this result succinctly is the following. In the (discrete) many-channel problem, all elements of the S matrix can in principle be determined from a knowledge of the sum of the eigenphase shifts for all energies. [We can calculate $D(E)$ from $\det S$ and then the $S_{\alpha\beta}$ by analytic continuation of $D(E)$ to its various sheets.] For continuous channels this is not true.

We now turn to the details of the substitution rule. As a preparation we outline first a convenient technique of deriving Eqs. (1) and (2) due to Sugar and Blankenbecler.[3] It obviates the necessity of introducing generalized Jost solutions and is much more direct than that of Ref. 1.

Let us first consider a simple one-channel problem of fixed angular momentum. Then $K = GV$ is the kernel of the radial Lippmann–Schwinger equation, G being the outgoing-wave Green's function

$$G(E) = (E + i\epsilon - H_0)^{-1}. \tag{5}$$

The Fredholm determinant is given by

$$D(E) = \det [\mathbf{1} - K(E)]. \tag{6}$$

Let $D^-(E)$ be the Fredholm determinant of $K^- = G^-V$,

$$G^-(E) = (E - i\epsilon - H_0)^{-1}. \tag{5'}$$

Now, if $\Delta K \equiv K - K^-$, then

$$\begin{aligned} D^- &= \det (\mathbf{1} - K^-) = \det (\mathbf{1} - K + \Delta K) \\ &= \det (\mathbf{1} - K) \det [\mathbf{1} + \Delta K(\mathbf{1} - K)^{-1}], \end{aligned}$$

or

$$D^-/D = \det [\mathbf{1} + \Delta K(\mathbf{1} - K)^{-1}]. \tag{7}$$

But

$$\begin{aligned} \Delta G &= (E + i\epsilon - H_0)^{-1} - (E - i\epsilon - H_0)^{-1} \\ &= -2\pi i\delta(E - H_0) = -2\pi iP(E), \end{aligned}$$

where $P(E)$ is the projection onto the unique[4] eigenstate of H_0 with eigenvalue E. In Dirac's notation

$$P(E) = |E\rangle\langle E|,$$

if

$$H_0 |E\rangle = E |E\rangle.$$

Hence

$$\Delta K = -2\pi iP(E)V. \tag{8}$$

Therefore,

$$\begin{aligned} D^-/D &= \det[\mathbf{1} - 2\pi iP(E)V(\mathbf{1} - K)^{-1}], \\ &= 1 - 2\pi i \langle E| V(\mathbf{1} - K)^{-1} |E\rangle, \\ &= 1 - 2\pi i \langle E| V + V(E - H)^{-1}V |E\rangle, \\ &= 1 - 2\pi iT(E) = S(E). \end{aligned} \tag{9}$$

Remembering that the Fredholm determinant equals the Jost function,[5] we have therefore found Jost's

[3] R. Sugar and R. Blankenbecler, Phys. Rev. **136**, B472 (1965).

[4] The angular momentum being fixed, there can be no degeneracy.

[5] If the latter is defined as in my book *Scattering Theory of Waves and Particles* (McGraw-Hill Book Co., New York, 1966), and not in the more usual fashion.

decomposition of the S matrix without using any of his apparatus.

The foregoing technique may now be applied in the full three-dimensional Schrödinger equation. To do that we consider the Fredholm determinant of the full Lippmann–Schwinger equation as a function of the wavenumber for each angular momentum state separately, with $k_l^2 = k_{l'}^2$, and we write

$$D_l^-(E) \equiv D(k_1, \cdots, -k_l, \cdots).$$

The corresponding ΔK then picks out the projection onto the lth angular momentum eigenstate of H_0, and we get

$$D_l^-/D = S_l.$$

In this case the result, of course, is just Eq. (9) again, because, l being conserved, the full Fredholm determinant is just the product of all the radial ones.

We now turn to the multichannel problem, where K is a matrix with channel indices α, β, of the structure

$$K_{\alpha\beta}(E) = G_\alpha(E - E_\alpha)V_{\alpha\beta}, \tag{10}$$

E_α being the threshold of the α channel. Then

$$D_\alpha^- = D(k_1, \cdots, -k_\alpha, \cdots)$$

is the Fredholm determinant of the operator obtained by changing G_α to G_α^-, but leaving all G_β, $\beta \neq \alpha$, alone. This means

$$\Delta K = -2\pi i P_\alpha(E)V,$$

where $P_\alpha(E)$ is the projection on the eigenstate of H_0 in the α channel:

$$P_\alpha(E) = |E\rangle_\alpha\,{}_\alpha\langle E|.$$

Consequently, we get

$$D_\alpha^- = 1 - 2\pi i\,{}_\alpha\langle E|\, V(\mathbf{1} - K)^{-1}\,|E\rangle_\alpha = S_{\alpha\alpha},$$

which is Eq. (1).

For inelastic collisions we write

$$D_{\alpha\beta}^- = D(k_1, \cdots, -k_\alpha, \cdots, -k_\beta, \cdots)$$

and get

$$\Delta K = -2\pi i P_{\alpha\beta}(E)V, \tag{11}$$

if $P_{\alpha\beta}(E)$ is the projection on the eigenstates of H_0 in the α and β channels.

Now if $P_{\mathcal{H}}$ is any orthogonal projection on a subspace $\mathcal{H}$, then

$$\det(\mathbf{1} - P_{\mathcal{H}}A) = \det_{\mathcal{H}}(\mathbf{1} - A_{\mathcal{H}}), \tag{12}$$

where

$$A_{\mathcal{H}} = P_{\mathcal{H}}AP_{\mathcal{H}},$$

and $\det_{\mathcal{H}}$ is the determinant of the operator on the subspace $\mathcal{H}$. Consequently, we get

$$D_{\alpha\beta}^-/D = \det{}_{\alpha\beta}\{1 - 2\pi i[\langle E|\, V(\mathbf{1} - K)^{-1}\,|E\rangle]_{\alpha\beta}\}$$
$$= \det[S]_{\alpha\beta} = \begin{vmatrix} S_{\alpha\alpha} & S_{\alpha\beta} \\ S_{\beta\alpha} & S_{\beta\beta} \end{vmatrix}.$$

Because of the symmetry of S and (1), this equation is identical with (2).

Now this procedure is readily extended to the reversal of more than two signs. Forming

$$D_{\alpha\beta}^- \ldots = D(k_1, \cdots, -k_\alpha, \cdots, -k_\beta, \cdots)$$

we need

$$\Delta K = -2\pi i P_{\alpha\beta}\ldots(E)V, \tag{13}$$

where $P_{\alpha\beta}\ldots(E)$ is the projection on the eigenstates of H_0 in the $\alpha, \beta, \cdots$ channels. The result is evidently

$$\frac{D_{\alpha\beta}^- \ldots}{D} = \det[S]_{\alpha\beta}\ldots = \begin{vmatrix} S_{\alpha\alpha} & S_{\alpha\beta} & \cdots \\ S_{\beta\alpha} & S_{\beta\beta} & \cdots \\ \cdot & \cdot\ \cdot & \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix}; \tag{14}$$

the right-hand side being the determinant of the $\alpha, \beta, \cdots$ submatrix of the S matrix. Since Eqs. (1) and (2) express each element on the right of Eq. (14) in terms of D, Eq. (14) is an equation between values of D with various sign reversals. The special case of three such sign changes is given by Eq. (4.4) of Ref. 1.

We now want to generalize these results to the case of continuum channels. Let us suppose that the channel indices form partly a continuum and partly a discrete set. When matrix multiplication calls for index summation, we may combine both into Stieltjes integrals. Therefore, K is now an integral operator not only in the sense of coordinate or momentum variables, but in the additional sense of Stieltjes channel integration with a given weight function. Its structure is (10). Let us then replace G by G^- in the region R of channel integration, and call the corresponding value of D, D_R^-. We then get

$$D_R^-/D = \det S_R, \tag{15}$$

where S_R is the submatrix of S that corresponds to the region R. We must now look at this in more detail.

Equation (15) is an exact relation, no matter how large the region R is. If R comprises all open channels, then D_R^- is obtained by replacing all positive-energy Green's functions G^+ by G^-, and leaving the negative-energy (closed-channel) Green's functions exponentially decreasing. Let us call the D_R^- so defined, simply D^-. Because of the reality of the closed-channel Green's functions, we have

$$D^- = D^*.$$

The right-hand side of (15) now is the product of all the eigenvalues of S; that is,

$$D^-/D = e^{2i\delta},$$

where δ is the sum of all the eigenphase shifts. Of course, there is a continuity of the latter. The spectrum of S is, at least partly, continuous in the present case.

We now want to let the extension of R shrink to zero. Let us first suppose R contains no discrete points. The range C of energies for which G is replaced by G^- is made small and we are interested in the leading terms in its extension only. The integral operator S_C will then differ only infinitesimally from the unit operator,

$$S_C(E; k, k') \equiv \delta(k - k') - T(E; k, k') \quad (16)$$

and we get

$$\frac{D_C^-(E)}{D(E)} = 1 - \int_C dk T(E; k, k) \quad (17)$$

with $k = (E - E_\alpha)^{\frac{1}{2}}$. The amplitude $T(E; k, k')$ is, of course, of no direct physical interest, because it leads from the continuum to the continuum and hence is the analog of three-particle to three-particle scattering, for example. But we need Eq. (17) in the next step.

Suppose now that R contains one discrete point α and a continuum C of small extension. Then

$$\begin{aligned} \text{Tr } T_R &= T_{\alpha\alpha} + \int_C dk T(k, k), \\ \text{Tr } T_R^2 &= T_{\alpha\alpha}^2 + 2\int_C dk [T_\alpha(k)]^2 + \cdots, \\ \text{Tr } T_R^n &= T_{\alpha\alpha}^n + n\int_C dk [T_\alpha(k)]^2 T_{\alpha\alpha}^{n-2} + \cdots, \end{aligned} \quad (18)$$

where $T_\alpha(k)$ is the off-diagonal element connecting the discrete α channel and the continuum k channel, and we have used the symmetry of T (assuming time reversal invariance). Because

$$\begin{aligned} \det(\mathbf{1} - T) &= \exp[\text{Tr ln}(\mathbf{1} - T)] \\ &= \exp\left\{-\sum_1^\infty \frac{1}{n} \text{tr } T^n\right\}, \end{aligned} \quad (19)$$

we therefore get

$$\begin{aligned} \det(\mathbf{1} - T_R) &= a \exp\{[1 - b + (c/a)] + \cdots\} \\ &= ab - c + \cdots, \end{aligned} \quad (20)$$

where

$$a = 1 - T_{\alpha\alpha}, \quad b = 1 - \int_C dk T(k, k),$$

$$c = \int_C dk [T_\alpha(k)]^2$$

and the result is exact to first order in the range of the continuum integration. Because of (17) and (1) we therefore find that

$$\int_C dk [T_\alpha(k)]^2 = \frac{D_\alpha^- D_C^-}{D^2} - \frac{D_{\alpha C}^-}{D}, \quad (21)$$

which is the analog of (2). It is an exact relation in the limit in which the range C of the integral on the left shrinks to zero. Since then $D_C^- \to D$ and $D_{\alpha C}^- \to D_\alpha^-$, the right-hand side also vanishes and we may write

$$[T_\alpha(k)] = \lim_{C \to 0} \left(\frac{D_\alpha^- D_C^-}{D^2} - \frac{D_{\alpha C}^-}{D}\right) \Big/ C, \quad (22)$$

if by D_C^- we mean, "Between k and $k + C$ replace G by G^-."

In Eq. (22) the continuum analog of Eq. (2) is exhibited in the form of a certain type of derivative. In this way it is related to the method of Blankenbecler[6] which uses variational derivatives with respect to the Green's function. Perhaps one should say that this is the underlying reason why that variational procedure works when it does.

It should be stressed that in the present paper the determinantal method has been generalized to *schematic* continuous channel problems only. Its use in an actual three (or more)-particle problem is still another matter. That will be pursued in a future publication.

[6] R. Blankenbecler, in *Strong Interaction and High Energy Physics*, R. G. Moorehouse, Ed. (Oliver and Boyd, Edinburgh, 1964), p. 411.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Existence and Bifurcation Theorems for the Ginzburg–Landau Equations

F. ODEH
*IBM Watson Research Center, Yorktown Heights, New York**

(Received 14 July 1966)

The second-order transition of a superconducting material from normal to superconducting state according to the Ginzburg–Landau theory is rigorously discussed. The bifurcation of a superconducting state is proved for both the Abrikosov mixed state and the case of a film in a parallel magnetic field when the flux or external field is slightly less than critical. The existence of a mixed state for all values of flux below the critical value is also proved.

INTRODUCTION

THE Ginzburg–Landau (GL) equations are a set of nonlinear partial differential equations which describe, within certain limitations on the temperature and impurity content, the behavior of superconducting materials in external magnetic fields. Although they were originally derived[1] via a phenomenological approach, they may be regarded as fundamental equations since Gor'kov[2] and others showed how to derive them from a microscopic theory of superconductivity under the assumption that the temperature is close to the transition temperature. Moreover, there is good agreement between the predictions of the GL equations and the experimental properties of both type-I and type-II superconductors.[3] However, except for approximate and numerical calculations, there seems to exist only one rigorous mathematical treatment of these interesting equations, by Carrol and Glick.[4] An existence and uniqueness theorem for the GL equations is given in Ref. 4 under the rather restrictive assumptions that the external magnetic field, as well as the parameter k in the GL equations (see Secs. 1, 2 below), are small. Of course, since the GL equations always possess a "trivial" solution corresponding to the normal state—in which the "order-parameter" function (see Sec. 1) vanishes identically—the solution referred to above is a "nontrivial" solution describing the superconducting state.

In this paper we rigorously discuss (i) the bifurcation of the nontrivial (superconducting) solution to the GL equations from the normal state as the external field, or the flux, becomes slightly less than its critical value which is determined by the linearized equations; and (ii) the existence of nontrivial solutions for all values of flux or field less than critical. We restrict the discussion to the two cases of a plane film in a parallel magnetic field and the Abrikosov mixed state[5] because of their physical significance, although the methods used are also applicable to an arbitrary smooth superconducting body in an external magnetic field. The following is an outline of the paper. In Sec. 1 we consider the ordinary differential equations form of the GL equations and give a simple proof of the existence of nontrivial solutions for values of the field near—and slightly below—the "second" critical field. The so-called "unsymmetric" or "surface-state" solutions of Saint James and De Gennes,[6] associated with the "third" critical field, are also discussed. In Sec. 2 a similar, but more complicated, proof is given to demonstrate the existence of the Abrikosov mixed state for values of flux slightly lower than the critical flux. The proofs in these two sections amount to proving the convergence of a perturbation procedure. In Sec. 3, we use the direct methods of calculus of variations to prove the existence of an Abrikosov state which minimizes the free energy for all values of flux less than critical.

1. BIFURCATION PHENOMENA FOR A FILM IN A PARALLEL MAGNETIC FIELD

Consider a superconducting slab of thickness d in a parallel magnetic field. In normalized units (see Marcus[7]) the GL equations are equivalent to

$$d^2\phi/dx^2 = k^2\phi[\phi^2 - 1 + \lambda B^2] \quad (-d < x < d), \tag{1.1}$$

$$d^2B/dx^2 = \phi^2 B, \quad (-d < x < d), \tag{1.2}$$

$$d\phi/dx = 0, \quad dB/dx = 1 \quad (x = \pm d), \tag{1.3}$$

where $\lambda = h^2$ is the square of external field, and hB = magnetic potential. It is clear that $\phi \equiv 0$,

* Present Address: Mathematics Department, American University of Beirut, Beirut, Lebanon.

[1] V. Ginzburg and L. Landau, Zh. Eksperim. i Teor. Fiz. **20**, 1064 (1950).

[2] L. Gor'kov, Zh. Eksperim. i Teor. Fiz. **36**, 1918 (1959) [English Transl.: Soviet Phys.—JETP **9**, 1364 (1959)].

[3] See, for example, A. M. Toxen and M. Burns, Phys. Rev. **130**, 1808 (1963), where other references are given.

[4] R. Carrol and A. Glick, Arch. Ratl. Mech. Anal. **16**, 373 (1964).

[5] A. Abrikosov, Zh. Eksperim. i Teor. Fiz. **32**, 1442 (1957) [English Transl.: Soviet Phys.—JETP **5**, 1174 (1957)].

[6] D. Saint-James and P. DeGennes, Phys. Letters **7**, 306 (1963).

[7] P. Marcus, Research Note NC-489, T. J. Watson Research Center, Yorktown Heights, N.Y. (1964). Also in *Proceedings of the Conference on the Physics of Type II Superconductivity* (Western Reserve University Press, Cleveland, Ohio, 1964), Vol. 3.

$B = (x + c)$, where c is an arbitrary constant, is a solution to (1.1)–(1.3). This solution corresponds to the normal state and will be referred to as the "trivial" solution. As in classical bifurcation theory, one seeks nontrivial solutions which "split off" this normal solution and hence are close enough to it. We therefore select a convenient parameter, call it ϵ, which measures the deviation of the nontrivial solution from the normal one; for example, we may choose ϵ to be proportional to

$$\|\phi\|^2 \equiv \int_{-d}^{+d} \phi^2 \, dx.$$

Introducing the functions $\psi = \epsilon^{-\frac{1}{2}}\phi$,

$$A = \epsilon^{-1}[B - (x + c)],$$

and substituting in Eqs. (1.1)–(1.3), we obtain

$$L_\epsilon \psi \equiv k^{-2}\psi'' + \{1 - \lambda(\epsilon)[x + c(\epsilon)]^2\}\psi = \epsilon f_1[\psi, A, \lambda, \epsilon] \quad (1.4)$$

and

$$MA \equiv A'' = N[x, \psi, c] + \epsilon f_2[\psi, A], \quad (1.5)$$

with the boundary conditions $\psi' = A' = 0$ at $\pm d$. Here, $N = (x + c)\psi^2$;

$$f_1 = \psi^3 + \lambda\psi[2(x + c)A + \epsilon A^2];$$

$f_2 = \psi^2 A$; and the primes denote differentiation with respect to x. It is clear that there is no loss of generality in restricting A to be orthogonal to constants, i.e., $\int A \, dx = 0$, since the mean value of A may be absorbed into the constant c. To discuss the existence of nontrivial solutions to Eqs. (1.1)–(1.3) it is sufficient to consider the equivalent equations (1.4)–(1.5). In the limit $\epsilon = 0$, (1.4) becomes a linear equation which has, for any fixed c, nonzero solutions for a discrete set of values $\lambda_n(c)$, of the eigenvalue parameter $\lambda = h^2$. It is easy to check that $\lambda_n \to -\infty$ as $n \to \infty$, but that there exists at least one positive eigenvalue. Let $\lambda_0 = h_0^2$ denote the largest positive eigenvalue and $\psi_0(x)$ the associated normalized eigenfunction which is clearly unique because it has no nodes. We can, at this stage, put $\epsilon = 0$ and $\psi = \psi_0$ in (1.5) and solve for the first approximations A_0, c_0 of A, c, but it will prove more convenient not to do so. Instead, we argue as follows: if the right side of (1.5) were orthogonal to 1, i.e., if

$$\int_{-d}^{d} \psi^2(c)[(x + c) + \epsilon A] = 0, \quad (1.6)$$

then one can invert (1.5) by means of a generalized Green's function for the operator M to get

$$A - \int_{-d}^{d} G(x, \xi)[(\xi + c)\psi^2 + \epsilon\psi^2 A] \, d\xi = 0. \quad (1.7)$$

We now discuss first the existence of solutions to the slightly different problem defined by (1.4) and (1.7) and then arrange for the constant c to satisfy (1.6), thus obtaining a solution to the original problem. If we set

$$A_0 = \int G(x, \xi)\psi_0^2(\xi)\xi \, d\xi,$$

then the vector function $u_0 = (\lambda_0, \psi_0, A_0)$ is a solution to Eqs. (1.4)–(1.7) when $\epsilon = 0$, and hence it corresponds to a trivial solution to the nonlinear problem. Following a standard argument we look, for small enough ϵ, for a solution $u(\epsilon) = [\lambda_0 + \epsilon\lambda_1(\epsilon), \psi_0 + \psi_1(\epsilon), A(\epsilon)]$, where ψ_1 is orthogonal to ψ_0. Substituting in (1.5) we find that ψ_1, λ_1 must satisfy

$$L_0\psi_1 = \epsilon[\lambda_1(\epsilon)(x + c)^2(\psi_0 + \psi_1) + f_1(u(\epsilon), \epsilon)]. \quad (1.8)$$

Since L_0 is self-adjoint, the right side of (1.8) must be orthogonal to ψ_0, which leads to the bifurcation equation

$$\int_{-d}^{d} \lambda_1(\epsilon)(x + c)^2\psi_0(\psi_0 + \psi_1) + \psi_0 f_1(u(\epsilon), \epsilon) = 0. \quad (1.9)$$

Inverting the left side of (1.8), we obtain

$$\psi_1 - \epsilon\int_{-d}^{d} K(x, \xi)[\lambda_1(x + c)^2(\psi_0 + \psi_1) + f_1(\psi, A)] = 0, \quad (1.10)$$

where K, the generalized Green's function for L_0, is a continuous function which may be expressed in terms of the eigenfunctions of L_0. Consider now the system of equations (1.9), (1.10), and (1.7). The left sides of these equations define a nonlinear map $N(u) = N[\lambda_1, \psi_1, A; \epsilon]$ in the Banach space $B = I \otimes C \otimes C$ where I denotes the real numbers and C the space of continuous functions on $[-d, d]$ with the maximum norm. The map N is bounded and continuously differentiable in B and ϵ. This follows from the continuity of the kernels K and G and the differentiability of the nonlinear perturbations f_1, f_2 in the space B. For $\epsilon = 0$, the equation $N(u) = 0$ has the solution $\psi_1 = 0$, $A = A_0$, and

$$\lambda_1 = -\left[\int_{-d}^{d} f_1(\psi_0, A_0, \lambda_0)\psi_0\right] \times \left[\int_{-d}^{d} \psi_0^2(x + c)^2\right]^{-1}.$$

To prove the existence of a unique solution of $N(u, \epsilon) = 0$ for small enough ϵ, we calculate the Frechet derivative (Jacobian) of the map N at this particular solution. This is obtained by functionally differentiating N with respect to λ_1, ψ_1, A and is given by the matrix of linear operators

$$J = \begin{pmatrix} \alpha_1 & R_1 & R_2 \\ 0 & I & 0 \\ 0 & P & I \end{pmatrix}, \quad (1.11)$$

where

$$\alpha_1 = \int_{-d}^{d} \psi_0^2(x + c)^2\, dx,$$

$$R_1 = \int_{-d}^{d} \{\lambda_1(0)\psi_0^2(x + c)^2 + \psi_0[3\psi_0^2 + 2\lambda_0 A_0(x + c)]\}\, dx,$$

$$R_2 = 2\lambda_0 \int_{-d}^{d} (x + c)\psi_0^2\, dx,$$

and P is the integral operator defined for any continuous function $f(x)$ by

$$Pf(x) = -2\int G(x, \xi)(\xi + c)\psi_0(\xi)f(\xi)\, d\xi,$$

and I is the identity operator. From the form of J and the fact that P is a bounded operator, J is clearly nonsingular. In fact, the form of J indicates the correct perturbation procedure. One solves for the second component ψ of u, then the third, A, then the first component, λ_1. Now we turn to the bifurcation equation (1.6) which is an algebraic equation in c, ϵ. For $\epsilon = 0$ it reduces to

$$\int_{-d}^{d} \psi_0^2(c_0)(x + c_0) = 0,$$

which is a rather complicated equation involving hypergeometric functions. By inspection, $c_0 = 0$ is a solution. A calculation by Saint James and De Gennes[6] shows that it has other solutions $c_{0,i}$. Moreover, for $c_0 = 0$, (1.6) is solvable, for small ϵ, since its derivative r with respect to c at $\epsilon = 0$ is given by

$$r(c_0) = \int_{-d}^{d} \psi_0^2(c_0) + 2c_0 \int_{-d}^{d} \psi_0 \frac{\partial \psi}{\partial c}, \qquad (1.12)$$

which is nonzero if $c_0 = 0$. We assume that $r(c_{0,i})$ also does not vanish; then we have the following:

Theorem 1: There exists an $\epsilon_0 > 0$ such that if $0 < \epsilon \leq \epsilon_0$, there exist nontrivial solutions of Eqs. (1.4)–(1.5).

Proof: Since J is nonsingular, the soft implicit function theorem in Banach spaces (see Ehrmann[8]) proves that the system (1.7), (1.9), (1.10) has a solution for small enough ϵ. This solution defines a nontrivial function $u(c, \epsilon)$. Since the orthogonality relation (1.6) is also solvable for c when ϵ is small, the function $u(c(\epsilon), \epsilon)$ provides a nontrivial solution to (1.4)–(1.5) which proves the theorem.

Remarks: (i) From the bifurcation equation (1.9), we notice that $\lambda_1(0)$ is negative, at least for $c_0 = 0$. Since λ is proportional to the square of the external field, the above theorem proves the existence of superconducting solutions of Eqs. (1.1)–(1.3) for fields slightly lower than the "critical field" $h = \lambda_0^{\frac{1}{2}}$.

(ii) If $c_0 = 0$, one gets the "symmetric states" (Marcus[7]) associated with the second critical field $\lambda_0^{\frac{1}{2}} = h_{c2}$. The other constants $c_{0,i}$ clearly give "unsymmetric" states which may be associated with higher critical fields. The calculations of Ref. 6 show that for some optimal $c_{0,i}$ one gets "surface states" where ψ_0 is concentrated near the boundary and a "third" critical field $h_{c3} > h_{c2}$.

(iii) The equations (1.1)–(1.3) have nontrivial solutions for all external fields $h < h_c$. This fact can be proved by the variational methods of Sec. 3. Since we prove the analogous theorem for the more complicated situation of Sec. 3, we give no proof here.

2. BIFURCATION OF ABRIKOSOV'S MIXED STATE

We consider now the partial differential equation form of the GL equations in the special "quasi-periodic" case discussed by Abrikosov.[5] The physical situation corresponds to a "vortex" structure in the x–y plane in which both the magnitude of the complex order parameter $\phi(x, y)$ as well as the magnetic field are periodic functions which do not depend on z. The magnetic field is taken to have one component, in the z direction, say. The equations, in appropriate units[5] are

$$[(i/k)\nabla + \mathcal{A}]^2\phi = \phi(1 - |\phi|^2) \qquad (2.1)$$

$$-\nabla \times \nabla \times \mathcal{A} = \mathrm{Re}\,\{\phi^*[(i/k)\nabla + \mathcal{A}]\phi\}. \qquad (2.2)$$

We seek solutions to (2.1), (2.2) such that both the magnetic field $\mathbf{B} = (0, 0, B)$ and $|\phi|^2$ are periodic with periods defined by lattice vectors $\mathbf{t}_j$, $j = 1, 2$. If such a solution exists, then (2.2) leads to the flux quantization condition[9]

$$k\bar{B}F = 2\pi n, \qquad (2.3)$$

where F is the area of a primitive cell, $\bar{B}$ is the average flux density, and n is an integer. Let $\mathbf{t}$ denote either of the lattice vectors $\mathbf{t}_j$, then to insure (2.3), one has to impose the "boundary" conditions

$$\phi(\mathbf{r} + \mathbf{t}) = \exp\{ikg(\mathbf{r}, \mathbf{t})\}\phi(\mathbf{r}), \qquad (2.4)$$

$$\mathcal{A}(\mathbf{r} + \mathbf{t}) = \mathcal{A}(\mathbf{r}) + \nabla g(\mathbf{r}, \mathbf{t}), \qquad (2.5)$$

where g is a smooth, single-valued, real function.[10] An elementary calculation shows that (2.4), (2.5) imply (2.3) for any "gauge" g. We now fix the gauge as follows. Since the problem (2.1)–(2.5) is invariant

[8] H. Ehrmann, Enseignement Math. (2) **9**, 129 (1963).

[9] G. Lasher, Phys. Rev. **140**, A523 (1965).

[10] G. Eilenberger, Z. Physik **180**, 32 (1964).

under the transformation $\mathcal{A}' = \mathcal{A} + \nabla\eta$, $\phi' = \phi e^{ik\eta}$, we may first choose η so that $\operatorname{div}\mathcal{A} = 0$. Now define $\mathbf{A}$ by $\mathcal{A} = \frac{1}{2}(\mathbf{r} \times \bar{\mathbf{B}}) + \mathbf{A}$. Then $\iint \operatorname{curl} \mathbf{A} = 0$. An elementary calculation shows that by adding a vector $\nabla\eta$, where η is an appropriate harmonic function, to the potential $\mathcal{A}$ we make $\mathbf{A}$ a periodic vector. Then $\mathcal{A}$ is defined up to an arbitrary constant; hence we may further assume that $\iint \mathcal{A}\, dx\, dy = 0$. A possible shift in the origin of coordinates then reduces $\mathcal{A}$ to the form

$$\mathcal{A} = \tfrac{1}{2}(\mathbf{r} \times \mathbf{B}) + \mathbf{A},$$

where $\operatorname{div} \mathbf{A} = 0$; $\mathbf{A}$ is periodic and with mean zero. The boundary conditions (2.4)–(2.5) are now fixed completely, with the gauge g given by $g = \frac{1}{2}(\mathbf{t} \times \bar{\mathbf{B}}) \cdot \mathbf{r}$. It is convenient at this point to change variables in the GL equations as in Lasher.[9] Let

$$(x', y') = (k\bar{B}/2\pi)^{\frac{1}{2}}(x, y); \quad \phi' = (\bar{B}/2\pi k)^{-\frac{1}{2}}\phi;$$
$$\mathcal{A}' = (\bar{B}/2\pi k)^{-\frac{1}{2}}\mathcal{A}.$$

Then a unit cell in the (x', y') system containing one quantum of flux [namely, when $n = 1$ in (2.3)] occupies a unit area. The period of the order parameter no longer depends on $\bar{B}$. Dropping the primes in the new coordinate system, we define $\mathbf{A}$ by $\mathcal{A} = \mathbf{A}_0 + \mathbf{A}$, where $\mathbf{A}_0 = \pi\mathbf{r} \times \mathbf{h}$ with $\mathbf{h}$ a unit vector in the z direction. Notice then that $\operatorname{curl} \mathbf{A}_0 = 2\pi$ and that $\mathbf{A}$ is periodic with zero mean. The gauge in the new coordinates is given by $g(\mathbf{r}, \mathbf{t}) = \pi(\mathbf{t} \times \mathbf{h}) \cdot \mathbf{r}$, and the GL equations reduce to

$$-[(i\nabla + \mathbf{A}_0)^2 - \lambda]\phi = \phi\,|\phi|^2 + 2\mathbf{A} \cdot (i\nabla + \mathbf{A}_0)\phi + \mathbf{A}^2\phi \quad (2.6)$$

$$-\nabla \times \nabla \times \mathbf{A} = \tfrac{1}{2}k^{-2}\,\mathrm{Re}\,[\phi^*(i\nabla + \mathcal{A})\phi], \quad (2.7)$$

where $\lambda = 2\pi k(\bar{B})^{-1}$ plays the role of an eigenvalue parameter. It is clear that $\phi \equiv 0$, $\mathbf{A} = 0$ is a solution to (2.6)–(2.7) for any value of λ. This is the "trivial" solution corresponding to the normal state. To prove the existence of nontrivial small solutions, we define $\phi = \epsilon^{\frac{1}{2}}\psi$, $\mathbf{a} = \epsilon\mathbf{A} \equiv \epsilon(\mathcal{A} - \mathbf{A}_0)$ and substitute in (2.6)–(2.7) to get

$$(L - \lambda)\psi \equiv \{(i\nabla + \mathbf{A}_0)^2 - \lambda\}\psi = -\epsilon f_1(\psi, \psi^*, \mathbf{a}), \quad (2.8)$$

$$M\mathbf{a} \equiv -\nabla \times \nabla \times \mathbf{a} = f_2(\psi, \psi^*, \mathbf{a}), \quad (2.9)$$

where

$$f_1 = \psi\,|\psi|^2 + 2\mathbf{a} \cdot (i\nabla + \mathbf{A}_0)\psi + \epsilon\mathbf{a}^2\psi, \quad (2.10)$$

$$f_2 = k^{-2}\,\mathrm{Re}\,[\psi^*(i\nabla + \mathbf{A}_0)\psi] + \epsilon k^{-2}\,\mathrm{Re}\,[\psi^*\,\mathbf{a}\psi]. \quad (2.11)$$

The boundary conditions on ψ, $\mathbf{a}$, derived from Eqs. (2.4)–(2.5), are

(i) $\psi(\mathbf{r} + \mathbf{t}) = \exp[i\pi(\mathbf{t} \times \mathbf{h}) \cdot \mathbf{r}]\psi(\mathbf{r})$ (2.12)

(ii) $\mathbf{a}$ is periodic with zero mean and divergence. (2.13)

For $\epsilon = 0$, (2.8) is solvable for a discrete set of positive values for λ (see, for example, Brown,[11] Sec. 4). The smallest possible λ, corresponding to greatest flux, is $\lambda_0 = 2\pi$, which is associated with one or more eigenfunctions ψ_0 depending on lattice symmetry. For the important cases of square and triangular lattices there exists only one eigenfunction[10,12] and we assume this in what follows. With ψ_0 given, Eq. (2.9) defines a unique vector function $\mathbf{a}_0$. The triple $(\lambda_0, \psi_0, \mathbf{a}_0)$ is a solution to (2.8)–(2.13) for the case $\epsilon = 0$ and hence still corresponds to a trivial solution. We now seek solutions to (2.8)–(2.13) for small enough ϵ in the form

$$(\lambda_0 + \epsilon\lambda_1(\epsilon), \psi_0 + \psi_1(\epsilon), \mathbf{a}_0 + \mathbf{a}_1(\epsilon))$$

where ψ_1 is orthogonal to ψ_0. Projecting (2.8) on the subspace E generated by ψ_0 and its orthogonal complement E', we obtain

$$\langle \psi_0, \lambda_1(\epsilon)\psi - f_1(\psi, \psi^*, \mathbf{a})\rangle = 0, \quad (2.14)$$

$$\psi_1 - \epsilon R[\lambda_1\psi_1 + f_1] = 0, \quad (2.15)$$

where $\langle\ ,\ \rangle$ denotes the scalar product and R is the pseudoinverse of $(L - \lambda_0)$. Equation (2.14) is the bifurcation equation and determines $\lambda_1(0)$ which may be seen to be positive. Inverting (2.9), we get

$$\mathbf{a} - Kf_2(\psi, \psi^*, \mathbf{a}) = 0, \quad (2.16)$$

where K is the inverse of M. In order to discuss the properties of these inverses (which may be written explicitly in terms of eigenfunction expansions connected with L, M), we introduce the Sobolev spaces $W_n(\Omega)$ where Ω is the primitive cell. The space W_n is the space of scalar or vector distributions whose derivatives up to the nth order belong to $L_2(\Omega)$, and thus is a Hilbert space with a natural scalar product. Let $\|u\|_n$ denote the norm of u in W_n; then we have

Lemma 1: The operators R and K are compact operators in W_n which satisfy the estimates

$$\|Rf\|_{n+2} \leq C_1\,\|f\|_n \quad (2.17)$$

$$\|K\mathbf{a}\|_{n+2} \leq C_2\,\|\mathbf{a}\|_n. \quad (2.18)$$

where C_1, C_2 are constants which depend only on Ω and $\mathbf{a}_0$.

[11] E. Brown, Phys. Rev. **133**, A1038 (1964).

[12] W. Kleiner, L. Roth, and S. Autler, Phys. Rev. **133**, A1226 (1964).

Proof: Consider the equation $(L - \lambda_0)u = f$, where L is defined by (2.8), (2.12), and u and f are orthogonal to ψ_0. To prove (2.17) we have to estimate u in terms of f. Let $v = \exp iQ(\mathbf{r})u(\mathbf{r})$ where Q may be chosen to make $v(\mathbf{r})$ periodic. Then v satisfies $(L' - \lambda_0)v = f' \equiv f \exp(iQ)$, where $L' = (i\nabla + \nabla Q + A_0)^2$. Since v is the solution to a uniquely invertible, uniformly elliptic equation with regular boundary conditions, the inequality $\|v\|_{n+2} \leq C\|f'\|_n$ follows from standard estimates (see, for example, Agmon, Douglis, and Nirenberg[13]). The same inequality, with a different constant depending on Q, holds therefore for u, which proves (2.17). A similar, but simpler, argument proves (2.18).

Now we consider the system of equations (2.14)–(2.16) for the unknown $(\lambda, \psi_1, \mathbf{a})$ where both ψ_1 and $\mathbf{a}$ belong to $W_3(\Omega)$. Then we have

Theorem 2: There exists an $\epsilon_0 > 0$ such that for $\epsilon \leq \epsilon_0$ the system (2.8)–(2.13) has a unique nontrivial solution.

Proof: From the discussion preceding Lemma 1, it is sufficient to discuss the system (2.14)–(2.16). Let $\mathbf{u}$ denote the vector $\mathbf{u} = (\lambda, \psi_1, \mathbf{a})$ and consider the left side of (2.14)–(2.16) as a nonlinear map $N(\mathbf{u}, \epsilon)$ of the space $B = I \otimes W_3(\Omega) \otimes W_3(\Omega)$, where I denotes the real numbers, into itself. Since the norm in W_3 dominates the maximum norm by Sobolev's imbedding lemma[14] and the nonlinear functions f_1, f_2 are smooth—in fact, analytic—in the maximum norm, it follows that the map N is bounded and continuously differentiable in B and ϵ. Since $\mathbf{u}_0 = (\lambda_1(0), 0, \mathbf{a}_0)$ is the unique solution of $N[\mathbf{u}, 0] = 0$, the above theorem becomes a consequence of the implicit function theorem if the Frechet derivative D of N at $\mathbf{u}_0$ is nonsingular. A simple calculation shows that D has the same form as J in Sec. 1 [Eq. (1.11)] and therefore is invertible, which proves the theorem.

Remarks: (i) The standard methods for proving the implicit function theorem in Banach spaces (see Ehrmann[8] or Schwartz[15]) may be used to give different perturbation procedures for calculating the solution (λ, ϕ, A) in terms of the parameter ϵ. The Newton method (Schwartz[15]), for example, gives a quadratically convergent procedure.

(ii) The method used in Lasher[9] is based on a formal expansion in terms of $\alpha = (2\pi)^{-1}(\lambda - \lambda_0)$. This procedure may be justified as follows: for small enough ϵ, λ is an increasing function of ϵ since $\lambda_1(0) > 0$. Hence, for some $\delta > 0$ and for $\epsilon \leq \delta$, we may consider ϵ as a unique function of α. Then ψ, $\mathbf{a}$ become functions of α also, and any convergent procedure in terms of ϵ produces a similar one in terms of α. However, the α region of convergence may be smaller than the ϵ region.

(iii) Equation (2.8), for $\epsilon = 0$, has solutions for a discrete set of eigenvalues $\lambda_n \to \infty$. Using the same method of Theorem 2, it is possible to prove the existence of nontrivial solutions to the GL equations for values of λ near (and slightly higher than) these eigenvalues.

3. EXISTENCE OF THE MIXED STATE FOR $\lambda > \lambda_0$

In this section we use the direct methods of calculus of variations to prove the existence of nontrivial solutions to the GL equations (2.6)–(2.7) for all $\lambda > \lambda_0 = 2\pi$. This means that the mixed vortex state exists for all values of the flux below the critical flux $\bar{B} = k$. As is well-known,[5] Eqs. (2.6)–(2.7) are the Euler equations for the energy

$$E = \int \{(i\nabla + \mathcal{A})\psi \cdot (-i\nabla + \mathcal{A})\psi^* + f(\psi, \psi^*, \lambda) + 2k^2 |\nabla \times \mathcal{A}|^2\}\, dx\, dy, \quad (3.1)$$

where $f = \frac{1}{2}|\psi|^4 - \lambda|\psi|^2$. The integration in (3.1) is over a fixed unit cell independent of λ because of the normalizations given in Sec. 2. To define the variational problem precisely we have to specify the admissible functions. As in Sec. 2 we first define $\mathbf{A} = \mathcal{A} - \pi\mathbf{r} \times \mathbf{h}$. Consider then smooth functions $\mathbf{u} \equiv (\psi, \mathbf{A})$ satisfying conditions (2.12)–(2.13). By completing these functions with respect to the norm

$$\|u\|^2 = \int \{|\mathbf{u}|^2 + |\nabla\mathbf{u}|^2\}\, dx\, dy,$$

where $\nabla\mathbf{u}$ denotes all the first-order derivatives of ψ, $\mathbf{A}$, we get a closed subspace H of the Hilbert space $W_1(\Omega)$. The variational problem E is then defined in H, and our object is to prove that if $\lambda > 2\pi$, there exists a nontrivial vector $v(\lambda) \in H$ which lies in the interior of some bounded region of H and minimizes E. Then $\mathbf{v}$ will be a variational solution to the GL equations corresponding to the mixed state. First we prove

Lemma 2: The free energy $E(\mathbf{u}) \to \infty$ as $\|\mathbf{u}\| \to \infty$ for all admissible $\mathbf{u}$.

[13] S. Agmon, A. Douglis, and L. Nirenberg, Commun. Pure Appl. Math. 17, 35 (1964).

[14] N. Dunford and J. T. Schwartz, *Linear Operators, Part II* (Interscience Publishers, Inc., New York, 1963), Chap. XIV, p. 1686.

[15] J. T. Schwartz, "Nonlinear Functional Analysis," Courant Institute of Mathematical Sciences Lecture Notes, New York (1965).

Proof: Using $\nabla \cdot \mathbf{A} = 0$, we rewrite (3.1) in the form

$$E - 8k^2\pi^2 F = \int \{|\nabla\psi|^2 - 2\,\mathrm{Im}\,(\psi^*\mathbf{A}\cdot\nabla\psi) + |\psi\mathbf{A}|^2 + f(\psi,\lambda) + 2k^2\,|\nabla\times\mathbf{A}|^2\}\,dx\,dy, \quad (3.2)$$

where F is the area of the unit cell and where we have used the fact that the average of $\nabla \times \mathbf{A}$ is zero. If $\|\mathbf{u}\| \to \infty$, then at least one of the L_2 norms of ψ, $\mathbf{A}$, $\nabla\psi$ or $\nabla\mathbf{A}$, which we denote by $\|\ \|_2$, must approach infinity. We check the behavior of E in all possible cases. First, if $\|\psi\|_2 \to \infty$, then, for any fixed λ, $\|f(\psi,\lambda)\|_2 \to \infty$ by the Schwarz inequality and hence, using (3.1), $E \to \infty$. We may henceforth assume $\|\psi\|_2$ bounded. Using Green's formula, the periodicity of $\mathbf{A}$ and $\mathrm{div}\,\mathbf{A} = 0$, we find $\|\nabla\times\mathbf{A}\|_2 = \|\nabla\mathbf{A}\|_2$. Hence, if $\|\nabla\mathbf{A}\|_2 \to \infty$, the last term in (3.1) tends to infinity by the Schwarz inequality. Since the first two terms are bounded from below, because $\|\psi\|_2$ is bounded, E must approach infinity. Therefore, we may assume now that $\|\nabla\mathbf{A}\|_2$ is bounded also. Because of the periodicity conditions (2.13), $\|\mathbf{A}\|_2$ is also bounded. By Sobolev's imbedding lemma, $\mathbf{A} \in L_p$ for all $p < \infty$. We now discuss the case when $\|\nabla\psi\|_2 \to \infty$. It is clearly sufficient to restrict attention to the case when $\psi \in L_4$ for otherwise (3.1) shows that $E = \infty$. Then, using the Schwarz inequality twice on the second term in (3.2), we find

$$E \geq \|\nabla\psi\|_2^2 - c_1\,\|\nabla\psi\|_2 - c_2, \quad (3.3)$$

where c_1, c_2 are some finite constants. Hence, $E \to \infty$ in this case also, and the lemma is proved.

Let B_r denote the ball of radius r around the origin in H. Then, for a given r, we have

Lemma 3: The functional E is lower semicontinuous on B_r with respect to weak convergence in H.

Proof: Consider the form (3.2) for E. The third and fourth terms are continuous in the weak topology. Consider, for example, the fourth term. Let ψ_n converge weakly to ψ. Then the H norms of ψ, ψ_n are bounded. Hence, a subsequence of the ψ_n, call it ψ_n again, converges to ψ in L_2 by Rellich's lemma.[16] Hence, if $\phi_n = |\psi_n|$, we have

$$\int |\phi_n^4 - \phi^4|\,dx\,dy = \int |(\phi_n - \phi)P_3(\phi_n,\phi)|\,dx\,dy, \quad (3.4)$$

where P_3 is a cubic in ϕ_n, ϕ. By Schwarz, the right side of (3.4) tends to zero if the L_2 norm of P_3 is bounded. But this is an immediate consequence of the boundedness of the H norm of ψ and Sobolev's imbedding lemma. A similar argument applies to the third term. The remaining terms in the integrand of (3.2) are, for fixed ψ, $\mathbf{A}$, convex in the first derivatives of the vector $\mathbf{u} = (\psi, \mathbf{A})$. By a theorem of Morrey[17] the integrals of these terms are then lower semicontinuous, which completes the proof. Now we state the main result.

Theorem 3: There exists at least one nontrivial variational solution of the GL equations (2.6)–(2.7) for all $\lambda > \lambda_0$

Proof: In view of Lemma 3, $E(\mathbf{u})$ achieves its minimum on every ball B_r because such a ball is weakly compact. By Lemma 2 we can choose r large enough so that this minimum, call it $\mathbf{v}$, does not occur on the boundary of B_r. Then $\mathbf{v}$ is a variational solution of the GL equations. It is then sufficient to check that $\mathbf{v} \not\equiv 0$. We calculate the second variation of E at $\mathbf{u} = 0$; i.e., at $\psi = 0$, $\mathcal{A} = A_0$, in the "direction" $\delta u = (\delta\psi, \delta\mathbf{A})$. If L denotes the operator

$$(i\nabla + \mathbf{A}_0)\cdot(-i\nabla + \mathbf{A}_0),$$

this variation is

$$\delta^2 E = \langle\delta\psi, (L - \lambda)\delta\psi\rangle + 2k^2\langle\delta\mathbf{A}, -\Delta\delta\mathbf{A}\rangle. \quad (3.5)$$

Choosing $\delta\mathbf{A} = 0$ and $\delta\psi$ to be the eigenfunction of L corresponding to the lowest eigenvalue $\lambda = \lambda_0$, we find that the second variation in this direction is negative for all $\lambda > \lambda_0$. Hence, $\mathbf{u} = 0$ is not even a local minimum for E and the theorem is proved.

ACKNOWLEDGMENTS

The author would like to thank his colleagues, I. Tadjbakhsh and G. Lasher, for helpful discussions.

[16] See, for example, Ref. 14, pp. 1692–1693.

[17] C. B. Morrey, Pacific J. Math. 2, 25 (1952).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Properties of the Wavefunction for Singular Potentials

RICHARD M. SPECTOR
Physics Department, Wayne State University, Detroit, Michigan 48202

(Received 4 May 1967)

For potentials more singular than the inverse square at the origin and with complex strength, the nature of the solutions to the Schrödinger equation is investigated. The difficulties occurring with attractive real potentials in the three-dimensional equation are discussed, and it is shown that no solutions exist in a domain including the origin. For complex or repulsive singular potentials with radial form varying as the inverse fourth or sixth power, a relatively simple series solution exists. This is also true for the singular Yukawa with inverse fourth power. These series are shown to be asymptotic by means of general theorems from the theory of ordinary differential equations. Other inverse powers have much more complicated solutions. In the Dirac case, all inverse powers and singular Yukawa forms have asymptotic series solutions, and the series is given explicitly for the singular powers. A form of the S-wave scattering length for the singular inverse fourth-power Yukawa $V = g^2e^{-i\Delta}e^{-\mu r}/r^4$ is derived which is valid for small values of μg.

I. INTRODUCTION

SINCE the pioneering paper of Case[1] many years ago, an extensive literature has developed on the subject of singular potentials. By singular potentials it is meant those potentials which have the property

$$\lim_{r\to 0} r^2V(r) \to \infty, \tag{1}$$

although sometimes the inverse-square potential is also regarded as singular. This particular potential[2] has a number of properties which set it apart from those characterized by Eq. (1), but nevertheless it does behave abnormally in enough ways to exclude it from being classified as a regular potential. All singular potentials are assumed to vanish at infinity fast enough to cause no trouble with ordinary scattering theory. This usually means they vanish faster than r^{-3} at infinity.[3]

The investigation of singular potentials covers a number of areas of both physical and mathematical interest. The literature in the last few years is too voluminous to reference here completely, but some of the important papers in each area can be mentioned. Vogt and Wannier[4] made early use of the singular potential behaving as r^{-4} in the study of gaseous ions moving through a gas whose molecules are not too large.

Tiktopoulous[5] and Kouris[6] have used the r^{-2} and r^{-4} potentials, respectively, with energy-dependent complex strengths to investigate the elastic differential cross sections for high energy p–p and p–π scattering. Their models, patterned somewhat after the optical model of Serber,[7] give reasonable results. Quite recently, Spector and Chand[8] have used the general form r^{-n} $(n > 3)$ with complex energy-independent strength to describe the KN and $\bar{K}N$ low-energy scattering lengths. Again the results are satisfactory.

On the theoretical side, early work was stimulated by a paper of Predazzi and Regge[9] and by the weak field theory papers of Feinberg and Pais.[10] The former authors conjectured that, inasmuch as the real world probably involved rather singular interactions, the best way to investigate Regge poles in field theory (assuming that it describes the real world) was to investigate the appearance of Regge poles in the scattering amplitudes of singular potentials. Work by Jakšić and Limić[11] and by Aly and Müller[12] was carried out along this line.

The Feinberg and Pais papers introduced the idea of peratization, which is a rather complicated procedure for summing divergent terms in weak field theory. Tiktopoulous and Treiman[13] and Khuri and Pais[14] shortly thereafter applied similar procedures to nonrelativistic singular potential theory in order to calculate the scattering lengths for singular potentials. These scattering lengths cannot be calculated in the usual Born series, since every term diverges, so that a

[1] K. M. Case, Phys. Rev. **80**, 797 (1950).
[2] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., Inc., New York, 1953), Chap. 5.
[3] We want the scattering lengths, forward scattering amplitude, and total cross section to exist. See, for example, L. Landau and E. Lifshitz, *Quantum Mechanics* (Pergamon Press, Inc., London, 1958), Chap. 14. See also Ref. 21.
[4] E. Vogt and G. Wannier, Phys. Rev. **95**, 1190 (1954)
[5] G. Tiktopoulous, Phys. Rev. **138**, B1550 (1965).
[6] C. B. Kouris, Nuovo Cimento **44**, 598 (1966).
[7] R. Serber, Phys. Rev. Letters **10**, 357 (1963).
[8] R. Spector and Ramesh Chand (to be published).
[9] E. Predazzi and T. Regge, Nuovo Cimento **24**, 518 (1962).
[10] G. Feinberg and A. Pais, Phys. Rev. **131**, 724 (1963); **133**, B477 (1964).
[11] B. Jakšić and N. Limić, J. Math. Phys. **7**, 88 (1966).
[12] H. Aly and H. Müller, J. Math. Phys. **7**, 1 (1966).
[13] G. Tiktopoulous and S. Treiman, Phys. Rev. **134**, B844 (1964).
[14] N. Khuri and A. Pais, Rev. Mod. Phys. **36**, 590 (1964).

special technique, called peratization, is utilized. A great number of workers have investigated peratization for particular potentials. Among these are Aly *et al.*,[15] Spector,[16] Wu,[17] and Calogero.[18]

Other workers, and in particular Cornille,[19] have been concerned with more general properties of singular potentials: such as when the peratization procedure can be expected to work in the general case; how the asymptotic high-energy scattering amplitudes behave[20]; what the effective range low-energy approximations look like[21]; what the formal expressions for the S matrix are[22], and the determination of various approximations to the scattering amplitudes.[23]

Despite the extensive literature on singular potentials, extremely few papers have been devoted to the actual solution of the nonrelativistic Schrödinger equation or the Dirac equation for singular potentials. Wannier,[24] Spector,[25] and Aly and Müller[12] are among the very few papers devoted to this subject, and they all deal with solutions to the radial Schrödinger equation for r^{-4} at arbitrary energy and angular momentum. The reason for this is that the r^{-4} potential is the only one that can be solved in terms of known functions for nonzero energy. Surprisingly, Plesset[26] did make a cursory study of singular potentials in the Dirac equation 35 years ago.

In this paper we intend to present an investigation and discussion of the solutions to the three-dimensional Schrödinger equation, the radial Schrödinger equation, and the Dirac equation. In Sec. II the difficulties caused by attractive real singular potentials are investigated in the three-dimensional equation. In Sec. III the general properties of the radial equation solutions are determined. In Sec. IV a more general class of singular potentials, namely the singular Yukawa forms, are discussed; and in Sec. V formal solutions are given to the Dirac equation. In the Appendix we derive an approximate form for the S-wave scattering length resulting from the singular Yukawa potential which varies as $e^{-\mu r}/r^4$.

II. THREE-DIMENSIONAL NONRELATIVISTIC SCHRÖDINGER EQUATION

In nonrelativistic quantum mechanics, a spinless particle of energy k^2 moving in a potential $V(\mathbf{r})$ has a wavefunction satisfying

$$\nabla^2\psi(\mathbf{r}) + [k^2 - V(\mathbf{r})]\psi(\mathbf{r}) = 0, \tag{2}$$

where ∇^2 is the Laplacian operator in three dimensions and our units are such that $\hbar = 2m = 1$. The conventional separation into partial waves (for a spherically symmetric potential) produces a one-dimensional radial equation for the lth partial wave of

$$\frac{d^2\phi}{dr^2} + \left[k^2 - \frac{l(l+1)}{r^2} - V(r)\right]\phi = 0, \tag{3}$$

where $\phi = r\psi$. Since singular potentials have stronger behavior at the origin than the angular momentum, it is easily seen from Eq. (3) that for many purposes it is sufficient to discuss only the zero-energy S-wave wavefunction. We now fix the form of out potential, in this part of the paper, to be

$$V(r) = \frac{g^2e^{-i\Delta}}{r^n}\, g > 0, \quad 0 \leq \Delta \leq \pi, \quad n > 2. \tag{4}$$

Notice that when $\Delta = 0$, the potential is repulsive, and when $\Delta = \pi$, it is attractive. The restricted range on Δ is necessary to keep Im $V \leq 0$ in accordance with the usual need of scattering theory to ensure that the scattering causes a loss of flux in the elastic channel rather than an increase. Although we admit $n > 2$ mathematically, we usually require $n > 3$ for the physical reasons mentioned earlier.

The solutions to Eq. (3), with potential Eq. (4) and for $k^2 = l = 0$, are easily obtained from the known solutions[27] for $\Delta = 0$. We have

$$\phi(r) = r^{\frac{1}{2}}K_p(2pge^{-i\Delta/2}r^{-1/(2p)}) \quad p = \frac{1}{n-2}, \tag{5}$$

where $K_p(x)$ is the modified Bessel function of the third kind with complex argument.[28]

The behavior near $r = 0$ for any k^2 and l is given by Eq. (5), as we mentioned above. Thus we obtain

$$\lim_{r\to 0} \phi(r) \sim r^{n/4} \exp\left(-\frac{2g}{n-2}\, e^{-i\Delta/2}r^{1-(n/2)}\right) \quad \text{any } k^2, l. \tag{6}$$

There is, of course, a second solution to Eq. (3) which exhibits a plus sign in the exponential in Eq. (6); but as $r \to 0$, only the form Eq. (6) makes ϕ

[15] H. Aly, Riazuddin, and A. Zimerman, Phys. Rev. **136**, B1174 (1964).
[16] R. Spector, J. Math. Phys. **7**, 2103 (1966).
[17] T. T. Wu, Phys. Rev. **136**, B1176 (1964).
[18] F. Calogero, Phys. Rev. **139**, B602 (1965).
[19] H. Cornille, Nuovo Cimento **38**, 1243 (1965); **39**, 557 (1965); **43**, 786 (1966).
[20] E. Del Giudice and E. Galzenati, Nuovo Cimento **38**, 443 (1965); N. Dombey, *ibid.* **37**, 1741 (1965).
[21] T. O'Malley, L. Spruch, and L. Rosenberg, J. Math. Phys. **2**, 491 (1961).
[22] R. Spector, Nuovo Cimento **45**, 924 (1966).
[23] F. Calogero, Nuovo Cimento **27**, 261 (1963). See also Ref. 27.
[24] G. Wannier, Quart. Appl. Math. **11**, 33 (1953).
[25] R. Spector, J. Math. Phys. **5**, 1185 (1965).
[26] M. Plesset, Phys. Rev. **41**, 278 (1932).
[27] F. Calogero, Phys. Rev. **135**, B693 (1964).
[28] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1958), 2nd ed.

vanish if $\Delta < \pi$. When $\Delta = \pi$, ϕ does not vanish exponentially fast, but rather much more slowly, as some positive power of r. In this case the second solution also vanishes, causing a long recognized quandry, first observed by Case,[1] concerning the choice of the correct linear combination of wavefunctions at the origin. As we shall see below, this quandry in fact has no meaning.

In order to investigate the case of $\Delta = \pi$ (attractive potential) more closely, we restrict ourselves for the present to $n = 4$ for $k^2 = l = 0$, where the two *exact* solutions for ϕ are

$$\phi = re^{\pm igr^{-1}}, \tag{7}$$

and Eq. (2) becomes

$$\nabla^2\psi + \frac{g^2}{r^4}\psi = 0. \tag{8}$$

Before looking at Eqs. (7) and (8), we recall that Laplace's equation

$$\nabla^2\psi = 0 \tag{9}$$

is *not* satisfied by the spherically symmetric $\psi = 1/r$ because $\nabla^2(r^{-1}) = -4\pi\delta(\mathbf{r})$, where $\delta(\mathbf{r})$ is the three-dimensional Dirac delta function. Yet the associated radial function $\phi(r) = 1$ does satisfy the radial equation $\phi'' = 0$.

In fact, $\psi = 1/r$ even *appears* to satisfy the three-dimensional equation when written out in spherically symmetric form

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial r^2} + \frac{2}{r}\frac{\partial\psi}{\partial r} = 0. \tag{10}$$

Equation (10) is satisfied by $\psi = 1/r$ at all points except $r = 0$, and it is tempting to say that $1/r$ is a solution there, too, in the sense of taking some suitably defined smooth limit as $r \to 0$.

Precisely what, then, is meant by saying that $\psi(\mathbf{r})$ is a solution to Eq. (8) or Eq. (9)? A function $\psi(\mathbf{r})$ is said to be a solution to Eq. (8) [or Eq. (9) in an obvious simplification of what follows below] if the following expression is valid:

$$\iiint_{\text{ALL SPACE}} \left[\psi(\mathbf{r})\nabla^2 u(\mathbf{r}) + \frac{g^2}{r^4}u(\mathbf{r})\psi(\mathbf{r})\right] dV = 0, \tag{11}$$

where $u(\mathbf{r})$ is any continuous function with compact support and continuous first and second derivatives.[29] Any ψ which satisfies the condition of Eq. (11) is said to be a solution in the weak sense,[30] although it may (but not necessarily) be a solution in the strong sense as well, i.e., satisfying Eq. (8) at *all* points in the domain $0 \leq r \leq \infty$ with the conventional differential form for ∇^2, the radial part of which is given in Eq. (10).

The expression Eq. (11) is obtained from

$$\iiint_{\text{ALL SPACE}} u(\mathbf{r})\left[\nabla^2\psi(\mathbf{r}) + \frac{g^2}{r^4}\psi(\mathbf{r})\right] dV \tag{12}$$

by integration by parts, utilizing the properties of $u(\mathbf{r})$, or just by recalling directly that the ∇^2 operator is self-adjoint. This conversion from Eq. (12) to Eq. (11) is necessary in order to allow ∇^2 to operate on $u(\mathbf{r})$ where ∇^2 may be expressed in its conventional differential operator form because $u(\mathbf{r})$ is well-behaved everywhere. On the other hand, when ∇^2 operates on $\psi(\mathbf{r})$, as it does in Eq. (12), its properties are uncertain if $\psi(\mathbf{r})$ or if its derivatives fail to exist anywhere. For instance, this occurs at the origin for $\psi = 1/r$.

If now we look at the spherically symmetric solutions given in Eq. (7) and use Eq. (11) [after integrating out the angles in Eq. (12) first], we require that the integral

$$\int_0^\infty e^{\pm igr^{-1}}\left(r^2u'' + 2ru' + \frac{g^2}{r^2}u\right) dr \tag{13}$$

be zero if Eq. (7) is to be a solution to Eq. (8). The u in Eq. (13) is now a radial function only, and differentiation is with respect to r. Rewriting Eq. (13) we have

$$\begin{aligned}\int_0^\infty &\left[(r^2e^{\pm igr^{-1}}u')' \pm ige^{\pm igr^{-1}}u' + \frac{g^2}{r^2}u\right] dr \\ &= r^2e^{\pm igr^{-1}}u'\Big|_0^\infty \pm ig\int_0^\infty (e^{\pm igr^{-1}}u)'\,dr \\ &= \lim_{r\to 0} \pm igu(r)e^{\pm igr^{-1}} \neq 0 \quad (\text{if } u(0) \neq 0),\end{aligned} \tag{14}$$

where both integrated terms vanish at infinity because u has compact support. Thus even though

$$\phi = re^{\pm igr^{-1}}$$

satisfies the radial equation, the three-dimensional wavefunction $\psi = \phi/r$ does *not* satisfy the three-dimensional equation. Such a state of affairs is noted briefly by Newton,[31] although he uses the less rigorous plausibility argument that ∇^2, operating on certain general functions of r^{-1}, produces delta-function terms. It is necessary, however, to show (as we have done above) that not only are these terms present but that they actually affect the solution. This must also be done in such a way as to bypass the actual use of the

[29] Roughly, a function with compact support must vanish outside of some bounded domain. The classic example of an infinitely differentiable function with compact support is $f(r) = \exp[-1/(1-r)]$, $0 \leq r \leq 1$; $f(r) = 0$, $r \geq 1$.

[30] See L. Hörmander, *Linear Partial Differential Operators* (Springer-Verlag, Berlin, 1963), especially Chap. IV.

[31] R. G. Newton, *Scattering Theory of Waves and Particle* (McGraw-Hill Book Co., Inc., New York, 1966), p. 390 ff.

delta function. In fact, it can be shown with a little effort that, by using a particular representation of the delta function, the integral

$$\int_0^\infty e^{\pm ig r^{-1}} u(r)\delta(r)\, dr$$

must vanish, which would lead to a result inconsistent with Eq. (14). This may also be seen less rigorously by using the property of $\delta(r)$ that

$$\int_0^\infty e^{\pm ig r^{-1}} u(r)\delta(r)\, dr = \lim_{r\to 0} u(r) e^{\pm ig r^{-1}} \neq 0.$$

(See the remarks below about the correct form of ∇^2.)

Since any solution of Eq. (8) must be equal to ϕ/r with ϕ given by Eq. (7) when $r > 0$, this means that there are no (continuous) solutions to Eq. (8) at all.[32] The quandry first mentioned by Case, and referred to earlier, has no meaning.

A number of comments are in order about this result. First, the limit in expression (14) failed to exist because we took $\Delta = \pi$. By repeating the process leading to Eq. (14) with $\Delta < \pi$, it is easily seen that the limit is zero. Thus, so long as $\Delta < \pi$, there is a three-dimensional solution in the weak sense—which is also a strong solution, of course.

Second, since the dangerous behavior of ∇^2 occurs only at the origin, our results obviously hold even if k^2 and l are not zero. The proof, analogous to that above, would no longer be so simple, but fortunately it is not necessary.

Third, the same results would also hold for $n > 4$, as can be seen by using the expression (6) and making use of the fact that, when operating on a radial function containing r^{-1}, we may write the Laplacian operator as[33]

$$\nabla^2 \to \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \delta(r)\frac{\partial}{\partial r}.$$

Thus, we are left with a sound mathematical solution over all space if $\Delta < \pi$, but also with solutions for $\Delta = \pi$ under certain circumstances. For instance, in a problem where the singular potential is cut off at some point $R > 0$, the problem is well-defined, and by matching interior ($r < R$) and exterior ($r > R$) solutions, consistent results may be obtained.

A simple example might be the problem of finding the S-wave scattering length for

$$V(r) = 0, \qquad r < R,$$

$$V(r) = -\frac{g^2}{r^4}, \quad r > R.$$

The interior solution is $\phi = r$, and the exterior is

$$\phi = B_+ r e^{ig r^{-1}} + B_- r e^{-ig r^{-1}}.$$

By continuity of ϕ and ϕ' at $r = R$, it is easy to determine that the scattering length for S waves is

$$\lim_{k\to 0} -\frac{\delta_0(k)}{k} = a = g \tan (g/R,) \tag{15}$$

and at $g/R = (2m + 1)\frac{1}{2}\pi$, bound states occur.

Another case where $\Delta = \pi$ solutions may be used is in the one-dimensional problem since the radial equation is then exact, and is not the reduction of a three-dimensional equation. Finally, it is possible to make these nonsolutions into correct three-dimensional solutions if a term is added to the potential which contains a $\delta(r)$ multiplied by just the right function to cancel out the delta-function term caused by ∇^2 in Eq. (11). This integral would then be zero by construction. This last point has been mentioned in a footnote by Newton.[31]

III. RADIAL SCHRÖDINGER EQUATION

What can be said of the solutions to Eq. (3) now that we know the circumstances under which it has meaning? We have already seen that the case $n = 4$ is rather special in that it can be solved exactly (for $k^2 = l = 0$) with a very simple function, rather than with the Bessel functions necessary for general n. Even if k^2 and l do not vanish, this potential has an exact solution while the other n's do not.[12,24,25] In this special case, the solutions are the modified Mathieu functions and are rather complicated due both to the general nature of Mathieu functions and the need to use complex parameters in some of the solutions.

The reasons for this unusual behavior, which is special to $n = 4$, will be seen below to follow simply from the general properties of the equation. We also see that $n = 6$ has some properties that set it apart from other values of n. In the material that follows, we follow closely the treatment of ordinary differential equations as given by Coddington and Levinson,[34] hereafter referred to as CL.

It is convenient to put $z = 1/r$ and transform the

[32] Though some form of distribution might be a solution, it is unlikely because, in the similar case of Laplace's equation, it is known that no such solutions exist. This is essentially the import of Hörmander's Theorems 4.1.4 and 4.1.5, which are extensions of Weyl's lemma. See Ref. 30, p. 96, 100, 101. In any case, in quantum mechanics we accept only continuous wavefunctions.

[33] I am indebted to Professor L. D. Favro for pointing this out to me.

[34] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill Book Co., Inc., New York, 1955), especially Chap. 5.

singularity in Eq. (3) to infinity so that we have

$$\frac{d^2\phi}{dz^2} + \frac{2}{z}\frac{d\phi}{dz} + \left(\frac{k^2}{z^4} - \frac{l(l+1)}{z^2} - \frac{V(z^{-1})}{z^4}\right)\phi = 0. \quad (16)$$

Then we write this as a second-order system:

$$y_1' = y_2,$$
$$y_2' = -\left[\frac{k^2}{z^4} - \frac{l(l+1)}{z^2} - \frac{V(z^{-1})}{z^4}\right]y_1 - \frac{2y_2}{z}, \quad (17)$$

where $y_1 = \phi$, $y_2 = \phi'$. Considering y_1 and y_2 as components of a two-dimensional vector, we write

$$\mathbf{y} = z^{\alpha}A(z)\mathbf{y}, \quad (18)$$

where $A(z)$ is a 2×2 matrix, each element of which is analytical in the neighborhood of $z = \infty$, $A(\infty) \neq 0$ (meaning that not every element of A vanishes), and $\alpha \geq 0$.

In particular, for the potential Eq. (4) with integral n,

$$\alpha = n - 4,$$
$$A(z) = \begin{bmatrix} 0 & z^{4-n} \\ -k^2z^{-n} + l(l+1)z^{-n+2} + g^2e^{-i\Delta} & -2z^{-n+3} \end{bmatrix}, \quad (19)$$

$$A(\infty) = \begin{bmatrix} 0 & \delta_{4,n} \\ g^2e^{-i\Delta} & 0 \end{bmatrix} \quad \text{if} \quad n \geq 4,$$

and

$$\alpha = 0,$$
$$A(z) = \begin{bmatrix} 0 & 1 \\ -k^2z^{-4} + l(l+1)z^{-2} + g^2e^{-i\Delta}z^{-1} & -2z^{-1} \end{bmatrix},$$
$$A(\infty) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{if} \quad n = 3.$$

Singularities of this type, with $\alpha \geq 0$, are known as singularities of the second kind and are enormously more difficult to analyze theoretically than those of the first kind. (Fuchian singularities, such as those that develop in the radial Schrödinger equation for regular potentials, are examples of singularities of the first kind.)

The nature of the solutions to the original Eqs. (3) or (16) breaks into two types, depending on whether the eigenvalues of $A(\infty)$ are degenerate or distinct. When they are distinct, it is possible to express the general solution to Eq. (3) as (we use here a special case of Theorem 2.1, Chap. 5 in CL)

$$\Phi = \left[\sum_{m=0}^{\infty} P_m r^m\right] r^{-M} e^{Q} \quad (P_m \text{ const}), \quad (20)$$

where M is a 2×2 diagonal matrix of complex constants and Q is a 2×2 matrix polynomial

$$Q = \frac{r^{-\alpha-1}}{\alpha+1}Q_0 + \frac{r^{-\alpha}}{\alpha}Q_1 + \cdots + r^{-1}Q_{\alpha},$$

with the Q_i complex 2×2 diagonal matrices and

$$Q_0 = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},$$

with λ_1 and λ_2 the distinct eigenvalues of $A(\infty)$. In Eq. (20) Φ is the 2×2 solution matrix, the first column being the first solution for the pair (y_1, y_2) and the second column being the other independent solution for (y_1, y_2).

From Eq. (19) we see that only for $n = 4$ does $A(\infty)$ have distinct eigenvalues and these are

$$\lambda_1 = ge^{-i\Delta/2}, \quad \lambda_2 = -ge^{-i\Delta/2},$$

and we have

$$Q = \frac{ge^{-i\Delta/2}}{r}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The quantity r^{-M} in (20) is defined as

$$r^{-M} = e^{-M\ln r} = \sum_{s=0}^{\infty}\frac{(-1)^s}{s!}(M\ln r)^s,$$

so that for diagonal M,

$$r^{-M} = \begin{bmatrix} r^{-m_1} & 0 \\ 0 & r^{-m_2} \end{bmatrix}$$

for some two constants m_1 and m_2. The quantity e^{Q} behaves similarly since Q is diagonal. Our two solutions are now

$$y_1^{\pm} = \phi_{\pm} = r\exp(\pm ge^{-i\Delta/2}r^{-1})\left[\sum_{m=0}^{\infty} P_m r^m\right], \quad (21)$$

where we know from Eq. (6) that $m_1 = m_2 = -1$. The interesting feature of the power series appearing in Eq. (21) is that it does *not* converge for any r, but is an asymptotic series. This is an unusual phenomenon for physicists who normally associate asymptotic series with behavior at infinity. However, as proved in Theorem 4.1 of Chap. 5 of CL, there do exist actual solutions to Eq. (3) for which the so-called formal solutions of Eq. (21) are their asymptotic expansions in the sense that

$$\lim_{r\to 0}\left|\phi^{\pm}_{\text{actual}} - r\exp(\pm ge^{-i\Delta/2}r^{-1})\sum_{m=0}^{N-1}P_m r^m\right| = O[r\exp(\pm ge^{-i\Delta/2}r^{-1})r^N].$$

It is possible to obtain the result (21) in a considerably less sophisticated manner by putting

$$\phi_{\pm} = r\exp(\pm ge^{-i\Delta/2}r^{-1})\chi(r)$$

and inserting this into Eq. (3) to obtain

$$\chi'' + \left[\frac{2}{r} + \frac{ge^{-i\Delta/2}}{r^2}\right]\chi' + \left[k^2 + \frac{l(l+1)}{r^2}\right]\chi = 0. \quad (22)$$

It is easy to see that a trial function of the form

$$\chi(r) = \sum_{m=0}^{\infty} P_m r^m \quad (23)$$

will formally solve Eq. (22), but it is not at all apparent that this series will be asymptotic rather than convergent. We note here the interesting point that if one equates the solution (21) with the solution given in terms of the Mathieu function as specified in Ref. 25, then one has automatically derived one of the lesser known asymptotic series for the Mathieu function.[35]

The asymptotic nature of the solution causes certain difficulties in any attempt to integrate the r^{-4} potential (or any other for that matter) by numerical methods.[36]

We turn now to the case where the eigenvalues of $A(\infty)$ are degenerate, or, in other words, $n \neq 4$. Even though asymptotic solutions still exist and are given in general form by Theorem 6.1 of Chap. 5 of CL, they are of greatly increased complexity compared with Eq. (21). The solution matrix Φ involves the matrix product of two factors. One of these is the exponentiation of a nondiagonal matrix, every element of which is a finite power series in a negative fractional power of r. The other factor is a matrix, every element of which is a product of a power of r times a finite power series in $\ln(r^{-1})$. In this power series every coefficient of each power of $\ln(r^{-1})$ is itself an infinite power series in a fractional power of r. Again a theorem exists relating this formal asymptotic solution to an actual solution.

It might appear that the complexity here is so great that nothing further can be done. However, $n = 6$ presents an interesting exception. The involved form for Φ just mentioned could simplify by accident in a given instance and this is precisely what happens for $n = 6$. By putting $r = t^{\frac{1}{2}}$ and $z = t^{-1}$, we may write Eq. (18) as

$$y_1' = y_2,$$
$$y_2' = -\left(\frac{k^2}{4z^3} - \frac{l(l+1)}{4z^2} - \tfrac{1}{4}g^2e^{-i\Delta}z^{(n/2)-3}\right)y_1 - \frac{3}{2z}y_2.$$

Thus, for $n = 6$,

$$\alpha = 0,$$
$$A(z) = \begin{bmatrix} 0 & 1 \\ -\tfrac{1}{4}[k^2z^{-3} - l(l+1)z^{-2} - g^2e^{-i\Delta}] & -\dfrac{3}{2z} \end{bmatrix}, \quad (24)$$
$$A(\infty) = \begin{bmatrix} 0 & 1 \\ \tfrac{1}{4}g^2e^{-i\Delta} & 0 \end{bmatrix}.$$

This is just the form of $A(\infty)$ required for the simpler type of solution in Eq. (20). However, in this case the argument of the solution is not r, but r^2. The reason we could not obtain this solution before is that Theorem 2.1 in CL gives sufficient, but not necessary, conditions on $A(z)$ in order that the solution be of the form of Eq. (20).

If we make the general substitution

$$\phi_{\pm}(r) = r^{n/4}\exp\left[\pm\frac{2g}{n-2}e^{-i\Delta/2}r^{1-(n/2)}\right]\chi(r), \quad (25)$$

and then put $r = t^{2/(n-2)}$, we can obtain the general equation corresponding to Eq. (22), which is

$$\chi'' + \left[\frac{2}{t} + \frac{4}{n-2}ge^{-i\Delta/2}t^{-2}\right]\chi' + \left(\frac{2}{n-2}\right)^2\left[k^2t^{\delta} - \frac{\lambda}{t^2}\right]\chi = 0, \quad (26)$$

where

$$\lambda = l(l+1) + \frac{n}{4}\left(\frac{n}{4} - 1\right) \quad \text{and} \quad \delta = \frac{8-2n}{n-2}.$$

We see that δ is an integer only for $n = 4$ or 6, and a trial solution of the form of Eq. (23) will produce a formal solution which, in fact, is not convergent.

Even though for all other n awkward solutions occur, it is always possible to obtain the first two terms of χ in Eq. (25). From Eq. (26) we easily find that, as $t \to 0$,

$$\chi(t) \sim 1 + \frac{\lambda}{g(n-2)}e^{i\Delta/2}t$$
$$\sim 1 + \frac{\lambda}{g(n-2)}e^{i\Delta/2}r^{(n-2)/2}.$$

Unfortunately, the next terms are likely to be exceedingly complex, as indicated above. Equation (26) indicates that no other n will produce a formal power series solution.

IV. SINGULAR YUKAWA POTENTIALS

Potentials of the form

$$V(r) = \frac{g^2e^{-i\Delta}e^{-\mu r}}{r^n} \quad (27)$$

[35] In particular, rather than the usual asymptotic series in inverse powers of cosh x, for large x, which is commonly given in the literature, the series derived is a power series in e^{-x}. See N. W. McLachlan, *Theory and Applications of Mathieu Functions* (Oxford University Press, Oxford, 1947).

[36] Near the origin, Eq. (22) involves large terms which diverge at $r = 0$. These terms analytically cancel, but they cause great trouble when numerical methods are used. Because the solution (21) is only asymptotic, it cannot be used satisfactorily to obtain a numerical start far enough away from the origin to render the divergent type terms harmless.

present more difficult problems since even at zero energy the solutions are not known. However, by repeating the analysis in Sec. III, we see that $A(\infty)$ is unaltered by the presence of e^{-ur}. Hence it is also possible to solve the singular fourth inverse-power Yukawa by an asymptotic power series. Such is not the case for $n = 6$, because the presence of $\exp(-\mu z^{-\frac{1}{2}})$ in $A(z)$ causes it to lose its analytic properties in the neighborhood of $z = \infty$.

For any n, however, it is possible to see that the behavior near the origin is

$$\lim_{r\to 0} \phi_{\pm} \sim V^{-\frac{1}{4}}(r) \exp\left[\pm \int^r V^{\frac{1}{2}}(r)\, dr\right]. \quad (28)$$

Note that, because of the rapid oscillation of the imaginary part of the exponent in Eq. (28), the radial behavior of the integrand cannot be replaced by $r^{-n/2}$. As $r \to 0$, $V^{-\frac{1}{4}}(r)$ may be replaced by $r^{n/4}$. For small enough r, it might be supposed that $V(r)$ could be represented by

$$V(r) \approx \frac{g^2 e^{-i\Delta}}{r^n} - \frac{\mu g^2 e^{-i\Delta}}{r^{n-1}};$$

and then the r^{-n+1} term neglected in the limit $r \to 0$. This cannot be done for the reason just mentioned. This nonanalytic type of behavior with respect to the r^{-n+1} term was first noted by Predazzi and Regge[9] for $n = 4$.

For small enough values of μg and $n = 4$, it is possible to approximate the S-wave scattering length by a Green's function method and then find the correction due to μ. We give the details in the Appendix which lead to the result (for $\Delta < \pi$)

$$a = g e^{-i\Delta/2}[\tfrac{1}{2} + \tfrac{3}{4} e^{-i\Delta/2}(\mu g) \ln(\mu g)]. \quad (29)$$

V. THE DIRAC EQUATION

As mentioned in the Introduction, Plesset[26] long ago sketched the solution for the radial Dirac equation when the potential was of the form of Eq. (4). It was not until some time later that Rose and Newton[37] pointed out that, as in the Schrödinger case, the wavefunctions were not usable for physical problems. This time both attractive and repulsive real potentials are excluded and for the different reason that unitarity is violated. We refer to their paper for the detailed statement and proof.

For complex strengths, or for the other uses indicated in Sec. II, we give the power series solutions to the two radial Dirac equations. (These were not given by Plesset.) First we write the two equations for $F = rf$ and $G = rg$, where f and g are the two standard radial Dirac functions for spherically symmetric potentials[38]:

$$\begin{aligned} G' &= -\frac{q}{r} G + (\alpha_1 - V(r))F, \quad \alpha_1 = E + mc^2, \\ F' &= \frac{q}{r} F - (\alpha_2 - V(r))G, \qquad \alpha_2 = E - mc^2. \end{aligned} \quad (30)$$

Putting $r = z^{-1}$, we obtain

$$\begin{aligned} G' &= \frac{q}{z} G - \left[\frac{\alpha_1}{z^2} - \frac{V(z^{-1})}{z^2}\right]F, \\ F' &= -\frac{q}{z} F + \left[\frac{\alpha_2}{z^2} - \frac{V(z^{-1})}{z^2}\right]G, \end{aligned} \quad (31)$$

or

$$\mathbf{y}' = z^{\alpha} A(z) \mathbf{y},$$

with

$$\alpha = n - 2,$$

$$A(z) = \begin{bmatrix} qz^{-n+1} & -\alpha_1 z^{-n} + g^2 e^{-i\Delta} e^{-\mu/z} \\ \alpha_2 z^{-n} - g^2 e^{-i\Delta} e^{-\mu/z} & -qz^{-n+1} \end{bmatrix},$$

$$A(\infty) = \begin{bmatrix} 0 & g^2 e^{-i\Delta} \\ -g^2 e^{-i\Delta} & 0 \end{bmatrix}.$$

We easily see from the results quoted in Sec. III that, for $n \geq 2$, power series solutions (multiplied by the appropriate exponential function) result for all n, even in the case of the Yukawa types.

With the aid of Eq. (31), we see that the solution to Eq. (30) is of the form

$$\begin{aligned} G &= \exp\left[\pm i \int^r V(r)\, dr\right] H_g(r), \\ F &= \mp i \exp\left[\pm i \int^r V(r)\, dr\right] H_f(r). \end{aligned} \quad (32)$$

Forming $S(r) = H_g + H_f$ and $D(r) = H_g - H_f$ and substituting Eq. (32) into Eq. (30), we obtain

$$\begin{aligned} S' + \left(\frac{q}{r} \pm i\lambda\right) D \pm i\delta S = 0, \\ D' + \left(\frac{q}{r} \mp i\lambda\right) S \pm i(2V - \delta) D = 0, \end{aligned} \quad (33)$$

with $\lambda = -mc^2$, $\delta = E$. Assuming a solution to Eq. (33) of the form

$$S = \sum_{m=0}^{\infty} s_m r^m, \quad D = \sum_{m=0}^{\infty} d_m r^{(n-1)+m},$$

[37] M. E. Rose and R. R. Newton, Phys. Rev. **82**, 470 (1951).

[38] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Co. Inc., New York, 1955), Chap. 12.

we may obtain, specializing to the non-Yukawa case $\mu = 0$ for $m \leq n-2$,

$$s_m = (\pm i)^m \frac{\delta^m}{m!} s_0,$$

$$d_0 = \pm i \frac{q e^{i\Delta}}{2g^2} s_0,$$

$$d_m = (\mp i)^{m-1} \delta^{m-1} \frac{e^{i\Delta}}{2g^2} \left[\frac{q\delta}{m!} + \frac{\lambda}{(m-1)!}\right] s_0, \quad m > 0.$$

When $m = n-1$, we have

$$s_{n-1} = \left[(\pm i)^m \frac{\delta^m}{m!} \mp \frac{iq^2 e^{i\Delta}}{2mg^2}\right] s_0,$$

$$d_{n-1} = \left[(\mp i)^{m-1} \delta^{m-1} \frac{e^{i\Delta}}{2g^2} \left(\frac{q\delta}{m!} + \frac{\lambda}{(m-1)!}\right) - \frac{qm}{4g^4} e^{2i\Delta}\right] s_0.$$

For $m > n$, the recursion relations are

$$s_m = \pm \frac{i\delta}{m} s_{m+1} - \frac{1}{m}(q d_{m-n+1} \pm i\lambda d_{m-n}),$$

$$d_m = \frac{\pm i e^{i\Delta}}{2g^2}(q s_m \mp i\lambda s_{m-1}) \pm \frac{i e^{i\Delta}}{2g^2}(m d_{m-n+1} \mp i\delta d_{m-n}).$$

We see that the functions H_g and H_f are of the form

$$H_g = \frac{1}{2}\sum_{m=0}^{n-2} s_m r^m + \frac{1}{2}\sum_{m=n-1}^{\infty} c_m r^m, \quad c_m = s_m + d_m,$$

$$H_f = \frac{1}{2}\sum_{m=0}^{n-2} s_m r^m + \frac{1}{2}\sum_{m=n-1}^{\infty} e_m r^m, \quad e_m = s_m - d_m.$$

These series are, of course, asymptotic, not convergent.

We note the interesting feature of Eq. (32): that for neither attractive nor repulsive potentials do the solutions go to zero exponentially at the origin; but rather, they oscillate wildly. This is similar to the attractive potential case in the Schrödinger equation and is related to the Klein paradox. This behavior results from the fact that near the origin a Dirac particle may have very large momentum and large negative kinetic energy even when the potential is strongly repulsive. As is well known, the physical one-particle interpretation of the Dirac equation must be altered in such situations.[39]

VI. CONCLUSION

There are a number of cases where the use of singular potentials is of interest physically. In addition, the mathematical behavior of the scattering by singular potentials is of interest in helping to study the divergences in field theory. In fact, the study of these singular systems sheds almost the only available light on what physics could be like if field theory or S-matrix theory do not have the nice analytic properties usually conjectured.

For the Schrödinger equation, it was seen that only r^{-4} and r^{-6} potentials lend themselves to simple asymptotic representations, while only the singular Yukawa behaving as $e^{-\mu r} r^{-4}$ has a simple type of solution. The leading two terms are available in all cases, but the general form of the solution is, usually, extremely complicated.

In the Dirac equation, all the potentials considered are amenable to an asymptotic series solution. In this case the behavior of repulsive potentials near the origin differs radically from the nonrelativistic behavior.

VII. ACKNOWLEDGMENTS

The author is deeply indebted to Professor Thomas Seidman of the Wayne State University Mathematics Department for acquainting him with Hörmander's book, as well as for some very helpful discussions about the material in it. He is also indebted to Professor L. D. Favro for helpful discussions and to Professor David Fradkin for his interest and useful suggestions.

APPENDIX

For

$$V(r) = g^2 e^{-i\Delta}/r^4,$$

the S-wave scattering length is given simply by

$$a = g e^{-i\Delta/2}. \tag{A1}$$

To obtain a result for

$$V(r) = \frac{g^2 e^{-i\Delta} e^{-\mu r}}{r^4},$$

we write the radial Schrödinger equation as

$$\phi'' - \frac{g^2 e^{-i\Delta}}{r^4}\phi = \frac{g^2 e^{-i\Delta}}{r^4}(e^{-\mu r} - 1)\phi. \tag{A2}$$

The Green's function for this equation is made from the solutions for $\mu = 0$, so that

$$G(r, r') = \frac{e^{i\Delta 2}}{2g} rr' \exp(-g e^{-i\Delta/2} r^{-1}) \times \exp(g e^{-i\Delta/2} r'^{-1}), \quad r < r',$$

$$G(r, r') = \frac{e^{i\Delta/2}}{2g} rr' \exp(g e^{-i\Delta/2} r^{-1}) \times \exp(-g e^{-i\Delta/2} r'^{-1}), \quad r > r'. \tag{A3}$$

[39] J. D. Bjorken and S. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill Book Co., Inc., New York, 1964).

The solution to Eq. (A2) may now be written as

$$\phi(r) = \phi_0(r) - \frac{ge^{-i\Delta/2}}{2} r \exp(ge^{-i\Delta/2}r^{-1})$$
$$\times \int_0^r r'^{-3}(e^{-\mu r'} - 1) \exp(-ge^{-i\Delta/2}r'^{-1})\phi(r')\, dr'$$
$$- \frac{ge^{-i\Delta/2}}{2} r \exp(-ge^{-i\Delta/2}r^{-1})$$
$$\times \int_r^\infty r'^{-3}(e^{-\mu r'} - 1) \exp(ge^{-i\Delta/2}r'^{-1})\phi(r')\, dr', \quad \text{(A4)}$$

where $\phi_0 = r \exp(-ge^{-i\Delta/2}r^{-1})$ is the solution to the Schrödinger equation for $\mu = 0$. The scattering length is obtained from Eq. (A4) by taking the limit

$$\lim_{r\to\infty} \phi(r) \sim r - ge^{-i\Delta/2} - \frac{ge^{-i\Delta/2}}{2}(r + ge^{-i\Delta/2})$$
$$\times \int_0^\infty r'^{-3}(e^{-\mu r'} - 1) \exp(-ge^{-i\Delta/2}r'^{-1})\phi(r')\, dr'$$
$$- \lim_{r\to\infty} \frac{ge^{-i\Delta/2}}{2}(r - ge^{-i\Delta/2}) \int_r^\infty r'^{-3}(e^{-\mu r'} - 1)$$
$$\times \exp(ge^{-i\Delta/2}r'^{-1})\phi(r')\, dr', \quad \text{(A5)}$$

and retaining constant terms and terms linear in r.

If μg is small enough, we may solve Eq. (A5) by iteration and put $\phi(r) = \phi_0(r)$ inside of the integrals. This gives us, for small μg, the first correction to Eq. (A1). We have

$$\lim_{r\to\infty} \phi(r) \sim r - ge^{-i\Delta/2} - \frac{ge^{-i\Delta/2}}{2}(r + ge^{-i\Delta/2})$$
$$\times \int_0^\infty r'^{-2}(e^{-\mu r'} - 1) \exp(-2ge^{-i\Delta/2}r'^{-1})\, dr'$$
$$- \lim_{r\to\infty} \frac{ge^{-i\Delta/2}}{2}(r - ge^{i\Delta/2}) \int_r^\infty r'^{-2}(e^{-\mu r'} - 1)\, dr'. \quad \text{(A6)}$$

As $r \to \infty$, it is easy to see that the only term in the second integral that causes a contribution to the scattering length is

$$-\int_r^\infty r^{-2}\, dr = -\frac{1}{r}.$$

The other term falls off exponentially rapidly since it is essentially the exponential integral. Thus the last term in Eq. (A6) contributes a constant term $\frac{1}{2}ge^{-i\Delta/2}$. Note that if $\mu = 0$, this term is not present since it is cancelled by the first term in the integral. But as long as $\mu > 0$, regardless of how small it is, we obtain this term.

The first integral in Eq. (A6) has a second part which is easily integrated:

$$-\int_0^\infty r^{-2} \exp(-2ge^{-\Delta/2}r^{-1})\, dr = -\frac{e^{i\Delta/2}}{2g}, \quad \Delta < \pi.$$

The first part,

$$\int_0^\infty r^{-2} \exp(-2ge^{-i\Delta/2}r^{-1})e^{-\mu r}\, dr, \quad \text{(A7)}$$

is the Laplace transform of $r^{-2}\exp(-2ge^{-i\Delta/2}r^{-1})$. The result is available from the general result[40] that

$$\int_0^\infty e^{-pt}(t^{\nu-1}e^{-b/(4t)})\, dt = 2\left[\frac{b}{2p}\right]^{\nu/2} K_\nu[(bp)^{\frac{1}{2}}] \begin{matrix}\text{Re } b > 0 \\ \text{Re } p > 0,\end{matrix}$$

where $K_\nu(x)$ is the modified Bessel function of the third kind.

So Eq. (A7) becomes

$$2\left[\frac{\mu}{2g}\right]^{\frac{1}{2}} e^{i\Delta/4} K_1(2e^{-i\Delta/4}(2\mu g)^{\frac{1}{2}}).$$

Collecting terms, we have

$$\lim_{r\to\infty} \phi(r) \sim r\left[1 - \left(\frac{\mu g}{2}\right)^{\frac{1}{2}} e^{-i\Delta/4} K_1(2e^{-i\Delta/4}(2\mu g)^{\frac{1}{2}}) + \tfrac{1}{4}\right]$$
$$- \left[ge^{-i\Delta/2} + g\left(\frac{\mu g}{2}\right)^{\frac{1}{2}} e^{-3i\Delta/4} K_1(2e^{-i\Delta/4}(2\mu g)^{\frac{1}{2}})\right.$$
$$\left. - \tfrac{3}{4}ge^{-i\Delta/2}\right]. \quad \text{(A8)}$$

Thus the scattering length is

$$a = \frac{\frac{1}{4}ge^{-i\Delta/2} + g(\frac{1}{2}\mu g)^{\frac{1}{2}} e^{-3i\Delta/4} K_1(2e^{-i\Delta/4}(2\mu g)^{\frac{1}{2}})}{\frac{5}{4} - (\frac{1}{2}\mu g)^{\frac{1}{2}} e^{-i\Delta/4} K_1(2e^{-i\Delta/4}(2\mu g)^{\frac{1}{2}})}. \quad \text{(A9)}$$

Since for small x, $K_1(x) \to 1/x + \frac{1}{2}x \ln x + O(x)$, we may approximate (A9) by keeping only the first two terms in the Bessel function expansion to obtain

$$a = \frac{\frac{1}{2}ge^{-i\Delta/2} + g(\mu g)e^{-i\Delta/2} \ln[2(2\mu g)^{\frac{1}{2}}]}{1 - (\mu g)e^{-i\Delta/2} \ln[2(2\mu g)^{\frac{1}{2}}]},$$

or, for small enough μg,

$$a = ge^{-i\Delta/2}[\tfrac{1}{2} + \tfrac{3}{4}(\mu g)e^{-i\Delta/2} \ln(\mu g)], \quad \text{(A10)}$$

which is the expression given in Eq. (29). Due to the extra term mentioned above, this expression does not reduce to Eq. (A1) in the limit $\mu g \to 0$. Thus the dominant effect of the Yukawa exponential term is to cut the scattering length in half. This type of behavior was mentioned in Sec. IV, where it affected the wavefunction itself.

[40] Bateman Manuscript Project, *Tables of Integral Transforms* A. Erdelyi, Ed. (McGraw-Hill Book Co., Inc., New York, 1954), Vol. I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Bethe–Salpeter Equation in Momentum Space*†

ROBERT M. SAENGER
Stanford Linear Accelerator Center, Stanford University, Stanford, California

(Received 10 December 1966)

The Wick transformation in momentum space is modified to include all scattering energies by use of a coordinate surface which possesses limited detours into the complex relative energy plane. This device retains the simple form of the equation for the Bethe–Salpeter amplitude $\psi(\mathbf{p}, p_0)$. It is shown that the transformation is valid if $\psi(\mathbf{p}, p_0)$ has the cut structure indicated by field theory, and this structure is shown to be consistent with the Bethe–Salpeter equation provided the interaction $V(x)$ satisfies a simple causality condition. It is further shown that the cut structure of a solution to the transformed equation can be deduced from the causal structure of the interaction alone without reference to field theory. Basic properties of the transformed equation are derived and a numerical treatment for purely elastic scattering is presented.

1. INTRODUCTION

THIS paper will be concerned with the Bethe–Salpeter equation[1] for two nonidentical spinless mesons ("a" and "b") of equal mass m interacting in the ladder diagram approximation via the field of a third spinless meson of mass μ. Principal attention will be paid to the scattering case in which E, the total center-of-mass frame energy, is greater than $2m$. In relative momentum four-space in the center-of-mass frame the equation will be taken as

$$\psi(p) = \psi_0(p) - \frac{1}{D(p)} \int V(p - \eta)\psi(\eta)\, d^4\eta, \quad (1.1)$$

where $p = (\mathbf{p}, p_0)$, $\eta = (\mathbf{n}, \eta_0)$, and with $q^2 = \mathbf{q}^2 - q_0^2 (\hbar = c = 1)$.

The function of $D(p)$ is

$$D(p) = [(p + \tfrac{1}{2}E)^2 + m^2][(p - \tfrac{1}{2}E)^2 + m^2] \quad (1.2)$$

$$= (p^2 - k^2)^2 - E^2 p_0^2, \quad (1.3)$$

$$k^2 \equiv (E/2)^2 - m^2, \quad (1.4)$$

$$V(p - \eta) = \pi^{-2}[(p - \eta)^2 + \mu^2]^{-1} \quad (1.5)$$

and, for the scattering case, $E > 2m$,

$$\psi_0(p) = \delta^3(\mathbf{p} - \mathbf{p}_{\text{in}})\delta(p_0) \quad (1.6)$$

with incident center-of-mass momenta

$$\mathbf{p}_{\text{in}} = (\mathbf{p}_a)_{\text{in}} - (\mathbf{p}_b)_{\text{in}}, \quad (1.7)$$

$$|\mathbf{p}_{\text{in}}| = k; \quad (1.8)$$

$\psi(x)$, the configuration space amplitude, is to be related to $\psi(p)$ by

$$\psi(x) = \int e^{i(\mathbf{p}\mathbf{x} - p_0 x_0)}\psi(p)\, d^4p, \quad (1.9)$$

and the Feynman convention $m \to m - i\epsilon$, $\mu \to \mu - i\epsilon'$ is then understood as the proper way of handling singularities in $1/[D(p)]$ and $V(p - \eta)$. For the case of purely elastic scattering, $2m < E < 2m + \mu$, Schwartz and Zemach[2] have presented an analysis of the Bethe–Salpeter equation in the circumstance considered here and have produced numerical scattering results. Their approach is related to the previous work on the bound-state case by Wick[3] in that, as Wick, they use the device of rotation to imaginary values of the relative time variable to produce a transformed integral equation with the standard Euclidean metric which is more susceptible to solution than the nonrotated equation with its Lorentz metric.[2a]

The treatment of Schwartz and Zemach for elastic scattering energies differs somewhat from the approach of Wick in that the former authors perform the sought-after rotation in configuration space, whereas the rotation of Wick for the bound state case ($E < 2m$, $\psi_0(p) \equiv 0$) is performed primarily in momentum space, i.e., is primarily concerned with rotation of the variables p_0 and η_0 in Eq. (1.1). For the scattering case, the relationship between rotations in these two forms of the equation is not trivial, as is apparent whenever calculations in momentum space become a practical necessity. Thus Schwartz and Zemach are forced to use a distorted contour in momentum space when evaluating certain integrals needed to obtain their numerical results.[4]

We present below a simple method for extending the Wick rotation in momentum space to scattering energies. In this modified procedure the necessary

* Work supported in part by the U.S. Atomic Energy Commission.

† All the numerical work and most of the theoretical work in this paper was performed while the author was a National Science Foundation Fellow at Massachusetts Institute of Technology in Cambridge, Mass.

[1] E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951); M. Gell-Mann and F. Low, *ibid.* **84**, 350 (1951); J. Schwinger, Proc. Acad. Sci. U.S. **37**, 455 (1951).

[2] C. Schwartz and C. Zemach, Phys. Rev. **141**, 1454 (1966). (a) Recent work on the momentum space aspects of the Bethe–Salpeter equation appears in the following: A. Pagnamenta and J. G. Taylor, Phys. Rev. Letters **17**, 218 (1966); M. I. Levine, J. A. Wright, and J. A. Tjon, University of California, preprint (September 1966).

[3] G. C. Wick, Phys. Rev. **96**, 1124 (1954).

[4] See Ref. 2, Fig. 2.

contour distortion presented by Schwartz and Zemach plays a fundamental role. Many features of the Bethe–Salpeter equation can be easily demonstrated for the resulting momentum space representation. Also, a method of obtaining numerical results in the elastic energy range from the equation in this form will be discussed, and some computed values presented which serve as a rough check on the Schwartz–Zemach procedure and other more recent calculations.[2a]

2. TRANSFORMATION IN MOMENTUM SPACE

As an initial step toward transformation of the Bethe–Salpeter equation in momentum space, one can refer back to Wick's analysis (based on the operators and states of field theory) of the structure of $\psi(\mathbf{p}, p_0)$ as a function of complex p_0. For $E > 2m$, as for $E < 2m$, the Wick analysis states that $\psi(\mathbf{p}, p_0)$ is analytic in p_0 everywhere except along two cuts. One cut lies just below the real p_0 axis extending from $\omega_{\min}(\mathbf{p}) - i\epsilon$ to $+\infty - i\epsilon$, while the other lies just above the real axis from $\omega_{\max}(\mathbf{p}) + i\epsilon$ to $-\infty + i\epsilon$, where (for the mass equal case)

$$\omega_{\min}(\mathbf{p}) = -\omega_{\max}(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{\frac{1}{2}} - E/2. \quad (2.1)$$

Although this analysis depends on field theory, it can be shown that the structure implied by Eq. (2.1) is consistent with the structure of the equation itself, as will be demonstrated later in this section. The location of the cuts in $\psi(\mathbf{p}, p_0)$ is shown in Fig. 1.

For $\mathbf{p}^2 > k^2$, $\omega_{\min}(\mathbf{p})$ is positive so that there is a gap between the two cuts and one can analytically continue $\psi(\mathbf{p}, p_0)$ counterclockwise from the real axis to the imaginary axis in complex p_0, as is indicated in Fig. 1(a). For $\mathbf{p}^2 \leq k^2$ the two cuts in $\psi(\mathbf{p}, p_0)$ as a function of p_0 overlap since then $\omega_{\min}(\mathbf{p})$ is negative. Thus for $\mathbf{p}^2 \leq k^2$ one cannot analytically continue counterclockwise from the real p_0 axis to the imaginary p_0 axis without encountering singularities in $\psi(\mathbf{p}, p_0)$, but one can continue from the real axis to a contour which follows the imaginary axis except for two detours around those portions of the cuts which protrude below and above $p_0 = 0$, continuing to $\omega_{\min}(\mathbf{p}) - i\epsilon$ and $-\omega_{\min}(\mathbf{p}) + i\epsilon$, respectively. The suggested contour for $\mathbf{p}^2 \leq k^2$ is shown in Fig. 1(b), and will be recognized as the contour that was introduced by Schwartz and Zemach. Since the suggested p_0 contour for consideration of the analytic continuation of $\psi(\mathbf{p}, p_0)$ depends on $\mathbf{p}^2$, it is called $C_{\mathbf{p}^2}$. For $\mathbf{p}^2 > k^2$ the contour $C_{\mathbf{p}^2}$ is just the imaginary axis [see Fig. 1(a)]. (The pathology of $\mathbf{p}^2 = k^2$ will be considered in more detail below.)

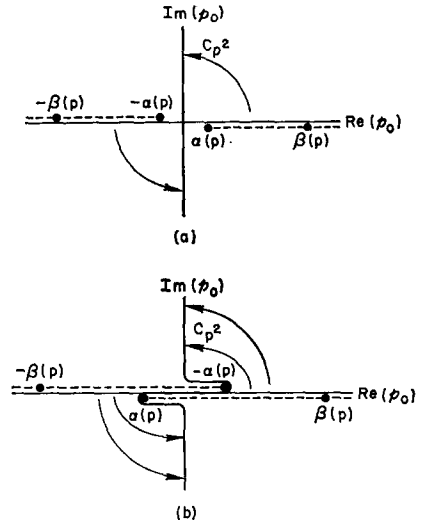


FIG. 1. (a) Cuts in $\psi(\mathbf{p}, p_0)$ for $\mathbf{p}^2 > k^2$ [cuts start at $\pm\alpha(\mathbf{p}) = \pm(\omega_{\min}(\mathbf{p}) - i\epsilon)$; arrows indicate continuation in p_0]. (b) Cuts in $\psi(\mathbf{p}, p_0)$ for $\mathbf{p}^2 \leq k^2$.

One can now convert Eq. (1.1) into an integral equation for the analytic continuation of $\psi(\mathbf{p}, p_0)$ from real four-space to a new coordinate surface which consists of all real values of $\mathbf{p}$, but only those complex values of p_0 which lie on $C_{\mathbf{p}^2}$. In order that this conversion be accomplished, one must analytically continue in p_0 all terms on the right-hand side of Eq. (1.1),

$$\psi_0(\mathbf{p}, p_0) - \frac{1}{D(p)} \iint V(p - \eta)\psi(\eta)\, d\eta_0\, d^3\mathbf{n} \quad (2.2)$$

from the real axis to $C_{\mathbf{p}^2}$, and also one must deform the η_0 contour in Eq. (2.2) from the real axis to the contour $C_{\mathbf{n}}$ for every $\mathbf{n}$.

Consider the first term in Eq. (2.2),

$$\psi_0(\mathbf{p}, p_0) = \delta^3(\mathbf{p} - \mathbf{p}_{\text{in}})\delta(p_0). \quad (2.3)$$

To treat this term consistently, one needs a representation of $\delta(p_0)$ which exhibits the singular structure indicated by the Wick analysis. Such a representation is readily achieved by writing

$$\delta(p_0) = \frac{1}{2\pi i}\left(\frac{1}{p_0 - i\epsilon} - \frac{1}{p_0 + i\epsilon}\right), \quad (2.4)$$

where ϵ is positive real and arbitrarily small. With $\delta(p_0)$ in this form, ψ_0 has singularities in p_0 at $p_0 = \pm i\epsilon$. But the points $\pm i\epsilon$ fall in the cut region allowed by Eq. (2.1) for $\mathbf{p}^2 \leq k^2$, and hence are avoided by counterclockwise continuation to $C_{\mathbf{p}^2}$ for $\mathbf{p}^2 \leq k^2$. Since the incident particles are strictly on the "mass shell" $\mathbf{p}^2 = k^2$, $\psi_0(\mathbf{p}, p_0) \equiv 0$, if $\mathbf{p}^2 \neq k^2$, and therefore the singularities in $\psi_0(\mathbf{p}, p_0)$ with $\delta(p_0)$ in form Eq. (2.4) conform to the Wick structure indicated by Eq. (2.1).

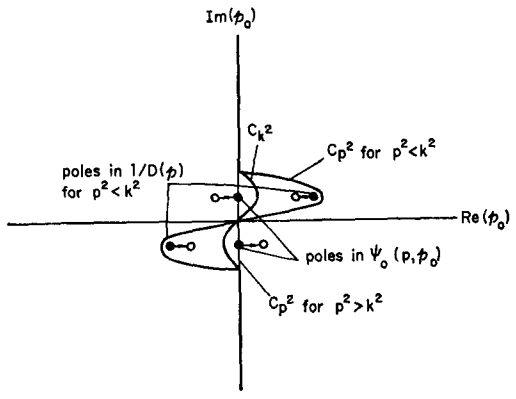


FIG. 2. Location of poles in $1/D(p)$ relative to C_{p^2} (arrows indicate relocation of poles under $\epsilon = \epsilon_1 \to \epsilon_1 + i\epsilon_2$).

Further, if one lets ϵ have an infinitesimal positive imaginary part $\epsilon \to \epsilon_1 + i\epsilon_2$, then the points $\pm i\epsilon$ are also avoided by the contour rotation $\mathbf{p}^2 > k^2$ and all ambiguity in defining C_{p^2} for $\mathbf{p}^2$ near k^2 is eliminated. With this convention for ϵ, the poles in Eq. (2.4) at $\pm i\epsilon$ remain on the same side of the p_0 contour throughout and the right-hand side of Eq. (2.4) satisfies

$$\int_C f(p_0) \cdot [\text{r.h.s. of (2.4)}]\, dp_0 \to f(0) \qquad (2.5)$$

in the limit $|\epsilon| \to 0$ whether C is the real axis or C_{p^2} for any $\mathbf{p}$ (provided f is analytic at $p_0 = 0$). Note (see Fig. 2) the convention $\epsilon \to \epsilon_1 + i\epsilon_2$, $i\epsilon \to i\epsilon_1 - \epsilon_2$ is a natural one for relocating poles relative to C_{p^2}. Thus, with Eq. (2.4) and $\epsilon \to \epsilon_1 + i\epsilon_2$, $\psi_0(\mathbf{p}, p_0)$ can be continued to C_{p^2} for all $\mathbf{p}$ and the formal identification $\psi_0(\mathbf{p}, p_0) = \delta^3(\mathbf{p} - \mathbf{p}_{in})\delta(p_0)$ can be retained for the analytically continued ψ_0.

It remains to analytically continue

$$-\frac{i}{D(p)} \int V(p - \eta)\psi(\eta)\, d\eta_0\, d^3\mathbf{n} \qquad (2.6)$$

in p_0 from the real axis to C_{p^2}.

Continuation of the factor

$$1/D(p) = [(p_0 - \alpha)(p_0 + \alpha)(p_0 - \beta)(p_0 + \beta)]^{-1}, \qquad (2.7)$$

where

$$\alpha(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{\frac{1}{2}} - (E/2) - i\epsilon, \qquad (2.8)$$

$$\beta(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{\frac{1}{2}} + (E/2) - i\epsilon, \qquad (2.9)$$

($-i\epsilon$ is assigned by the Feynman convention and $\epsilon = \epsilon_1 + i\epsilon_2$ is understood to be available if needed) is easy since the poles at $\pm\beta(\mathbf{p})$ never obstruct continuation [note Im $(\beta) < 0$, Re $(\beta) > 0$)] while $\alpha(\mathbf{p})$ and $-\alpha(\mathbf{p})$ lie at $\omega_{\min}(\mathbf{p}) - i\epsilon$ and $-\omega_{\min}(\mathbf{p}) +$

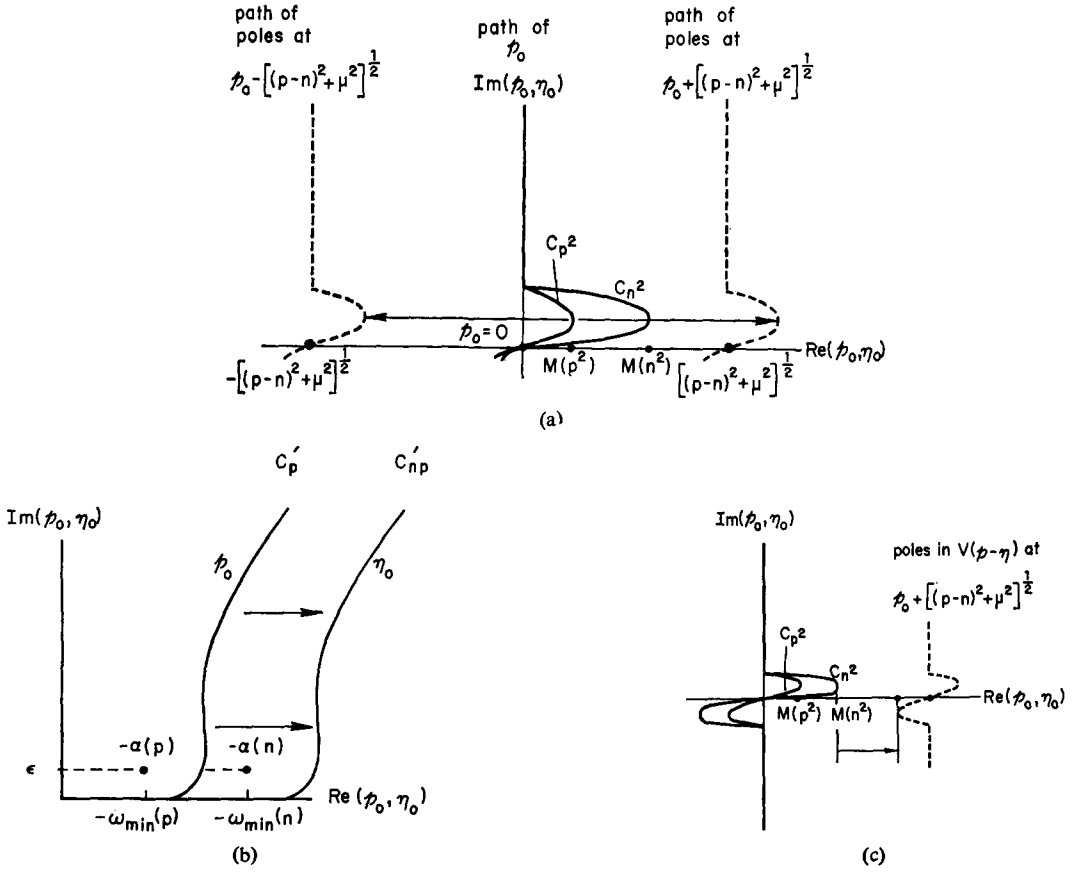


FIG. 3 (a), (b), (c). (See legend opposite.)

$i\epsilon$, respectively, and hence are specifically avoided by C_{p^2} [see Fig. 1(b)].

Finally, as we now show, simultaneous continuation in p_0 and deformation of the η_0 contour in the term

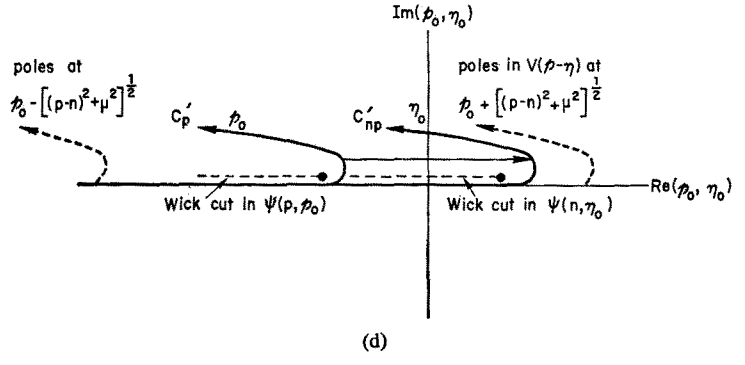
$$\iint V(p-\eta)\psi(n,\eta_0)\,d\eta_0\,d^3\mathbf{n} \equiv T(p) \qquad (2.10)$$

can be accomplished in the manner presented by Wick for the bound-state case.[3] In order to simplify the discussion of this process, one may consider transformation of the η_0 contour for the special value of $p_0 = 0$ which lies on both the real axis and C_{p^2}. Subsequently, one can consider analytic continuation of this transformed integral to all p_0 on C_{p^2}.

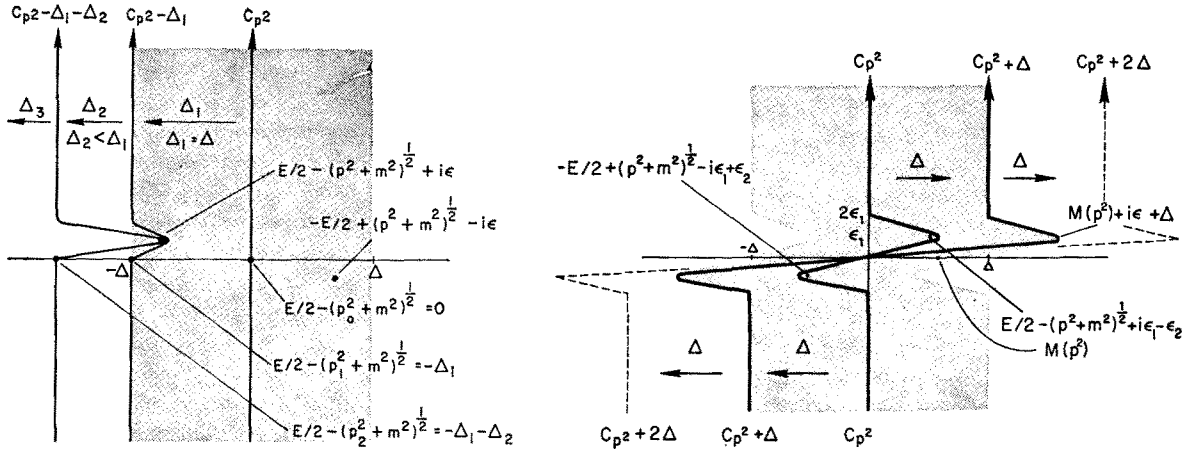
It is established by assumption of the Wick structure that singularities in $\psi(\mathbf{n}, \eta_0)$ will not interfere with η_0 contour deformation from the real axis (counterclockwise) to C_{n^2}. That the poles in $V(p-\eta)$ at $\pm[(\mathbf{p}-\mathbf{n})^2+\mu^2]^{\frac{1}{2}}$ will not interfere with the desired rotation is illustrated in Fig. 3(a). Hence, the η_0 contour change is valid for $p_0 = 0$. Now one wants to move p_0 from $p_0 = 0$ along C_{p^2}, thus analytically continuing the transformed integral

$$\int_{C_{n^2}} \frac{\psi(\mathbf{n},\eta_0)\,d\eta_0}{(\mathbf{p}-\mathbf{n})^2 + (p_0-\eta_0)^2 + \mu_2} \qquad (2.11)$$

to all p_0 on C_{p^2}. This process of analytic continuation can be taken trivially under the integral sign provided the poles at $r_\pm = p_0 \pm [(\mathbf{p}-\mathbf{n})^2+\mu^2]^{\frac{1}{2}}$ never cross the contour C_{n^2}. The two paths followed by these



(d)



(e) (f)

FIG. 3. (a) Continuation of $T(\mathbf{p}, p_0)$ by moving p_0 from $p_0 = 0$ along C_{p^2}. All η_0 contours have previously been deformed to C_{n^2} after using $\mu \to \mu - i\epsilon'$ to locate poles in $V(p-\eta)$. (b) Continuation along C'_p Arrows are of length $M(\mathbf{n}^2) - M(\mathbf{p}^2)$. (In the lower half plane the arrows would point in the opposite direction.) (c) Separation between η_0 on C_{n^2} and poles in V for $E < 2m + \mu$. Arrow is of length $\{[(\mathbf{p}-\mathbf{n})^2+\mu^2]^{\frac{1}{2}} - M(\mathbf{p}^2) - M(\mathbf{n}^2)\}$. (d) Continuation of $T(\mathbf{p}, p_0)$ up to the Wick cuts in $\psi(\mathbf{p}, p_0)$. Arrow length is $M(\mathbf{n}^2) - M(\mathbf{p}^2)$. (e) Shifting the p_0 contour for a specific $\mathbf{p}$, $\mathbf{p}^2 < k^2$, in steps of size Δ through the first and third quadrants. Shaded area indicates initial strip of analyticity in $T(\mathbf{p}, p_0)$ for this $\mathbf{p}$. [The validity of the procedure shown here is generally independent of small details in contour shape. However, two criteria must be satisfied: (1) The maximum horizontal separation between any two of the original or shifted contours, e.g., C_{p^2} and C_{n^2}, must be given by the difference between their maximum real part extensions, $|M(\mathbf{p}^2) - M(\mathbf{n}^2)|$. (2) In shifting horizontally from C_{p^2} to $C_{p^2} + \Delta$, some points on C_{p^2} may be displaced less than Δ but none may be displaced more than Δ. These criteria are easy to meet, as the above choice of linearly segmented contours demonstrates.] (f) Shifting the p_0 contour for a specific $\mathbf{p}$, $\mathbf{p}_1^2 > \mathbf{p}^2 > \mathbf{p}_0^2 = k^2$, in steps of size Δ_i through the left-half plane. Shading indicates initial strip of analyticity in $T(\mathbf{p}, p_0)$. [The criteria mentioned under Fig. 3(e) are again observed.]

poles as p_0 moves along $C_{\mathbf{p}^2}$ are shown in Fig. 3(a). Defining $M(\mathbf{p}^2)$ to be the maximum real part extension of $C_{\mathbf{p}^2}$,

$$M(\mathbf{p}^2) = 0 \quad \text{if} \quad \mathbf{p}^2 > k^2,$$
$$M(\mathbf{p}^2) = \tfrac{1}{2}E - (\mathbf{p}^2 + m^2)^{\frac{1}{2}} \quad \text{if} \quad \mathbf{p}^2 \leq k^2, \tag{2.12}$$

it will be seen from Fig. 3(a) that the poles in question always stay on the same respective sides of $C_{\mathbf{n}^2}$ (although they may come infinitesimally close to it) provided

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > |M(\mathbf{p}^2) - M(\mathbf{n}^2)| \tag{2.13a}$$

for all $\mathbf{p}$ and $\mathbf{n}$. If $\mathbf{p}^2$ and $\mathbf{n}^2$ are greater than k^2, Eq. (2.13a) holds trivially and rotation is possible as in the bound-state case. Now consider $\mathbf{n}^2 < \mathbf{p}^2$, $\mathbf{n}^2 < k^2$ (the case $\mathbf{p}^2 < \mathbf{n}^2$, $\mathbf{p}^2 < k^2$ will be covered by symmetry). Then, Eq. (2.13a) becomes

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > \tfrac{1}{2}E - (\mathbf{n}^2 + m^2)^{\frac{1}{2}} - M(\mathbf{p}^2). \tag{2.14}$$

But $M(\mathbf{p}) \geq 0$, so it suffices to show

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > \tfrac{1}{2}E - (\mathbf{n}^2 + m^2)^{\frac{1}{2}}. \tag{2.15}$$

Since

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} \geq \mu \tag{2.16}$$

and $(\mathbf{n}^2 + m^2)^{\frac{1}{2}} \geq m$, it suffices that

$$\mu > \tfrac{1}{2}E - m \quad \text{or} \quad E < 2m + 2\mu. \tag{2.17}$$

Thus, Eq. (2.13) is easily verifiable if E is less than the energy necessary to produce two real mesons of mass μ by inelastic scattering. To show Eq. (2.14) holds for arbitrary E, note

$$M(\mathbf{p}^2) \geq \tfrac{1}{2}E - (\mathbf{p}^2 + m^2)^{\frac{1}{2}} \tag{2.18}$$

for all $\mathbf{p}$, so Eq. (2.14) follows if

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > (\tfrac{1}{2}E - (\mathbf{n}^2 + m^2)^{\frac{1}{2}}) - (\tfrac{1}{2}E - (\mathbf{p}^2 + m^2)^{\frac{1}{2}}), \tag{2.19a}$$

i.e., if

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > (\mathbf{p}^2 + m^2)^{\frac{1}{2}} - (\mathbf{n}^2 + m^2)^{\frac{1}{2}}, \tag{2.19b}$$

for all $\mathbf{n}^2 < k^2$ and $\mathbf{p}^2 > \mathbf{n}^2$, and Eq. (2.14) certainly follows if Eq. (2.19b) can be shown for arbitrary $\mathbf{n}$, $\mathbf{p}$. Setting $w = |\mathbf{p}|$, $e = |\mathbf{n}|$, then

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > |e - w|, \tag{2.20a}$$

while

$$(\mathbf{p}^2 + m^2)^{\frac{1}{2}} - (\mathbf{n}^2 + m^2)^{\frac{1}{2}} \leq |(w^2 + m^2)^{\frac{1}{2}} - (e^2 + m^2)^{\frac{1}{2}}| = \frac{|w^2 - e^2|}{(w^2 + m^2)^{\frac{1}{2}} + (e^2 + m^2)^{\frac{1}{2}}} \leq \frac{|w^2 - e^2|}{w + e} = |w - e|, \tag{2.20b}$$

verifying Eq. (2.19b). [Inequality (2.19b) has been fully considered in the general context of Fourier transforms and causality by Dyson.[5] The relationship of causality to the Bethe–Salpeter equation will be emphasized below.]

In a manner analogous to Wick's rotation, one can now continue $T(p)$ in Eq. (2.10) to any p_0 which lies in the region swept by counterclockwise contour deformation from the real p_0 axis to $C_{\mathbf{p}^2}$: Call the partially deformed contour $C_{\mathbf{p}}'$ [as shown in Fig. 3(b)]. For each n in Eq. (2.10), deform the η_0 contour from the real axis to a contour $C_{\mathbf{p}n}'$ horizontally displaced from $C_{\mathbf{p}}'$ by $M(\mathbf{n}) - M(\mathbf{p})$ as shown in Fig. 3(b). Finally, Eq. (2.10) with this choice of η_0 contours can be continued in p_0 along $C_{\mathbf{p}}'$ in the manner just considered for continuation, Eq. (2.11), along $C_{\mathbf{p}^2}$. Thus the extension of the Wick rotation in momentum space to all values of E greater than $2m$ is complete.

It is significant to note there are circumstances under which $C_{\mathbf{n}^2}$ will never come infinitesimally close to the poles in $V(p - \eta)$ for p_0 on $C_{\mathbf{p}^2}$. As shown in Fig. 3(c), such separation will occur if

$$[(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} > M(\mathbf{p}^2) + M(\mathbf{n}^2) \tag{2.13b}$$

for all $\mathbf{p}$, $\mathbf{n}$, and thus will occur if $\mu > (\frac{1}{2}E - m) + (\frac{1}{2}E - m) = E - 2m$, or $E < 2m + \mu$. This will be recognized as the energy restriction for purely elastic scattering and the momentum-space transformation in this case corresponds to the transformation of Schwartz and Zemach in configuration space.

The transformed equation reads

$$\psi(\mathbf{p}, p_0) = \delta^3(\mathbf{p} - \mathbf{p}_{1n})\delta(p_0) - \frac{1}{D(p)} \iint_{C_{\mathbf{n}^2}} V(p - \eta)\psi(\eta)\, d\eta_0\, d^3\mathbf{n}, \tag{2.21}$$

valid for all real $\mathbf{p}$ and all p_0 on $C_{\mathbf{p}^2}$ for any value of $E > 2m$.

A. Singularities in the Transformed Equation

The advantage of transforming to the new coordinate surface is that the singularities in $1/D(p)$ and $V(p - \eta)$ which occur in Eq. (1.1) over an unbounded region of real four-space are now confined to a limited region on the new coordinate surface. Singularities in $1/D(p)$ only occur when the contour $C_{\mathbf{p}^2}$ touches one of the poles in $1/D(\mathbf{p}, p_0)$ at

$$p_0 = \pm((\mathbf{p}^2 + m^2)^{\frac{1}{2}} - \tfrac{1}{2}E - i\epsilon)$$

and

$$p_0 = \pm((\mathbf{p}^2 + m^2)^{\frac{1}{2}} + \tfrac{1}{2}E - i\epsilon).$$

But the latter set of poles are never touched by $C_{\mathbf{p}^2}$

[5] F. J. Dyson, Phys. Rev. **110**, 1470 (1958).

(see Fig. 1), while the former are only touched for $\mathbf{p}^2 \leq k^2$ by the "detour" segments of $C_{\mathbf{p}^2}$. Thus, outside the limited region of detour on the transformed coordinate surface, $1/D(p)$ is free of singularity and vanishes as $(1/p)^4$ as $p^2 \to \infty$.

Now consider singularities in η_0 in $V(p-\eta)$ for fixed p_0 on $C_{\mathbf{p}^2}$. As discussed above, poles occur at $r_\pm = p_0 \pm [(\mathbf{p}-n)^2+\mu^2]^{\frac{1}{2}}$. In order for these poles to touch $C_{\mathbf{n}^2}$ at some η_0, it must be that $\mathrm{Im}\,(\eta_0) = \mathrm{Im}\,(p_0) = 0$, i.e., p_0 must be zero ($\mathbf{p}^2 > k^2$) or on a detour section of $C_{\mathbf{p}^2}$ with $\mathbf{p}^2 < k^2$ and η_0 must be similarly restricted. Now for $\mathbf{p}^2 > k^2$, inequality (2.13a) excludes the possibility of $C_{\mathbf{n}^2}$ touching a pole. By symmetry, no poles can be touched if $\mathbf{n}^2 > k^2$. Thus, singularities can only occur in $V(p-\eta)$, when $\mathbf{p}^2 < k^2$, $\mathbf{n}^2 < k^2$, and p_0 and η_0 are on the detour portions of $C_{\mathbf{p}^2}$ and $C_{\mathbf{n}^2}$, respectively. In the limit $p^2 \to \infty$ outside the region of detour, $V(p-\eta)$ vanishes as $(\lambda/\pi^2)\cdot(1/p)^2$ for any fixed η. [It has already been commented that for $E < 2m+\mu$, $V(p-\eta)$ is completely free of singularity on the whole coordinate surface.] The regularity of $1/D(p)$ and $V(p'-\eta')$ away from $\mathbf{p}^2 < k^2$, $\mathrm{Im}\,(p_0) = \mathrm{Im}\,(\eta_0) = 0$ allows one to form an intuitive picture of the transformed equation. Asymptotically, it is the same regular equation considered by Wick for $E < 2m$.

Since for most of the coordinate surface, p_0 and η_0 are purely imaginary, it is useful to write the transformed equation in terms of rotated coordinates

$$p_0' = p_0/i, \quad p' = (\mathbf{p}, p_0'), \quad p'^2 = \mathbf{p}^2 + p_0'^2 \tag{2.22}$$

with analogous definitions for η_0', η', η'^2. Writing $\psi'(p') = i\psi(p)$ and noting $\delta(p_0) = \delta(ip_0') = -i\delta(p_0')$, Eq. (2.21) becomes

$$\psi'(p') = \delta^3(\mathbf{p}-\mathbf{p}_{\mathrm{in}})\delta(p_0') + \frac{1}{D(p')}\iint_{C'_{\mathbf{n}^2}} V(p'-\eta')\psi'(\eta')\,d\eta_0'\,d^3\mathbf{n}, \tag{2.23}$$

with

$$D(p') = (p'^2-k^2)^2 + E^2p_0'^2, \tag{2.24}$$

$$V(p'-\eta') = \lambda/\pi^2[(p'-\eta')^2+\mu^2]^{-1}; \tag{2.25}$$

$C'_{\mathbf{n}^2}$ is $C_{\mathbf{n}^2}$ rotated clockwise by 90°. The complex p' coordinate surface on which Eq. (2.23) applies can be sketched in terms of $|p|$, $\mathrm{Re}\,(p_0')$, and $\mathrm{Im}\,(p_0')$ as shown in Fig. 4.

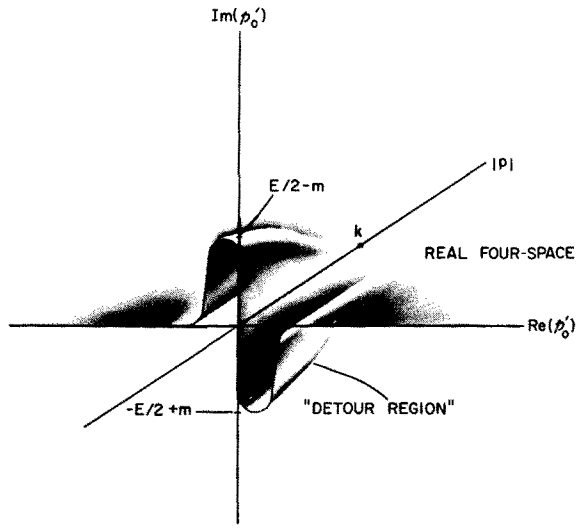


FIG. 4. Transformed coordinate surface.

B. The Wick Frequency Assumption—Causality

It should be noted that the ability to make the contour transformations in complex p_0, η_0, demonstrated above, only depends on the Wick frequency analysis (based on field theory) through the assumption that $\psi(\mathbf{n}, \eta_0)$ in Eq. (1.1) has the structure indicated by Eq. (2.1). But the only assumption really needed is that $\psi(\mathbf{n}, \eta_0)$ be analytic in the region swept by counterclockwise deformation of the η_0 contour from the real axis to $C_{\mathbf{n}^2}$. The demonstrated ability to continue the right-hand side of Eq. (1.1) in complex p_0 throughout the region of contour distortion between the real axis and $C_{\mathbf{p}^2}$, proves that this structure, which was assumed for $\psi(\mathbf{n}, \eta_0)$, although suggested by field theory, is fundamentally consistent with the structure of Eq. (1.1).

One may go further and ask whether the full Wick structure implied by Eq. (2.1) can be shown consistent with the structure of Eq. (1.1) in this way. The above analysis has shown that $\psi_0(\mathbf{p}, p_0)$ and $1/D(p)$ are consistent with Eq. (2.1). It remains to show that

$$T(p) = \iint V(p-\eta)\psi(\eta)\,d^4\eta$$

can be continued throughout the complex p_0 plane up to the cuts described by Eq. (2.1), assuming that $\psi(\mathbf{n}, \eta_0)$ has the cuts described by Eq. (2.1) in complex η_0.

As illustrated in Fig. 3(d), the required continuation, accompanied by the appropriate η_0 contour distortion, can be accomplished if

$$[(\mathbf{p}-\mathbf{n})^2+\mu^2]^{\frac{1}{2}} > |(\mathbf{p}^2+m^2)^{\frac{1}{2}} - (\mathbf{n}^2+m^2)^{\frac{1}{2}}| \tag{2.19c}$$

for all $\mathbf{p}$, $\mathbf{n}$. (One first chooses the path of p_0 continuation $C_{\mathbf{p}}'$, then deforms the η_0 path of integration to $C'_{n\mathbf{p}}$ for each $\mathbf{n}$, and finally continues p_0 along $C_{\mathbf{p}}'$.) But this inequality is equivalent to inequality (2.19b). Thus, not surprisingly, ability to perform the p_0, η_0 contour transformations in Eq. (1.1) for arbitrary E

follows from the consistence of Eq. (1.1) with the structural requirements of Eq. (2.1) for all $\mathbf{p}$. It is significant to note that in the above proof of this consistency, as indicated by Eq. (2.20a), it suffices that V satisfies a simple condition: $V(\mathbf{q}, q_0)$ can only have singularities in q_0 which are at $(r - i\epsilon)$ and/or $-(r - i\epsilon')$, where r is real and $r \geq |\mathbf{q}|$. In configuration space, $r \geq \mathbf{q}$ implies that $V(\mathbf{x}, x_0)$ for $x_0 > 0$ and $x_0 < 0$ is a superposition of waves which move with a velocity less than or equal to 1, the velocity of light. This inequality places a very strong restriction on V since it must be satisfied for all $\mathbf{q}$. On the other hand, as indicated by Eq. (2.16), for the restricted range $E < 2m + 2\mu$, the contour transformation to $C_{\mathbf{p}^2}$ proposed here only requires $r \geq \mu$, i.e., that $V(\mathbf{x}, x_0)$ have x_0 frequencies greater than μ for $x_0 > 0$, and less than $-\mu$ for $x_0 < 0$ ("causality" need only be obeyed for $|\mathbf{q}| \leq \mu$).

One naturally asks whether the cut structure of ψ can be deduced from the Bethe–Salpeter equation alone without reference to field theory. Toward this end, note that if ψ is a solution to the *transformed* Eq. (2.21), then the known separation of the poles in $V(p - \eta)$ from η_0 on $C_{\mathbf{n}^2}$ indicates that

$$T(\mathbf{p}, p_0) = \iint_{C_{\mathbf{n}^2}} V(p - \eta)\psi(\eta)\, d^4\eta,$$

appearing in Eq. (2.21) will be an analytic function of p_0 at least within a strip about $C_{\mathbf{p}^2}$ [see Fig. 3(e)] of half-width

$$\begin{aligned}\Delta_{\mathbf{p}} &= \min_{\mathbf{n}} ([(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} - |M(\mathbf{p}^2) - M(\mathbf{n}^2)|) \\ &\geq \min_{\mathbf{n}^2 \leq k^2, \mathbf{p}^2 \leq k^2} ([(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} \\ &\quad - |(\mathbf{p}^2 + m^2)^{\frac{1}{2}} - (\mathbf{n}^2 + m^2)^{\frac{1}{2}}|) \equiv \Delta > 0.\end{aligned}$$

[A closer inspection shows that

$$\Delta = (k^2 + (m + \mu)^2)^{\frac{1}{2}} - (k^2 + m^2)^{\frac{1}{2}}.]$$

But noting the analytic structure of $1/D(\mathbf{p}, p_0)$, Eq. (2.21) shows that $\psi(\mathbf{p}, p_0)$ can be continued in p_0 (for any $\mathbf{p}$) to the right in the upper-half plane (and/or left in the lower half) through a distance Δ from $C_{\mathbf{p}^2}$ to "$C_{\mathbf{p}^2} + \Delta$," where Δ is independent of $\mathbf{p}$. Then putting p_0 on $C_{\mathbf{p}^2} + \Delta$, and noting that the transformed equation indicates $\psi(\mathbf{p}, p_0) \sim O(1/p_0^2)$ as $p_0/i \to \pm\infty$ on $C_{\mathbf{p}^2}$, one can deform all the η_0 contours in Eq. (2.21) from $C_{\mathbf{n}^2}$ to $C_{\mathbf{n}^2} + \Delta$. Now one can repeat the argument and continue $\psi(\mathbf{p}, p_0)$ to $C_{\mathbf{p}^2} + 2\Delta$, $C_{\mathbf{p}^2} + 3\Delta$, etc.

If, however, one applies this procedure to the left in both upper- and lower-half planes [see Fig. 3(f)], one cannot continue $\psi(\mathbf{p}, p_0)$ through the poles in $1/D(\mathbf{p}, p_0)$. For any $\mathbf{p}^2$ such that these poles are encountered, the shifted contour must detour around them. One thus "discovers" the Wick cuts in $\psi(\mathbf{p}, p_0)$.

As one shifts successively to the left, adding detours at each step [Fig. 3(f)], the length Δ_i through which one can move the nonobstructed parts of all the p_0 contours in the ith shift decreases as i increases. After $i - 1$ shifts, the p_0 contours in the upper-half plane must detour whenever $\mathbf{p}^2 < \mathbf{p}_{i-1}^2$, where

$$\mathbf{p}_{i-1}^2 = \left(E/2 + \sum_{j=1}^{i-1} \Delta_j\right)^2 - m^2 > \mathbf{p}_{i-2}^2 > \cdots > \mathbf{p}_0^2 = k^2.$$

Thus

$$\begin{aligned}\Delta_i = \min_{\mathbf{n}^2 < \mathbf{p}_{i-1}^2, \mathbf{p}^2 < \mathbf{p}_{i-1}^2} &([(\mathbf{p} - \mathbf{n})^2 + \mu^2]^{\frac{1}{2}} \\ &- |(\mathbf{p}^2 + m^2)^{\frac{1}{2}} - (\mathbf{n}^2 + m^2)^{\frac{1}{2}}|) \\ &< \Delta_{i-1} < \cdots < \Delta_1 = \Delta.\end{aligned}$$

[More specifically,

$$\Delta_i = (\mathbf{p}_{i-1}^2 + (m + \mu)^2)^{\frac{1}{2}} - (\mathbf{p}_{i-1}^2 + m^2)^{\frac{1}{2}}.]$$

However, since Δ_i is bounded away from zero for any possible bound on

$$\sum_{j=1}^{i-1} \Delta_j,$$

the process of continuation illustrated in Fig. 3(f) can be extended throughout the left-half plane (and similarly throughout the right one) despite the decreases in Δ_i. Thus, any solution of the transformed equation which has the asymptotic behavior indicated by the equation itself must have the analytic structure indicated by field theory.

One notes again that the full Wick cut structure only results if V possesses the complete causal structure indicated by Eq. (2.19c).

C. Properties of $1/D(p')$—Relationship to Nonrelativistic Theory

As in the case of the nonrelativistic scattering equation, many important features of Eq. (2.23) are determined by properties of the factor $1/D(p')$ [see Eqs. (2.7)–(2.9)] which can be written

$$\begin{aligned}\frac{1}{D(p')} = &\frac{1}{4E(\mathbf{p}^2 + m^2)^{\frac{1}{2}}} \\ &\times \left[\frac{1}{\beta(\mathbf{p})} i\left(\frac{1}{p_0' - i\beta(\mathbf{p})} - \frac{1}{p_0' + i\beta(\mathbf{p})}\right)\right. \\ &\left. - \frac{1}{\alpha(\mathbf{p})} i\left(\frac{1}{p_0' - i\alpha(\mathbf{p})} - \frac{1}{p_0' + i\alpha(\mathbf{p})}\right)\right]. \quad (2.27)\end{aligned}$$

This expression corresponds to the nonrelativistic term $1/(\mathbf{p}^2 - k^2)$, and noting that

$$\alpha(\mathbf{p}) \cdot \beta(\mathbf{p}) = \mathbf{p}^2 - k^2 \quad (2.28)$$

one can write Eq. (2.27) in analogy to the nonrelativistic factor:

$$[D(p')]^{-1} = (\mathbf{p}^2 - k^2)^{-1} \cdot d(\mathbf{p}^2, p_0'), \quad (2.29)$$

where, with $\omega(\mathbf{p}) = (\mathbf{p}^2 + m^2)^{\frac{1}{2}}$,

$$d(\mathbf{p}^2, p_0') = (4E)^{-1}([\alpha(\mathbf{p})/\omega(\mathbf{p})]i\{[p_0' - i\beta(\mathbf{p})]^{-1} - [p_0' + i\beta(\mathbf{p})]^{-1}\} - [\beta(\mathbf{p})/\omega(\mathbf{p})] \times i\{[p_0' - i\alpha(\mathbf{p})]^{-1}[p_0' + i\alpha(\mathbf{p})]^{-1}\}). \quad (2.30)$$

Now note that at $\mathbf{p}^2 = k^2$, the following hold: $\omega = E/2$, $\alpha(\mathbf{p}) = -i\epsilon \to 0$, $\beta(\mathbf{p}) = E - i\epsilon \to E$, and thus

$$d(k^2, p_0') = (4E)^{-1}\{(-i\epsilon/E)2i[(p_0' - i\epsilon)^{-1} - (p_0' + i\epsilon)^{-1}] - 2i[(p_0' - \epsilon)^{-1} - (p_0 + \epsilon)^{-1}]\} \quad (2.31a)$$

$$\to -(i/2E)[(p_0' - \epsilon)^{-1} - (p_0' + \epsilon)^{-1}]. \quad (2.31b)$$

Setting $\epsilon = \epsilon_1 + i\epsilon_2$, $\epsilon_1 \to 0$, and taking C'_{k^2} as the real axis (see Fig. 2),

$$d(k^2, p_0') \to -(i/2E)[(p_0' - i\epsilon_2)^{-1} - (p_0' + i\epsilon_2)^{-1}], \quad (2.32)$$

so that, with arbitrarily small $|\epsilon|$ understood,

$$d(k^2, p_0) = (\pi/E)\delta(p_0'). \quad (2.33)$$

Now using the familiar formula

$$(\mathbf{p}^2 - k^2)^{-1} = P(p^2 - k^2)^{-1} + i\pi\delta(\mathbf{p}^2 - k^2), \quad (2.34)$$

one has[6]

$$\frac{1}{D(p')} = P\frac{1}{\mathbf{p}^2 - k^2} d(\mathbf{p}^2, p_0') + \frac{i\pi^2}{E}\delta(\mathbf{p}^2 - k^2)\delta(p_0'). \quad (2.35)$$

It is easy to check that $d(\mathbf{p}^2, p_0)$ is a real factor in the sense that integrals of the sort $\int C'_{\mathbf{p}^2} U(p_0')\, d(\mathbf{p}^2, p_0')\, dp_0'$ will be real if $U^*(p_0') = U(p_0'^*)$ and U is analytic on $C'_{\mathbf{p}^2}$. Thus Eq. (2.35) effectively splits $[D(p')]^{-1}$ into real and imaginary parts in analogy to Eq. (2.34) for the nonrelativistic case.

Equations (2.29) and (2.30) can be used to obtain the nonrelativistic limit of $[D(p')]^{-1}$. Note one can consider the nonrelativistic limit to correspond to the limit $m \to E/2 \to \infty$ for fixed k, since this implies $mc \gg mv$, i.e., $c \gg v$ where v is the incident particle velocity. But then for any finite $\mathbf{p}^2$,

$$\alpha(\mathbf{p}) = [E/2)^2 + (\mathbf{p}^2 - k^2)]^{\frac{1}{2}} - E/2 \to (\mathbf{p}^2 - k^2)/E \to 0,$$

$$\beta(\mathbf{p}) \to 2m \to E,\ \omega(p) \to m \to E/2,$$

so that, very much in the manner just considered,

$$d(\mathbf{p}^2, p_0) \to \frac{i}{2E}\left(\frac{1}{p_0' - i\alpha(\mathbf{p})} - \frac{1}{p_0' + i\alpha(\mathbf{p})}\right) \quad (2.36)$$

in the limit $E \to \infty$, k fixed. Note that although $\alpha(\mathbf{p})$ becomes a negative infinitesimal for $\mathbf{p}^2 < k^2$, the contour $C'_{\mathbf{p}^2}$ specifically detours for $p^2 < k^2$ so that the poles in Eq. (2.36) at $\pm i\alpha(\mathbf{p})$ stay on the same side of $C'_{\mathbf{p}^2}$ for $\mathbf{p}^2 < k^2$ as for $\mathbf{p}^2 > k^2$ when $\alpha(\mathbf{p}) > 0$. Thus in Eq. (2.36), $\alpha(\mathbf{p})$ can be treated as a positive (real) infinitesimal for all finite $\mathbf{p}^2$, giving

$$d/(\mathbf{p}^2, p_0') \to \pi/E\delta(p_0') \quad (2.37)$$

and

$$1/D(p') \to \frac{\pi}{E}\frac{1}{\mathbf{p}^2 - k^2}\delta(p_0') \quad (2.38)$$

in the nonrelativistic limit $E \to \infty$, k fixed.[7]

Having analyzed the location of the poles in $d(\mathbf{p}^2, p_0)$ in detail, note that Eq. (2.35) and Eq. (2.38) can be derived simply by writing

$$\frac{1}{D(p')} = \frac{1}{(p'^2 - k^2)}\rho(p'^2, p_0),$$

where

$$\rho(p'^2, p_0') = \frac{1}{E}\left[\frac{(p'^2 - k^2)/E}{[(p'^2 - k^2)/E]^2 + p_0'^2}\right]$$

and $\rho(p'^2, p_0') \to \pi/E\delta(p_0')$ for $p'^2 - k^2 = \epsilon_2 \to 0$, or $E \to \infty$ by inspection.

Finally, Eqs. (1.1), (2.29), and (2.33) yield a simple derivation of the scattering amplitude. Setting

$$\psi(x) = \int e^{i(\mathbf{p}\cdot\mathbf{x} - p_0 x_0)}\psi(p)\, d^4p$$

and using Eq. (1.1),

$$\psi(\mathbf{x}, 0) = e^{i\mathbf{p}_{1n}\cdot\mathbf{x}} - i\int e^{i\mathbf{p}\cdot\mathbf{x}}\int\frac{1}{D(p)} \times \int V(p - \eta)\, d^4\eta\, dp_0\, d^3\mathbf{p}. \quad (2.39)$$

Rotating contours in the manner introduced above, using $\psi'(\eta') = i\psi(\eta)$ and inserting Eq. (2.29) for $1/D(p')$,

$$\psi(\mathbf{x}, 0) = e^{i\mathbf{p}_{1n}\mathbf{x}} + \int e^{i\mathbf{p}\cdot\mathbf{x}}\frac{1}{\mathbf{p}^2 - k^2} \cdot \left(\int_{C'_{\mathbf{p}^2}} d(\mathbf{p}^2, p_0') \cdot T'(\mathbf{p}, p_0')\, dp_0'\right) d^3\mathbf{p} \quad (2.40)$$

with

$$T'(p') = \iint_{C'_{\mathbf{n}^2}} V(p' - \eta')\psi'(\eta')\, d\eta_0'\, d^3\mathbf{n} = T(p). \quad (2.41)$$

Now it is well known from studies of nonrelativistic scattering that in the limit $|\mathbf{x}| \to \infty$,

$$\int e^{i\mathbf{p}\cdot\mathbf{x}}\frac{1}{\mathbf{p}^2 - k^2}Q(\mathbf{p})\, d^3\mathbf{p} \to f(\Omega_f)e^{ikR}/R, \quad (2.42)$$

[6] This result corresponds to Schwartz and Zemach's equation (3.11) in Ref. 2.

[7] Again the result agrees with Schwartz and Zemach's equations (4.2) and (4.6).

where $R = |\mathbf{x}|$, f refers to the $\mathbf{x}$ direction, and $f(\Omega_f) = 2\pi^2 Q(\mathbf{p}_f)$ with $\mathbf{p}_f = k\hat{x}$. But the second term on the right of Eq. (2.40) is precisely of the form of the term on the left in Eq. (2.42), so that the usual nonrelativistic analysis yields[8] for $x_0 = 0$, $|\mathbf{x}| \to \infty$

$$\psi(\mathbf{x}, 0) \to e^{ip_{1n}\cdot x} + f(\Omega_f)e^{ikR}/R, \tag{2.43}$$

where

$$\begin{aligned} f(\Omega_f) &= \int_{C'_{k^2}} d(k^2, p_0')T'(\mathbf{p}_f, p_0')\, dp_0' \\ &= \int_{C'_{k^2}} \frac{\pi}{E}\delta(p_0')T'(\mathbf{p}_f, p_0')\, dp_0' \\ &= (2\pi^3/E)T'(\mathbf{p}_f, 0) = (2\pi^3/E)T'(p_f). \end{aligned} \tag{2.44}$$

3. A METHOD OF APPROXIMATE SOLUTION

Since the method to be presented here does not produce highly precise results but does serve as a rough check on other methods, the following discussion is brief.[9] Outside the restricted region of detours in the coordinate surface, Eq. (2.23) indicates that $|\psi'(p')|$ will vanish rapidly as $p' \to \infty$. Specifically, $1/D(p')$ will vanish as $|1/p'|^4$, while

$$|T'(p')| = \int_{C'_{\mathbf{n}^2}} \frac{\lambda}{(p' - \eta')^2 + \mu^2}\, \psi'(\eta')\, d^4\eta'$$

should vanish as $1/|p'|^2$ (assuming the integral over η' is amply convergent). Thus in the limit $|p'| \to \infty$ one expects $|\psi'(p')|$ to vanish as $1/|p'|^6$ (reinforcing the assumption that the integral for T' is amply convergent). This fact suggests that one can obtain good approximations to $\psi'(\mathbf{p}, p_0')$ by solving a cutoff version of Eq. (2.23), i.e., by considering the restriction of Eq. (2.23) to a large sphere of radius $\Lambda > k$ in p', η' four-space. For $E < 2m + \mu$ it has been commented that $V(p' - \eta')$ is never singular for p' and η' on the coordinate surface. Then for p' and η' within a finite cutoff sphere, $V(p' - \eta')$ can be well approximated by a finite sum of separable potentials. Thus in the elastic scattering range one can take

$$V(p' - \eta') \simeq \sum_l \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} V_{ij}^l g_i^l(|\mathbf{p}|, p_0') g_j^l(|\mathbf{n}|, \eta_0') P_l(\hat{\mathbf{p}}) P_l(\hat{\mathbf{n}}), \tag{3.1}$$

where V_{ij}^l are real and g_i are real for real p' and analytic on the whole coordinate surface. But with the right-hand side of (3.1) substituted for V in Eq. (2.23), the new equation is as easy to solve as the nonrelativistic equation with a finite sum of (three dimensions) separable potentials. Using matrix algebra analogous to the nonrelativistic case, setting

$$f(\Omega_f) = \sum f_l P_l(\hat{\mathbf{p}}_f), \quad f_l = \frac{2l+1}{k}\, e^{i\delta_l} \sin\delta_l$$

and with Eq. (2.35) for Im $[1/D(p')]$, one finds

$$\tan\delta_l = \{(2\pi^3 k)/[(2l+1)E]\} G_l^t V^l (I - U^l V^l)^{-1} G_l, \tag{3.2}$$

(with all factors real so elastic unitarity is valid) where matrix notation is used; G is a column matrix:

$$(G_l)_{j,1} = g_j^l(k, 0), \quad I_{ij} = \delta_{ij},$$

and

$$U_{ij}^l = P\int \frac{P_l^2(q) g_i^l(q') g_j^l(q')}{D(q')}\, d^4q'. \tag{3.3}$$

The range of q' integration in Eq. (3.3) is limited to the interior of the cutoff sphere of radius Λ.

One method of getting an approximation to V of form Eq. (3.1) inside a sphere of radius Λ_f is to take a least-square polynomial fit to $(1 + x^2)^{-1}$ on the interval $-2\Lambda_f \le x \le 2\Lambda_f$, i.e.,

$$\frac{1}{[1 + x^2]} \cong \sum_{i=0}^{M} C_i (x^2)^i. \tag{3.4}$$

Then in units such that $\mu = 1$,

$$V(p' - \eta') \cong \frac{\lambda}{\pi^2} \sum_{i=0}^{M} C_i [(p' - \eta')^2]^i$$

for p', η' inside a sphere of radius Λ_f.

Now one can expand

$$(p' - \eta')^{2i} = \sum_{j=0}^{i} \mathbf{p}^j \mathbf{n}^j z^j \sum_{\gamma,s,\tau,q} A^j_{\gamma,s,\tau,q} (p'^2)^\gamma (p_0'^2)^s (\eta_0'^2)^q (\eta'^2)^\tau, \tag{3.5}$$

$$s; q = 0, 1, \cdots, \tfrac{1}{2}[i - j],$$

$$\gamma; \tau = 0, 1, \cdots, i - j - 2s; i - j - 2s,$$

and since

$$z^j = \sum_{l=0}^{j} a_{jl} P_l(z) \to \sum_{l=0}^{j} a_{jl} P_l(\hat{\mathbf{p}}) P_l(\hat{\mathbf{n}}),$$

one has

$$V \cong \sum_{l=0}^{M} P_l(\hat{\mathbf{p}}) \mathbf{p}^l P_l(\hat{\mathbf{n}}) \mathbf{n}^l \sum_{i=1}^{N_{M-l}} \sum_{j=1}^{N_{M-l}} V_{ij}^l g_i(p_0'^2, p_0'^2) g_i(\eta'^2, \eta_0'^2), \tag{3.6}$$

where $g_\gamma(q'^2, q_0'^2)$ belong to the set of functions $(q'^2)^s (q_0'^2)^t$ for $\tau = 0, 1, \cdots, [(M - l)/2]$ and $s = 0, 1, \cdots, M - 1 - 2\tau$. The dimension of V^l is N_{M-l}, where $N_q = ([q/2] + 1)(q + 1 - [q/2])$.

The constants V_{ij}^l can be expressed in terms of the constants C_i, $A^j_{\gamma,s,\tau,q}$, and a_{jl} by simple arithmetic. The required operations for evaluating V_{ij}^l can be

[8] This result has been previously obtained via analysis of the Fourier transform of $1/D$ [see Ref. 2, and A. R. Swift and B. W. Lee, J. Math. Phys. 5, 908 (1964)]. Momentum space arguments allow particular rigor in the elastic range because of the known analyticity of $V(p' - \eta')$ and $T'(p')$ (see Ref. 9).

[9] R. M. Saenger, Ph.D. thesis, Massachusetts Institute of Technology (1966).

carried out rapidly by machine. Also, the integrals U_{ij}^{l} can be done in closed form and expression Eq. (3.2) for $\tan \delta_l$ evaluated by machine.[9]

To investigate convergence as a function of the cutoff Λ, the fit to V was kept fixed with $\Lambda_f = 4.0$ and polynomial degree $M = 10$. The resulting $\tan \delta_l$ was evaluated for several values of Λ up to and including Λ_f. Convergence as a function of M was investigated by setting $\Lambda = \Lambda_f = \Lambda_{\min}$ and varying M up to 10. To standardize this procedure, $\Lambda_{\min}$ was chosen to satisfy $(\Lambda_{\min}^2 - k^2)^2 = E^2\Lambda_{\min}^2$, since then $|p'| = \Lambda_{\min}$ roughly marks the beginning of the asymptotic region of $1/D(p')$.

From the results with variable Λ and fixed M, an estimate was made of the error incurred by cutting off the equation at $\Lambda = \Lambda_{\min}$. The results with $\Lambda = \Lambda_{\min}$ and variable M were used to estimate the precise cutoff result at $\Lambda = \Lambda_{\min}$. Finally, the estimated result at $\Lambda = \Lambda_{\min}$ was combined with the estimated error due to cutoff to give an estimate of the non-approximated result. This scheme was only carried out fully for $m = \mu = 1$, $l = 0$. The results for $\lambda = 1$, $k^2 = 0.4$ are listed in Table I. The estimated value of $E/2k \tan \delta_0$ for Eq. (2.21) cutoff at $\Lambda = \Lambda_{\min} = 2.52$ is 3.4 ± 0.2. The estimated error due to this cutoff is -0.2. Thus, with generous allowance for possible error, one can set $E/2k \tan \delta_0 = 3.6 \pm 0.4$. Schwartz and Zemach find for this same case $E/2k \tan \delta_0 = 3.5640 \pm 0.0002$.

TABLE I. Computed values of $(E/2k) \tan \delta_0$ at $k^2 = 0.4$, $\lambda = 1$, $m = \mu = 1$.

Values for variable Λ with $M = 10$		Values for variable M at $\Lambda = \Lambda_{\min}$	
Λ	$(E/2k) \tan \delta_0$	M	$(E/2k) \tan \delta_0$
$\Lambda_{\min} = 2.52$	3.133	7	3.149
3.0	3.219	8	3.208
3.5	3.266	9	3.268
3.75	3.280	10	3.312
4.0	3.290		
Limit	3.32 ± 0.03	Limit	3.4 ± 0.2

The principal difficulty in the approach used here in that polynomial fits to $(1 + x^2)^{-1}$ are difficult to handle numerically. Thus the degree of convergence was not great, and in some cases it was nonexistent. The method worked best at $k^2 = 0$ where for the least positive bound state λ it yielded 0.765 ± 0.021, whereas Schwartz and Zemach find a value of 0.76222.

4. CONCLUSION

The principal advantage of viewing the Bethe–Salpeter equation in momentum space is that $1/D(p)$ and $V(p - \eta)$ have a simple algebraic form, in sharp contrast to the form of their transforms in configuration space. The only difficulty with these functions in momentum space is the abundance of poles they possess due to the Lorentz metric. By making the modified Wick rotation suggested here one does not eliminate these singularities, but one does restrict their occurrence in the equation to a bounded region. All remaining singularities must be considered carefully in order to arrive at meaningful results. This has only been attempted here for purely elastic scattering $E < 2m + \mu$ where only singularities in $1/D(p)$ must be considered. For higher values of E the task of analysis is more complex. However, the technique of modified rotation, in that it minimizes the added complexity, and emphasizes the simple analytic structure of the equation for all values of E, should provide a useful point of view for further analysis.

ACKNOWLEDGMENTS

The author would like to thank Professor Herman Feshbach for his advice and encouragement, and to thank the Stanford Linear Accelerator Center for giving him the opportunity to put this paper in final form.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Dynamical Content of Differential Geometry

ALVIN MECKLER
University of Maryland, Baltimore County, Catonsville, Maryland

(Received 20 April 1967)

A formulation of the notion of a Cartesian basis for a vector space suitably adapted to the local differential vector space at a space–time point leads to this description of and derivation for the law of motion of a mass point: There exists a class of equivalent 4-dimensional Cartesian frames in which the trajectory is a straight line, the velocity being constant along the trajectory. The lived-in coordinate system, in which the trajectory is in general not a straight line, is in general not a Cartesian coordinate system. The equation for the motion of the point is purely a statement of differential geometry—an integrability condition—and the forces which are responsible for the deviation from constant velocity are immediately derivable from the metric tensor of the non-Cartesian system. The equation of motion and the relations among dynamical variables are more relativistic than Newtonian, but the metric tensor is positive definite; no indefinite metric is needed. A Newtonian equation of motion is an approximation based on small velocity and it covers the general case of motion in a noninertial system in an external field of force. Intrinsic mass is identified as a constant of the motion and, although the relevancy of mass is somewhat limited in this description of the motion of a single point in an immutable force field, it is seen that the effective mass of the point contains a contribution by the potential energy.

THE mathematics of differential forms is well-established in the expression of analytical mechanics. Aptly described as intrinsic tensor analysis, it has been applied to relativity theory.[1] The connotation *intrinsic* suggests the freedom and consequent incisiveness of the mathematical technique, by virtue of which it has been used with great efficiency in the presentation of Hamiltonian theory in mechanics and in the restatement of electromagnetic and relativistic equations. In this paper, the theory of differential forms—exterior calculus—is used from the start in the formulation of kinematics, which is the differential geometry of space–time. The beginning is a trajectory in 3-dimensional space. A trajectory is a flow line through tangent vectors attached to each point of the 3-dimensional space. A tangent vector is a vector in a tangent space attached to each point of the 3-dimensional space. It is easy enough to recall how it has been done. In 3-dimensional space, one sets up, following Gibbs, three-unit orthogonal vectors (i, j, k) at each point (x, y, z). A tangent vector is a linear combination $(ai + bj + ck)$. If the coefficients (a, b, c) are functions of (x, y, z), then one has a vector field over the 3-dimensional space.

Now it is a matter of experience that a richer treatment of kinematics results from the introduction of time as a fourth coordinate. (Originally, it was simply the parameter to describe the 1-dimensional trajectory.) Given, then, a 4-dimensional space, how does one set up the tangent space? The answer to that question is the theory of differential forms and the development of exterior calculus. The (i, j, k) system is then seen to be an abridged version of a more general description of local structure. A naive extension of it would involve an overcommitment. It turns out to be rather unappealing to establish a tangent space at each space–time point based on a tetrad of vectors (i_x, i_y, i_z, i_t), orthogonal or oblique. It is better to re-examine the notion of a Cartesian coordinate system and then to show that there does exist a 4-dimensional coordinate system, not (x, t), which can be established as Cartesian. In the Cartesian coordinate system, there is no dynamics beyond the kinematics: the trajectory tangent vector has unchanged components along the Cartesian axes from point to point. In the 4-dimensional Cartesian system the trajectory is a straight line. In the (x, t) system, a non-Cartesian system, the trajectory is described as curved—the deviation from straightness being attributed to forces, the forces now clearly ascribable to the geometry. It remains true that there is no dynamics beyond the kinematics.

Since the notion of Cartesian equivalence is basic to this theory, it seems best to review that part of the theory of linear vector spaces which leads to the notion. This is done in the first section. Following that, the theory of differential forms and exterior calculus, as used in this paper, is developed. By this means the subject of kinematics is taken up.

[1] C. W. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) 2, 525 (1957); reprinted in J. A. Wheeler, *Geometrodynamics* (Academic Press Inc., New York, 1962). There are many references to the literature of modern differential geometry and differential forms in the paper. Later publications are: H. Flanders, *Differential Forms* (Academic Press Inc., New York, 1963); and S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Inc., New York, 1964). A comprehensive treatment of classical mechanics in the framework of differential forms without the explicit use of exterior calculus is given in L. A. Pars, *A Treatise on Analytical Dynamics* (John Wiley & Sons, Inc., New York, 1965).

I. LINEAR VECTOR SPACES

A. Definitions, Bases

An element of a linear vector space[2] is denoted by $|x)$. Here the essential symbolism is the semibracket $|\ )$. The letter inside is merely the tag, the name, for the particular vector. No arithmetic is to be done inside the semibracket.[3] The addition of two vectors is a vector. Addition is commutative and associative. There exists a unique vector, the identity element under addition, called the null vector and denoted simply by 0. To each vector there is associated a unique vector—its negative—and the sum of a vector and its negative is the null vector.

The vector space is defined over a field of scalars which, for this paper, need be no more than the field of real numbers. If α is a scalar, $\alpha\,|x)$ is a vector, and

$$\alpha[\beta\,|x)] = (\alpha\beta)\,|x),$$
$$1\,|x) = |x),$$
$$\alpha[|x) + |y)] = \alpha\,|x) + \alpha\,|y),$$
$$(\alpha + \beta)\,|x) = \alpha\,|x) + \beta\,|x).$$

A sum $\sum_i \alpha_i\,|x_i)$ with not all $\alpha_i = 0$ is called a linear combination of the set of vectors labeled as $|x_i)$. A finite set of vectors is termed linearly dependent if there exists a linear combination equal to the null vector. If such a null linear combination does not exist, the set is termed linearly independent.

If every vector of the space can be expressed as a linear combination of a certain set, then that set is said to span the space. A basis is a linearly independent set of nonnull vectors which spans the space: If $|e^i)$ is a typical member of a basis, then $|x) = \sum_i |e^i)\alpha_i$ for every vector $|x)$ and, by definition of a basis, the (α_i)'s are uniquely determined by $|x)$. The scalar α_i is called a component of the vector $|x)$ in the particular basis.

Given one basis, other bases can be constructed. In fact, every linearly independent set of vectors may be extended into a basis by augmenting the set with other vectors of the space. All bases, it then follows, have the same number of elements. This number is called the dimension of the space. In this paper, the relevant vector spaces have finite dimension.

Let $|e^i)$ be a member of a basis. Any vector $|x) = \sum_i |e^i)\alpha_i$. A notation is now introduced which indicates explicitly that a component is a function of the particular basis vector and the particular vector $|x)$:

$$\alpha_i = (e_i\,|\,x),$$

so that

$$|x) = \sum_i |e^i)(e_i\,|\,x).$$

Let the summation convention now be adopted wherein each repeated index in an expression is to be summed over the range of the index. Thus

$$|x) = |e^i)(e_i\,|\,x).$$

Now consider the case of an $|f^j)$ which is of another basis:

$$|f^j) = |e^i)(e_i\,|f^j).$$

The bracket of two basis vectors is called a transformation amplitude. In the above equation, one may expand $|e^i)$ in terms of $|f^k)$:

$$|f^j) = |f^k)(f_k\,|\,e^i)(e_i\,|\,f^j), \tag{1}$$

which forces the conclusion

$$(f_i\,|\,e^k)(e_k\,|f^j) = \delta_i^j, \tag{2}$$

with $\delta_i^j = 0$ for $i \neq j$ and $\delta_i^j = 1$ for $i = j$.

The particularization of Eq. (1) to the statement

$$|e^j) = |e^i)(e_i\,|\,e^j)$$

reveals that

$$(e_i\,|\,e^j) = \delta_i^j. \tag{3}$$

The transformation amplitude $(f_i\,|\,e^j)$ serves to connect the components of a vector in one basis to its components in another:

$$(f_i\,|\,x) = (f_i\,|\,e^j)(e_j\,|\,x).$$

The assignment of the scalar $(e_i\,|\,x)$ to every vector of the space is an example of a linear functional. Instead of the common functional notation $y(x)$, the notation $(y\,|\,x)$ is used to denote the value of the functional at the vector $|x)$. A linear functional on a vector space is an assignment of a scalar to every vector with the qualification that, if $|z) = \alpha\,|u) + \beta\,|v)$, then $(y\,|\,z) = \alpha(y\,|\,u) + \beta(y\,|\,v)$. A linear functional is completely determined by its values on a basis. That is, if $(y\,|\,e^i) = \alpha^i$ and $|x) = |e^i)\beta_i$, then $(y\,|\,x) = \alpha^i\beta_i$. Thus the linear functional $(e_i\,|\,x)$ is completely determined by Eq. (3).

On the set of linear functionals on a vector space, addition and scalar multiplication may be defined by means of the condition that the value of a linear combination of functional is the linear combination of values. The null linear functional is the assignment of zero to every vector of the space. In this way, the set of linear functionals on a vector space itself becomes a vector space. The elements of this vector

[2] The treatment here, as well as that of multilinear functions later on, follows that of P. R. Halmos, *Finite Dimensional Vector Spaces* (Princeton University Press, Princeton, New Jersey, 1942).

[3] The notation is based on that of P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, London, 1947).

space, called the dual vector space, are denoted by $(u|$, where the value of the functional $(u|$ at a point $|x)$ is $(u \mid x)$. A basis is evident as the set of $(e_i|$ for the following.

(*i*) *The set is linearly independent:* if $\alpha^i(e_i \mid x) = 0$ for all $|x)$, then, for $|x) = |e^j)$,

$$0 = \alpha^i(e_i \mid e^j) = \alpha^j.$$

(*ii*) *The set spans the space*, for

$$(y \mid x) = (y \mid e^j)(e_j \mid x) = \alpha^j(e_j \mid x).$$

The two bases $|e)$ and $(e|$ are called dual bases; corresponding basis vectors are given the same letter tag with superior indices for the one case and inferior indices for the other. Any vector of the dual space can be written as $(x| = (x \mid e^i)(e_i|$. Here the bracket notation has been extended to the dual space. The extension of the bracket symbol allows the definition of the bracket of two arbitrary vectors, one from each space:

$$(x \mid y) = (x \mid e^i)(e_i \mid y).$$

This definition is independent of the basis, for, by Eqs. (1) and (2),

$$\begin{aligned}(x \mid y) &= (x \mid e^i)(e_i \mid y),\\ &= (x \mid f^j)(f_j \mid e^i)(e_i \mid f^k)(f_k \mid y),\\ &= (x \mid f^j)(f_j \mid y).\end{aligned}$$

Duality is reciprocal. One may identify the dual of the dual as the original space.

Previously, a theorem was quoted to the effect that a linearly independent set could be extended into a basis. The realization of the dual space permits an explicit algebraic procedure for this extension. In fact, one can extend a set in an infinite variety of ways. The idea is to extend the set while simultaneously constructing the dual basis. The sole condition is to ensure that $(f_i \mid f^j) = \delta_i^j$ on the new bases. For example, let

$$|y^1) = |e^1) - |e^2)$$

and

$$|y^2) = |e^2) + |e^3)$$

in a 4-dimensional space with $|e)$ as a basis. Thus $|y^1)$ and $|y^2)$ are to be two basis vectors of a new basis. Choose

$$(y_1| = (e_1|$$

and

$$(y_2| = (e_3|.$$

This choice guarantees that $(y_1 \mid y^1) = 1$, $(y_1 \mid y^2) = 0$, $(y_2 \mid y^1) = 0$, and $(y_2 \mid y^2) = 1$. Now choose

$$|y^3) = |e^3)$$

and

$$(y_3| = (e_2| + (e_1| - (e_3|.$$

A simple choice for $|y^4)$ is

$$|y^4) = |e^4),$$

and thus

$$(y_4| = (e_4|.$$

It is true that $(y_i \mid y^j) = \delta_i^j$. The problem could have been solved by setting $(y_1| = (e_3| - (e_2|$, $(y_2| = (e_2|$, and so on. Any choice would do, provided that $(y_i \mid y^j) = \delta_i^j$. The procedure works because, if $(y_i \mid y^j) = \delta_i^j$, the set $|y^i)$ is linearly independent as is the set $(y_i|$: if

$$|y^j)\alpha_j = 0,$$

then

$$0 = (y_i \mid y^j)\alpha_j = \alpha_i.$$

For an n-dimensional space, the sets $|y^i)$ and $(y_i|$ need to contain n vectors each to complete the conditions for a basis.

B. The Cartesian Identification

A vector space and its dual are related to no more nor no less an extent than that defined by the condition $(e_i \mid e^j) = \delta_i^j$. No correspondence need be established between vectors of the two spaces beyond $(e_i \mid e^j) = \delta_i^j$. However, there does exist a particular correspondence which leads to the notion of an inner product and the notion of a Cartesian set of axes.

One associates a vector of the space with a vector of the dual in a one-to-one fashion in this manner: If $|x) = |e^i)(e_i \mid x)$, the Cartesian adjoint of $|x)$ is defined to be the vector in the dual space $\langle x| = (e_i \mid x)(e_i|$. An *angular* semibracket is written to indicate the special relationship. The process is termed a *Cartesian identification*, and it is established on a particular basis.

Given two vectors $|x)$ and $|y)$, there is the bracket

$$\begin{aligned}\langle x \mid y\rangle &= \langle x \mid y),\\ &= (e_i \mid x)(e_i \mid y),\\ &= \langle y \mid x\rangle,\end{aligned}$$

which can be thought of as a bilinear function on the vector space. The two properties

$$\langle x \mid y\rangle = \langle y \mid x\rangle$$

and

$$\langle x \mid x\rangle = (e_i \mid x)(e_i \mid x) \geq 0$$

give rise to the Schwartz inequality:

$$(\langle x \mid y\rangle)^2 \leq \langle x \mid x\rangle\langle y \mid y\rangle.$$

The bracket $\langle x \mid y\rangle$ is called the inner product of the

two vectors. The inner product, the Cartesian identification, permits the discussion of metric properties of the vector space. The requirements of a distance function are met by the inner product: if $|z) = |x) - |y)$, then the distance between $|x)$ and $|y)$ may be defined as $\langle z \mid z)^{\frac{1}{2}}$. The length of a vector may be defined as the distance between it and the null vector.

It is of interest to relate the components of a Cartesian adjoint of a vector in an arbitrary basis to the components of the vector in that basis. Let $|g)$ be a basis for a Cartesian identification. Let $|e)$ be any basis. For any vector $|x)$,

$$\langle x| = (g_m \mid x)(g_m|,$$

so that

$$\begin{aligned}\langle x \mid e^i\rangle &= (g_m \mid x)(g_m \mid e^i),\\ &= (g_m \mid e^j)(e_j \mid x)(g_m \mid e^i),\\ &= g^{ij}(e_j \mid x),\end{aligned}$$

with

$$g^{ij} = g^{ji} = (g_m \mid e^i)(g_m \mid e^j).$$

Similarly,

$$(e_i \mid x) = \langle x \mid e^j\rangle g_{ji}$$

with

$$g_{ij} = g_{ji} = (e_i \mid g^m)(e_j \mid g^m).$$

Note that g^{ij} and g_{ij} are elements of inverse matrices:

$$\begin{aligned}g^{ij}g_{jk} &= (gm \mid e^i)(g_m \mid e^j)(e_j \mid g^n)(e_k \mid g^n),\\ &= (g_m \mid e^i)(e_k \mid g^m),\\ &= \delta^i_k.\end{aligned}$$

The arrays $\|g^{ij}\|$ and $\|g_{ij}\|$ are called the metric tensors. They connect the components in a non-Cartesian basis to the components of the adjoint. To go further,

$$\langle e^i \mid e^j\rangle = g^{jm}(e_m \mid e^i) = g^{ji},$$

and, if the Cartesian identification had been made out of the dual space, then

$$\langle e_i \mid e_j\rangle = g_{ij}.$$

Also

$$\langle x \mid y\rangle = \langle x \mid e^i\rangle\langle y \mid e^j\rangle g_{ij}.$$

Bases can be grouped into classes of Cartesian equivalent bases. Let two bases be termed Cartesian equivalent if

$$(e_i \mid f^j) = (f_j \mid e^i).$$

An equivalence relation is so defined for

(i) $(e_i \mid e^j) = \delta^j_i$
$= (e_j \mid e^i)$. REFLEXIVITY

(ii) If $(e_i \mid f^j) = (f_j \mid e^i)$, then
$(f_i \mid e^j) = (e_j \mid f^i)$. SYMMETRY

(iii) If $(e_i \mid f^j) = (f_j \mid e^i)$ and
$(f_i \mid g^j) = (g_j \mid f^i)$, then
$$\begin{aligned}(e_i \mid g^j) &= (e_i \mid f^k)(f_k \mid g^j),\\ &= (g_j \mid f^k)(f_k \mid e^i),\\ &= (g_j \mid e^i).\end{aligned}$$ TRANSITIVITY

If $|e)$ and $|f)$ are Cartesian equivalent, then the metric tensor of the $|e)$ basis with $|f)$ identified as Cartesian is

$$\begin{aligned}g^{ij} &= (f_m \mid e^i)(f_m \mid e^j),\\ &= (e_i \mid f^m)(f_m \mid e^j),\\ &= \delta^j_i.\end{aligned}$$

The metric tensor is unity. Conversely, if the metric tensor is unity, then

$$(f_j \mid e^i) = g^{ik}(e_k \mid f^j) = (e_i \mid f^j).$$

The metric tensor of any other basis is independent of the choice of Cartesian equivalent bases: If $|e)$ and $|f)$ are Cartesian equivalent, then

$$\begin{aligned}(e_m \mid h^i)(e_m \mid h^j) &= (e_m \mid f^k)(f_k \mid h^i)(e_m \mid h^j),\\ &= (f_k \mid e^m)(e_m \mid h^j)(f_k \mid h^i),\\ &= (f_k \mid h^i)(f_k \mid h^j).\end{aligned}$$

For an element of a Cartesian basis,

$$\langle g^i| = (g_i|$$

and thus $\langle g^i \mid g^j\rangle = \delta^{ij}$. The basis, identified with its dual via the Cartesian identification, is the model of an orthogonal system of unit vectors. Any Cartesian equivalent basis forms a similar orthonormal framework.

I. DIFFERENTIAL FORMS

A. Tangent and Cotangent Space

Coordinate space is simply the space of ordered n-tuples $(x^1, x^2, \cdots, x^n)$, where the x^i are real numbers. Enough of a metric is introduced to allow the definition of the derivative. Consider the set of all real-valued functions of the n real variables, which are as continuous and differentiable as are needed. (At any particular point, each function considered shall have a Taylor series convergent in some neighborhood of the point.) Let two functions be termed equivalent at a point in the coordinate space if their first-order partial derivatives are equal, one for one, at that point,

$$\partial f/\partial x^i = \partial g/\partial x^i$$

at the point $(x^1_0, x^2_0, \cdots, x^n_0)$. This condition certainly defines an equivalence relation. Let the class of all functions equivalent to the function f at a point be labeled as $|df)$. The set of equivalence classes can be

made into a vector space: Let the sum of two classes be the class of the sum of a function from one class and a function from the other. More generally, let

$$\alpha\,|df) + \beta\,|dg) = |d[\alpha f + \beta g]).$$

This definition of linear combination is independent of the choice of functions taken from the classes. The null class is the class of all functions whose first-order partial derivatives vanish at the point. A basis can be found, for, by Taylor's theorem,

$$f(x) = f(x_0) + \sum_i \frac{\partial f}{\partial x^i}\bigg|_0 (x^i - x_0^i) + \cdots,$$

so that

$$|df) = |dx^i)(\partial f/\partial x^i),$$

where $|dx^i)$ is the class of all functions for which the ith derivative is unity and all others are zero at the point—the class of the function $x^i - x_0^i$. This vector space is called the cotangent space at the point and it is n-dimensional in an n-dimensional coordinate space.

The basis dual to $|dx)$ is written, in accord with the convention of the previous section, as $(dx|$. That is,

$$(dx_i \mid dx^j) = \delta_i^j.$$

Furthermore,

$$(dx_i \mid df) = \partial f/\partial x^i.$$

The dual space is called the tangent space at that point. A vector of the cotangent space, a linear combination

$$|u) = |dx^i)u_i,$$

is called a linear differential form. A vector of the tangent space retains the title of vector.

Consider now a coordinate transformation viewed as either a mapping of the coordinate space onto itself or a relabeling of the point via the n functions $u^i = u^i(x)$. The transformation is to be one-to-one: no two distinguishable points are to be transformed into a single point. This means that the set of $|du^i)$ is not dependent. If it were, one could select an independent subset, say $(|du^1), \cdots, |du^k))$, and one could certainly find a vector $(\delta x| = \delta x^j\,(dx_j|$ such that

$$(\delta x \mid du^j) = 0 \quad \text{for} \quad j = 1, \cdots, k.$$

It would then be true, because of the dependence, that $(\delta x \mid du^j) = 0$ for $j = 1, \cdots, n$. Thus, in a neighborhood of the point,

$$\delta u^i = \delta x^j(\partial u^i/\partial x^j) = 0,$$

contrary to the stipulation that the coordinate transformation be one-to-one. Therefore, for a one-to-one coordinate transformation, the set of $|du^i)$ is a linearly independent set and can be chosen as a basis. The consequent expansion,

$$|dx^i) = |du^j)(du_j \mid dx^i),$$

indicates the local behavior of the x^i as functions of the u:

$$\partial x^i/\partial u^j = (du_j \mid dx^i).$$

The transformation amplitude relations

$$(dx_i \mid du^k)(du_k \mid dx^j) = \delta_i^j$$

and

$$(dx_i \mid df) = (dx_i \mid du^j)(du_j \mid df)$$

are seen to be the familiar statements

$$\frac{\partial u^k}{\partial x_i}\frac{\partial x^j}{\partial u^k} = \delta_i^j$$

and

$$\frac{\partial f}{\partial x^i} = \frac{\partial u^j}{\partial x^i}\frac{\partial f}{\partial u^j}.$$

B. Exterior Algebra

Although differential forms were introduced via the classes

$$|df) = |dx^i)(\partial f/\partial x^i),$$

it is not true that every differential form $|v)$ can be expressed as $|v) = |df)$ for some class of functions. Alternatively, while it is true that

$$(dx_i \mid df) = \partial f/\partial x^i,$$

it is not true that, for any basis $(u|$, $(u_i \mid dt)$ is the partial derivative of f with respect to some variable. The testing of either $|v) = |df)$ or $(u_i \mid df) = \partial_i f$ would involve successive partial derivatives, a multiplicative process. To handle such questions it is necessary to build a hierarchy of higher forms and vectors. This can be done by the consideration of multilinear functionals on a vector space.

A multilinear functional on a vector space of degree k is a scalar-valued function of k vectors which is linear in each of the arguments. A linear functional is of degree *one*. As with linear functionals, the set of all multilinear functionals of degree k can be made into a vector space.

An alternating multilinear functional of degree k has the added property that it vanishes if two of the arguments are equal. It then follows that it changes sign if two of the arguments are interchanged. It follows also that it vanishes on a dependent set. The alternating property does not affect either the linearity or the feasibility of construction of a vector space of alternating multilinear functionals.

Let $|e)$ be a basis for the vector space. A selection of

k basis vectors as arguments of a k-degree alternating functional is nontrivial only if the vectors be different. The order may as well be a standard one. A block index notation can be used whereby $|e^A)$ indicates a selection of the k basis vectors; the numerical word A is formed of the k different single-index numerals arranged in ascending order from left to right. The manipulation of block indices is facilitated by the use of the symbol ϵ_A^B, which is zero—unless the index A, as a single properly ordered word, is equal to the index B, a single properly ordered word. For equal words, the symbol has the value unity. The symbol ϵ_{AB}^{CD} is zero—unless, while the letters in A and B are distinct and the letters in C and D are distinct, the single word AB is equal to some permutation of the word CD. For an even permutation, the value is $+1$ for an odd permutation the value is -1. The definitions allow the theorems:

$$\epsilon_{DE}^{A}\epsilon_{A}^{BC} = \epsilon_{DE}^{BC},$$
$$\epsilon_{A}^{BC}\epsilon_{C}^{DE} = \epsilon_{A}^{BDE}.$$

An alternating multilinear functional of degree k is defined everywhere if it is defined on the set $|e^A)$, for

$$(u \mid x^2, x^2 \cdots) = (u \mid e^A)\epsilon_A^{i,j\cdots}(e_i \mid x^1)(e_j \mid x^2)\cdots.$$

Therefore, a basis for the vector space of k-degree alternating functionals is the set $(e_A|$ where

$$(e_A \mid e^B) = \epsilon_A^B.$$

The space has the dimension $n!/k!\,(n-k)!$, where n is the dimension of the vector space. Clearly, a dual basis is the set $|e^A)$.

Thus, from a vector space of dimension n with basis of $|e^i)$ and with dual basis of $(e_i|$, a ladder of vector spaces can be constructed for $0 \leq k \leq n$, with each being of dimension $n!/k!\,(n-k)!$ and with bases and duals represented by $|e^A)$ and $(e_A|$. (If $k = 0$, the vector space is the space of scalars. In no case is k greater than n because no linearly independent set can contain more than n vectors.) A total vector space can be taken as the direct sum of all the vector spaces.

In the context of differential forms, the scalars, 1-forms, 2-forms, up to n-forms, are in correspondence with scalars, 1-vectors, 2-vectors, and so on. The process of the creation of higher vectors and forms can be thought of and handled as a process of multiplication—*exterior multiplication*. The exterior product of two basis 1-forms $|dx^i)$ and $|dx^j)$ is defined by

$$|dx^i\,dx^j) = |dx^A)\epsilon_A^{ij}.$$

This being done, the exterior product of a p-form basis vector and a q-form basis vector is defined as the $(p+q)$-form

$$|dx^A)\,|dx^B) = |dx^A\,dx^B) = |dx^C)\epsilon_C^{AB}.$$

The exterior product is to be linear in all the factors so that the exterior product of a p-form and a q-form is the $(p+q)$-form:

$$|u)\,|v) = |uv) = |dx^A)\epsilon_A^{BC}(dx_B \mid u)(dx_C \mid v).$$

Similarly, a p-vector and a q-vector can be multiplied to give a $(p+q)$-vector:

$$(u|\,(v| = (uv| = (u \mid dx^A)(v \mid dx^B)\epsilon_{AB}^{C}(dx_C|.$$

How do higher forms transform under a coordinate transformation? Let $u^i = u^i(x)$ be the new coordinates. Then

$$|du^i\,du^j) = |dx^A)\epsilon_A^{mn}(dx_m \mid du^i)(dx_n \mid du^j).$$

But

$$|du^i\,du^j) = |du^A)\epsilon_A^{ij}.$$

Therefore,

$$(dx_B \mid du^A)\epsilon_A^{ij} = \epsilon_B^{mn}(dx_m \mid du^i)(dx_n \mid du^j). \quad (4)$$

Going up the ladder, for n-forms, therefore,

$$|du) = |dx)(dx \mid du)$$

(where no index symbol is used to represent the word 1, 2, 3, $\cdots$, n). The transformation amplitude $(dx \mid du)$ is called the determinant of the matrix $\|(dx_j \mid du^i)\|$.

Now: higher forms can be built out of lower forms, so that

$$|du^A\,du^B) = |du)\epsilon^{AB},$$

where A is a k-word, B is an $(n-k)$-word. Therefore,

$$(dx \mid du)\epsilon^{AB} = (dx \mid du^A\,du^B)$$
$$= \epsilon^{CD}(dx_C \mid du^A)(dx_D \mid du^B). \quad (5)$$

The matrix $\|(dx_B \mid du^A)\|$, where A and B are k-words, is called the kth compound of the matrix $\|(dx_j \mid du^i)\|$. Equations (4) and (5) indicate that the elements of the kth compound are the determinants formed from submatrices of order k. Equation (5) is an expression of the Laplace expansion of a determinant.

C. Exterior Derivation

The 1-form $|df)$ can be thought of as the creation of a 1-form from a 0-form f by a process of derivation. Based on

$$|df) = |dx^i)(\partial f/\partial x^i), \quad (6)$$

the exterior derivative of a k-form can be defined recursively as a linear operation which transforms a k-form into a $(k+1)$-form via

$$d\,|u\,dv) = |du\,dv).$$

The definition implies that $d\,|dv) = 0$, setting $u = 1$. Because of linearity and Eq. (6), it follows that, in any coordinate system,

$$\begin{aligned} d\,|u) &= |du) = d[|dx^A)(dx_A \mid u)], \\ &= |dx^i\, dx^A)\frac{\partial}{\partial x^i}(dx_A \mid u), \\ &= |dx^B)\epsilon_B^{iA}\frac{\partial}{\partial x^i}(dx_A \mid u). \end{aligned}$$

If $|u)$ is a p-form and $|v)$ a q-form, then

$$|uv) = (-1)^{pq}\,|vu)$$

and

$$d\,|uv) = |du)v + (-1)^{pq}\,|dv)\,|u).$$

The exterior derivative of a 0-form, a scalar function, is a 1-form; but not every 1-form is the exterior derivative of a function. If $|u) = |df)$, then certainly it is necessary that $d\,|u) = 0$. This is also a sufficient condition: If, in the neighborhood of a point at which it does not vanish identically, $d\,|u) = 0$ for a 1-form $|u)$, then there exists a function f such that $|u) = |df)$. The proof presented here consists of the construction of the Taylor series for the function f. Let the point in question be taken as the origin and let

$$|u) = |dx^i)u_i.$$

The condition $|du) = 0$ becomes

$$\begin{aligned} 0 &= |dx^i\, du_i), \\ &= |dx^i\, dx^j)(dx_j \mid du_i), \end{aligned}$$

which implies that

$$(dx_j \mid du_i) = (dx_i \mid du_j).$$

Let

$$u_{ij} = (dx_j \mid du_i) = u_{ji},$$

so that

$$|du_i) = |dx^j)u_{ji}.$$

Taking the derivative,

$$0 = |dx^j\, dx^k)(dx_k \mid du_{ji}).$$

This implies that

$$(dx_k \mid du_{ji}) = (dx_j \mid du_{ki}),$$

which, when combined with the previously proven symmetry, demands that

$$u_{ijk} = (dx_i \mid du_{jk})$$

be fully symmetric in all its indices. Clearly, one goes on to construct

$$u_{ij\cdots n} = (dx_i \mid du_{j\cdots n}),$$

which is fully symmetric.

Now, set

$$f(x) = x^i u_i\,|_0 + \frac{x^i x^j}{2!}u_{ij}\,|_0 + \cdots + \frac{x^i x^j \cdots x^n}{n!}u_{ij\cdots n}\,|_0 + \cdots,$$

where $u_{ij\cdots n}|_0$ is $u_{ij\cdots n}$ evaluated at the origin. Taking the derivative, we find that

$$|df) = |dx^i)u_i\,|_0 + |dx^i)x^j u_{ij}\,|_0 + \cdots + \frac{|dx^i)x^j \cdots x^n}{(n-1)!}u_{ij\cdots n}\,|_0 + \cdots.$$

The full symmetry of $u_{ij\cdots n}$ has been of essential use. Going on we find that

$$|df) = |dx^i)\Big[u_i\,|_0 + x^1 u_{ji}\,|_0 + \cdots + \frac{x^j \cdots x^n}{(n-1)!}u_{ij\cdots n}\,|_0 + \cdots\Big].$$

But the series in the bracket is $u_i(x)$. Therefore,

$$|df) = |dx^i)u_i = |u).$$

Of course, f is indeterminate to an additional constant.

D. Surfaces

An m-dimensional surface in an n-dimensional coordinate space, with $m < n$, is the mapping of an m-dimensional space into the n-dimensional space via $x^i = x^i(t^1, t^2, \cdots, t^m)$. The parameters $(t^1, \cdots, t^m)$ are the coordinate variables on the surface. Associated with this mapping is the mapping of the differential forms:

$$\begin{aligned} |dx^i) &= |dt^j)(dt_j \mid dx^i), \\ &= |dt^j)(\partial x^i/\partial t^j). \end{aligned} \tag{7}$$

The use of the equality sign in the above relation requires some explanation or qualification. Clearly, in the m-dimensional cotangent space spanned by the $|dt^j)$, the full set of n $|dx^i)$ cannot be independent. There must be $(n - m)$ independent combinations of the $|dx^i)$ which vanish, and this contradicts the fundamental fact that the $|dx^i)$'s form an independent set. The matter can be resolved by the concept of a quotient space.

Let $|z)$ be a vector in a vector space. Two vectors of the space are defined to be equivalent if their difference is a multiple of $|z)$. The set of classes resulting from this equivalence relation can easily be made into a vector space. So easily can this be done that it is not worth while to change the notation; thus the new vectors are usually represented by the same symbols as the vectors of the original space. The original

vector space has been reduced by the imposition of addition modulo $|z)$. In effect, $|z) = 0$. The new vector space is called the quotient space of the original with respect to $|z)$. The procedure can be generalized for the case of a quotient space with respect to the vector space spanned by a set of vectors. Addition is modulo any linear combination of the selected vectors. In effect, each vector selected is equal to the null vector.

A more cautious way, then, of defining a surface would be to have a proper coordinate transformation $x^i = x^i(t^1, \cdots, t^n)$, and then to define the surface cotangent space as the quotient space with respect to the set $|dt^{m+1}), \cdots, |dt^n)$. Equation (7) would be written as

$$|dx^i) = |dt^j)(dt_j \mid dx^i) \quad \text{and} \quad \text{mod } |dt^{m+1}), \cdots, |dt^n).$$

The apparent linear dependence would be resolved by statements such as

$$|dx^j)a_j = 0 \text{ mod } |dt^{m+1}), \cdots, |dt^n).$$

This is the same as saying that

$$|dx^j)a_j = |dt^{m+1})b_1 + \cdots + |dt^n)b_n.$$

One need not take so cautious an approach. Given only the limited coordinate transformation $x^i = x^i(t^1, \cdots, t^m)$, one can find the unmentioned functions $t^{m+1}, \cdots, t^n$.

The $n - m$ differential forms which vanish in the quotient space lie in the extension of the set of $|dt^i)$, $i = 1, \cdots, m$. That is, if $|u^k)$ be such a form, then

$$(dt_i \mid u^k) = 0 \quad \text{and} \quad i = 1, \cdots, m, \tag{8}$$

where $(dt_i|$ is calculable as

$$(dt_i| = (dt_i \mid dx^j)(dx_j|.$$

Now,

$$(dt_i\, dt_j \mid du^k) = (dt_i \mid d(dt_j \mid u^k)) - (dt_j \mid d(dt_i \mid u^k)),$$

with both i and j in the range $(1, \cdots, m)$. Because of Eq. (8), therefore,

$$(dt_i\, dt_j \mid du^k) = 0, \quad \text{for} \quad i, j = 1, \cdots, m.$$

It follows that

$$|du^k) = |u^m\omega_m^k), \tag{9}$$

where $|\omega_m^k)$ is some 1-form. Thus, in the space of 2-forms built on the quotient space with respect to the $|u^k)$, we see that $|du^k) = 0$.

In general we find that

$$|u^k) = |dx^i)u_i^k.$$

The set of $|u^k)$ can be replaced by a set of 1-forms of a certain standard form. Consider the set of $(n - m) + n$ forms

$$|u^1), |u^2), \cdots, |u^{n-m}), |dx^1), \cdots, |dx^n).$$

This set certainly spans the cotangent space and is certainly a linearly dependent set. Proceeding from left to right, successively strike out those $|dx^i)$'s which are linear combinations of the preceding $|dx)$ and of the set of $|u)$'s. In this way, a basis, an independent set which spans the space, will be found consisting of the full set of $(n - m)$ $|u)$ and m of the $|dx)$. Let the forms which were stricken from the list be labeled as $|dz^i)$ where $i = 1, \cdots, n - m$. Let the remaining ones be labeled as $|dx^i)$ where $i = 1, \cdots, m$. Then

$$|dz^i) = |u^k)\alpha_k^i + |dx^k)\lambda_k^i. \tag{10}$$

Certainly, if each $|u^k)$ does not vanish at the point in question, the matrix $\|\alpha_i^k\|$ possesses an inverse:

$$\alpha_i^j\beta_j^k = \beta_i^j\alpha_j^k = \delta_i^k.$$

Let

$$|v^i) = |u^k)\alpha_k^i.$$

Taking the exterior derivative,

$$\begin{aligned} |dv^i) &= |du^k)\alpha_k^i - |u^k\, d\alpha_k^i), \\ &= |u^m\omega_m^k)\alpha_k^i - |u^k\, d\alpha_k^i), \\ &= |v^j\beta_j^m\omega_m^k)\alpha_k^i - |v^j\beta_j^k\, d\alpha_k^i), \\ &= |v^j\tilde{\omega}_j^i). \end{aligned}$$

This is the same sort of equation as was Eq. (9). The quotient space with respect to the set of $|v^i)$'s is identical to that with respect to the set of $|u^i)$'s and, furthermore,

$$|dv^i) = 0 \text{ mod } |v^1), \cdots, |v^m).$$

Equation (10) can be written as

$$|dz^i) = |v^i) + |\lambda^i), \tag{11}$$

with

$$|\lambda^i) = |dx^k)\lambda_k^i.$$

Taking the exterior derivative,

$$0 = |dv^i) + |d\lambda^i).$$

On the quotient space, therefore,

$$|d\lambda^i) = 0 \text{ mod } |v^1), \cdots.$$

The meaning of this last statement is this: While it is true that $d\,|dz^i) = 0$ when $|dz^i)$ is considered as a 1-form in the full n-dimensional cotangent space, only because $|dv^i) = 0$ mod $|v^1)$, $\cdots$ does the exterior derivative of $|dz^i)$ vanish when $|dz^i)$ is considered as a 1-form in the m-dimensional cotangent space. That is, the set of equations

$$|dz^i) = |\lambda^i) = |dx^k)\lambda_k^i \tag{12}$$

can be solved in the same manner as the equation

$|df) = |u)$ of the previous section was solved. This yields

$$z^i = z^i(x, c), \tag{13}$$

where the c^i, m in number, are integration constants. The solution, Eq. (13), is such that

$$\partial z^i/\partial x^k = \lambda^i_k,$$

as was demanded by Eq. (12). The key point is that the integration constants, c^i, can be taken as the unmentioned coordinates—which are indeed constants on the surface. Treatment of the (c^i)'s as coordinates in Eq. (13) takes the $|dz^i)$ out of the surface cotangent space via

$$|dz^i) = |dx^k)\lambda^i_k + |dc^k)\frac{\partial z^i}{\partial c_k},$$

where the invertibility of $\|\partial z^i/\partial c_j\|$ is guaranteed by the independence of the set of $|dz^i)$ and $|dx^i)$. If this expression is put into Eq. (11) we find that

$$\begin{aligned} |v^i) &= |dz^i) - |\lambda^i), \\ &= |dc^j)(\partial z^i/\partial c_j). \end{aligned}$$

And, finally, we see that

$$\begin{aligned} |u^i) &= |dc^j)(\partial z^k/\partial c^j)\beta^i_k, \\ |u^i) &= |dc^j)p^i_j \end{aligned}$$

with

$$\|p^i_j\| = \|(\partial z^k/\partial c^j)\beta^i_k\|,$$

an invertible matrix.

The essential condition is given by Eq. (9). Given any independent set of 1-forms, k in number, with each

$$|du^i) = |u^j\omega^i_j),$$

then there exist functions c^i and an invertible matrix of functions $\|p^j_i\|$ such that

$$|u^i) = |dc^j)p^i_j.$$

This is the theorem of Frobenius. There are other conditions equivalent to Eq. (9) which are stated here without proof.

(i) If $|U) = |u^1u^2 \cdots u^k)$, then $|u^i\, dU) = 0$ for all values of i.

(ii) $|dU) = |U\Omega)$.

An important case is that of $k = n - 1$. Any linearly independent set of $(n - 1)$ 1-forms in an n-dimensional space will satisfy the conditions of Frobenius' theorem.

Frobenius' theorem is part of the theory of integrating factors, and it is perhaps worthwhile to follow the steps of the proof in the simple classic example: In 2-dimensional space, let

$$|u) = -y\,|dx) + x\,|dy) \quad \textit{where} \quad x \neq 0.$$

This is the forced situation. Explicitly,

$$\begin{aligned} |du) &= 2\,|dx\, dy) \\ &= (2/x)\,|dx)\,|u). \end{aligned}$$

The list-striking process results in

$$|dy) = x^{-1}\,|u) + (y/x)\,|dx),$$

so the standard form is

$$|v) = x^{-1}\,|u) = -(y/x)\,|dx) + |dy).$$

On the quotient space with respect to $|v)$,

$$|dy) = (y/x)\,|dx) \quad \textit{or} \quad y = cx,$$

so that

$$\begin{aligned} |v) &= x\,|dc) \\ &= x\left|d\frac{y}{x}\right), \end{aligned}$$

and

$$|u) = x^2\left|d\frac{y}{x}\right).$$

E. Metric Definitions

Let a k-dimensional surface in an n-dimensional space be defined via $x^i = x^i(t^1, \cdots, t^k)$. There is the associated mapping of 1-vectors:

$$(dt_i| = (dt_i \mid dx^j)(dx_j|.$$

That is, even on the k-dimensional tangent space, one may refer, as it were, to the full n-dimensional tangent space. In the k-dimensional tangent space, the maximum vector is the k-vector:

$$(dt_1 \cdots dt_k| = (dt| = (dt \mid dx^A)(dx_A|.$$

One may establish a Cartesian identification common to tangent spaces of every degree. Let $(dx_i|$ be the chosen Cartesian system for 1-vectors. Then $(dx_A|$, where A is a k-word, is taken to be the Cartesian system for k-vectors. A unified metric is effected. On a k-dimensional tangent space spanned by $(dt_i|$, the k-dimensional volume element appropriate to the t coordinate system is defined to be

$$\langle dt \mid dt\rangle^{\frac{1}{2}} = [(dt \mid dx^A)(dt \mid dx^A)]^{\frac{1}{2}}.$$

For example, the n-dimensional volume element is

$$(dt \mid dx) = [\det \|g_{ij}\|]^{\frac{1}{2}},$$

with

$$g_{ij} = (dt_i \mid dx^k)(dt_j \mid dx^k).$$

The 1-dimensional volume element, the element of length, is then

$$[(dt \mid dx^i)(dt \mid dx^i)]^{\frac{1}{2}}$$

and so on.

III. KINEMATICS

A. Coordinate Systems

Given a 3-dimensional coordinate space, a trajectory is determined by a vector field, a continuum of tangent vectors defined at each point of the space. Let x^i be the coordinates and let

$$(dt| = (dt \mid dx^i)(dx_i|$$

be the vector field. The equation of the trajectory is obtained by the integration of the equations

$$dx^i/dt = (dt \mid dx^i) = v^i. \tag{14}$$

In this paper, the components of the tangent vector will be functions of the space coordinates alone:

$$v^i = v^i(x).$$

Equations (14) certainly can be integrated—at least in the neighborhood of a point. In fact, a formal solution can be written:

$$dx^i/dt = v^i = v^j(\partial/\partial x^j)x^i$$

so that

$$x^i(t) = \exp\,[t(v^j(\partial/\partial x^j))_0]x^i\,(0),$$

where, in the exponent, the differential operator acts on the coordinates $x^i\,(0)$.

So far, t is simply the parameter of the trajectory. Consider now an extension of the 3-space into a $3 + 1$ space with coordinates (x, t). At every point, a 4-dimensional cotangent space can be attached so that, if $f = f(x, t)$,

$$|df) = |dx^i)(\partial f/\partial x^i) + |dt)(\partial f/\partial t).$$

[From now on, the partial derivative symbol will apply to the (x, t) set of coordinates alone.]

The 1-forms $|dx^i)$ and $|dt)$ form a basis. Let the dual basis be represented as $(\partial/\partial x^i|$ and $(\partial/\partial t|$ so that

$$\left(\frac{\partial}{\partial x^i}\middle| dx^j\right) = \delta_i^j, \qquad \left(\frac{\partial}{\partial t}\middle| dx^i\right) = 0,$$

$$\left(\frac{\partial}{\partial x^i}\middle| dt\right) = 0, \qquad \left(\frac{\partial}{\partial t}\middle| dt\right) = 1.$$

and

$$\left(\frac{\partial}{\partial x^i}\middle| df\right) = \frac{\partial f}{\partial x^i}, \qquad \left(\frac{\partial}{\partial t}\middle| df\right) = \frac{\partial f}{\partial t}.$$

Note that $(\partial/\partial t|$ is not $(dt|$, the tangent vector which defines the trajectory. The trajectory, as a mapping of a 1-dimensional space onto the $(3 + 1)$-dimensional space, is defined by

$$t = t, \quad x^i = x^i(t),$$

so that the tangent vector

$$(dt| = (dt \mid dx^i)\left(\frac{\partial}{\partial x^i}\middle|\right) + (dt \mid dt)\left(\frac{\partial}{\partial t}\middle|\right)$$

or

$$(dt| = v^i\left(\frac{\partial}{\partial x^i}\middle|\right) + \left(\frac{\partial}{\partial t}\middle|\right). \tag{15}$$

As a differential operator, $(dt|\,d$ is the total derivation with respect to t, time:

$$(dt \mid df) = \dot{f} = v^i\frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial t}.$$

In fluid mechanics, this expression is referred to as the comoving derivative.

This extension of the 3-space into a $3 + 1$ space, the transition from geometry to kinematics, is not merely a matter of elegant display. Concealed here is a premise involving the existence of a standard trajectory in 3-space to which the position of any other point traveling in space is correlated. It is the purpose of mechanics to determine the position of a point as a function of distance travelled by the standard point along the standard trajectory. The function is the set of ordered pairs (x, t), among which no value of t is repeated; and the velocity of a point is the limiting ratio of the difference in its positions to the difference in the associated positions of the standard point. The standard trajectory appears to be that traced by a light ray and so, by definition, the velocity of light is unity. The time axis must be laid down as a straight line or, in terms of this paper, $|dt)$ must be taken as one of a Cartesian basis in 4-dimensional cotangent space. The question arises concerning the other Cartesian axes, the spatial axes. Setting aside their determination for the moment, consider the meaning of a rotation of the four Cartesian axes. The time axis is changed but, as distance is preserved under the conditions of a rotation, the unit of time is preserved: the velocity of light is still unity. In the original Cartesian frame, a point at rest in space is seen as a straight line parallel to the time axis. With respect to the rotated axes that line is not parallel to the time axis, but it is straight and represents motion with uniform velocity: The preservation of 4-dimensional distance, Cartesian distance, is the basis for the constancy of the velocity of light in 3-dimensional frames moving with uniform velocity with respect to one another.

Concerning the choice of the Cartesian space axes: are the $|dx^i)$'s suitable? Should a Cartesian identification be established on the

$$\left[\left(\frac{\partial}{\partial x^i}\middle|, \left(\frac{\partial}{\partial t}\middle|\right]\right. \text{ and } [|dx^i), |dt)]$$

bases? A trajectory in the (x, t) frame is not necessarily a straight line. Differentially, the trajectory is a quotient space in which

$$|dx) - v\,|dt) = 0.$$

(For purposes of illustration let there be only one spatial dimension.) Let a rotated set of axes be

$$|dz) = c\,|dx) + s\,|dt)$$

and

$$|d\tau) = -s\,|dx) + c\,|dt).$$

On the trajectory, we find that

$$|dz) = (cv + s)\,|dt),$$
$$|d\tau) = (-sv + c)\,|dt),$$

so that

$$(-sv + c)\,|dz) - (cv + s)\,|d\tau) = 0.$$

The new velocity is

$$cv + s/-sv + c,$$

which in no way introduces or preserves simplicity of description. There is, however, a coordinate system in which the trajectory is described simply, and this simplicity is preserved under rotation. In this coordinate system the trajectory is a straight line—a straight line in *all* Cartesian equivalent systems—and the motion is described simply as that of uniform velocity, velocity which can be anything from zero to unity in magnitude, the particular magnitude and direction being but the label of the particular set of Cartesian axes. The system is discovered in the following way.

The trajectory is a 1-dimensional surface. The cotangent at each point should be a quotient space with respect to three independent 1-forms. Consider the set

$$|\sigma^i) = |dx^i) - v^i\,|dt). \tag{16}$$

In the quotient space with respect to these 1-forms,

$$|dx^i) = v^i\,|dt), \quad [\text{mod } |\sigma^i)],$$

which is an expression of the differential equations of the trajectory. Another expression of the situation is

$$(dt \mid \sigma^i) = 0.$$

Now, the set $|\sigma^i)$ is an independent set, for

$$|\sigma^1\sigma^2\sigma^3) = |\sigma) = |dx) - \epsilon_{iA}v^i\,|dx^A\,dt).$$

Clearly $|\sigma)$ does not vanish, and so the set $|\sigma^i)$ is independent. Here then is the forced situation for the application of Frobenius's theorem—three independent forms in a 4-dimensional space. There must exist three independent functions $\bar{x}^i = \bar{x}^i(x, t)$ such that

$$|\sigma^i) = |d\bar{x}^j)p^i_j, \tag{17}$$

where the p^i_j are certain functions. The inverse relation is

$$|d\bar{x}^i) = |\sigma^j)q^i_j \tag{18}$$

with

$$p^k_i q^j_k = q^k_i p^j_k = \delta_i{}^j. \tag{19}$$

The following bracket relations hold:

$$(dt \mid \sigma^i) = 0,$$
$$\left(\frac{\partial}{\partial x^i} \,\middle|\, \sigma^j\right) = \delta^j_i,$$
$$(dt \mid d\bar{x}^i) = 0.$$

Define

$$(d\bar{x}_i| = p^j_i(\partial/\partial x^j| \tag{20}$$

so that

$$(d\bar{x}_i \mid dt) = 0,$$
$$(d\bar{x}_i \mid d\bar{x}^j) = p^k_i(\partial/\partial x^k \mid \sigma^m)q^j_m = \delta^j_i.$$

This confusion of forms and vectors can be relieved somewhat by grouping into bases. Firstly, there is the $[|dx^i), |dt)]$ basis with dual $[(\partial/\partial x^i|, (\partial/\partial t|]$. Secondly, there is the $[|\sigma^i), |dt)]$ basis with dual $[(\partial/\partial x^i|, (dt|]$. Thirdly, there is the $[|d\bar{x}^i), |dt)]$ basis with dual $[(d\bar{x}_i|, (dt|]$. The relevant transformation functions are

$$\begin{aligned}(\partial/\partial x^i \mid dx^j) &= \delta^j_i, \quad (\partial/\partial x^i \mid dt) = 0,\\ (\partial/\partial x^i \mid d\bar{x}^j) &= q^j_i, \quad (d\bar{x}_i \mid dx^j) = p^j_i.\end{aligned} \tag{21}$$

A trajectory is the curve $\bar{x} =$ constant, a straight line in the $(\bar{x}, t)$ system. In that system the velocity is zero. If the $(\bar{x}, t)$ system is rotated—that is, in a frame Cartesian equivalent to the $(\bar{x}, t)$ frame—the velocity is constant along a trajectory. What better way is there to find a Cartesian system? It is the system laid out by the moving point; it is the system in which, by its very definition, the point is undisturbed and moves with constant velocity. It is the inertial system and all systems Cartesian equivalent to it are coinertial systems.

B. Kinematical Equations

With a Cartesian identification established on the basis of $[(d\bar{x}_i|, (dt|]$, how are measurements effected in the lived-in system, the

$$\left[\left(\frac{\partial}{\partial x^i}\right|, \left(\frac{\partial}{\partial t}\right|\right]$$

system? A measurement of spatial distance is accomplished by a small displacement δx^j developed by the tangent vector

$$(\delta s| = \delta x^j\left(\frac{\partial}{\partial x^j}\right|.$$

There is no time change:

$$\delta t = (\delta s \mid dt) = 0.$$

The square of the element of length is

$$\begin{aligned}(\delta s \mid d\bar{x}^i)(\delta s \mid d\bar{x}^i) + (\delta s \mid dt)(\delta s \mid dt) \\ = (\delta s \mid \sigma^j)(\delta s \mid \sigma^k) q_j^i q_k^i, \\ = \delta x^j \delta x^k \gamma_{jk}, \end{aligned} \tag{22}$$

with

$$\gamma_{ij} = q_i^m q_j^m = \gamma_{ji}. \tag{23}$$

A measurement of the time interval is accomplished via the tangent vector

$$(\delta s| = \delta t\, (\partial/\partial t|.$$

The square of the length of the time interval is calculated to be equal to

$$(\delta t)^2(1 + v^i v^j \gamma_{ij}). \tag{24}$$

The volume element is given by

$$(\partial/\partial x \mid d\bar{x}) = \gamma, \tag{25}$$

with

$$\det \|\gamma_{ij}\| = \gamma^2. \tag{26}$$

The inverse of $\|\gamma_{ij}\|$ is

$$\gamma^{ij} = p_m^i p_m^j = \gamma^{ji}. \tag{27}$$

The tensor $\|\gamma_{ij}\|$ appears as the relevant metric tensor for spatial measurements. Its variation is revealed by its exterior derivative—as is the variation of any quantity. The development of a set of equations of variation can begin with

$$|d\sigma^i) = -|d\bar{x}^j)\, |dp_j^i),$$

which is reducible, by Eq. (18), to

$$\begin{aligned}|d\sigma^i) &= -|\sigma^k) q_k^j\, |dp_j^i) \\ &= |\sigma^k \omega_k^i) \end{aligned} \tag{28}$$

with

$$|\omega_i^j) = -q_i^k\, |dp_k^j), \tag{29}$$

because

$$q_i^k p_k^j = \delta_i^j,$$

$$|dq_i^k) p_k^j + q_i^k\, |dp_k^j) = 0,$$

so that

$$|\omega_i^j) = |dq_i^k) p_k^j. \tag{30}$$

It is not always necessary to show the indices in an expression. With the understanding that summation occurs whenever an upper index is to the left of a lower index, the formulas can be written with suppressed indices as

$$qp = pq = 1,$$

$$|d\sigma) = |\sigma\omega),$$

$$|\omega) = -q\, |dp) = |dq)p.$$

Alternatively,

$$|dp) + p\, |\omega) = 0, \tag{31}$$

$$|dq) - |\omega)q = 0. \tag{32}$$

To obtain the variation of the velocity, begin with

$$v = vqp.$$

Take the derivative:

$$\begin{aligned}|dv) &= |dvq)p + vq\, |dp) \\ &= |dvq)p - vqp\, |\omega) \end{aligned}$$

or

$$|dv) + v\, |\omega) = |dvq)p. \tag{33}$$

Equation (33) is the inhomogeneous associate of the homogeneous Eq. (31). Let

$$|\rho) = |dvq)p, \tag{34}$$

so that

$$|dv) + v\, |\omega) = |\rho). \tag{35}$$

No special property of v has been used in the derivation of Eq. (35). By the defining equation [Eq. (16)]:

$$\begin{aligned}|d\sigma) &= -|dv\, dt) \\ &= v\, |\omega\, dt) - |\rho\, dt), \end{aligned}$$

and because

$$\begin{aligned}|d\sigma) &= |\sigma\omega) \\ &= |dx\omega) - v\, |dt\omega), \end{aligned}$$

therefore

$$v\, |\omega\, dt) - |\rho\, dt) = |dx\omega) - v\, |dt\omega)$$

or

$$|dx\omega) = -|\rho\, dt).$$

The $\left(\dfrac{\partial}{\partial x}\dfrac{\partial}{\partial t}\right|$ component of the last equation is

$$\left(\frac{\partial}{\partial x}\frac{\partial}{\partial t}\right| dx\omega\Big) = -\left(\frac{\partial}{\partial x}\frac{\partial}{\partial t}\right| \rho\, dt\Big), \tag{36}$$

$$(\partial/\partial t \mid \omega) = -(\partial/\partial x \mid \rho).$$

However,

$$\partial v/\partial t = (\partial/\partial t \mid dv) = 0,$$

and so the $(\partial/\partial t|$ component of Eq. (35) is

$$v(\partial/\partial t \mid \omega) = (\partial/\partial t \mid \rho).$$

This, combined with Eq. (36), reveals that

$$(\partial/\partial t \mid \rho) + v(\partial/\partial x \mid \rho) = 0,$$

or

$$(dt \mid \rho) = 0. \tag{37}$$

The converse of Eq. (34) is

$$|dvq) = |\rho)q.$$

Therefore,

$$(dt \mid dvq) = d|dt(vq) = 0. \tag{38}$$

Along a trajectory, vq is a constant. Its square is also a constant:

$$v^i q^i_j v^k q^i_k = v^i v^k \gamma_{jk} = \text{const.}$$

It is convenient to define

$$\xi_i = v^j \gamma_{ji}, \tag{39}$$

so that

$$v^i = \gamma^{ij} \xi_j. \tag{40}$$

The exterior derivative of $\|\gamma_{ij}\|$ follows easily from its definition in terms of $\|q^j_i\|$:

$$|d\gamma_{ij}) = |\omega^k_i)\gamma_{kj} + |\omega^k_j)\gamma_{ki}. \tag{41}$$

Therefore

$$\begin{aligned} |d\xi_i) &= |dv^j)\gamma_{ji} + v^j\,|d\gamma_{ji}) \\ &= |\omega^k_i)\xi_k + |\rho^j)\gamma_{ji}. \end{aligned} \tag{42}$$

From Eqs. (35) and (42),

$$|dv\xi) = 2\,|\rho)\xi, \tag{43}$$

and it is confirmed that

$$\begin{aligned} (dt \mid dv\xi) &= 2(dt \mid \rho)\xi \\ &= 0. \end{aligned}$$

C. Dynamical Equations

A Cartesian coordinate system has been chosen, the $(\bar{x}, t)$ system. What is needed now is a function $\bar{t}$ such that $(x, \bar{t})$ can be used as a 4-dimensional coordinate system. Curves of constant x in the $(\bar{x}, t)$ system are not necessarily parallel to the t-axis, nor are they necessarily straight lines. $\bar{t}$ is to serve as the variable coordinate along curves of constant x. It is to be a function of (x, t) so that

$$|d\bar{t}) = |dt)H - |dx^i)\Pi_i, \tag{44}$$

where H and Π are as yet undetermined. By stipulation,

$$(d\bar{t} \mid dx^i) = 0,$$

and therefore,

$$(d\bar{t}| = 1/H\,(\partial/\partial t|. \tag{45}$$

The 1-vectors

$$(dx_i| = \left(\frac{\partial}{\partial x^i}\right| + \frac{\Pi_i}{H}\left(\frac{\partial}{\partial t}\right| \tag{46}$$

satisfy the conditions

$$(dx_i \mid d\bar{t}) = 0$$

and

$$(dx_i \mid dx^j) = \delta^j_i.$$

A new 1-form basis and dual is that of $[|dx^i), |d\bar{t})$ and $[(dx_i|, (d\bar{t}|]$.

That Eq. (44) is truly an expression of an exterior derivative is certified by

$$d\,|d\bar{t}) = 0,$$

or

$$|dH\,dt) + |dx^i\,d\Pi_i) = 0,$$

$$|dx^i\,dt)\frac{\partial H}{\partial x^i} + |dx^i\,dt)\frac{\partial \Pi_i}{\partial t} + |dx^i\,dx^j)\frac{\partial \Pi_i}{\partial x^j} = 0.$$

From that two sets of equations are obtained:

$$\partial \Pi_i/\partial x^j = \partial \Pi_j/\partial x^i \tag{47}$$

and

$$\partial \Pi_i/\partial t = -\partial H/\partial x^i. \tag{48}$$

The equations of motion are contained in Eqs. (47) and (48). It remains to be shown that the motion is determined by the metric tensor of the $(x, \bar{t})$ system.

Because

$$\begin{aligned} |dx^i) &= |dt)v^i + |d\bar{x}^j)p^i_j, \\ \langle dx^i| &= v^i(dt| + p^i_j(d\bar{x}_j|. \end{aligned}$$

Thus

$$\begin{aligned} g^{ij} &= \langle dx^i \mid dx^j\rangle, \\ &= v^i v^j + p^i_m p^j_m, \end{aligned}$$

or

$$g^{ij} = v^i v^j + \gamma^{ij}. \tag{49}$$

By the definition, Eq. (44),

$$\langle d\bar{t}| = H(dt| - \Pi_i\,\langle dx^i|,$$

so that

$$\begin{aligned} g^{0i} &= \langle d\bar{t} \mid dx^i\rangle \\ &= Hv^i - g^{ij}\Pi_j, \end{aligned} \tag{50}$$

and

$$\begin{aligned} g^{00} &= \langle dt \mid d\bar{t}) \\ &= H^2 - Hv\Pi - g^{0i}\Pi_i. \end{aligned}$$

From Eq. (50),

$$g^{0i}\Pi_i = Hv\Pi - g^{ij}\Pi_i\Pi_j,$$

so that

$$g^{00} = H^2 - 2g^{0i}\Pi_i - g^{ij}\Pi_i\Pi_j.$$

The relation of inverse metric tensors implies that

$$g^{ik}g_{k0} + g^{i0}g_{00} = 0,$$

so that

$$-g^{i0} = (g^{ik}g_{k0})/g_{00}.$$

Let

$$A_i = g_{i0}/g_{00}. \tag{51}$$

Then

$$g^{ij}A_j = -g^{i0}, \tag{52}$$

and

$$\begin{aligned} g^{00} &= H^2 + 2g^{ij}A_j\Pi_i - g^{ij}\Pi_i\Pi_j \\ &= H^2 - g^{ij}(\Pi_i - A_i)(\Pi_j - A_j) + g^{ij}A_iA_j. \end{aligned}$$

From Eq. (52)

$$\begin{aligned} g^{ij}A_iA_j &= -g^{0i}A_i \\ &= -(g^{0i}g_{i0})/g_{00} \\ &= -(1 - g^{00}g_{00})/g_{00} \\ &= g^{00} - 1/g_{00}, \end{aligned}$$

where, again, the relation of inverse metric tensors has been used. Thus

$$g^{00} = H^2 - g^{ij}(\Pi_i - A_i)(\Pi_j - A_j) - 1/g_{00} + g^{00}$$

or

$$H^2 = 1/g_{00} + g^{ij}(\Pi_i - A_i)(\Pi_j - A_j). \quad (53)$$

Going on to the lower index metric tensor,

$$\begin{aligned}(d\bar{t}| &= 1/H\,(\partial/\partial t| \\ &= 1/H\,(dt| - v^i/H\,(\partial/\partial x^i| \\ &= 1/H\,(dt| - vq/H\,(d\bar{x}|,\end{aligned}$$

so that

$$\begin{aligned}|d\bar{t}\rangle &= |dt)1/H - |d\bar{x}^i)v^i q_i^j/H \\ &= |dt)1/H - |\sigma)\xi/H.\end{aligned}$$

Therefore,

$$\begin{aligned}g_{00} &= (d\bar{t}\,|\,dt\rangle \\ &= 1/H^2 + v\xi/H^2,\end{aligned}$$

or

$$H^2 g_{00} = 1 + v\xi. \quad (54)$$

Going on,

$$\begin{aligned}g_{0i} &= \langle dx_i\,|\,d\bar{t}\rangle \\ &= (dx_i\,|\,dt)1/H - (dx_i\,|\,\sigma^k)\xi_k/H.\end{aligned}$$

Because

$$(dx_i\,|\,d\bar{t}) = 0,$$

it follows from Eq. (44) that

$$(dx_i\,|\,dt) = \Pi_i/H, \quad (55)$$

and so

$$g_{0i} = \Pi_i/H^2 - \xi_i/H + v\xi\Pi_i/H^2.$$

By Eqs. (54) and (51), this becomes

$$\xi = Hg_{00}(\Pi - A) \quad (56)$$

Finally,

$$\begin{aligned}(dx_i| &= (\partial/\partial x^i| + \Pi_i/H\,(\partial/\partial t| \\ &= (\partial/\partial x^i| + \Pi_i(d\bar{t}|,\end{aligned}$$

so that

$$|dx_i\rangle = |\sigma^k)\gamma_{ki} + |d\bar{t}\rangle\Pi_i$$

and

$$g_{ij} = \gamma_{ij} - \xi_j\Pi_i/H + g_{0i}\Pi_j,$$

which, combined with previously derived relations, becomes

$$g_{ij} = \gamma_{ij} - \xi_i\xi_j/H^2 g_{00} + g_{00}A_iA_j. \quad (57)$$

D. Equations of Motion

The equations of motion can now be given explicitly in terms of the metric tensor and the various dynamical variables which are connected by the metric tensors. To repeat Eq. (48),

$$\partial\Pi_i/\partial t = -\partial H/\partial x^i.$$

From Eq. (53),

$$\begin{aligned}2H\frac{\partial H}{\partial x^i} &= \frac{\partial}{\partial x^i}\left(\frac{1}{g_{00}}\right) + 2g^{jk}(\Pi_j - A_j)\frac{\partial}{\partial x^i}(\Pi_k - A_k) \\ &\quad + \frac{\partial g^{jk}}{\partial x^i}(\Pi_j - A_j)(\Pi_k - A_k) \\ &= \frac{\partial}{\partial x^i}\left(\frac{1}{g_{00}}\right) + 2Hv^j\frac{\partial}{\partial x^i}\Pi_j - 2Hv^j\frac{\partial}{\partial x^i}A_j \\ &\quad + \frac{\partial g^{jk}}{\partial x^i}(\Pi_j - A_j)(\Pi_k - A_k)\end{aligned}$$

or

$$\begin{aligned}\frac{\partial H}{\partial x^i} &= \frac{1}{2H}\frac{\partial}{\partial x^i}\left(\frac{1}{g_{00}}\right) + v^j\frac{\partial\Pi_j}{\partial x^i} - v^j\frac{\partial A_j}{\partial x^i} \\ &\quad + \frac{1}{2H}\frac{\partial g^{jk}}{\partial x^i}(\Pi_j - A_j)(\Pi_k - A_k).\end{aligned}$$

Therefore, using Eq. (47),

$$\begin{aligned}(\Pi_i - A_i) = &-\frac{1}{2H}\frac{\partial}{\partial x^i}\left(\frac{1}{g_{00}}\right) - \frac{\partial A_i}{\partial t} \\ &+ v^j\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right) \\ &- \frac{1}{2H}\frac{\partial g^{jk}}{\partial x^i}(\Pi_j - A_j)(\Pi_k - A_k). \quad (58)\end{aligned}$$

Before going on to an interpretation of the last equation, let us consider the case of (x, t) being Cartesian. In that case,

$$\begin{aligned}g^{ij} = \delta^{ij}, &\quad g_{ij} = \delta_{ij}, \\ g^{00} = 1, &\quad g_{00} = 1, \\ A_i = 0.&\end{aligned}$$

Equation (58) states that Π is constant along the trajectory. The general connection between v and Π can be derived from Eqs. (50) and (52):

$$Hv^i = g^{ij}(\Pi_j - A_j), \quad (59)$$

which in this special case of Cartesian equivalence reduces to

$$Hv^i = \Pi_i.$$

Therefore, Eq. (53) reduces to

$$H^2 = 1 + H^2v^2,$$

or

$$H^2 = 1/(1 - v^2),$$

so that

$$\Pi = v/(1 - v^2)^{\frac{1}{2}},$$

and

$$\xi = v/(1 - v^2).$$

How are measurements effected in this case? The tensor $\|\gamma_{ij}\|$ reduces to

$$\gamma_{ij} = \delta_{ij} + v^i v^j/(1 - v^2)$$

and so Eq. (22), the square of the element of length, becomes

$$(\delta x)^2 + (v\delta x)^2/(1 - v^2).$$

For a measurement perpendicular to the velocity $v\delta x = 0$, the length is calculable as δx; for a measurement along the velocity, $(v\delta x)^2 = v^2(\delta x)^2$, the length is calculable as $\delta x(1 - v^2)^{-\frac{1}{2}}$. The interpretation is this: In the x-system a spatial distance is always measured as δx—the world is *seen* as Cartesian. For two points moving along parallel paths having equal speeds and with the line connecting them being perpendicular to the path, the measured separation is the same as the one for stationary points. However, for two points moving one behind the other along a common path with equal speeds, the measured separation is $(1 - v^2)^{\frac{1}{2}}$ times the separation of a similar set of points at rest. This, of course, is an expression of the Lorentz contraction.

In the Cartesian case, the element of time interval, Eq. (24), reduces to $\delta t(1 - v^2)^{-\frac{1}{2}}$. If a ticking clock is attached to the point on the trajectory, the time between ticks is longer than the time between ticks of a stationary clock by a factor of $(1 - v^2)^{-\frac{1}{2}}$. This, of course, is an expression of the relativistic time dilation.

It is a matter of rotating a coordinate system to bring the differential equations, now immediately integrable, into the familiar Lorentz transformation form. For a single spatial dimension,

$$q = p^{-1} = \gamma = (1 - v^2)^{-\frac{1}{2}},$$

and so

$$\bar{x} = \gamma(x - vt),$$
$$\bar{t} = \gamma(t - vx).$$

In terms of a rotation through a real angle whose sine is equal to v,[4]

$$x = (1 - v^2)^{\frac{1}{2}}\bar{x} + vt,$$
$$t = -v\bar{x} + (1 - v^2)^{\frac{1}{2}}t.$$

E. Newtonian Approximation

By means of Eq. (58), the forces responsible for deviation from constant velocity are deriveable from the metric tensor. The last term from the right-hand side of Eq. (58),

$$-\frac{1}{2H}\frac{\partial g^{jk}}{\partial x^i}(\Pi_j - A_j)(\Pi_k - A_k),$$

is the type of force which appears in the expression for acceleration in nonrectangular coordinates. To see that this is so, one must play havoc with the theory of this paper. Newtonian kinematics is no more than 3-dimensional geometry with time being merely a parameter which is in no way correctly incorporated into a proper 4-dimensional formalism. Thus, in the plane, the conversion from rectangular coordinates—they may as well be rectangular—to polar coordinates is a time-independent transformation:

$$x = r\cos\theta,$$
$$y = r\sin\theta,$$

which yields

$$|dr) = |dx)\cos\theta + |dy)\sin\theta,$$
$$|d\theta) = -|dx)(\sin\theta/r) + |dy)(\cos\theta/r).$$

Based on the (x, y) rectangular system,

$$\langle dr| = \cos\theta\ \langle dx| + \sin\theta\ \langle dy|,$$
$$\langle d\theta| = -(\sin\theta/r)\ \langle dx| + (\cos\theta/r)\ \langle dy|,$$

so that

$$g^{11} = \langle dr \mid dr) = 1,$$
$$g^{12} = \langle dr \mid d\theta) = 0,$$
$$g^{22} = \langle d\theta \mid d\theta) = 1/r^2.$$

The Newtonian mistake is, of course, the neglect of g^{i0} and g_{i0} through ignorance. There is no A in Newtonian kinematics. Therefore, by Eq. (59),

$$H\dot{r} = g^{11}\Pi_1 + g^{12}\Pi_2,$$
$$H\dot{r} = \Pi_1,$$

and

$$H\dot{\theta} = (r^2)^{-1}\Pi_2.$$

Now

$$H^2 g_{00} = 1 + v\xi,$$

which is constant along a trajectory. In the Newtonian view, $g_{00} = 1$; thus H is a constant. The left-hand side of Eq. (58) becomes

$$H\ddot{r}$$

and

$$H\frac{d}{dt}(r^2\dot{\theta}).$$

The last term on the right becomes

$$-\frac{1}{2H}\frac{\partial g^{jk}}{\partial \gamma}\Pi_j\Pi_k = Hr\dot{\theta}^2$$

and

$$-\frac{1}{2H}\frac{\partial g^{jk}}{\partial \theta}\Pi_j\Pi_k = 0.$$

Equation (58), now considerably weakened, becomes

$$\ddot{r} - r\dot{\theta}^2 = 0$$

[4] One of the stimulants for this work was the observation that if $x^2 - t^2 = \bar{x}^2 - \bar{t}^2$, then $x^2 + \bar{t}^2 = \bar{x}^2 + t^2$.

and

$$\frac{d}{dt}(r^2\dot{\theta}) = 0$$

in which can be recognized the components of acceleration in polar coordinates. (The θ component appears quite naturally as the rate of change of angular momentum.) Any reference to "true" force which may have been contained in Eq. (58) has been completely eliminated by the severe Newtonian prejudice—that kinematics and dynamics are separate. One might be tempted to write

$$H\ddot{r} = Hr\dot{\theta}^2$$

and

$$H\ddot{\theta} = -(\partial H/r)\dot{r}\dot{\theta}.$$

But, as it is a definite pronouncement that r and θ are *not* rectangular coordinates, it would be a violation of the rules to read the equation as $F = ma$. The path of a free particle is not a linear relation between r and θ, but this is not attributed to forces. The present theory accepts the fact that the (x, t) system is neither Cartesian nor even associated with a constant metric tensor. Forces are either all "true" or all "fictitious": kinematical relations are dynamical relations.

An illustration of this is given by a more cautious approach to the Newtonian case. Let the velocity be small and let the metric tensor stay close to unity. The last term in Eq. (58) can be neglected. Because g_{00} is a positive number it can be expressed as

$$g_{00} = e^{-2\varphi}. \tag{60}$$

The constant of motion, $1 + v\xi$, can be labeled as

$$m^2 = 1 + v\xi, \tag{61}$$

so that

$$H^2 = m^2e^{2\varphi} \tag{62}$$

by Eq. (54). With $g^{ij} \approx \delta^{ij}$, velocity and momentum are related by

$$Hv = \Pi - A.$$

Equation (58) can be written now as

$$\frac{d}{dt}(Hv^i) = -\frac{1}{2me^{\varphi}}\frac{\partial}{\partial x^i}e^{2\varphi} - \frac{\partial A_i}{\partial t} + v^j\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right)$$
$$= -\frac{e^{\varphi}}{m}\frac{\partial \varphi}{\partial x^i} - \frac{\partial A_i}{\partial t} + v^j\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right).$$

With φ small, this becomes

$$m\frac{d}{dt}(1 + \varphi)v^i = -\frac{1}{m}\frac{\partial \varphi}{\partial x^i} - \frac{\partial A_i}{\partial t} + v^j\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right).$$

If m were spatially constant, or if the term

$$\varphi(\partial/\partial x^i)(1/m)$$

were negligible, the equation of motion would be that of a particle of intrinsic mass m moving in a potential field φ/m with an effective mass $m(1 + \varphi)$. The terms involving A would be interpreted as effects of a noninertial frame. To clarify the last point, recall that

$$g^{ij}A_iA_j = g^{00} - 1/g_{00}.$$

The present approximation requires that

$$A^2 = g^{00} - 1/g_{00},$$

with

$$g^{00} = e^{2V}$$
$$\approx 1 + 2V,$$
$$\varphi = V - \tfrac{1}{2}A^2.$$

The equation of motion is now

$$m\frac{d}{dt}(1 + \varphi)v^i = -\frac{1}{m}\frac{\partial V}{\partial x^i} + \frac{1}{2m}\frac{\partial A^2}{\partial x^i} - \frac{\partial A_i}{\partial t} + v^j\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right).$$

The simplest situation is that in which

$$A_i = m\epsilon_{ijk}x^j\omega^k$$

with m and ω constant. The equation of motion can be written as

$$m\frac{d}{dt}(1 + \varphi)v^i = -\frac{\partial}{\partial x^i}\frac{V}{m} + \frac{m}{2}\frac{\partial}{\partial x^i}[x^2\omega^2 - (x\omega)^2] + 2m\epsilon_{ijk}\omega^jv^k.$$

The Newtonian interpretation is:

(i) V/m is the potential energy;

(ii) ω is the angular velocity of the rotating frame;

(iii) $(m/2)[x^2\omega^2 - (x\omega)^2]$ is the centripetal potential energy;

(iv) $2m\epsilon_{ijk}\omega^jv^k$ is the Coriolis force.

The constant of motion m, called the intrinsic mass, is intrinsic to a particular trajectory; it varies from trajectory to trajectory. In general,

$$m^2 = 1 + (vq)^2,$$

and, in the case of a Cartesian (x, t) system,

$$m^2 = 1/(1 - v^2).$$

The number 1 in the numerator of the last equation is as good as any other number insofar as the present theory can do anything about it. To anticipate the closing remarks: there is little of significant result from the conception of mass for a single particle in an imposed field.

IV. CONCLUDING REMARKS

The metric tensor determines the motion, but what determines the metric tensor? The equation $ma = F$—while *mathematically* symmetric, $(F = ma)$—is not

physically so. F is to be given, and the acceleration is fixed by it. So the theory of this paper is incomplete in so far as it describes the motion of a single particle against an immutable background. The origin of the background is unknown; the metric tensor is to be given. Newtonian mechanics can be completed by a prescription for the determination of the force which is independent of the detailed observation of a point moving under that force—Newton's law of gravitational attraction. Any other truly Newtonian force is independent of velocity and so can be determined by opposition to a gravitational force, where both forces act on a stationary point—e.g., the linear spring extended by a hanging weight. However, for a large enough system of particles, there is no force *external* to the system; thus the force acting on a particular particle is a summarization—approximate or not—of the mutable configuration of the other particles. So perhaps the metric tensor is set by the configuration of other particles. Through a proper treatment of the differential geometry of many trajectories it may be possible to eliminate the metric tensor formally, retaining it only for the very practical function of an intermediary.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Weak Correspondence Principle

JOHN R. KLAUDER*
Bell Telephone Laboratories, Incorporated Murray Hill, New Jersey

(Received 5 April 1967)

The weak correspondence principle (WCP) for a scalar field states that the diagonal matrix elements

$$G(f,g) \equiv \langle f,g|\,\mathcal{G}\,|f,g\rangle$$

of a quantum generator $\mathcal{G}$ necessarily have the form of the appropriate classical generator G in which $f(\mathbf{x})$ and $g(\mathbf{x})$ are interpreted as the classical momentum and field, respectively. For a field operator $\varphi(\mathbf{x})$ and its canonically conjugate momentum $\pi(\mathbf{x})$ the states in question are given by

$$|f,g\rangle \equiv \exp\{i\int[\varphi(\mathbf{x})f(\mathbf{x}) - \pi(\mathbf{x})g(\mathbf{x})]\,d\mathbf{x}\}\,|0\rangle,$$

where $|0\rangle$ denotes the vacuum. The validity of the WCP is established for the six Euclidean generators (plus the Hamiltonian) of a Euclidean-invariant theory, and for the ten Poincaré generators of a Lorentz-invariant theory. Only general properties and certain operator domain conditions are essential to our argument. The WCP holds whether the representation of π and φ is irreducible or reducible; in the latter case, the WCP holds even if the vectors $|f,g\rangle$ do not span the Hilbert space, or even if the generator $\mathcal{G}$ is not a function solely of π and φ. Thus, the WCP is an exceedingly general and completely representation-independent connection between a classical theory and its quantum generators which is especially useful in the formulation of nontrivial, Euclidean-invariant quantum field theories.

1. INTRODUCTION

BASIC to any quantization procedure is a prescription for relating the quantum problem to its classical counterpart. The traditional guide in this respect has been a prescription which involves, essentially, a straightforward operator substitution for "coordinates" and "momenta" in the classical generators. In field theory, a normal ordering and subsequent renormalization is often necessary to make any sense of such a prescription. In this paper we shall establish a more general correspondence rule—the weak correspondence principle (WCP)—and show how it accounts for normal ordering and certain renormalizations in field theory. More significantly, however, we show that the WCP is far more general and that it applies in cases where the traditional prescription is manifestly incorrect. This latter aspect has been concretely demonstrated in recent analyses of the soluble "rotationally-symmetric" models.[1,2]

We confine our attention to a neutral scalar quantum field $\varphi(\mathbf{x}, t)$ in the presence of Euclidean-invariant self-interactions. In Sec. 2 some basic definitions and the general statement of the WCP for

* Present address: Department of Physics, Syracuse University, Syracuse, N.Y.

[1] J. R. Klauder, J. Math. Phys. **6**, 1666 (1965).
[2] H. D. I. Abarbanel, J. R. Klauder, and J. G. Taylor, Phys. Rev. **152**, 1198 (1966).

physically so. F is to be given, and the acceleration is fixed by it. So the theory of this paper is incomplete in so far as it describes the motion of a single particle against an immutable background. The origin of the background is unknown; the metric tensor is to be given. Newtonian mechanics can be completed by a prescription for the determination of the force which is independent of the detailed observation of a point moving under that force—Newton's law of gravitational attraction. Any other truly Newtonian force is independent of velocity and so can be determined by opposition to a gravitational force, where both forces act on a stationary point—e.g., the linear spring extended by a hanging weight. However, for a large enough system of particles, there is no force *external* to the system; thus the force acting on a particular particle is a summarization—approximate or not—of the mutable configuration of the other particles. So perhaps the metric tensor is set by the configuration of other particles. Through a proper treatment of the differential geometry of many trajectories it may be possible to eliminate the metric tensor formally, retaining it only for the very practical function of an intermediary.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Weak Correspondence Principle

JOHN R. KLAUDER*
Bell Telephone Laboratories, Incorporated Murray Hill, New Jersey

(Received 5 April 1967)

The weak correspondence principle (WCP) for a scalar field states that the diagonal matrix elements

$$G(f,g) \equiv \langle f,g|\,\mathcal{G}\,|f,g\rangle$$

of a quantum generator $\mathcal{G}$ necessarily have the form of the appropriate classical generator G in which $f(\mathbf{x})$ and $g(\mathbf{x})$ are interpreted as the classical momentum and field, respectively. For a field operator $\varphi(\mathbf{x})$ and its canonically conjugate momentum $\pi(\mathbf{x})$ the states in question are given by

$$|f,g\rangle \equiv \exp\{i\int[\varphi(\mathbf{x})f(\mathbf{x}) - \pi(\mathbf{x})g(\mathbf{x})]\,d\mathbf{x}\}\,|0\rangle,$$

where $|0\rangle$ denotes the vacuum. The validity of the WCP is established for the six Euclidean generators (plus the Hamiltonian) of a Euclidean-invariant theory, and for the ten Poincaré generators of a Lorentz-invariant theory. Only general properties and certain operator domain conditions are essential to our argument. The WCP holds whether the representation of π and φ is irreducible or reducible; in the latter case, the WCP holds even if the vectors $|f,g\rangle$ do not span the Hilbert space, or even if the generator $\mathcal{G}$ is not a function solely of π and φ. Thus, the WCP is an exceedingly general and completely representation-independent connection between a classical theory and its quantum generators which is especially useful in the formulation of nontrivial, Euclidean-invariant quantum field theories.

1. INTRODUCTION

BASIC to any quantization procedure is a prescription for relating the quantum problem to its classical counterpart. The traditional guide in this respect has been a prescription which involves, essentially, a straightforward operator substitution for "coordinates" and "momenta" in the classical generators. In field theory, a normal ordering and subsequent renormalization is often necessary to make any sense of such a prescription. In this paper we shall establish a more general correspondence rule—the weak correspondence principle (WCP)—and show how it accounts for normal ordering and certain renormalizations in field theory. More significantly, however, we show that the WCP is far more general and that it applies in cases where the traditional prescription is manifestly incorrect. This latter aspect has been concretely demonstrated in recent analyses of the soluble "rotationally-symmetric" models.[1,2]

We confine our attention to a neutral scalar quantum field $\varphi(\mathbf{x}, t)$ in the presence of Euclidean-invariant self-interactions. In Sec. 2 some basic definitions and the general statement of the WCP for

* Present address: Department of Physics, Syracuse University, Syracuse, N.Y.

[1] J. R. Klauder, J. Math. Phys. **6**, 1666 (1965).
[2] H. D. I. Abarbanel, J. R. Klauder, and J. G. Taylor, Phys. Rev. **152**, 1198 (1966).

such fields are formulated. Applications to Euclidean- and Lorentz-invariant interactions are carried out in Secs. 3 and 4, respectively. Finally, in Sec. 5, we discuss the significance and utility of the WCP. In order to ensure clarity of the concepts involved we adopt a semiheuristic approach in this paper ignoring questions of operator domains, test-function smearing, and the like.

2. BASIC DEFINITIONS AND GENERAL STATEMENT OF WEAK CORRESPONDENCE PRINCIPLE

We deal with a neutral scalar field $\varphi(\mathbf{x})$ and its conjugate momentum $\pi(\mathbf{x})$, which at a common time satisfy the usual canonical commutation relations (CCR),

$$[\varphi(\mathbf{x}), \pi(\mathbf{x}')] = i\delta(\mathbf{x} - \mathbf{x}'), \tag{1}$$

where $\mathbf{x}$, $\mathbf{x}'$ are points in a three-dimensional Euclidean space. Unless otherwise stated, all operators are to be evaluated at time t, a variable which is generally suppressed. We especially use the unitary Weyl operators

$$U[f, g] \equiv \exp\left\{i\int [\varphi(\mathbf{x})f(\mathbf{x}) - \pi(\mathbf{x})g(\mathbf{x})]\, d\mathbf{x}\right\}, \tag{2}$$

which are defined for sufficiently many well-behaved functions $f(\mathbf{x})$ and $g(\mathbf{x})$. To fix the idea, it may be assumed that $f(\mathbf{x})$ and $g(\mathbf{x})$ are infinitely differentiable and fall off at infinity faster than any inverse power of $|\mathbf{x}|$. Evidently, we have $U[f, g]\dagger = U[-f, -g]$. Some straightforward consequences of the CCR which we frequently use are

$$U[f, g] = \exp\left\{\tfrac{1}{2}i\int f(\mathbf{x})g(\mathbf{x})\, d\mathbf{x}\right\}U[0, g]U[f, 0] \tag{3a}$$

$$= \exp\left\{-\tfrac{1}{2}i\int f(\mathbf{x})g(\mathbf{x})\, d\mathbf{x}\right\}U[f, 0]U[0, g], \tag{3b}$$

and the "translation property"

$$U[f, g]^\dagger\{\alpha\varphi(\mathbf{x}) + \beta\pi(\mathbf{x})\}U[f, g] = \alpha\{\varphi(\mathbf{x}) + g(\mathbf{x})\} + \beta\{\pi(\mathbf{x}) + f(\mathbf{x})\}, \tag{4}$$

where α and β are complex numbers.

We introduce a distinguished set of unit vectors according to the definition

$$|f, g\rangle \equiv U[f, g]\,|0\rangle \tag{5}$$

for sufficiently many f and g. In our subsequent applications we regard $|0\rangle$ as the unique translationally invariant state, or equivalently as the unique ground state for the problem at hand. Generally, the vectors $|f, g\rangle$ are not mutually orthogonal for different arguments; on the contrary, telling information regarding the problem and the CCR representation leaves its imprint in the overlap $\langle f, g \mid f', g'\rangle$ between such vectors. While the states $|f, g\rangle$ are most useful when they span the Hilbert space $\mathfrak{H}$, the basic statements embodied in the weak correspondence principle do *not* require that they span $\mathfrak{H}$; the WCP remains valid in the subspace that these vectors do span. For notational convenience we shall not distinguish these cases and we refer simply to the collection of states $|f, g\rangle$ as the overcomplete family of states (OFS).

We are primarily interested in the *diagonal* OFS matrix elements

$$G(f, g) \equiv \langle f, g|\, \mathcal{G}\, |f, g\rangle$$

for various operators interpreted as quantum generators. Additionally, from (3), we may deduce that

$$G(f, g) = \langle 0|\, U[0, -g]U[-f, 0]\mathcal{G}\, U[f, 0]U[0, g]\,|0\rangle, \tag{6a}$$

$$= \langle 0|\, U[-f, 0]U[0, -g]\mathcal{G}\, U[0, g]U[f, 0]\,|0\rangle, \tag{6b}$$

where we note that

$$U[f, 0] = \exp\left\{i\int \varphi(\mathbf{x})f(\mathbf{x})\, d\mathbf{x}\right\}, \tag{7a}$$

$$U[0, g] = \exp\left\{-i\int \pi(\mathbf{x})g(\mathbf{x})\, d\mathbf{x}\right\}. \tag{7b}$$

Suppose, now, one seeks to choose an operator $\mathcal{G}$ to associate with some classical generator (e.g., Hamiltonian) $G(\pi_{cl}, \varphi_{cl})$ depending on the classical field $\varphi_{cl}(\mathbf{x})$ and its conjugate momentum $\pi_{cl}(\mathbf{x})$. In the traditional approach one adopts

$$\mathcal{G} = :G(\pi, \varphi): \tag{8}$$

for the quantum generator where the colons denote some sort of normal ordering. With this choice we see from (4) that

$$\langle 0|\, U[f, g]^\dagger :G(\pi, \varphi):U[f, g]\,|0\rangle = \langle 0|\, :G(\pi + f, \varphi + g):\,|0\rangle = G(f, g), \tag{9}$$

with all other terms vanishing. Thus one consequence of the traditional identification (8) is that diagonal OFS matrix elements of the quantum generator yield the classical generator with the understanding that

$$f(\mathbf{x}) \equiv \pi_{cl}(\mathbf{x}), \quad g(\mathbf{x}) \equiv \varphi_{cl}(\mathbf{x}). \tag{10}$$

This property is not limited to the usual definition of normal ordering (all creation operators to the left of all annihilation operators), but applies to the generalized normal ordering implicit in the generating functional

$$:U[f, g]: \equiv U[f, g]/\langle 0|\, U[f, g]\,|0\rangle,$$

whatever form the c-number denominator may take.

If the CCR representation is irreducible, then the prescription (8) is generally correct. However, if the CCR representation is *reducible*, then for many important operators the preceding prescription *necessarily fails*. This failure is often so complete that $\mathcal{G}$ cannot be expressed as *any* function solely of π and φ. Such is the case, for example, for the three generators of space translations $\mathcal{P}_k$, $k = 1, 2, 3$, and the Hamiltonian $\mathcal{H}$, if we assume that $|0\rangle$ is a nondegenerate eigenstate of these operators.[3] The space–time translation generators for a generalized free field are a case in point.

It is the purpose of the WCP to shed light on the situation even in those cases where (8) breaks down. The WCP states that the diagonal OFS matrix elements of quantum generators have the proper classical functional form independent of whether Eq. (8) holds or not. In symbols, if $\mathcal{G}$ is the quantum generator associated with the classical generator G, then the WCP states that these quantities are connected by the relation

$$G(f, g) = \langle f, g| \mathcal{G} |f, g\rangle \tag{11}$$

not only when (8) is true but even in cases where $\mathcal{G}$ is not any function solely of the π and φ.[4]

Roughly speaking, the WCP does not precast the CCR representation into the narrow confinements of an "irreducible mold"; on the contrary, the WCP leaves completely open the ultimate nature of the CCR representation. As was shown elsewhere,[1] such liberalism is necessary to achieve any solution at all in the case of the "rotationally-symmetric" models.

The remainder of this paper is devoted to verifying (11) for the principal generators in Euclidean- and Lorentz-invariant problems. We treat the simpler Euclidean-invariant cases first as preliminaries to the more interesting Lorentz-invariant cases.

3. WEAK CORRESPONDENCE PRINCIPLE IN EUCLIDEAN-INVARIANT THEORIES

For a given Euclidean-invariant theory we assume there exist three space-translation generators $\mathcal{P}_k$, three infinitesimal rotation generators $\mathcal{J}_k$, $k = 1, 2, 3$, having traditional commutation properties, and a positive Hamiltonian operator $\mathcal{H}$, which commutes with the generators of the Euclidean group. We assume the state $|0\rangle$ is a simultaneous eigenstate such that (for all k)

$$\mathcal{P}_k |0\rangle = \mathcal{J}_k |0\rangle = \mathcal{H} |0\rangle = 0.$$

From Euclidean invariance of $|0\rangle$ it follows that

$$\nabla_k \langle 0| \varphi(\mathbf{x}) |0\rangle = \nabla_k \langle 0| \pi(\mathbf{x}) |0\rangle = 0. \tag{12}$$

Momentum Operators

To illustrate our basic idea most simply let us first consider the space-translation generators $\mathcal{P}_k$. Almost instinctively these operators are identified with the quantities $\int \pi \nabla_k \varphi \, d\mathbf{x}$ suitably normal ordered. Yet, as we have noted,[3] for a reducible CCR representation and unique translationally invariant state, the above identification is manifestly false.

We proceed to determine the three functionals

$$P_k(f, g) = \langle f, g| \mathcal{P}_k |f, g\rangle. \tag{13}$$

From the condition $\mathcal{P}_k |0\rangle = 0$, we conclude that $P_k(0, 0) = 0$. From $[\mathcal{P}_k, \varphi(\mathbf{x})] = -i\nabla_k \varphi(\mathbf{x})$ coupled with (6a), (4), and (12), we learn that

$$\begin{aligned}[\delta/\delta f(\mathbf{x})] P_k(f, g) &= i\langle f, g| [\mathcal{P}_k, \varphi(\mathbf{x})] |f, g\rangle \\ &= \langle f, g| \nabla_k \varphi(\mathbf{x}) |f, g\rangle = \nabla_k g(\mathbf{x}).\end{aligned}$$

In similar fashion, $[\mathcal{P}_k, \pi(\mathbf{x})] = -i\nabla_k \pi(\mathbf{x})$ coupled with (6b), (4), and (12) leads to

$$\begin{aligned}[\delta/\delta g(\mathbf{x})] P_k(f, g) &= -i\langle f, g| [\mathcal{P}_k, \pi(\mathbf{x})] |f, g\rangle \\ &= -\langle f, g| \nabla_k \pi(\mathbf{x}) |f, g\rangle \\ &= -\nabla_k f(\mathbf{x}).\end{aligned}$$

The only functionals consistent with these three conditions are easily seen to be

$$\langle f, g| \mathcal{P}_k |f, g\rangle = P_k(f, g) = \int f(\mathbf{x}) \nabla_k g(\mathbf{x}) \, d\mathbf{x}. \tag{14}$$

Clearly if we interpret $g(\mathbf{x})$ as a *classical* c-number field and $f(\mathbf{x})$ as its conjugate momentum—as we hereafter shall—then (14) states that the diagonal expectation value of the quantum generator of space translations in the states $|f, g\rangle$ yields the classical generator of space translations. This is just the WCP as applied to the space-translation generators. In obtaining this result we note that no functional form for $\mathcal{P}_k$ was assumed, nor was it assumed that the states $|f, g\rangle$ span $\mathfrak{H}$.

[3] The field and momentum operators are reducible if they admit a common decomposition $\varphi = \varphi_1 \oplus \varphi_2$, $\pi = \pi_1 \oplus \pi_2$. If $\mathcal{H}$ (say) were a function solely of π and φ, then $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ which would violate the assumption that $|0\rangle$ is a nondegenerate eigenstate of $\mathcal{H}$. Consequently, $\mathcal{H}$ cannot be a function solely of π and φ. A similar argument applies to $\mathcal{P}_k$.

[4] The word "classical" as used here need only be taken to infer that the diagonal OFS matrix elements have a functional dependence on the fields f and g similar to that of the appropriate generator of the classical theory. No deterministic interpretation of the fields f and g is implied or intended; indeed, $\hbar$ is not set to zero but remains unaltered in the WCP (having the value one in the units chosen here). Further discussion of the WCP including the $\hbar$ dependence is contained in a related analysis for particle mechanics by J. R. Klauder, J. Math. Phys. **4**, 1058 (1963); **5**, 177 (1964).

Angular Momentum Operators

A similar computation determines the diagonal matrix elements

$$J_k(f, g) \equiv \langle f, g| \mathfrak{J}_k |f, g\rangle$$

for the rotation group generators. From $\mathfrak{J}_k |0\rangle = 0$ it follows that $J_k(0, 0) = 0$. Since

$$[\mathfrak{J}_k, \varphi(\mathbf{x})] = -i\epsilon_{klm}x_l \nabla_m \varphi(\mathbf{x}),$$

and similarly for $\pi(\mathbf{x})$, it follows that

$$\begin{aligned}[\delta/\delta f(\mathbf{x})]J_k(f, g) &= i\langle f, g| [\mathfrak{J}_k, \varphi(\mathbf{x})] |f, g\rangle \\ &= \epsilon_{klm}x_l \nabla_m g(\mathbf{x}),\end{aligned}$$

and

$$\begin{aligned}[\delta/\delta g(\mathbf{x})]J_k(f, g) &= -i\langle f, g| [\mathfrak{J}_k, \pi(\mathbf{x})] |f, g\rangle \\ &= -\epsilon_{klm}x_l \nabla_m f(\mathbf{x}).\end{aligned}$$

These conditions uniquely fix J_k and we find that

$$\langle f, g| \mathfrak{J}_k |f, g\rangle = J_k(f, g) = \int \epsilon_{klm} f(\mathbf{x}) x_l \nabla_m g(\mathbf{x})\, d\mathbf{x}. \tag{15}$$

As is readily seen, the J_k are just the classical generators of infinitesimal rotations as required by the WCP.[5]

Hamiltonian

We now turn our attention to the generator of time translations, the Hamiltonian $\mathfrak{H}$. We seek to study the functional

$$H(f, g) \equiv \langle f, g| \mathfrak{H} |f, g\rangle \tag{16}$$

which is evidently real and (granted suitable spectral conditions on $\mathfrak{H}$) is positive subject to the single exception $H(0, 0) = 0$ since $\mathfrak{H} |0\rangle = 0$. From the condition $[\mathfrak{H}, \varphi(\mathbf{x})] = -i\pi(\mathbf{x})$ coupled with (6a) and (4), it follows that

$$\begin{aligned}[\delta/\delta f(\mathbf{x})]H(f, g) &= i\langle f, g| [\mathfrak{H}, \varphi(\mathbf{x})] |f, g\rangle \\ &= \langle f, g| \pi(\mathbf{x}) |f, g\rangle \\ &= f(\mathbf{x}),\end{aligned}$$

since $\langle 0| \pi(\mathbf{x}) |0\rangle = i\langle 0| [\mathfrak{H}, \varphi(\mathbf{x})] |0\rangle \equiv 0$. Consequently we can put

$$H(f, g) = \frac{1}{2}\int f^2(\mathbf{x})\, d\mathbf{x} + W(g), \tag{17}$$

with $W(g) > 0$ excepting the case $g = 0$, where $W(0) = 0$. Euclidean invariance of $\mathfrak{H}$ implies that $W(g)$ is invariant under spatial rotations and translations of $g(\mathbf{x})$. Further information about W is of course difficult to deduce in general but some features may be inferred from plausible requirements on $\mathfrak{H}$.

If we introduce the notation

$$\pi(g) = \int \pi(\mathbf{x}) g(\mathbf{x}) d\mathbf{x},$$

then we may set

$$\begin{aligned}W(g) &= H(0, g) = \langle 0| e^{i\pi(g)} \mathfrak{H} e^{-i\pi(g)} |0\rangle \\ &= \sum_{n=2}^{\infty} (n!)^{-1} i^n \langle 0| [\pi(g), [\pi(g), \cdots, [\pi(g), \mathfrak{H}] \cdots] |0\rangle \\ &\equiv \sum_{n=2}^{\infty} (n!)^{-1} W_n(g) \\ &\equiv \sum_{n=2}^{\infty} (n!)^{-1} \int \cdots \int w_n(\mathbf{x}_1, \cdots, \mathbf{x}_n) \\ &\qquad \times g(\mathbf{x}_1) \cdots g(\mathbf{x}_n)\, d\mathbf{x}_1 \cdots d\mathbf{x}_n,\end{aligned} \tag{18}$$

where $W_n(g)$ is homogeneous of degree n. If $W(-g) = W(g)$ is a plausible *even symmetry*, then $W_n(g) = 0$ for all odd integral n. For a *polynomial-type interaction* the series in (18) should terminate; indeed, for example, a general φ^4-type interaction Hamiltonian should at least be sensitive to the power *four* in such a way that a *five* fold multiple commutator of $\pi(g)$ with $\mathfrak{H}$ vanishes identically. If we combine these arguments, then a φ^4-type theory should involve only $W_2(g)$ and $W_4(g)$.

In addition to limiting the number of terms which contribute to $W(g)$, some general properties of $w_n(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ may be postulated in special cases. In particular, besides Euclidean invariance of these quantities, an *essentially local* theory may be defined as one for which all nonvanishing $w_n(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ are distributions with but a single point of support at $\mathbf{x}_1 = \mathbf{x}_2 = \cdots = \mathbf{x}_n$ for all n. This is an important classification and, as we note below, it includes the relativistic interactions.

Let us examine the specific term $W_2(g)$ in somewhat greater detail. Evidently we have the several relations

$$\begin{aligned}W_2(g) &= -\langle 0| [\pi(g), [\pi(g), \mathfrak{H}]] |0\rangle \\ &= 2\langle 0| \pi(g)\mathfrak{H}\pi(g) |0\rangle \\ &= \iint w_2(\mathbf{x}_1, \mathbf{x}_2) g(\mathbf{x}_1) g(\mathbf{x}_2)\, d\mathbf{x}_1\, d\mathbf{x}_2 \\ &\equiv \int \tilde{w}_2(k) |\tilde{g}(\mathbf{k})|^2\, d\mathbf{k}.\end{aligned} \tag{19}$$

In this expression,

$$w_2(\mathbf{x}_1, \mathbf{x}_2) = 2\langle 0| \pi(\mathbf{x}_1)\mathfrak{H}\pi(\mathbf{x}_2) |0\rangle,$$

$$\tilde{g}(\mathbf{k}) = (2\pi)^{-\frac{3}{2}} \int e^{-i\mathbf{k}\cdot\mathbf{x}} g(\mathbf{x})\, d\mathbf{x};$$

[5] It is evident from the preceding examples that any "kinematic" generator (i.e., one whose commutator with an arbitrary linear sum of φ and π is again a linear sum of φ and π) which annihilates the state $|0\rangle$ will fulfill the WCP.

also

$$\tilde{w}_2(k) = \int e^{-i\mathbf{k}\cdot\mathbf{x}} w_2(\mathbf{x}, 0)\, d\mathbf{x},$$

which by rotational invariance is a function only of $k = |\mathbf{k}|$. In the case of a unique ground state, then $\tilde{w}_2(k) > 0$ (almost everywhere). For an essentially local theory, $\tilde{w}_2$ is a polynomial in k^2; for a relativistic theory we show below that $\tilde{w}_2(k) = k^2 + m_0^2$.

In the simplest of examples, $W_2(g)$ is the only nonvanishing term in (18), and a free-field Hamiltonian results. As a "collection of independent oscillators" such a free-field theory admits a straightforward irreducible quantization whose CCR representation is characterized by the fact that

$$\langle f, g \mid 0\rangle = \exp\left\{-\frac{1}{4}\int [\tilde{w}_2(k)^{-\frac{1}{2}} |\tilde{f}(\mathbf{k})|^2 + \tilde{w}_2(k)^{\frac{1}{2}} |\tilde{g}(\mathbf{k})|^2]\, d\mathbf{k}\right\}.$$

In the present case, $\mathcal{K} = :H(\pi, \varphi):$ so that the validity of the WCP follows from Eq. (9). It may also be worth noting in this case that each distinct $\tilde{w}_2(k)$ corresponds to a unitarily inequivalent CCR representation.

Concrete examples of Euclidean-invariant models with nonvanishing interaction are the "rotationally symmetric" models.[1,2] Only discrete energy levels arise in these models so no scattering takes place. Elsewhere, we will present other Euclidean-invariant models with nonvanishing interaction [e.g., $W_4(g) \neq 0$] which exhibit both scattering and production. Some of these latter examples are essentially local in the sense discussed above. Each of these models exploits a reducible CCR representation, and although they satisfy no conventional prescription like Eq. (8), they nonetheless all fulfill the WCP.

4. WEAK CORRESPONDENCE PRINCIPLE IN RELATIVISTIC INVARIANT THEORIES

We divide our relativistic discussion into two parts: the first part is an extension of the arguments in the preceding section, while the second part is a covariant reformulation of the results of the first part.

Quite clearly the previous discussion of the momentum and angular momentum generators applies in the present case. We begin by examining the Hamiltonian somewhat further.

Hamiltonian

We note initially that local commutativity on a single spacelike surface ($t = \text{const}$) requires that

$$[\pi(\mathbf{x}_1), [\pi(\mathbf{x}_2), \mathcal{K}]] = 0; \quad \mathbf{x}_1 \neq \mathbf{x}_2,$$

and similarly for additional commutators with $\pi(\mathbf{x})$. Thus relativistic fields are essentially local fields in the previous sense. We next show in a two-stage analysis that covariance arguments enable us to determine the functional form of

$$W(g) = \sum_{n=2}^{\infty} (n!)^{-1} W_n(g)$$

in Eq. (18).

We may determine the second-order term $W_2(g)$ with the aid of the Lehmann representation for the two-point function. In particular, relativistic invariance leads to the well-known expression[6]

$$\langle 0 |\, \varphi(\mathbf{x}) e^{-i\mathcal{K}t} \varphi(\mathbf{y})\, | 0\rangle = \frac{1}{2(2\pi)^3} \iint e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) - i\omega t} \times \frac{d\mathbf{k}}{\omega} \rho(m^2)\, dm^2, \quad (20)$$

where $\omega^2 = k^2 + m^2$, and (since we assume CCR)

$$\int \rho(m^2)\, dm^2 = 1; \quad \rho(m^2) \geq 0.$$

Combining the relation

$$\langle 0 |\, \varphi(\mathbf{x}) \mathcal{K}^3 \varphi(\mathbf{y})\, | 0\rangle = \langle 0 |\, [\varphi(\mathbf{x}), \mathcal{K}]\mathcal{K}[\mathcal{K}, \varphi(\mathbf{y})]\, | 0\rangle = \langle 0 |\, \pi(\mathbf{x})\mathcal{K}\pi(\mathbf{y})\, | 0\rangle = \tfrac{1}{2} w_2(\mathbf{x}, \mathbf{y})$$

with the result of three time derivatives of (20) at the origin, we find that

$$w_2(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^3} \iint e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} (k^2 + m^2)\, d\mathbf{k} \rho(m^2)\, dm^2.$$

Evidently, in the relativistic case, we may conclude that

$$\tilde{w}_2(k) = k^2 + m_0^2, \quad (21)$$

where

$$m_0^2 \equiv \int m^2 \rho(m^2)\, dm^2. \quad (22)$$

The latter expression is recognized as the usual definition of the bare mass, although it is deduced by different arguments.[7] In the present formulation, Eq. (22) becomes a consequence of the WCP. By way of summary at this point we note that (19) and (21) lead to

$$W_2(g) = \int [(\nabla g)^2 + m_0^2 g^2]\, d\mathbf{x}, \quad (23)$$

as befits the second-order contribution to a relativistic Hamiltonian.

The remainder of $W(g)$ is best treated as a unit.

[6] See, e.g., S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, Publishers, Inc., New York, 1962), p. 659.

[7] S. S. Schweber, Ref. 6, p. 667.

Let us introduce

$$V(g) \equiv W(g) - \tfrac{1}{2}W_2(g) = \sum_{n=3}^{\infty} (n!)^{-1} W_n(g) \quad (24)$$

as the classical interaction potential. When we discuss the relativity generators below we shall be able to show that $V(g)$ has a form given by

$$V(g) = \int V[g(\mathbf{x})]\, d\mathbf{x} = \sum_{n=3}^{\infty} (n!)^{-1} c_n \int g^n(\mathbf{x})\, d\mathbf{x}, \quad (25)$$

where c_n are constants.

Accepting this result for $V(g)$ temporarily, we learn in the relativistic case that the diagonal OFS matrix elements of the Hamiltonian have the form

$$\begin{aligned} H(f, g) &= \langle f, g | \mathcal{H} | f, g \rangle \\ &= \tfrac{1}{2}\int \{f^2(\mathbf{x}) + [\nabla g(\mathbf{x})]^2 + m_0^2 g^2(\mathbf{x})\}\, d\mathbf{x} \\ &\qquad + \int V[g(\mathbf{x})]\, d\mathbf{x}, \\ &\equiv \int H(\mathbf{x})\, d\mathbf{x}. \end{aligned} \quad (26)$$

Here, $H(\mathbf{x})$ is a nonnegative density defined in obvious fashion from $f(\mathbf{x})$, $\nabla g(\mathbf{x})$, and $g(\mathbf{x})$. If in addition we assume the conditions of (say) a φ^4-type theory as discussed in Sec. 3, then $H(f, g)$ would have the form (26) in which $V[g(\mathbf{x})] = \lambda g^4(\mathbf{x})$ for some positive constant λ. Evidently Eq. (26) gives just the desired family of classical relativistic Hamiltonians, and so fulfills the WCP in splendid fashion.

Relativity Transformations

We may adjoin three relativity-transformation generators $\mathcal{K}_k$, $k = 1, 2, 3$ (all defined at $t = 0$) to our Euclidean generators and Hamiltonian discussed above and make a set of Poincaré generators. The three functionals

$$K_k(f, g) = \langle f, g | \mathcal{K}_k | f, g \rangle$$

may be found as before. From $\mathcal{K}_k |0\rangle = 0$ there follows the condition $K_k(0, 0) = 0$. From

$$[\mathcal{K}_k, \varphi(\mathbf{x})] = -i[t\nabla_k \varphi(\mathbf{x}) + x_k \dot{\varphi}(\mathbf{x})],$$

we learn that

$$\begin{aligned} \frac{\delta}{\delta f(\mathbf{x})} K_k(f, g) &= i\langle f, g | [\mathcal{K}_k, \varphi(\mathbf{x})] | f, g \rangle \\ &= \langle f, g | t\nabla_k \varphi(\mathbf{x}) + i x_k [\mathcal{H}, \varphi(\mathbf{x})] | f, g \rangle \\ &= t\nabla_k g(\mathbf{x}) + x_k \frac{\delta}{\delta f(\mathbf{x})} H(f, g). \end{aligned}$$

Likewise,

$$[\mathcal{K}_k, \pi(\mathbf{x})] = -i[t\nabla_k \pi(\mathbf{x}) + \nabla_k \varphi(\mathbf{x}) + x_k \dot{\pi}(\mathbf{x})],$$

which is just the time derivative of the previous commutator, leads to

$$\begin{aligned} \frac{\delta}{\delta g(\mathbf{x})} K_k(f, g) &= -i\langle f, g | [\mathcal{K}_k, \pi(\mathbf{x})] | f, g \rangle \\ &= \langle f, g | - t\nabla_k \pi(\mathbf{x}) - \nabla_k \varphi(\mathbf{x}) \\ &\qquad - i x_k [\mathcal{H}, \pi(\mathbf{x})] | f, g \rangle \\ &= -t\nabla_k f(\mathbf{x}) - \nabla_k g(\mathbf{x}) \\ &\qquad + x_k \frac{\delta}{\delta g(\mathbf{x})} H(f, g). \end{aligned} \quad (27)$$

In terms of $H(\mathbf{x})$ [implicitly defined in (26)], the solution for the functionals K_k is readily seen to be

$$\begin{aligned} \langle f, g | \mathcal{K}_k | f, g \rangle &= K_k(f, g) \\ &= \int [f(\mathbf{x}) t\nabla_k g(\mathbf{x}) + x_k H(\mathbf{x})]\, d\mathbf{x}. \end{aligned} \quad (28)$$

This expression makes explicit use of the $(\nabla g)^2$ term in $H(\mathbf{x})$ [to win the term $-\nabla_k g(\mathbf{x})$ in (27)], and makes implicit use of the absence of any other field gradient in $H(\mathbf{x})$. Stated otherwise, if some term in $V(g)$ contained a field gradient, then the solution for $K_k(f, g)$ would differ from that in (28) over and above a mere redefinition of $H(\mathbf{x})$ to reflect the gradients in $V(g)$. However, with a standard argument we next show that only (28) can be correct thereby clinching the fact that $V(g)$ has no gradients and hence has the form stated in (25).

We recall the basic commutation relation

$$[\mathcal{K}_k, \mathcal{J}_l] = -i\delta_{kl}\mathcal{H}$$

among the Poincaré generators, which we can immediately restate as

$$e^{-ia_l \mathcal{J}_l} \mathcal{K}_k e^{ia_l \mathcal{J}_l} = \mathcal{K}_k + a_k \mathcal{H},$$

where a_k are the components of a three vector $\mathbf{a}$. Since

$$e^{ia_l \mathcal{J}_l} | f(\mathbf{x}), g(\mathbf{x}) \rangle = | f(\mathbf{x} - \mathbf{a}), g(\mathbf{x} - \mathbf{a}) \rangle,$$

it follows that $[f = f(\mathbf{x}), g = g(\mathbf{x})]$

$$\begin{aligned} K_k(f(\mathbf{x} - \mathbf{a}), g(\mathbf{x} - \mathbf{a})) &= \langle f, g | e^{-ia_l \mathcal{J}_l} \mathcal{K}_k e^{ia_l \mathcal{J}_l} | f, g \rangle \\ &= \langle f, g | (\mathcal{K}_k + a_k \mathcal{H}) | f, g \rangle \\ &= K_k(f, g) + a_k H(f, g). \end{aligned}$$

This relation is fulfilled by (28) since $H(\mathbf{x})$ is the same density involved in defining $H(f, g)$. In turn, in order for this to be true, it is required that Eq. (25) hold, thus establishing its validity. This is the same argument used to prove Eq. (25) in a strictly classical theory.

In summary, we may combine the present results with those of Sec. 3 to conclude that in a relativistic theory all ten generators of the Poincaré group fulfill the WCP for some local interaction density $V[g(\mathbf{x})]$.

Additional conditions—such as those corresponding to a φ^4-type theory—may be imposed to restrict the form of the interaction potential.

Covariant Formulation of the Weak Correspondence Principle

In this section we recast the previous results into a Lorentz covariant form. We employ a metric with signature -2, and a unit surface-normal vector n^μ and three-volume element $d\sigma^\mu$, which in an appropriate frame have vanishing spacelike components and timelike components of 1 and $d\mathbf{x}$, respectively. Let σ denote the spacelike surface which in the same appropriate Lorentz frame is just the surface $t =$ const. Then we may relate the function pair $f(\mathbf{x})$ and $g(\mathbf{x})$ used previously to a space–time function $g(x)$ through the relations

$$g(\mathbf{x}) = g(x); \quad x \in \sigma,$$
$$f(\mathbf{x}) = n^\mu \partial_\mu g(x); \quad x \in \sigma.$$

Note by this that all previously used spatial coordinates are reinterpreted as contravariant components x^k as usual. With this notation, the covariant definition of the vectors in the OFS is given by

$$|f, g\rangle \equiv \exp\left\{i \int \varphi \overleftrightarrow{\partial}_\mu g \, d\sigma^\mu\right\} |0\rangle, \tag{29}$$

which in the appropriate frame just reduces to (5).

The ten generators of the Poincaré group are given by the skew tensor $\mathcal{M}^{\alpha\beta}$ and four-momentum $\mathcal{P}^\mu$ which fulfill the standard commutation rules

$$[\mathcal{P}^\mu, \mathcal{M}^{\alpha\beta}] = -i(g^{\beta\mu}\mathcal{P}^\alpha - g^{\alpha\mu}\mathcal{P}^\beta),$$
$$[\mathcal{M}^{\alpha\beta}, \mathcal{M}^{\sigma\tau}] = -i(g^{\alpha\sigma}\mathcal{M}^{\beta\tau} - g^{\beta\sigma}\mathcal{M}^{\alpha\tau} + g^{\beta\tau}\mathcal{M}^{\alpha\sigma} - g^{\alpha\tau}\mathcal{M}^{\beta\sigma}).$$

The connection of these expressions with our previous generators is given by

$$\mathcal{P}_k = -\mathcal{P}^k, \quad \mathcal{H} = \mathcal{P}_0 = (\mathcal{P}^0), \quad \mathcal{K}_k = \mathcal{M}_{0k} \,(= \mathcal{M}^{0k}),$$
$$\mathcal{J}_k = \mathcal{M}_{lm} \,(= \mathcal{M}^{lm}); \; k, l, m = 1, 2, 3 \text{ cyclic}.$$

We may immediately generalize the diagonal OFS matrix elements to covariant form from their values in a specific frame. In particular, the covariant WCP for the ten Poincaré generators reads

$$\langle f, g| \mathcal{P}^\mu |f, g\rangle = P^\mu(f, g) = \int T^{\mu\nu}(x) \, d\sigma_\nu \tag{30}$$

and

$$\langle f, g| \mathcal{M}^{\alpha\beta} |f, g\rangle = M^{\alpha\beta}(f, g) = \int [x^\alpha T^{\beta\nu}(x) - x^\beta T^{\alpha\nu}(x)] \, d\sigma_\nu. \tag{31}$$

In these expressions, $T^{\mu\nu}(x)$ is the classical stress–energy tensor given by

$$T^{\mu\nu}(x) = \partial^\mu g(x) \partial^\nu g(x) - g^{\mu\nu} L(x),$$

where

$$L(x) \equiv \tfrac{1}{2}\{[\partial_\varkappa g(x)]^2 - m_0^2 g^2(x)\} - V[g(x)]$$

as follows from our analysis in a specific frame coupled with Lorentz covariance.

In summary, the diagonal OFS matrix elements of the ten Poincaré generators yield the ten generators of a classical covariant theory expressed in the traditional form with the aid of a stress–energy tensor.

5. SUMMARY AND DISCUSSION

In the previous sections, we have shown that diagonal matrix elements of quantum generators in the states $|f, g\rangle$ yield the appropriate classical generators as required by the weak correspondence principle. Basically, these results are only contingent on certain operator domain conditions, and are not wedded to a three-dimensionality of space nor to a complete expandability of $W(g)$ in a power series. If the canonical operators π and φ are irreducible, then the WCP is equivalent to normal ordering. However, if π and φ are reducible—independent of whether the states $|f, g\rangle$ span $\mathcal{H}$ or not—the WCP is far more general. As already noted in such a case,[3] $\mathcal{H}$ cannot be solely a function of π and φ. It must clearly be understood that, in general, this property implies that no conventional field equations (i.e., $\ddot{\varphi} = F\{\varphi\}$ for some functional F of φ at a fixed time) can hold; hence no conventional τ-function equations, nor conventional Schwinger equations can hold. These features are not related to renormalizations and are easily illustrated for generalized free fields.

The utility of the WCP stems from the constraints it places on the CCR representation and the quantum generators, constraints which can serve as a guide in formulating the quantum theory. For example, given a representation of the operators π and φ, the WCP can help to test which operators if any can serve as generators appropriate to a specified classical theory. Generally, for a given CCR representation, it happens that *no* operator can be found which fulfills the WCP for a prescribed classical Hamiltonian and different CCR representations must be tried. Once a representation and generators compatible with the WCP are found, it follows that a consistent quantum theory exists. Sometimes it ends up that there are several quantum theories compatible with the WCP as, e.g., happens with a generalized free field in which only the first moment of the spectral weight is specified classically [cf., Eq. (22)]. For the "rotationally

symmetric" models there are a two-parameter family of quantum solutions compatible with the WCP (these solutions may be labeled by the mass values of the two asymptotic "one-particle" states that arise in these models). Such nonuniqueness is by no means unexpected since the WCP, by itself, generally provides only a *partial* constraint on the quantum generator, especially if the states $|f, g\rangle$ do not even span $\mathfrak{H}$.

Because the reducible representations for π and φ which we advocate are somewhat unconventional, let us note here several features in their favor. For example, reducible representations readily take care of Haag's theorem which requires that the CCR representation of a nontrivial Euclidean-invariant theory be unitarily inequivalent to an irreducible, free-field Fock representation.[8] One fashionable irreducibility axiom is the so-called time-slice axiom, which roughly states that the operators $\varphi(\mathbf{x}, t)$ for all $\mathbf{x}$ and $|t| < \epsilon$, $\epsilon > 0$, are irreducible. From this axiom it need not follow that just $\varphi(\mathbf{x})$ and $\dot{\varphi}(\mathbf{x}) = \pi(\mathbf{x})$ at $t = 0$ (say) are irreducible. Also, since the commutation relations for fields are known to have uncountably many inequivalent irreducible representations, it is conceivable that a reducible representation could arise, loosely speaking, so as to be able to display some of this variety as so often occurs, e.g., with the rotation group.

Finally, although not directly related to this paper, it is interesting to note that recent results for relativistic Fermi fields (depending likewise on a few domain conditions) imply that a nontrivial relativistic theory fulfilling traditional anticommutation rules must necessarily employ reducible representations of the spinor field and its adjoint.[9] We hope to develop a weak correspondence principle for Euclidean- and Lorentz-invariant Fermi fields in a subsequent paper.

[8] A. S. Wightman, *Lecture Notes at the French Summer School of Theoretical Physics, Cargese Corsica, July*, 1964 (Gordon & Breach, to be published)

[9] R. T. Powers, Commun. Math. Phys. **4**, 145 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Internal Dynamics of Particlelike Solutions to Nonlinear Field Theories

GERALD ROSEN
Drexel Institute of Technology, Philadelphia, Pennsylvania

(Received 14 September 1966)

We set up a Rayleigh–Ritz procedure for determining the approximate classical and quantum internal motion of a singularity-free solution to a nonlinear field theory. The dynamical approximation procedure is applied to the particlelike solutions of a nonlinear model scalar field theory and to the particlelike solutions of a class of nonlinear model scalar–spinor theories. Dynamically stable, metastable, periodic, and unstable particlelike solutions all follow for certain ranges of the physical parameters from the classical theory, but only corresponding metastable and unstable associated quantum stationary states are obtained for the nonlinear model theories considered here. The functional dependence of the decay constant for the metastable quantum stationary states is such that very long-lived, practically stable, states are admissible. Group-theoretic techniques for the systematic derivation of rigorous particlelike solutions are also described and illustrated with examples.

I. INTRODUCTION

TIME-INDEPENDENT singularity-free particlelike solutions to model Lorentz-covariant nonlinear field theories have been the subject of recent papers.[1–3] Rigorous examples of such spatially localized solutions of finite energy are reported in these papers for certain model field theories, along with general criteria for the existence and stability of a time-independent singularity-free particlelike solution to any nonlinear field theory.[1,2] A Rayleigh–Ritz approximation theory has been formulated for deriving the motion of such a "particle" in a superimposed small-amplitude external field, that is, the over-all motion induced in a singularity-free particlelike solution by virtue of the nonlinearity of the field equations.[3]

In the present work we consider the internal dynamical theory associated with particlelike solutions, according to both classical and quantum mechanics. Our analysis is again based on a Rayleigh–Ritz approximation procedure, but in this case a procedure for calculating *internal* dynamical changes of the "particle." Theory for the internal quantum motion of the particlelike solution is then formulated in analogy to the classical Rayleigh–Ritz approximation theory by adapting and refining an idea of Petiau.[4] The latter approximation method offers a new and practical way to obtain physically reasonable quantum dynamical predictions from an essentially nonlinear field theory. In this paper, we apply the classical and quantum approximation theory to the model particlelike solutions obtained previously[1,2] and give special attention to the dynamical stability of these solutions.

[1] G. Rosen, J. Math. Phys. **6**, 1269 (1965).
[2] G. Rosen, J. Math. Phys. **7**, 2066 (1966).
[3] G. Rosen, J. Math. Phys. **8**, 573 (1967).
[4] G. Petiau, Ann. Inst. H. Poincaré **3**, 127 (1965); Nuovo Cimento **40**, 84 (1965).

II. GENERAL APPROXIMATION METHOD

Consider a Lorentz-covariant field theory with the scalar-invariant Lagrangian density

$$\mathfrak{L} = \mathfrak{L}(\phi, \dot{\phi}, \nabla\phi) \tag{2.1}$$

in which $\phi = \phi(\mathbf{x}, t)$ is a generic (multicomponent) real field and $\dot{\phi} \equiv \partial\phi/\partial t$. The action principle

$$\delta \int L\, dt = 0, \quad L \equiv \int \mathfrak{L}\, d^3\mathbf{x} \tag{2.2}$$

produces the field equations

$$\frac{\partial \mathfrak{L}}{\partial \phi} - \frac{\partial}{\partial t}\left(\frac{\partial \mathfrak{L}}{\partial \dot{\phi}}\right) - \nabla \cdot \left(\frac{\partial \mathfrak{L}}{\partial(\nabla\phi)}\right) = 0. \tag{2.3}$$

With the assumption that the classical field equations (2.3) admit an approximate solution of the form $\phi = \bar{\phi}(\mathbf{x}; \xi_1, \cdots, \xi_N)$, where the ξ's are certain functions of t, it follows that the functions $\xi_k = \xi_k(t)$ must satisfy the Rayleigh–Ritz equations derived from Eq. (2.2),

$$\frac{\partial L}{\partial \xi_k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\xi}_k}\right) = 0, \quad [k = 1, \cdots, N], \tag{2.4}$$

equations of the Euler–Lagrange form with the Lagrangian given by

$$L = \int \mathfrak{L}(\bar{\phi}, \dot{\bar{\phi}}, \nabla\bar{\phi})\, d^3\mathbf{x} = L(\xi_1, \cdots, \xi_N, \dot{\xi}_1, \cdots, \dot{\xi}_N). \tag{2.5}$$

Modulo an additive constant volume integral which does not involve the ξ's or the $\dot{\xi}$'s, the Lagrangian (2.5) must exist as a well-defined (finite) function for applicability of the Rayleigh–Ritz equations (2.4), requiring a singularity-free (integrable) $\bar{\phi}$ and conditioning a suitable functional dependence of $\bar{\phi}$ on the ξ's. Approximate dynamical solutions to the field

theory are then obtainable by integrating the Euler–Lagrange equations (2.4). Moreover, corresponding and approximate quantum states are also obtainable by an elementary adaptation[3,4] of the preceding classical method, the Lagrangian (2.5) describing the physical system accurately to within the approximation of N dynamical degrees of freedom. Associated with (2.5) is the Hamiltonian

$$H = H(\xi_1, \cdots, \xi_N, \eta_1, \cdots, \eta_N) \equiv \sum_{k=1}^{N} \dot{\xi}_k \eta_k - L, \quad \eta_k \equiv \frac{\partial L}{\partial \dot{\xi}_k}, \tag{2.6}$$

and standard canonical quantization leads to the Schrödinger equation

$$i\hbar \partial \psi / \partial t = H(\xi_1, \cdots, \xi_N, -i\hbar \partial/\partial \xi_1, \cdots, -i\hbar \partial/\partial \xi_N)\psi \tag{2.7}$$

for the wavefunction $\psi = \psi(\xi_1, \cdots, \xi_N; t)$. Again, as in the approximate classical dynamical theory, the Lagrangian (2.5) must exist as a well-defined (finite) function of the ξ's and $\dot{\xi}$'s (modulo an additive constant volume integral) for applicability of the method. Then by integrating the Schrödinger equation (2.7), quantum states are obtained for the reduced field theory, approximated as a system with N dynamical degrees of freedom.

To illustrate the classical and quantum approximation methods discussed above, let us consider the theory based on the Lagrangian density

$$\mathfrak{L} = (\dot{\phi})^2 - (\nabla \phi)^2 - \mu^2 \phi^2 \tag{2.8}$$

for a real scalar meson field $\phi = \phi(\mathbf{x}, t)$ of constant positive mass $\hbar\mu$. We seek approximate solutions[5] to the associated linear field equation in the form of plane waves,

$$\phi = \bar{\phi} \equiv \begin{cases} \xi \sin \mathbf{k} \cdot \mathbf{x} & \text{for} \quad \mathbf{x} \text{ in } V \\ 0 & \text{for} \quad \mathbf{x} \text{ not in } V \end{cases} \tag{2.9}$$

with $\mathbf{k}$ a constant and $\xi = \xi(t)$. With Eqs. (2.8) and (2.9) the Lagrangian (2.5) is evaluated as

$$L = \tfrac{1}{2}[\dot{\xi}^2 - (\mathbf{k}^2 + \mu^2)\xi^2]V \tag{2.10}$$

for $V \gg |\mathbf{k}|^{-3}$. Hence, the classical dynamical equation (2.4)

$$\ddot{\xi} + (\mathbf{k}^2 + \mu^2)\xi = 0 \tag{2.11}$$

is an exact relation for this simple example. Likewise, the quantum dynamical equation (2.7)

$$i\hbar \partial \psi / \partial t = \left[-\frac{\hbar^2}{2V} \frac{\partial^2}{\partial \xi^2} + \frac{V}{2}(\mathbf{k}^2 + \mu^2)\xi^2 \right] \psi \tag{2.12}$$

is a precise relation, yielding the exact energy eigenvalues and eigenstates for the quantized simple harmonic oscillator associated with the plane wave (2.9) in the quantum field theory for (2.8).

Another illustration of the approximation methods, a very simple nonlinear example, is provided by the spatially homogeneous solutions to the generic nonlinear field theory with (2.1), solutions of the form

$$\phi = \bar{\phi} \equiv \begin{cases} \xi = \xi(t) & \text{for} \quad \mathbf{x} \text{ in } V \\ 0 & \text{for} \quad \mathbf{x} \text{ not in } V. \end{cases} \tag{2.13}$$

The Lagrangian (2.5) is

$$L = \mathfrak{L}(\xi, \dot{\xi}, \mathbf{0})V, \tag{2.14}$$

and hence the classical dynamical equation (2.4)

$$\frac{\partial \mathfrak{L}}{\partial \xi} - \frac{d}{dt}\left(\frac{\partial \mathfrak{L}}{\partial \dot{\xi}}\right) = 0 \tag{2.15}$$

is in fact the exact relation obtained by putting Eq. (2.13) into the field equations (2.3). Furthermore, the quantum dynamical equation (2.7) obtained from Eq. (2.14) is a physically reasonable one-degree-of-freedom correspondent to the Schrödinger functional differential equation for the quantum field theory.[6]

III. DYNAMICS OF PARTICLELIKE SOLUTIONS TO A NONLINEAR MODEL SCALAR THEORY

In this section, we use the Rayleigh–Ritz approximation method to analyze the internal dynamics of particlelike solutions to the solvable nonlinear model theory based on the Lagrangian density[1]

$$\mathfrak{L} = (\dot{\theta})^2 - (\nabla \theta)^2 + g\theta^6 \tag{3.1}$$

with θ a real scalar field and g a positive physical constant. The associated field equation

$$\ddot{\theta} - \nabla^2 \theta - 3g\theta^5 = 0 \tag{3.2}$$

is satisfied rigorously by the singularity-free spherically symmetric static particlelike solutions

$$\theta = \theta_0(\mathbf{x}) \equiv Z(Z^4 g + |\mathbf{x}|^2)^{-\frac{1}{2}} \tag{3.3}$$

in which the "size parameter" Z is a free nonzero

[5] It is important to note that the approximate solution must correspond qualitatively to a rigorous solution of the field theory, for otherwise false results follow. This is exemplified by the spurious approximate particlelike "solutions" to the linear field equation derived from (2.8), say approximate "solutions" of the form

$$\phi = \bar{\phi} \equiv (\text{const})\, \xi^{-1} \exp -|\mathbf{x}|/\xi$$

with $\xi = \xi(t)$; formal application of the method leads to bounded periodic classical dynamics for ξ [whereas spatially localized and temporally periodic solutions to the theory based on Eq. (2.8) are precluded rigorously] and to similarly false results for the associated quantum states.

[6] G. Rosen, Phys. Rev. Letters **16**, 704 (1966); Phys. Rev. **160**, 1278 (1967).

constant. By direct integration, the total field energy or classical "particle rest mass" is obtained as

$$\int [(\nabla\theta_0)^2 - g\theta_0^6]\, d^3\mathbf{x} = \frac{\pi^2}{2g^{\frac{1}{2}}} \equiv m_0, \tag{3.4}$$

a quantity independent of Z. We give a systematic derivation of the static particlelike solutions (3.3) in Appendix A and a systematic derivation of the time-dependent self-similar particlelike solutions to the field equation (3.2) in Appendix B, employing appropriate group-theoretic techniques in order to obtain these exact solutions.

Now consider an approximate solution to Eq. (3.2) of a form that generalizes (3.3),

$$\theta = \bar{\theta} \equiv Z(\xi^2 + |\mathbf{x}|^2)^{-\frac{1}{2}}, \tag{3.5}$$

in which the "effective particle radius" $\xi = \xi(t)$ is introduced as a positive function of time. By putting (3.5) into (3.1), the Lagrangian density is expressed as

$$\mathfrak{L} = [Z^2(\xi^2\dot{\xi}^2 - |\mathbf{x}|^2) + Z^6 g](\xi^2 + |\mathbf{x}|^2)^{-3} \tag{3.6}$$

and thus the Lagrangian

$$L = \frac{(\pi Z)^2}{4}\left(\frac{\dot{\xi}^2}{\xi} - \frac{3}{\xi} + \frac{Z^4 g}{\xi^3}\right) \tag{3.7}$$

is obtained by integration over all $\mathbf{x}$. From Eq. (3.7) it follows that the classical dynamical equation (2.4) is in this instance

$$2\xi\ddot{\xi} - \dot{\xi}^2 + 3\left(\frac{Z^4 g}{\xi^2} - 1\right) = 0 \tag{3.8}$$

with the unstable equilibrium value $\xi \equiv Z^2 g^{\frac{1}{2}}$ evidently a solution, producing (3.3) from (3.5). More generally, the energy integral of (3.8)

$$\frac{(\pi Z)^2}{4}\left[\frac{\dot{\xi}^2}{\xi} + \frac{3}{\xi} - \frac{Z^4 g}{\xi^3}\right] = m \equiv \text{const} \tag{3.9}$$

shows that ξ eventually goes to either zero or infinity as t increases and that no periodic classical motion is possible.[7]

With the simplest Hermitian ordering of the momentum term, (3.7) yields the Schrödinger equation

$$i\hbar\partial\psi/\partial t = \left[-\frac{\hbar^2}{(\pi Z)^2}\frac{\partial}{\partial\xi}\,\xi\,\frac{\partial}{\partial\xi} + \frac{(\pi Z)^2}{4}\left(\frac{3}{\xi} - \frac{Z^4 g}{\xi^3}\right)\right]\psi. \tag{3.10}$$

We have the boundary conditions $\psi = 0$ at $\xi = 0$ and $\xi = \infty$ since negative and infinite values of ξ are excluded. Stationary quantum states derived with (3.10) are of the form

$$\psi = \omega(\xi; m)e^{-imt/\hbar}, \tag{3.11}$$

where m denotes the eigenvalue of total energy or particle rest mass and $\omega(\xi; m)$ satisfies the equation

$$\left[\frac{d^2}{d\xi^2} + \frac{1}{\xi}\frac{d}{d\xi} - \frac{(\pi Z)^4}{4\hbar^2}\left(\frac{3}{\xi^2} - \frac{Z^4 g}{\xi^4}\right) + \frac{(\pi Z)^2 m}{\hbar^2\xi}\right]\omega(\xi; m) = 0 \tag{3.12}$$

for $0 \leq \xi < \infty$, subject to the boundary conditions $\omega(0; m) = \omega(\infty; m) = 0$. An acceptable physical interpretation is not available for states with m negative in the context of the closed theory being studied, and therefore such states are excluded from consideration here.[8] The special academic solution to (3.12) with $m = 0$ is given exactly by

$$\omega(\xi; 0) = (\text{const})\, J_{\gamma/2}(\pi^2 Z^4 g^{\frac{1}{2}}/2\hbar\xi), \qquad \gamma \equiv \sqrt{3}\,(\pi Z)^2/\hbar, \tag{3.13}$$

that is, in terms of a Bessel function of order $\frac{1}{2}\gamma$. Since Eq. (3.12) implies the asymptotic behavior $\omega(\xi; 0) \cong (\text{const})\xi^{-\gamma/2}$ for $\xi \gg (\pi^2 Z^4 g^{\frac{1}{2}}/2\hbar)$, it follows that Eq. (3.13) is in $L^2(0, \infty)$ with the Hilbert norm $\|\omega\| \equiv (\int_0^\infty |\omega|^2\, d\xi)^{\frac{1}{2}}$ [concomitant with the Hermitian ordering in (3.10)] provided that $\gamma > 1$. For the cases with $m \geq m_0 \equiv \pi^2/2g^{\frac{1}{2}}$, Eq. (3.12) admits free waves and short-lived resonances, quantum states of no practical interest in the present theory. Turning to the physically interesting cases with m greater than zero but less than m_0, we find that the wavefunction is contained in the region $0 < \xi \lesssim \xi'$ by a "potential barrier" in the region $\xi' \lesssim \xi \lesssim \xi''$, where

$$\xi' \equiv \frac{1}{\sqrt{3}}\, Z^2 g^{\frac{1}{2}} \quad \text{and} \quad \xi'' \equiv 3(\pi Z)^2/4m. \tag{3.14}$$

The physically admissible solutions to (3.12) with $0 < m < m_0$ are thus *not* in $L^2(0, \infty)$, having the asymptotic behavior

$$\omega(\xi; m) \cong \begin{cases} (\text{const})\, J_{\gamma/2}(\pi^2 Z^4 g^{\frac{1}{2}}/2\hbar\xi) & \text{for } \xi \lesssim \xi'' \\ (\text{const})\, \xi^{-\frac{1}{4}} \sin\left[(2\pi Z(m\xi)^{\frac{1}{2}}/\hbar) + (\text{const})\right] & \text{for } \xi \gtrsim \xi'', \end{cases} \tag{3.15}$$

and hence there are no absolutely stable "bound" stationary states with m positive. However, for values of γ greater than unity metastable states are obtainable

[7] A spatially localized and temporally periodic solution is not precluded by the recent theorem [G. Rosen, J. Math. Phys. 7, 2071 (1966)], because the energy density associated with (3.1) is not positive definite.

[8] Since Eq. (3.10) is dominated by an effective ξ^{-4} "attractive potential" about $\xi = 0$, there are an infinite number of purely formal (unphysical) "bound" states with the total energy eigenvalue m negative [e.g., R. M. Spector, J. Math. Phys. 5, 1185 (1964)].

by matching the asymptotic forms (3.15) at $\xi = \xi''$ and by taking a value of m for which $|\omega(\xi; m)|^2$ is relatively small [although only asymptotic to (const) $\xi^{-\frac{1}{2}}$] for values of $\xi > \xi''$. We estimate λ, the decay constant or characteristic rate of exponential dissolution of a metastable state, by means of a Laue-type formula familiar from the elementary theory of alpha disintegration,

$$\lambda \cong (m/\hbar)\,|\omega(\xi''; m)/\omega(\xi'; m)|^2$$
$$\equiv (\text{FREQUENCY OF ASSAULT ON THE "POTENTIAL BARRIER"}) \times (\text{PROBABILITY OF PENETRATION}). \quad (3.16)$$

Thus, using the first asymptotic expression in (3.15) and the values (3.14), we have the approximation

$$\lambda \cong (m/\hbar)[J_{\gamma/2}(2Z^2 g^{\frac{1}{2}} m/3\hbar)/J_{\gamma/2}(\sqrt{3}\,(\pi Z)^2/2\hbar)]^2$$
$$= (m/\hbar)[J_{\gamma/2}(\gamma m/3\sqrt{3}\,m_0)/J_{\gamma/2}(\gamma/2)]^2. \quad (3.17)$$

Now with $m < m_0$ the finite zeros of the Bessel function of order $\frac{1}{2}\gamma$ are not accessible to the argument $\gamma m/3\sqrt{3}\,m_0$ in Eq. (3.17), and so the mass eigenvalues for physically admissible metastable states comprise a continuum of positive $m \ll m_0$ for which (3.17) reduces to the simple expression

$$\lambda \cong (m/\hbar)(2m/3\sqrt{3}\,m_0)^{\gamma}. \quad (3.18)$$

Hence, in the semiclassical range $\gamma \gg 1$ the metastable states are characterized by a very small decay constant; the classical theory[1] only admits $m = m_0$ and gives $\lambda \cong 6.6(m_0/\hbar)\gamma^{-1}$.

In summary, the metastable quantum states associated with particlelike solutions to the nonlinear model scalar theory based on (3.1) correspond qualitatively to the metastable particlelike solutions of the classical theory with Z^2 large. The principal effect of quantization is to suppress the decay rate substantially for large values of the dimensionless parameter $\gamma \equiv \sqrt{3}\,(\pi Z)^2/\hbar$. With the admissibility of a positive continuum of values for the eigenvalue mass $m \ll m_0$, such metastable quantum states are not in the appropriate $L^2(0, \infty)$, and no absolutely stable "bound" stationary states are realized by the theory. Extended analysis shows that these conclusions hold true for any alternative Hermitian ordering of the momentum term in Eq. (3.10), with or without an alternative Hilbert-norm definition (concomitant with certain Hermitian orderings). Further analysis demonstrates that quantization may enhance metastability but does not give rise to absolute stability for other approximate particlelike solutions which differ from Eq. (3.5) in functional form.[9] In the following section we investigate the approximate effects of quantization on the particlelike solutions to a class of nonlinear model *multicomponent* field theories, some of the latter particlelike solutions exhibiting absolute stability on the level of the classical theory.

IV. DYNAMICS OF PARTICLELIKE SOLUTIONS TO NONLINEAR MODEL SCALAR-SPINOR THEORIES

We now consider a class of more complicated nonlinear model theories based on the generic Lagrangian density[2]

$$\mathfrak{L} = (\dot\theta)^2 - (\nabla\theta)^2 + \frac{i}{2}(\psi^\dagger\dot\psi - \dot\psi^\dagger\psi) + i\psi^\dagger\boldsymbol{\sigma}\cdot\nabla\psi + G\theta^2\psi^\dagger\psi \quad (4.1)$$

with θ a real scalar field, ψ a two-component complex Weyl spinor field ($\psi^\dagger$ its Hermitian adjoint, $\boldsymbol{\sigma}$ the Pauli matrices), and $G = G(\kappa)$ an arbitrary differentiable real function of the positive invariant scalar $\kappa \equiv \psi^\dagger\psi/\theta^4$. The associated field equations admit rigorous singularity-free spherically symmetric static solutions of the form

$$\theta = \theta_0(\mathbf{x}) \equiv \pm(3aH^{-1})^{\frac{1}{2}}(a^2 + |\mathbf{x}|^2)^{-\frac{1}{2}},$$
$$\psi = \psi_0(\mathbf{x}) \equiv 3aH^{-1}\kappa^{\frac{1}{2}}(a^2 + |\mathbf{x}|^2)^{-\frac{3}{2}}(a + i\boldsymbol{\sigma}\cdot\mathbf{x})u,$$
$$H = H(\kappa) \equiv G + \kappa\frac{dG}{d\kappa}, \quad (4.2)$$

where u is a constant Weyl spinor normalized to unity, $u^\dagger u = 1$, κ is a constant positive root of the equation

$$H^2 + 3\kappa(2H - 3G) = 0 \quad (4.3)$$

for which $H \neq 0$, and the "size parameter" a in Eq. (4.2) is a free nonzero constant with the sign of H, so that (aH^{-1}) is a positive quantity. The total field energy or classical "particle rest mass" is obtained as

$$\int[(\nabla\theta_0)^2 - i\psi_0^\dagger\boldsymbol{\sigma}\cdot\nabla\psi_0 - G\theta_0^2\psi_0^\dagger\psi_0]\,d^3\mathbf{x} = \tfrac{3}{2}\pi^2(1 + \tfrac{3}{2}\kappa H^{-1})\,|H^{-1}| \equiv m_0, \quad (4.4)$$

a quantity independent of the size parameter a. If the

[9] For instance, particlelike solutions of the spherically symmetric self-similar form (see Appendix B) $\theta = \xi^{-\frac{1}{2}}\chi(|\mathbf{x}|/\xi)$, in which χ is a prescribed function, give $L = \frac{1}{2}m_*\dot\xi^2 - m_0$,

$$m_* \equiv \tfrac{1}{2}\int_0^\infty [2\lambda\chi'(\lambda) + \chi(\lambda)]^2 4\pi\lambda^2\,d\lambda,$$
$$m_0 \equiv \int_0^\infty [\chi'(\lambda)^2 - g\chi(\lambda)^6]4\pi\lambda^2\,d\lambda.$$

It follows that admissible stationary states with $\psi = 0$ at $\xi = 0$ are sinusoidal in ξ and therefore not in $L^2(0, \infty)$.

root of (4.3) is such that $-3\kappa < H < -3\kappa/2$, or equivalently such that $-\kappa < G < -3\kappa/4$, then the particlelike solutions (4.2) are energetically stable with respect to changes in their functional form induced by infinitesimal spatial dilatations.

A dynamical generalization of (4.2) is given by

$$\theta = \pm(3aH^{-1})^{\frac{1}{2}}(\xi^2 + |\mathbf{x}|^2)^{-\frac{1}{2}},$$
$$\psi = 3\kappa^{\frac{1}{2}}H^{-1}(\xi^2 + |\mathbf{x}|^2)^{-\frac{3}{2}}(|a|\,\xi + ia\boldsymbol{\sigma}\cdot\mathbf{x})u, \tag{4.5}$$

where the "effective particle radius" $\xi = \xi(t)$ is introduced as a positive function of time and the quantity $\kappa \equiv \psi^\dagger\psi/\theta^4$ is taken to be a constant positive root of Eq. (4.3), as in the case of the rigorous static solutions (4.2). We evaluate the Lagrangian density explicitly by putting (4.5) into (4.1),

$$\mathfrak{L} = 3aH^{-1}[\xi^2\dot{\xi}^2 - |\mathbf{x}|^2 + 3\,|a|\,\kappa H^{-1} \times (u^\dagger\boldsymbol{\sigma}\cdot\mathbf{x}u\dot{\xi} - 3\xi) + 9a^2H^{-2}\kappa G] \times (\xi^2 + |\mathbf{x}|^2)^{-3}, \tag{4.6}$$

and the Lagrangian

$$L = \frac{3\pi^2aH^{-1}}{4}\left(\frac{\dot{\xi}^2}{\xi} - \frac{3}{\xi} + \frac{a^2}{\xi^3} + 3H^{-1}\kappa\left(\frac{2a^2}{\xi^3} - \frac{3\,|a|}{\xi^2}\right)\right) \tag{4.7}$$

follows by integrating (4.6) over all $\mathbf{x}$ and using (4.3) to eliminate G. In this instance the associated classical dynamical equation (2.4) takes the form

$$2\xi\ddot{\xi} - \dot{\xi}^2 + 3\left(\frac{a^2}{\xi^2} - 1\right) + 18H^{-1}\kappa\left(\frac{a^2}{\xi^2} - \frac{|a|}{\xi}\right) = 0 \tag{4.8}$$

with the equilibrium value $\xi \equiv |a|$ evidently a solution, producing (4.2) from (4.5). Elementary analysis applied to (4.8) indicates that the $\xi \equiv |a|$ solution is dynamically stable for $-3\kappa < H < 0$,[10] and furthermore if the latter condition is satisfied by H at the root of (4.3), then (4.8) also admits the unstable equilibrium solution $\xi \equiv (6\,|H^{-1}|\,\kappa - 1)\,|a|$. For $-6\kappa < H < -3\kappa$, Eq. (4.8) admits the *stable* equilibrium solution $\xi \equiv (6\,|H^{-1}|\,\kappa - 1)\,|a|$ as well as the *unstable* equilibrium solution $\xi \equiv |a|$. Moreover, the general energy integral

$$\frac{3\pi^2aH^{-1}}{4}\left[\frac{\dot{\xi}^2}{\xi} + \frac{3}{\xi} - \frac{a^2}{\xi^3} + 3H^{-1}\kappa\left(\frac{3\,|a|}{\xi^2} - \frac{2a^2}{\xi^3}\right)\right] = m \equiv \text{const} \tag{4.9}$$

shows that periodic classical motion about the stable equilibrium $\xi = |a|$ for $-3\kappa < H < 0$ is possible with m positive if $m' < m < m''$, where

$$m' \equiv \tfrac{3}{4}\pi^2\,|H^{-1}|\,(2 - 3\,|H^{-1}|\,\kappa)$$

and

$$m'' \equiv \tfrac{3}{4}\pi^2\,|H^{-1}|\,(9\,|H^{-1}|\,\kappa - 2)(6\,|H^{-1}|\,\kappa - 1)^{-2};$$

as $H \to -3\kappa$ this bounding condition on m reduces to $\pi^2/4\kappa < m < \pi^2/4\kappa$, thereby precluding periodic motion, while as $H \to 0$, the bounding condition relaxes to $m < 3\pi^2/16\kappa$. The more stringent bounding condition with m' and m'' interchanged, $m'' < m < m'$, follows from (4.9) for periodic motion about the stable equilibrium $\xi = (6\,|H^{-1}|\,\kappa - 1)\,|a|$ for $-6\kappa < H < -3\kappa$. In all other cases ξ eventually goes to either zero or infinity as t increases, *a fortiori* and irrespective of the value of m if either $H \leq -6\kappa$ or $H > 0$.

With the simplest Hermitian ordering of the momentum term, the Schrödinger equation obtained from (4.7) admits stationary quantum states of the form

$$\psi = \omega(\xi; m)e^{-imt/\hbar}, \tag{4.10}$$

where m denotes the eigenvalue of total energy or particle rest mass and $\omega(\xi; m)$ satisfies the equatior

$$\left\{\frac{d^2}{d\xi^2} + \frac{1}{\xi}\frac{d}{d\xi} - \left(\frac{3\pi^2a}{2\hbar H}\right)^2\left[\frac{3}{\xi^2} - \frac{a^2}{\xi^4} + 3H^{-1}\kappa\left(\frac{3\,|a|}{\xi^3} - \frac{2a^2}{\xi^4}\right)\right] + \frac{3\pi^2am}{\hbar^2H\xi}\right\}\omega(\xi; m) = 0, \tag{4.11}$$

for $0 \leq \xi < \infty$, subject to the boundary conditions $\omega(0; m) = \omega(\infty; m) = 0$. For small values of ξ physically admissible solutions to (4.11) have the asymptotic behavior

$$\omega(\xi; m) \cong (\text{const})\begin{cases} \xi^{\frac{1}{2}}\sin\,[(H^{-1}\kappa + \frac{1}{6})^{\frac{1}{2}}(\gamma\,|a|/\sqrt{2}\,\xi) + (\text{const})] \\ \qquad\qquad \text{for} \quad H^{-1}\kappa > -\tfrac{1}{6} \\ J_\gamma(\gamma(|a|/2\xi)^{\frac{1}{2}}) \quad \text{for} \quad H^{-1}\kappa = -\tfrac{1}{6} \\ \xi^{\frac{1}{2}}\exp\,[-(|H^{-1}|\,\kappa - \tfrac{1}{6})(\gamma\,|a|/\sqrt{2}\,\xi)] \\ \qquad\qquad \text{for} \quad H^{-1}\kappa < -\tfrac{1}{6}, \end{cases} \tag{4.12}$$

where $\gamma \equiv 3\sqrt{3}\,\pi^2aH^{-1}/\hbar$. On the other hand, for large values of ξ, physically admissible solutions to Eq. (4.11) with $m > 0$ take the asymptotic form

$$\omega(\xi; m) \cong (\text{const})\ \xi^{-\frac{1}{4}}\sin\,[(2\pi(3aH^{-1}m\xi)^{\frac{1}{2}}/\hbar) + (\text{const})], \tag{4.13}$$

and therefore the physically admissible solutions are not in $L^2(0, \infty)$. Hence, there are no absolutely stable

[10] Energetic stability of the particlelike solution also requires the condition $0 > H > -3\kappa$, and the agreement here for the energetic and dynamical stability conditions lends support to the conjectured general equivalence of the stability criteria (Ref. 2).

"bound" stationary quantum states with m positive, notwithstanding the fact that some stable particlelike solutions appear on the level of the classical theory. Metastable states are obtainable, however, for values of γ greater than unity if $H^{-1}\kappa \geq -\frac{1}{6}$, that is if either $H \leq -6\kappa$ or $H > 0$, and for positive mass eigenvalues

$$m \ll \min\{3\pi^2/4\kappa,\ 9\pi^2\,|H^{-1}|/4(\tfrac{1}{3} + 2H^{-1}\kappa)^{\frac{1}{2}}\}$$

[mass eigenvalues generally very small compared to the classical quantity (4.4) with $H^{-1}\kappa \geq -\frac{1}{6}$, as shown by simple analysis]. With the latter conditions satisfied, the wavefunction is contained in the region $0 \precsim \xi \precsim \xi'$ by a "potential barrier" in the region $\xi' \precsim \xi \precsim \xi''$, where

$$\xi' \equiv \max\{3aH^{-1}\kappa,\ (\tfrac{1}{3} + 2H^{-1}\kappa)^{\frac{1}{2}}\,|a|\}$$
and (4.14)
$$\xi'' \equiv 9\pi^2 aH^{-1}/4m,$$

the solutions to Eq. (4.11) decreasing like $\omega(\xi; m) \cong (\text{const})\xi^{-\gamma/2}$ in the region of the "potential barrier" $\xi' \precsim \xi \precsim \xi''$. The decay constant or characteristic rate of exponential dissolution of such a metastable state is estimated by the Laue-type formula

$$\begin{aligned}\lambda &\cong (m/\hbar)\,|\omega(\xi''; m)/\omega(\xi'; m)|^2 \\ &\cong (m/\hbar)(\xi'/\xi'')^{\gamma} \\ &= \max\{(m/\hbar)(4\kappa m/3\pi^2)^{\gamma}, \\ &\qquad (m/\hbar)[4(\tfrac{1}{3} + 2H^{-1}\kappa)^{\frac{1}{2}}m/9\pi^2\,|H^{-1}|]^{\gamma}\},\end{aligned} \quad (4.15)$$

which shows that in the semiclassical range $\gamma \gg 1$, the metastable states are characterized by a very small decay constant.

In summary, the quantization of particlelike solutions to the nonlinear model scalar–spinor theories based on (4.1) produces metastable stationary states but no absolutely stable "bound" stationary states. The metastable states are obtained for values of κ [a positive root of (4.3)] such that $H^{-1}\kappa \geq -\frac{1}{6}$, positive values of the eigenvalue mass

$$m \ll \min\{3\pi^2/4\kappa,\ 9\pi^2\,|H^{-1}|/4(\tfrac{1}{3} + 2H^{-1}\kappa)^{\frac{1}{2}}\},$$

and values of the dimensionless parameter $\gamma \equiv 3\sqrt{3}\,\pi^2 aH^{-1}/\hbar$ greater than unity, large values of γ giving a very small decay rate. It is interesting that all of the quantum states associated with absolutely stable particlelike solutions in the classical theory, solutions for which $H^{-1}\kappa < -\frac{1}{6}$, are unstable, quantization effecting the rapid dissolution of such particlelike concentrations of field energy. Of course other nonlinear field theories, perhaps of a more physical and less academic character, may admit absolutely stable stationary quantum states associated with particlelike solutions.

ACKNOWLEDGMENT

This work was supported in part by a National Science Foundation grant.

APPENDIX A: CONJUGATION GROUP INVARIANCE AND RIGOROUS SINGULARITY-FREE STATIC PARTICLELIKE SOLUTIONS

Let us consider the time-independent solutions $\theta = \theta_0(\mathbf{x})$ to the field equation (3.2), the solutions which satisfy

$$\nabla^2\theta_0 + 3g\theta_0^5 = 0. \quad (A1)$$

Associated with any $\theta_0(\mathbf{x})$ is a *conjugate solution*

$$\bar{\theta}_0 = \bar{\theta}_0(\mathbf{x}) \equiv \frac{\alpha}{|\mathbf{x}|}\,\theta_0(\bar{\mathbf{x}}), \quad \bar{\mathbf{x}} \equiv \alpha^2\mathbf{x}/|\mathbf{x}|^2 \quad (A2)$$

in which α is a fixed nonzero constant, because the operator relation

$$\begin{aligned}\nabla^2 &\equiv \frac{\partial^2}{\partial x_i \partial x_i} = \frac{|\bar{\mathbf{x}}|^4}{\alpha^4}\frac{\partial^2}{\partial \bar{x}_i \partial \bar{x}_i} - \frac{2\,|\bar{\mathbf{x}}|^2}{\alpha^4}\,\bar{x}_i\frac{\partial}{\partial \bar{x}_i} \\ &\equiv \frac{|\bar{\mathbf{x}}|^5}{\alpha^4}\frac{\partial^2}{\partial \bar{x}_i \partial \bar{x}_i}\cdot\frac{1}{|\bar{\mathbf{x}}|}\end{aligned} \quad (A3)$$

implies that Eq. (A2) also satisfies Eq. (A1):

$$\begin{aligned}\nabla^2\bar{\theta}_0 &= \frac{|\bar{\mathbf{x}}|^5}{\alpha^4}\frac{\partial^2}{\partial \bar{x}_i \partial \bar{x}_i}\cdot\frac{1}{|\bar{\mathbf{x}}|}\cdot\frac{\alpha}{|\mathbf{x}|}\,\theta_0(\bar{\mathbf{x}}) \\ &= \frac{|\bar{\mathbf{x}}|^5}{\alpha^5}\cdot\frac{\partial^2\theta_0(\bar{\mathbf{x}})}{\partial \bar{x}_i \partial \bar{x}_i} \\ &= -3g\,\frac{|\bar{\mathbf{x}}|^5}{\alpha^5}\,\theta_0(\bar{\mathbf{x}})^5 \equiv -3g\bar{\theta}_0^5.\end{aligned} \quad (A4)$$

From Eq. (A2), it follows that the conjugation operation squares to the identity $\bar{\bar{\theta}}_0 \equiv \theta_0$, and so the identity operator and conjugation operation are representative elements for the discrete group of order two for any fixed (nonzero) value of α. A solution of Eq. (A1) is *self-conjugate* or invariant with respect to this conjugation group if $\bar{\theta}_0 = \theta_0$, that is, if

$$\frac{\alpha}{|\mathbf{x}|}\,\theta_0\left(\frac{\alpha^2\mathbf{x}}{|\mathbf{x}|^2}\right) = \theta_0(\mathbf{x}). \quad (A5)$$

The fact that Eq. (A1) implies Eq. (A4) guarantees the existence of self-conjugate solutions with the functional property (A5). Moreover, it is practical to seek rigorous *closed-form* self-conjugate solutions to (A1).

To illustrate the last remark and demonstrate the utility of the condition (A5), we derive the self-conjugate spherically symmetric solutions to Eq. (A1), necessarily of the functional form

$$\theta_0(\mathbf{x}) = |\mathbf{x}|^{-\frac{1}{2}}\tau(\eta), \quad \eta \equiv \frac{|\mathbf{x}|}{\alpha} + \frac{\alpha}{|\mathbf{x}|}, \quad [\alpha > 0], \quad (A6)$$

where $\tau(\eta)$ is a function of the conjugation-invariant η. By putting (A6) into (A1), we obtain a total differential equation for $\tau(\eta)$,

$$(\eta^2 - 4)\frac{d^2\tau}{d\eta^2} + \eta\frac{d\tau}{d\eta} - \tfrac{1}{4}\tau + 3g\tau^5 = 0 \quad \text{(A7)}$$

with the immediate first integral

$$(\eta^2 - 4)\left(\frac{d\tau}{d\eta}\right)^2 - \tfrac{1}{4}\tau^2 + g\tau^6 = \text{const.} \quad \text{(A8)}$$

For a finite solution (A6) at $\mathbf{x} = 0$, we have $\tau(\infty) = 0$, and thus the constant of integration on the right side of (A8) vanishes. Then the final quadrature is performed easily to yield

$$\tau(\eta) = \pm g^{-\frac{1}{4}}[(\cosh C)\eta + (\sinh C)(\eta^2 - 4)^{\frac{1}{2}}]^{-\frac{1}{2}}, \quad \text{(A9)}$$

where C is a constant of integration. The solution (3.3) is thereby derived systematically by substituting Eq. (A9) into Eq. (A6) and setting $Z \equiv \pm g^{-\frac{1}{4}}\alpha^{\frac{1}{2}}e^{-C/2}$.

APPENDIX B: DILATATION GROUP INVARIANCE AND RIGOROUS SINGULARITY-FREE SELF-SIMILAR PARTICLELIKE SOLUTIONS

If $\theta = \theta_1(\mathbf{x}, t)$ satisfies the field equation (3.2), then

$$\theta_\lambda = \theta_\lambda(\mathbf{x}, t) \equiv \lambda^{\frac{1}{2}}\theta_1(\lambda\mathbf{x}, \lambda t) \quad \text{(B1)}$$

also satisfies Eq. (3.2) for all real $\lambda > 0$. Hence, the rigorous solutions to Eq. (3.2) fall into equivalence classes, solutions in an equivalence class being generated by the continuous one-parameter group of dilatation transformations. A general theorem[11] guarantees the existence of dilatation group invariant or so-called *self-similar solutions to* (3.2), for which

$$\lambda^{\frac{1}{2}}\theta_1(\lambda\mathbf{x}, \lambda t) = \theta_1(\mathbf{x}, t) \quad \text{(B2)}$$

for all real $\lambda > 0$; these self-similar solutions take the general form prescribed by (B2),

$$\theta_1(\mathbf{x}, t) = t^{-\frac{1}{2}}\chi(\mathbf{x}/t) \quad [t > 0], \quad \text{(B3)}$$

where χ satisfies the equation obtained by putting Eq. (B3) into (3.2).

Let us now consider the spherically symmetric specialization of (B3),

$$\theta_1(\mathbf{x}, t) = \pm(4g)^{-\frac{1}{4}}|\mathbf{x}|^{-\frac{1}{2}}\gamma(\zeta), \quad \zeta \equiv |\mathbf{x}|^2/t^2, \quad \text{(B4)}$$

with γ satisfying

$$(\zeta^3 - \zeta^2)\frac{d^2\gamma}{d\zeta^2} + (\tfrac{3}{2}\zeta^2 - \zeta)\frac{d\gamma}{d\zeta} + \tfrac{1}{16}\gamma - \tfrac{3}{16}\gamma^5 = 0 \quad \text{(B5)}$$

as a consequence of (3.2). The first integral of (B5) is obtained immediately as

$$16(\zeta^3 - \zeta^2)\left(\frac{d\gamma}{d\zeta}\right)^2 + \gamma^2 - \gamma^6 = A \quad \text{(B6)}$$

in which A is identically constant for all values of ζ if the physical solutions are required to be of function class C^1, piecewise C^2, but only piecewise constant if wave-crested solutions of function class C^0, piecewise C^2, are also admissible.[12] In the former case with $A \equiv$ const and with a finite solution (B4) at $\mathbf{x} = 0$, we have $\gamma(0) = 0$, and hence the constant A vanishes. Then the final quadrature is performed easily to yield

$$\gamma(\zeta) = \zeta^{\frac{1}{4}}[B + (1 - B^2)^{\frac{1}{2}}(\zeta - 1)^{\frac{1}{2}}]^{-\frac{1}{2}} \quad \text{(B7)}$$

in which $B\,(> -1)$ is a constant of integration. The spherically symmetric self-similar solution obtained by putting (B7) into (B4) is physically admissible in the region

$$|\mathbf{x}| \geq (1 - B^2)^{-\frac{1}{2}}t \quad \text{for} \quad -1 < B < 0,$$

in the region

$$|\mathbf{x}| > t \quad \text{for} \quad B = 0,$$

in the region

$$|\mathbf{x}| \geq t \quad \text{for} \quad 0 < B < 1,$$

and in the region

$$|\mathbf{x}| \leq t \quad \text{for} \quad B > 1,$$

only the trivial (spatially homogeneous) solution for $B = 1$ being admissible for all values of $|\mathbf{x}|$. However, singularity-free solutions of function class C^0, piecewise C^2, are obtainable for all values of $|\mathbf{x}|$ with A in (B6) piecewise constant; for example, with

$$A \equiv \begin{cases} 0 & \text{for} \quad 0 \leq \zeta < \zeta_{\text{cr}}, \\ A_{\text{cr}} & \text{for} \quad \zeta_{\text{cr}} < \zeta, \end{cases} \quad \text{(B8)}$$

where the positive constant $A_{\text{cr}}\,(> 2/3\sqrt{3})$ is related to $\zeta_{\text{cr}}\,(> 1)$ by

$$\int_0^{\zeta_{\text{cr}}^{\frac{1}{4}}} \frac{d\gamma}{(A_{\text{cr}} - \gamma^2 + \gamma^6)^{\frac{1}{2}}} = \tfrac{1}{2}\sin^{-1}\zeta_{\text{cr}}^{-\frac{1}{2}}, \quad \text{(B9)}$$

we obtain the wave-crested family of solutions

$$\gamma(\zeta) = \zeta^{\frac{1}{4}} \quad \text{for} \quad 0 \leq \zeta \leq \zeta_{\text{cr}},$$

$$\int_0^{\gamma(\zeta)} \frac{d\gamma}{(A_{\text{cr}} - \gamma^2 + \gamma^6)^{\frac{1}{2}}} = \tfrac{1}{2}\sin^{-1}\zeta^{-\frac{1}{2}} \quad \text{for} \quad \zeta_{\text{cr}} \leq \zeta. \quad \text{(B10)}$$

In contrast to the solutions (B7), which at best are asymptotically constant as $\zeta \to \infty$, the singularity-free solutions (B10) give $\gamma(\zeta) \to \frac{1}{2}A_{\text{cr}}^{\frac{1}{2}}\zeta^{-\frac{1}{2}}$ as $\zeta \to \infty$, and so the associated particlelike solutions (B4) are well-localized in space, falling off as $|\mathbf{x}|^{-\frac{3}{2}}$ for large values of $|\mathbf{x}|$. Note that the spherical surface $\zeta = \zeta_{\text{cr}}$ has the dynamical character of a supersonic shock wave with the velocity $|\mathbf{x}|/t = \zeta_{\text{cr}}^{\frac{1}{2}}$ greater than unity.

[11] A. J. A. Morgan, Quart. J. Math. (Oxford) **2**, 250 (1952).

[12] Rigorous examples of self-similar wave-crested solutions to a nonlinear partial differential equation are also reported in: G. Rosen, J. Math. and Phys. **45**, 235 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Proof that Successive Derivatives of Boltzmann's H Function for a Discrete Velocity Gas Alternate in Sign

STEWART HARRIS
College of Engineering, State University of New York at Stony Brook, Stony Brook, New York

(Received 13 January 1967)

It has been conjectured by McKean that the particular property of Boltzmann's H function which singles it out from a wide class of functionals of the Boltzmann solution may be that its successive derivatives alternate in sign. We consider here the proof of this alternating property for a discrete velocity gas. For the linearized-model Boltzmann equation, the proof is trivial. For the full (i.e., nonlinear) model Boltzmann equation, the proof is shown to be equivalent to demonstrating the positivity of a particular polynomial. The proof of this property is then demonstrated. It is also shown that $H^{(n)}$, like $H^{(1)}$, is zero only for the equilibrium distribution.

I. INTRODUCTION

IT is generally recognized that Boltzmann's H theorem stands as one of the great results of the kinetic theory. Boltzmann's somewhat imprecise interpretation of this result set the stage for one of the great controversies of physics history,[1] which included among its participants such notable non-physicists as the philosophers Nietzche and Spencer. In their masterful *Encyklopedia* article,[2] the Ehrenfests provided for the first time a clear understanding of what was, and what was not, implied by the H theorem, thus bringing to a close the controversy which had surrounded it since its birth. Until recently, the only efforts toward developing a deeper understanding of Boltzmann's result were confined almost exclusively to the related problems of ergodic theory.[3] However, in the past few years these problems have been re-examined by Kac[4] and by Grad,[5–7] who have brought to light several important points not considered by the Ehrenfests relating to the general nature of irreversibility.

More directly related to Boltzmann's result, which pertains only to systems described by the Boltzmann equation (i.e., a dilute, neutral gas) are the works of Moran,[8] McKean,[9,10] and the later work of Kac[11] and Grad.[12] Here some of the deeper problems associated with the H theorem are posed and solved for the first time for particular models of, and approximations to, the full (i.e., nonlinear) Boltzmann equation. One such problem which is stated, but only partially solved, is the following: what are the particular properties of the H function which single it out from a wide class of functionals of the Boltzmann solution? McKean[10] has conjectured that such a property may be that the successive derivatives of the H function alternate in sign; i.e., $dH/dt \leq 0$, $d^2H/dt^2 \geq 0$, $d^3H/dt^3 \leq 0$, etc. In order for this conjecture to be meaningful, it must first be established that the H function in fact possesses this alternating property. No such result is known either for the full Boltzmann equation or any of the models of, or approximations to it.

For the full Boltzmann equation, or even for the much more tractable linearized Boltzmann equation, the alternating property described above appears very difficult to show. We have, however, been able to prove the general result for what we believe is a reasonable, nonlinear model of the full Boltzmann equation; this is the Boltzmann equation for a discrete velocity gas. This model of a gas was first introduced, with surprisingly good quantitative results, by Maxwell[13] to obtain the equation of state for an ideal gas. It has subsequently been used with great success in describing boundary-value problems in

[1] An excellent account of this controversy and many of the original papers are contained in the book by S. G. Brush, *Kinetic Theory* (Pergamon Press, Inc., New York, 1966), Vol. 2.

[2] P. Ehrenfest and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics*, translated by M. J. Moravcsik (Cornell University Press, Ithaca, New York, 1959).

[3] Ja. G. Sinai, Dokl. Akad. Nauk SSSR **153**, 1261 (1963) [English transl.: Soviet Phys—Doklady **4**, 1818 (1963)] and references therein contains the most recent results. For an account of, and references to, earlier work, see I. E. Farquhar, *Ergodic Theory in Statistical Mechanics* (Interscience Publishers, Inc., New York, 1964).

[4] M. Dresden, in *Studies in Statistical Mechanics* (North-Holland Publishing Company, Amsterdam, 1962), Vol. I.

[5] H. Grad, J. Chem. Phys. **33**, 1342 (1960).

[6] H. Grad, Comm. Pure Appl. Math. **14**, 323 (1961).

[7] H. Grad, "Levels of Description in Statistical Mechanics and Thermodynamics," to appear in *The Delaware Seminar in the Foundations of Physics*. We are grateful to Professor Grad for sending us a preprint of this article.

[8] P. Moran, Proc. Cambridge Phil. Soc. **57**, 833 (1961).

[9] H. P. McKean Jr., Z. Wahrsheinlichkeitstheorie **2**, 167 (1963).

[10] H. P. McKean Jr., Arch. Ratl. Mech. Anal. **21**, 343 (1966).

[11] M. Kac, *Probability and Related Topics in the Physical Sciences* (Interscience Publishers, Inc., New York, 1959).

[12] H. Grad, J. Soc. Indust. Appl. Math. **13**, 259 (1965).

[13] J. C. Maxwell, *Scientific Papers II* (Cambridge University Press, Cambridge, England, 1890), p. 26.

kinetic theory,[14,15] and we have also previously used it to consider the extension of the H theorem to a moderately dense gas.[16]

In the next section we briefly describe this model, and formulate the proposition we are considering in terms of the model variables. In Sec. III we prove for the model that $d^{2k-1}H/dt^{2k-1} \leq 0$, $d^{2k}H/dt^{2k} \geq 0$ $(k = 1, 2, \cdots)$, where H is the analog for the model of Boltzmann's functional, $H[f] = \int \mathbf{dv} f(\mathbf{v}) \ln f(\mathbf{v})$.

II. FORMULATION OF THE PROPOSITION FOR THE DISCRETE VELOCITY GAS

A. Discrete Velocity Gas

The model we consider is a rarefied system of n particles which move with constant speed in a fixed number of directions (a discrete-velocity gas). Neither the number of velocity directions allowable nor the dimensionality of the system affects the qualitative results we are interested in (neglecting degenerate situations, e.g., one dimension or two dimensions and either one or two velocity directions). For this reason we restrict ourselves to the simplest meaningful case, which is two dimensions and four (orthogonal) velocity directions.

Let the number of particles moving in the direction of the positive y axis (the 1 direction) be N_1, so that N_3 is the number moving in the direction of the negative y axis, and N_2, N_4 are the number moving, respectively, in the direction of the positive and negative x axis (so that we have a standard xy coordinate system with directions numbered clockwise starting with the positive y direction). Then the "Boltzmann equation" for the model is[16]

$$\begin{aligned} \frac{dn_1}{dt} &= \frac{dn_3}{dt} = (n_2 n_4 - n_1 n_3), \\ \frac{dn_2}{dt} &= \frac{dn_4}{dt} = (n_1 n_3 - n_2 n_4) = -\frac{dn_1}{dt} = -\frac{dn_3}{dt}. \end{aligned} \tag{1}$$

Here $n_i = N_i/n$, and we have chosen our units so that $\frac{1}{2}B = 1$, where B is the mutual collision frequency for "head on" collisions. This "Boltzmann equation" has as its basis the two conservation laws (conservation of number implies conservation of energy in a collision), a stosszahlansatz for about-to-collide particles, and the assumption of symmetrical scattering through binary collisions, and thus corresponds for our model to the Boltzmann-equation description for an actual gas.

This model is used in the spirit one must generally adopt when attempting so simple a description of a complicated system: We consider the proof of a given proposition to be an *indication* that the more complicated system also exhibits this behavior (and we must not let the venerability of the Boltzmann equation awe us into forgetting that it too is only a model, albeit a very good model, in a particular well-defined regime). How strong this indication is, of course, depends on how well the model reproduces those features of the complicated system which are of particular interest. Although at first sight the discrete velocity gas seems quite simple, it has a decided advantage over many other more elegant models in that (i) the kinetic equation (1) is nonlinear, yet (ii) for a wide class of problems (1) is exactly solvable.

B. Formulation of the Problem

As we have shown elsewhere,[16] the proper analog of Boltzmann's H function for the discrete velocity gas model is

$$H[n_i] = \sum_{\text{(all velocities)}} n_i \ln n_i .$$

Differentiating this expression, we obtain for the model we are considering[17]

$$\frac{dH}{dt} = \sum \left(\frac{dn_i}{dt}\right) \ln n_i \leq 0,$$

$$\frac{d^2H}{dt^2} = \sum \left[\frac{d^2n_i}{dt^2} \ln n_i + \frac{1}{n_i}\left(\frac{dn_i}{dt}\right)^2\right].$$

The inequality indicated in the first equation is the statement of the H theorem for the model, the proof of which is trivial. That $d^2H/dt^2 \geq 0$ then follows immediately; using (1) and the normalization requirement $\sum n_i = 1$, we have

$$\frac{d^2H}{dt^2} = -\frac{dH}{dt} + \sum \frac{1}{n_i}\left(\frac{dn_i}{dt}\right)^2 \equiv -\frac{dH}{dt} + A[n_i] \geq 0. \tag{2}$$

The inequality is a direct result of the H theorem since $n_i \geq 0$. Also, since $A[n_i]$ is nonnegative, we see that the equality is only obtained for the equilibrium distribution.

An expression for $d^{k+2}H/dt^{k+2}$ can be obtained by differentiating (2) k times:

$$d^{k+2}H/dt^{k+2} \equiv H^{(k+2)} = -H^{(k+1)} + A^{(k)}[n_i]. \tag{3}$$

Since $H^{(1)} \leq 0$, showing $A^{(2k)} \geq 0$, $A^{(2k+1)} \leq 0$, for all k, will then constitute a proof, for the discrete velocity gas, that the successive derivatives of H possess the alternating property.

[14] J. Broadwell, J. Fluid Mech. **19**, 401 (1964).
[15] J. Broadwell, Phys. Fluids **7**, 1243 (1964).
[16] S. Harris, Phys. Fluids **9**, 1328 (1966).

[17] When the limits of summation are omitted, it is to be understood that the subscript i is summed from 1 to 4.

In the next section we first digress to prove (trivially) the desired result for the *linearized*-model Boltzmann equation. We then extend this result to the *full* model Boltzmann equation.

III. PROOF OF THE ALTERNATING PROPERTY

A. Proof for the Linearized-Model Boltzmann Equation

Let us first consider the linearized version of the model Boltzmann equation. Writing $n_i = n_{0i}(1 + 4h_i)$, substituting into (1), and retaining terms linear in h_i, we obtain the analog for the discrete velocity gas of the linearized Boltzmann equation:

$$\frac{dh_1}{dt} = \frac{dh_3}{dt} = -\frac{dh_2}{dt} = -\frac{dh_4}{dt} = (h_2 + h_4 - h_1 - h_3). \tag{4}$$

n_{0i} is the equilibrium value of n_i, which, for the case we are restricting ourselves to, a spatially homogeneous system in the absence of external fields, is $\frac{1}{4}$.[16] The corresponding linearized H function for the model is

$$H = H^\circ + \sum 2h_i^2$$

so that

$$\frac{dH}{dt} = 4\sum h_i \frac{dh_i}{dt} = -4(h_1 + h_3 - h_2 - h_4)^2 \leq 0,$$

$$\frac{d^2H}{dt^2} = 4\sum\left[\left(\frac{dh_i}{dt}\right)^2 + h_i\frac{d^2h_i}{dt^2}\right]$$

$$= -H^{(1)} + \sum\left(\frac{dh_i}{dt}\right)^2 \geq 0.$$

Differentiating this last equation k times, we obtain

$$H^{(k+2)} = -H^{(k+1)} + (-1)^{k+2}2^k\sum (dh_i/dt)^2, \tag{5}$$

so that, since $H^{(2)} \geq 0$, it is at once obvious that for this case the alternating property is established. Further we see also that since both terms on the right-hand side of (5) have the same sign for all k, $H^{(k+2)} = 0$ only for the equilibrium distribution.

B. Proof for the Full Model Boltzmann Equation

We wish to show that for dn_i/dt given by the full model Boltzmann equation (1),

$$A^{(2k)}[n_i] \geq 0, \quad A^{(2k-1)}[n_i] \leq 0, \quad k = 1, 2, \cdots,$$

where from (2)

$$A[n_i] = \sum (1/n_i)(dn_i/dt)^2 \equiv \sum A_i.$$

It is of course sufficient to show the desired result for the summand; consider A_1 as an example (the following results will also be valid for the other A_i).

Let us consider the kth derivative of (n_1A_1):

$$\frac{d^k(n_1A_1)}{dt^k} = \sum_{j=0}^{k} n_1^{(k-j)}A_1^{(j)}\binom{k}{j},$$

where $\binom{k}{j}$ is the binomial coefficient, $k!/j!\,(k-j)!$. The above equation can be rewritten as

$$A_1^{(k)} = \frac{1}{n_1}\left[(-2)^k\left(\frac{dn_1}{dt}\right)^2 - \sum_{j=0}^{k-1} n_1^{(k-j)}A_1^{(j)}\binom{k}{j}\right] \tag{6}$$

since

$$d^k(n_1A_1)/dt^k = (-2)^k(dn_1/dt)^2.$$

Equation (6) can be solved by iteration; doing this, we find

$$A_1^{(k)} = (-1)^k\left(\frac{1}{n_1}\frac{dn_1}{dt}\right)^2 \sum_{j=0}^{k} C_j^k\left[\frac{1}{n_1}\frac{dn_1}{dt}\right]^j, \tag{7}$$

where

$$C_{j+1}^k = \sum_{r=0}^{k-1} C_j^r\binom{k}{r}$$

and $C_0^k = 2^k$. Equation (7) is valid for all the $A_i^{(k)}$, i.e., if the 1 subscript is replaced *throughout* the equation by any of the i, the resulting equation will be valid.

From (7) it is clear that we need now only show that the polynomial in $(1/n_1)(dn_1/dt)$,

$$P(k) \equiv \sum_{j=0}^{k} C_j^k\left[\frac{1}{n_1}\frac{dn_1}{dt}\right]^j,$$

is nonnegative. Further we must only consider values of $x \equiv (1/n_1)\,dn_1/dt$ which are negative, since the C_j^k are positive. Since the minimum value which x can assume is -1 (this occurs when $n_2n_4 \to 0$, so that $x \to -n_3 \geq -1$), we must then consider the positivity of $P(k)$ for $-1 \leq x \leq 0$.

Writing out the first three and the rth term of $P(k)$, we have

$$P(k) = 2^k - (3^k - 2^k)\,|x| + (4^k - (2)3^k + 2^k)\,|x|^2$$

$$- \cdots + \sum_{j=1}^{r}(j+1)^k\binom{r-1}{j-1}(-1)^{j+1}(-|x|)^{r-1} + \cdots$$

$$= \sum_{r=1}^{k+1}\sum_{j=1}^{r}(j+1)^k\binom{r-1}{j-1}(-1)^{j+1}(-|x|)^{r-1}. \tag{8}$$

The positivity of this sum can be shown most simply by making use of a probabilistic interpretation. To do this we expand $(j+1)^k$, change indices on the j summation, and expand the resulting $(j+1)^s$ term, so that we have

$$P(k) = \sum_{r=1}^{k+1}\sum_{s=0}^{k}\sum_{t=0}^{s}|x|^{r-1}\binom{k}{s}\binom{s}{t}\sum_{j=0}^{r-1}\binom{r-1}{j}(-1)^{r-1-j}(j)^t,$$

$$= \sum_{r=1}^{k+1}\sum_{s=0}^{k}\sum_{t=0}^{s}|x|^{r-1}\binom{k}{s}\binom{s}{t}A^*(t, r-1), \tag{9}$$

where $A^*(t, r-1)$ is the *number* of possible distributions of t objects among $r-1$ cells which leave none of the cells empty.[18] The crucial point for our present interest is that $A^*(t, r-1) \geq 0$, and therefore $P(k) \geq 0$, which completes our proof of the alternating property for the full model Boltzmann equation.

It is possible to present a more explicit proof of the positivity property; this exploits the following identity:

$$\sum_{t=0}^{\infty} \frac{x^t}{t!} A^*(t, r) = (e^x - 1)^r.$$

Taking the nth derivative of both sides at $x = 0$ then gives the desired result:

$$A(n, r) = (\partial^n/\partial x^n)[(e^x - 1)^r]_{x=0} \geq 0.$$

We again note that $H^{(k)}$ will only be zero for the equilibrium distribution.

[18] See, e.g., W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley & Sons, Inc., New York, 1957), Vol. I, Chap. 2.

ACKNOWLEDGMENTS

The author would like to take this opportunity to gratefully acknowledge several discussions with Professor H. Grad during the course of the past several years, which have served to clarify for him many of the problems relating to the nature of irreversibility. The author thanks Professor G. Stell for helpful discussions concerning certain of the technical details in Sec. IIIB.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Energy Requirement for Nonlinear Density Fluctuations in a Vlasov Plasma

ROBERT L. W. CHEN
Physics Department, Emory University, Atlanta, Georgia

(Received 5 June 1967)

Exact solutions of an isoperimetric problem are presented which lead to a lower bound on the kinetic energy of the particles in a Vlasov plasma, in terms of the mean-density fluctuation:

$$E \gtrapprox \tfrac{3}{2}nkT\left[1 + 0.4\left(\frac{\overline{|\Delta n|}}{n}\right)^2\right],$$

where E is the kinetic energy per unit volume, n the mean number density, k the Boltzmann constant, T the temperature, and Δn the deviation of density. This energy requirement associated with Δn must be taken into account when considering the energetic aspect of the growth of instabilities. The present work, which relies entirely on mathematical analysis, confirms the main results of an earlier nonrigorous calculation.

1. INTRODUCTION

IN plasma researches, it is useful to obtain theoretic bounds on the nonlinear turbulent fluctuations that grow out of a given unstable velocity distribution. This can be accomplished by means of energetic considerations, as has been demonstrated in some recent plasma literature.[1-4] The growth of the instabilities requires the conversion of particle kinetic energy into electromagnetic field energy. The portion of kinetic energy available for this process can be readily calculated in accordance with the Liouville theorem, which is valid for a collisionless plasma. The ensuing field fluctuations are, therefore, necessarily limited. The writer previously observed[2] that on the same basis of Liouville invariance, a certain amount of kinetic energy must accompany the *density* fluctuations. This energy requirement must be reckoned with in addition to the field energy we have mentioned. Its exact amount can be found by solving an appropriate isoperimetric problem. A proper mathematical treatment of this problem is presented here for the first time, although a nonrigorous calculation has been made earlier.[2]

The basic nature of this problem can be understood in qualitative terms. Consider the various states accessible to a plasma with a given initial condition in the sense that they comply with the constraint of

[1] T. K. Fowler, J. Math. Phys. **4**, 559 (1963).
[2] R. L. W. Chen, Phys. Fluids **9**, 761 (1966).
[3] T. K. Fowler, Phys. Fluids **8**, 459 (1965); **9**, 1437 (1966).
[4] C. S. Gardner, Phys. Fluids **6**, 839 (1963).

where $A^*(t, r-1)$ is the *number* of possible distributions of t objects among $r-1$ cells which leave none of the cells empty.[18] The crucial point for our present interest is that $A^*(t, r-1) \geq 0$, and therefore $P(k) \geq 0$, which completes our proof of the alternating property for the full model Boltzmann equation.

It is possible to present a more explicit proof of the positivity property; this exploits the following identity:

$$\sum_{t=0}^{\infty} \frac{x^t}{t!} A^*(t, r) = (e^x - 1)^r.$$

Taking the nth derivative of both sides at $x = 0$ then gives the desired result:

$$A(n, r) = (\partial^n/\partial x^n)[(e^x - 1)^r]_{x=0} \geq 0.$$

We again note that $H^{(k)}$ will only be zero for the equilibrium distribution.

[18] See, e.g., W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley & Sons, Inc., New York, 1957), Vol. I, Chap. 2.

ACKNOWLEDGMENTS

The author would like to take this opportunity to gratefully acknowledge several discussions with Professor H. Grad during the course of the past several years, which have served to clarify for him many of the problems relating to the nature of irreversibility. The author thanks Professor G. Stell for helpful discussions concerning certain of the technical details in Sec. IIIB.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Energy Requirement for Nonlinear Density Fluctuations in a Vlasov Plasma

ROBERT L. W. CHEN
Physics Department, Emory University, Atlanta, Georgia

(Received 5 June 1967)

Exact solutions of an isoperimetric problem are presented which lead to a lower bound on the kinetic energy of the particles in a Vlasov plasma, in terms of the mean-density fluctuation:

$$E \gtrapprox \tfrac{3}{2}nkT\left[1 + 0.4\left(\frac{\overline{|\Delta n|}}{n}\right)^2\right],$$

where E is the kinetic energy per unit volume, n the mean number density, k the Boltzmann constant, T the temperature, and Δn the deviation of density. This energy requirement associated with Δn must be taken into account when considering the energetic aspect of the growth of instabilities. The present work, which relies entirely on mathematical analysis, confirms the main results of an earlier nonrigorous calculation.

1. INTRODUCTION

IN plasma researches, it is useful to obtain theoretic bounds on the nonlinear turbulent fluctuations that grow out of a given unstable velocity distribution. This can be accomplished by means of energetic considerations, as has been demonstrated in some recent plasma literature.[1-4] The growth of the instabilities requires the conversion of particle kinetic energy into electromagnetic field energy. The portion of kinetic energy available for this process can be readily calculated in accordance with the Liouville theorem, which is valid for a collisionless plasma. The ensuing field fluctuations are, therefore, necessarily limited. The writer previously observed[2] that on the same basis of Liouville invariance, a certain amount of kinetic energy must accompany the *density* fluctuations. This energy requirement must be reckoned with in addition to the field energy we have mentioned. Its exact amount can be found by solving an appropriate isoperimetric problem. A proper mathematical treatment of this problem is presented here for the first time, although a nonrigorous calculation has been made earlier.[2]

The basic nature of this problem can be understood in qualitative terms. Consider the various states accessible to a plasma with a given initial condition in the sense that they comply with the constraint of

[1] T. K. Fowler, J. Math. Phys. **4**, 559 (1963).
[2] R. L. W. Chen, Phys. Fluids **9**, 761 (1966).
[3] T. K. Fowler, Phys. Fluids **8**, 459 (1965); **9**, 1437 (1966).
[4] C. S. Gardner, Phys. Fluids **6**, 839 (1963).

Liouville invariance. The state with the minimum kinetic energy is a state whose higher densities *in phase space* are located as near as possible to $v = 0$.[4] This would be a state of uniform *spatial* density. If we ask: What is the minimum kinetic energy of states of a given degree of density (spatial) inhomogeneity? The answer must be some value *above* that of the uniform state, since some portion of the plasma must give up its "preferred" positions in $\mathbf{r}$, $\mathbf{v}$ space in order to realize the inhomogeneity. A quantitative treatment of this problem leads to the result that the additional energy is approximately $\frac{3}{5}nkT(\overline{|\Delta n|}/n)^2$ per unit volume, where $\overline{|\Delta n|}$ is the average density deviation.

2. ISOPERIMETRIC PROBLEM

We consider a one-component electronic plasma which obeys periodic boundary conditions within a large box of dimensions L^3. To obtain a lower bound on the kinetic energy for a given value of $\overline{|\Delta n|}$, we pose the following mathematical problem.

Let $f(\mathbf{r}, \mathbf{v})$ be a nonnegative function defined for $\mathbf{r}$ within a cubic volume L^3 of $\mathbf{r}$ space, and for $\mathbf{v}$ at any point throughout $\mathbf{v}$ space. Determine $f(\mathbf{r}, \mathbf{v})$ [or a class of $f(\mathbf{r}, \mathbf{v})$] which *minimizes* $\mathcal{E}$:

$$\mathcal{E} \equiv \int\int_{L^3} \tfrac{1}{2}mv^2 f(\mathbf{r}, \mathbf{v})\, d\mathbf{r}\, d\mathbf{v},$$

under the constraints that the following quantities be kept constant:

$$\Phi(\alpha) = \text{const}, \tag{2.1a}$$

$$\overline{|\Delta n|} = \text{const}, \tag{2.1b}$$

where $\Phi(\alpha)$ is defined as the measure of the point set in the six-dimensional $\mathbf{r}$, $\mathbf{v}$ product space wherein f is greater than α, i.e., the set $\{\mathbf{r}, \mathbf{v}: f(\mathbf{r}, \mathbf{v}) > \alpha\}$. Then Δn is the deviation of the number density at $\mathbf{r}$ from the mean number density:

$$\Delta n(\mathbf{r}) \equiv \int f(\mathbf{r}, \mathbf{v})\, d\mathbf{v} - L^{-3}\int\int_{L^3} f(\mathbf{r}, \mathbf{v})\, d\mathbf{r}\, d\mathbf{v},$$

and the average is

$$\overline{|\Delta n|} \equiv L^{-3}\int_{L^3} |\Delta n(\mathbf{r})|\, d\mathbf{r}.$$

We could have chosen the *root-mean-square* value of Δn as the quantity to be fixed in (2.1b). The present form of the problem, however, proves to be more amenable to analytic treatment. For density distributions that are not sharply peaked, the values of these two quantities, the mean $\overline{|\Delta n|}$ and the root mean square of Δn, do not differ greatly from each other.

The other constraint (2.1a) sets $\Phi(\alpha)$ equal to a given function of α. The latter is designated as $\Phi_0(\alpha)$ and is chosen to be the $\Phi(\alpha)$ of a uniform unperturbed Maxwellian of temperature T and density n:

$$f_0(v) = (2\pi kT/m)^{-\frac{3}{2}} n e^{-mv^2/2kT}.$$

It is straightforward to find that

$$\Phi_0(\alpha) = \tfrac{4}{3}\pi L^3 (2kT/m)^{\frac{3}{2}}\{-\log\,[(\alpha/n)(2\pi kT/m)^{\frac{3}{2}}]\}^{\frac{3}{2}}. \tag{2.2}$$

Thus T and n, together with $\overline{|\Delta n|}$, form a set of constant parameters for the problem. The minimum $\mathcal{E}$ that we are seeking should be a function of these parameters.

A. Reduction of the Problem

It is difficult to solve the above problem *directly*. With the introduction of another constraint, however, we can reduce it to a form for which exact solutions become possible. This extraneous constraint, which is not wanted, can be removed at an appropriate time. (See the last section of this paper.)

Consider two point sets in $\mathbf{r}$ space, in which $\Delta n(\mathbf{r})$ is negative and nonnegative, respectively; i.e.,

I. $\{\mathbf{r}: \Delta n(\mathbf{r}) < 0\}$,

II. $\{\mathbf{r}: \Delta n(\mathbf{r}) \geq 0\}$.

The volumes of these sets are then functionals of $f(\mathbf{r}, \mathbf{v})$ and will be denoted by $\mathcal{V}_1$ and $\mathcal{V}_2$. Our new constraint takes the form

$$\mathcal{V}_1/\mathcal{V}_2 = \text{const} = w. \tag{2.3}$$

In seeking $f(\mathbf{r}, \mathbf{v})$ that minimizes $\mathcal{E}$ under constraints (2.2) and (2.3), we may consider only f of the form

$$f(\mathbf{r}, \mathbf{v}) = \begin{cases} f_1(v) & \text{for } \mathbf{r} \text{ in Set I,} \\ f_2(v) & \text{for } \mathbf{r} \text{ in Set II,} \end{cases} \tag{2.4}$$

where $f_1(v)$ and $f_2(v)$ are *monotonically decreasing* functions of v $(0 \leq v < \infty)$. The reason is as follows: Given an arbitrary $f(\mathbf{r}, \mathbf{v})$, the two sets in $\mathbf{r}$ space, I and II, are defined accordingly. Now consider two domains in $\mathbf{r}$, $\mathbf{v}$ phase space in which $\mathbf{r} \in$ I, $\mathbf{r} \in$ II, respectively. In accordance with Gardner's observation,[4] one can distort $f(\mathbf{r}, \mathbf{v})$ in *each* of the domains to the forms $f_1(v)$ and $f_2(v)$, respectively, in such a way that *their respective ϕ measures remain the same while their $\mathcal{E}$ values become less than or equal to the original values.* In the mean time, the distortion also *leaves* $\mathcal{V}_1/\mathcal{V}_2$ *unchanged*, since the total number of particles in each set remains the same and the density at any point $\mathbf{r}$ in Set I remains *lower* than that in Set II. Thus, we conclude that, for any given $f(\mathbf{r}, \mathbf{v})$, we can always find an f of the form (2.4) which has a smaller or equal value of total energy $\mathcal{E}$ while sharing the same $\Phi(\alpha)$ and $\mathcal{V}_1/\mathcal{V}_2$ with the original $f(\mathbf{r}, \mathbf{v})$.

We note further that $f_1(v)$ and $f_2(v)$ are not independent of each other; and both can be derived from an appropriately chosen *single* function. Consider the measures in the *velocity* space of sets $\{\mathbf{v}: f_1(v) > \alpha\}$ and $\{\mathbf{v}: f_2(v) > \alpha\}$. These can be expressed as $\phi_0(\alpha) - \phi(\alpha)$ and $\phi_0(\alpha) + w\phi(\alpha)$, respectively, where $\phi_0(\alpha) \equiv L^{-3}\Phi_0(\alpha)$. In these forms $\Phi(\alpha)$ is always $\Phi_0(\alpha)$ for arbitrary $\phi(\alpha)$, as is required by our constraints. Since $f_1(v), f_2(v)$ are monotonically decreasing functions of α, f_1 (or f_2) is equal to α on the surface of a sphere, in $\mathbf{v}$ space, of volume $\phi_0(\alpha) - \phi(\alpha)$ [or $\phi_0(\alpha) + w\phi(\alpha)$] centered at the origin $v = 0$. The total kinetic energy can now be readily expressed in terms of $\phi(\alpha)$:

$$E = -2^{-\frac{7}{5}}\left(\frac{3}{\pi}\right)^{\frac{2}{5}}\frac{m}{1+w}\int\{w(\phi_0-\phi)^{\frac{2}{5}}(\phi_0'-\phi') + (\phi_0+w\phi)^{\frac{2}{5}}(\phi_0'+w\phi')\}\alpha\,d\alpha, \quad (2.5)$$

where $E \equiv L^{-3}\mathcal{E}$, the average energy per unit volume in L^3. The unknown function $\phi(\alpha)$ is to be varied to attain minimization of E. Since $\phi_0 - \phi$ and $\phi_0 + w\phi$ are, by their nature, nondecreasing functions of α, just as $\Phi(\alpha)$ is, $\phi(\alpha)$ is subject to the bounds

$$-\frac{1}{w}\phi_0'(\alpha) \geq \phi'(\alpha) \geq \phi_0'(\alpha). \quad (2.6a)$$

By definition, the density in I is smaller than that in II, and we must have

$$\int \phi'(\alpha)\alpha\,d\alpha < 0.$$

$\overline{|\Delta n|}$ can be found as

$$\overline{|\Delta n|} = \frac{-2w}{1+w}\int\phi'(\alpha)\alpha\,d\alpha.$$

Constraint (2.1b) now takes the form

$$\int\phi'(\alpha)\alpha\,d\alpha = \text{const} < 0. \quad (2.6b)$$

With suitable changes of variables, the preceding problem may be recast into a form which proves to be more convenient. Introducing

$$x \equiv -\log\,[(\alpha/n)(2\pi kT/m)^{\frac{3}{2}}]$$

and

$$\mathcal{F} \equiv -(\alpha^2/n)\phi'(\alpha),$$

(2.5) becomes

$$E = \tfrac{3}{2}nkT\Psi,$$

$$\Psi \equiv \frac{8}{15}\frac{\pi^{\frac{1}{2}}}{(1+w)}\int_0^\infty e^{-x}[u_1^{\frac{5}{2}}(x) + u_2^{\frac{5}{2}}(x)]'\,dx,$$

$$\begin{cases} u_2(x) \equiv \left[x^{\frac{3}{2}} + w\tfrac{3}{4}\pi^{\frac{1}{2}}\int_0^x e^y\mathcal{F}(y)\,dy\right]^{\frac{2}{3}}, \\ u_1(x) \equiv \left[x^{\frac{3}{2}} - \tfrac{3}{4}\pi^{\frac{1}{2}}\int_0^x e^y\mathcal{F}(y)\,dy\right]^{\frac{2}{3}}. \end{cases} \quad (2.7)$$

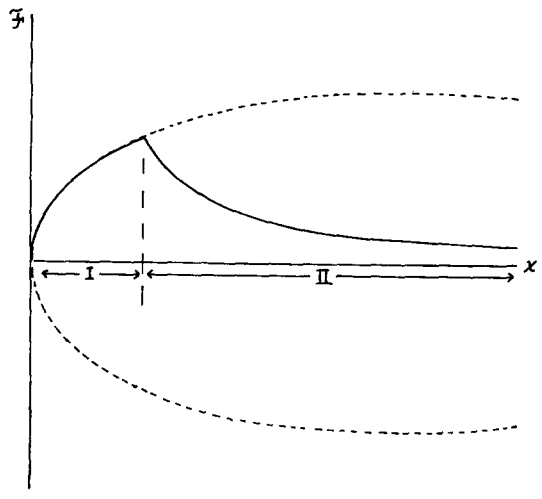


FIG. 1. $\mathcal{F}(x)$ and its bounds; the figure shows (I) "constrained interval" and (II) "free intervals" solid line represents $\mathcal{F}(x)$; broken lines represent the upper and lower bounds $\pm 2/\pi^{\frac{1}{2}}e^{-x}x^{\frac{1}{2}}$.

Primes indicate differentiation with respect to x. Equations (2.6a) and (2.6b) become

$$\begin{cases} -\frac{1}{w}\frac{2}{\pi^{\frac{1}{2}}}x^{\frac{1}{2}}e^{-x} \leq \mathcal{F}(x) \leq \frac{2}{\pi^{\frac{1}{2}}}x^{\frac{1}{2}}e^{-x}, & (2.8a) \\ \int_0^\infty \mathcal{F}(x)\,dx = \text{const} = c > 0. & (2.8b) \end{cases}$$

The problem now reduces to this: *Find the* $\mathcal{F}(x)$ *which minimizes* Ψ *of* (2.7), *subject to the conditions* (2.8).

3. EXACT SOLUTIONS

The preceding problem would be solvable in a straightforward manner by the usual Euler–Lagrange methods except for the bound (2.8a) imposed on the solution. $\mathcal{F}(x)$ may coincide with the bounds in some interval, say, $x_1 < x < x_2$. $x_1 - x_2$ will be referred to as the "constrained interval," and the remaining range, the "free interval." (See Fig. 1.)

We wish to find an $\mathcal{F}(x)$ which satisfies these criteria: (1) $\delta\Psi = 0$ and $\delta^2\Psi > 0$ when the variation $\delta\mathcal{F}(x)$ is confined to within the *free* interval; (2) $\delta\Psi > 0$ when the variation involves moving $\mathcal{F}(x)$ inward and *away* from the boundary in the *constrained interval.* For this purpose it is convenient to consider an elemental form of $\delta\mathcal{F}(x)$:

$$\delta(x) = \delta s[d\mathcal{F}(x-b) - d(x-a)], \quad (3.1)$$

which is consistent with the constraint (2.8b). Herein d denotes the delta function, while δ is reserved for the *variation.* A general $\delta\mathcal{F}(x)$ can be built up from superposition of (3.1) of different a and b. Equation (3.1) signifies a transport of quantity δs from $x = a$

to $x = b$. The criteria for (3.1) become:

$$\left.\begin{array}{ll}(1)\ \delta\Psi = 0 \\ (2)\ \delta^2\Psi > 0\end{array}\right\}\ \text{when both } a \text{ and } b \text{ lie in the } \textit{free} \text{ interval,}$$

$$(3)\ \delta\Psi > 0 \quad \text{when } a \text{ lies in the } \textit{constrained} \text{ interval and } b \text{ lies in the } \textit{free} \text{ interval.} \tag{3.2}$$

When (1) is assumed to be valid for *arbitrary* a and b within a certain interval x_2–x_3, it leads directly to a first integral

$$u_2(x) - u_1(x) = \text{const} = k. \tag{3.3}$$

It can be readily verified that the use of Lagrange multipliers lead to an identical result. Equation (3.3) governs $\mathcal{F}(x)$ in the *free* interval. It will be convenient for later purposes to introduce a new variable

$$\eta \equiv \tfrac{3}{4}\pi^{\frac{1}{2}}x^{-\frac{3}{2}}\int_0^x e^{x'}\mathcal{F}(x')\,dx'$$

from which $\mathcal{F}(x)$ may be recovered:

$$\mathcal{F}(x) = \frac{2}{\pi^{\frac{1}{2}}}e^{-x}x^{\frac{1}{2}}\left(\eta(x) + \tfrac{2}{3}x\frac{d\eta}{dx}\right).$$

Equation (3.3) now takes the form

$$[(1 + w\eta)^{\frac{2}{3}} - (1 - \eta)^{\frac{2}{3}}]x = \text{const} = k. \tag{3.4}$$

For any positive k, η decreases monotonically. The corresponding $\mathcal{F}(x)$ reaches the boundary at $2^{-\frac{3}{2}}k$ and remains between the bounds from $2^{-\frac{3}{2}}k$ to ∞. Thus Eq. (3.4) is acceptable as a solution in this range of x, provided that a suitable solution can be found outside of this range, namely 0–$2^{-\frac{3}{2}}k$. It turns out that the only admissable one is $\eta(x) = 1$ *identically*. Any other $\eta(x)$ can be shown to violate either the bounds of (2.8a) or fail to match the value of Eq. (3.4) at $2^{-\frac{3}{2}}k$. Thus we obtain the solution in the whole range of x:

$$\begin{cases}\eta = 1, & (0 \le x \le 2^{-\frac{3}{2}}k); \\ [(1 + w\eta)^{\frac{2}{3}} - (1 - \eta)^{\frac{2}{3}}]x = k & (2^{-\frac{3}{2}}k < x < \infty).\end{cases} \tag{3.5}$$

The first part of Eq. (3.5) yields $\mathcal{F}(x)$ lying on the boundary. Then $0 - 2^{-\frac{3}{2}}k$ is what we call the *constrained* interval, and $2^{-\frac{3}{2}}k - \infty$ is called the free *interval.*

Now it remains to test tne other two conditions (2) and (3) in (3.2). For condition (2), we obtain the second variation by treating the delta functions in (3.1) as appropriate step functions. Upon taking suitable limits, we obtain

$$\frac{\delta^2\Psi}{\delta s^2} = N(a, a) + N(b, b) - 2N(a, b), \qquad (2^{-\frac{3}{2}}k < a < b < \infty),$$

where

$$N(a, b) \equiv \frac{\pi^{\frac{1}{2}}w}{3(1 + w)}e^a[u_1^{-\frac{1}{2}}(a) + wu_2^{-\frac{1}{2}}(b)] - \frac{2\pi^{\frac{1}{2}}}{3}\frac{w}{(1 + w)}e^{a+b}\int_a^\infty e^{-x}[wu_2^{-\frac{1}{2}}(x) + u_1^{-\frac{1}{2}}(x)]'\,dx. \tag{3.6}$$

In accordance with Eq. (3.4), it can be shown that both $u_1(x)$ and $u_2(x)$ are *positive–definite and monotonically increasing* functions of x in the range $2^{-\frac{3}{2}}k < x < \infty$. The integrand in (3.6) is therefore negative. These lead to the following inequalities:

$$(1)\quad e^a[u_1(a)^{-\frac{1}{2}} + wu_2(a)^{-\frac{1}{2}}] > e^a[u_1(b)^{-\frac{1}{2}} + wu_2(b)^{-\frac{1}{2}}] > e^b[u_1(b)^{-\frac{1}{2}} + wu_2(b)^{-\frac{1}{2}}]; \tag{3.7}$$

$$(2)\quad 2e^{a+b}\int_b^\infty \cdots dx > (e^{2a} + e^{2b})\int_b^\infty \cdots dx > e^{2a}\int_a^\infty \cdots dx + e^{2b}\int_b^\infty \cdots dx, \tag{3.8}$$

where $\int \cdots dx$ stands for the integral in Eq. (3.6). With these relations, $\delta^2\Psi/\delta s^2$ of Eq. (3.6) becomes positive–definite: $\delta^2\Psi/\delta s^2 > 0$ (for a and b both in the interval: $2^{-\frac{3}{2}}k - \infty$).

To test condition (3) of Eq. (3.2), a and b are now placed in the intervals $0 < a < 2^{-\frac{3}{2}}k < b < \infty$. We obtain $\delta\Psi/\delta s$ in terms of u_1 and u_2:

$$\frac{\delta\Psi}{\delta s} = M(a) - M(b),$$

where

$$M(a) \equiv \frac{2w}{3(1 + w)}[u_2(a) - u_1(a)] - \frac{2e^a w}{3(1 + w)}\int_a^\infty [u_2'(x) - u_1'(x)]e^{-x}\,dx.$$

Evaluating the above in accordance with (3.5), we obtain

$$\frac{\delta\Psi}{\delta s} = \tfrac{2}{3}w(1 + w)^{-\frac{1}{3}}[(2^{-\frac{3}{2}}k - a) + \exp(-2^{-\frac{3}{2}}k + a) - 1].$$

Since $2^{-\frac{3}{2}}k - a > 0$, we may use the inequality

$$h + e^{-h} - 1 > 0 \quad (\text{for } h > 0).$$

Thus we have

$$\delta\Psi/\delta s > 0 \quad (\text{for } a, b \text{ in the ranges } 0 < a < 2^{-\frac{3}{2}}k < b < \infty).$$

We therefore conclude that (3.5) is the minimum solution we are seeking.

4. APPROXIMATE EXPRESSION FOR THE BOUND

On the basis of the preceding analysis, we may proceed to calculate the minimum E as a function of n, T, and $\overline{|\Delta n|}$. It will be useful to find an expression in terms of these three parameters. The original objective of this paper was to exhibit an additional energy requirement for a state of density fluctuation in comparison with the uniform state.

In accordance with the solution (3.5), the minimum Ψ and $\overline{|\Delta n|}$ may be equated to the following quadratures:

$$\Psi = \frac{4(1+w)^{\frac{3}{2}}}{3\pi^{\frac{1}{2}}}\int_0^{2^{-\frac{3}{2}}k} x^{\frac{3}{2}}e^{-x}\,dx + \frac{8(1+w)}{9\pi^{\frac{1}{2}}}\int_0^1 \exp\frac{-k}{P^{\frac{2}{3}}-Q^{\frac{2}{3}}} \cdot [P^{-\frac{1}{3}}+Q^{-\frac{1}{3}}][P^{\frac{2}{3}}-Q^{\frac{2}{3}}]^{-\frac{7}{2}}[P^{\frac{1}{3}}+wQ^{\frac{1}{3}}]^{-1}\,d\eta, \tag{4.1}$$

$$\overline{|\Delta n|} = \frac{4wn}{\pi^{\frac{1}{2}}(1+w)}\int_0^{2^{-\frac{3}{2}}k} x^{\frac{1}{2}}e^{-x}\,dx + \frac{8wnk^{\frac{3}{2}}}{3\pi^{\frac{1}{2}}(1+w)}\int_0^1 \exp\frac{-k}{P^{\frac{2}{3}}-Q^{\frac{2}{3}}} \cdot [P^{\frac{1}{3}}-Q^{\frac{1}{3}}] \cdot [P^{-\frac{1}{3}}+Q^{-\frac{1}{3}}][P^{\frac{2}{3}}-Q^{\frac{2}{3}}]^{-\frac{5}{2}}[P^{\frac{1}{3}}+wQ^{\frac{1}{3}}]^{-1}\,d\eta, \tag{4.2}$$

where $\begin{cases} P \equiv 1 + w\eta, \\ Q \equiv 1 - \eta. \end{cases}$

Thus Ψ and $\overline{|\Delta n|}/n$ depend only on w and k. These integrals have been computed electronically for different sets of values of w and k (see Table I). For a given fixed w, different k yield different pairs of Ψ and $\overline{|\Delta n|}/n$. A connection between Ψ and $\overline{|\Delta n|}/n$ is thereby established. We denote this dependence by

$$\Psi_w\left(\frac{\overline{|\Delta n|}}{n}\right).$$

The minimum E is then equal to

$$\tfrac{3}{2}nkT\Psi_w,$$

according to Eq. (2.7). As yet, there is a dependence on w due to the *additional* constraint introduced early in Sec. 2, namely $\mathfrak{V}_1/\mathfrak{V}_2 = w$. To return to solving the original problem of minimizing E *without* this extraneous constraint, we can seek w that minimizes $\Psi_w(\overline{|\Delta n|}/n)$ for each value of $\overline{|\Delta n|}/n$. It turns out that this lies in the range 1–1.4. Within this range, $\Psi_w(\overline{|\Delta n|}/n)$ is only slightly dependent on w, differing by a few percent. Thus, for the absolute minimum, we have

$$\Psi_{\min}\left(\frac{\overline{|\Delta n|}}{n}\right) \approx \Psi_w\left(\frac{\overline{|\Delta n|}}{n}\right)\Big|_{w=1}.$$

TABLE I. Samples of actual computed values of $\Psi_w(\overline{|\Delta n|}/n)$ and their percent differences from the approximate formula $\Psi = 1 + 0.4(\overline{|\Delta n|}/n)^2$, taken from extensive data of electronic computations.

w	Ψ	$\frac{\overline{\lvert\Delta n\rvert}}{n}$	% difference from formula
1.0	1.0108	0.1542	0.1
1.0	1.0337	0.2886	0.1
1.0	1.1384	0.5877	0.1
1.0	1.2382	0.7416	1.4
1.0	1.3429	0.8543	2.8
1.2	1.0231	0.2271	0.3
1.2	1.1764	0.6477	0.8
1.2	1.2866	0.8129	1.8
1.4	1.0229	0.2239	0.3
1.4	1.1976	0.6965	0.3
1.4	1.3102	0.8486	0.9

Based on our numerical results, $\Psi_w(\overline{|\Delta n|}/n)$ is expressed in the form of a power series. It turns out to be

$$\Psi_w\left(\frac{\overline{|\Delta n|}}{n}\right)\Big|_{w=1} \approx 1 + 0.4\left(\frac{\overline{|\Delta n|}}{n}\right)^2. \tag{4.3}$$

We have adjusted the constants in the definition of Ψ so that it normalizes to 1 for $\Delta n = 0$. The vanishing of the linear term in Eq. (4.3) is attributable to the stationarity of $\Psi_w|_{w=1}$ at $\Delta n = 0$. The contribution from the third- and higher-order terms is less than 3 or 4% of the second term, for $\overline{|\Delta n|}/n$ as large as two-thirds. Equation (4.3) therefore omits these higher terms. The minimum E becomes, to a good approximation,

$$E_{\min} \cong \tfrac{3}{2}nkT\left[1 + 0.4\left(\frac{\overline{|\Delta n|}}{n}\right)^2\right]. \tag{4.4}$$

Equation (4.4) *gives the lower bound on the average kinetic energy per unit volume for a plasma of mean number density* n, *temperature* T, *and average density deviation* $\overline{|\Delta n|}$.

ACKNOWLEDGMENTS

This research is supported in part by the McCandless Fund of Emory University. The numerical work was done at Emory Biomedical Data Processing and Analysis Center. The author wishes to thank its director, Dr. Buell Evan.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Results on the Analyticity of Many-Body Scattering Amplitudes in Perturbation Theory

PAUL FEDERBUSH
Department of Mathematics, University of Michigan, Ann Arbor, Michigan

(Received 8 February 1967)

The problem of analytic continuation of the many-body scattering amplitude associated with a perturbation-theory diagram under the rotation of the final momenta from real to complex momenta, $\mathbf{k} \to (1 + i\theta)\mathbf{k}$, is studied. It is shown that the contour of integration over internal momenta can be distorted avoiding singularities of the integrand, as θ varies for small enough θ. If the diagram is connected enough, the potentials are Yukawa-type, Re $E > 0$, and Im $E < 0$. The rotation angle can be picked independently of Im E.

I. INTRODUCTION

We intend to study the analytic continuation in θ of a many-body scattering diagram, where the final momenta, all initially real, are rotated under the transformation $\mathbf{k} \to (1 + i\theta)\mathbf{k}$. A more general type of continuation will also be studied. We restrict our attention to Re $E > 0$, Im $E < 0$, the most interesting situation. The heart of the study is an analysis of a peculiar geometrical problem. The type of geometrical analysis involved clearly can be applied to more general problems than those studied here, about which a few comments will be made later.

II. SCATTERING DIAGRAM

We deal with an N-particle scattering situation. There will then be N masses, $M_1, M_2, \cdots, M_N$, and a Hamiltonian

$$H = \sum_i \frac{\mathbf{k}_i^2}{2M_i} + \sum_{i<j} V_{ij}(|\mathbf{x}_i - \mathbf{x}_j|). \tag{1}$$

Clearly we limit ourselves to two-body interactions. We change momentum variables so that H_0 can be written

$$H_0 = \sum_i \mathbf{k}_i^2, \tag{2}$$

and there is then the equation of momentum conservation

$$\sum_i \mathbf{k}_i M_i^{\frac{1}{2}} = \text{const.} \tag{3}$$

Now consider a perturbation-theory diagram specified by a sequence of interactions and internal momenta (before integration). We write the momentum of each state as a vector in $3N$-dimensional Euclidean space $(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_N)$. By working in the center-of-mass system, one can take

$$\sum_i M_i^{\frac{1}{2}} \mathbf{k}_i = 0. \tag{4}$$

If there are S intermediate states in the diagram, there is then the following sequence of $(S + 2)$ vectors:

$$\begin{aligned}
&(\mathbf{k}_1^{\text{in}}, \mathbf{k}_2^{\text{in}}, \cdots, \mathbf{k}_N^{\text{in}}) \\
&(\mathbf{k}_1^{1}, \mathbf{k}_2^{1}, \cdots, \mathbf{k}_N^{1}) \\
&\quad\cdot \\
&\quad\cdot \\
&\quad\cdot \\
&(\mathbf{k}_1^{s}, \mathbf{k}_2^{s}, \cdots, \mathbf{k}_N^{s}) \\
&(\mathbf{k}_1^{\text{fin}}, \mathbf{k}_2^{\text{fin}}, \cdots, \mathbf{k}_N^{\text{fin}}).
\end{aligned} \tag{5}$$

For each of the vectors, Eq. (4) holds. If the interaction between the ith intermediate state and the $(i + 1)$th intermediate state is V_{LT}, the vector difference in $3N$-dimensional space

$$(\mathbf{k}_1^{i+1}, \cdots, \mathbf{k}_N^{i+1}) - (\mathbf{k}_1^{i}, \cdots, \mathbf{k}_N^{i})$$

is parallel to the vector

$$(0, 0, \cdots, 0, \underset{(L\text{th place})}{M_T^{\frac{1}{2}}\mathbf{a}}, 0, \cdots, 0, \cdots, \underset{(T\text{th place})}{-M_L^{\frac{1}{2}}\mathbf{a}}, 0, \cdots, 0)$$

for some three-dimensional vector $\mathbf{a}$. We prefer to think of this difference as the appropriate succession of three displacements in the directions

$$\begin{aligned}
&(0, 0, \cdots, M_T^{\frac{1}{2}}\mathbf{i}, 0, \cdots, 0, -M_L^{\frac{1}{2}}\mathbf{i}, 0, \cdots, 0), \\
&(0, 0, \cdots, M_T^{\frac{1}{2}}\mathbf{j}, 0, \cdots, 0, -M_L^{\frac{1}{2}}\mathbf{j}, 0, \cdots, 0), \\
&(0, 0, \cdots, M_T^{\frac{1}{2}}\mathbf{k}, 0, \cdots, 0, -M_L^{\frac{1}{2}}\mathbf{k}, 0, \cdots, 0).
\end{aligned} \tag{6}$$

Thus we have a $3(N - 1)$-dimensional vector space (the subspace of E^{3N} with $\sum M_i^{\frac{1}{2}}\mathbf{k}_i = 0$), and a sequence of $4 + 3S$ vectors in this space (breaking the momentum transfer at each interaction into a sequence of three momentum transfers, as indicated above). In the sequence, two successive vectors differ by a vector parallel to one of the $3[N(N - 1)/2]$ possible directions (three directions are associated to every interaction). These $3[N(N - 1)/2]$ directions will be called preferred directions.

The integration over internal momenta is an integration over all possible sequences of $3S + 2$

vectors consistent with the direction restrictions associated with the interactions.

We define, in the usual sense, the connectivity properties of the perturbation diagram. The diagram is M-connected if the sequence of $3S + 3$ vectors can be partitioned into M subsequences, and at most M subsequences, such that the vectors in each subsequence are a spanning set. (The sequence of $3S + 3$ vectors is the sequence of preferred directions associated with the diagram.) This is equivalent to being able to subdivide the diagram into a sequence of M diagrams, and at most M diagrams, such that each subdiagram is connected.

As we will be concerned with analytic continuation, we will allow the momenta to become complex. We restrict our attention to distortions of the original real contour that project one-to-one onto the original contour under the projection that sends each complex momentum onto its real part. Call these semiflat contours. Such a contour is specified by associating with each sequence of $3S + 2$ real vectors in the $3(N - 1)$-dimensional space, subject to the conditions that successive vector differences are parallel to the appropriate preferred directions, a second sequence of $3S + 2$ vectors (the imaginary parts of the momenta), with the same conditions on successive vector differences.

The contour of integration may be distorted in any bounded region provided the integrand is analytic through the region of distortion. We will also allow distortions of the contour at infinity, a procedure that must be studied separately. The analyticity of the integrand involves the analyticity of the potentials and the analyticity of the energy denominators. For the first result we aim at below, we require the analyticity properties of the potential in momentum space to be that of a superposition of Yukawa potentials with a minimum mass greater than zero. We will also give a similar result for the case of momentum-space analyticity of the potentials in a strip about the real momentum axis as is obtained if, in coordinate space,

$$|V(x)| < ce^{-\alpha|x|}, \quad \alpha > 0 \tag{7}$$

for some c and α. The essential difficulties and interest arise from consideration of analyticity properties of the energy denominators.

The two results we obtain are among many that are obtainable by a similar procedure. In the conclusion we will indicate further directions. The type of theorems we now aim at are chosen through a consideration of those that seem most useful in the study of many-body scattering theory along the lines of Faddeev's study of the three-body system.[1] The type of rotation of the momentum we consider in the theorems below, applied in a more general context, (outside of perturbation theory) could (hopefully) provide a method for dealing with the singular limit of the energy approaching the real axis.

III. GEOMETRICAL THEOREMS

We consider the following basic situation, motivated as an abstraction of the considerations of the last section. There is a Euclidean space E and a finite set P of unit vectors in E (specifying preferred directions) that together span E.

Lemma 1: There is a constant c such that if any vector $\mathbf{v}$ in E is expressed as a linear combination of linearly independent vectors, $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_g$ from P, each of the expansion coefficients is less than $c\,|\mathbf{v}|$.

Proof: Writing

$$\mathbf{v} = \alpha_1\mathbf{a}_1 + \cdots + \alpha_g\mathbf{a}_g,$$

we take the inner product of $\mathbf{v}$ with each of the unit vectors $\mathbf{a}_1, \cdots, \mathbf{a}_g$:

$$\mathbf{a}_i \cdot \mathbf{v} = \sum_k \alpha_k \mathbf{a}_i \cdot \mathbf{a}_k, \quad i = 1, 2, \cdots, g.$$

This set of equations has a unique solution for the $\mathbf{a}_k$, since the matrix of inner products $(\mathbf{a}_i \cdot \mathbf{a}_k)$ is nonsingular. Because the value of the determinants of all such matrices formed from linearly independent subsets of P, being finite in number, is bounded away from zero, the lemma follows.

Lemma 2: Let $\mathbf{u}$ be an arbitrary unit vector in E, and let $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_g$ be any subset of vectors from P. Let

$$\mu = \max_{i=1,\cdots,g} (|\mathbf{u} \cdot \mathbf{a}_i|),$$

and let θ be the angle between $\mathbf{u}$ and the subspace spanned by $\mathbf{a}_1, \cdots, \mathbf{a}_g$ (i.e., the smallest angle possible between $\mathbf{u}$ and a vector in the subspace). Then

$$|\cos\theta| \leq gc\mu,$$

with c the constant of Lemma 1.

Proof: This follows immediately from Lemma 1 upon writing any vector in the subspace as a combination of the $\mathbf{a}_k$.

[1] L. D. Faddeev, "Mathematical Problems of the Quantum Theory of Scattering for a Three Particle System," AERE-Trans 1002, United Kingdom Atomic Energy Authority Translation.

Lemma 2': There is a minimum angle between the subspaces spanned by any subset of P and any vector from P not in the subspace.

Proof: This follows since there are only a finite number of such pairs of subspaces and vectors not in the subspace formable from P.

Main Theorem: Let $(\mathbf{a}_{\alpha(1)}, \mathbf{a}_{\alpha(2)}, \cdots, \mathbf{a}_{\alpha(g)})$ be a sequence of vectors from the set P, possibly with repetitions. The connectivity of this sequence is defined as before as the greatest number of subsequences the sequence can be partitioned into:

$$(\mathbf{a}_{\alpha(1)}, \mathbf{a}_{\alpha(2)}, \cdots, \mathbf{a}_{\alpha[\beta(1)]})(\mathbf{a}_{\alpha[\beta(1)+1]}, \cdots) \cdots (\mathbf{a}_{\alpha[\beta(M-1)+1]}, \cdots, \mathbf{a}_{\alpha(g)}), \quad (8)$$

such that each subsequence contains a spanning subset. Let $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{g+1})$ be a sequence of vectors from E with the property that $\mathbf{x}_{i+1} - \mathbf{x}_i$ is parallel to $\mathbf{a}_{\alpha(i)}$. Let a be a fixed number greater than zero. There is a $\delta > 0$ and an M_0 such that, for every two sequences related as the above, there is a sequence $(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{g+1})$ of vectors with

(a) $\mathbf{y}_{i+1} - \mathbf{y}_i$ parallel to $\mathbf{a}_{\alpha(i)}$,

(b) $\mathbf{y}_1 = 0$,

$\mathbf{y}_{g+1} = \mathbf{x}_{g+1}$,

(c) $\mathbf{y}_i \cdot \mathbf{x}_i \geq 0$ if $a - \delta \leq |\mathbf{x}_i| \leq a + \delta$,

provided the connectivity of the $\mathbf{a}$ sequence is greater than M_0.

It is easy to show that, in general, the theorem would not be true if instead of (c) above we tried

(c') $\mathbf{y}_i \cdot \mathbf{x}_i \geq 0$ if $|\mathbf{x}_i| \leq a$.

However, it does not seem difficult to modify the proof to include the stronger theorem obtained with (c) replaced by

(c'') $\mathbf{y}_i \cdot \mathbf{x}_i \geq 0$ if $0 < b \leq |\mathbf{x}_i| \leq a$,

a and b given, limited only by $0 < b \leq a$. This generalization, important for some further applications, is not explored here.

Lemma 3: If $|\mathbf{x}_i - \mathbf{y}_i| < a - \delta$, the inner product condition (c) is automatically satisfied.

Proof: This is immediate.

Lemma 4: If $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{\beta(1)}$ all lie outside the shell $a - \delta \leq |\mathbf{x}| \leq a + \delta$ [$(\mathbf{a}_{\alpha(1)}, \cdots, \mathbf{a}_{\alpha[\beta(1)]})$ a spanning set as indicated in the statement of the theorem], then $\mathbf{y}_i$ can be chosen equal to $\mathbf{x}_i$ for $i = \beta_1 + 1, \beta_1 + 2, \cdots, g$ and $\mathbf{y}_1, \cdots, \mathbf{y}_{\beta(1)}$ can be chosen so that this set of $\mathbf{y}_i$ satisfies conditions of the theorem.

Proof: Expand $\mathbf{x}_{\beta(1)+1}$ in terms of $\mathbf{a}_1, \cdots, \mathbf{a}_{\alpha[\beta(1)]}$,

$$\mathbf{x}_{\beta(1)+1} = \alpha_1 \mathbf{a}_1 + \cdots + \alpha_{\beta(1)} \mathbf{a}_{\alpha[\beta(1)]}.$$

Pick

$$\mathbf{y}_1 = 0, \quad \mathbf{y}_2 = \alpha_1 \mathbf{a}_1, \quad \mathbf{y}_3 = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2,$$
$$\mathbf{y}_{\beta(1)+1} = \mathbf{x}_{\beta(1)+1}.$$

These choices satisfy the necessary conditions.

Lemma 5: There is an ε such that, if $|\mathbf{x}_i - \mathbf{y}_i| < \varepsilon$ for some i and if $\mathbf{a}_{\alpha(i)}, \cdots, \mathbf{a}_{\alpha(s)}$ form a spanning set, then $\mathbf{y}_{s+1}$ can be chosen equal to $\mathbf{x}_{s+1}$ and intermediate y's can be found consistent with the conditions of the theorem.

Proof: Pick a linearly independent set of spanning vectors among $\mathbf{a}_{\alpha(i)}, \cdots, \mathbf{a}_{\alpha(s)}$. Now expand

$$\mathbf{x}_i - \mathbf{y}_i = \alpha_i \mathbf{a}_{\alpha(i)} + \cdots + \alpha_s \mathbf{a}_{\alpha(s)} \quad (9)$$

with coefficients different from zero only among the linearly independent spanning vectors. Pick

$$\begin{aligned} \mathbf{y}_{i+1} &= \mathbf{y}_i + \alpha_i \mathbf{a}_{\alpha(i)} + (\mathbf{x}_{i+1} - \mathbf{x}_i), \\ \mathbf{y}_{i+2} &= \mathbf{y}_{i+1} + \alpha_{i+1} \mathbf{a}_{\alpha(i+1)} + (\mathbf{x}_{i+2} - \mathbf{x}_{i+1}), \\ \mathbf{y}_{s+1} &= \mathbf{x}_{s+1}. \end{aligned} \quad (10)$$

It follows that

$$\begin{aligned} \mathbf{y}_{i+1} - \mathbf{x}_{i+1} &= (\mathbf{y}_i - \mathbf{x}_i) + \alpha_i \mathbf{a}_{\alpha(i)}, \\ \mathbf{y}_{i+2} - \mathbf{x}_{i+2} &= (\mathbf{y}_i - \mathbf{x}_i) + \alpha_i \mathbf{a}_{\alpha(i)} + \alpha_{i+1} \mathbf{a}_{\alpha(i+1)}. \end{aligned} \quad (11)$$

Thus

$$|\mathbf{y}_{i+r} - \mathbf{x}_{i+r}| \leq |\mathbf{y}_i - \mathbf{x}_i| + \sum |\alpha_i|. \quad (12)$$

Since $|\alpha_i| < c\,|\mathbf{x}_i - \mathbf{y}_i|$ by Lemma 1, if D is the dimension of the space,

$$|\mathbf{y}_{i+r} - \mathbf{x}_{i+r}| \leq (1 + Dc)\,|\mathbf{y}_i - \mathbf{x}_i|, \quad r = 0, 1, \cdots. \quad (13)$$

If $|\mathbf{y}_i - \mathbf{x}_i|$ is small enough, then $|\mathbf{y}_{i+r} - \mathbf{x}_{i+r}| < a - \delta$, which, by Lemma 3, guarantees that these y's work.

Lemma 6: If $|\mathbf{x}_i - \mathbf{y}_i| < a - \delta$ for some i, then a sequence of y's can be found starting with $\mathbf{y}_i$ which satisfy the conditions of the theorem such that $|\mathbf{x}_k - \mathbf{y}_k|$ is monotonically decreasing; and each time the $\mathbf{a}$'s pass through a spanning sequence, the $|\mathbf{x}_k - \mathbf{y}_k|$ decreases at least by some factor r, $r < 1$. r depends only on the vectors in P.

Proof: Pick y's successively by minimizing $|\mathbf{x}_k - \mathbf{y}_k|$ at each stage; that is, $\mathbf{y}_{k+1} - \mathbf{y}_k = \lambda_k \mathbf{a}_{\alpha(k)}$ with λ_k chosen to minimize $|\mathbf{x}_{k+1} - \mathbf{y}_{k+1}|$. It is clear that $|\mathbf{x}_k - \mathbf{y}_k| < a - \delta$ for all k, so the conditions of the theorem are satisfied. It remains to see that $|\mathbf{x}_k - \mathbf{y}_k|$ decreases by some factor as the $\mathbf{a}$'s go through a spanning set. We

omit the proof of this fact; it may be constructed along the lines of the following lemma.

Lemma 7: If $a - \delta \leq |\mathbf{x}_1| \leq a + \delta$ (with $\mathbf{y}_1 = 0$), an allowable sequence of $\mathbf{y}$'s can be found such that $|\mathbf{x}_{s+1} - \mathbf{y}_{s+1}| < a - \delta$, provided $\mathbf{a}_1, \cdots, \mathbf{a}_s$ are a spanning set (from P) and δ is small enough. δ may be chosen, depending on a and the set P, but independently of the $\mathbf{x}$'s.

The main geometric theorem follows from Lemmas 3–7. If $\mathbf{x}$ never hits the spherical shell as the $\mathbf{a}$'s pass through the first spanning set, Lemma 4 shows the existence of the $\mathbf{y}$'s. If some $\mathbf{x}$ hits the spherical shell as the $\mathbf{a}$'s pass through the first spanning set, use Lemma 7 to obtain $|\mathbf{x}_j - \mathbf{y}_j| < a = \delta$. Then use Lemma 6 to get $|\mathbf{x}_k - \mathbf{y}_k|$ small enough to use Lemma 5, completing the sequence of $\mathbf{y}$'s.

Proof of Lemma 7: Let D be the dimension of the Euclidean space, $\theta_{\min}$ be the angle of Lemma 2′, and c be the constant of Lemma 1. Choose $\theta_1, \theta_2, \cdots, \theta_{D-1}$ such that

$$\begin{aligned} \theta_1 &= \phi_1 > 0, \\ \theta_{i+1} - 2\sin^{-1}(Dc \sin\theta_i) &= \phi_{i+1} > 0, \\ \theta_{\min} - 2\sin^{-1}(Dc \sin\theta_{D-1}) &= \phi_D > 0. \end{aligned} \tag{14}$$

Let $\phi_{\min} = \min(\phi_1, \phi_2, \cdots, \phi_D)$; pick δ such that $(a + \delta)\cos\phi_{\min} < a - \delta$. By assumption, $a - \delta \leq |\mathbf{x}_1| \leq a + \delta$. Let $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_R$ be a spanning set, but $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{R-1}$ not. Let $\mathbf{b}_1 = \mathbf{a}_1$, $\mathbf{b}_i = \mathbf{a}_{\beta(i)}$, $\mathbf{b}_i$ the first of the $\mathbf{a}_j$'s linearly independent of $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_{i-1}$. $\mathbf{b}_D$ will equal $\mathbf{a}_R$. Let $\psi_i = \pi/2 - \angle(\mathbf{b}_i, \mathbf{x}_1)$. By $\angle(\mathbf{b}_i, \mathbf{x}_1)$ we mean the angle between the vectors $\mathbf{b}_i$ and $\mathbf{x}_1$ (minus their orientation), an angle between 0 and $\pi/2$. Let ψ_k be the first of the ψ's with $\psi_k \geq \phi_k$. Then $\psi_1 < \phi_1, \psi_2 < \phi_2, \cdots, \psi_{k-1} < \phi_{k-1}$, and $\psi_k \geq \phi_k$. Such a k must exist due to the nature of the ϕ's. If $|\mathbf{x}_{\beta(k)}| \leq a + \delta$, let

$$\mathbf{y}_1 = \mathbf{y}_2 = \cdots = \mathbf{y}_{\beta(k)} = 0, \quad \mathbf{y}_{\beta(k)+1} = \lambda\mathbf{b}_k,$$

λ chosen to minimize $|\mathbf{x}_{\beta(k)+1} - \mathbf{y}_{\beta(k)+1}|$. If $|\mathbf{x}_{\beta(k)}| > a + \delta$, with

$$\begin{aligned} |\mathbf{x}_g| \leq a + \delta, \quad |\mathbf{x}_{g+1}| > a + \delta, \\ |\mathbf{x}_{g+2}| > a + \delta, \quad \cdots, \quad |\mathbf{x}_{\beta(k)}| > a + \delta, \end{aligned}$$

then pick

$$\begin{aligned} \mathbf{y}_1 = \mathbf{y}_2 = \cdots = \mathbf{y}_g = 0, \\ \mathbf{y}_{g+1} = \mathbf{x}_{g+1} - \mathbf{x}_g, \\ \vdots \\ \mathbf{y}_{\beta(k)} = \mathbf{x}_{\beta(k)} - \mathbf{x}_g, \\ \mathbf{y}_{\beta(k)+1} = \lambda\mathbf{b}_k + \mathbf{y}_{\beta(k)}, \end{aligned}$$

λ chosen to minimize $|\mathbf{x}_{\beta(k)+1} - \mathbf{y}_{\beta(k)+1}|$. One can check that $|\mathbf{y}_{\beta(k)+1} - \mathbf{x}_{\beta(k)+1}| < a - \delta$ in both cases.

IV. ANALYTIC CONTINUATION

Theorem 1: If $E = a^2 - i\epsilon$, $a > 0$, $\epsilon > 0$, and if all the potentials involved have the analyticity of a superposition of Yukawa potentials with a minimum mass greater than zero, then there is an M_0 and an η such that any perturbation-theory diagram with connectivity $\geq M_0$ considered a function of θ, $\mathbf{k}_{\text{fin}} \to (1 + i\theta)\mathbf{k}_{\text{fin}}$[$\mathbf{k}_{\text{in}}$ and $\mathbf{k}_{\text{fin}}$ being real] has the property that the intermediate-state integration contour can be distorted through analytic regions of the integrand as θ varies from $\theta = 0$ to $\theta = \eta$. η will depend on:

(a) N, (b) $M_1, M_2, \cdots, M_N$, (c) the minimal mass in the Yukawa decomposition of the potentials, (d) a, (e) $\mathbf{k}_{\text{in}}$, but not on $\mathbf{k}_{\text{fin}}$.

It is important to notice that the requirement of high connectivity is essential. For example, with $N = 3$, the simple second-order diagram with interactions V_{12} followed by V_{23} cannot be continued as above. With $\mathbf{k}_{\text{in}}$ and $\mathbf{k}_{\text{fin}}$ specified in this case, the possible rotation angle may approach zero as $\epsilon \to 0$. The existence of the rotation in the limit $\epsilon \to 0$ is most crucial.

It is reasonable to suppose that by increasing the connectivity requirements the amount of rotation can be increased beyond that allowed by the theorem. This is an interesting question, but possibly not important to the proposed application of the theorem.

To prove Theorem 1, we observe that the following bounds can be put on $|\mathbf{y}_i|$ and $|\mathbf{y}_{i+1} - \mathbf{y}_i|$ by examining the construction of Lemmas 3–7:

$$\begin{aligned} |\mathbf{y}_i| &\leq c_1 |\mathbf{x}_1| + c_2 |\mathbf{x}_i|, \\ |\mathbf{y}_{i+1} - \mathbf{y}_i| &\leq c_3 |\mathbf{x}_{i+1} - \mathbf{x}_i| + c_4 |\mathbf{x}_1|. \end{aligned} \tag{15}$$

We identify the a of the theorem and the a of the geometric construction, and associate the $\mathbf{x}$'s with the real parts of the momenta, the $\mathbf{y}$'s with a constant λ times the imaginary part of the momenta. It is easy to see that, with a fixed a, fixed initial momenta, and minimum mass of the Yukawa potentials, if λ is large enough ($\lambda \geq \lambda_0$, say), the real and imaginary parts form a complex momentum always avoiding the singularities of the potentials and the energy denominators. The inner product condition of the geometric construction guarantees that the denominators are never singular. If Im E is fixed at some negative value, then the sequences of $\mathbf{y}$'s satisfying the conditions of the geometric theorem form a convex set under vector-by-vector addition. We can impose (15) and still maintain a convex set of solutions. Since Im $E < 0$, we can

replace the inner product inequality $\mathbf{x}_i \cdot \mathbf{y}_i \geq 0$ by $\mathbf{x}_i \cdot \mathbf{y}_i > -\tilde{\epsilon}$, $\tilde{\epsilon} > 0$ for some small $\tilde{\epsilon}$, a condition that also possesses a convex solution set. We observe the following lemma.

Contour Construction Lemma: Let E^{R+S} be Euclidean $(R + S)$-dimensional space considered as a product of E^R and E^S, with π the projection onto E^R. Let U be an open set in E^{R+S} with the following two properties:

(a) U projects, under π, onto all of E^R;

(b) The inverse image of each point under the map $\pi: u \to E^R$ is a convex subset of E^S.

Then there exists a differentiable (even C^∞) cross section, i.e., there is a differentiable map $F: E^R \to U$ such that $\pi F: E^R \to E^R$ is the identity map.

Proof: Pick a point $[x, t(x)]$ in U for each point x in E^R. Each such $[x, t(x)]$ is contained in a product neighborhood $u_1(x) \times u_2(x) \subset u$. In the neighborhood $u_1(x)$ of x in E^R, the map $f_x: y \to [y, t(x)]$ is a C^∞ cross section. There is a locally finite refinement V_α of the $u_1(x)$ and a subordinate C^∞ partition of unity ϕ_α. The convexity of the fibers allows this C^∞ partition of unity to provide a global cross section:

$$f = \sum_\alpha \phi_\alpha f_{x(\alpha)},$$

where $V_\alpha \subset u_1[x(\alpha)]$.

This lemma enables us to construct a global contour of integration. $(1/\lambda_0)$ is the η of the theorem. For $0 \leq \theta \leq \eta$, the contour of integration may be chosen to be the momentum surface associated with $\lambda = 1/\theta$. (The lemma is applied with E^R the space of $\mathbf{x}$ sequences and E^S the space of $\mathbf{y}$ sequences.)

If the potentials have analyticity in a strip about real momentum values, and not in the full Yukawa region, the above construction must be modified to bound $|\mathbf{y}_i|$. We restrict ourselves to the following result.

Theorem 2: Let $\mathbf{k}_{in}$, $\mathbf{R}_{fin}$, and $\mathbf{I}_{fin}$ be real vectors, with $\mathbf{R}_{fin}$ and $\mathbf{I}_{fin}$ satisfying the following conditions:

(a) $|\mathbf{I}_{fin}| < 1$;

(b) If $\mathbf{k}_0$ is any vector in the shell $(a - \delta \leq |\mathbf{k}_0| \leq a + \delta)$ and $\mathbf{R}_{fin} - \mathbf{k}_0 = \sum_{i \in J} \alpha_i \mathbf{a}_i$ for some subset indexed by J of vectors from P, then $\mathbf{I}_{fin} - \mathbf{k}_0 = \sum_{i \in J} \beta_i \mathbf{a}_i$ for some β_i.

Then, given any $B > 0$, there is a λ_0 such that the amplitude can be continued in the final momenta written as $\mathbf{k}_{fin} = \mathbf{R}_{fin} + i\lambda \mathbf{I}_{fin}$ from $\lambda = 0$ to $\lambda = \lambda_0$, and $|\mathrm{Im}\, \mathbf{k}_i| < B$ for all intermediate stresses. λ_0 will depend on: (a) $M_1, M_2, \cdots, M_N$, (b) N, (c) B, (d) Re E.

Proof: Let Q be a sphere of radius greater than the maximum of $a + \delta$ and 1, in the setup of the geometric theorem before. As before, solve the geometric problem for $\mathbf{x}_1 = \mathbf{k}_{in}$ and $\mathbf{x}_{fin} = \mathbf{R}_{fin}$. By Lemma 2 there is a sphere Q' greater than Q such that if any subsequence of vectors in the solution set begin with a vector in Q and end with a vector in Q, the solution set can be modified to keep all vectors inside Q'. With such modifications this theorem follows from Theorem 1.

V. CONCLUSION

We indicate problems remaining to be studied. First, the behavior at infinity must be studied sufficiently to justify the contour distortion at infinity. Second, the equations for many-body scattering amplitudes should be continued similarly to the perturbation-theory case. The geometry of this problem seems treatable with only a slight generalization of the theorems included here. If the analytic questions relating to threshhold behavior and behavior at infinity for the integral equations can be treated systematically, a rigorous treatment of the problem of asymptotic completeness may be achieved.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Daughter Regge Trajectory in a Field Theory Model

ARTHUR R. SWIFT*
Department of Physics, University of Wisconsin, Madison, Wisconsin

(Received 27 April 1967)

All contributions of ladder diagrams in a $\lambda\phi^3$ theory with unequal masses which are asymptotically of order $t^{-2}\ln^m t$ are summed to give a family of secondary Regge trajectories. The analysis is carried out with the Mellin transformed scattering amplitude. The daughter Regge trajectory predicted by Freedman and Wang is identified. For equal internal masses on the sides of the ladder, the daughter pole has the following properties: (i) The trajectory function $\alpha(s)$ is constant when calculated to lowest order in the coupling constant; that is, the pole is fixed. (ii) The pole moves in higher order, but the trajectory function has no two-particle cut. (iii) The residue of the pole has no two-particle cut, and vanishes to all orders for equal external masses. The amplitude with unequal internal masses is considered in lowest order, and it is shown that three-particle scattering should continue to dominate the daughter trajectory function although there is a two-particle cut. The daughter pole moves towards the physical region less rapidly than the leading pole, and it develops a smaller imaginary part at the two-particle threshold. It is concluded that a detailed statement about the motion of the daughter pole requires an accurate treatment of three-particle scattering. As a by-product of this work an earlier error in a treatment of the mixing of Regge poles and cuts is corrected.

1. INTRODUCTION

RECENTLY Freedman and Wang[1] have shown that for unequal mass kinematics the requirements of analyticity of the scattering amplitude in the energy and momentum-transfer variables s and t coupled with Regge pole dominance of the asymptotic behavior implies the existence of a whole new set of subsidiary Regge trajectories. For an s-channel Regge pole the positions and residues of the secondary poles, called daughter poles, are determined at $s = 0$ by the requirement that they exactly cancel the $s = 0$ singularities of the leading pole. These singularities occur because the momentum p^2 becomes infinite at $s = 0$, rather than remaining finite as in the equal mass case. The residues of the daughter poles have corresponding singularities adjusted to cancel the leading pole singularity. If $\alpha(0)$ is the position of the leading pole, the nth daughter pole has $\alpha_n^d(0) = \alpha(0) - n$; and the *singular portion* of its reduced residue is proportional to s^{-n} and vanishes, if n is odd, for coupling to equal mass particles. If $\tilde{\beta}(s)$ is the reduced residue of a Regge pole, then near $s = 0$ the reduced residue of the first daughter pole is given by[1]

$$\tilde{\beta}_d(s) = -[(m_1^2 - m_2^2)^2/4s](2\alpha(0) + 1)\tilde{\beta}_l(0) + \text{N.S.}, \tag{1}$$

where N.S. denotes nonsingular terms. The subscripts d and l refer to the daughter and leading poles, respectively. The external masses in the s channel are m_1 and m_2. The nonsingular terms in (1) do not necessarily vanish for $m_1 = m_2$. Durand[2] has extended the analysis of Freeman and Wang to show that daughter poles are a more general phenomenon which should also occur in the scattering of particles with spin in order to cancel kinematical singularities from leading poles; the residues are again required to be singular. These general discussions of daughter poles provide no information about their behavior away from $s = 0$ or about the nonsingular portion of the residue. The physical importance of these poles depends to a great extent on their position as s becomes physical.[1,2] If the daughters move more or less parallel to the leading pole, they should lead to a whole new set of physical particles or resonances with unusual quantum numbers.[1] On the other hand, some mechanism might act to prevent these poles from ever reaching the physical region. They might move more slowly with increasing s and never reach $l = 0$ or they might interact with other singularities and disappear from the physical sheet of the amplitude. Although the existence of these new poles was deduced on kinematical ground, the answers to questions about behavior away from $s = 0$ apparently involve dynamics.

One approach to the problem involves solving the Bethe–Salpeter equation, usually in the ladder approximation. If the well-known exact solutions to the bound state equations[3] are reinterpreted in terms of Regge poles, it is possible to argue that the daughter poles move more or less parallel to the leading poles and reach physical values of the energy and angular

* Work supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the U.S. Atomic Energy Commission under contract AT(11-1)-881, #COO-881-105.

[1] D. Freedman and J. Wang, Phys. Rev. Letters **17**, 569 (1966); Phys. Rev. **153**, 1596 (1967).

[2] L. Durand, III, Phys. Rev. Letters **18**, 58 (1967).

[3] R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954); J. S. Goldstein, *ibid.* **91**, 1516 (1953).

momentum. However, some of the residues of these daughter poles have the wrong sign. Since this approach does not provide an analytic expression for the trajectory functions, there is some difficulty in determining why the ladder approximation should fail here. Recently, the Bethe–Salpeter equation has been solved numerically and the daughter poles followed away from their $s = 0$ position.[4] This work has shown that the daughter poles move somewhat more slowly than the leading poles, but does not provide any further information. In particular it does not lead to an analytic expression for the trajectory function and it says nothing about the residues.

In this paper we apply the techniques of high-energy perturbation theory to an infinite sum of ladder diagrams.[5] The justification for using perturbation theory as a laboratory for investigating questions of this kind has been given many times before. Regge poles, Regge cuts, asymptotic behavior of production amplitudes, and many other problems have been investigated by this technique, and it has been found to be quite useful and not misleading. The model considered here satisfies all the postulates necessary to prove the existence of daughter poles, contains a well-studied leading Regge pole, and leads to an analytic expression for the trajectory and residue functions of the first daughter pole. Having these analytic expressions, we are able to show that its residue vanishes to all order in the coupling constant, for all s, in the limit that $m_1 = m_2$. As a corollary to this result, the motion of the daughter poles is dominated by the effects of three-particle scattering. Since both this model and the ladder approximation Bethe–Salpeter equation treat three-particle intermediate states very badly, any detailed statement about daughter-pole motion based on these models is unlikely to be correct. The main results of this paper have been published in a letter[6]; we now present the details of the calculation.

Polkinghorne[7] has studied ladder diagrams and, by summing all terms of the form $(\ln t)^m/t$, obtained the complete expression for a Regge pole which approaches $l = -1$ in the weak coupling or infinite s limits. That analysis was carried out for diagrams in which all masses, internal and external, were equal;

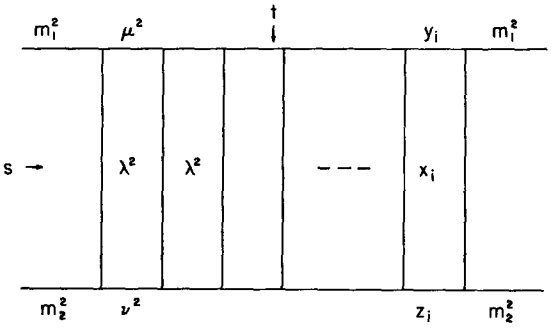


FIG. 1. A general ladder diagram of type studied with the various masses carried by each line labeled. In addition the Feynman parameters y_i, z_i, x_i belonging to the diagram are shown.

however, it is trivially extended to a ladder diagram of the type shown in Fig. 1. These ladder diagrams will not satisfy two-particle unitarity unless $m_1^2 = \mu^2$ and $m_2^2 = \nu^2$; however, the existence proof for the daughter poles is independent of unitarity. In addition, the daughter pole is presumably a dynamical entity whose coupling to two particles depends on kinematics but whose motion is largely independent of this coupling. We have generalized the approach of Polkinghorne[7] to sum all contributions of the ladder diagrams which are proportional asymptotically to $(\ln t)^m/t^2$ in order to obtain a set of Regge poles near $l = -2$. A lowest-order summation (in the coupling constant) has been performed previously in connection with Regge cuts.[8] There is a minor error in this earlier analysis due to an improper ordering of limits; in addition, it was carried out for all masses equal. Since the method of that paper could not easily be generalized to problems involving unequal masses, a completely different approach has been developed.

Of the several techniques developed for extracting the asymptotic behavior of Feynman integrals, the most powerful has proved to be that utilizing the Mellin transform of the scattering amplitude.[5] The Mellin transform of the scattering amplitude is discussed fully in I and Ref. 5, and the relation between Mellin poles and Regge poles is discussed thoroughly in II. As shown in these articles, the leading pole is obtained by summing all portions of the amplitude singular near $\alpha = -1$, where α is the Mellin transform variable. Since every Mellin pole can be identified with a Regge pole,[8] a Mellin trajectory is also a Regge trajectory.

The secondary poles are obtained by summing terms singular near $\alpha = -2$. The summation procedure for the secondary poles is considerably more complicated than that used near $\alpha = -1$, and much

[4] D. Freedman and J. Wright (private communication).

[5] For a complete discussion of the techniques and justification of high-energy perturbation theory, as well as an account of the various problems to which it has been applied, see R. Eden, P. Landshoff, D. Olive, and J. Polkinghorne, *The Analytic S-Matrix* (Cambridge University Press, New York, 1966), Chap. 3. An extensive list of references is given there.

[6] A. R. Swift, Phys. Rev. Letters **18**, 813 (1967).

[7] J. C. Polkinghorne, J. Math. Phys. **5**, 431 (1964). Our notation generally follows this paper, which will be referred to as I.

[8] A. R. Swift, J. Math. Phys. **6**, 1472 (1965). This paper will be referred to as II.

of this paper is concerned with resolving this difficulty. The result is an amplitude that has four poles near $\alpha = -2$. One is a trivial recurrence of the leading pole, and one is the daughter pole. The other two are new dynamical poles which are present even in the equal mass problem. We concentrate on isolating and identifying the daughter pole. In the limit that the masses on the sides of the ladder are equal ($\mu = \nu$), the daughter pole can be isolated completely and its residue determined. The trajectory and residue functions have their first singularity at the three-particle threshold. To lowest order, the daughter trajectory is a constant for $\mu = \nu$. If $\mu \neq \nu$, the situation is quite complicated. The formal solution to all orders does not factor; even in lowest order the daughter pole is one of three roots of a cubic equation. In general it is one of the zeros of a 4×4 determinant. However, we are able to show that to lowest order the daughter pole still does not move parallel to the leading pole.

In the next section we discuss briefly the calculation of the leading Regge pole for a sum of ladder diagrams; we obtain the residue and trajectory function for the leading pole for comparison later with the daughter pole. The third section is devoted to a detailed treatment of the complete summation of terms of the scattering amplitude singular near $\alpha = -2$. In Sec. 4 we set $\mu = \nu$ and discuss the amplitude in lowest order and identify the daughter pole. Then we look at the complete amplitude and isolate the daughter trajectory and its residue to all orders. Section 5 treats the case with unequal masses ($\mu \neq \nu$) to lowest order to check the conclusions reached in Sec. 4. Appendix A contains a number of definitions and identities for various functions of the Feynman parameters that occur in Sec. 4. Appendices B and C present explicit expressions for some of the matrices and energy-dependent functions used in Secs. 4 and 5. In Appendix D the work of II on the mixing of poles and cuts is corrected for the error mentioned above.

2. THE LEADING REGGE POLE

We start with the Mellin transform of the scattering amplitude for a ladder diagram with $N + 1$ rungs as given by Eq. (6) of I.

$$L_N(\alpha, s) = \Gamma(-\alpha) g^2 \left(\frac{g^2}{16\pi^2}\right)^N \int_0^\infty \prod_{i=1}^N dy_i\, dz_i \times \prod_{i=1}^{N+1} dx_i (x_1 \cdots x_{N+1})^\alpha \frac{e^{-Q_N}}{\Delta_N^{\alpha+2}}. \quad (2)$$

The parameters x_i, y_i, z_i are labeled in Fig. 1. The function Q_N is linear in s and the various masses[5]:

$$Q_N = \mu^2 \sum_1^N y_i + \nu^2 \sum_1^N z_i + \lambda^2 \sum_1^{N+1} x_i - m_1^2[(g_N^y + h_N^y)/\Delta_N] - m_2^2[(g_N^z + h_N^z)/\Delta_N] - s f_N/\Delta_N. \quad (3)$$

The determinant for the ladder Δ_N is well known,[5] and the other functions of the parameters occurring in (3) are given explicitly in Appendix A. Those portions of the scattering amplitude singular near $\alpha = -1$ have been summed in I. The full amplitude is given by the sum of $L_N(\alpha, s)$, $N \geq 0$. We repeat the procedure here briefly to emphasize the difficulties encountered near $\alpha = -2$. Equation (2) is integrated by parts over the x_i; the result is an integrand of the form

$$(-1)^{N+1} \left\{ \prod_{i=1}^{N+1} \frac{[x_i^{\alpha+1} - (\sigma_i \sigma_{i-1})^{\alpha+1}]}{\alpha + 1} + \frac{(\sigma_i \sigma_{i-1})^{\alpha+1}}{\alpha + 1} \right\} \times \frac{\partial^{N+1}}{\partial x_1 \cdots \partial x_{N+1}} \left(\frac{e^{-Q_N}}{\Delta_N^{\alpha+2}} \right), \quad (4)$$

where $\sigma_i = y_i + z_i$, $\sigma_0 = \sigma_{N+1} = 1$. We have added and subtracted a term $(\sigma_i \sigma_{i-1})^{\alpha+1}$ in the $(N + 1)$-fold product. For this leading pole we could have added and subtracted $(1)^{\alpha+1}$ as was done in I, but for $\alpha = -2$ several unpleasant divergences appear if this is done. The term $[x_i^{\alpha+1} - (\sigma_i \sigma_{i-1})^{\alpha+1}]/(\alpha + 1)$ is regular at $\alpha = -1$. The product in (4) can be expanded to give a sum of 2^{N+1} terms. If $(\sigma_i \sigma_{i-1})^{\alpha+1}/(\alpha + 1)$ occurs in the sum, the corresponding x_i will be absent and the integral over x_i can be done immediately to give

$$-\frac{e^{-Q_{1N}}}{(\Delta_{1N})^{\alpha+2}}\bigg|_{x_i=0} = -\frac{e^{-Q_{1i-1}}}{(\Delta_{1i-1})^{\alpha+2}} \times \frac{e^{-Q_{iN}}}{(\Delta_{iN})^{\alpha+2}}.$$

The double subscript notation used here is explained in the first paragraph of Appendix A. This simple factorization property allows the singular part of the amplitude to be summed readily. In $L_N(\alpha, s)$ there are terms with zero to $N + 1$ rungs contracted, and these are distributed all possible ways. [A diagram with the nth rung contracted[5] has the corresponding Feynman parameter x_n, set equal to zero in (2) rather than integrated over.] Let us define the functions

$$F(\alpha, s) = -\sum_{N=1}^\infty \left(\frac{-g^2}{16\pi^2}\right)^N \int_0^\infty \prod_{i=1}^N dy_i\, dz_i \times \prod_{j=1}^{N-1} \frac{[x_j^{\alpha+1} - (\sigma_{j+1}\sigma_j)^{\alpha+1}]}{\alpha + 1} dx_j \times (\sigma_1 \sigma_N)^{\alpha+1} \frac{\partial^{N-1}}{\partial x_1 \cdots \partial x_{N-1}} \left(\frac{e^{-Q_{1N}}}{\bar{\Delta}_N^{\alpha+1}} \right), \quad (5)$$

and

$$G(\alpha, s) = 1 + \sum_{N=1}^{\infty} \left(\frac{-g^2}{16\pi^2}\right)^N \int_0^\infty \prod_{i=1}^{N} dy_i\, dz_i \times \prod_{j=1}^{N} \left[\frac{x_j^{\alpha+1} - (\sigma_{j-1}\sigma_j)^{\alpha+1}}{\alpha+1}\right] dx_j \times (\sigma_N)^{\alpha+1} \frac{\partial^N}{\partial x_1 \cdots \partial x_N}\left[\frac{\exp(-\bar{Q}_{1N})}{(\bar{\Delta}_{1N})^{\alpha+2}}\right], \quad (6)$$

where $\bar{Q}_{1N}, \bar{\Delta}_{1N}$ are the Feynman functions for a ladder diagram with both ends contracted and $N-1$ internal rungs; $\bar{Q}'_{1N}, \bar{\Delta}'_{1N}$ refer to a ladder with just one end contracted. If the sum over N of $L_N(\alpha, s)$ is interchanged with the sum over m, the number of contracted lines, it is easy to see that

$$\sum_{N=0}^{\infty} L_N(\alpha, s) = \frac{g^2\Gamma(-\alpha)G^2(\alpha, s)}{\alpha+1} \sum_{m=1}^{\infty}\left[\frac{F(\alpha, s)}{\alpha+1}\right]^{m-1} = \frac{g^2\Gamma(-\alpha)G^2(\alpha, s)}{\alpha+1-F(\alpha, s)}, \quad (7)$$

where we have discarded a term that is nonsingular in the neighborhood of $\alpha = -1$. The leading Regge pole is given by the solution of

$$\alpha + 1 = F(\alpha, s). \quad (8)$$

The extra $\sigma_i^{\alpha+1}$ factors in (5) do not change the solution of (8) from that given in I. The Mellin residue of the pole in (7) is given by

$$\gamma_i(s) = \frac{g^2\Gamma(-\alpha)G^2(\alpha, s)}{1 - (d/d\alpha)F(\alpha, s)}, \quad (9)$$

where α is given by the solution of (8). The Mellin residue is related to the Regge residue by[8]

$$\gamma(s) = -\frac{(2\alpha+1)[\Gamma(-\alpha)]^2}{4^\alpha 2\pi \cos \pi\alpha \Gamma(-2\alpha)}\beta(s). \quad (10)$$

Since it is more convenient, we shall work with Mellin residues.

Equations (6) and (8) constitute an equation for $\alpha + 1$. If we define a function $Y(\alpha, s)$ by

$$Y(\alpha, s) = -\frac{1}{(\alpha+2)} \sum_{N=1}^{\infty}\left(\frac{g^2}{16\pi^2}\right)^N \frac{1}{(\alpha+1)^N} \times \int_0^\infty dy_j\, dz_j \prod_{i=1}^{N-1} dx_i \frac{[x_i^{\alpha+2} - (\sigma_i\sigma_{i+1})^{\alpha+2}]}{\alpha+2} \times \frac{\partial^{2(N-1)}}{\partial x_1^2 \cdots \partial x_{N-1}^2}\left[(\sigma_1\sigma_N)^{\alpha+2}\frac{e^{-Q_N}}{\bar{\Delta}_N^{\alpha+2}}(x_1 \cdots x_{N-1})\right], \quad (11)$$

then a term-by-term analysis shows that

$$\alpha + 2 = -(\alpha+2)Y(\alpha, s) \quad (12)$$

has the same solution for $\alpha + 2$ that (8) has for $\alpha + 1$. Equation (12) occurs in the fourth section and is associated with the Mellin recurrence of the leading Regge pole. From the formulas given in II we see that for every leading Regge pole there will occur a series of secondary Mellin poles, the first of which is given by

$$L'(\alpha, s) = \frac{-2\alpha_l(s)\gamma_l(s)P^2}{\alpha + 1 - \alpha_l(s)}. \quad (13)$$

Finally, we give the first-order solution of (8).

$$\alpha + 1 = \frac{g^2}{16\pi^2}K(s) = \frac{g^2}{16\pi^2}\int_0^\infty \frac{dy\, dz}{\sigma} \exp\left[(-\mu^2 y - \nu^2 z) + \frac{sxy}{\sigma}\right]. \quad (14)$$

The trajectory function of the leading pole has its first singularity at the two-particle threshold.

3. SUMMATION OF $\alpha = -2$ SINGULARITIES

To sum those portions of $L_N(\alpha, s)$ singular near $\alpha = -2$, we integrate $L_N(\alpha, s)$ by parts twice, and replace (4) by

$$\frac{1}{(\alpha+1)^{N+1}} \prod_{i=1}^{N+1}\left(\frac{[x_i^{\alpha+2} - (\sigma_i\sigma_{i-1})^{\alpha+2}]}{\alpha+2} + \frac{(\sigma_i\sigma_{i-1})^{\alpha+2}}{\alpha+2}\right) \times \frac{\partial^{2(N+1)}}{\partial x_1^2 \cdots \partial x_{N+1}^2}\left(\frac{e^{-Q_N}}{\Delta_N^{\alpha+2}}\right). \quad (15)$$

The product is expanded in the same manner as before and those terms not containing x_i are integrated immediately to give

$$-\frac{\partial}{\partial x_i}(e^{-Q_N}/\Delta_N^{\alpha+2})_{x_i=0}. \quad (16)$$

Unfortunately, (16) does not factor simply, and herein lies the great complication in summing $L_N(\alpha, s)$ near $\alpha = -2$. Since we want to obtain equations similar to (5) and (6) valid to all orders in the coupling constant, the approach we shall use is totally different from that developed in II. In general we have to consider a series of derivatives with respect to $x_i, x_j, \cdots, x_m, x_k$. We order the x's so that $i < j < , \cdots, < m < k$ and then perform the differentiations in the order $i, k, j, \cdots, m$. The reason for this ordering will become apparent. Using the various functions and relations given in Appendix A, we find that (16) becomes

$$\frac{\partial}{\partial x_i}\left(\frac{e^{-Q_N}}{\Delta_N^{\alpha+2}}\right)_{x_i=0} = \left(\frac{e^{-Q_{1i-1}}}{\Delta_{1i-1}^{\alpha+2}}\right) \times \left(\frac{e^{-Q_{iN}}}{\Delta_{iN}^{\alpha+2}}\right) \times M_{1iN}, \quad (17)$$

where

$$M_{1iN} = -(\alpha+2)\left(\frac{\Delta_{1i-2}}{\Delta_{1i-1}} + \frac{\Delta_{i+1N}}{\Delta_{iN}}\right) - \lambda^2 + m_1^2\left(K^v_{1i-1} + \frac{G_{1i-1}}{\Delta_{1i-1}}\frac{\bar{g}^v_{iN}}{\Delta_{iN}} + H^v_{iN} + \frac{\bar{h}^v_{1i-1}}{\Delta_{1i-1}}\frac{G_{i+1N+1}}{\Delta_{iN}}\right) + m_2^2(y \to z) + s\left(F^1_{iN} + F^2_{1i-1} + \frac{[\bar{h}^v_{1i-1}\bar{g}^z_{iN} + \bar{h}^z_{1i-1}\bar{g}^y_{iN}]}{\Delta_{1i-1}\Delta_{iN}}\right). \tag{18}$$

The only function used in Eq. (18) and not defined in Appendix A is G_{ij}; it is given by

$$G_{ij} = \prod_{k=i}^{j} x_k, \quad G_{ii-1} = 1, \quad G_{ii-2} = 0. \tag{19}$$

Equations (17) and (18) are quite complicated, and further derivatives will be even more complicated. Hence, we introduce a matrix notation and write

$$M_{1iN} = \bar{G}_{1i-1}(a)A_{ab}G_{iN}(b). \tag{20}$$

The elements of the 8×8 matrix A_{ab} depend on s, the various external masses λ^2 and $\alpha + 2$. The column (row) matrix $G_{iN}(b)(\bar{G}_{1i-1}(a))$ is a function of the Feynman parameters only. In Appendix B, A_{ab}, $G(b)$, $\bar{G}(a)$ are given explicitly.

Next, keeping in mind the ordering procedure, we differentiate with respect to x_k:

$$\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_i}\left(\frac{e^{-Q_N}}{\Delta_N^{\alpha+2}}\right)_{x_k,x_i=0} = [P_{1i-1}\bar{G}_{1i-1}(a)]A_{ab}\left\{\left[\frac{\partial}{\partial x_k}G_{iN}(b)\right](P_{ik-1})(P_{kN}) + \left(\frac{\partial}{\partial x_k}P_{iN}\right)[G_{iN}(b)_{x_k=0}]\right\}, \tag{21}$$

where $P = [\exp(-Q)]/\Delta^{\alpha+2}$. Referring again to Appendix A, we find that

$$\frac{\partial}{\partial x_k}G_{iN}(b) = (E_{ik-1})_{bc}G_{kN}(c). \tag{22}$$

The matrix E_{ik-1} is given in Appendix B. It has the following properties:

$$[(E_{ik-1})_{bc}]_{x_j=0} = 0, \tag{23a}$$

$$\frac{\partial}{\partial x_j}(E_{ik-1})_{bc} = (E_{ij-1})_{bd}(E_{jk-1})_{dc}. \tag{23b}$$

These properties of E_{ik-1} enable us to immediately carry out all subsequent derivatives with respect to $x_j, \cdots, x_m$ on the first term on the right of (21). The result is

$$[P\bar{G}(a)]_{1i-1}A_{ab}(PE_{bc})_{ij-1}, \cdots, (PE_{de})_{mk-1}[PG(e)]_{kN}. \tag{24}$$

The factorization of (24) allows the corresponding part of $L_N(\alpha, s)$ to be summed quite easily.

The second term in (21) can be written in the form

$$[P\bar{G}(a)]_{1i-1}A_{ab}[F(b)P\bar{F}(c)]_{ik-1}A_{cd}[PG(d)]_{kN}. \tag{25}$$

$F(b)[\bar{F}(c)]$ is a column (row) matrix closely related to $G(b)$ and is defined and discussed in Appendix B following Eq. (B2); the 3, 6, 7 elements are identically zero. All further derivatives of (25) operate only on the $(FP\bar{F})$ factor; $F\bar{F}$ is an 8×8 matrix formed from the outer product of F and $\bar{F}$. Suppressing the matrix subscripts, we find that

$$\frac{\partial}{\partial x_j}(FP\bar{F})_{ik-1} = (E^1_{ik-1}F_{jk-1}\bar{F}_{jk-1} + F_{ij-1}\bar{F}_{ij-1}E^2_{jk-1} + F_{ij-1}\bar{F}_{jk-1}\bar{F}_{ij-1}HF_{jk-1})P_{ij-1}P_{jk-1}, \tag{26}$$

where we have used

$$\frac{\partial}{\partial x_j}F = E^1F, \tag{27a}$$

$$\frac{\partial}{\partial x_j}\bar{F} = \bar{F}E^2, \tag{27b}$$

$$\frac{\partial}{\partial x_j}P = (P\bar{F})H(FP), \tag{27c}$$

and

$$F_{ik-1}(\bar{F}_{ik-1})|_{x_j=0} = F_{ij-1}(\bar{F}_{jk-1}). \tag{28}$$

All of these relations are easily derived from Appendices A and B. The matrix H is just A with all the external masses set equal to zero. The matrices E^1 and E^2 are given in Appendix B. They are closely related to E and satisfy (23a, b). Let us form a two-component supermatrix $\hat{X}$ from 8×8 matrices:

$$\hat{X} = \begin{pmatrix} FP\bar{F} \\ E^2P \end{pmatrix}. \tag{29}$$

Then from (26) and (23a, b) we find that

$$(\partial/\partial X_m)\hat{X}_{ik-1} = \hat{Z}_{im-1}\hat{X}_{mk-1}, \tag{30}$$

where

$$\hat{Z} = \begin{pmatrix} E^1P + FP\bar{F}H & FP\bar{F} \\ 0 & E^2P \end{pmatrix}. \tag{31}$$

All further derivatives are given by

$$\frac{\partial^n}{\partial x^n}\hat{X} = (\hat{Z})^n\hat{X}. \tag{32}$$

Since only the upper component of this two-component supermatrix is of interest, we project it out with

$$\hat{Y} = (1, 0). \tag{33}$$

Equations (24) and (32) represent a factorization of the amplitude similar to that found near $\alpha = -1$.

The scattering amplitude $L_N(\alpha, s)$ is given by an integral over (15). In the expansion of the product, there will occur terms where $x_{i+1}, \cdots, x_{j-1}$ are not trivially integrated to give (16). We replace the functions of the Feynman parameters that appear in (24) and (32) by the integrals over these functions and label them by the same symbol. In other words, we have

$$f_{ij-1} = \int (\sigma_i \sigma_{j-1})^{\alpha+2} \left\{ \prod_{r=i+1}^{j-1} [x_r^{\alpha+2} - (\sigma_r \sigma_{r-1})^{\alpha+2}] \frac{\partial}{\partial x_r^2} \right\} \times \left[\frac{\exp(-\bar{Q}_{ij-1})}{(\bar{\Delta}_{j-1})^{\alpha+2}} f_{ij-1} \right], \tag{34}$$

where $\bar{Q}$ and $\bar{\Delta}$ are the same functions that appeared in (5). The elements of $G(a)$ are given by expressions similar to (34) with $\bar{Q}'$ and $\bar{\Delta}'$. If I_N^m is that portion of $L_N(\alpha, s)$ with m of the x's contracted, we see that

$$\begin{aligned} I_N^m = - \sum_{r_1=1}^{N+2-m} \sum_{r_2=r_1+1}^{N+3-m} \cdots \sum_{r_m=r_{m-1}+1}^{N+1} & (-\bar{G})_{1r_1-1} A \\ & \times [(-E)_{r_1 r_2-1} \cdots (-E)_{r_{m-1} r_m-1} \\ & + \hat{Y}(-\hat{Z})_{r_1 r_2-1} \cdots (-\hat{Z})_{r_{m-2} r_{m-1}-1} \\ & \times (-\hat{X})_{r_{m-1} r_m-1} A] \times (-G)_{r_m N}. \end{aligned} \tag{35}$$

The total scattering amplitude is given by

$$L(\alpha, s) = \sum_{N=0}^{\infty} L_N(\alpha, s) = \frac{\Gamma(-\alpha) g^2}{(\alpha+1)(\alpha+2)} \times \sum_{N=0}^{\infty} \left[\frac{g^2}{16\pi^2(\alpha+1)(\alpha+2)} \right]^N \sum_{m=0}^{N+1} I_m^N. \tag{36}$$

The sum over N is performed in exactly the same way as for the $\alpha = -1$ case.

$$L(\alpha, s) = \frac{\Gamma(-\alpha) g^2}{(\alpha+1)(\alpha+2)} [\Lambda - \bar{G} A (1+E)^{-1} G + \bar{G} A \hat{Y} (1+\hat{Z})^{-1} \hat{X} A G]. \tag{37}$$

The matrices E, $\hat{Z}$, and $\hat{X}$ are obtained by summing (34) over all values of $n = j - i - 1$. For example,

$$E_{ab} = \frac{1}{(\alpha+2)} \sum_{s=1}^{\infty} \left[\frac{g^2}{16\pi^2(\alpha+1)} \right]^s \frac{1}{(\alpha+2)^{s-1}} (E_{ab})_{1s-1}. \tag{38}$$

G and $\bar{G}$ are given by expressions similar to (34) and (38), but involving functions for ladders with only one end contracted. The term $\Lambda/(\alpha+2)$ in (31) is a regular function of α near $\alpha = -2$; since it has no poles, we discard it.

Equation (37) constitutes the complete solution to the problem of summing the ladder diagrams near $\alpha = -2$. The poles in the α plane come from the zeros of the determinants of $(1 + E)$ and $(1 + \hat{Z})$. Ostensibly there are a very large number of poles. However, as we shall see in the next section, there are at most five poles in α. As the final step in this section, we rewrite the third term on the right of (37) in terms of a more useful set of functions.

$$\bar{G} A \hat{Y} (1+\hat{Z})^{-1} \hat{X} A G = \bar{G} A (1 + E^1 + F\bar{F}H)^{-1} F\bar{F} (1+E^2)^{-1} A G. \tag{39}$$

The zeros of the determinants of $(1 + E^2)$ and $(1 + E)$ are just recurrences of the leading Regge pole, while the determinant of $(1 + E^1 + F\bar{F}H)$ leads to at least two new dynamical poles and the pole to be identified as the daughter pole. In the next section we try to unravel our solution and investigate these various poles.

4. DAUGHTER POLES IN THE EQUAL INTERNAL MASS LIMIT

The complete amplitude in the form given by (37) is not very useful; it is too complicated. In order to understand the nature of the solution, we make the simplifying assumption that $\mu = \nu$. Although the resulting amplitude does not satisfy two-particle unitarity, the existence of the daughter pole is independent of unitarity. Even with $\mu = \nu$, the complete amplitude is still quite complicated. Thus, we first look at the amplitude to lowest order. In other words, we assume that $\alpha + 2$ is of order g^2, expand every function appearing in (37) and (39) in powers of g^2, and keep just the leading power. In Appendix C we define a set of 15 functions of s in terms of which the various matrices can be written. In lowest order for $\mu = \nu$, there are just five independent functions of s and μ^2. The only tricky step occurs in the lowest-order evaluation of the integral over the function X defined in (A17). From Table I of Appendix C we have $X_{ii} = -\sigma_i^{-2}$, and

$$X = -\int_0^{\infty} dy\, dz\, \frac{\sigma^{2\alpha+4} e^{-\mu^2 \sigma + (syz/\sigma)}}{\sigma^{\alpha+4}}, \tag{40}$$

where $\sigma = y + z$. Since this integral is multiplied by g^2, the natural procedure would be to set $\alpha = -2$ under the integral, only to discover that the result is divergent. The $(\sigma)^{2(\alpha+2)}$ factor in the numerator of the integrand is necessary to keep the integral convergent for $\alpha > -2$. Since the whole procedure for isolating the singular portion of the amplitude near $\alpha = -2$

involves finding a form valid for $\alpha > -2$ which can be analytically continued to $\alpha < -2$, this factor, or one like it, is necessary. This is the reason that we added and subtracted $(\sigma_i\sigma_{i-1})^{\alpha+2}$ rather than $(1)^{\alpha+2}$ in manipulating Eq. (15). The correct prescription for evaluating (40) is to do the integral for $\alpha > -2$ and analytically continue the result. To lowest order (40) is $-(\alpha+2)^{-1}$; since X is always multiplied by $g^2(\alpha+2)$, this procedure yields a term of order g^2. In the analysis of II, this contribution was overlooked, and an incorrect equation for the position of the poles was obtained. In Appendix D, we consider the effect of this change on the mixing of poles and cuts.

With the matrices and functions of Appendix C it is a straightforward but tedious problem in matrix algebra to calculate $L(\alpha, s)$ to lowest order. The result is

$$L(\alpha,s) = -g^2\left[\frac{-2p^2}{\alpha+2-G^2K(s)} + \frac{(m_1^2-m_2^2)^2/2s}{\alpha+2-G^2/\mu^2} - \frac{N(\alpha,s)}{D(\alpha,s)}\right], \quad (41)$$

where

$$N(\alpha,s) = (\alpha+2)\left(m_1^2+m_2^2-\frac{s}{2}-\lambda^2\right) + \frac{G^2}{\mu^2}\left(m_1^2+m_2^2-\frac{s}{2}\right) - G^2\frac{(m_1^2+m_2^2-2\mu^2)^2}{8q^2}\left[\frac{1}{\mu^2}+K(s)\right],$$

$$D(\alpha,s) = (\alpha+2)^2 - G^2(\alpha+2)\left[K(s)(1+\lambda^2/2q^2) - \frac{1}{\mu^2}\left(1-\frac{\lambda^2}{2q^2}\right)\right] - G^4\frac{K(s)}{\mu^2},$$

and $4p^2 = s - 2(m_1^2+m_2^2) + (m_1^2-m_2^2)^2/s$, $4q^2 = s - 4\mu^2$, and $G^2 = g^2/(16\pi^2)$. The trajectory function of the leading Regge pole is $K(s)$, where $K(0) = 1/\mu^2$. Its explicit form is given by (14) with $\mu = \nu$.

The first term in (41) is the recurrence of the leading Regge pole which arises from the use of Mellin rather than Legendre transforms. The second term has the correct residue and position at $s = 0$ to be the daughter pole. If the leading and daughter Regge poles are related by (1), the corresponding relation for the residues of Mellin poles is

$$\gamma_d(s) = \alpha(0)\frac{(m_1^2-m_2^2)^2}{2s}\gamma_l(0) + \text{N.S.} \quad (42)$$

To lowest order $\alpha(0)$ is replaced in this expansion by -1. The third term in (35) contains two new poles near $\alpha = -2$. These poles are present in the equal mass limit. The significance of these new poles is uncertain. They do mix with the Regge cuts. To lowest order in this model, the daughter pole is a fixed pole. This is the first indication that the daughter pole behaves very differently from the leading pole.

Keeping $\mu = \nu$, we follow the first daughter pole to higher order in the coupling constant. The daughter pole is one of the zeros of the determinant of $(1 + E^1 + F\bar{F}H)$. As shown in Appendix C, this determinant factors, for $\mu = \nu$, into the product of

$$1 + Y(s,\alpha) + s[2g_0(s,\alpha) - \tfrac{1}{2}J(s,\alpha)], \quad (43)$$

and the determinant of a 3×3 matrix. The functions in (43) are defined in Appendix C. In Eq. (12) we saw that the recurrence of the leading pole is given by the solution of $1 + Y(\alpha, s) = 0$. Since g_0 and J are finite at $s = 0$, the pole in $L(\alpha, s)$ defined by setting (43) equal to zero has the correct $s = 0$ limit to be the daughter pole; this result is true to all orders in the coupling constant. To lowest order it agrees with the daughter pole given in (41). Once (43) is identified as the daughter pole, we can calculate its residue to all orders, keeping only those terms in $L(\alpha, s)$ that contain this pole. The result is

$$L_\alpha(\alpha,s) = \frac{g^2\Gamma(-\alpha)}{(\alpha+1)2s} \times \frac{\{(m_2^2-m_1^2)G(3) + s[G(4)-G(5)]\}^2}{(\alpha+2)[1+Y+s(2g_0-\frac{1}{2}J)]}. \quad (44)$$

$G(a)$ refers to the appropriate integrals over the functions introduced in (20) and given explicitly by (B2). We have used the left–right symmetry of ladder diagrams to replace $\bar{G}(a)$ by $G(a)$. An explicit evaluation of the first few terms of $G(3)$ and $Y(s, 0)$ shows that

$$\frac{G(3)^2}{1+(d/d\alpha)[(\alpha+2)Y(\alpha,0)]} = \frac{-\alpha G(\alpha,0)^2}{1-(d/d\alpha)F(\alpha,0)}, \quad (45)$$

where common factors have been discarded. $F(\alpha, s)$ and $G(\alpha, s)$ are given by (5) and (6). Therefore, the daughter pole has the correct residue.

The trajectory function for the daughter pole is given to all orders in the coupling constant by the solution of

$$\alpha+2 = -\sum_{N=1}^{\infty}\left(\frac{g^2}{16\pi^2}\right)^N\frac{1}{(\alpha+1)^N}\,dy_i\,dz_i \times \prod_{i=1}^{N-1}dx_i\left[x_i^{\alpha+2} - (\sigma_i\sigma_{i+1})^{\alpha+2}\frac{\partial^2}{\partial x_i^2}\right] \times (\sigma_1\sigma_N)^{\alpha+2}\frac{e^{-Q_N}}{\bar{\Delta}_N^{\alpha+3}} \times \left[G_{1N-1} + \frac{S}{2\Delta_{1N}}(\bar{g}_{1N}^y - \bar{g}_{1N}^z)(h_{1N}^y - h_{1N}^z)\right]. \quad (46)$$

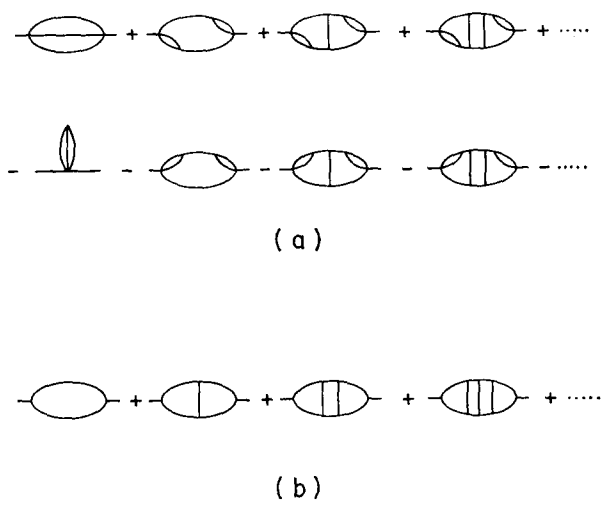


FIG. 2. (a) The diagram sum which describes graphically the analytic structure of the daughter trajectory function. (b) The corresponding diagram sum for the leading trajectory.

This expression can be simplified by using the variables $\sigma_i = y_i + z_i$ and $\delta_i = y_i - z_i$. The part of the integrand of (46) in square brackets has the form necessary to trivially carry out the δ_1 and δ_N integrations[9]; the effect of these integrations is to replace (46) by

$$\alpha+2 = -\frac{g^2}{16\pi^2}\frac{\Gamma(\alpha+3)}{(\alpha+1)(\mu^2)^{\alpha+3}} + \sum_{N=2}^{\infty}\left(\frac{g^2}{16\pi^2}\right)^N \frac{1}{(\alpha+1)^N} \times \frac{1}{s}\int_0^{\infty}\prod_{i=1}^{N=1} dx_i \frac{[x_i^{\alpha+2} - (\sigma_i\sigma_{i+1})^{\alpha+2}]}{\alpha+2}\frac{\partial^2}{\partial x_i^2} \times (\sigma_1\sigma_N)^{\alpha+2}\frac{e^{-\bar{Q}_N}}{\bar{\Delta}_N^{\alpha+2}}\{\delta(z_N)\delta(y_1) - \delta(z_N)\delta(z_1)\}. \tag{47}$$

This expression should be compared with (5) for the leading pole. The daughter pole first begins to move to the right in the complex α plane in order g^4. The slope of the leading pole at $s = 0$ is $+G^2/(6\mu^4)$ and that of the daughter pole is $+0.09G^4/(\mu^6)$. The effect of the delta functions inside the final curly brackets of (47) is to give the trajectory function a singularity structure which is identical to that arising from the sum of diagrams in Fig. 2(a). The corresponding sum for the leading trajectory is given in Fig. 2(b). Because of the minus signs in Fig. 2(a), all two-particle cuts cancel out term-by-term from the daughter trajectory function.[10] We show in the next section that if $\mu \neq \nu$ the two-particle cuts enter with coefficients proportional to $(\mu - \nu)^2$.

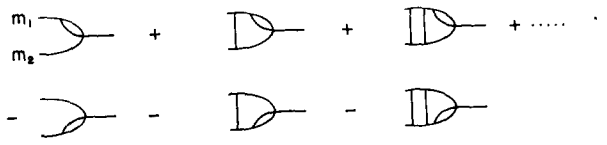


FIG. 3. The diagram sum which describes graphically the analytic structure of the daughter residue function.

The residue of the daughter pole in (44) has the property that it vanishes to all orders if the external masses are equal. The same manipulations used to go from (46) to (47) lead to an expression for the form factors coupling the Regge pole to two-particle states that has only three-particle singularities:

$$(m_2^2 - m_1^2)^2 G(3) + s[G(4) - G(5)] = (m_2^2 - m_1^2) + \sum_{N=1}^{\infty}\left(\frac{g^2}{16\pi^2}\frac{1}{(\alpha+1)}\right)^N \times \int dy_i\, dz_i \prod_{i=1}^{N}\frac{(x_i^{\alpha+2} - (\sigma_i\sigma_{i-1})^{\alpha+2})}{\alpha+2} \times \sigma_N^{\alpha+2}\frac{\partial^{2N}}{\partial x_1^2 \partial x_N^2}\left\{[\delta(y_N) - \delta(z_N)]\frac{e^{-Q_{1N}}}{\Delta_{1N}^{\alpha+2}}\right\}. \tag{48}$$

The corresponding diagram sum is given in Fig. 3. Again the two-particle cuts cancels out.[10]

Equations (44), (47), and (48) completely describe the first daughter pole in our model. The first comment to be made is that the daughter pole is very different from the leading pole. Its motion is determined by the scattering of three (or more) particles. Since the three-particle contributions of ladder diagrams are not those that would be expected to dominate a three-particle scattering amplitude, we conclude that our model is useless for determining the actual motion of the daughter pole. Such a determination would require an accurate treatment of three-particle scattering. This objection applies also to Bethe–Salpeter calculations using the ladder approximation.[4] In the next section we see to what extent relaxing the $\mu = \nu$ constraint changes these conclusions. The second point to be made is that the residue of the daughter pole vanishes to all orders in the coupling constant, for all energies, if the external masses are equal. The general arguments on the existence of daughter poles only show that the singular part of the daughter residue vanishes in the equal mass limit.

With the advantage of hindsight we can argue on general grounds that the daughter pole in this model should contain no two-particle singularities. The leading Regge trajectory function satisfies a dispersion relation in s with the cut beginning at the two-particle

[9] The dependence of the coefficient of s in Q_{1N} can be written as $\delta^t A\delta$, with A an $N \times N$ matrix and δ a column matrix of the δ_i. In terms of this matrix, the coefficient of s in (46) is proportional to $(A_{1k}\delta_k)(\delta_k A_{kN}) = [(\partial/\partial\delta_1)\delta^t A\delta][(\partial/\partial\delta_N)\delta^t A\delta]$. If δ_N is integrated by parts, only the end points contribute. The resulting integral over δ_1 is done exactly.

[10] The analyticity properties of perturbation graphs are discussed in detail in Ref. 5, Chap. 2.

threshold; the imaginary part, up to kinematical factors, is given by the form factor squared $|G(\alpha, s)|^2$. If this relation between the imaginary part of the trajectory function and the square of the form is assumed to be a general property of Regge poles in a unitary theory, then, since the daughter pole does not couple at all to equal mass two-particle states, it will not have a two-particle cut. If the coupling to equal mass particles dominates the trajectory function, three-particle intermediate states would be expected to be very important. Presumably the coupling of a daughter pole to a three-particle amplitude could be described by an expression similar to that involved in the coupling to unequal mass two-particle states, only with a variable mass. This argument on three-particle dominance requires that the nonsingular terms in (1) vanish for equal masses.

5. UNEQUAL INTERNAL MASSES IN LOWEST ORDER

In the previous section we extracted the daughter pole from the complete scattering amplitude and studied its properties to all orders in the coupling constant. This separation was possible only with the simplifying assumption that the masses on the sides of the ladder were equal. We now want to see to what extent the conclusions obtained with this assumption depend upon it. If $m_1 = \mu \neq \nu = m_2$, then the scattering amplitude represented by ladder diagrams will satisfy two-particle unitarity. The problem of separating out the daughter pole to all orders becomes algebraically very formidable. Rather than attempt it, we look at the zeros of the determinant of $(1 + E^1 + F\bar{F}H)$ to lowest order in the coupling constant and investigate the daughter pole only to this order. From the relations given in Appendix C, we see that

$$\det(1 + E^1 + F\bar{F}H) = [\epsilon - K(s)]\left\{\epsilon^3 - \epsilon^2\left[K(s)\left(1 + \frac{\lambda^2}{2q^2}\right) + \Gamma\frac{\lambda^2}{2q^2}\right] + \epsilon\left[\frac{\lambda^2\Gamma K(s)}{2q^2} - \frac{1}{\mu^2\nu^2}\left(1 - \frac{\lambda^2}{2q^2}\right)\right] + \frac{K(s)}{\mu^2\nu^2}\right\} = 0, \tag{49}$$

where $\alpha + 2 = G^2\epsilon$ and $K(s)$ is again the trajectory function of the leading pole. The mass of the exchanged particle is λ, and Γ is given by

$$\Gamma = \frac{(\mu^2 + \nu^2)s - (\mu^2 - \nu^2)^2}{2s\mu^2\nu^2}, \tag{50}$$

and $4q^2 = s - 2(\mu^2 + \nu^2) + (\mu^2 - \nu^2)^2/s$. The first factor in (49) is just the recurrence of the leading Regge pole; we neglect it now. The daughter pole is one of three roots of a cubic equation; the other roots are the two new dynamical poles which appear in (41). If $\mu = \nu$, one of the roots of the cubic equation is μ^{-2}, in agreement with (41), For $\mu \neq \nu$ and $s = 0$, $q^2 = \infty$ and the cubic equation becomes

$$[\epsilon - K(0)]\left(\epsilon^2 + \frac{\lambda^2}{4\mu^2\nu^2}\epsilon - \frac{1}{\mu^2\nu^2}\right) = 0, \tag{51}$$

where

$$K(0) = \frac{1}{\mu^2 - \nu^2}\ln\left(\frac{\mu^2}{\nu^2}\right).$$

Thus, it contains a zero whose position satisfies the condition for a daughter pole. If $s = \infty$, we find that

$$\epsilon\left(\epsilon - \frac{1}{\mu\nu}\right)\left(\epsilon + \frac{1}{\mu\nu}\right) = 0. \tag{52}$$

By comparison with the equal mass limit in (41), we see that the middle root of (52) is the daughter trajectory. Thus the daughter pole certainly moves to lowest order in the coupling constant in contrast to its behavior for $\mu = \nu$. For $s \neq 0, \infty$ there is no simple analytic solution in general. However, if $\lambda^2 = 0$, we find that

$$[\epsilon - K(s)]\left(\epsilon - \frac{1}{\mu\nu}\right)\left(\epsilon + \frac{1}{\mu\nu}\right) = 0. \tag{53}$$

In this limit, the daughter trajectory is parallel to that of the leading pole. Since the three-particle thresholds in (47) become degenerate with the two-particle thresholds, it is not surprising that to lowest order the trajectories should be parallel. On the other hand the transition to $\lambda^2 = 0$ is apparently singular. Moreover, there are problems with infrared divergences in higher orders if $\lambda^2 = 0$.

For λ the same order of magnitude as μ and ν, we can obtain an expansion of the daughter pole around $s = 0$. For both s and $\delta = (\mu - \nu)^2$ small, we find that

$$\epsilon_d = \frac{1}{\mu^2}\left[1 + \frac{1}{12}\frac{\delta}{\mu^2} + \frac{\delta^2}{80\mu^4} + \frac{s\delta}{\mu^4}\left(\frac{1}{45} - \frac{\mu^2}{18\lambda^2}\right) + \cdots\right] \tag{54}$$

while the leading pole is given by

$$\epsilon_l = \frac{1}{2}\left[1 + \frac{1}{\mu^2}\left(\frac{s}{6} + \frac{\delta}{12}\right) + \frac{1}{\mu^4}\left(\frac{s^2}{30} + \frac{s\delta}{60} + \frac{\delta^2}{80}\right) + \cdots\right], \tag{55}$$

where $\alpha_l + 1 = G^2\epsilon_l$. Note the singular dependence on λ^2 in (55). The daughter pole moves more slowly than the leading pole in this limit; hence, three-particle effects would still be expected to dominate. If δ is

large, the daughter pole can also be expanded in the neighborhood of $s = 0$.

$$\epsilon_d = K(0) - \frac{2s}{(\mu^2 - \nu^2)^2}\left[1 - \frac{K(0)}{2}(\mu^2 + \nu^2 + 2\lambda^2) + \frac{K(0)^2\lambda^2}{\mu^2\nu^2K(0)^2 + \lambda^2K(0) - 1}\right]. \quad (56)$$

This equation should be compared with

$$\epsilon_l = K(0) - \frac{2s}{(\mu^2 - \nu^2)^2}\left[1 - \frac{K(0)}{2}(\mu^2 + \nu^2)\right]. \quad (57)$$

If $\lambda^2 = \nu^2 < \mu^2$, we find from (56) and (57) that the daughter pole has a smaller slope than the leading pole at $s = 0$ for $\nu^2/\mu^2 > \frac{1}{8}$. We conjecture that for all values of ν^2/μ^2 the leading pole will eventually move faster than the daughter pole.

Equation (49) is not valid near the two-particle threshold since $K(s)$ becomes infinite there. On the other hand, it is valid for s approaching $\pm\infty$. A solution can be obtained for ϵ_d in powers of s^{-1}; the zero-order solution is given by (52). In this large s limit the daughter pole is identified by extrapolation to equal masses.

$$\epsilon_d = \frac{1}{\mu\nu}\left\{1 + \frac{\lambda^2\delta}{2s\mu\nu}\left[1 + \frac{\delta}{s} + \frac{\lambda^2(\mu+\nu)^2}{4s\mu\nu} + \frac{\delta^2}{s^2} + \frac{\lambda^2(\mu+\nu)^2}{2s^2\mu\nu}(\mu^2+\nu^2) - \frac{\lambda^4(\mu+\nu)^2}{4s^2\mu\nu}\right] + \frac{\lambda^4\delta}{s^3}\left[\ln\frac{s^2}{\mu^2\nu^2} + 2\pi i\theta(s)\right]\right\}. \quad (58)$$

In this expansion there is no restriction on the magnitude of δ. The corresponding expression for the leading trajectory is

$$\epsilon_l = \frac{1}{s}\left[\ln\frac{s^2}{\mu^2\nu^2} + 2\pi i\theta(s)\right]. \quad (59)$$

Notice that the imaginary part of ϵ_d has the same sign as that of ϵ_l, but is smaller by a factor of s^{-2}. Parenthetically we remark that one of the other two dynamical poles is given by (58) with ν replaced by $-\nu$. Since the imaginary part of this pole is negative, there would be difficulties in interpreting it if it becomes physical. The third pole is asymptotically parallel to the leading pole and has an imaginary part proportional to s^{-1}; this third pole may be identified as the normal secondary Regge trajectory, since it behaves very similarly to the leading pole. Inasmuch as the expansion of the daughter trajectory function in powers of s^{-1} converges for s above threshold, (58) may be taken as an indication that the imaginary part of the daughter trajectory is less than that of the leading pole for all s. Hence, our conclusion that the daughter pole is strongly affected by three-particle scattering still holds. Although for unequal internal masses the daughter moves to lowest order in g^2 and has a two-particle cut in contrast to the results of the preceding section, the effect of the two-particle singularities is substantially less important for the daughter pole than for the leading pole. In general, the daughter pole does not move parallel to the leading pole. It moves to the right less rapidly and develops a smaller imaginary part at the two-particle threshold.

ACKNOWLEDGMENT

The author wishes to thank Professor L. Durand for numerous discussions and for reading the manuscript.

APPENDIX A

In this Appendix we give both explicit definitions and various identities for the functions of the Feynman parameters that are used in Sec. 3. The determinant Δ_N is given by

$$\Delta_N = \det\begin{vmatrix} \sum_1 & x_1 & 0 & 0 & \cdots & & \\ x_1 & \sum_2 & x_2 & 0 & \cdots & & \\ 0 & x_2 & \cdot & & & & \\ 0 & 0 & & \cdot & & & \\ \cdot & \cdot & & & \cdot & & \\ \cdot & \cdot & & & & \cdot & x_N \\ \cdot & \cdot & & & & x_N & \sum_N \end{vmatrix}, \quad (A1)$$

where $\sum_i = x_i + y_i + z_i + x_{i+1}$. From (A1) it is readily seen that if $x_i = 0$, Δ_N factors as discussed in I and II. It will be convenient to relabel the determinant for a ladder by its first and next-to-last rungs. In this notation Δ_N becomes Δ_{1N}. If $x_i = 0$, Δ_{1N} factors into the product $(\Delta_{1i-1})(\Delta_{iN})$ with the understanding that $x_i = 0$ in the determinants. In addition it follows from (A1) that

$$(\partial/\partial x_i)\Delta_{iN} = \Delta_{1i-2}\Delta_{iN} + \Delta_{1i-1}\Delta_{i+1N}. \quad (A2)$$

In (A2) and all further equations involving derivatives with respect to x_i, it is understood that x_i is set equal to zero after differentiation. Equation (A2) is valid for all $1 \leq i \leq N+1$, if we set $\Delta_{ii-1} = 1$ and $\Delta_{ii-n} = 0$, $n \geq 2$.

The functions g_N^ν, h_N^ν are treated as a pair.

$$h_N^\nu = \sum_{k=1}^{N}\Delta_{1k-1}y_kG_{k+1N+1},$$
$$g_N^\nu = \sum_{k=1}^{N}G_{1k}y_k\Delta_{k+1N}. \quad (A3)$$

If the superscript is z, z_k replaces y_k in (A3). The function G_{ij} is defined in (19). If $x_i = 0$, these functions factor to give

$$g^v_{1N} = g^v_{1i-1}\Delta_{iN},$$
$$h^v_{1N} = \Delta_{1i-1}h^v_{iN}. \tag{A4}$$

The notation in (A4) follows the rules used in labeling Δ_{ij}. We define another set of functions $\bar{g}^v_{ik}$, $\bar{h}^v_{ik}$ by

$$\bar{h}^v_{ik} = \sum_{j=i}^{k}\Delta_{ij-1}y_jG_{j+1k},$$
$$\bar{g}^v_{ik} = \sum_{j=i}^{k}G_{i+1j}y_j\Delta_{j+1k}. \tag{A5}$$

Equation (A5) differs from (A3) by the external subscript on the function G_{ik}. For $x_j = 0$ these new functions factor according to the rule given in (A4); they also obey the identity $\bar{g}^v_{ik} + \bar{g}^z_{ik} = \Delta_{ik}$.

The derivatives of g^v_{1N} and h^v_{1N} are given by

$$\frac{\partial}{\partial x_k}\left(\frac{h^v_{1N}}{\Delta_{1N}}\right) = \frac{h^v_{1k-1}G_{k+1N+j}}{\Delta_{1k-1}\Delta_{kN}} + H^v_{kN},$$
$$\frac{\partial}{\partial x_k}\left(\frac{g^v_{1N}}{\Delta_{1N}}\right) = \frac{G_{1k-1}g^v_{kN}}{\Delta_{1k-1}\Delta_{kN}} + K^v_{1k-1}, \tag{A6}$$

where

$$H_{kN} = \frac{h_{k+1N}\Delta_{kN} - \Delta_{k+1N}h_{kN}}{\Delta^2_{kN}},$$
$$K_{1i-1} = \frac{g_{1i-2}\Delta_{1i-1} - \Delta_{1i-2}g_{1i-1}}{\Delta^2_{1i-1}}. \tag{A7}$$

The corresponding derivative of $\bar{g}(\bar{h})$ is given by (A6) and (A7) with $\bar{K}(\bar{H})$ replacing $K(H)$ in (A6) and $\bar{g}(\bar{h})$ replacing $g(h)$ in (A7). The functions H, K, $\bar{H}$, and $\bar{K}$ have the property that they vanish if an internal x_j is zero. Moreover, their derivatives are given by

$$(\partial/\partial x_j)H_{kN} = \bar{H}_{kj-1}G_{j+1\,N+1},$$
$$(\partial/\partial x_j)K_{1k-1} = G_{1j-1}\bar{K}_{jk-1}. \tag{A8}$$

The derivatives of $\bar{H}$ and $\bar{K}$ are also given by (A8); in addition, $\bar{K}^v_{ij} + \bar{K}^z_{ij} = -G_{ij}/\Delta_{ij}$.

Another sequence of functions to be studied in this Appendix is obtained from $f_N = f_{1N}$.

$$f_{1N} = \sum_{k=1}^{N}\Delta_{1k-1}y_kz_k\Delta_{k+1N} + \sum_{k=1}^{N-1}\sum_{j=k+1}^{N}\Delta_{1k+1}(y_kz_j + z_ky_j)G_{k+1j}\Delta_{j+1N}. \tag{A9}$$

If $x_i = 0$, we have

$$f_{1N} = \Delta_{1i-1}f_{iN} + f_{1i-1}\Delta_{iN}. \tag{A10}$$

The derivative if f_{1N} with respect to x_i is given by

$$\frac{\partial}{\partial x_i}f_{1N} = F^2_{1i-1} + F^1_{iN} + \frac{(\bar{h}^v_{1i-1}\bar{g}^z_{iN} + \bar{h}^z_{1i-1}\bar{g}^v_{iN})}{\Delta_{1i-1}\Delta_{iN}}, \tag{A11}$$

where

$$F^1_{iN} = \frac{f_{i+1N}\Delta_{iN} - f_{iN}\Delta_{i+1N}}{\Delta^2_{iN}},$$
$$F^2_{1i-1} = \frac{f_{1i-2}\Delta_{1i-1} - f_{1i-1}\Delta_{1i-2}}{\Delta^2_{1i-1}}.$$

If $x_k = 0$, we see that

$$F^1_{iN} = F^1_{ik-1},$$
$$F^2_{1i-1} = F^2_{ki-1}. \tag{A12}$$

In addition we need the relation

$$\frac{\partial}{\partial x_k}F^1_{iN} = F_{ik-1} + \bar{H}^v_{ik-1}\frac{\bar{g}^z_{iN}}{\Delta_{iN}} + \bar{H}^z_{ik-1}\frac{\bar{g}^v_{iN}}{\Delta_{iN}}, \tag{A13}$$

where

$$F_{ik-1} = \frac{f_{i+1k-2}\Delta_{ik-1} - f_{ik-2}\Delta_{i+1k-2}}{\Delta^2_{ik-1}} + \frac{f_{i+1k-1}\Delta_{ik-2} - f_{ik-1}\Delta_{i+1k-2}}{\Delta^2_{ik-1}} - 2\frac{\Delta_{ik-2}}{\Delta_{ik-1}}\left(\frac{f_{i+1k-1}\Delta_{ik-1} - f_{ik-1}\Delta_{i+1k-1}}{\Delta^2_{ik-1}}\right). \tag{A14}$$

If $x_j = 0$, $F_{ik-1} = 0$. One more relation involving F_{ik-1} is needed:

$$(\partial/\partial x_j)F_{ik-1} = \bar{H}^v_{ij-1}\bar{K}^z_{jk-1} + \bar{H}^z_{ij-1}\bar{K}^v_{jk-1}. \tag{A15}$$

The final functions that need be considered are the pair $(\Delta_{1i-2}/\Delta_{1i-1})$ and $(\Delta_{i+1N}/\Delta_{iN})$. The $x_k = 0$ properties of these are obvious. Their derivatives are given by

$$\frac{\partial}{\partial x_k}\left(\frac{\Delta_{1i-2}}{\Delta_{1i-1}}\right) = X_{ki-1},$$
$$\frac{\partial}{\partial x_k}\left(\frac{\Delta_{i+1N}}{\Delta_{iN}}\right) = X_{ik-1}, \tag{A16}$$

where

$$X_{ik-1} = \frac{\Delta_{i+1k-2}\Delta_{ik-1} - \Delta_{i+1k-2}\Delta_{ik-2}}{(\Delta_{ik-1})^2}. \tag{A17}$$

Both X_{ik-1} and its derivative with respect to x_j vanish if $x_j = 0$.

APPENDIX B

In this Appendix the various matrices used in Secs. 3 and 4 are given explicitly.

$$A_{ab}=\begin{pmatrix} -\lambda_2 & -(\alpha+2) & 0 & 0 & 0 & m_1^2 & m_1^2 & s \\ -(\alpha+2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_1^2 & m_1^2 & 0 & 0 & 0 \\ 0 & 0 & m_1^2 & 0 & s & 0 & 0 & 0 \\ 0 & 0 & m_2^2 & s & 0 & 0 & 0 & 0 \\ m_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ m_2^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{B1}$$

The matrix H is obtained from A by setting $m_1^2 = m_2^2 = 0$. The two row matrices $\bar{G}(a)$ and $G(a)^t$ are

$$\bar{G}(a)_{1i-1}=\left(1,\frac{\Delta_{1i-2}}{\Delta_{1i-1}},\frac{G_{1i-1}}{\Delta_{1i-1}},\frac{\bar{h}^y_{1i-1}}{\Delta_{1i-1}},\frac{\bar{h}^z_{1i-1}}{\Delta_{1i-1}},K^y_{1i-1},K^z_{1i-1},F^2_{1i-j}\right),$$

$$G^t(a)_{iN}=\left(1,\frac{\Delta_{i+1N}}{\Delta_{iN}},\frac{G_{i+1N+1}}{\Delta_{iN}},\frac{\bar{g}^y_{iN}}{\Delta_{iN}},\frac{h^z_{iN}}{\Delta_{iN}},H^y_{iN},H^z_{iN},F^1_{iN}\right). \tag{B2}$$

If $x_j = 0$, $i \leq j \leq N+1$, we find from the formulas of Appendix A that $G_{iN}(a) = F_{ij-1}(a)$, where $F_{ij-1}(a)$ is obtained from $G_{iN}(a)$ by replacing N by $j-1$ in (B2) and setting $F(3) = F(6) = F(7) = 0$. There is a similar relationship between $\bar{G}_{1i-1}(a)$ and $\bar{F}_{ji-1}(a)$ with j replacing 1 and the 3, 6, 7, elements zero.

The matrix E_{ab} is given by

$$(E_{ab})_{ik-1}=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ X_{ik-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{G_{i+1k-1}}{\Delta_{ik-1}} & 0 & 0 & 0 & 0 & 0 \\ \bar{K}_{ik-1} & 0 & 0 & \frac{G_{i+1k-1}}{\Delta_{ik-1}} & 0 & 0 & 0 & 0 \\ K_{ik-1} & 0 & 0 & 0 & \frac{G_{i+1k-1}}{\Delta_{ik-1}} & 0 & 0 & 0 \\ 0 & 0 & \bar{H}^y_{ik-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{H}^z_{ik-1} & 0 & 0 & 0 & 0 & 0 \\ F_{ik-1} & 0 & 0 & \bar{H}^z_{ik-1} & \bar{H}^y_{ik-1} & 0 & 0 & 0 \end{pmatrix}. \tag{B3}$$

The matrix E^1 is obtained from E by setting all elements of column three equal to zero; E^2 is obtained from E^1 by transposition and interchanging $\bar{K}$ and $\bar{H}$.

TABLE I. The fifteen functions of s used in calculating the scattering amplitude are listed together with the corresponding functions of the Feyman parameters. The correspondence is defined by Eq. (C1). Also given are the functions of the Feyman parameters used in the calculations of the lowest order amplitude.

$T\{f_{ik-1}\}$	f_{ik-1}	f_{ii}
J	1	1
Y	G_{i+1k-1}/Δ_{ik-1}	1^{-1}
Δ	$\Delta_{ik-2}/\Delta_{ik-1}$, $\Delta_{i+1k-1}/\Delta_{ik-1}$	σ^{-1}
g^y	$\bar{g}_{ik-1}/\Delta_{ik-1}$, h_{ik-1}/Δ_{ik-1}	y/σ
K^y	$\bar{K}^y_{ik-1}$, $\bar{H}^y_{hk-1}$	$-y/\sigma^2$
X	X_{i-k1}	$-1/\sigma^2$
F	F_{ik-1}	$2yz/\sigma^3$
F^1	F^1_{ik-1}, F^2_{ik-1}	$-yz/\sigma^2$
$\bar{\Delta}$	$(\Delta_{i+1k-1}\Delta_{ik-2})/(\Delta_{ik-1})^2$	$1/\sigma^2$
$g_\Delta{}^y$	$(\bar{g}_{ik-1}\Delta_{ik-2})/(\Delta_{ik-1})^2$, $(\bar{h}^y_{ik-1}\Delta_{i+1k-1})/(\Delta_{ik-1})^2$	y/σ^2
$F_\Delta{}^1$	$(F^1_{ik-1}\Delta_{ik-2})/(\Delta_{ik-1})$, $(F^2_{ik-1}\Delta_{i+1k-1})/(\Delta_{ik-1})$	$-yz/\sigma^3$
g_e^y	$(\bar{g}^y_{ik-1}\bar{h}^y_{ik-1})/(\Delta_{ik-1})^2$	y^2/σ^2
g_0	$(\bar{g}^y_{ik-1}\bar{h}^z_{ik-1})/(\Delta_{ik-1})^2$, $(y \leftrightarrow z)$	yz/σ^2
f^y	$(\bar{g}^y F^2)_{ik-1}/(\Delta_{ik-1})$, $(\bar{h}^y F^1)_{ik-1}/(\Delta_{ik-1})$	$-y^2z/\sigma^3$
$\bar{F}$	$(F^1_{ik-1})(F^2_{ik-1})$	y^2z^2/σ^4

APPENDIX C

The matrices that occur in the complete solution for $L(\alpha, s)$ near $\alpha = -2$ contain functions of s which are defined by (34) and (38) as integrals over combinations of the functions discussed in Appendix A. Let us rewrite the operation described by (34) and (38) in symbolic form

$$E_{ab} = T[(E_{ab})_{ik-1}]. \tag{C1}$$

There will be another operation for the functions involved in $G(a)$. Each of the Feynman functions in Appendix A will generate a corresponding function of s by (C1). In Table I we list the 15 ostensibly independent functions that can occur. The left–right symmetry of ladder diagrams has been used to reduce the number. There are probably several general identities relating these functions, but we have not investigated this question. In cases where the functions are labeled by a y superscript, there is also a corresponding z superscript which we do not include. When $\mu = \nu$, y and z functions are equal; this is the simplification that enables the daughter pole to be factored out so easily. The product functions, such as g^y_Δ, occur in the matrix $(F\bar{F})$. Using Table I we generate the matrix E from (B3) by direct substitution. We will not write E and the other matrices explicitly. The Regge poles occur as zeros of the determinants

arising from the various inverse matrices in (31) and (33). These determinants are

$$\det(1 + E) = (1 + Y)^3, \tag{C2}$$

$$\det(1 + E^2) = (1 + Y)^2, \tag{C3}$$

$$\det(1 + E^1 + F\bar{F}H) = \tfrac{1}{2}(1 + Y)\det\begin{vmatrix} 1 - (\alpha+2)\Delta + sF^1 & \tfrac{1}{2}s(g^y - g^z) & sJ \\ (K^z + sf^z - (\alpha+2)g^z_\Delta) - (z \to y) & 1 + Y + s(2g_0 - \tfrac{1}{2}J) & s(g^z - g^y) \\ \left[\dfrac{\lambda^2}{s} - \dfrac{(\alpha+2)}{s}X + F + (\alpha+2)^2/s\bar{\Delta} + s\bar{F} - 2(\alpha+2)F^1_\Delta - \tfrac{1}{2}(1+Y)\right] & -\tfrac{1}{2}[K^z + sf^z - (\alpha+2)g^z_\Delta] + \tfrac{1}{2}(z \to y) & 1 - (\alpha+2) + sF^1 \end{vmatrix}.$$

If $\mu = \nu$, (C4) can be written in the form

$$\det(1 + E^1 + F\bar{F}H) = [1 + Y + s(2g_0 - \tfrac{1}{2}J)]D(1 + Y), \tag{C5}$$

where D is the determinant of a 2×2 matrix. As discussed in Sec. 4, the solution of

$$[1 + Y + s(2g_0 - \tfrac{1}{2}J)] = 0$$

is the daughter trajectory to all orders in the coupling constant. Since $(1 + Y) = 0$ is the recurrence of the leading pole, only (C4) will contain new poles. Given the factorization (C5) it is straightforward to invert $(1 + E^1 + F\bar{F}H)$ and keep only those elements that contain the daughter pole.

In Table I we also list the values of the Feynman functions for $k = i + 1$. These are the lowest-order values and are used in the manipulations leading to (35). If $(\alpha + 2)$ is set equal to zero under the integral, in lowest order the operation $T\{f\}$ becomes

$$T\{f\} = \int_0^\infty dy\, dz \exp[-\mu^2 y - \nu^2 z + (syz/\sigma)]f. \tag{C6}$$

This procedure does not apply to X and $\bar{\Delta}$ which are singular in this limit. These are discussed separately. With the change of variables $\sigma = y + z$, $\sigma\epsilon = y - z$, Eq. (C6) can be written as a sum of terms of the form

$$J^n_m = \frac{1}{2}\int_{-1}^{1} d\epsilon \frac{\epsilon^n}{[M]^m}, \tag{C7}$$

where $M = \tfrac{1}{2}[\mu^2 + \nu^2 + (\mu^2 - \nu^2)\epsilon - s/2(1 - \epsilon^2)]$. To express the functions in Table I we need $m = 1$, $n = 0, 1, 2$, and $m = 2, n = 0, 1, 2, 3, 4$. If $\mu = \nu$, integrals with odd values of n vanish; the other integrals are the five functions introduced in II and denoted by $A, \cdots, E$. The trajectory function of the leading pole is given by $J^0_1 = K(s)$. All other functions can be expressed in terms of this one, plus various rational functions of the masses and s. If $\Delta = \tfrac{1}{2}(\mu^2 - \nu^2)$ and $\sum = \tfrac{1}{2}(\mu^2 + \nu^2)$, the identities necessary to obtain (41) and (49) are

$$I^0_2 = \frac{2\Delta - \Sigma s}{2p^2 s(\Sigma^2 - \Delta^2)} - \frac{1}{2p^2}K(s), \tag{C8}$$

$$I^1_2 = \frac{\Delta}{p^2 s}\left[K(s) - \frac{2\Sigma - s}{2(\Sigma^2 - \Delta^2)}\right], \tag{C9}$$

$$I^2_2 = \frac{4K(s)}{s} - \frac{4\Delta}{s}I^1_2 - \frac{(4\Sigma - s)}{s}I^0_2, \tag{C10}$$

$$I^3_2 - \frac{4}{s}I^1_1 = -\frac{(4\Sigma - s)}{s}I^1_2 - \frac{16\Delta}{s^2}K(s) + \frac{16\Delta^2}{s^2}I^1_2 + \frac{4\Delta(4\Sigma - s)}{s^2}I^0_2, \tag{C11}$$

$$-\tfrac{1}{2}I^2_1 + \frac{s}{16}I^4_2 = -\frac{1}{s} + \frac{1}{s}[\Delta^2 I^2_2 + 2\Delta(\Sigma - s/4)I^1_2 + (\Sigma - s/4)^2 I^0_2]. \tag{C12}$$

These relations, together with the expressions for lowest-order functions given in Table I, plus lots of algebra, enable us to obtain the results quoted in Secs. 4 and 5. In these calculations we use the fact that $\alpha + 2$ is of order g^2 to discard terms like $(\alpha + 2)g^y_\Delta$ in (C4).

The correct evaluation of X (and $\bar{\Delta}$) was discussed in Sec. 4. In terms of the notation of (C7), X is just

$$X = -\Gamma(\alpha + 2)I^0_{\alpha+2} \approx -(\alpha + 2)^{-1} = -\bar{\Delta}, \tag{C13}$$

Thus, in (C4) we retain $(\alpha + 2)X$, but not $(\alpha + 2)^2\bar{\Delta}$.

The derivation of (41) requires the lowest-order expressions for $G(a)$. An analysis of the simple box diagram shows that the proper choice is $G(1) = G(3) = 1$ and all other $G(a) = 0$. In Appendix D we

need the $G(a)$ appropriate to a ladder terminated at one end by a contracted ring. The correct prescription is again found by referring to a low-order diagram. In this case we find that, up to a coupling constant,

$$\begin{aligned} G(1) &= J = I_2^0, \\ G(2) &= \Delta = K(s), \\ G(4) &= g^y = \tfrac{1}{2}[I_2^0 + I_2^1], \\ G(5) &= g^z = \tfrac{1}{2}[I_2^0 - I_2^1], \\ G(8) &= F^1 = -\tfrac{1}{4}(I_2^0 - I_2^2). \end{aligned} \tag{C14}$$

The other $G(a)$ are zero. The masses in the integrals in (C14) are m_1 and m_2. If $m_1 = m_2$, $G(4) = G(5)$ and the daughter pole does not mix with the cut.

APPENDIX D: MIXING OF POLES AND CUTS

The previous work on the mixing of Regge pole and cuts contained a minor error. The nature of this error was discussed in Sec. 4. In this Appendix we repeat the calculation correcting the earlier error and including the daughter pole. The diagram we consider is given by Fig. 3 of II. As discussed there, it generates a Regge cut. The only change we make is to let the external masses be different; in addition, we take the masses to be different in the lines coupling the external ladders to the cut generating diagram. (This means that in Fig. 3 of II the lines denoted by α_1, δ_1 carry mass m_1 and the lines denoted by α_3, δ_3 carry m_2.) All other internal masses are equal to μ. As shown in (45) of II, extracting the most singular part of the cut-generating diagram near $\alpha = -2$ leads to a diagram with the external ladders having one end terminated by a contracted rung. The sum of these ladders is obtained from (37) by an appropriate choice of $G(a)$ or $\bar{G}(a)$. This choice of $G(a)$ is discussed in Appendix C. We consider only the lowest-order solution. $G(3)$, $G(6)$, and $G(7)$ are zero identically, and the other $G(a)$ are expressible in terms of the set of basic functions discussed previously. A lengthy problem in matrix algebra, similar to that leading to (41), results in the following factor for an external ladder:

$$\frac{\alpha + 2 + G^2/\mu^2}{D(\alpha_1 s)} - \frac{(m_1^2 - m_2^2)^2}{4p^2 s} \times \frac{\hat{K}(s) - (m_1^2 + m_2^2 - s)/(2m_1^2 m_2^2)}{\alpha + 2 - (G^2/\mu^2)} - 1. \tag{D1}$$

This equation replaces (46) of II. In the first term of (D1), $D(\alpha, s)$ is given in (41); and for simplicity we have set all masses in its residue equal to μ^2, since this term is present even in the equal mass limit. In the second term of (D1), $\hat{K}(s)$ is just $K(s)$ with the masses m_1 and m_2 replacing μ and ν. We see that the daughter pole mixes with the Regge cut. However, since $4p^2 s$ is finite at $s = 0$, the residue of the daughter pole is nonsingular. When (D1) is combined, the cut contribution given by (47) of II and terms with one or no external ladders are added, the total amplitude becomes

$$L(\alpha, s) = \frac{ig^4}{4\pi}\left[\frac{\alpha + 2 + G^2/\mu^2}{D(\alpha, s)} - \frac{N_D(s)}{\alpha + 2 - G^2/\mu^2}\right]^2 \times \int_{\lambda \le 0} \frac{\rho(s, s_1, s_2)\, ds_1\, ds_2}{\alpha + 2 - G^2 K(s_1)}, \tag{D2}$$

where

$$\rho(s, s_1, s_2) = \frac{G^4[K(s_1)]^2}{[\lambda(s, s_1, s_2)]^{\frac{1}{2}}(s_2 - \mu^2)},$$

$$\lambda = s^2 + s_1^2 + s_2^2 - 2ss_1 - 2ss_2 - 2s_1 s_2,$$

and $N_d(s)$ is the residue of the daughter pole in (D1). The conclusion reached in II that the leading pole does not mix with the cut is correct, even with the mass differences included. However, the daughter pole mixes with the cut, as do the other two dynamical poles near $\alpha = -2$. There is no singularity at $s = 0$ associated with the daughter pole when it mixes with the cut. The leading singularity in the α plane remains a pure pole.[11]

[11] A conjecture based on analytic continuation of multiparticle unitarity suggests that the leading Regge pole should appear in the curly brackets of (D2). V. N. Gribov, I. Ya. Pomeranchuk, and K. A. Ter-Martirosyan, Phys. Rev. 139, 184 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Selection Rules and the Decomposition of the Kronecker Square of Irreducible Representations*

P. R. SMITH AND B. G. WYBOURNE
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 20 May 1967)

The irreducible representations occurring in the decomposition of the Kronecker squares of irreducible representations of finite and continuous groups are shown to be readily separable into symmetric and antisymmetric parts using Littlewood's method of plethysm. Particular applications of the rotation and symplectic groups, together with selection rules for isoscalar factors, are given.

I. INTRODUCTION

THE Kronecker square $\Gamma \times \Gamma$ of an irreducible representation Γ of a group G is always reducible to the sum of a symmetric product representation $[\Gamma^2]$ and an antisymmetric product representation $\{\Gamma^2\}$, such that[1,2]

$$\Gamma \times \Gamma = [\Gamma^2] + \{\Gamma^2\}. \tag{1}$$

The resolution of Kronecker squares into symmetric and antisymmetric product representations is of particular significance in determining selection rules over and above those normally found by the decomposition of the appropriate triple Kronecker products.[1-3]

Judd and Wadzinski[3] have recently discussed the resolution of the Kronecker squares of the irreducible representations of the continuous groups R_7 and G_2. Their method, which basically makes use of a chain calculation starting with the trivial resolution of a few simple representations, rapidly becomes excessively tedious. The procedure of resolving Kronecker squares may be greatly simplified by using the more direct method of Littlewood's[4-9] operation of plethysm. With seemingly only two exceptions,[10,11] the powerful technique of plethysm has been unrecognized by physicists. In the present paper, we briefly review Littlewood's method and then consider its application to our central problem of resolving the Kronecker squares of irreducible representations into their symmetric and antisymmetric parts for both finite and continuous groups.

II. PLETHYSM AND THE GENERAL LINEAR GROUP $GL(n)$

The construction of basis functions $\phi_\alpha^{(\Gamma)}$ that transform according to an irreducible representation Γ of a group G is a common problem in group theory. If the irreducible representation Γ is of degree n, then the basis function $\phi_\alpha^{(\Gamma)}$ also spans the representation $\{1\}$ of the general linear group $GL(n)$. The products $[\phi_\alpha^{(\Gamma)}]^r$ of the basis functions $\phi_\alpha^{(\Gamma)}$ form bases for the different irreducible representations $\{\lambda\}$ of $GL(n)$ contained in the Kronecker product

$$\{1\}^r = \sum g_\lambda \{\lambda\}, \tag{2}$$

where the λ's are partitions of the integer r, and g_λ is the number of times a given irreducible representation $\{\lambda\}$ occurs in the decomposition of the product. The irreducible representations $\{\lambda\}$ of $GL(n)$ generally will be reducible under restriction to the group G.

The product functions constructed from the $\phi_\alpha^{(\Gamma)}$'s that form a basis for the $\{\lambda\}$ representation of $GL(n)$ have the symmetry associated with the corresponding representation (λ) of the symmetric group of order $r!$. In practice we are usually interested in constructing product functions of a particular symmetry type, and hence, in picking out those product functions forming a basis for the representation $\{\nu\}$ of $GL(n)$ which turns up in the sum $\sum g_\lambda\{\lambda\}$ of Eq. (2). The decomposition of the representation $\{\lambda\}$ of $GL(n)$, on restriction to G, can then be studied.

Let us suppose that Γ corresponds to the irreducible representation $\{\mu\}$ of $GL(m)$ and that the functions $\phi_\alpha^{\{\mu\}}$ form a basis for this representation. If the dimension of $\{\mu\}$ is n ($m \leq n$ for $\{\mu\} \neq \{0\}$), then this set of functions also forms a basis for the representation $\{1\}$ of $GL(n)$. We now construct powers and

* Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant No. 1275-67.

[1] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962).

[2] J. S. Griffith, *The Theory of Transition—Metal Ions* (Cambridge University Press, New York, 1961).

[3] B. R. Judd and H. T. Wadzinski, J. Math. Phys. **8**, 2125 (1967).

[4] D. E. Littlewood, J. Lond. Math. Soc. **11**, 49 (1936).

[5] D. E. Littlewood, Phil. Trans. Roy. Soc. London **A239**, 305 (1944).

[6] D. E. Littlewood, Phil. Trans. Roy. Soc. London **A239**, 387 (1944).

[7] D. E. Littlewood, *Theory of Group Characters and Matrix Representations of Groups* (Oxford University Press, New York 1958), 2nd ed.

[8] D. E. Littlewood, *A University Algebra* (William Heinemann Ltd., London, 1950).

[9] D. E. Littlewood, *The Skeleton Key of Mathematics* (Harper and Row, New York, 1960).

[10] M. Kretzschmar, Z. Physik **158**, 284 (1960).

[11] J. P. Elliott, Proc. Roy. Soc. (London) **A245**, 128 (1958).

products of the functions $\phi_\alpha^{\{\mu\}}$ of degree r, and from these product functions we choose a basis for the representation $\{\nu\}$ of $GL(n)$, where (ν) is a partition of r. In general, these product functions form a basis for a reducible representation of $GL(m)$. If we denote the set of functions forming a basis for the $\{\lambda\}$ representation of $GL(m)$ by $\phi^{\{\lambda\}}$, and if we denote the set of powers and products of the functions $\phi^{\{\mu\}}$ forming a basis for the $\{\nu\}$ representation of $GL(n)$ by $|\phi^{\{\mu\}}|^{\{\nu\}}$, then clearly

$$|\phi^{\{\mu\}}|^{\{\nu\}} = \sum \phi^{\{\lambda\}}. \tag{3}$$

The above result can be expressed equivalently in terms of transformation matrices. The functions $\phi^{\{\mu\}}$ may be expressed as linear combinations of powers and products of functions $\phi^{\{1\}}$ that form a basis for the $\{1\}$ representation of $GL(m)$. The matrix $A^{\{\mu\}}$, which transforms the $\phi^{\{\mu\}}$'s, is then said to be an induced matrix of the matrix A which transforms the functions $\phi^{\{1\}}$. The operation $|\phi^{\{\mu\}}|^{\{\nu\}}$, which we perform on the functions $\phi^{\{\mu\}}$, is equivalent to forming the induced matrix $|A^{\{\mu\}}|^{\{\nu\}}$ of the induced matrix $A^{\{\mu\}}$. An induced matrix of an induced matrix is, in general, reducible to the direct sum of other induced matrices $A^{\{\lambda\}}$, each of which is the transformation matrix appropriate to the irreducible representation $\{\lambda\}$ with basis functions $\phi^{\{\lambda\}}$. Equation (3) is thus equivalent to the equation

$$|A^{\{\mu\}}|^{\{\nu\}} = \dot{\sum} A^{\{\lambda\}} \tag{4}$$

written in terms of the transformation matrices with $\dot{\sum}$ denoting a direct sum.

In Littlewood's development of the character theory of continuous groups, the symbol $\{\lambda\}$ is used to represent an S function which is defined as the spur[9] of the induced matrix $A^{\{\lambda\}}$. The operations physicists usually associate with representations of continuous groups reflect the properties of the S functions corresponding to the various matrices transforming the elements of the representations. For example, the decomposition of a product representation $\{\lambda\} \times \{\mu\}$ into its simple components is equivalent to the expression of the product of two S functions as the sum of S functions. Equation (4) may be taken as defining a particular type of multiplication of S functions.[4] Taking the spurs of the matrices, we have the definitive equation for the plethysm of S functions:

$$\{\mu\} \otimes \{\nu\} = \sum \{\eta\}, \tag{5}$$

where the symbol $\otimes$ is used to indicate the operation of plethysm[12] and $\{\mu\} \otimes \{\nu\}$ is read as "$\{\mu\}$ plethys $\{\nu\}$."

[12] The symbol $\otimes$ is frequently used to designate the Kronecker outer product; here we reserve it solely for the operation of plethysm. In general, we shall follow Littlewood's notation (Ref. 7) throughout.

The basic algebra of plethysm has been developed by Littlewood, and here we simply state the principal results. Plethysm is distributive on the right with respect to both multiplication and addition, i.e.,

$$A \otimes (B + C) = A \otimes B + A \otimes C, \tag{6}$$

and

$$A \otimes (BC) = (A \otimes B)(A \otimes C) = A \otimes BA \otimes C. \tag{7}$$

For addition, subtraction, and multiplication to the left we have:

$$(A + B) \otimes \{\lambda\} = \sum \Gamma_{\nu\mu\lambda}(A \otimes \{\mu\})(B \otimes \{\nu\}), \tag{8}$$

where $\Gamma_{\mu\nu\lambda}$ is the coefficient of $\{\lambda\}$ in $\{\mu\}\{\nu\}$;

$$(A - B) \otimes \{\lambda\} = \sum (-1)^r \Gamma_{\mu\nu\lambda}(A \otimes \{\mu\})(B \otimes \{\tilde{\nu}\}), \tag{9}$$

where $\{\tilde{\nu}\}$ is the partition of r conjugate to $\{\nu\}$;

$$(AB) \otimes \{\lambda\} = \sum g_{\mu\nu\lambda}(A \otimes \{\mu\})(B \otimes \{\nu\}), \tag{10}$$

where $g_{\mu\nu\lambda}$ is the coefficient of the character $\chi^{(\lambda)}$ of the symmetric group on n symbols, where $\chi^{(\mu)}\chi^{(\nu)} = \sum g_{\mu\nu\lambda}\chi^{(\lambda)}$, and (μ), (ν), and (λ) are all partitions of n; and

$$(A \otimes B) \otimes C = A \otimes (B \otimes C). \tag{11}$$

These formulas may all be readily extended by repeated application of the basic formulas. For example, we readily obtain

$$\begin{aligned}(A + B - C) \otimes \{\lambda\} &= \sum (-1)^r \Gamma_{\mu\nu\lambda}((A + B) \otimes \{\mu\})(C \otimes \{\tilde{\nu}\}), \\ &= \sum (-1)^r \Gamma_{\mu\nu\lambda}\Gamma_{\eta\tau\mu}(A \otimes \{\eta\})(B \otimes \{\tau\})(C \otimes \{\tilde{\nu}\}),\end{aligned} \tag{12}$$

where (ν) is a partition of r. It is also useful to note that, if (λ) is a partition of r and

$$\{\lambda\} \otimes \{\mu\} = \sum \{\nu\},$$

then, if r is *even*,

$$\{\tilde{\lambda}\} \otimes \{\mu\} = \sum \{\tilde{\nu}\}, \tag{13}$$

and, if r is *odd*,

$$\{\tilde{\lambda}\} \otimes \{\tilde{\mu}\} = \sum \{\tilde{\nu}\}, \tag{14}$$

where the sign $\tilde{}$ denotes that the conjugate partition is taken.

The following theorem is particularly useful in the construction of a plethysm: If

$$\{\lambda\} \otimes \{\pi\} = \sum \{\nu\}, \tag{15a}$$

then

$$\sum_{\zeta,\nu} \Gamma_{1\zeta\nu}\{\zeta\} = \left[\sum_{\mu} \Gamma_{1\gamma\pi}\{\lambda\} \otimes \{\gamma\}\right]\left[\sum_{\mu} \Gamma_{1\mu\lambda}\{\mu\}\right]. \tag{15b}$$

If $\{\pi\} \equiv \{n\}$ has only one part, then Eq. (15b) reduces to

$$\sum_{\zeta,\nu} \Gamma_{1\zeta\nu}\{\zeta\} = \{\lambda\} \otimes \{n-1\}\left[\sum_{\mu} \Gamma_{1\mu\lambda}\{\mu\}\right]. \tag{15c}$$

This last equation forms the basis of Littlewood's "third method." [5] The right-hand sides of Eqs. (15b) and (15c) are usually readily expandable; one is left only with the task of selecting a suitable set of $\{\nu\}$'s to match the left-hand side.

A theorem due to Ibrahim[13] is particularly useful in defining the choice of $\{\nu\}$ in Littlewood's third method. The *principal part* of a product of two S functions $\{\alpha\}$ and $\{\beta\}$ is defined as $\{\alpha_1+\beta_1, \alpha_2+\beta_2, \alpha_3+\beta_3, +\cdots,\}$. Ibrahim then proves the following: The principal part of the products of terms in the expansion $(\{\lambda\}\otimes\{\omega\})(\{\mu\}\otimes\{\eta\})$, appear as terms in the expansion of

$$\{\lambda_1+\mu_1, \lambda_2+\mu_2, \lambda_3+\mu_3, \cdots,\}\otimes\{\nu\},$$

wherever $\chi^{(\omega)}\chi^{(\eta)}=\chi^{(\nu)}$, where (ω), (η), and (ν) are all partitions of n, and the χ's are characters of $S_{n!}$. Three special cases found by Ibrahim[14] are of use here.

1. The principal parts in the product

$$(\{\lambda\}\otimes\{n\})(\{\mu\}\otimes\{1^n\})$$

are the terms in the expansion of $\{\lambda_1+\mu_1, \lambda_2+\mu_2, \cdots,\}\otimes\{1^n\}$, where (n) is a partition into one part.

2. The principal parts in the product

$$(\{\lambda\}\otimes\{n\})(\{\mu\}\otimes\{n\})$$

are the terms in the expansion of $\{\lambda_1+\mu_1, \lambda_2+\mu_2, \cdots,\}\otimes\{n\}$.

3. The principal parts in the product

$$(\{\lambda\}\otimes\{1^n\})(\{\mu\}\otimes\{1^n\})$$

are the terms in the expansion of $\{\lambda_1+\mu_1, \lambda_2+\mu_2, \cdots,\}\otimes\{n\}$.

These three results provide a list of S functions which certainly belong to the reduction of the plethysm. While this list is not necessarily complete, it usually permits an unambiguous selection of the $\{\nu\}$'s of Eq. (15a) to be made.

Littlewood[5] has given two results that are of assistance in establishing further plethysms. If n is an integer, then

$$\{n\}\otimes\{2\}=\{2n\}+\{2n-2,2\}+\{2n-4,4\}+\cdots \tag{16}$$

to $(n+1)/2$ or $(n+2)/2$ terms, and

$$\{n\}\otimes\{1^2\}=\{2n-1,1\}+\{2n-3,3\}+\cdots \tag{17}$$

to $\frac{1}{2}(n+1)$ or $\frac{1}{2}n$ terms. Using these two results, together with the conjugation theorems of Eqs. (12) and (13), it is readily deduced for n *odd* that

$$\{1^n\}\otimes\{2\}=\{2^1,1^{2n-2}\}+\{2^3,1^{2n-6}\}+\cdots \tag{18}$$

and

$$\{1^n\}\otimes\{1^2\}=\{1^{2n}\}+\{2^2,1^{2n-4}\}+\{2^4,1^{2n-8}\}+\cdots, \tag{19}$$

both to $\frac{1}{2}(n+1)$ terms, while for n *even*

$$\{1^n\}\otimes\{2\}=\{1^{2n}\}+\{2^2,2^{2n-4}\}+\{2^4,1^{2n-8}\}+\cdots \tag{20}$$

to $\frac{1}{2}(n+2)$ terms, and

$$\{1^n\}\otimes\{1^2\}=\{2^1,1^{2n-2}\}+\{2^3,1^{2n-6}\}+\{2^5,1^{2n-10}\}+\cdots \tag{21}$$

to $\frac{1}{2}n$ terms.

Our problem of resolving the Kronecker square of a representation Γ into its symmetric and antisymmetric product representations amounts to forming the plethysms

$$[\Gamma^2]=\Gamma\otimes\{2\} \tag{22}$$

for the symmetric representations and

$$\{\Gamma^2\}=\Gamma\otimes\{1^2\} \tag{23}$$

for the antisymmetric representation.

Equations (16)–(21) allow many of the Kronecker squares of irreducible representations of the general linear group to be decomposed into their symmetric and antisymmetric representations immediately. To extend these results, it is necessary to make use of Eq. (15b), which reduces to

$$\sum_{\zeta,\nu}\Gamma_{1\zeta\nu}\{\zeta\}=\{\lambda\}\sum_{\mu}\Gamma_{1\mu\lambda}\{\mu\}, \tag{24}$$

if $\{\pi\}\equiv\{2\}$ or $\{1^2\}$ in Eq. (15a). There are two distinct choices for the series of S functions $\{\nu\}$ both of which yield the right-hand side. Ibrahim's three results give the distinction between the two sets.

As an example, consider the case

$$\{21\}\otimes\{2\}=\sum\{\nu\}.$$

Equation (15) gives

$$\begin{aligned}\sum\Gamma_{1\zeta\nu}\{\zeta\}&=\{21\}[\textstyle\sum\Gamma_{1\mu(21)}\{\mu\}]\\&=\{21\}[\{1^2\}+\{2\}]\\&=\{41\}+2\{32\}+2\{31^2\}+2\{2^21\}+\{21^3\}.\end{aligned}$$

Applying Ibrahim's theorem with $\{\lambda\}\equiv\{1\}$ and $\{\mu\}\equiv\{1^2\}$, we find that $(\{1\}\otimes\{2\})(\{1^2\}\otimes\{2\})$ has principal parts $\{42\}$ and $\{31^3\}$, allowing us to establish the complete list of $\{\nu\}$'s by noting that

$$\begin{aligned}\{42\}&\rightarrow\{32\}+\{41\},\\\{31^3\}&\rightarrow\{31^2\}+\{21^3\},\\\{2^3\}&\rightarrow\{2^21\},\\\{321\}&\rightarrow\{2^21\}+\{31^2\}+\{32\}\end{aligned}$$

[13] E. M. Ibrahim, Oxford Quart. J. Math. 3, 50 (1952).
[14] E. M. Ibrahim, Am. Math. Soc. Proc. 7, 199 (1956).

TABLE I. Decomposition of the Kronecker squares of irreducible representations of the general linear group.

$\{\lambda\}$	$\{\lambda\} \otimes \{2\}$	$\{\lambda\} \otimes \{1^2\}$
$\{21\}$	$\{42\} + \{31^3\} + \{321\} + \{2^3\}$	$\{41^3\} + \{3^3\} + \{321\} + \{2^21^3\}$
$\{21^3\}$	$\{42^3\} + \{41^4\} + \{3^21^3\} + \{321^4\} + \{2^4\} + \{2^21^4\}$	$\{421^2\} + \{3^32\} + \{321^3\} + \{31^5\} + \{2^31^3\}$
$\{2^3\}$	$\{4^3\} + \{42^3\} + \{2^21^3\} + \{2^4\}$	$\{431\} + \{32^21\}$
$\{2^21\}$	$\{4^3\} + \{4321\} + \{431^3\} + \{42^3\} + \{3^31\} + \{32^31\} + \{32^21^3\} + \{3^221^3\} + \{2^5\}$	$\{4^21^3\} + \{43^3\} + \{4321\} + \{42^21^3\} + \{3^32^3\} + \{3^221^3\} + \{3^21^4\} + \{32^31\} + \{2^41^3\}$
$\{2^3\}$	$\{4^3\} + \{4^22^3\} + \{43^21^3\} + \{42^4\} + \{3^32^21^3\} + \{2^6\}$	$\{4^331\} + \{432^21\} + \{3^31^3\} + \{32^41\}$

to give

$$\{21\} \otimes \{2\} = \{42\} + \{31^3\} + \{2^3\} + \{321\}.$$

Using Eq. (14) and the result for $\{21\} \otimes \{2\}$ immediately gives

$$\{21\} \otimes \{1^2\} = \{41^2\} + \{3^2\} + \{321\} + \{2^21^2\}.$$

This result may be checked by noting that

$$\{\lambda\} \otimes (\{2\} + \{1^2\}) = \{\lambda\}\{\lambda\}. \tag{25}$$

The method just outlined allows the rapid decomposition of the Kronecker squares of the irreducible representations of the general linear group given in Table I. If $\{\lambda\}$ is an irreducible representation of $GL(m)$, then all partitions occurring in the decomposition that have more than m parts will be null. Individual decompositions may be checked by noting that if the irreducible representation $\{\lambda\}$ of $GL(m)$ is of degree n, then the sum of the dimensions of the symmetric representations must be equal to that of the representation $\{2\}$ of $GL(n)$, i.e., $n[\frac{1}{2}(n+1)]$ while the sum of the antisymmetric representations must be equal to that of the representation $\{1^2\}$ of $GL(n)$, i.e., $n[\frac{1}{2}(n-1)]$. Since the representations of $GL(m)$ remain irreducible under restriction to the unitary group $U(m)$, the entries in Table I may be equally well applied to the decomposition of the Kronecker squares of the irreducible representations of the unitary groups.

III. PLETHYSMS FOR RESTRICTED GROUPS

It is well known that when a group is restricted to a subgroup, a set of functions which transforms irreducibly under the full group is, in general, reducible under the subgroup. Thus, the S functions $\{\lambda\}$ corresponding to the irreducible representation $\{\lambda\}$ of $GL(m)$ will be reducible under restriction to the rotation of symplectic groups. This means that the transformation matrix $A^{\{\lambda\}}$ for the representation $\{\lambda\}$ can be expressed as a direct sum of matrices irreducible under the lower group, e.g.,

$$A^{\{\lambda\}} = \dot{\sum} A^{[\alpha]}$$

for the rotation group, and

$$A^{\{\lambda\}} = \dot{\sum} A^{\langle\alpha\rangle}$$

for the symplectic group.

Following Littlewood,[7] we denote $[\alpha]$ and $\langle\alpha\rangle$ as the spurs of the irreducible matrices $A^{[\alpha]}$ and $A^{[\alpha]}$, respectively, i.e., $[\alpha]$ is an S function for the rotation group and $\langle\alpha\rangle$ for the symplectic group. Rules for performing the decompositions $\{\lambda\} \to \sum [\alpha]$ and $\{\lambda\} \to \sum [\alpha]$ have been given by Littlewood.[7] In many cases, the decomposition of a representation $\{\lambda\}$ of $GL(n)$ or $U(n)$ gives rise to nonstandard symbols containing more parts than allowed for the representations of the appropriate subgroup. Methods of expressing these nonstandard symbols in terms of standard symbols have been discussed by several authors.[6,7,15–17]

Plethysms for the rotation group $R(n)$ may be obtained from those established for the unitary group $U(n)$ by noting that

$$[\lambda] = \{\lambda\} + \sum (-1)^{\rho/2} \Gamma_{\nu\mu\lambda} \{\mu\}, \tag{26}$$

where $\Gamma_{\nu\mu\lambda}$ is the coefficient of $\{\lambda\}$ in the product $\{\nu\}\{\mu\}$ and the sum is taken over all partitions of ρ for which the Frobenius notation is[6]

$$\binom{r+1}{r}, \quad \binom{r+1, s+1}{r \ , \ s}, \quad \binom{r+1, s+1, t+1}{r \ , \ s \ , \ t} \cdots.$$

These partitions appear in the expansion

$$1 + \sum (-1)^{\rho/2} \{\nu\} = 1 - \{2\} + \{31\} - \{41^2\} - \{3^2\} + \cdots.$$

The plethysm is then given in terms of plethysms for $U(n)$ by the expression

$$[\lambda] \otimes \{\eta\} = [\{\lambda\} + \sum (-1)^{\rho/2} \Gamma_{\nu\mu\lambda} \{\mu\}] \otimes \{\eta\}. \tag{27}$$

[15] F. D. Murnaghan, *The Theory of Group Representations* (The Johns Hopkins Press, Baltimore, 1938).
[16] B. H. Flowers, Proc. Roy. Soc. (London) **A212**, 248 (1952).
[17] J. D. Darling and R. G. Seyler, Acta Phys. Acad. Sci. Hung. **21**, 33 (1966).

TABLE II. Decomposition of the Kronecker squares of irreducible representations of the general rotation group.

$[\lambda]$	$[\lambda] \otimes \{2\}$	$[\lambda] \otimes \{1^2\}$
[1]	$\{2\}$	$\{1^2\}$
[2]	$\{4\} + \{22\} - \{2\}$	$\{31\} + \{0\} - \{2\}$
[3]	$\{6\} + \{42\} + \{1^2\} - \{4\} - \{31\}$	$\{51\} + \{3^2\} + \{2\} - \{4\} - \{31\}$
[4]	$\{8\} + \{62\} + \{44\} + \{31\} - \{6\} - \{51\} - \{42\}$	$\{71\} + \{53\} + \{4\} + \{2^2\} - \{6\} - \{51\} - \{42\}$
[21]	$\{42\} + \{321\} + \{31^3\} + \{2^3\} + \{1^2\} - \{31\} - \{22\} - \{211\}$	$\{41^2\} + \{3^2\} + \{321\} + \{2^21^2\} + \{2\} - \{31\} - \{2^2\} - \{21^2\}$
[2²]	$\{44\} + \{42^2\} + \{3^21^2\} + \{31\} + \{2^4\} - \{42\} - \{321\} - \{2^3\}$	$\{431\} + \{4\} + \{32^21\} + \{2^2\} - \{42\} - \{321\} - \{2^3\}$
[21²]	$\{42^2\} + \{41^4\} + \{3^21^2\} + \{321^3\} + \{2^4\} + \{2^21^4\} + \{21^2\} - \{321\} - \{31^3\} - \{2^3\} - \{2^21^2\} - \{21^4\}$	$\{421^2\} + \{3^22\} + \{321^3\} + \{31^5\} + \{2^31^2\} + \{2^2\} + \{1^4\} - \{321\} - \{31^3\} - \{2^3\} - \{2^21^2\} - \{21^4\}$
[2²1]	$\{4^32\} + \{4321\} + \{431^3\} + \{42^3\} + \{41^2\} + \{3^31\} + \{3^221^2\} + \{3^3\} + \{32^31\} + \{32^21^3\} + \{321\} + \{2^5\} + \{2^31^2\} - \{431\} - \{42^2\} - \{421^2\} - \{3^32\} - \{32^21\} - \{2^4\} - \{2^31^2\}$	$\{4^21^2\} + \{43^2\} + \{4321\} + \{42^21^2\} + \{42\} + \{3^2Z^4\} + \{3^221^2\} + \{3^21^4\} + \{32^31\} + \{321\} + \{31^3\} + \{2^41^2\} + \{2^3\} - \{431\} - \{42^2\} - \{421^2\} - \{3^32\} - \{32^21\} - \{2^4\} - \{2^31^2\}$
[2³]	$\{4^3\} + \{4^22^2\} + \{43^21^2\} + \{431\} + \{42^4\} + \{3^22^21^2\} + \{32^21\} - \{4^32\} - \{4321\} - \{3^221^2\} - \{32^21\} - \{2^5\}$	$\{4^231\} + \{4^2\} + \{432^21\} + \{42^3\} + \{3^31^3\} + \{3^21^2\} + \{32^41\} + \{2^4\} - \{4^32\} - \{4321\} - \{3^221^2\} - \{32^21\} - \{2^5\}$

The plethysms for $\{\eta\} \equiv \{2\}$ or $\{1^2\}$ may then be readily calculated, using the results of Table I together with the special forms of Eqs. (7) and (8),

$$(A - B) \otimes \{2\} = A \otimes \{2\} + B \otimes \{1^2\} - AB$$

and

$$(A - B) \otimes \{1^2\} = A \otimes \{1^2\} + B \otimes \{2\} - AB,$$

and their extensions. For example,

$$\begin{aligned}[21] \otimes \{2\} &= [\{21\} - \{1\}] \otimes \{2\} \\ &= \{21\} \otimes \{2\} + \{1\} \otimes \{1^2\} - \{21\}\{1\} \\ &= \{42\} + \{31^3\} + \{321\} + \{2^3\} + \{1^2\} \\ &\quad - \{31\} - \{2^2\} - \{21^2\}.\end{aligned}$$

The decomposition of the Kronecker squares of irreducible representations of the rotation group into their symmetric and antisymmetric parts is given in Table II in terms of representations of the corresponding unitary groups. To obtain these decompositions for a particular rotation group $R(n)$, we first strike out all representations of $U(n)$ that contain more than n parts, and then decompose the remaining representations into those of $R(n)$, as in Table III.

Plethysms for the rotation and symplectic groups can also be found using the rather remarkable theorem due to Littlewood,[18] which states that if

$$\sum \Gamma_{\xi\zeta\lambda}\Gamma_{\xi\nu\mu}\{\zeta\}\{\nu\} = \sum K_{\lambda\mu\rho}\{\rho\}, \tag{28a}$$

the summation on the left being with respect to all possible S functions including $\{\xi\} = \{0\}$, then

$$[\lambda][\mu] = \sum K_{\lambda\mu\rho}[\rho] \tag{28b}$$

and

$$\langle\lambda\rangle\langle\mu\rangle = \sum K_{\lambda\mu\rho}\langle\rho\rangle. \tag{28c}$$

Two special cases[18] of direct relevance to the present problem may be derived from the above result. If (μ) is a partition of 2, then:

(1)
$$[\lambda] \otimes \{\mu\} = \sum H_{\lambda\mu\nu}[\nu], \tag{29a}$$

where

$$\sum H_{\lambda\mu\nu}\{\nu\} = \sum (\Gamma_{\xi\eta\lambda}\{\eta\}) \otimes \{\mu\} + \sum \Gamma_{\xi\eta\lambda}\Gamma_{\xi\zeta\lambda}\{\eta\}\{\zeta\}, \quad (\eta) \neq (\zeta) \tag{29b}$$

summed for all suitable S functions $\{\xi\}$, $\{\eta\}$, $\{\zeta\}$, the last term not being repeated for the interchange of $\{\eta\}$ and $\{\zeta\}$;

(2)
$$\langle\lambda\rangle \otimes \{\mu\} = \sum J_{\lambda\mu\nu}\langle\nu\rangle, \tag{30a}$$

where

$$\sum J_{\lambda\mu\nu}\{\nu\} = \sum (\Gamma_{\xi\eta\lambda}\{\eta\}) \otimes (\{\mu\} \cdot \{\epsilon\}) + \sum \Gamma_{\xi\eta\lambda}\Gamma_{\xi\zeta\lambda}\{\eta\}\{\zeta\}, \quad (\eta) \neq (\zeta), \tag{30b}$$

in which $(\epsilon) = (2)$ if $\{\xi\}$ is of *even* weight, but $(\epsilon) = (1^2)$ if $\{\xi\}$ is of *odd* weight. $(\{\mu\} \cdot \{\epsilon\})$ denotes an inner product of S functions.

The above two results do have the advantage of yielding the decomposition of the Kronecker square directly in terms of the representations appropriate to the rotation or symplectic group concerned, rather than in terms of representations of the general linear group (which must then be reduced.) It is somewhat surprising that Eqs. (28a)–(28c) have not been applied to the general problem of decomposing Kronecker products of representations of the rotation and symplectic groups more frequently. Examples of plethysms for the symplectic groups are given in Tables IV and V.

IV. KRONECKER SQUARES FOR THE GROUP G_2

Judd and Wadzinski[3] have considered the resolution of the irreducible representations contained in the decomposition of the Kronecker squares of the representations of G_2 into their symmetric and antisymmetric results. The results they give for the

[18] D. E. Littlewood, Can. J. Math. 10, 17 (1958).

TABLE III. Decomposition of the Kronecker squares of irreducible representations of the rotation group R_5.

$[\lambda]$	$[\lambda] \otimes \{2\}$	$[\lambda] \otimes \{1^2\}$
[0]	[0]	
[1]	[2] + [0]	[11]
$[1^2]$	[22] + [2] + [1] + [0]	[21] + [11]
[2]	[4] + [22] + [2] + [0]	[31] + [11]
[21]	[42] + [40] + [32] + [31] + [30] + 2[22] + [21] + 2[2] + [11] + [1] + [0]	[41] + [33] + [32] + 2[31] + 2[21] + 2[11]
[22]	[44] + [42] + [4] + [32] + [3] + [22] + [2]	[43] + [41] + [33] + [31] + [21]

representations $(\omega 0)$ of G_2 may be obtained readily by noting that the irreducible representations of R_7 that contain only one part are irreducible under restriction to G_2; hence

$$(\omega 0) \otimes \{\lambda\} = [\omega 00] \otimes \{\lambda\} = (\{\omega\} - \{\omega - 2\}) \otimes \{\lambda\}.$$

The remaining entries in their Table I may be found directly from the plethysms found for the group R_7 by noting that if Γ^a labels the representations of a group G, γ^α those of a subgroup g, and $\Gamma^a \to \sum_\alpha \gamma_a^\alpha$ under restriction to g, and, if

$$\Gamma^a \otimes \{\lambda\} = \sum_b \Gamma^b \to \sum_{b,\beta} \gamma_b^\beta, \tag{31a}$$

where $\{\lambda\} \equiv \{2\}$ or $\{1^2\}$, then

$$\left[\sum_\alpha \gamma_a^\alpha\right] \otimes \{\lambda\} = \sum_\alpha (\gamma_a^\alpha \otimes \{\lambda\}) + \sum_{\alpha<\rho} \gamma_a^\alpha \gamma_a^\rho = \sum_{b,\beta} \gamma_b^\beta. \tag{31b}$$

For example, to evaluate $(11) \otimes \{1^2\}$, we use Eq. (31a) to give

$$[110] \otimes \{1^2\} = [110] + [211] \to 2(10) + 2(11) + (20) + (21) + (30),$$

and then Eq. (31b) to yield

$$[(10) + (11)] \otimes \{1^2\} = (10) \otimes \{1^2\} + (11) \otimes \{1^2\} + (10) \times (11).$$

Using the fact that $(10) \otimes \{1^2\} = (10) + (11)$ and $(10) \times (11) = (10) + (20) + (21)$, we deduce that $(11) \otimes \{1^2\} = (11) + (30)$. The other entries of Judd and Wadzinski's Table I follow in a similar manner.

V. KRONECKER SQUARES FOR FINITE GROUPS

The separation of the Kronecker squares of the irreducible representations of a finite group G into their symmetric and antisymmetric parts has been discussed by Hamermesh[1] and Griffith.[2] The Kronecker squares of the irredicible representations $D^{(J)}$ of the three-dimensional rotation group R_3 may be readily

TABLE IV. Decomposition of the Kronecker squares of irreducible representations of the general symplectic group.

$\langle\sigma\rangle$	$\langle\sigma\rangle \otimes \{2\}$	$\langle\sigma\rangle \otimes \{1^2\}$
$\langle 0\rangle$	$\{0\}$	
$\langle 1\rangle$	$\{2\}$	$\{1^2\}$
$\langle 1^2\rangle$	$\{2^2\} + \{1^4\} - \{1^2\}$	$\{21^2\} + \{0\} - \{1^2\}$
$\langle 1^3\rangle$	$\{2^3\} + \{21^4\} + \{1^7\} - \{21^2\} - \{1^4\}$	$\{2^21^2\} + \{2\} + \{1^6\} - \{21^2\} - \{1^4\}$
$\langle 1^4\rangle$	$\{2^4\} + \{2^21^4\} + \{21^8\} + \{1^8\} - \{2^21^2\} - \{21^4\} - \{1^6\}$	$\{2^21^2\} + \{2^2\} + \{21^6\} + \{1^4\} - \{2^21^2\} - \{21^4\} - \{1^6\}$
$\langle 21\rangle$	$\{42\} + \{321\} + \{31^3\} + \{2^3\} + \{1^2\} - \{31\} - \{2^2\} - \{21^2\}$	$\{41^2\} + \{3^2\} + \{321\} + \{2^21^2\} + \{2\} - \{31\} - \{2^2\} - \{21^2\}$
$\langle 2^2\rangle$	$\{4^2\} + \{42^2\} + \{3^21^2\} + \{2^4\} + \{21^2\} - \{3^2\} - \{321\} - \{2^21^2\}$	$\{431\} + \{32^21\} + \{2^2\} + \{1^4\} - \{3^2\} - \{321\} - \{2^21^2\}$
$\langle 21^2\rangle$	$\{42^2\} + \{41^4\} + \{3^21^2\} + \{321^3\} + 2\{31\} + \{2^4\} + \{2^21^4\} + 3\{21^2\} + \{0\} - \{41^2\} - 2\{321\} - 2\{31^3\} - \{2^3\} - 2\{2^21^2\} - \{21^4\} - \{2\} - \{1^2\}$	$\{421^2\} + \{4\} + \{3^22\} + \{321^3\} + \{31^5\} + \{31\} + \{2^21^2\} + 2\{22\} + 2\{21^2\} + \{1^4\} - 2\{321\} - 2\{31^3\} - \{2^3\} - \{2^21^2\} - \{21^4\} - \{2\} - \{1^2\}$
$\langle 2^21\rangle$	$\{4^22\} + \{4321\} + \{431^3\} + \{42^3\} + \{41^2\} + \{3^21\} + \{3^221^2\} + \{3^3\} + \{32^31\} + \{32^21^3\} + 3\{321\} + \{31^3\} + \{2^5\} + \{2^3\} + 3\{2^21^2\} + \{21^4\} + \{2\} + \{1^6\} - \{431\} - \{42^3\} - \{421^2\} - 2\{3^22\} - \{3^21^2\} - 3\{32^31\} - \{321^3\} - \{31\} - \{2^4\} - 2\{2^21^2\} - \{2^21^4\} - \{21^2\} - \{1^4\}$	$\{4^21^2\} + \{43^2\} + \{4321\} + \{42^21^2\} + \{42\} + \{3^22^2\} + \{3^221^2\} + \{3^21^4\} + \{32^31\} + 3\{321\} + 2\{31^3\} + \{2^41^2\} + 3\{2^3\} + 2\{2^21^2\} + \{21^4\} + \{1^2\} - \{431\} - \{42^3\} - \{421^2\} - 2\{3^22\} - \{3^21^2\} - 3\{32^31\} - \{321^3\} - \{31\} - \{2^4\} - 2\{2^21^2\} - \{2^21^4\} - \{2^2\} - 2\{21^3\} - \{1^4\}$
$\langle 2^3\rangle$	$\{4^3\} + \{4^22^2\} + \{43^21^2\} + \{42^4\} + \{421^2\} + \{3^22^21^2\} + 2\{3^32\} + \{32^31\} + \{321^3\} + \{31^5\} + \{2^6\} + 2\{2^21^2\} + \{2^3\} + \{2^2\} + \{1^4\} + \{1^2\} + \{0\} - \{432\} - \{431^2\} - \{42^21\} - \{421^3\} - \{3^3\} - \{3^221\} - \{32^21^2\} - \{321\} - \{31^3\} - \{2^41\} - \{2^21^3\} - \{2^3\} - \{2^21^2\} - \{21^4\} - \{21^2\}$	$\{4^231\} + \{432^21\} + \{42^2\} + \{41^4\} + \{3^21^3\} + \{3^32\} + \{3^21^2\} + \{32^41\} + \{32^21\} + \{321^3\} + \{2^4\} + \{2^21^2\} + \{2^3\} + \{2^21^4\} + \{1^2\} - \{432\} - \{431^2\} - \{42^21\} - \{421^3\} - \{3^3\} - \{3^221\} - \{32^21^2\} - \{321\} - \{31^3\} - \{2^41\} - \{2^21^3\} - \{2^3\} - \{2^21^2\} - \{21^4\}$

TABLE V. Decomposition of the Kronecker squares of the irreducible representations of the symplectic group Sp_8.

$\langle\sigma\rangle$	$\langle\sigma\rangle \otimes \{2\}$	$\langle\sigma\rangle \otimes \{1^2\}$
$\langle 0\rangle$	$\langle 0\rangle$	
$\langle 1\rangle$	$\langle 2\rangle$	$\langle 1^2\rangle + \langle 0\rangle$
$\langle 1^2\rangle$	$\langle 2^2\rangle + \langle 1^4\rangle + \langle 1^2\rangle + \langle 0\rangle$	$\langle 21^2\rangle + \langle 2\rangle$
$\langle 1^3\rangle$	$\langle 2^3\rangle + \langle 21^2\rangle + \langle 2\rangle$	$\langle 2^21^2\rangle + \langle 2^2\rangle + \langle 1^2\rangle + \langle 0\rangle$
$\langle 1^4\rangle$	$\langle 2^4\rangle + \langle 2^2\rangle + \langle 0\rangle$	$\langle 2^3\rangle + \langle 2\rangle$

decomposed into their symmetric and antisymmetric parts by first noting that for integral J

$$[D^{(J)}] = D^{(J)} \otimes \{2\} = 0, 2, \cdots, 2J$$

and

$$\{D^{(J)}\} = D^{(J)} \otimes \{1^2\} = 1, 3, \cdots, 2J - 1,$$

while for half integer J,

$$[D^{(J)}] = D^{(J)} \otimes \{2\} = 1, 3, \cdots, 2J,$$

$$\{D^{(J)}\} = D^{(J)} \otimes \{1^2\} = 0, 2, \cdots, 2J - 1,$$

and then decomposing the irreducible representations $D^{(J)}$ into those of the finite group G, followed by use of Eqs. (31a) and (31b).

For example, in the case of the icosahedral group K, we have $D^{(1)} \to T_1$.

$$D^{(1)} \otimes \{2\} = D^{(0)} + D^{(2)} \to A + V,$$

and

$$D^{(1)} \otimes \{1^2\} = D^{(1)} \to T_1,$$

giving

$$T_1 \otimes \{2\} = A + V \quad \text{and} \quad T_1 \otimes \{1^2\} = T_1.$$

The other Kronecker squares for the icosahedral group may be evaluated in a similar manner to give the results of Table VI. Similar results may be readily obtained for any other finite group that is a subgroup of R_3.

VI. SELECTION RULES AND PLETHYSM

Judd and Wadzinski[3] have considered the application of the resolution of the Kronecker squares of irreducible representations into their symmetric and antisymmetric parts to the determination of selection rules. Let us consider an operator $f^{(\Gamma_2\gamma_2)}$ that transforms as the Γ_2 representation of a group G and as the γ_2 representation of a subgroup g of G. The matrix element

$$\langle \psi^{(\Gamma_1\gamma_1)}, f^{(\Gamma_2\gamma_2)} \phi^{(\Gamma_2\gamma_2)} \rangle$$

will certainly vanish, unless either

$$\Gamma_2 \otimes \{2\} \supset \Gamma_1 \quad \text{and} \quad \gamma_2 \otimes \{2\} \supset \gamma_1$$

TABLE VI. Decomposition of the Kronecker squares of the irreducible representations of the icosahedral group K.

Γ	$\Gamma \otimes \{2\}$	$\Gamma \otimes \{1^2\}$
A	A	
T_1	$A + V$	T_1
V	$A + 2V + U$	$T_1 + T_2 + U$
T_2	$A + V$	T_2
U	$A + V + U$	$T_1 + T_2$
E'	T_1	A
E''	T_2	A
U'	$T_1 + T_2 + U$	$A + V$
W'	$2T_1 + 2T_2 + U + V$	$A + 2V + U$

or

$$\Gamma_2 \otimes \{1^2\} \supset \Gamma_1 \quad \text{and} \quad \gamma_2 \otimes \{1^2\} \supset \gamma_1.$$

These dual conditions make use of the fact that functions that transform as Γ_1 under G must be of the same symmetry classification as those spanning the γ_1 representation of the subgroup g.

As an example, consider the isoscalar factor

$$\langle [22]L \mid [20]k + [20]k \rangle,$$

where [22] and [20] are irreducible representations of R_5 and L and k are the irreducible representations $D^{[L]}$ and $D^{[k]}$ of R_3. We have

$$[20] \otimes \{2\} \supset [22] \quad \text{and} \quad [k] \otimes \{2\} \supset 0, 2, \cdots, 2k.$$

Hence, we conclude that the isoscalar factor must vanish for *odd* values of L.

Similar applications to finite groups are possible, some of which have been discussed by Hamermesh[1] and Griffith.[2]

VII. CONCLUSION

In the preceding, we have attempted to indicate some of the applications of Littlewood's technique of plethysm to the resolution of the Kronecker squares of irreducible representations into their symmetric and antisymmetric parts. Plethysm plays an important part in physics wherever the theory of groups enters. Two typical problems, where plethysm introduces remarkable simplifications, are the group theoretical analysis of the n-particle operators that may be constructed from a basic set of single-particle operators and the general problem of the classification of states of n-particle systems. Examples of these two applications will be considered in a later paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Exact Occupation Statistics for One-Dimensional Arrays of Dumbbells

D. LICHTMAN AND R. B. MCQUISTAN
University of Wisconsin and Laboratory for Surface Studies, Milwaukee, Wisconsin

(Received 17 March 1966)

Exact relationships are developed which describe the occupation statistics for one-dimensional arrays of dumbbells. It is shown that in the limit, as the number of compartments per array tends to infinity, these relationships reduce to those calculated, using the Bethe approximation when the number of nearest neighbors is two. A partition function, which includes the influence of the configurational correlation inherent in a one-dimensional array of dumbbells, is also derived and discussed.

I. INTRODUCTION

THERE exists in the literature a large number of papers[1–6] concerned with the statistical analysis of one-dimensional systems. The rationale for treating such systems is that it is often possible to perform a relatively thorough investigation of their statistical-mechanical properties, and knowledge thus gained may be of value when considering systems of higher dimensionality.

In the present paper we investigate the occupation statistics of particles which occupy two adjacent compartments in a linear array of compartments. In such systems there is configurational correlation in the sense that if a compartment is occupied then at least one of its nearest neighbors is also occupied. Thus, there does not exist a random distribution of occupied compartments but rather a random distribution of *pairs* of occupied compartments.

Utilizing the Bethe approximation[7] and methods developed by Peierls[8] and Fowler,[9,10] Roberts and Miller[11,12] have treated the statistical problem of dumbbells in connection with the adsorption of diatomic molecules. Within the limits imposed by the Bethe approximation, they were able to determine (1) the probability that a vacant compartment is isolated, (2) the probability that if a compartment is occupied a given nearest neighbor is occupied by part of another dumbbell, and (3) the probability of success in placing an additional dumbbell on a linear array of N compartments containing q dumbbells. The present paper is an attempt to calculate these quantities without recourse to the Bethe approximation and to determine the exact statistics which can be used to describe linear arrays of dumbbells. Gornick and Jackson[13] have treated the mathematically related problem of the sequence selection in the crystallization of linear polymer chains.

II. THE NUMBER OF WAYS OF PLACING q DUMBBELLS ON A LINEAR ARRAY OF N COMPARTMENTS

The statistical system to be investigated here has the following character: (a) A linear array of N identical compartments; (b) indistinguishable particles (dumbbells) which occupy contiguous pairs of compartments; (c) the occupancy of each compartment is either 0 or 1.

The first quantity which must be determined is $W(q, N)$, the number of ways in which q indistinguishable dumbbells can be arranged on a linear array of N compartments. If the number of dumbbells on a linear array of N compartments is q, then there are $(N - 2q)$ vacant compartments and $(N - 2q) + q = N - q$ individual "things" to be permuted of which q are of one kind and $N - 2q$ of another. Thus, $W(q, N)$ is given by

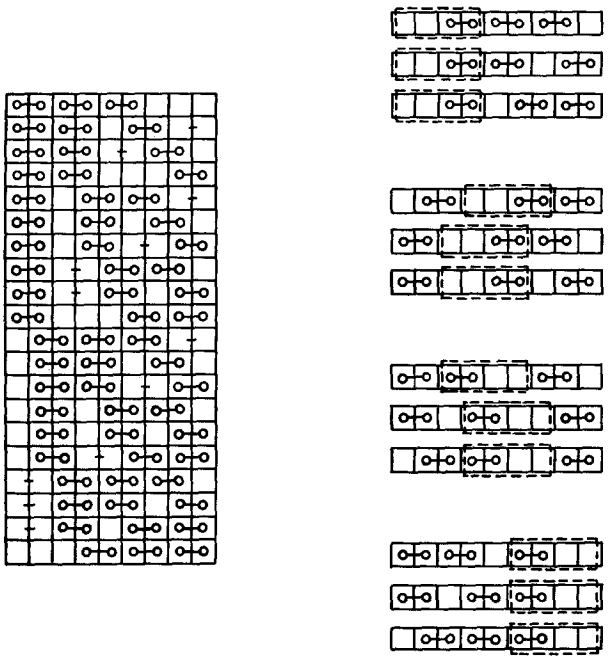
$$W(q, N) = \frac{(N - q)!}{q!\,(N - 2q)!} = \binom{N - q}{q}. \qquad (1)$$

In Fig. 1(a) we see, for example, that there are 20 distinguishable ways of arranging 3 dumbbells on a linear array of 9 compartments. For large values of N, $W(q, N)$ becomes

$$W(q, N) = 2^{N-q}\left[\frac{2}{[\pi(N-q)]^{\frac{1}{2}}}\right] \exp\left[-\frac{(N - 3q)^2}{2(N - q)}\right] \qquad (2)$$

which is maximized for a given value of N at $q = N/3$.

[1] E. Ising, Z. Physik **31**, 253 (1925).
[2] A. Lenard, J. Math. Phys. **2**, 682 (1961).
[3] M. Kac, Phys. Fluids **2**, 8 (1959).
[4] S. F. Edwards and A. Lenard, J. Math. Phys. **3**, 778 (1962).
[5] J. M. J. Van Leeuwen, J. Chem. Phys. **41**, 2959 (1964).
[6] I. P. Pavlotski, Dokl. Akad. Nauk SSSR **161**, 1055 (1965) [English transl.: Soviet Phys.—Doklady **10**, 301 (1965)].
[7] H. A. Bethe, Proc. Roy. Soc. (London) **A150**, 59 (1935).
[8] R. Peierls, Proc. Cambridge Phil. Soc. **32**, 471 (1936).
[9] R. H. Fowler, Proc. Cambridge Phil. Soc. **31**, 260 (1935).
[10] R. H. Fowler, Proc. Cambridge Phil. Soc. **32**, 144 (1936).
[11] J. K. Roberts, Proc. Roy. Soc. (London) **A161**, 141 (1937).
[12] J. K. Roberts and A. R. Miller, Proc. Cambridge Phil. Soc. **35**, 293 (1939).
[13] F. Gornick and J. L. Jackson, J. Chem. Phys. **38**, 1150 (1963).



(a) (b)

FIG. 1. (a) The number of distinguishable arrangements when 3 dumbbells are placed on a linear array of 9 compartments. (b) Total number of double vacancies ($p = 2$) when 3 dumbbells are placed on a linear array of 9 compartments in all distinguishable arrangements.

III. THE NUMBER OF p-TUPLE CONTIGUOUS VACANCIES

We must also determine $N_p(q, N)$, the number of p-tuple contiguous vacant compartments created when q dumbbells are placed on a linear array of N compartments in all possible arrangements. If we consider the p-contiguous vacant compartments which make up the p-tuple vacancy *plus* the dumbbell which terminates it at one end as a unit, then there are a total of $(N - q - p - 1)$ individuals remaining which can be permuted. Of these, $(q - 1)$ are dumbbells because one of the dumbbells serves to terminate the p-tuple vacancy and $N - 2q - p$ are vacant compartments. Each of the arrangements arising from the permutation of $(N - q - p - 1)$ individuals may be created in $(q - 1) + 2 = (q + 1)$ ways because the p-tuple vacancy may be interspersed between the q dumbbells in $q - 1$ ways and also on each end of the array [see Fig. 1(b)] so that

$$N_p(q, N) = \frac{(q + 1)(N - q - p - 1)!}{(q - 1)!\,(N - 2q - p)!} = (q + 1)\binom{N - q - p - 1}{q - 1}. \quad (3)$$

Thus $P_1(q, N)$, the probability that a compartment is vacant and isolated is given by

$$P_1(q, N) = \frac{N_1(q, N)}{NW(q, N)}, \quad (4)$$

where N_1 is obtained from Eq. (3) when $p = 1$; then

$$P_1(q, N) = \frac{q(q + 1)(N - 2q)}{N(N - q)(N - q - 1)}. \quad (5)$$

If we define $\theta \equiv 2q/N$ and let N tend to infinity, we obtain

$$P_1(\theta) = \lim_{N\to\infty} P_1(q, N) = (1 - \theta)\left(\frac{\theta}{2 - \theta}\right)^2 \quad (6)$$

which is *precisely* the result obtained when the Bethe approximation[12] is utilized in this problem. Figure 2 shows P_1 as a function of θ. $P_{1\,\max}$ is ≈ 0.0925 and occurs at $\theta_{\max} = 3 - 5^{\frac{1}{2}} \approx 0.766$.

The probability of success, $S(q, N)$, when attempting to place in a random manner an additional dumbbell on a linear array of N compartments containing q dumbbells is given by

$$S(q, N) = \frac{N_{2v}(q, N)}{(N - 1)W(q, N)}, \quad (7)$$

where $N_{2v}(q, N)$ is the number of different ways in which one additional dumbbell may be placed on all the distinguishable arrangements created when previously q dumbbells are placed on a linear array of N compartments in all possible ways. The factor $(N - 1)$ arises because there are $(N - 1)$ possible places for a dumbbell on a linear array of N compartments. Since each p-tuple vacancy can accommodate $(p - 1)$ dumbbells, $N_{2v}(q, N)$ is given by

$$N_{2v}(q, N) = \sum_{p=1}^{N-2q} (p - 1)N_p(q, N) = \frac{(N - q - 1)!}{q!\,(N - 2q - 2)!}, \quad (8)$$

so that $S(q, N)$ becomes

$$S(q, N) = \frac{(N - 2q)(N - 2q - 1)}{(N - q)(N - 1)}. \quad (9)$$

In the limit, as N tends to infinity, Eq. (9) may be written

$$S(\theta) = \lim_{N\to\infty} S(q, N) = \frac{2(1 - \theta)^2}{(2 - \theta)}, \quad (10)$$

where $\theta \equiv 2q/N$.

Again this result is *precisely* that given by the Bethe approximation.[12]

IV. THE NUMBER OF CONTIGUOUS DUMBBELLS

The number of runs of r contiguous dumbbells, $N_r(q, N)$, may be determined in the following manner.

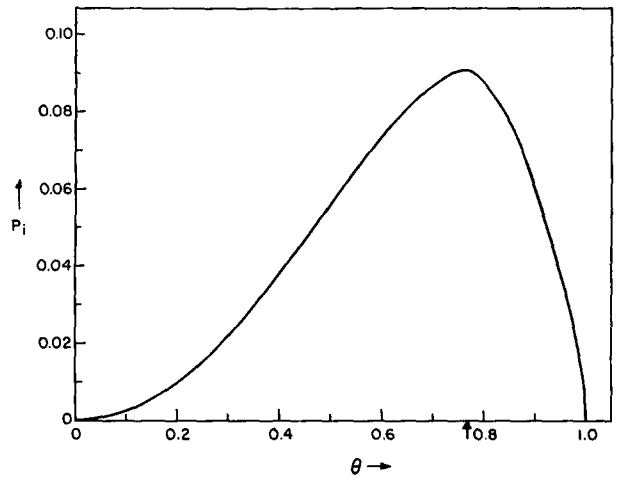


FIG. 2. P_i, the probability that a site is isolated and vacant as a function of coverage.

If we consider the run of r contiguous dumbbells *and* the vacancy that terminates the run as a unit, then there are $(N-q-r-1)$ individuals remaining. Of these, $(q-r)$ are dumbbells and $(N-2q-1)$ are vacancies. Each of the arrangements arising from the permutation of $(N-q-r-1)$ individuals may be created in $(N-2q+1)$ ways. Thus,

$$N_r(q, N) = \frac{(N-2q+1)(N-q-r-1)!}{(N-2q-1)!\,(q-r)!} = (N-2q+1)\binom{N-q-r-1}{q-r}. \quad (11)$$

If a compartment is occupied, $P_n(q, N)$, the probability that a nearest neighbor is occupied by part of another dumbbell is given by

$$P_n(q, N) = \frac{1}{NW(q, N)} \sum_{r=1}^{q} (r-1) N_r(q, N), \quad (12)$$

because each run of r dumbbells contains $(r-1)$ pairs of nearest-neighbor compartments which are occupied by part of other dumbbells. Equation (12) yields

$$P_n(q, N) = \frac{q(q-1)}{N(N-q)}. \quad (13)$$

In the limit, as N tends to infinity we obtain

$$P_n(\theta) = \lim_{N\to\infty} P_n(q, N) = \frac{(\theta)^2}{2(2-\theta)}, \quad (14)$$

which again is the result obtained utilizing the Bethe approximation.[12]

V. DISTRIBUTION FUNCTION WITH CONFIGURATIONAL CORRELATION

Theoretical treatments of superconductivity, the statistics of electrons in semiconductors, cooperative phenomena, etc., involve the consideration of several types of correlation. If the nature of the correlation is such that the occupation of a cell in phase space precludes the occupation of a nearest-neighbor cell, or if the cells are always occupied in adjacent pairs, then there exists configurational correlation of the kind discussed in the present paper. Therefore, even for a one-dimensional space, it is interesting to examine the influence of configurational correlation on the form of the distribution function.

If W_j, the number of ways of arranging q_j dumbbells on the jth cell composed of a linear array of N compartments is given by [see Eq. (1)]

$$W_j = \frac{(N-q_j)!}{q_j!\,(N-2q_j)!}, \quad (15)$$

then the probability of a given macrostate in these dumbbell statistics is

$$W = \prod_j W_j = \prod_j \frac{(N-q_j)!}{(q_j)!\,(N-2q_j)!}, \quad (16)$$

where q_j is the number of dumbbells on the jth array of N compartments. As in any type of statistics, we assume that the entropy is proportional to the logarithm of the thermodynamic probability and that $\ln W$ is a maximum at the equilibrium state of maximum entropy. To determine the maximum value of $\ln W$, we proceed in the following manner. The logarithm of Eq. (16) yields

$$\ln W = \sum_j \{\ln (N-q_j)! - \ln (q_j)! - \ln (N-2q_j)!\}. \quad (17)$$

Using the Stirling approximation yields

$$\ln W = \sum_j \{(N-q_j)\ln(N-q_j) - q_j \ln q_j - (N-2q_j)\ln(N-2q_j)\}. \quad (18)$$

If W_m represents the maximum probability and at this condition q_{jm} is the corresponding number of dumbbells in the jth array, then

$$\delta \ln W = \sum_j \ln\left[\frac{(N-2q_{jm})^2}{q_{jm}(N-q_{jm})}\right]\delta q_j = 0. \quad (19)$$

If the number of particles and system energy are constant, then

$$\delta q = \sum_j \delta q_j = 0, \quad (20)$$

$$\delta U = \sum_j E_j \delta q_j = 0. \quad (21)$$

Adding Eqs. (19)–(21) yields

$$\sum_j \left\{\ln\left[\frac{(N-2q_{jm})^2}{q_{jm}(N-q_{jm})}\right] - \ln B - \beta E_j\right\}\delta q_j = 0, \quad (22)$$

where $\ln B$ and β are the undetermined multipliers

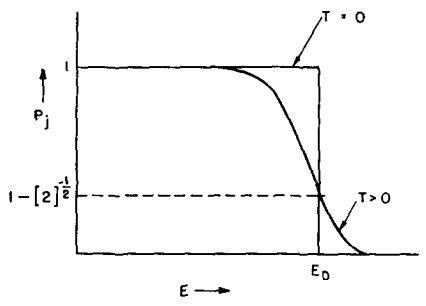


FIG. 3. P_j, the probability that a level of energy E_j is occupied as a function of E for various temperatures.

and where q_{jm} is the number of dumbbells in the jth array when W is a maximum. Equation (22) yields

$$\frac{(N-2q_{jm})^2}{q_{jm}(N-q_{jm})} = B\exp[\beta E_j]. \tag{23}$$

Thus,

$$\frac{2q_{jm}}{N} = 1 \pm \left[1 + \frac{1}{4B}\exp\left(-\frac{E_j}{kT}\right)\right]^{-\frac{1}{2}}, \tag{24}$$

where, as usual, β is determined to be $1/kT$ by the fact that in a constant volume process

$$dU = T\,ds. \tag{25}$$

Thus from Eq. (24), we see that P_j, the relative occupation of the jth array is given by

$$P_j = 1 - \{1 + \exp[(E_D - E_j)/kT]\}^{-1}, \tag{26}$$

where $E_D \equiv -kT\ln 4B$ and the minus sign is chosen because $P_j \leq 1$. At $T = 0$, this distribution function has the Fermi-like character shown in Fig. 3. Thus, if $E_j < E_D$, the partition function is unity. If $E_j > E_D$, the value of the partition function is zero. This distribution function also exhibits behavior similar to classical statistics when $(E_D - E_j)/kT \ll 0$, i.e.,

$$P_j \simeq \tfrac{1}{2}\exp[(E_D - E_j)/kT]. \tag{27}$$

VI. CONCLUSION

It is shown that an exact treatment of the occupation statistics of dumbbells on a linear array of N compartments leads, in the limit as N tends to infinity, to the same results obtained when the Bethe approximation is used.

A partition function is derived on the basis of an exact treatment of the occupation statistics of linear arrays of dumbbells.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to T. Casselman and F. N. Simon for many fruitful discussions, and to L. P. Levine for the numerical calculations necessary for this study.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Time-Dependent Green's Function for Electromagnetic Waves in Moving Conducting Media*

IOANNIS M. BESIERIS†‡ AND RALPH T. COMPTON, JR.
Case Institute of Technology, Cleveland, Ohio

(Received 15 March 1967)

This paper treats the problem of radiation from sources of arbitrary time dependence in a moving conducting medium. The medium is assumed to be homogeneous and isotropic, with permittivity ϵ, permeability μ, and conductivity σ, and to move with constant velocity $\mathbf{v}$ with respect to a given inertial reference frame. $\mathbf{v}$ may have any value up to the speed of light. It is shown how the Maxwell–Minkowski equations for the electromagnetic fields in the moving medium can be integrated by means of a pair of vector and scalar potential functions analogous to those commonly used with stationary media. The wave equation associated with these potential functions is derived, and an associated scalar Green's function is defined. The solution for the Green's function is obtained in closed form, by means of a technique making use of the relation between the fundamental solution of a radiation problem and that of a corresponding Cauchy initial-value problem. The resulting Green's function is found to consist of an oblate spheroidal shell, similar to that which occurs in a lossless medium, plus a residue which persists after the shell. In addition, the Green's function is exponentially damped in both space and time, an effect not present in a lossless medium.

I. INTRODUCTION

ALTHOUGH Minkowski's theory of the electrodynamics of moving media[1,2] was first introduced in 1908, it was not until recently that workers have begun to make much use of Minkowski's results to solve electromagnetic boundary value problems. Thus, for example, the problem of Čerenkov radiation from a uniformly moving point charge is usually treated as a moving charge in a stationary medium,[3] although the same problem viewed as a stationary charge in a moving medium is mathematically simpler, being an electrostatics problem in Minkowski's theory.[4–6] In a series of recent articles,[7–9] Tai has related Minkowski's results to other formulations of this subject, and also a number of papers on various specialized problems involving moving media have appeared.[10–22]

In two earlier papers,[23,24] the present authors determined the time-dependent Green's function for a moving lossless medium of arbitrary velocity up to c, the speed of light, and for a slowly moving ($v \ll c$) conducting medium. (The propagation of plane waves in a slowly moving lossy medium has also previously been discussed by Collier and Tai.[25]) The present paper discusses the time-dependent Green's function for a moving conducting medium where the velocity of the medium may have any value up to c. The medium is assumed to be homogeneous and isotropic, to be of infinite extent in all directions, and to be moving with a uniform velocity $\mathbf{v}$ with respect to a given inertial reference frame. The electrical properties of the medium are characterized by a permittivity ϵ, a permeability μ, and a conductivity σ, which are constant at all frequencies. (ϵ, μ, and σ are measured in a reference frame attached to the medium.)

II. THE MAXWELL–MINKOWSKI EQUATIONS FOR A CONDUCTING MEDIUM

We consider two inertial reference frames K and K' in relative motion. The primed coordinate system is at

* This work was supported by the U.S. Air Force Cambridge Research Laboratories, Office of Aerospace Research, under contract AF19(628)1699.

† This paper is based on part of I. M. Besieris' Ph.D. dissertation submitted to Case Institute of Technology, Cleveland, Ohio.

‡ Present address: Bell Telephone Laboratories, Inc., Whippany, New Jersey.

[1] H. Minkowski, Nachr. Kgl. Ges. Wiss. Göttingen **1**, 53 (1908).
[2] A. Sommerfeld, *Electrodynamics* (Academic Press Inc., New York, 1964), pp. 280–290.
[3] J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, Inc., New York, 1963), pp. 494–499.
[4] B. D. Nag and A. M. Sayied, Proc. Roy. Soc. (London) **A235**, 544 (1956).
[5] V. P. Pyati, Ph.D. Thesis, University of Michigan (1966).
[6] R. T. Compton, Jr., "On Čerenkov Radiation" (to be published).
[7] C. T. Tai, Proc. IEEE **52**, 685 (1964).
[8] C. T. Tai, Radiation Laboratory, University of Michigan, Report RL-310 (1965).
[9] C. T. Tai, IEEE Intern. Conv. Rec. Pt. 7, 148 (1966).
[10] K. S. H. Lee and C. H. Papas, J. Math. Phys. **5**, 1668 (1964).
[11] K. S. H. Lee and C. H. Papas, IEEE Trans. **AP-13**, 799 (1965).
[12] R. T. Compton, Jr., and C. T. Tai, IEEE Trans. **AP-13**, 574 (1965).
[13] C. T. Tai, IEEE Trans. **AP-13**, 322 (1965).
[14] C. T. Tai, Radio Sci. **69D**, 401 (1965).
[15] L. J. Chu, H. A. Haus, and P. Penfield, Jr., M.I.T. Research Laboratory of Electronics, Technical Rept. 433 (1965).
[16] L. J. Du, Ph.D. thesis, Ohio State University (1965).
[17] R. C. Costen and D. Adamson, Proc. IEEE **53**, 1181 (1965).
[18] J. R. Collier and C. T. Tai, IEEE Trans. **MTT-13**, 441 (1965).
[19] L. J. Du and R. T. Compton, Jr., IEEE Trans. **MTT-14**, 358 (1966).
[20] C. T. Tai, Radiation Laboratory, University of Michigan, Rept. RL-353 (1966).
[21] S. W. Lee and Y. T. Lo, Radio Sci. **1**, 313 (1966).
[22] I. M. Besieris, Ph.D. Thesis, Case Institute of Technology (1966).
[23] R. T. Compton, Jr., J. Math. Phys. **7**, 2145 (1966).
[24] I. M. Besieris, J. Math. Phys. **8**, 409 (1967).
[25] J. R. Collier and C. T. Tai, IEEE Trans. **AP-12**, 375 (1964).

rest with respect to a homogeneous, nondispersive, time-invariant, isotropic, conducting linear medium of infinite extent which moves with a uniform velocity $\mathbf{v}$ relative to the laboratory system K.

Although no change whatsoever is required in the well-known laws of stationary electrodynamics for an observer "riding" with the medium, a radically new formulation is necessary in the case of an observer at rest in the laboratory system. In this coordinate frame the electromagnetic fields must satisfy Maxwell's equations

$$\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t, \tag{1a}$$

$$\nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t + \mathbf{J}_f + \mathbf{J}, \tag{1b}$$

$$\nabla \cdot \mathbf{D} = \rho_f + \rho, \tag{1c}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{1d}$$

where $\mathbf{E}$, $\mathbf{H}$ are respectively the electric and magnetic field intensities; $\mathbf{D}$, $\mathbf{B}$ the electric and magnetic displacements; ρ_f, $\mathbf{J}_f$ the free charge and current densities and, finally, ρ, $\mathbf{J}$ the externally applied charge and current densities, all referred to the mks system of units.

Since all the fundamental laws of physics are covariant in the light of the theory of special relativity, Maxwell's equations must have the same form in all inertial frames of reference. This set of equations in the K' frame is not closed and, therefore, does not permit a solution; consequently it must be supplemented by auxiliary relations between the electric displacement, the magnetic displacement, the free-current density, and the electric and magnetic field intensities. It will be assumed that the following constitutive relations are specified in K':

$$\mathbf{D}' = \epsilon' \mathbf{E}', \tag{2a}$$

$$\mathbf{B}' = \mu' \mathbf{H}', \tag{2b}$$

$$\mathbf{J}_f' = \sigma' \mathbf{E}', \tag{2c}$$

where ϵ', μ', and σ' are the permittivity, the permeability and the conductivity, respectively. All these quantities are taken to be independent of the space coordinates and time. In other words, the medium under investigation is considered to be homogeneous and stationary.[26] In addition, it is specified that the medium is neither spatially nor temporally dispersive.

The approach enunciated by Minkowski in 1908 was to assume that the properties of the medium, which are expressed in terms of the constitutive relations in K', are known. The transformations of the theory of special relativity are then applied to obtain relations which are valid in the laboratory system relative to which the material is moving with uniform, but otherwise arbitrary, velocity.

According to Minkowski's theory, Eqs. (2a)–(2c) are sufficient to determine the constitutive expressions for the unprimed field quantities if the Lorentz relativistic transformations are known. If it is assumed that the coordinate frames K and K' are coincident at $t = t'$, have the same orientation, and move with a uniform velocity $\mathbf{v} = v\mathbf{a}_z$ with respect to each other, the following formulas link the primed and unprimed field quantities[27]:

$$\mathbf{E}' = \bar{\nu} \cdot (\mathbf{E} + \bar{v} \cdot \mathbf{B}), \tag{3a}$$

$$\mathbf{B}' = \bar{\nu} \cdot (\mathbf{B} - c^{-2}\bar{v} \cdot \mathbf{E}), \tag{3b}$$

$$\mathbf{D}' = \bar{\nu} \cdot (\mathbf{D} + c^{-2}\bar{v} \cdot \mathbf{H}), \tag{3c}$$

$$\mathbf{H}' = \bar{\nu} \cdot (\mathbf{H} - \bar{v} \cdot \mathbf{D}), \tag{3d}$$

where

$$\bar{\nu} = \gamma \bar{I}_t + \mathbf{a}_z\mathbf{a}_z, \quad \bar{I}_t = \mathbf{a}_x\mathbf{a}_x + \mathbf{a}_y\mathbf{a}_y,$$

$$\gamma = (1 - \beta^2)^{-\frac{1}{2}}, \quad \beta = v/c.$$

The antisymmetric tensor $\bar{v}$ is defined in such a way that $\bar{v} \cdot \mathbf{F} = \mathbf{v} \times \mathbf{F}$ for an arbitrary vector $\mathbf{F}$.

In addition to the relativistic transformations of the field intensities given above, the transformation laws of the 4-vector $(\mathbf{J}_f', ic\rho_f')$ assume the following form:

$$\mathbf{J}_f' = \gamma \bar{\nu}^{-1} \cdot (\mathbf{J}_f - \rho_f \mathbf{v}), \tag{3e}$$

$$\rho_f' = \gamma(\rho_f - c^{-2}\mathbf{v} \cdot \mathbf{J}_f). \tag{3f}$$

Even if $\rho_f' = 0$ in the rest frame of the medium, so that $\mathbf{J}_f'$ is a pure conduction current in this system, the charge density ρ_f in K will be different from zero, which means that a convection current is present.

In the ensuing discussion, it is assumed that ρ_f'—the free-charge-density distribution as measured by an observer moving with the material medium in the same direction and the same speed—is very small so that it can be neglected. In view of the brevity of the relaxation time, this does not constitute an essentially serious restriction.[28–30] If this condition is borne in

[26] The stationarity of the medium pertains to the time invariance of its characteristic properties (permittivity, permeability, etc.) and should not be confused with the motion of the material.

[27] For convenience, the velocity of the medium is taken to be in the z direction, i.e., $\mathbf{v} = v\mathbf{a}_z$. This condition does not detract from the generality of the problem since a coordinate transformation of the final solution can be used to treat the more general case.

[28] A self-consistent scheme for characterizing charges and currents by sources and response terms, which does not require the simplified assumption that the free charge is zero in the K' frame, is being presently investigated by Tai *et al.* (cf. Refs. 29 and 30).

[29] C. T. Tai, V. P. Pyati, and R. M. Kalafus, Radiation Laboratory, University of Michigan, Report 7322-I-F (1966).

[30] R. M. Kalafus and C. T. Tai, Union Radioscientifique Internationale Fall Meeting, Pt. 6, 96 (1966).

mind, by substituting Eqs. (3a)–(3f) into Eqs. (2a)–(2c) one obtains the constitutive relations

$$\mathbf{D} = \epsilon'\bar{\alpha}\cdot\mathbf{E} + \mathbf{\Omega}\times\mathbf{H}, \tag{4a}$$

$$\mathbf{B} = -\mathbf{\Omega}\times\mathbf{E} + \mu'\bar{\alpha}\cdot\mathbf{H}, \tag{4b}$$

$$\mathbf{J}_f = \bar{\gamma}\cdot\mathbf{E} + \bar{\delta}\cdot\mathbf{H}, \tag{4c}$$

$$\rho_f = c^{-2}\mathbf{v}\cdot\mathbf{J}_f, \tag{4d}$$

where the following definitions have been used:

$$\bar{\alpha} = a\bar{I}_t + \mathbf{a}_z\mathbf{a}_z, \quad a = (1-\beta^2)/(1-n^2\beta^2),$$
$$n = (\mu'\epsilon'/\mu_0\epsilon_0)^{\frac{1}{2}}, \quad \mathbf{\Omega} = [\beta(n^2-1)/c(1-n^2\beta^2)]\mathbf{a}_z,$$
$$\bar{\gamma} = \sigma'\gamma\bar{\alpha}, \quad \bar{\sigma} = \sigma'\gamma\mu' a\bar{v}.^{31}$$

Substituting Eqs. (4a)–(4d) into the initial set of Maxwell's equations in K frame, one obtains a definite form of the Maxwell–Minkowski equations which may be written in the following convenient fashion:

$$\mathbf{D}_0\times\mathbf{E} = -(\partial/\partial t)\mu'\bar{\alpha}\cdot\mathbf{H}, \tag{5a}$$

$$\mathbf{D}_0\times\mathbf{H} = (\partial/\partial t)\epsilon'\bar{\alpha}\cdot\mathbf{E} + \sigma'\gamma\bar{\alpha}\cdot\mathbf{E} + (1/a)\boldsymbol{\delta}\times(\bar{\alpha}\cdot\mathbf{H}) + \mathbf{J}, \tag{5b}$$

$$\mathbf{D}_0\cdot(\epsilon'\bar{\alpha}\cdot\mathbf{E}) = \epsilon'\boldsymbol{\delta}\cdot\bar{\alpha}\cdot\mathbf{E} + \rho + \mathbf{\Omega}\cdot\mathbf{J}, \tag{5c}$$

$$\mathbf{D}_0\cdot(\mu'\bar{\alpha}\cdot\mathbf{H}) = 0. \tag{5d}$$

$\mathbf{D}_0$ stands for the differential operator $\nabla - \mathbf{\Omega}(\partial/\partial t)$, and $\boldsymbol{\delta} = \sigma'\gamma\mu' a\mathbf{v}$.

III. SCALAR AND VECTOR POTENTIAL EQUATIONS

It will be shown in this section that the electromagnetic field intensities satisfying the Maxwell–Minkowski set are expressible in terms of appropriately defined scalar and vector potentials, in a manner analogous to that followed in stationary electromagnetic theory. The introduction of the potential functions is made possible by a generalization of the Helmholz theorem. The partial differential equations arising from this procedure are of the second order with respect to time, and are initially "coupled" in the sense that they contain both the scalar and vector potentials. The final conversion to the "uncoupled" scalar and vector inhomogeneous equations is achieved by means of a gauge transformation.

A vector potential $\mathbf{A}$ may be defined so that

$$\mu'\bar{\alpha}\cdot\mathbf{H} = \mathbf{D}_0\times\mathbf{A}. \tag{6}$$

This follows from Eq. (5d) in view of the identities

$$\mathbf{D}_0\cdot\mathbf{D}_0\times\mathbf{F} = 0, \tag{7a}$$

$$\mathbf{D}_0\times\mathbf{D}_0\Phi, \tag{7b}$$

which hold for all twice-differentiable scalar and vector functions Φ and $\mathbf{F}$, and permit, in turn, the formulation of a generalized Helmholz theorem. Equations (5a) and (6) suggest that

$$\mathbf{E} = -\partial\mathbf{A}/\partial t - \mathbf{D}_0\Psi, \tag{8}$$

where Ψ is a suitably chosen scalar potential function.

These expressions for $\mathbf{E}$ and $\mathbf{H}$ in terms of the scalar and vector potential functions are introduced first into the second equation of the Maxwell–Minkowski set to obtain

$$\begin{aligned}\mathbf{D}_0\times[\bar{\alpha}^{-1}\cdot(\mathbf{D}_0\times\mathbf{A})] = {}& -\epsilon'\mu'\bar{\alpha}\cdot(\partial^2\mathbf{A}/\partial t^2) - \epsilon'\mu'\bar{\alpha}\cdot(\partial/\partial t)\mathbf{D}_0\Psi \\ & -\mu'\sigma'\gamma\bar{\alpha}\cdot(\partial/\partial t)\mathbf{A} - \mu'\sigma'\gamma\bar{\alpha}\cdot\mathbf{D}_0\Psi \\ & + (1/a)\boldsymbol{\delta}\times(\mathbf{D}_0\times\mathbf{A}) + \mu'\mathbf{J}.\end{aligned} \tag{9}$$

A new vector function $\mathbf{A}_0$ and a new differential operator $\mathbf{D}_a$ are defined as follows:

$$\mathbf{A}_0 = \bar{\alpha}\cdot\mathbf{A}, \quad \mathbf{D}_a = (1/a)\bar{\alpha}\cdot\mathbf{D}_0. \tag{10}$$

If these definitions are taken into consideration, the following equation holds identically[13]:

$$\begin{aligned}\mathbf{D}_0\times[\bar{\alpha}^{-1}\cdot(\mathbf{D}_0\times\bar{\alpha}^{-1}\cdot\mathbf{A}_0)] &= (1/a)[\mathbf{D}_a(\mathbf{D}_0\cdot\mathbf{A}_0) - (\mathbf{D}_a\cdot\mathbf{D}_0)\mathbf{A}_0]. \\ &\end{aligned} \tag{11}$$

The validity of this identity can be checked by writing out both sides in a Cartesian coordinate system. By an extension of a standard vector identity, since $\boldsymbol{\delta}$ is a constant vector, it is seen that

$$\begin{aligned}\boldsymbol{\delta}\times(\mathbf{D}_0\times\mathbf{A}) &= \mathbf{D}_0(\boldsymbol{\delta}\cdot\mathbf{A}) - (\boldsymbol{\delta}\cdot\mathbf{D}_0)\mathbf{A} - (\mathbf{A}\cdot\mathbf{D}_0)\boldsymbol{\delta} \\ &\quad - \mathbf{A}\times(\mathbf{D}_0\times\boldsymbol{\delta}) \\ &= \mathbf{D}_0(\boldsymbol{\delta}\cdot\mathbf{A}) - (\boldsymbol{\delta}\cdot\mathbf{D}_0)\mathbf{A}.\end{aligned} \tag{12}$$

Upon substituting Eqs. (10)–(13) into Eq. (9) one has

$$\begin{aligned}(\mathbf{D}_a\cdot\mathbf{D}_0)\mathbf{A}_0 - \mu'\epsilon' a(\partial^2\mathbf{A}_0/\partial t^2) - \mu'\sigma'\gamma a(\partial\mathbf{A}_0/\partial t) & \\ - (\boldsymbol{\delta}\cdot\mathbf{D}_a)\mathbf{A}_0 & \\ = \mathbf{D}_a[\mathbf{D}_0\cdot\mathbf{A}_0 + \epsilon'\mu' a^2(\partial\Psi/\partial t) + \mu'\sigma'\gamma a^2\Psi & \\ - \boldsymbol{\delta}\cdot\mathbf{A}_0] - \mu' a\mathbf{J}. &\end{aligned} \tag{13}$$

In a similar manner it can be shown that (5c) and (8) imply the following equation:

$$(\mathbf{D}_0\cdot\mathbf{D}_a)\Psi - (\boldsymbol{\delta}\cdot\mathbf{D}_a)\Psi + \frac{1}{a}\frac{\partial}{\partial t}\mathbf{D}_0\cdot\mathbf{A}_0 - \mu'\sigma'\gamma\mathbf{v}\cdot\frac{\partial\mathbf{A}_0}{\partial t} = -\frac{\rho + \mathbf{\Omega}\cdot\mathbf{J}}{a\epsilon'} \tag{14}$$

Thus, the Maxwell–Minkowski equations have been reduced to the two partial differential equations given in (13) and (14). However, they are coupled. The uncoupling, as in the conventional case, can be accomplished by investigating the arbitrariness in defining the potentials. Making use of this freedom, a set of potentials Ψ and $\mathbf{A}_0$ will be selected so that they

[31] The notational definitions here have been chosen to conform with already existing ones (cf. Refs. 8 and 13).

satisfy the "extended" Lorentz condition

$$\mathbf{D}_0 \cdot \mathbf{A}_0 + \epsilon'\mu' a^2(\partial\Psi/\partial t) + \mu'\sigma'\gamma a^2\Psi - \boldsymbol{\delta} \cdot \mathbf{A}_0 = 0. \quad (15)$$

It should be stressed here that the modified Lorentz condition is not an arbitrary, subsidiary condition. On the contrary, besides effecting the desired uncoupling, it introduces complete symmetry between the scalar and vector potentials, i.e., it makes both of them satisfy the same inhomogeneous differential equation. More specifically,

$$[\mathbf{D}_a \cdot \mathbf{D}_0 - \epsilon'\mu' a(\partial^2/\partial t^2) - \mu'\sigma' a\gamma(\partial/\partial t) - \boldsymbol{\delta} \cdot \mathbf{D}_a]\Psi = -(\rho + \boldsymbol{\Omega} \cdot \mathbf{J})/a\epsilon', \quad (16)$$

$$[\mathbf{D}_a \cdot \mathbf{D}_0 - \epsilon'\mu' a(\partial^2/\partial^2) - \mu'\sigma' a\gamma(\partial/\partial t) - \boldsymbol{\delta} \cdot \mathbf{D}_a]\, A_0 = -\mu' a\mathbf{J}.$$

To solve for these potentials, we define the time-dependent Green's function $G(\mathbf{r}, t/\mathbf{r}', t')$ as the solution of the equation

$$\left[\mathbf{D}_a \cdot \mathbf{D}_0 - \epsilon'\mu' a\frac{\partial^2}{\partial t^2} - \mu'\sigma' a\gamma\frac{\partial}{\partial t} - \boldsymbol{\delta} \cdot \mathbf{D}_a\right] G(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') \quad (17)$$

subject to the causality condition, viz., $G(\mathbf{r}, t/\mathbf{r}', t')$ satisfies Eq. (17) for $t \geq t'$ but vanishes identically for $t < t'$. The electromagnetic potentials are given in terms of the Green's function by integral expressions over the source distributions.

IV. THE EXPLICIT SOLUTION FOR THE TIME-DEPENDENT GREEN'S FUNCTION

Instead of using a combination of fourfold space–time Fourier transformation and residue theory to determine the time-dependent Green's function satisfying Eq. (17), use will be made of the relation between the fundamental solution of a radiation problem and that of a corresponding Cauchy initial-value problem. This alternative technique has been discussed in detail and exemplified in Ref. (24).

Consider the Cauchy initial-value problem

$$\left[\nabla_t^2 + \frac{1}{a}\frac{\partial^2}{\partial z^2} - 2\frac{\Omega}{a}\frac{\partial^2}{\partial z \partial t} - \left(\epsilon'\mu' - \frac{\Omega^2}{a}\right)\frac{\partial^2}{\partial t^2} + (v\Omega - \gamma a)\mu'\sigma'\frac{\partial}{\partial t} - \mu'\sigma' v\frac{\partial}{\partial z}\right]\Psi(\mathbf{r}, t) = 0, \quad (18a)$$

$$\Psi(\mathbf{r}, t)\big|_{t=t'} = 0, \quad (18b)$$

$$(\partial/\partial t)\Psi(\mathbf{r}, t)\big|_{t=t'} = g(\mathbf{r}, t'). \quad (18c)$$

Its solution can be simplified by the substitution

$$\Phi(\mathbf{r}, t) = \Psi(\mathbf{r}, t)\exp(pz + qt). \quad (19)$$

With the scalars p, q appropriately specified, the initial-value problem reduces to the "semicanonical" form[32,33]

$$\left[\nabla_t^2 + \frac{1}{a}\frac{\partial^2}{\partial z^2} - 2\frac{\Omega}{\alpha}\frac{\partial^2}{\partial z \partial t} - \left(\epsilon'\mu' a - \frac{\Omega^2}{a}\right)\frac{\partial^2}{\partial t^2} + l^2\right]\Phi(\mathbf{r}, t) = 0. \quad (20a)$$

The correct choice for p and q is

$$p = -(\sigma'/2\epsilon')(\mu'\epsilon' va - \gamma\Omega), \quad q = \sigma'\gamma/2\epsilon'$$

for which

$$l^2 = -p^2/a + q^2(\epsilon'\mu' a - \Omega^2/a) + 2pq(\Omega/a).$$

The initial conditions must be modified as follows:

$$\Phi(\mathbf{r}, t)\big|_{t=t'} = 0, \quad (20b)$$

$$(\partial/\partial t)\Phi(\mathbf{r}, t)\big|_{t=t'} = g(\mathbf{r}, t')\exp(pz + qt') \equiv h(\mathbf{r}, t'). \quad (20c)$$

We shall next convert Eq. (20a) into a system of two partial differential equations of the first order with respect to time. Towards this goal we define

$$u_1 = \Phi, \quad u_2 = \partial\Phi/\partial t, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \quad (21)$$

We then write

$$(\partial/\partial t)\mathbf{u}(\mathbf{r}, t) = \mathfrak{P}\mathbf{u}(\mathbf{r}, t) \quad (22)$$

where

$$\mathfrak{P} = \begin{bmatrix} 0 & 1 \\ \kappa^{-1}(\nabla_t^2 + a^{-1}\partial^2/\partial z^2) + \kappa^{-1}l^2 - \kappa^{-1}(2\Omega/a^{-1}\partial/\partial z & \end{bmatrix}$$

$$\kappa = (\epsilon'\mu' a - \Omega^2 a^{-1}).$$

Operating next with a threefold spatial Fourier transform, we find the relation

$$(\partial/\partial t)\mathbf{u}(\mathbf{s}, t) = \mathfrak{P}(\mathbf{s})\mathbf{u}(\mathbf{s}, t) \quad (23)$$

in which

$$\mathfrak{P}(\mathbf{s}) = \begin{bmatrix} 0 & 1 \\ \kappa^{-1}(-s_t^2 - s_z^2/a) + \kappa^{-1}l^2 & -\kappa^{-1}(2\Omega/a)is_z \end{bmatrix},$$

$$s_t^2 = s_x^2 + s_y^2.$$

The solution to Eq. (22) is now given by[34]

$$\mathbf{u}(\mathbf{r}, t) = \int_{E_3} \mathfrak{R}'(\mathbf{r}, t/\mathbf{r}', t')\mathbf{u}_0(\mathbf{r}', t')\, d\mathbf{r}', \quad (24)$$

[32] This second-order partial differential equation would be in the "canonical" form if the term involving the mixed derivatives with respect to z and t were absent (cf. Ref. 33).

[33] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, New York, 1962), Vol. II, pp. 180–184.

[34] The integration here extends over a three-dimensional Euclidean space containing the space coordinates.

where $\mathbf{u}_0(\mathbf{r}, t')$ is the value of $\mathbf{u}(\mathbf{r}, t)$ at $t = t'$, and the matrix function $\mathfrak{R}'(\mathbf{r}, t/\mathbf{r}', t')$, the *Riemann matrix*, is given by

$$\mathfrak{R}'(\mathbf{r}, t/\mathbf{r}', t') = \frac{1}{(2\pi)^3}\int_{E_3} \exp\,[i(\mathbf{r}-\mathbf{r}')\cdot\mathbf{s}\mathfrak{I}] \times \sum_{k=1}^{2} e^{\lambda_k t}\left[\prod_{\substack{j=1\\ j\neq k}}^{2} \frac{\mathfrak{P}(\mathbf{s})(t-t') - \lambda_j\mathfrak{I}}{\lambda_k - \lambda_j}\right] d\mathbf{s}. \quad (25)$$

The solution of the characteristic equation

$$\text{Det}\,[\mathfrak{P}(\mathbf{s})t - \lambda\mathfrak{I}] = 0 \quad (26)$$

yields the two eigenvalues

$$\lambda_{1,2} = (-\omega_0 \pm \omega_1)t, \quad (27)$$

where

$$\omega_0 = \kappa^{-1}(\Omega/a)s_z,$$
$$\omega_1 = \kappa^{-\frac{1}{2}}[s_t^2 + (1/a + \kappa^{-1}\Omega^2/a^2)s_z^2 - l^2]^{\frac{1}{2}}.$$

By virtue of the definitions in Eq. (21) and the specified initial conditions [cf. Eqs. (20b) and (20c)], we need only investigate the (12) entry of $\mathfrak{R}'$, viz.,

$$(\mathfrak{R}')_{12} = \frac{1}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}\cdot\mathbf{R}} \times \left[e^{-i(\Omega/a)u^2 s_z\tau}\,\frac{\sin u\tau(s_t^2 + s_z^2/b^2 - l^2)^{\frac{1}{2}}}{u(s_t^2 + s_z^2/b^2 - l^2)^{\frac{1}{2}}}\right] d\mathbf{s}. \quad (28)$$

The following abbreviations have been used:

$$\mathbf{R} = \mathbf{r} - \mathbf{r}', \quad \tau = t - t', \quad u = \kappa^{-\frac{1}{2}},$$
$$b = (1/a + \kappa^{-1}\Omega^2/a^2)^{-\frac{1}{2}}.$$

Let $\mathbf{s}$ and $\mathbf{R}$ be subjected to the following linear transformation:

$$\mathbf{s}_0 = \mathfrak{A}\mathbf{s}, \quad (29a)$$

$$\mathfrak{R}_0 = \mathfrak{A}^{-1}\mathfrak{R}, \quad (29b)$$

with

$$\mathfrak{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/b \end{bmatrix}, \quad \mathbf{s}_0 = \sum_{j=1}^{3} s_j\mathbf{a}_j, \quad \mathbf{R}_0 = \sum_{j=1}^{3} x_j\mathbf{a}_j.$$

With this transformation it develops that

$$(\mathfrak{R}')_{12} = \frac{b}{(2\pi)^3}\int_{E_3} e^{i\mathbf{s}_0\cdot\mathbf{R}_1}\,\frac{\sin u\tau(s_0^2 - l^2)^{\frac{1}{2}}}{u(s_0^2 - l^2)^{\frac{1}{2}}}\,d\mathbf{s}_0, \quad (30)$$

in which

$$\mathbf{R}_1 = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + (x_3 - \gamma'\tau)\mathbf{a}_3, \quad \gamma' = (\Omega/a)bu^2.$$

Choosing next a spherical coordinate system with its polar axis along the $\mathbf{R}_1$ direction and proceeding exactly as in Ref. 24, one establishes that

$$(\mathfrak{R}')_{12} = \frac{b}{4\pi u}\frac{1}{R_1}\,\delta(R_1 - u\tau) + \frac{b}{4\pi}\frac{1}{(R_1^2 + u^2\tau^2)^{\frac{1}{2}}}\,J_1[l(R_1^2 - u^2\tau^2)^{\frac{1}{2}}]\mathbf{1}_+(R_1 - u\tau) \quad (31)$$

where $\mathbf{1}_+$ designates the Heaviside unit step function, and

$$(\mathfrak{R})_{12} = \exp\,(-p\zeta - q\tau)(\mathfrak{R}')_{12}, \quad \zeta = z - z'.^{35} \quad (32)$$

Finally,

$$G(\mathbf{r}, t/\mathbf{r}', t') = 0 \quad (33a)$$

for $t < t'$, and

$$G(\mathbf{r}, t/\mathbf{r}', t') = \kappa^{-1}(\mathbf{R})_{12} = \kappa^{-1}\exp\,(-p\zeta - q\tau)\left\{\frac{b}{4\pi u}\frac{1}{R_1}\,\delta(R_1 - u\tau) - \frac{b}{4\pi}\frac{1}{(R_1^2 - u^2\tau^2)^{\frac{1}{2}}}\,J_1[l(R_1^2 - u^2\tau^2)^{\frac{1}{2}}]1_+(R_1 - u\tau)\right\} \quad (33b)$$

for $t \geq t'$. In this equation,

$$R_1 = [(x - x')^2 + (y - y')^2 + b^2(\zeta - \gamma'\tau/b)^2]^{\frac{1}{2}}. \quad (34)$$

Before an attempt is made to interpret the solution in Eq. (33), we first direct our attention to the following interesting special cases:

Case I. If $v \neq 0$ and $\sigma' = 0$, the generalized function

$$G(\mathbf{r}, t/\mathbf{r}', t') = \frac{b}{4\kappa\pi u}\frac{1}{R_1}\,\delta(R_1 - u\tau) \quad (35)$$

satisfies the equation

$$\left[\nabla_t^2 + \frac{1}{a}\frac{\partial^2}{\partial z^2} - 2\frac{\Omega}{a}\frac{\partial^2}{\partial z\partial t} + \left(\frac{\Omega^2}{a} - \epsilon'\mu' a\right)\frac{\partial^2}{\partial t^2}\right]G(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') \quad (36)$$

for $t \geq t'$. Since

$$u = v_0\left[\frac{1 - \beta^2}{1 - (\beta/n)^2}\right]^{\frac{1}{2}}, \quad v_0 = (\mu'\epsilon')^{-\frac{1}{2}},^{36}$$
$$b = \left[\frac{n^2 - \beta^2}{n^2(1 - \beta^2)}\right]^{\frac{1}{2}}, \quad \frac{\gamma'}{b} = \frac{v(n^2 - 1)}{n^2 - \beta^2},$$
$$\kappa = \frac{n^2 - \beta^2}{c^2(1 - \beta^2)},$$

[35] $\mathfrak{R}$ and $\mathfrak{R}'$ are used in connection with the scalar functions $\Psi(\mathbf{r}, t)$ and $\Phi(\mathbf{r}, t)$, respectively.

[36] v_0 is the phase velocity of a wave in the medium, as seen by an observer in the rest frame of the medium.

it can be shown that the Green's function can be rewritten as

$$G(\mathbf{r}, t/\mathbf{r}', t') = \frac{1}{4\pi}\left[\frac{1-(\beta/n)^2}{1-\beta^2}\right]^{\frac{1}{2}}\frac{1}{R_1}\,\delta\left\{\tau - \frac{R_1}{v_0}\left[\frac{1-(\beta/n)^2}{1-\beta^2}\right]^{\frac{1}{2}}\right\} \quad (37)$$

with

$$R_1 = \left[(x-x')^2 + (y-y')^2 + \frac{n^2-\beta^2}{n^2(1-\beta^2)}\left(\zeta - \frac{n^2-1}{n^2-\beta^2}\,v\tau\right)^2\right]^{\frac{1}{2}}, \quad (38)$$

which is the solution obtained previously by Compton (cf. Ref. 23).

Case II. If $\sigma' \neq 0$, but we are considering the nonrelativistic limit, i.e., the case where v/c and $v/v_0 \ll 1$, the generalized function given by (33) with

$$p = -\tfrac{1}{2}(\sigma'\mu' v) + \mu'\sigma\Lambda/2\epsilon', \quad q = \sigma/2\epsilon',$$
$$\Lambda = (\mu'\epsilon' - \mu_0\epsilon_0)v,$$
$$b = (1 + \Lambda^2/\mu'\epsilon')^{-\frac{1}{2}}, \quad u = v_0, \quad \kappa = v_0^{-2}, \quad \gamma' = u^2\Lambda,$$

and $l^2 = \mu'\epsilon' q^2 + 2\Lambda pq - p^2$ is the solution to the problem

$$\left(\nabla^2 - \epsilon'\mu'\frac{\partial^2}{\partial t^2} - \mu'\sigma'\frac{\partial}{\partial t} - 2\Lambda\frac{\partial^2}{\partial z\partial t} - \mu'\sigma'\frac{\partial}{\partial z}\right) \times G(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \quad (39)$$

for $t \geq t'$, which has been investigated previously by Besieris (cf. Ref. 24).

Case III. Lastly, if $v = 0$ (or $\mu\epsilon = \mu_0\epsilon_0$) and $\sigma = 0$,

$$G(\mathbf{r}, t/\mathbf{r}', t') = (v_0/4\pi R)\delta(R - v_0\tau) \quad (40)$$

is the well-known solution of the simple wave equation

$$\left(\nabla^2 - \frac{1}{v_0^2}\frac{\partial^2}{\partial t^2}\right)G(\mathbf{r}, t/\mathbf{r}', t') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t'), \quad t \geq t'. \quad (41)$$

V. INTERPRETATION OF THE SOLUTION

The part of the Green's function $G(\mathbf{r}, t/\mathbf{r}', t')$ [cf. Eq. (33)] containing the Dirac delta function can be interpreted as an expanding wavefront which arrives at $R_1 = u\tau$ diminished by the geometrical factor $1/R_1$ and modified by the exponential term $\exp(-p\zeta - q\tau)$. R_1, as given in Eq. (34), can be taken as the "radial distance" between the point $(x', y', z' + \gamma'\tau/b)$ and the observation point $\mathbf{r}(x, y, z)$, with a scaling of the z-axis dimension by the factor b due to the Lorentz contraction along this axis.

Whereas for a "stationary" medium ($v = 0$ or $\mu'\epsilon' = \mu_0\epsilon_0$) the expanding wavefronts are spheres centered at the spatial position of the source, in the more general problem under consideration here, apart from the multiplicative factors $1/R_1$ and $\exp(-p\zeta - q\tau)$, the wavefronts obey the equation

$$R_1 = u\tau. \quad (42)$$

It is quite easy to show that, for constant values of τ, the wavefronts are actually spheroidal surfaces with semiaxes $u\tau$, $u\tau$, and $u\tau/b$ along the directions of the x, y, and z axes, respectively. Before any more remarks are made regarding the nature of these wavefronts, it should be noted that the wavefront center $(x', y', z' + \gamma'\tau/b)$ moves along the z direction at the speed γ'/b. Since, however,

$$\gamma'/b = v(n^2-1)/(n^2-\beta^2) < v,$$

the center of the spheroid moves slowly, and one might say that it "cannot keep up with the medium."

Two special cases are considered next, by assigning specific ranges to the velocity of the medium:

Case I. When the medium moves slowly enough so that $n\beta = v/v_0 < 1$, since $b > 1$, it follows that $u\tau/b < u\tau$ which implies, in turn, that the expanding shells are oblate spheroids with respect to the z axis. Consider next the ratio of the distances associated, first with the semiaxis along the z direction and secondly the position of the wavefront center relative to the source point $\mathbf{r}'$, viz.,

$$\frac{u\tau}{b}\Big/\frac{\gamma'\tau}{b} = \frac{u}{\gamma'} = \frac{v_0}{v}\,\frac{n^2-(v/v_0)^2}{n^2-1}.$$

Since by assumption $v/v_0 < 1$, this ratio is greater than unity. This shows clearly that the shell encloses the source point; that is, the source radiates in all directions.[37]

The effect of a pulse at a distance R_1 and at a time τ after its onset vanishes for $R_1 > u\tau$, that is, as long as the wave initiated by the pulse has not had sufficient time to reach the observation point $\mathbf{r}$.

Case II. If now the material moves with high enough speed so that $v/v_0 > 1$, the entire shell—which is still an oblate spheroid with respect to the z axis—is "dragged" away from the source point. The source point is outside the shell, and the shell is, at all times, tangent to a cone of interior half-angle θ_0 specified by the relation

$$\cos\theta_0 = [(n^2\beta^2 - 1)/\beta^2(n^2-1)]^{\frac{1}{2}}.$$

[37] This is always the case in the nonrelativistic limit.

An observer positioned outside the cone will not experience any radiation effects. However, someone located inside the conical region will detect two discontinuities, caused first by one side of the expanding and moving wavefront and then the other. This effect is the well-known *Čerenkov phenomenon.*

In either Case I and Case II, at $R_1 = u\tau$, the original pulse arrives diminished by the geometrical factor $1/R_1$. The wave then leaves in its wake a residue, or "tail," which persists for an infinite time at points which have been traversed by the wavefronts. This contribution is represented by the second part of $G(\mathbf{r}, t/\mathbf{r}', t')$ in Eq. (33). The entire solution is, of course, attenuated exponentially in the z direction Moreover, it subsides exponentially with respect to time.

ACKNOWLEDGMENT

The authors gratefully acknowledge the advice, guidance and continued interest of Professor R. E. Collin.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Infinite Systems of Classical Particles*

AMNON KATZ†
Argonne National Laboratory, Argonne, Illinois

(Received 13 March 1967)

In this paper, infinite translation-invariant and periodic systems of classical particles are treated directly—not as limits of finite systems. A formalism of classical creation and annihilation operators that create and destroy classical identical particles at points of phase space is employed. By use of this formalism, classical dynamical variables are expressed without reference to the canonical coordinates and momenta of individual particles, or to the number of particles. Different physical situations are described by different representations of the algebra of creation and annihilation operators. The concept of thermal equilibrium is generalized so as to be meaningful in infinite systems. Stationary states and thermal-equilibrium states for an infinite system of noninteracting particles (possibly in a periodic external potential) are exhibited explicitly.

1. INTRODUCTION

WITH the possible exception of the "true vacuum" of the theories of elementary particles, all systems of physical interest are finite. Still, infinite systems have always held great attraction for the theorist. The simplifying features that go with translation invariance apply only in the infinite case. Also, phase transitions and broken symmetries, although apparent to the experimenter in finite samples, become truly qualitative phenomena only in infinite systems.

The traditional approach to infinite systems is to regard them as limits of finite systems as the volume and number of particles tend to infinity with the density remaining constant. One speaks of "quasi-averages" [1] which are limits of expectation values in finite systems.

The alternative of directly treating infinite systems has received increasing attention in recent years.[2–6] Although global dynamical variables such as the total number of particles or the total energy become infinite (and hence meaningless) in infinite systems, such local dynamical variables as the density of particles or of energy remain meaningful. In the quantum theory of relativistic fields, a great deal of attention has been paid to "algebras of local observables." [7] A *state* of the system is then defined as a positive, normalized (continuous) linear functional on the algebra; in other words, a *state* of the system is the set of expectation values of all dynamical variables.[8] States corresponding to infinite translation-invariant (or periodic)

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

† On leave from the Weizmann Institute, Rehovoth, Israel. Present address: Department of Physics, University of Washington, Seattle, Washington.

[1] N. N. Bogoliubov, Physica Suppl. **26**, S1 (1960).

[2] D. W. Robinson, Commun. Math. Phys. **1**, 159 (1965).

[3] G. F. Dell'Antonio, S. Doplicher, and D. Ruelle, Commun. Math. Phys. **2**, 223 (1966).

[4] D. Ruelle, Commun. Math. Phys. **3**, 133 (1966).

[5] D. Kastler and D. W. Robinson, Commun. Math. Phys. **3**, 151 (1966).

[6] R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).

[7] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964), Secs. I and II.

[8] The concept goes back to J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1965)—although he did not call it a "state."

An observer positioned outside the cone will not experience any radiation effects. However, someone located inside the conical region will detect two discontinuities, caused first by one side of the expanding and moving wavefront and then the other. This effect is the well-known *Čerenkov phenomenon.*

In either Case I and Case II, at $R_1 = u\tau$, the original pulse arrives diminished by the geometrical factor $1/R_1$. The wave then leaves in its wake a residue, or "tail," which persists for an infinite time at points which have been traversed by the wavefronts. This contribution is represented by the second part of $G(\mathbf{r}, t/\mathbf{r}', t')$ in Eq. (33). The entire solution is, of course, attenuated exponentially in the z direction Moreover, it subsides exponentially with respect to time.

ACKNOWLEDGMENT

The authors gratefully acknowledge the advice, guidance and continued interest of Professor R. E. Collin.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Infinite Systems of Classical Particles*

AMNON KATZ†
Argonne National Laboratory, Argonne, Illinois

(Received 13 March 1967)

In this paper, infinite translation-invariant and periodic systems of classical particles are treated directly—not as limits of finite systems. A formalism of classical creation and annihilation operators that create and destroy classical identical particles at points of phase space is employed. By use of this formalism, classical dynamical variables are expressed without reference to the canonical coordinates and momenta of individual particles, or to the number of particles. Different physical situations are described by different representations of the algebra of creation and annihilation operators. The concept of thermal equilibrium is generalized so as to be meaningful in infinite systems. Stationary states and thermal-equilibrium states for an infinite system of noninteracting particles (possibly in a periodic external potential) are exhibited explicitly.

1. INTRODUCTION

WITH the possible exception of the "true vacuum" of the theories of elementary particles, all systems of physical interest are finite. Still, infinite systems have always held great attraction for the theorist. The simplifying features that go with translation invariance apply only in the infinite case. Also, phase transitions and broken symmetries, although apparent to the experimenter in finite samples, become truly qualitative phenomena only in infinite systems.

The traditional approach to infinite systems is to regard them as limits of finite systems as the volume and number of particles tend to infinity with the density remaining constant. One speaks of "quasi-averages" [1] which are limits of expectation values in finite systems.

The alternative of directly treating infinite systems has received increasing attention in recent years.[2–6] Although global dynamical variables such as the total number of particles or the total energy become infinite (and hence meaningless) in infinite systems, such local dynamical variables as the density of particles or of energy remain meaningful. In the quantum theory of relativistic fields, a great deal of attention has been paid to "algebras of local observables." [7] A *state* of the system is then defined as a positive, normalized (continuous) linear functional on the algebra; in other words, a *state* of the system is the set of expectation values of all dynamical variables.[8] States corresponding to infinite translation-invariant (or periodic)

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

† On leave from the Weizmann Institute, Rehovoth, Israel. Present address: Department of Physics, University of Washington, Seattle, Washington.

[1] N. N. Bogoliubov, Physica Suppl. **26**, S1 (1960).

[2] D. W. Robinson, Commun. Math. Phys. **1**, 159 (1965).

[3] G. F. Dell'Antonio, S. Doplicher, and D. Ruelle, Commun. Math. Phys. **2**, 223 (1966).

[4] D. Ruelle, Commun. Math. Phys. **3**, 133 (1966).

[5] D. Kastler and D. W. Robinson, Commun. Math. Phys. **3**, 151 (1966).

[6] R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).

[7] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964), Secs. I and II.

[8] The concept goes back to J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1965)—although he did not call it a "state."

situations with a finite average density of particles exist.[2–5,9,10] Every state can be implemented by a procedure known as the Gel'fand construction[11] in a suitable separable Hilbert space which is a representation space of the algebra of local observables. There exist many inequivalent representations.[11] The choice of a state implies the choice of a representation in which it can be implemented. The choice of representation therefore embodies a large part of the physics.

A rigorous evaluation of quasi-averages as limits ought to be quite a complicated endeavor. It should involve the dynamical description of the boundary of the system for every finite volume. The direct treatment of infinite systems avoids this as well as the limiting procedure. It remains, of course, to be shown that the two approaches are equivalent. It should be remarked, however, that people who claim to be calculating quasi-averages usually avoid a detailed description of the boundary and take various short cuts—some of which amount to a direct calculation in the infinite case. The calculation of Martin and Schwinger,[12] who made use of thermal Green's functions, is one such calculation in an infinite system. The only appeal to finite systems is for the derivation of the "boundary condition" $\langle AB\rangle_\beta = \langle B(-i\hbar\beta)A\rangle_\beta$, which is then applied in the infinite case. The "boundary condition" is in effect used as the *definition* of thermal equilibrium in the infinite system, for which the construction involving $\exp(-\beta H)$ no longer applies. In fact, some later authors[6] accept the boundary condition of Martin and Schwinger as the definition of thermal equilibrium in an infinite system. (Other authors,[9,10] conversely, *define* the state of thermal equilibrium by the limit of expectation values in a finite system and later endeavor to implement the state so defined.)

The quantum mechanical treatment of infinite systems is made possible by the "second-quantized" formalism which expresses dynamical variables without reference to individual particles or to their number. In spite of its name, "second quantization" (which does not involve Planck's constant $\hbar$) has nothing to do with physical quantization. The creation and annihilation operators merely count particles occupying first-quantized single-particle states. A similar formalism for classical identical particles has been put forward by this author.[13] It involves operators that create and destroy particles at various points in the phase space of a single particle. This formalism makes it possible to treat infinite systems of classical particles by considering suitable representations of the classical creation and annihilation operators. This is the subject of the present paper.

2. SCOPE AND PLAN OF PRESENTATION

The next two sections review the formalism describing a system of an indefinite (finite) number of particles and the formalism of classical creation and annihilation operators. The space in which a state of an indefinite number of identical particles can be represented is discussed in Sec. 3. This is the classical analog of the quantum mechanical Fock space. In Sec. 4 it serves as the defining space for the algebra of creation and annihilation operators. Derivations are omitted in Secs. 3 and 4. They all follow directly from the definitions; some of them may be found in Ref. 13. Section 5 introduces the algebra of local dynamical variables. Section 6 offers an example of a class of representations that can describe translation-invariant or periodic situations with finite average density of particles—in other words, nontrivial infinite systems. For noninteracting particles possibly in an external potential, this class of representations allows us to represent explicitly stationary states in Sec. 7 and thermal-equilibrium states in Sec. 9. General properties and characterization of stationary states and of thermal equilibrium (for particles with interaction) are discussed in Secs. 7 and 8, respectively. In the definition of thermal equilibrium put forward in Sec. 8, the construction involving $\exp(-\beta H)$ is generalized to infinite systems. This definition may be considered the classical analog of the "boundary condition" of Martin and Schwinger. Since the purpose of the present paper is merely to present and illustrate the formalism, only problems of noninteracting particles have been worked out.

3. GRAND CANONICAL APPROACH

This section briefly summarizes the grand canonical treatment of identical classical particles.[13,14] In this treatment the state of the system is represented by a "statistical element" f which is a sequence

$$f = (f_0, f_1, f_2, \cdots, f_N, \cdots), \tag{3.1}$$

where f_N is a nonnegative measure on the phase space of N classical particles and is symmetric in all particles.[15] The phase space in question is the Euclidean space of all coordinates and canonically conjugate

[9] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).
[10] H. Araki and W. Wyss, Helv. Phys. Acta **37**, 136 (1964).
[11] I. E. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, Rhode Island, 1963). Further references may be found in this book.
[12] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).
[13] A. Katz, *Principles of Statistical Mechanics (The Information Theory Approach)* (W. H. Freeman and Company, San Francisco, 1967), Appendix B and Sec. 3 of Chap. VII.
[14] J. M. Cook, Institut D'Études Scientifiques de Cargèse, Cargèse, France, lecture notes, 1965.
[15] In the special case of the grand canonical ensemble, $f_N = \exp(\Omega - \beta H_N - \alpha N)$, where H_N is the N-particle Hamiltonian.

momenta of N particles: $\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N$. The measure density $f_N(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N)$ must be unchanged by any permutation of the particle indices, i.e.,

$$f_N(\mathbf{q}_{a_1}, \mathbf{p}_{a_1}, \cdots, \mathbf{q}_{a_N}, \mathbf{p}_{a_N}) = f_N(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N) \quad (3.2)$$

for all permutations $n \to a_n$. The total weight of the entire N-particle phase space is

$$P_N \equiv \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_N\, d\mathbf{p}_N \left(\frac{1}{N!}\right) f_N, \quad (3.3)$$

where f_0 is a non-negative constant $P_0 = f_0$, and P_N is interpreted as the probability that the number of particles in the system is N. Consequently, we demand that statistical elements representing physical states should satisfy the relation

$$\sum_{N=0}^{\infty} P_N = 1. \quad (3.4)$$

Equations (3.3) and (3.4) constitute a normalization condition on f. In the following we shall, in addition, consider unnormalized and not necessarily positive objects $f = (f_0, f_1, f_2, \cdots)$, such that

$$\|f\| \equiv \sum_{N=0}^{\infty} \frac{1}{N!} \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_N\, d\mathbf{p}_N\, |f_N|$$

is finite but not necessarily equal to 1. Here f_N is a measure over the N-particle phase space. The form $\|f\|$ may serve as a norm in the space of objects f (state representatives), which becomes a normed linear space. We also define

$$\mathrm{Tr}\, f \equiv \sum_{N=0}^{\infty} \frac{1}{N!} \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_N\, d\mathbf{p}_N f_N. \quad (3.5)$$

Thus, $\mathrm{Tr}\, f$ is a linear form on the space of objects f. For the representatives of physical states, one has $\mathrm{Tr}\, f = \|f\| = 1$. The form $\mathrm{Tr}\, f$ is useful in expressing expectation values of dynamical variables.

The weight accorded to a subset of the phase space of N particles by the measure density $(1/N!)f_N$ belonging to a physical (positive and normalized) f together with all the subsets resulting from it by permutations of the N particles is interpreted as the probability that there are N (identical) particles in the system and that they occupy the positions and momenta in the subset.[16] Both pure and mixed states may be represented in this way. Pure states are represented by all f_N vanishing save one, which is a symmetrized delta function.

[16] As long as the measure densities are integrable functions, the $N!$ subsets obtained from each other by permutations of the N particles intersect each other on sets of measure zero; counting them all is the same as counting just one and multiplying by $N!$. One may then integrate on one subset only and drop the factor $(1/N!)$.

Dynamical variables are also represented as sequences

$$F = (F_0, F_1, F_2, \cdots, F_N, \cdots). \quad (3.6)$$

This time F_N is a continuous function on the phase space of N particles which is symmetric under permutations of the particles. The expectation value of the dynamical quantity F in the state f is

$$(f.F) \equiv \sum_{N=0}^{\infty} \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_N\, d\mathbf{p}_N \left(\frac{1}{N!}\right) f_N F_N. \quad (3.7)$$

For arbitrary f, the existence of $(f.F)$ can be guaranteed only for those functions F that are bounded i.e., only if there exists a bound A such that $|F_N| < A$ for all N and all $\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N$.[17] The space of bounded dynamical variables may be taken as a Banach space with the lowest bound for norm. The space of all representatives f of states (not necessarily normalized to 1) is then contained in its strong dual—the space of all continuous linear forms on the space of bounded dynamical variables. In this way we see how our f connect to the abstract concept of a state as a continuous linear form.

Most of the dynamical variables of interest are not bounded. An unbounded F may be connected to the bounded dynamical variable

$$\exp(iF) \equiv (\exp\{iF_0\}, \exp\{iF_1\}, \cdots). \quad (3.8)$$

Dynamical variables may depend on parameters as functions or as generalized functions.[18]

We define the product of a statistical element f by a dynamical variable F to be

$$Ff = (F_0 f_0, F_1 f_1, F_2 f_2, \cdots, F_N f_N, \cdots), \quad (3.9)$$

which is a new statistical element. In this way the dynamical variables appear as (an Abelian algebra of) linear operators on the space of statistical elements. The bounded dynamical variables become bounded (and therefore continuous) operators. The expectation value of F in the state represented by f may now be written as

$$(f.F) = \mathrm{Tr}\, Ff. \quad (3.10)$$

A dynamical variable is considered to be a *single-particle variable* if it is of the form

$$F^1 = \left(0, g^1(\mathbf{q}_1, \mathbf{p}_1), g^1(\mathbf{q}_1, \mathbf{p}_1) + g^1(\mathbf{q}_2, \mathbf{p}_2), \cdots, \sum_{a=1}^{N} g^1(\mathbf{q}_a, \mathbf{p}_a), \cdots\right). \quad (3.11)$$

[17] Here, as in several other places, an attempt is made to classify the mathematical structures involved. This is done for the purpose of future applications to problems that are not explicitly solvable. All of the illustrations in the present paper are worked out explicitly by construction and can be followed without appeal to "highbrow" mathematics.

[18] For example, the density of particles at the point $\mathbf{x}$ in space is represented by

$$F_N = \sum_{a=1}^{N} \delta(\mathbf{x} - \mathbf{q}_a),$$

which depends on $\mathbf{x}$ as a generalized function.

It is a *two-particle variable* if it is of the form

$$F^2 = (0, 0, g^2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2), g^2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2) + g^2(\mathbf{q}_2, \mathbf{p}_2, \mathbf{q}_3, \mathbf{p}_3) + g^2(\mathbf{q}_3, \mathbf{p}_3, \mathbf{q}_1, \mathbf{p}_1), \cdots); \quad (3.12)$$

and similarly for an *n-particle variable. A zero-particle variable* is a constant, i.e.,

$$F^0 = (g^0, g^0, g^0, \cdots), \quad (3.13)$$

where g^0 is a number. The general dynamical variable F of Eq. (3.5) may be expanded in dynamical variables of a definite number of particles, i.e.,

$$F = \sum_{n=0}^{\infty} F^n, \quad (3.14)$$

where

$$g^0 = F_0, \quad (3.15)$$

$$g^1(\mathbf{q}_1, \mathbf{p}_1) = F_1(\mathbf{q}_1, \mathbf{p}_1) - F_0, \quad (3.16)$$

$$g^2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2) = F_2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2) - F_1(\mathbf{q}_1, \mathbf{p}_1) - F_1(\mathbf{q}_2, \mathbf{p}_2) + F_0, \quad (3.17)$$

and so on.

Dynamical variables in classical mechanics also play a role as generators of canonical transformations. A dynamical quantity $W = (W_0, W_1, W_2, \cdots)$ may generate a transformation in F of Eq. (3.5) through

$$dF/d\alpha = \{W, F\} \equiv (\{W_0, F_0\}, \{W_1, F_1\}, \{W_2, F_2\}, \cdots), \quad (3.18)$$

where $\{W_N, F_N\}$ is the N-particle Poisson bracket

$$\{W_N, F_N\} \equiv \sum_{a=1}^{N} \frac{\partial W_N}{\partial \mathbf{p}_a} \cdot \frac{\partial F_N}{\partial \mathbf{q}_a} - \frac{\partial W_N}{\partial \mathbf{q}_a} \cdot \frac{\partial F_N}{\partial \mathbf{p}_a} \quad (3.19)$$

and $\{W_0, F_0\} \equiv 0$. The transformation depends on the continuous parameter α. The time development is generated in this way by the Hamiltonian. We are representing all canonical transformations in the Heisenberg picture in which the dynamical variables are mapped onto themselves and the states and their representatives are left unchanged. A Schrödinger picture, in which the states are transformed, is also possible but will not be discussed here.

4. CREATION AND ANNIHILATION OPERATORS

In this section we define creation and annihilation operators[13] which can act on state representatives and turn them into new (not necessarily normalized) state representatives. The creation and annihilation operators are labeled by points $\mathbf{q}, \mathbf{p}$ in a single-particle phase space and are interpreted as creating or destroying particles at that point in phase space. The *annihilation operator* $\phi(\mathbf{q}, \mathbf{p})$ is defined by

$$\phi(\mathbf{q}, \mathbf{p})f = f', \quad (4.1)$$

with

$$f'_N(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N) \equiv f_{N+1}(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N, \mathbf{q}, \mathbf{p}). \quad (4.2)$$

The *creation operator* is defined by

$$\phi^\dagger(\mathbf{q}, \mathbf{p})f = f'', \quad (4.3)$$

with

$$f''_N(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N) \equiv \sum_{\nu=0}^{N-1} f_{N-1}(\mathbf{q}_{1+\nu}, \mathbf{p}_{1+\nu}, \cdots, \mathbf{q}_{N-1+\nu}, \mathbf{p}_{N-1+\nu}) \times \delta(\mathbf{q}_\nu - \mathbf{q})\delta(\mathbf{p}_\nu - \mathbf{p}) \quad (4.4)$$

for $N > 0$ and

$$f''_0 \equiv 0. \quad (4.5)$$

All the particle indices on the right-hand side of Eq. (4.4) are taken modulo N. The right-hand side of Eq. (4.2) is, in general, a measure in $\mathbf{q}, \mathbf{p}$. The annihilation operators are therefore operator-valued measures and must be smeared with a continuous test function before a proper operator is obtained. The right-hand side of Eq. (4.4) is a measure in $\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_N, \mathbf{p}_N$ even for fixed $\mathbf{q}, \mathbf{p}$. The creation operator is therefore a proper operator even without smearing. The product $\phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})$ is an operator-valued measure.

We next proceed to state the most important properties of the creation and annihilation operators. All these properties are straightforward consequences of the definitions of ϕ and $\phi^\dagger$. Some of the derivations may be found in Ref. 13. The operators ϕ and $\phi^\dagger$ satisfy the commutation relations

$$[\phi(\mathbf{q}, \mathbf{p}), \phi^\dagger(\mathbf{q}', \mathbf{p}')] = \delta(\mathbf{q} - \mathbf{q}')\delta(\mathbf{p} - \mathbf{p}'),$$
$$[\phi(\mathbf{q}, \mathbf{p}), \phi(\mathbf{q}', \mathbf{p}')] = [\phi^\dagger(\mathbf{q}, \mathbf{p}), \phi^\dagger(\mathbf{q}', \mathbf{p}')] = 0. \quad (4.6)$$

They have the Hermiticity property

$$(F \cdot \phi^\dagger(\mathbf{q}, \mathbf{p})f) = (\phi(\mathbf{q}, \mathbf{p})F \cdot f),$$
$$(F \cdot \phi(\mathbf{q}, \mathbf{p})f) = (\phi^\dagger(\mathbf{q}, \mathbf{p})F \cdot f), \quad (4.7)$$

where the action of the creation and annihilation operators on the dynamical variable F is defined in complete analogy to their action on state representatives f; every lower-case f is just replaced by a capital F in Eqs. (4.1)–(4.5).[19]

To each single-particle dynamical variable F^1 we

[19] However, when acting on the dynamical variable F, whose components are continuous functions, the annihilation operator is the proper operator, while the creation operator is an operator-valued measure.

now associate two operators:

$$\tilde{F}^1 \equiv \int d\mathbf{q}\, d\mathbf{p} g^1(\mathbf{q}, \mathbf{p}) \phi^\dagger(\mathbf{q}, \mathbf{p}) \phi(\mathbf{q}, \mathbf{p}), \tag{4.8}$$

$$\hat{F} \equiv \int d\mathbf{q}\, d\mathbf{p} g^1(\mathbf{q}, \mathbf{p}) \times \left(\frac{\partial \phi^\dagger(\mathbf{q}, \mathbf{p})}{\partial \mathbf{p}} \cdot \frac{\partial \phi(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} - \frac{\partial \phi^\dagger(\mathbf{q}, \mathbf{p})}{\partial \mathbf{p}} \cdot \frac{\partial \phi(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} \right)$$

$$\equiv -\int d\mathbf{q}\, d\mathbf{p} g^1(\mathbf{q}, \mathbf{p}) \{\phi^\dagger(\mathbf{q}, \mathbf{p}), \phi(\mathbf{q}, \mathbf{p})\}. \tag{4.9}$$

Both of these operators involve the product of a creation operator and an annihilation operator at the same point in phase space smeared by the continuous function g^1.

The operator $\tilde{F}^1$ has the property

$$\tilde{F}^1 f = F^1 f; \tag{4.10}$$

its action on a statistical element f is identical to that of the original linear operator associated with F^1. The operator $\tilde{F}^1$ is just a way to write the operator F^1 without reference to the phase spaces of various numbers of particles. The expectation value of F^1 may now be expressed as

$$\mathrm{Tr}\, F^1 f = \mathrm{Tr}\, \tilde{F}^1 f = \int d\mathbf{q}\, d\mathbf{p} g^1(\mathbf{q}, \mathbf{p})\, \mathrm{Tr}\, (\phi^\dagger(\mathbf{q}, \mathbf{p}) \phi(\mathbf{q}, \mathbf{p}) f). \tag{4.11}$$

The expectation value of $\phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})$ thus determines the expectation value of any single-particle dynamical variable.

The operator $\hat{F}^1$ serves to generate in $\tilde{G}^1$ the canonical transformation that the dynamical variable F^1 generates in the dynamical variable G^1. That is,

$$d\tilde{G}^1/d\alpha = [\hat{F}^1, \tilde{G}^1] = \{F^1, G^1\}^\sim, \tag{4.12}$$

where $\{F^1, G^1\}^\sim$ is the operator constructed from the single-particle dynamical variable which is the Poisson bracket of F^1 and G^1 in the manner of Eq. (4.8).

Everything said above generalizes immediately to dynamical variables of any number of particles. For a two-particle dynamical variable F^2, we define

$$\tilde{F}^2 \equiv \frac{1}{2} \int d\mathbf{q}_1\, d\mathbf{p}_1\, d\mathbf{q}_2\, d\mathbf{p}_2 g^2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2) \times \phi^\dagger(\mathbf{q}_1, \mathbf{p}_1) \phi^\dagger(\mathbf{q}_2, \mathbf{p}_2) \phi(\mathbf{q}_2, \mathbf{p}_2) \phi(\mathbf{q}_1, \mathbf{p}_1), \tag{4.13}$$

$$\hat{F}^2 \equiv -\frac{1}{2} \int d\mathbf{q}_1\, d\mathbf{p}_1\, d\mathbf{q}_2\, d\mathbf{p}_2 g^2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2) \times \{\phi^\dagger(\mathbf{q}_1, \mathbf{p}_1) \phi^\dagger(\mathbf{q}_2, \mathbf{p}_2), \phi(\mathbf{q}_2, \mathbf{p}_2) \phi(\mathbf{q}_1, \mathbf{p}_1)\}, \tag{4.14}$$

and similarly for dynamical variables of any number of particles. We now find

$$\tilde{F}^n f = F^n f. \tag{4.15}$$

The expectation value of F^n is

$$\mathrm{Tr}\, F^n f = \mathrm{Tr}\, \tilde{F}^n f = \frac{1}{n!} \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_n\, d\mathbf{p}_n g^n(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_n, \mathbf{p}_n) \times \mathrm{Tr}\, \phi^\dagger(\mathbf{q}_1, \mathbf{p}_1) \cdots \phi^\dagger(\mathbf{q}_n, \mathbf{p}_n) \phi(\mathbf{q}_n, \mathbf{p}_n) \cdots \phi(\mathbf{q}_1, \mathbf{p}_1) f, \tag{4.16}$$

and F^n generates canonical transformations in G^m through

$$d\tilde{G}^m/d\alpha = [\hat{F}^n, \tilde{G}^m] = \{F^n, G^m\}^\sim. \tag{4.17}$$

In the last equation the tilde is applied to $\{F^n, G^m\}$, which is not ordinarily a dynamical variable of a definite number of particles. When F is not a dynamical variable of a definite number of particles, $\tilde{F}$ means $\tilde{F}^0 + \tilde{F}^1 + \tilde{F}^2 + \cdots$, where $F^0 + F^1 + F^2 + \cdots$ is the expansion of F in dynamical variables of a definite number of particles [Eqs. (3.14)–(3.17)].

5. LOCAL DYNAMICAL VARIABLES AND INFINITE SYSTEMS

We consider a single-particle dynamical quantity F^1 to be *local* if the corresponding $g^1(\mathbf{q}, \mathbf{p})$ has its support in the product of a compact set of position ($\mathbf{q}$) space by the whole of momentum ($\mathbf{p}$) space. An n-particle dynamical quantity F^n is *local* if the corresponding $g^n(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_n, \mathbf{p}_n)$ has its support in the product of compact sets in the n position spaces by the whole of the corresponding momentum spaces. Many important dynamical variables are not local. The total number of particles $g^1(\mathbf{q}, \mathbf{p}) = 1$, total momentum $\mathbf{g}^1(\mathbf{q}, \mathbf{p}) = \mathbf{p}$, and total kinetic energy $g^1(\mathbf{q}, \mathbf{p}) = \mathbf{p}^2/(2m)$, among many others, are not local. It is these global quantities that usually lose their meaning in an infinite translation invariant (or periodic) system. If the density of particles is finite, the total number of particles is infinite. Local quantities usually retain their meaning even in infinite systems. We now consider only local dynamical variables; and this will open the way to the treatment of infinite translation invariant systems. The restriction of the space of dynamical variables naturally enlarges the space of states. We seek to exhibit some of the new states by implementation of the algebra of local dynamical variables.

Since we have now reexpressed our dynamical variables in a language that makes no reference to phase spaces of finite numbers of particles, it is no longer necessary to express the statistical element f in terms of phase spaces of particles. Since our dynamical variables are now expressed in terms of the creation and annihilation operators, what is now necessary is a

"traced" representation space (i.e., a space with a real linear functional $f \to \mathrm{Tr} f$) for the algebra of creation and annihilation operators, so that expressions (4.12) and (4.17) for the expectation values of operators can be employed. A few simple examples are presented in the following sections.

6. SIMPLE EXAMPLES OF INFINITE SYSTEMS WITH FINITE AVERAGE DENSITY

When the original space of state representatives $f = (f_0, f_1, f_2, \cdots)$ is considered as a representation space for the algebra of creation and annihilation operators, it may be characterized in the following way. The space contains a unique translation invariant element

$$v \equiv (1, 0, 0, 0, \cdots), \tag{6.1}$$

which represents the vacuum—a state without particles. The element v is cyclic in the sense that all other elements of the space may be constructed from v by application of creation operators. All annihilation operators destroy v, i.e.,

$$\phi(\mathbf{q}, \mathbf{p})v = 0. \tag{6.2}$$

Let us now consider an algebra of operators

$$\begin{aligned} \psi(\mathbf{q}, \mathbf{p}) &\equiv \phi(\mathbf{q}, \mathbf{p}) - s(\mathbf{q}, \mathbf{p}), \\ \psi^\dagger(\mathbf{q}, \mathbf{p}) &\equiv \phi^\dagger(\mathbf{q}, \mathbf{p}) - s(\mathbf{q}, \mathbf{p}), \end{aligned} \tag{6.3}$$

where $s(\mathbf{q}, \mathbf{p})$ is a real function such that the integral of $s(s+1)$ over all of $\mathbf{p}$ space and over a bounded region of $\mathbf{q}$ space exists and converges. The algebra generated by ψ and $\psi^\dagger$ is isomorphic to the algebra of the original creation and annihilation operators. One may therefore construct for them a representation space with a cyclic vector g satisfying

$$\psi(\mathbf{q}, \mathbf{p})g = 0, \tag{6.4}$$

or

$$\phi(\mathbf{q}, \mathbf{p})g = s(\mathbf{q}, \mathbf{p})g. \tag{6.5}$$

The representation of the algebra generated by ψ and $\psi^\dagger$ in the space constructed from g is in every way equivalent to the representation of the original algebra generated by ϕ and $\phi^\dagger$ in the space constructed from v. We, however, consider the space based on g as a new representation of the original algebra. This new representation is quite different from the original representation.

In the new representation we have

$$\begin{aligned} \mathrm{Tr}\, \phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})g &= \mathrm{Tr}\, \psi^\dagger(\mathbf{q}, \mathbf{p})\psi(\mathbf{q}, \mathbf{p})g \\ &\quad + s(\mathbf{q}, \mathbf{p})\, \mathrm{Tr}\, \psi(\mathbf{q}, \mathbf{p})g + s(\mathbf{q}, \mathbf{p})\, \mathrm{Tr}\, \psi^\dagger(\mathbf{q}, \mathbf{p})g \\ &\quad + s(\mathbf{q}, \mathbf{p})^2\, \mathrm{Tr}\, g. \end{aligned} \tag{6.6}$$

Now $\mathrm{Tr}\, g = 1$ because of its normalization; Eq. (6.4) makes the first two terms on the right-hand side of Eq. (6.6) vanish; and a straightforward calculation in the original representation leads to $\mathrm{Tr}\, (\psi^\dagger(\mathbf{q}, \mathbf{p})g) = 1$. Eq. (6.6) is thus reduced to

$$\begin{aligned} \mathrm{Tr}\, \phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})g \\ = s(\mathbf{q}, \mathbf{p})[s(\mathbf{q}, \mathbf{p}) + 1] \equiv n(\mathbf{q}, \mathbf{p}). \end{aligned} \tag{6.7}$$

According to Eq. (4.12) this determines all the expectation values of single-particle dynamical variables in the state represented by g.[20] The number density of particles at point $\mathbf{x}$ in space is a single-particle dynamical quantity with $g^1(\mathbf{q}, \mathbf{p}) = \delta(\mathbf{q} - \mathbf{x})$. Consequently the expectation value of the density at $\mathbf{x}$ is

$$n(\mathbf{x}) = \int d\mathbf{q}\, d\mathbf{p} \delta(\mathbf{q} - \mathbf{x}) n(\mathbf{q}, \mathbf{p}) = \int d\mathbf{p} n(\mathbf{x}, \mathbf{p}). \tag{6.8}$$

The $n(\mathbf{x}, \mathbf{p})$ describes the distribution of momenta at $\mathbf{x}$. A translation invariant situation is obtained by choosing $s(\mathbf{q}, \mathbf{p})$ [and therefore $n(\mathbf{q}, \mathbf{p})$] independent of $\mathbf{q}$. A periodic situation results from a choice of $n(\mathbf{q}, \mathbf{p})$ which is periodic in $\mathbf{q}$. Expectation values of dynamical variables of more than one particle may be similarly evaluated. Also, states that differ from the one represented by g only locally may be represented by applying creation operators to g.

7. STATIONARY TRANSLATION INVARIANT OR PERIODIC STATES

Stationary situations, and situations not far removed from being stationary,[21] are of particular interest in physics. We, therefore, concentrate in this section on representations of the creation and annihilation operators based on a cyclic element g that represents a state with a finite average density of particles, one that is either translation invariant or periodic, and that is at the same time stationary.

A *stationary state* is one in which all expectation values remain constant in time. In the original grand canonical approach (Sec. 4) time development is governed by a Hamiltonian

$$H = H^1 + H^2 + \cdots + H^n, \tag{7.1}$$

where the term H^ν in the Hamiltonian is a ν-particle dynamical variable. The operator $\hat{H} = \hat{H}^1 + \hat{H}^2 + \cdots + \hat{H}^n$ generates the time development in the operator $\tilde{F}$ corresponding to any dynamical variable F by commutation. We retain this time development in the new representation, even though $\tilde{H}$ and $\hat{H}$ may have lost their status as operators.

[20] A choice of $s(\mathbf{q}, \mathbf{p})$ between 0 and -1 leads to a negative density of particles in phase space. We discard such a choice as unphysical.

[21] Such as quantum states represented by state vectors expandable in terms of stationary-state vectors.

For illustration let us consider a Hamiltonian with $n = 2$, whose first term H^1 is a single-particle variable characterized by

$$h^1(\mathbf{q}_1, \mathbf{p}_1) = \frac{\mathbf{p}^2}{2m} + u(\mathbf{q}_1), \tag{7.2}$$

and whose second term H^2 is a two-body interaction depending on interparticle distance only, so that it is given by

$$h^2(\mathbf{q}_1, \mathbf{p}_1, \mathbf{q}_2, \mathbf{p}_2) = V(|\mathbf{q}_1 - \mathbf{q}_2|). \tag{7.3}$$

In this case

$$\begin{aligned}\hat{H} = &- \int d\mathbf{q}\, d\mathbf{p} \left\{\frac{\mathbf{p}^2}{2m} + u(\mathbf{q})\right\} \{\phi^\dagger(\mathbf{q}, \mathbf{p}), \phi(\mathbf{q}, \mathbf{p})\} \\ &- \frac{1}{2} \int d\mathbf{q}_1\, d\mathbf{p}_1\, d\mathbf{q}_2\, d\mathbf{p}_2 V(|\mathbf{q}_1 - \mathbf{q}_2|) \\ &\times \{\phi^\dagger(\mathbf{q}_1, \mathbf{p}_1)\phi^\dagger(\mathbf{q}_2, \mathbf{p}_2), \phi(\mathbf{q}_2, \mathbf{p}_2)\phi(\mathbf{q}_1, \mathbf{p}_1)\}. \end{aligned} \tag{7.4}$$

Since all operators $\tilde{F}$ may be expressed in terms of the creation and annihilation operators, all the information about time development is contained in the time development of the creation and annihilation operators themselves. We find

$$\begin{aligned}\frac{d}{dt}\phi(\mathbf{q}, \mathbf{p}) = &\left(-\frac{\mathbf{p}}{m} \cdot \frac{\partial}{\partial \mathbf{q}} + \frac{\partial u(\mathbf{q})}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}}\right. \\ &\left. + \int d\mathbf{q}'\, d\mathbf{p}' \frac{\partial V(|\mathbf{q} - \mathbf{q}'|)}{\partial \mathbf{q}} \phi^\dagger(\mathbf{q}', \mathbf{p}')\phi(\mathbf{q}', \mathbf{p}') \frac{\partial}{\partial \mathbf{p}}\right) \phi(\mathbf{q}, \mathbf{p}), \end{aligned} \tag{7.5}$$

and a corresponding equation for $\phi^\dagger(\mathbf{q}, \mathbf{p})$ in which $\phi(\mathbf{q}, \mathbf{p})$ is replaced by $\phi^\dagger(\mathbf{q}, \mathbf{p})$. It is these equations of motion that we take over into the new representation. If the new representation is based on a cyclic element w representing a stationary state, we must have

$$\begin{aligned}&\sum_{a=1}^{n} \left(-\frac{\mathbf{p}_a}{m} \cdot \frac{\partial}{\partial \mathbf{q}_a} + \frac{\partial u(\mathbf{q}_a)}{\partial \mathbf{q}_a} \cdot \frac{\partial}{\partial \mathbf{p}_a}\right) \\ &\times \operatorname{Tr} A(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_n, \mathbf{p}_n)w \\ &+ \sum_{a=1}^{n} \int d\mathbf{q}'\, d\mathbf{p}' \frac{\partial V(|\mathbf{q}_a - \mathbf{q}'|)}{\partial \mathbf{q}_a} \cdot \frac{\partial}{\partial \mathbf{p}_a} \\ &\times \operatorname{Tr} \phi^\dagger(\mathbf{q}', \mathbf{p}')\phi(\mathbf{q}', \mathbf{p}')A(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_n, \mathbf{p}_n)w' = 0, \end{aligned} \tag{7.6}$$

where $A(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_n, \mathbf{q}_n)$ is any polynomial in creation and annihilation operators labeled by the points $(\mathbf{q}_1, \mathbf{p}_1), \cdots, (\mathbf{q}_n, \mathbf{p}_n)$.

An explicit characterization of the stationary w is, of course, impossible to work out in practice, as is the determination of any representation of a stationary state of many interacting particles. This problem is outside our scope in the paper. However, for noninteracting particles [i.e., taking $V(|\mathbf{q} - \mathbf{q}'|) = 0$ and $H = H^1$], the cyclic element g of the representation discussed in the last section can be made to represent a stationary state. To see this, choose $n = 1$ and $A(\mathbf{q}, \mathbf{p}) = \phi(\mathbf{q}, \mathbf{p})$. This choice turns Eq. (7.6) into

$$\left(-\frac{\mathbf{p}}{m} \cdot \frac{\partial}{\partial \mathbf{q}} + \frac{\partial u(\mathbf{q})}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}}\right) s(\mathbf{q}, \mathbf{p}) = 0. \tag{7.7}$$

If the single-particle potential $u(\mathbf{q})$ vanishes, this condition is satisfied for every choice of $s(\mathbf{q}, \mathbf{p})$ that is independent of $\mathbf{q}$, i.e., in every translation invariant situation. For a periodic $u(\mathbf{q})$, Eq. (7.7) is satisfied by a suitable class of $s(\mathbf{q}, \mathbf{p})$ that are periodic in $\mathbf{q}$. Since any $A(\mathbf{q}_1, \mathbf{p}_1, \cdots, \mathbf{q}_n, \mathbf{p}_n)$ is expressible as a function of $s(\mathbf{q}_1, \mathbf{p}_1), \cdots, s(\mathbf{q}_n, \mathbf{p}_n)$, it is easy to see that once Eq. (7.7) is satisfied, the more general Eq. (7.6) (with V put equal to zero) also is automatically satisfied. For particles that are either free or move in a periodic single-particle potential, we thus have explicit representations featuring a cyclic element representing a stationary state that is either translation invariant or periodic, respectively, and that has a finite average density of particles.

8. THERMAL EQUILIBRIUM

The state of thermal equilibrium is of great importance to physics. Here we do not go into any deep discussion of the meaning and significance of thermal equilibrium. Wherever it may apply, we take thermal equilibrium to be a state that, in the original grand canonical approach of Sec. 3, is represented by $f = (f_1, f_2, \cdots, f_N, \cdots)$ with

$$f_N = A_N \exp(-\beta H_N), \tag{8.1}$$

where H_N is the N-particle Hamiltonian and $\beta = 1/(k_B T)$, k_B being the Boltzmann factor and T the temperature. Equation (8.1) includes the canonical ensemble (for which $A_N = A\delta_{N,N_0}$) and the grand canonical ensemble [for which $A_N = A \exp(-\alpha N)$] as special cases.

Strictly speaking, the applicability of Eq. (8.1) is very limited indeed. So long as the Hamiltonian is translation invariant (or even periodic), the statistical element described in Eq. (8.1) cannot be normalized unless A_N is proportional to $\delta_{N,0}$ and the system in thermal equilibrium is the vacuum. The usual way to use the construction equation (8.1) to describe systems containing particles in thermal equilibrium is to supplement the Hamiltonian by introducing external potentials (representing the walls of a container) that surround a certain volume and slope up to infinity. The external potentials are not customarily treated consistently; one tends to remember them for certain purposes (e.g., normalization)

and forget them for others (e.g., translation invariance).

Since this paper is concerned with infinite translation-invariant systems, we are obviously in need of a generalization of the concept of thermal equilibrium suggested by Eq. (8.1). We need some property which is equivalent to Eq. (8.1) wherever the latter applies, but which retains its meaning in other cases as well. Consider the expectation value of a Poisson bracket $\{U, V\}$ of two dynamical variables taken in the state represented by Eq. (8.1). It is given by

$$\begin{aligned}\mathrm{Tr}\,\{U, V\}f &= \sum_{N=0}^{\infty} \frac{A_N}{N!} \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_N\, d\mathbf{p}_N e^{-\beta H_N} \\ &\quad \times \sum_{a=1}^{N} \left(\frac{\partial U_N}{\partial \mathbf{p}_a} \cdot \frac{\partial V_N}{\partial \mathbf{q}_a} - \frac{\partial U_N}{\partial \mathbf{q}_a} \cdot \frac{\partial V_N}{\partial \mathbf{p}_a} \right) \\ &= \sum_{N=0} \frac{A_N}{N!} \int d\mathbf{q}_1\, d\mathbf{p}_1 \cdots d\mathbf{q}_N\, d\mathbf{p}_N \beta e^{-\beta H} \\ &\quad \times \sum_{a=1}^{N} \left(\frac{\partial H_N}{\partial \mathbf{p}_a} \cdot \frac{\partial V_N}{\partial \mathbf{q}_a} - \frac{\partial H_N}{\partial \mathbf{q}_a} \cdot \frac{\partial V_N}{\partial \mathbf{p}_a} \right) U_N \\ &= \beta\, \mathrm{Tr}\,\{H, V\}Uf, \end{aligned} \tag{8.2}$$

the transition from the second stage to the third being made *via* integrations by parts. The equality between the initial and final expressions in Eq. (8.2) is already meaningful beyond the scope of Eq. (8.1), because now H appears explicitly only as a dynamical variable and not as part of the statistical element. However, remembering the role of H as the generator of time development, we may rewrite Eq. (8.2) as

$$\mathrm{Tr}\,\{U, V\}f = \beta\, \mathrm{Tr}\, U(dV/dt)f. \tag{8.3}$$

This last form makes no explicit use of the Hamiltonian; it uses the time development itself. It is therefore immediately adaptable to our program of treating infinite systems by exploring new representation spaces for the creation and annihilation operators. It immediately becomes

$$\mathrm{Tr}\,[\tilde{U}, \tilde{V}]f = \beta\, \mathrm{Tr}\, \tilde{U}(d\tilde{V}/dt)f, \tag{8.4}$$

where f now stands for the representative of the state of thermal equilibrium.

To sum up: as our *new definition of the state of thermal equilibrium*,[22] we now adopt the condition (8.3) or (8.4), or, in more common notation,

$$\langle\{U, V\}\rangle_\beta = \beta\langle U(dV/dt)\rangle_\beta, \tag{8.5}$$

where $\langle\ \rangle_\beta$ denotes the expectation value for thermal equilibrium at temperature $1/(k_B\beta)$. This definition is equivalent to the original definition (8.1) where the latter applies, but has a much wider range of applicability.

Some general properties of thermal equilibrium may readily be derived from Eq. (8.5)—or from the preceding equivalent equations. In this way, by choosing U as the constant 1, we find that

$$\langle dV/dt\rangle_\beta = 0, \tag{8.6}$$

which shows that *the state of thermal equilibrium is a stationary state.*

Consider now a canonical transformation generated by a dynamical variable W through Eq. (3.18). The change of a thermal expectation value under the canonical transformation is

$$\begin{aligned} d\langle U\rangle_\beta/d\alpha = \langle\{W, U\}\rangle_\beta &= -\langle\{U, W\}\rangle_\beta \\ &= -\beta\langle U(dW/dt)\rangle_\beta, \end{aligned} \tag{8.7}$$

Therefore, if W is a constant of the motion, all expectation values are invariant under the canonical transformation which it generates.[23] The converse is also true in a space in which thermal equilibrium is represented by a cyclic element. It must be observed, however, that upon transition to a representation of the creation and annihilation operators corresponding to an infinite system with finite density, many canonical transformations loose their generators. This means that although the transformation of the creation and annihilation operators is still meaningful and conserves Poisson bracket relations, the generator (or rather $\tilde{W}$) is not a local quantity and hence is no longer an operator. In such cases no conclusion may be reached from Eq. (8.7)—since in this case the W on the right-hand side of Eq. (8.7) is actually $\tilde{W}$. It is then quite possible that expectation values should vary when the transformation is applied, although the generator was a constant of the motion when it existed. Such cases are known as "broken symmetries."

Explicit calculations involving systems of interacting particles in thermal equilibrium are naturally major undertakings and will not be attempted here. Some properties of infinite systems of noninteracting particles in thermal equilibrium are worked out in the next section.

9. IDEAL GAS

In this section we again consider a one-particle Hamiltonian given by $h^1(\mathbf{q}_1, \mathbf{p}_1) = \mathbf{p}_1^2/2m + u(\mathbf{q}_1)$. We investigate the question whether a suitable choice of

[22] In quantum mechanics, the "boundary condition" derived by Martin and Schwinger (Ref. 12) $\langle UV\rangle_\beta = \langle V(-i\hbar\beta)U\rangle_\beta$ may serve as the generalized definition of thermal equilibrium. This condition may be equivalently stated as

$\langle(i/\hbar)[U, V]\rangle_\beta = (i/\hbar)\langle[V(-i\hbar\beta) - V]U\rangle_\beta = \beta\int_0^1 du\langle\dot{V}(-i\hbar\beta u)U\rangle_\beta$.

In the classical limit $\hbar \to 0$, the last condition goes over into Eq. (8.5).

[23] This does not apply to those constants of the motion that depend explicitly on time.

$s(\mathbf{q}, \mathbf{p})$ in the representation of Sec. 6 could describe thermal equilibrium with this Hamiltonian. Let us first assume that a suitable $s_\beta(\mathbf{q}, \mathbf{p})$ exists. Consider a single-particle dynamical variable G^1, of which the $g^1(\mathbf{q}, \mathbf{p})$ is $x(\mathbf{q})y(\mathbf{p})$. Now choosing $U = \phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})$ and $V = G^1$ in Eq. (8.5), we find that

$$\left(y(\mathbf{p})\frac{\partial x(\mathbf{q})}{\partial \mathbf{q}}\cdot\frac{\partial}{\mathbf{p}} - x(\mathbf{q})\frac{\partial y(\mathbf{p})}{\partial \mathbf{p}}\cdot\frac{\partial}{\partial \mathbf{q}}\right)\langle\phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})\rangle_\beta$$
$$= \beta\int d\mathbf{q}'\, d\mathbf{p}'\left(\frac{\mathbf{p}'}{m}\cdot\frac{\partial x(\mathbf{q}')}{\partial \mathbf{q}'}\, y(\mathbf{p}') - \frac{\partial u(\mathbf{q}')}{\partial \mathbf{q}'}\cdot\frac{\partial y(\mathbf{p}')}{\partial \mathbf{p}'}\, x(\mathbf{q}')\right)$$
$$\times\, \langle\phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})\phi^\dagger(\mathbf{q}', \mathbf{p}')\phi(\mathbf{q}', \mathbf{p}')\rangle_\beta. \quad (9.1)$$

In the particular representation of Sec. 6, the expectation values in the last equation may be evaluated explicitly. As was already given in Eq. (6.7), one of them is

$$\langle\phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})\rangle_\beta = s_\beta(\mathbf{q}, \mathbf{p})[s_\beta(\mathbf{q}, \mathbf{p}) + 1] \equiv n_\beta(\mathbf{q}, \mathbf{p}). \quad (9.2)$$

The other turns out to be

$$\langle\phi^\dagger(\mathbf{q}, \mathbf{p})\phi(\mathbf{q}, \mathbf{p})\phi^\dagger(\mathbf{q}', \mathbf{p}')\phi(\mathbf{q}', \mathbf{p}')\rangle_\beta = n_\beta(\mathbf{q}, \mathbf{p})n_\beta(\mathbf{q}', \mathbf{p}') + \delta(\mathbf{q} - \mathbf{q}')\delta(\mathbf{p} - \mathbf{p}')n_\beta(\mathbf{q}, \mathbf{p}). \quad (9.3)$$

When Eqs. (9.3) and (9.4) are substituted into Eq. (9.1), the latter equation becomes

$$\left(y(\mathbf{p})\frac{\partial x(\mathbf{q})}{\partial \mathbf{q}}\cdot\frac{\partial}{\partial \mathbf{p}} - x(\mathbf{q})\frac{\partial y(\mathbf{p})}{\partial \mathbf{p}}\cdot\frac{\partial}{\partial \mathbf{q}}\right)n(\mathbf{q}, \mathbf{p})$$
$$= \beta n(\mathbf{q}, \mathbf{p})\int d\mathbf{q}'\, d\mathbf{p}'\left(\frac{\mathbf{p}'}{m}\cdot\frac{\partial x(\mathbf{q}')}{\partial \mathbf{q}'}\, y(\mathbf{p}')\right.$$
$$\left. - \frac{\partial u(\mathbf{q}')}{\partial \mathbf{q}'}\cdot\frac{\partial y(\mathbf{p}')}{\partial \mathbf{p}'}\, x(\mathbf{q}')\right)n(\mathbf{q}', \mathbf{p}')$$
$$+ \beta\left(\frac{\mathbf{p}'}{m}\cdot\frac{\partial x(\mathbf{q})}{\partial \mathbf{q}}\, y(\mathbf{p}) - \frac{\partial u(\mathbf{q})}{\partial \mathbf{q}}\cdot\frac{\partial y(\mathbf{p})}{\partial \mathbf{p}}\, x(\mathbf{q})\right)n(\mathbf{q}, \mathbf{p}). \quad (9.4)$$

Equation (9.4) must hold for any choice of $x(q)$ and $y(p)$—provided, however, that $x(\mathbf{q})$ has compact support and that both functions are differentiable and that $\int d\mathbf{p}y(\mathbf{p})$ is finite. By choosing a point $(\mathbf{q}, \mathbf{p})$ where $\partial x/\partial \mathbf{q} = \partial y/\partial \mathbf{p} = 0$, we find that

$$\int d\mathbf{q}'\, d\mathbf{p}'\left(\frac{\mathbf{p}'}{m}\cdot\frac{\partial x(\mathbf{q}')}{\partial \mathbf{q}'}\, y(\mathbf{p}')\right.$$
$$\left. - \frac{\partial u(\mathbf{q}')}{\partial \mathbf{q}'}\cdot\frac{\partial y(\mathbf{p}')}{\partial \mathbf{p}'}\, x(\mathbf{q}')\right)n(\mathbf{q}', \mathbf{p}') = 0 \quad (9.5)$$

for every admissible $x(\mathbf{q})$ and $y(\mathbf{p})$ whose derivative vanishes somewhere. Using Eq. (9.5), we may separate Eq. (9.4) [which holds for arbitrary choice of $x(\mathbf{q})$ and $y(\mathbf{p})$] into

$$-\frac{\partial}{\partial \mathbf{p}}\, n_\beta(\mathbf{q}, \mathbf{p}) = \beta\frac{\mathbf{p}}{m}\, n_\beta(\mathbf{q}, \mathbf{p}), \quad (9.6)$$

$$\frac{\partial}{\partial \mathbf{q}}\, n_\beta(\mathbf{q}, \mathbf{p}) = -\beta\frac{\partial u(\mathbf{q})}{\partial \mathbf{q}}\, n_\beta(\mathbf{q}, \mathbf{p}). \quad (9.7)$$

The last pair of equations are readily solved to yield

$$n_\beta(\mathbf{q}, \mathbf{p}) = A\exp\{-\beta[(\mathbf{p}^2/2m) + u(\mathbf{q})]\}. \quad (9.8)$$

The remaining check is to determine whether the last form that $n_\beta(\mathbf{q}, \mathbf{p})$ must take actually satisfies Eq. (9.5) and whether the thermal equilibrium condition Eq. (8.5) is satisfied also for arbitrary choices of U and V. These checks are left to the reader.

Two remarks are pertinent in closing.

(1) For the treatment of the present section, it was essential to consider a single-particle Hamiltonian. It was not essential, however, to give it a particular form. We could have more generally found

$$n_\beta(\mathbf{q}, \mathbf{p}) = A\exp[-\beta h^1(\mathbf{q}, \mathbf{p})] \quad (9.9)$$

by following the same procedure.

(2) Although $n_\beta(\mathbf{q}, \mathbf{p})$ is determined uniquely, $s_\beta(\mathbf{q}, \mathbf{p})$ is not. Equation (9.3), connecting $s_\beta(\mathbf{q}, \mathbf{p})$ with $n_\beta(\mathbf{q}, \mathbf{p})$, is quadratic and admits two solutions:

$$s_\beta(\mathbf{q}, \mathbf{p}) = \tfrac{1}{2} \pm [\tfrac{1}{4} + n_\beta(\mathbf{q}, \mathbf{p})]^{\frac{1}{2}}. \quad (9.10)$$

These characterize two *different* representations of the creation and annihilation operators that describe the *same* physical situation.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge discussions with and comments by J. M. Cooper, H. Ekstein, Y. Frishman, and M. Peshkin.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

Single-Particle Green's Function for a One-Dimensional Many-Fermion System*

ALBA THEUMANN
Belfer Graduate School of Science, Yeshiva University, New York, New York

(Received 17 May 1967)

In this paper we compute the single-particle Green's function for a system of N-interacting fermions in one dimension. The problem is exactly soluble in the case of a δ-function potential. We consider also the case of a small modification to a δ-function potential and we compute the Green's function up to first order in a perturbation expansion. In both cases we find that the Green's function has a branch cut in the complex energy plane but it has no single-particle poles.

INTRODUCTION

SOME years ago, Luttinger[1] proposed an exactly soluble model for a system of N interacting fermions in one dimension. Later, Mattis and Lieb[2] reconsidered the problem. Finding that the solution given by Luttinger was incorrect, they were able to give an exact solution in terms of boson-like collective excitations. They calculated the spectrum, the free energy, and the dielectric constant and verified the existence of a nonanalyticity at the Fermi surface.

A different kind of information can be obtained through the calculation of the Green's functions of the model. As is very well known,[3] the poles of the single-particle Green's function in the complex energy plane determine the spectrum of the elementary excitations.

In the present paper, we compute the single-particle Green's function for Luttinger's Hamiltonian, expressing it explicitly in terms of the collective plasmons modes of Ref. 2. Due to the exact diagonalization of the Hamiltonian, the calculation can be carried through exactly, and the result is expressed as a function of the arbitrary potential $V(x)$.

In the limit $V(x) = \delta(x)$, Luttinger's model becomes relativistic and coincides with Thirring's model[4] of a two-dimensional field theory. Hence, it is not surprising that our result coincides with the single-particle Green's function for the Thirring model, computed by Johnson,[5] when we replace our general potential by a δ-function potential.

We show in Sec. 2 the general result and in Sec. 3 the exact relationship to Johnson's solution. In Sec. 4 we compute the modification to the Green's function for a potential that differs slightly from a δ function, in first order in a perturbation expansion.

We find in both cases that the single-particle Green's function is analytic on the complex energy plane except for a branch cut along the real axis. The absence of complex poles is remarkable, showing that the system has a discontinuous behavior when the interaction is switched on, regardless of the actual value of the coupling constant.

1. GENERAL EXPRESSION FOR $G(x, t)$

To start, we recall some of the results in the work of Mattis and Lieb. From now on we refer to this paper as ML. The system is initially considered to be enclosed in a one-dimensional "box" of length L.

$\psi(x)$ is a two-component, one-dimensional fermion field,

$$\psi(x) = \frac{1}{(L)^{\frac{1}{2}}} \sum_k e^{ikx} \begin{pmatrix} a_{1k} \\ a_{2k} \end{pmatrix},$$
$$\psi^+(x) = \frac{1}{(L)^{\frac{1}{2}}} \sum_k e^{-ikx} (a_{1k}^+ a_{2k}^+), \tag{1.1}$$

where

$$\{a_{jk}, a_{j'k'}\} = 0 = \{a_{jk}^+, a_{j'k'}^+\},$$
$$\{a_{jk}, a_{j'k'}^+\} = \delta_{jj'}\delta_{kk'}, \tag{1.2}$$

$$\rho_1(+p) = \sum_k a_{1k+p}^+ a_{1k}, \quad \rho_1(-p) = \sum_k a_{1k}^+ a_{1k+p},$$
$$\rho_2(+p) = \sum_k a_{2k+p}^+ a_{2k}, \quad \rho_2(-p) = \sum_k a_{2k}^+ a_{2k+p}, \tag{1.3}$$

$$H = H_0 + H = \sum_k (a_{1k}^+ a_{1k} - a_{2k}^+ a_{2k})k$$
$$+ \frac{2\lambda}{L} \sum_{\substack{k_1 k_2 \\ k_3 k_4}} \delta_{k_1+k_3, k_2+k_4} v(k_3 - k_4) a_{1k_1}^+ a_{1k_2} a_{2k_3}^+ a_{2k_4}. \tag{1.4}$$

Making a canonical transformation to particle–hole

* This research is partially sponsored by U.S. Air Force Office of Scientific Research, Office of Aerospace Research Grant AF107566 and National Science Foundation Grant GP 6058.

[1] J. M. Luttinger, J. Math. Phys. **4**, 1154 (1963).
[2] D. C. Mattis and E. H. Lieb, J. Math. Phys. **6**, 304 (1965).
[3] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinsky, *Quantum Field Theory in Statistical Physics* (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1963).
[4] W. Thirring, Ann. Phys. (N.Y.) **3**, 91 (1958).
[5] K. Johnson, Nuovo Cimento **20**, 773 (1961).

language, the former expressions can be written

$$a_{1k} = \begin{cases} b_k & k \geq 0 \\ c_k^+ & k < 0 \end{cases} \qquad a_{2k} = \begin{cases} b_k & k < 0 \\ c_k^+ & k \geq 0 \end{cases}$$

$$H = \sum_k |k| (b_k^+ b_k + c_k^+ c_k) + \frac{2\lambda}{L} \sum_{p>0} [v(p)\rho_1(-p)\rho_2(p) + v(-p)\rho_1(p)\rho_2(-p)],$$

$$v(p) = \text{real, even function of } p. \quad (1.5)$$

It was proved in ML that the commutators of the $\rho_i(p)$ do not vanish identically, but satisfy the relations

$$[\rho_1(-p), \rho_1(p')] = [\rho_2(p), \rho_2(-p')] = (pL/2\pi)\delta_{pp'},$$
$$[\rho_1(p), \rho_2(p')] = 0, \quad (1.6)$$

$$[H_0, \rho_1(\pm p)] = \pm p\rho_1(\pm p),$$
$$[H_0, \rho_2(\pm p)] = \mp p\rho_2(\pm p). \quad (1.6a)$$

It is possible to define a boson field that satisfies the canonical commutation relations

$$\alpha^+(p) = (2\pi/pL)^{\frac{1}{2}}\rho_1(p), \quad \alpha(p) = (2\pi/pL)^{\frac{1}{2}}\rho_1(-p),$$
$$\beta^+(p) = (2\pi/pL)^{\frac{1}{2}}\rho_2(-p), \quad \beta(p) = (2\pi/pL)^{\frac{1}{2}}\rho_2(p). \quad (1.7)$$

We refer the reader to the original paper for more details of the diagonalization of H. What is important for us to know is that after performing the canonical transformation:

$$H_D = e^{iS} H e^{-iS},$$

where

$$S = i \sum_{p>0} \varphi(p)[\alpha^+(p)\beta^+(p) - \alpha(p)\beta(p)]. \quad (1.8)$$

H is brought to diagonal form

$$H_D = H_1 + H_2,$$

$$H_1 = \sum_k |k| (b_k^+ b_k + c_k^+ c_k) - \sum_{p>0} p[\alpha^+(p)\alpha(p) + \beta^+(p)\beta(p)],$$

$$H_2 = \sum_{p>0} p\omega_p[\alpha^+(p)\alpha(p) + \beta^+(p)\beta(p)];$$
$$\omega_p = \operatorname{sech} 2\varphi(p). \quad (1.9)$$

The vacuum renormalization energy has been subtracted out. The boson operators transform according to

$$e^{iS}\alpha^+(p)e^{-iS} = \alpha^+(p)\cosh\varphi(p) + \beta(p)\sinh\varphi(p),$$
$$e^{iS}\beta(p)e^{-iS} = \beta(p)\cosh\varphi(p) + \alpha^+(p)\sinh\varphi(p). \quad (1.10)$$

$\varphi(p)$ is related to the Fourier transform of the potential through

$$\tanh 2\varphi(p) = -\lambda v(p)/\pi. \quad (1.11)$$

The one-particle Green's function is defined by the relation

$$G(x, t) = -i\langle\Phi_H| T\psi_H(x, t)\psi_H^+(0)|\Phi_H\rangle$$
$$= -i\theta(t)\langle\Phi_H|\psi_H(x, t)\psi_H^+(0)|\Phi_H\rangle + i\theta(-t)\langle\Phi_H|\psi_H^+(0)\psi_H(x, t)|\Phi_H\rangle,$$
$$\theta(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}, \quad (1.12)$$

$|\Phi_H\rangle$: true ground state $= e^{-iS}|\Phi\rangle$.

$|\Phi\rangle$: noninteracting ground state, filled with b particles from $-k_F$ to k_F,

$$\psi_H(x, t) = e^{iHt}\psi(x)e^{-iHt},$$
$$\alpha(p)|\Phi\rangle = \beta(p)|\Phi\rangle = 0, \quad (1.13)$$
$$\psi_H\psi_H^+ = \psi_{1H}\psi_{1H}^+ + \psi_{2H}\psi_{2H}^+.$$

The system is symmetric in k and $-k$ and the field $\psi_1(x)$ represents "particles" for $k > 0$ and the field $\psi_2(x)$ represents particles for $k < 0$, thus we need only compute

$$G_1(x, t) = -i\langle\Phi_H| T\psi_{1H}(x, t)\psi_{1H}^+(0)|\Phi_H\rangle.$$

Let us consider the function

$$A(x, t) = \langle\Phi_H|\psi_{1H}(x, t)\psi_{1H}^+(0)|\Phi_H\rangle$$
$$= \langle\Phi| e^{iS}e^{iHt}\psi_1(x)e^{-iHt}\psi_1^+(0)e^{-iS}|\Phi\rangle$$
$$= \langle\Phi| e^{iH_Dt}e^{iS}\psi_1(x)e^{-iS}e^{-iH_Dt}e^{iS}\psi_1^+(0)e^{-iS}|\Phi\rangle. \quad (1.14)$$

Expressions of the type $e^{iS}\psi_1(x)e^{-iS}$ were computed in ML, Sec. 5. The result is

$$e^{iS}\psi_1(x)e^{-iS} = \psi_1(x)\exp[Q^+(x) - Q(x)]$$
$$= \exp[Q^+(x) - Q(x)]\psi_1(x), \quad (1.15)$$

$$Q^+(x) = \sum_{p>0}\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}}\{\sinh\varphi(p)e^{ipx}\beta^+(p) - [\cosh\varphi(p) - 1]e^{-ipx}\alpha^+(p)\}. \quad (1.16)$$

Inserting (15) into (14) we find

$$A(x, t) = \langle\Phi| e^{iH_Dt}\psi_1(x)\exp[Q^+(x) - Q(x)] \times \exp[-iH_Dt]\exp[Q(0) - Q^+(0)]\psi_1^+(0)|\Phi\rangle$$
$$= \langle\Phi| e^{iH_Dt}\psi_1(x)\exp[-iH_Dt] \times \exp[Q^+(x, t) - Q(x, t)] \times \exp[Q(0) - Q^+(0)]\psi_1^+(0)|\Phi\rangle, \quad (1.17)$$

$$Q^+(x, t) = e^{iH_Dt}Q^+(x)\exp[-iH_Dt] = \sum_{p>0}\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}} \times \{\sinh\varphi(p)e^{ip(x+\omega_pt)}\beta^+(p) - [\cosh\varphi(p) - 1]e^{-ip(x-\omega_pt)}\alpha^+(p)\}. \quad (1.18)$$

Considering that

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]} \text{ if } [A,[A,B]] = [B,[A,B]] = 0,$$

we have

$$\exp[Q^+(x,t) - Q(x,t)] \exp[Q(0) - Q^+(0)]$$
$$= \exp[Q^+(x,t) - Q(x,t) + Q(0) - Q^+(0)]$$
$$\times \exp \tfrac{1}{2}[Q^+(x,t) - Q(x,t), Q(0) - Q^+(0)]$$

and

$$A(x,t) = \langle\Phi| e^{iH_D t}\psi_1(x)e^{-iH_D t}$$
$$\times \exp[Q^+(x,t) - Q^+(0) - Q(x,t) + Q(0)]$$
$$\times \psi_1^+(0)|\Phi\rangle \cdot \exp\{i \operatorname{Im} R(x,t)\}, \quad (1.19)$$

where

$$R(x,t) = [Q^+(x,t), Q(0)]$$
$$= -\sum_{p>0} \frac{2\pi}{pL}\{\sinh^2\varphi(p) \exp[ip(x+\omega_p t)]$$
$$+ [\cosh\varphi(p) - 1]^2 \exp[-ip(x-\omega_p t)]\}. \quad (1.20)$$

In the same way as before

$$e^{iH_D t}\psi_1(x)e^{-iH_D t} = e^{iH_0 t}\psi_1(x)$$
$$\times \exp[i\Gamma t + P^+(x,t) - P(x,t)] \exp(-iH_0 t),$$
$$\Gamma = \sum_{p>0} \frac{2\pi}{L}(\omega_p - 1), \quad (1.21)$$

$$P^+(x,t) = -\sum_{p>0} 2i\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}} \sin\tfrac{1}{2}p(\omega_p - 1)t$$
$$\times \exp\left[-ip\left(x - \frac{\omega_p - 1}{2}t\right)\right]\alpha^+(p), \quad (1.22)$$

$$e^{iH_D t}\psi_1(x)e^{-iH_D t} = e^{iH_0 t}\psi_1(x)e^{-iH_0 t}$$
$$\times \exp[i\Gamma t + \Lambda^+(x,t) - \Lambda(x,t)],$$
$$\Lambda^+(x,t) = e^{iH_0 t}P^+(x,t)e^{-iH_0 t}$$
$$= -\sum_{p>0} 2i\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}} \sin\tfrac{1}{2}p(\omega_p - 1)t$$
$$\times \exp\{-ip[x - \tfrac{1}{2}(\omega_p - 1)t]\}$$

because

$$e^{iH_0 t}\alpha^+(p)e^{-iH_0 t} = e^{ipt}\alpha^+(p).$$

Looking at (16), we can see that

$$Q^+(x,t) = Q_\beta^+(x,t) - Q_\alpha^+(x,t),$$
$$Q_\alpha^+(x,t) = \sum_{p>0}\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}}[\cosh\varphi(p) - 1]$$
$$\times \exp[-ip(x - \omega_p t)]\alpha^+(p). \quad (1.23)$$

It was proved in ML that the state $|\Phi\rangle$ can be considered a product state $|\Phi_\alpha\rangle\,|\Phi_\beta\rangle$. Since the β operators commute with α and a_{1k}, we move them to the edges of $A(x,t)$ and let them act on $|\Phi_\beta\rangle$.

$$e^{Q_\beta}|\Phi_\beta\rangle = |\Phi_\beta\rangle, \qquad \langle\Phi_\beta| e^{Q_\beta{}^+} = \langle\Phi_\beta|.$$

One must be careful to take into account the fact that $[Q_\beta, Q_\beta^+] \neq 0$:

$$\exp[Q_\beta^+(x,t) - Q_\beta^+(0) - Q_\beta(x,t) + Q_\beta(0)]$$
$$= \exp[Q_\beta^+(x,t) - Q_\beta^+(0)] \exp[-Q_\beta(x,t) + Q_\beta(0)]$$
$$\times \exp \tfrac{1}{2}[R_\beta(0) - \operatorname{Re} R_\beta(x,t)].$$

From (1.20) we have

$$R_\beta(x,t) = -\sum_{p>0} \frac{2\pi}{pL} \sinh^2\varphi(p) e^{ip(x+\omega_p t)}. \quad (1.24)$$

Introducing Eqs. (1.21)–(1.24) into Eq. (1.19) we obtain

$$A(x,t) = \langle\Phi_\alpha| e^{iH_0 t}\psi_1(x)e^{-iH_0 t} \exp[\Lambda^+(x,t) - \Lambda(x,t)]$$
$$\times \exp[-Q_\alpha^+(x,t) + Q_\alpha^+(0) + Q_\alpha(x,t) - Q_\alpha(0)]$$
$$\times \psi_1^+(0)|\Phi_\alpha\rangle$$
$$\times \exp\{R_\beta(0) - \operatorname{Re} R_\beta(x,t) + i\Gamma t + i \operatorname{Im} R(x,t)\} \quad (1.25)$$

and

$$A(x,t) = \langle\Phi_\alpha|\, \psi_{10}(x,t)e^{F^+(x,t)}e^{-F(x,t)}\psi_{10}^+(0)|\Phi_\alpha\rangle$$
$$\times \exp\{i\Gamma t + i \operatorname{Im} R(x,t) + i \operatorname{Im} U(t) - i \operatorname{Im} V(x,t)$$
$$- W(x,t) + R_\beta(0) - \operatorname{Re} R_\beta(x,t)\}, \quad (1.26)$$

where

$$\psi_{10}(x,t) = e^{iH_0 t}\psi_1(x)e^{-iH_0 t} = \sum_k e^{ik(x-t)}a_{1k}, \quad (1.27)$$

$$U(t) = [\Lambda(x,t), Q_\alpha^+(x,t)]$$
$$= \sum_{p>0} 2i\frac{2\pi}{pL} \sin\tfrac{1}{2}p(\omega_p - 1)t$$
$$\times [\cosh\varphi(p) - 1]e^{\frac{1}{2}ip(\omega_p - 1)t}, \quad (1.28)$$

$$V(x,t) = [\Lambda(x,t), Q_\alpha^+(0)]$$
$$= \sum_{p>0} 2i\frac{2\pi}{pL} \sin\tfrac{1}{2}p(\omega_p - 1)t$$
$$\times [\cosh\varphi(p) - 1]e^{ip[x - \frac{1}{2}(\omega_p + 1)t]}, \quad (1.29)$$

$$F(x,t) = \sum_{p>0}\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}}\{2i[\sin\tfrac{1}{2}p(\omega_p - 1)t]e^{ip[x - \frac{1}{2}(\omega+1)t]}$$
$$+ [\cosh\varphi(p) - 1][1 - e^{ip(x-\omega_p t)}]\}\alpha(p)$$
$$= \sum_{p>0} F_p\alpha(p), \quad (1.30)$$

$$2W(x,t) = [F(x,t), F^+(x,t)]$$
$$= \sum_{p>0} 2\frac{2\pi}{pL}\{1 - \cos p(\omega_p - 1)t$$
$$+ [\cosh\varphi(p) - 1][-\cos p(x - \omega_p t)$$
$$+ \cos p(x-t) + 1 - \cos p(\omega_p - 1)t]$$
$$+ [\cosh\varphi(p) - 1]^2[1 - \cos p(x - \omega_p t)]\}. \quad (1.31)$$

Finally, from (1.30) and (1.27)

$$[\psi_{10}(x,t), F^+(x,t)] = \psi_{10}(x,t)\sum_{p>0} F_p e^{ip(x-t)}\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}},$$

$$\begin{aligned} A(x,t) &= \langle\Phi_\alpha|\ \psi_{10}(x,t)\psi_{10}^+(0)\,|\Phi_\alpha\rangle \\ &\times \exp\Big\{i\Gamma t + i\,\mathrm{Im}\,R(x,t) + i\,\mathrm{Im}\,U(t) + R_\beta(0) \\ &- \mathrm{Re}\,R(x,t) - i\,\mathrm{Im}\,V(x,t) - W(x,t) \\ &+ \sum_p F_p e^{ip(x-t)}\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}} - \sum_p F_p\left(\frac{2\pi}{pL}\right)^{\frac{1}{2}}\Big\}, \end{aligned} \tag{1.32}$$

and if we call

$$\begin{aligned} \langle\Phi_\alpha|\ \psi_{10}(x,t)\psi_{10}^+(0)\,|\Phi_\alpha\rangle &= A_0(x,t), \\ A(x,t) &= A_0(x,t)\exp\,[Z(x,t)], \end{aligned} \tag{1.33}$$

$$\begin{aligned} Z(x,t) = \frac{2\pi}{L}\sum_{p>0}\frac{1}{p}\Big\{ & e^{ipx}[\cosh^2\varphi(p)e^{-ip\omega_p t} - e^{-ipt}] \\ & + \sinh^2\varphi(p)\exp\,[-ip(x+\omega_p t)]\Big\} \\ & - \frac{4\pi}{L}\sum_{p>0}\frac{1}{p}\sinh^2\varphi(p) \\ & - i\frac{2\pi}{L}\sum_{p>0}\frac{1}{p}\sin p(\omega_p - 1)t + i\Gamma t. \end{aligned} \tag{1.34}$$

The expression $A(x,0) = A_0(x,0)\exp\,[Z(x,0)]$ coincides with the similar expression $I(s-t) = I(x)$ of ML, Sec. 5.

It is easy to see, following the same steps as before, that

$$\begin{aligned} B(x,t) &= \langle\Phi_H|\ \psi_{1H}^+(0)\psi_{1H}(x,t)\,|\Phi_H\rangle \\ &= \exp\,[Z^+(x,t)]\langle\Phi_H|\ \psi_{10\,H}^+(0)\psi_{10H}(x,t)\,|\Phi_H\rangle \\ &= \exp\,[Z^+(x,t)]B_0(x,t). \end{aligned} \tag{1.35}$$

Thus, in (12)

$$\begin{aligned} G_1(x,t) &= -i\theta(t)A(x,t) + i\theta(-t)B(x,t) \\ &= -i\theta(t)e^{Z(x,t)}A_0(x,t) + i\theta(-t)e^{Z^+(x,t)}B_0(x,t). \end{aligned} \tag{1.36}$$

At this point, if we recall that ω_p and $\varphi(p)$ depend on the interaction potential $v(p)$, we can see that it is impossible to obtain any information from $G(x,t)$ without being more explicit about the interaction. However, we can anticipate by looking at (1.34) that the different exponential terms give rise to an *infinite series of poles on the real energy axis which in limit $L \to \infty$ coalesce forming a branch cut, while there are no single complex poles, that is, there are no single particle excitations.* We show in the next section that this result is confirmed in the case of a δ-function potential.

2. δ-FUNCTION POTENTIAL

$$v(p) = \int_{-\infty}^{\infty} V(x)e^{-ipx}dx = \int_{-\infty}^{\infty}\delta(x)e^{-ipx}dx = 1. \tag{2.1}$$

Replacing $v(p) = 1$ in Eq. (1.11) we obtain

$$\tanh 2\varphi(p) = -\lambda/\pi, \tag{2.2}$$

$$\omega_p = \mathrm{sech}\,2\varphi(p) = [1 - (\lambda^2/\pi^2)]^{\frac{1}{2}} = \Omega, \tag{2.3}$$

$$\begin{aligned} \cosh^2\varphi(p) &= \tfrac{1}{2}[\cosh 2\varphi(p) + 1] = (1+\Omega)/2\Omega, \\ \sinh^2\varphi(p) &= \tfrac{1}{2}[\cosh 2\varphi(p) - 1] = (1-\Omega)/2\Omega. \end{aligned} \tag{2.4}$$

Introducing Eq. (2.3) and (2.4) in Eq. (1.34) and replacing

$$\frac{2\pi}{L}\sum_{p>0} \quad \text{by} \quad \int_0^\infty dp,$$

$$\begin{aligned} Z(x,t) = {} & \frac{1+\Omega}{2\Omega}\int_0^\infty \frac{1}{p}e^{ip(x-\Omega t)}\,dp \\ & + \frac{1-\Omega}{2\Omega}\int_0^\infty e^{-ip(x+\Omega t)}\frac{1}{p}\,dp - \int_0^\infty e^{ip(x-t)}\frac{1}{p}\,dp \\ & - \frac{1-\Omega}{\Omega}\int_0^\infty\frac{1}{p}\,dp - i\int_0^\infty \sin\,[p(\Omega-1)t]\frac{1}{p}\,dp \\ & + it\int_0^\infty(\Omega-1)\,dp. \end{aligned} \tag{2.5}$$

The integrals appearing in $Z(x,t)$ are highly divergent, but we are able to separate out the divergencies, as usual, and to concentrate the divergencies into an infinite renormalization constant.

Considering a general integral of the type

$$I = \int_0^\infty e^{i\mu p}\frac{1}{p}\,dp,$$

we can integrate by parts to separate the divergence at $p = 0$ and we can introduce a small imaginary part to μ to make it convergent when $p \to \infty$,

$$\begin{aligned} I = \int_0^\infty e^{i\mu p}\frac{1}{p}\,dp &= \lim_{\epsilon\to 0}\int_0^\infty e^{i(\mu+i\epsilon)p}\frac{1}{p}\,dp \\ &= \lim_{\epsilon\to 0}[e^{i\mu p}e^{-\epsilon p}\log(p)]_0^\infty \\ &\quad - \lim_{\epsilon\to 0}(\mu + i\epsilon)\int_0^\infty e^{i(\mu+i\epsilon)p}\log(p)\,dp. \end{aligned} \tag{2.6}$$

In the following, we suppress the expression $\lim_{\epsilon\to 0}$. Since

$$\lim_{p\to\infty} e^{-p\epsilon}\log(p) = 0,$$

therefore, the first term in (2.6) contributes an infinite constant log (0):

$$\begin{aligned} \int_0^\infty e^{-(\epsilon-i\mu)p}\log(p)\,dp &= -\frac{1}{\epsilon - i\mu}[\gamma + \log(\epsilon - i\mu)] \\ &\to \frac{1}{i\mu}[\gamma + \log(-i\mu)], \end{aligned} \tag{2.7}$$

where γ is Euler's constant.

Our complete expression thus becomes

$$I = \int_0^\infty e^{i\mu p} \frac{1}{p}\, dp = \log(0) - [\gamma + \log(-i\mu)]. \quad (2.8)$$

Performing in this way the remaining integrals, in (5), we find

$$Z(x, t) = \frac{1-\Omega}{\Omega} \log(0) - i\frac{\pi}{2} + i\Gamma t + \log \frac{(x-t)}{[(x-\Omega t)^{[(1+\Omega)/2\Omega]}(x+\Omega t)^{[(1-\Omega)/2\Omega]}]}. \quad (2.9)$$

In this expression Γ is an infinite constant $= \int_0^\infty (\Omega - 1)\, dp$ which gives rise to a uniform contribution to the energy and

$$\int_0^\infty \sin p(\Omega - 1)t \frac{1}{p}\, dp = \frac{\pi}{2} \operatorname{sgn}[(\Omega - 1)t].$$

From Eq. (1.33) we have

$$A(x, t) = -iZ_0 e^{i\Gamma t} A_0(x, t)(x - t) \times [(x - \Omega t)^{(1+\Omega)/2\Omega}(x + \Omega t)^{(1-\Omega)/2\Omega}]^{-1}$$

where

$$Z_0 = \exp[(1 - \Omega)\log(0)/\Omega] \quad (2.10)$$

$$A_0(x, t) = \int_{-\infty}^\infty dk \int_{-\infty}^\infty d\omega \frac{\theta(k - k_F) e^{ikx} e^{-i\omega t}}{\omega - k + i\epsilon} = i\frac{e^{ik_F(x-t)}}{(x-t)}. \quad (2.11)$$

Introducing Eq. (2.11) into Eq. (2.10) and suppressing for simplicity Z_0 and Γ we have

$$A(x, t) = \frac{\exp[ik_F(x-t)]}{(x-\Omega t)^{(1+\Omega)/2\Omega}(x+\Omega t)^{(1-\Omega)/2\Omega}} \theta(t) \quad (2.12)$$

$$B_0(x, t) = \int_{-\infty}^\infty dk \int_{-\infty}^\infty d\omega \frac{\theta(k - k_F)}{\omega - k - i\epsilon} e^{ikx} e^{-i\omega t} = i \exp[ik_F(x-t)]\theta(-t)/(x-t). \quad (2.13)$$

From Eqs. (2.9) and (2.13), the expression for $G_1(x, t)$ in (1.36) becomes

$$G_1(x, t) = i \frac{\exp[ik_F(x-t)]}{(x-\Omega t)^{(1+\Omega)/2\Omega}(x+\Omega t)^{(1-\Omega)/2\Omega}},$$

$$G_1(k, \omega) = i\int_{-\infty}^\infty dx \int_{-\infty}^\infty dt \frac{e^{ik_F(x-t)} e^{-ikx} e^{i\omega t}}{(x-\Omega t)^\beta (x+\Omega t)^{\beta-1}},$$

$$\beta = \frac{1}{2}\left(1 + \frac{1}{\Omega}\right). \quad (2.14)$$

Introducing the new variables $u = \frac{1}{2}(x + \Omega t)$; $v = \frac{1}{2}(x - \Omega t)$, the integral (14) can be separated and expressed in the form

$$G_1(k, \omega) = \frac{2^{-\beta} i}{4\Omega} \int_{-\infty}^\infty du \frac{\exp[-iu(\bar{k} - \bar{\omega})]}{u^{\beta-1}} \times \int_{-\infty}^\infty \frac{\exp[-iv(\bar{k} + \bar{\omega})]}{v^\beta}\, dv,$$

$$\bar{k} = k - k_F, \quad \bar{\omega} = (\omega - k_F)/\Omega. \quad (2.15)$$

The integral (2.15) is highly divergent and does not have any meaning in terms of ordinary functions. However, like other divergent expressions in field theory, it arises because we are actually dealing with generalized functions or "distributions." [6] In the language of Ref. 6, each one of the integrals in (2.15) is a linear combination of Fourier transforms of the distributions x_+^λ and x_-^λ.

That is,

$$G_1(k, \omega) = i\frac{2^{2\beta}}{4\Omega}\left[\int_0^\infty du e^{-isu} u^{1-\beta} + (-)^{1-\beta}\int_{-\infty}^0 |u|^{1-\beta} e^{-isu}\, du\right] \times \left[\int_0^\infty v^{-\beta} e^{-irv}\, dv + (-)^{-\beta}\int_{-\infty}^0 |v|^{-\beta} e^{-irv}\, dv\right]$$

$$= i\frac{2^{2\beta}}{4\Omega}[F_s(u_+^{1-\beta}) - e^{\pi i\beta} F_s(u_-^{1-\beta})] \times [F_r(v_+^{-\beta}) + e^{\pi i\beta} F_r(v_-^{-\beta})], \quad (2.16)$$

where

$$s = \bar{k} - \bar{\omega}, \qquad r = \bar{k} + \bar{\omega}.$$

$$F_s(x_+^\lambda) = \lim_{\tau \to 0^+} \int_0^\infty x^\lambda e^{-isx} e^{-\tau x}\, dx = ie^{i\lambda\frac{1}{2}\pi}\Gamma(\lambda + 1)(s + i0)^{-\lambda-1}$$

$$F_s(x_-^\lambda) = \lim_{\tau \to 0^-} \int_{-\infty}^0 |x|^\lambda e^{-isx} e^{-\tau x}\, dx = -ie^{-i\lambda\frac{1}{2}\pi}\Gamma(\lambda + 1)(s - i0)^{-\lambda-1}. \quad (2.17)$$

Using (2.17) and (2.16) becomes

$$G_1(k, \omega) = (2^{2\beta}/\Omega)e^{\pi i\beta} \sin^2 \pi\beta\Gamma(2 - \beta) \times \Gamma(1 - \beta)s^{\beta-2} r^{\beta-1},$$

$$\Gamma(1 - z)\Gamma(z) = (\sin \pi z)^{-1}. \quad (2.18)$$

Finally

$$G_1(k, \omega) = \frac{(-)2^{2\beta}}{\Omega} e^{\pi i\beta} \sin \pi\beta \times \frac{\Gamma(2 - \beta)}{\Gamma(\beta)} \frac{[\bar{k}^2 - \bar{\omega}^2]^{\beta-1}}{[\bar{k} - \bar{\omega}]}. \quad (2.19)$$

As we predicted, $G_1(k, \omega)$ has a branch cut in the complex energy plane but has no other singularity.

[6] I. M. Gel'fand and G. E. Chilov, *Les Distributions* (Dunod Cie., Paris, 1962), especially Chap. 2. paragraph 2. Also E. C. Titchmarsh, Ed., *The Theory of Functions* (Oxford University Press, New York, 1939), in the regularization of the function $\Gamma(\lambda)$.

We therefore conclude that our system does not contain single-particle excitations in its energy spectrum.

Now it is very simple to compare Eq. (2.19) with Eq. (37) of Ref. 5. We can see that both expressions agree if we make the identifications

$$p_0 = \bar{\omega} = (\omega - k_F)/\Omega, \quad p_1 = \bar{k} = k - k_F,$$
$$\alpha = \beta - 1, \quad \Gamma(2-\beta) = (1-\beta)! = -\alpha!,$$
$$\Gamma(\beta) = (\beta - 1)! = \alpha!.$$

The denominator $p_0 - p_1 = [\gamma \cdot \mathbf{p}]_{21}$ results from considering the Green's function $\langle \Phi_H | T\psi_{1H}\psi^+_{1H} | \Phi_H \rangle$ and making use of the identities $\psi_1\psi_1^+ = i\psi_1\bar{\psi}_2$ because $\psi^+ = (i\bar{\psi}_2 - i\bar{\psi}_1)$. We have used the same set of γ matrices as in Ref. 5:

$$\gamma^0 = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix} = i\sigma_1.$$

The dependence of α on the coupling constant λ is slightly different than that in Johnson's work. We are unable at this moment to give a reason for the difference.

3. DEVIATION FROM A δ-FUNCTION POTENTIAL

In this section we are interested in the structure of the Green's function in the case that the interaction differs from the δ-function interaction.

We compute the first-order correction to $G(x, t)$ for the potential

$$V(x) = \delta(x) + (\mu/\lambda)v_1(x), \quad \text{where} \quad v_1(x) = \sin x/x,$$
$$v(p) = 1 + (\mu/\lambda)v_1(p), \quad v_1(p) = \begin{cases} 1 & |p| \le 1 \\ 0 & |p| > 1 \end{cases}. \tag{3.1}$$

According to the previous sections, $G(x, t)$ can be expressed in the form

$$G(x, t) = G_0(x, t)e^{Z(\mu,x,t)}, \tag{3.2}$$

where

$$G_0(x, t) = i\frac{e^{ik_F(x-t)}}{(x-t)} \tag{3.3}$$

and we have made the dependence of Z on μ explicit.

From Eq. (1.34) the expression for $Z(\mu, x, t)$ is

$$\begin{aligned} &Z(\mu, x, t) \\ &= \int_0^\infty \cosh^2\varphi(p) e^{ipx} e^{-ip\omega_p t}\frac{1}{p}\,dp \\ &\quad + \int_0^\infty \sinh^2\varphi(p) e^{-ipx} e^{-ip\omega_p t}\frac{1}{p}\,dp \\ &\quad + it\int_0^\infty (\omega_p - 1)\,dp - \int_0^\infty e^{ip(x-t)}\frac{1}{p}\,dp \\ &\quad - 2\int_0^\infty \sinh^2\varphi(p)\frac{1}{p}\,dp + i\int_0^\infty \frac{\sin p(\omega_p - 1)t}{p}\,dp, \end{aligned} \tag{3.4}$$

$$\begin{aligned} \tanh 2\varphi(p) &= -(\lambda/\pi)v(p), \\ \omega_p = \operatorname{sech} 2\varphi(p) &= \{1 - [\lambda v(p)/\pi]^2\}^{\frac{1}{2}}, \\ \cosh^2\varphi(p) &= (1 + \omega_p)/2\omega_p, \\ \sinh^2\varphi(p) &= (1 - \omega_p)/2\omega_p. \end{aligned}$$

Reordering terms in Eq. (3.4),

$$\begin{aligned} &Z(\mu, x, t) \\ &= \int_0^\infty \cos px \frac{e^{-ip\omega_p t}}{\omega_p}\frac{1}{p}\,dp + i\int_0^\infty \sin pxe^{-ip\omega_p t}\frac{1}{p}\,dp \\ &\quad - \int_0^\infty e^{ip(x-t)}\frac{1}{p}\,dp - \int_0^\infty \frac{1-\omega_p}{\omega_p}\frac{1}{p}\,dp \\ &\quad - i\int_0^\infty \sin p(1-\omega_p)t\frac{1}{p}\,dp + it\int_0^\infty (\omega_p - 1)\,dp, \end{aligned} \tag{3.5}$$

$$G(\mu, x, t) = G(0, x, t) + \frac{\mu}{\lambda}\left[\frac{\partial G}{\partial(\mu/\lambda)}\right]_{\mu/\lambda=0} + \cdots,$$

$$\frac{\partial G}{\partial(\mu/\lambda)} = G_0 e^{Z(\mu,x,t)}\frac{\partial Z}{\partial(\mu/\lambda)},$$

$$\left[\frac{\partial Z}{\partial(\mu/\lambda)}\right]_{\mu/\lambda=0} = \left[\frac{\partial Z}{\partial\omega_p}\frac{\partial\omega_p}{\partial(\mu/\lambda)}\right]_{\mu/\lambda=0}, \tag{3.6}$$

$$\omega_p = \left[1 - \frac{(\lambda + \mu v_1)^2}{\pi^2}\right]^{\frac{1}{2}}, \quad \left[\frac{\partial\omega_p}{\partial(\mu/\lambda)}\right]_{\mu/\lambda=0} = -\frac{\lambda}{\pi}\frac{v_1(p)}{\Omega}, \tag{3.7}$$

$$\begin{aligned} &\left[\frac{\partial Z}{\partial(\mu/\lambda)}\right]_{\mu/\lambda=0} \\ &\quad = \frac{\lambda}{\pi\Omega^3}\int_0^1 \frac{\cos pxe^{-ip\Omega t} - 1}{p}\,dp \\ &\quad + it\frac{\lambda}{\pi\Omega^2}\int_0^1 \cos pxe^{-ip\Omega t}\,dp \\ &\quad - \frac{\lambda}{\pi\Omega}t\int_0^1 \sin pxe^{-ip\Omega t}\,dp \\ &\quad - it\frac{\lambda}{\pi\Omega}\int_0^1 \cos p(1-\Omega t)\,dp + it\int_0^1 \frac{\lambda}{\pi\Omega}\,dp, \end{aligned}$$

$$\begin{aligned} \left[\frac{\partial Z}{\partial(\mu/\lambda)}\right]_{\mu/\lambda=0} &= \frac{\lambda}{2\pi\Omega^3}\int_0^1 \frac{e^{ip(x-\Omega t)} + e^{-ip(x+\Omega t)} - 2}{p}\,dp \\ &\quad + it\frac{\lambda}{2\pi} + \frac{\lambda}{2\pi\Omega^2}t\Big\{(1+\Omega)\frac{e^{i(x-\Omega t)}}{(x-\Omega t)} \\ &\quad + (\Omega - 1)\frac{e^{-i(x+\Omega t)}}{(x+\Omega t)} - \frac{2}{(x-\Omega t)}\Big\} \\ &\quad - i\frac{\lambda}{\pi\Omega(1-\Omega)}\sin(1-\Omega)t, \end{aligned} \tag{3.8}$$

$$\omega(\mu, k, \omega) \simeq \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dt \Big\{ G_0 e^{Z(0,x,t)} + \frac{\mu}{\lambda} G_0 e^{Z(0,x,t)} \Big[\frac{\partial Z}{\partial(\mu/\lambda)} \Big]_{\mu/\lambda=0} \Big\} e^{-ikx} e^{i\omega t}, \quad (3.9)$$

$$\omega(\mu, k, \omega) \simeq G(0, k, \omega) + \frac{\mu}{\lambda} \int_{-\infty}^{\infty} dx \times \int_{-\infty}^{\infty} dt \frac{e^{-ikx} e^{i\omega t} e^{ik_F(x-t)}}{(x - \Omega t)^{\beta} (x + \Omega t)^{\beta-1}} \Big[\frac{\partial Z}{\partial(\mu/\lambda)} \Big]_{\mu/\lambda=0}.$$

Changing variables as in (2.14),

$$\frac{e^{-ikx} e^{i\omega t} e^{ik_F(x-t)}}{(x - \Omega t)^{\beta} (x + \Omega t)^{\beta-1}} = \frac{2}{2^{\beta}} \frac{e^{-iu\rho}}{u^{\beta-1}} \frac{e^{-iv\tau}}{v^{\beta}},$$

$$u = \tfrac{1}{2}(x + \Omega t), \quad v = \tfrac{1}{2}(x - \Omega t), \quad (3.10)$$

$$\rho = k - k_F - [(\omega - k_F)/\Omega],$$

$$\tau = k - k_F + [(\omega - k_F)/\Omega],$$

$$\Big[\frac{\partial G(\mu, k, \omega)}{\partial(\mu/\lambda)} \Big]_{\mu/\lambda=0} = \frac{1}{2^{\beta}\Omega} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv \frac{e^{-iu\rho}}{u^{\beta-1}} \frac{e^{-iv\tau}}{v^{\beta}} \times \Big\{ \frac{\lambda}{2\pi\Omega^3} \int_0^1 \frac{e^{2ivp} - 1}{p} dp + \frac{\lambda}{2\pi\Omega^3} \int_0^1 \frac{e^{-2iup} - 1}{p} dp + \frac{\lambda}{2\pi\Omega^3} (u - v) \Big[(\Omega + 1) \frac{e^{2iv}}{2v} + (\Omega - 1) \frac{e^{-2iu}}{2u} - \frac{1}{v} \Big] - \frac{\lambda}{2\pi\Omega(\Omega - 1)} \Big[\exp\Big[i \frac{(1 - \Omega)}{\Omega} (u - v) \Big] - \exp\Big[-i \frac{(1 - \Omega)}{\Omega} (u - v) \Big] \Big] + i \frac{\lambda}{2\pi\Omega} (u - v) \Big\}. \quad (3.11)$$

We now have to evaluate integrals of the following type:

$$I = \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv \frac{e^{-iu\rho}}{u^{\beta-1}} \frac{e^{-iv\tau}}{v^{\beta}} \int_0^1 \frac{e^{2ipv} - 1}{p} dp. \quad (3.12)$$

Interchanging the order of integration we have

$$I = \int_0^1 \frac{1}{p} dp \int_{-\infty}^{\infty} \frac{e^{-iu\rho}}{u^{\beta-1}} du \int_{-\infty}^{\infty} \frac{e^{-iv(\tau-2p)} - e^{-iv\tau}}{v^{\beta}} dv,$$

which according to the considerations of Sec. II, is equal to

$$I = C\rho^{\beta-2} \int_0^1 \frac{(\tau - 2p)^{\beta-1} - \tau^{\beta-1}}{p} dp. \quad (3.13)$$

The evaluation of the rest of the integrals is now straightforward.

The result is

$$G(\mu, k, \omega) \simeq G(0, k, \omega) + (\mu/\lambda)\{c_1\rho^{\beta-2}F(\tau/2, \beta - 1) + c_2\tau^{\beta-1}F(-\rho/2, \beta - 2) + c_3\rho^{\beta-3}(\tau - 2)^{\beta} + c_4(\rho + 2)^{\beta-2}\tau^{\beta-1} + c_5\rho^{\beta-3}\tau^{\beta} + c_6\rho^{\beta-2}(\tau - 2)^{\beta-1} + c_7(\rho + 2)^{\beta-1}\tau^{\beta-2} + c_8(\rho - 2\beta + 2)^{\beta-2}(\tau + 2\beta - 2)^{\beta-1} + c_9(\rho + 2\beta - 2)^{\beta-2}(\tau - 2\beta + 2)^{\beta-1} + c_{10}\rho^{\beta-3}\tau^{\beta-1}\}. \quad (3.14)$$

The c_m are numerical constants depending on the coupling strength λ, which originate in the evaluation of integrals similar to (2.16).

The function $F(z, \nu)$ is defined by the integral

$$F(z, \nu) = \int_0^1 \frac{(z - t)^{\nu} - z^{\nu}}{t} dt, \quad (3.15)$$

which in the region $|z| < 1$ has the expansion

$$F(z, \nu) = \sum_{m=0}^{\infty} (-)^{\nu-m} \frac{\nu(\nu - 1) \cdots (\nu - m + 1)}{\nu - m} \frac{z^m}{m!}. \quad (3.16)$$

If we write $\nu = -a$, the coefficient of z^m is

$$b_m = \frac{\nu(\nu - 1) \cdots (\nu - m + 1)(-)^{\nu-m}}{\nu - m} = \frac{a(a + 1) \cdots (a + m - 1)}{a + m} \frac{(-)^m(-)^{-a-m}}{(-)},$$

$$b_m = \frac{\Gamma(a + m)}{\Gamma(a)} \cdot \frac{\Gamma(a + m)}{\Gamma(a + m + 1)} = \frac{1}{a} \Big\{ \frac{\Gamma(a + m)}{\Gamma(a)} \frac{\Gamma(a + m)}{\Gamma(a)} \frac{\Gamma(a + 1)}{\Gamma(a + m + 1)} \Big\}. \quad (3.17)$$

Thus, substituting Eq. (15) into Eq. (14) we find

$$F(z, \nu) = F(z, -a) = \frac{1}{a}(-)^{a-1} \times \sum_m \frac{\Gamma(a + m)}{\Gamma(a)} \frac{\Gamma(a + m)}{\Gamma(a)} \frac{\Gamma(a + 1)}{\Gamma(a + m + 1)} \frac{z^m}{m!},$$

which coincides with the hypergeometric series[7]

$$F(z, \nu) = F(z, -a) = \frac{(-)^{a-1}}{a} F(a, a, a + 1, z) = \frac{(-)^{\nu}}{\nu} F(-\nu, -\nu, -\nu + 1, z). \quad (3.18)$$

We see that, even in the case of a more general potential, $G(k, \omega)$ has a cut along the real axis in the complex energy plane. This branch cut is the only

[7] Bateman Manuscript Project, *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I.

singularity of G, confirming our earlier discussion concerning the energy spectrum of our system.

4. CONCLUSIONS

The absence of single-particle excitations in the *one*-particle Green's function, shows that the collective boson excitations completely determine the spectrum of the system.

It was also shown that the solution to the problem of interacting fermions given by Mattis and Lieb coincides with the solution of the relativistic Thiring model in the case of a δ-function potential.

ACKNOWLEDGMENTS

The author would like to thank Professor Daniel C. Mattis for the idea which led to this paper and for his continuous and encouraging advice. Dr. Theumann also wishes to thank Professor Leon F. Landovitz for making available to her his notes on which Sec. 1 is based.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

The Half-Space Green's Function for an Atmosphere with a Polarized Radiation Field*

O. J. SMITH† AND C. E. SIEWERT
Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina

(Received 21 February 1967; revision received 27 April 1967)

This paper presents a rigorous solution for the half-space Green's function of the matrix transport equation that describes the flow of polarized radiation in a free-electron atmosphere. The singular normal modes expansion technique is used to construct the solution in such a manner that the expansion coefficients can be determined by applying the proper boundary conditions. The necessary completeness and orthogonality theorems are proved, and thus all expansion coefficients are found by simply taking scalar products. In addition, the albedo problem for a semi-infinite half space is solved explicitly.

I. INTRODUCTION

IN a recent paper, hereafter referred to as I, Siewer and Fraley found the set of normal modes to the homogeneous matrix transport equation for the radiative-transfer problem in a free-electron atmosphere.[1] Also in I: the half-range completeness theorem was proved, the half-range adjoint functions were presented, the half-range normalization integrals were calculated, and the classical Milne problem was solved. The technique used by Siewert and Fraley was based upon Case's method of singular eigenfunctions that was developed for problems in one-speed neutron transport theory.[2,3] The formulation of the matrix transport equation, as given by Chandrasekhar, was reviewed briefly in I.[4,5] In addition, Kuščer and Ribarič have extended the matrix formulation in the theory of diffusion of light to include more generalized scattering laws.[6] The polarized light problem has been investigated recently by several authors.[7–10]

In this paper we extend the procedures discussed in I and solve explicitly for the half-space Green's function. In addition, the complete solution to the half-space albedo problem is obtained. Our procedure for finding the half-space Green's function is to develop a pseudoinfinite-medium Green's function from which subtractions can be made in order to meet the boundary condition of zero re-entrant radiation. The infinite-medium solution that we use is nonphysical, since it is allowed to diverge at an optical distance of minus infinity. The fact that this infinite-medium solution is nonphysical is a consequence of the conservative nature of our system and causes no real concern, since it is used only as an intermediate

* Based on part of a Ph.D. thesis to be submitted by one of the authors (O. J. S.) to North Carolina State University.
† United States Atomic Energy Commission Predoctoral Fellow.
[1] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[2] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[3] K. M. Case and P. F. Zweifel, *An Introduction to Linear Transport Theory* (Addison-Wesley Publishing Co., Reading, Mass, 1967).
[4] S. Chandrasekhar, Astrophys. J. **103**, 351 (1946).
[5] S. Chandrasekhar, Astrophys. J. **105**, 164 (1947).
[6] I. Kuščer and M. Ribarič, Opt. Acta, **6** 42 (1959).
[7] G. M. Simmons, thesis, Stanford University (1966).
[8] J. V. Dave, Astrophys. J. **140**, 1292 (1964).
[9] J. V. Dave and W. H. Walker, Astrophys. J. **144**, 798 (1966).
[10] T. W. Mullikin, Astrophys. J. **145**, 886 (1966).

singularity of G, confirming our earlier discussion concerning the energy spectrum of our system.

4. CONCLUSIONS

The absence of single-particle excitations in the *one*-particle Green's function, shows that the collective boson excitations completely determine the spectrum of the system.

It was also shown that the solution to the problem of interacting fermions given by Mattis and Lieb coincides with the solution of the relativistic Thiring model in the case of a δ-function potential.

ACKNOWLEDGMENTS

The author would like to thank Professor Daniel C. Mattis for the idea which led to this paper and for his continuous and encouraging advice. Dr. Theumann also wishes to thank Professor Leon F. Landovitz for making available to her his notes on which Sec. 1 is based.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 12 DECEMBER 1967

The Half-Space Green's Function for an Atmosphere with a Polarized Radiation Field*

O. J. SMITH† AND C. E. SIEWERT
Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina

(Received 21 February 1967; revision received 27 April 1967)

This paper presents a rigorous solution for the half-space Green's function of the matrix transport equation that describes the flow of polarized radiation in a free-electron atmosphere. The singular normal modes expansion technique is used to construct the solution in such a manner that the expansion coefficients can be determined by applying the proper boundary conditions. The necessary completeness and orthogonality theorems are proved, and thus all expansion coefficients are found by simply taking scalar products. In addition, the albedo problem for a semi-infinite half space is solved explicitly.

I. INTRODUCTION

IN a recent paper, hereafter referred to as I, Siewer and Fraley found the set of normal modes to the homogeneous matrix transport equation for the radiative-transfer problem in a free-electron atmosphere.[1] Also in I: the half-range completeness theorem was proved, the half-range adjoint functions were presented, the half-range normalization integrals were calculated, and the classical Milne problem was solved. The technique used by Siewert and Fraley was based upon Case's method of singular eigenfunctions that was developed for problems in one-speed neutron transport theory.[2,3] The formulation of the matrix transport equation, as given by Chandrasekhar, was reviewed briefly in I.[4,5] In addition, Kuščer and Ribarič have extended the matrix formulation in the theory of diffusion of light to include more generalized scattering laws.[6] The polarized light problem has been investigated recently by several authors.[7–10]

In this paper we extend the procedures discussed in I and solve explicitly for the half-space Green's function. In addition, the complete solution to the half-space albedo problem is obtained. Our procedure for finding the half-space Green's function is to develop a pseudoinfinite-medium Green's function from which subtractions can be made in order to meet the boundary condition of zero re-entrant radiation. The infinite-medium solution that we use is nonphysical, since it is allowed to diverge at an optical distance of minus infinity. The fact that this infinite-medium solution is nonphysical is a consequence of the conservative nature of our system and causes no real concern, since it is used only as an intermediate

* Based on part of a Ph.D. thesis to be submitted by one of the authors (O. J. S.) to North Carolina State University.
† United States Atomic Energy Commission Predoctoral Fellow.
[1] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[2] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[3] K. M. Case and P. F. Zweifel, *An Introduction to Linear Transport Theory* (Addison-Wesley Publishing Co., Reading, Mass, 1967).
[4] S. Chandrasekhar, Astrophys. J. **103**, 351 (1946).
[5] S. Chandrasekhar, Astrophys. J. **105**, 164 (1947).
[6] I. Kuščer and M. Ribarič, Opt. Acta, **6** 42 (1959).
[7] G. M. Simmons, thesis, Stanford University (1966).
[8] J. V. Dave, Astrophys. J. **140**, 1292 (1964).
[9] J. V. Dave and W. H. Walker, Astrophys. J. **144**, 798 (1966).
[10] T. W. Mullikin, Astrophys. J. **145**, 886 (1966).

step in the construction of the half-space solution. This same procedure has been used for other physical models.[11,12]

In Sec. II we briefly review the cogent results formulated in I. In Sec. III, the necessary full-range completeness theorem is proved, since it is needed for the determination of the pseudo infinite-medium Green's function. Section IV is devoted to the full-range orthogonality theorem and the calculation of the full-range normalization integrals. Finally, in Sec. V the half-space Green's function and the albedo problem are solved explicitly.

II. THE NORMAL MODES OF THE BASIC EQUATION

The equation that mathematically describes the scattering of radiation in a free-electron atmosphere may be written in matrix notation as[1,4]:

$$\mu \frac{\partial}{\partial x} \Psi(x, \mu) + \Psi(x, \mu) = \frac{1}{2} \int_{-1}^{1} \mathbf{x}(\mu, \mu') \Psi(x, \mu') \, d\mu', \tag{1}$$

where Ψ is a vector whose two components represent the parallel and perpendicular states of the polarized radiation field. The transfer matrix is

$$\mathbf{x}(\mu, \mu') = \tfrac{3}{4} \begin{bmatrix} 2(1 - \mu'^2)(1 - \mu^2) + \mu'^2\mu^2 & \mu^2 \\ \mu'^2 & 1 \end{bmatrix}; \tag{2}$$

x is the optical distance measured in units of the Thomson scattering coefficient, and μ is the direction cosine measured from the inward normal to the free surface.[13]

In order to separate the variables in Eq. (1), we seek solutions of the form

$$\Psi(x, \mu) = e^{-x/\eta} \Phi(\eta, \mu). \tag{3}$$

This *ansatz* leads to the following equation for the determination of $\Phi(\eta, \mu)$:

$$(\eta - \mu)\Phi(\eta, \mu) = \frac{\eta}{2} \int_{-1}^{1} \mathbf{x}(\mu, \mu')\Phi(\eta, \mu') \, d\mu'. \tag{4}$$

Siewert and Fraley found the solutions of Eq. (4) to be[1]:

$$\Phi_+ = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{5a}$$

$$\Phi_1(\eta, \mu) = \begin{bmatrix} \frac{3\eta}{2}(1 - \mu^2) \dfrac{P}{\eta - \mu} + \lambda_1(\eta)\delta(\eta - \mu) \\ 0 \end{bmatrix}, \quad \eta \in [-1, 1], \tag{5b}$$

and

$$\Phi_2(\eta, \mu) = \begin{bmatrix} -\frac{3\eta}{2}(\eta + \mu) \\ \frac{3\eta}{2}(1 - \eta^2) \dfrac{P}{\eta - \mu} + \lambda_2(\eta)\delta(\eta - \mu) \end{bmatrix}, \quad \eta \in [-1, 1]. \tag{5c}$$

In addition, they found a solution to Eq. (1) of the form

$$\Psi_-(x, \mu) = (x - \mu) \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{6}$$

In Eqs. (5) and throughout this work, the symbol P denotes that integrals involving these functions are to be carried out in the Cauchy principal value sense. Also $\delta(x)$ is the Dirac delta function,

$$\lambda_1(\eta) = -1 + 3(1 - \eta^2)[1 - \eta T(\eta)], \tag{7a}$$

$$\lambda_2(\eta) = 1 + 3(1 - \eta^2)[1 - \eta T(\eta)], \tag{7b}$$

and $T(x)$ denotes $\tanh^{-1}x$.

Note that Φ_+, $\Phi_1(\eta, \mu)$, and $\Phi_2(\eta, \mu)$ are solutions of Eq. (4); whereas, $\Psi_-(x, \mu)$ is a solution only of Eq. (1). The complete solution to Eq. (1) can thus be written as

$$\Psi(x, \mu) = A_+\Phi_+ + A_-\Psi_-(x, \mu) + \int_{-1}^{1} \alpha(\eta)\Phi_1(\eta, \mu) \, e^{-x/\eta} \, d\eta + \int_{-1}^{1} \beta(\eta) e^{-x/\eta} \Phi_1(\eta, \mu) \, d\eta, \tag{8}$$

where A_+, A_-, $\alpha(\eta)$, and $\beta(\eta)$ are arbitrary expansion coefficients.

Since in later sections we will need it, we state the theorem proved in I regarding the half-range completeness of the elementary solutions, Eqs. (5).

Theorem I: The eigensolutions Φ_+, $\Phi_1(\eta, \mu)$, and $\Phi_2(\eta, \mu)$ are complete on the half range, $\mu \in [0, 1]$, in the sense that an arbitrary two-component vector $\Psi(\mu)$ defined for $0 \leq \mu \leq 1$ can be expanded in the form

$$\Psi(\mu) = A_+\Phi_+ + \int_0^1 \alpha(\eta)\Phi_1(\eta, \mu) \, d\eta + \int_0^1 \beta(\eta)\Phi_2(\eta, \mu) \, d\eta. \tag{9}$$

[11] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) 36, 61 (1966).
[12] N. J. McCormick and I. Kuščer, J. Math. Phys. 6, 1939 (1965).
[13] We choose to measure the velocity vector from the inward normal rather than from the outward one in order that Case's method of normal modes may be directly applicable.

Also in I: the half-range adjoint functions were found to be

$$\mathbf{\Phi}_+^\dagger = \begin{bmatrix} \gamma_1(\mu) \\ -\gamma_2(\mu)(a + b\mu) \end{bmatrix}, \tag{10a}$$

$$\mathbf{\Phi}_1^\dagger(\eta, \mu) = \begin{bmatrix} \gamma_1(\mu)[\frac{3}{2}\eta(1 - \eta^2)\dfrac{P}{\eta - \mu} + \lambda_1(\eta)\delta(\eta - \mu) + \frac{3}{2}\eta(c + \eta)] \\ (15\eta/2b)\gamma_2(\mu) \end{bmatrix}, \tag{10b}$$

and

$$\mathbf{\Phi}_2^\dagger(\eta, \mu) = \begin{bmatrix} (3\eta/2b)\gamma_1(\mu) \\ \gamma_2(\mu)[\frac{3}{2}\eta(1 - \mu^2)\dfrac{P}{\eta - \mu} + \lambda_2(\eta)\delta(\eta - \mu) - \frac{3}{2}\eta(c + \mu)] \end{bmatrix}, \tag{10c}$$

where

$$\gamma_1(\mu) = \mu[X^+(\mu)/\Omega^+(\mu)], \tag{11a}$$

$$\gamma_2(\mu) = \mu[Y^+(\mu)/\Lambda^+(\mu)], \tag{11b}$$

$$\Omega(z) = -1 + 3(1 - z^2)[1 - zT(1/z)], \tag{11c}$$

and

$$\Lambda(z) = 1 + 3(1 - z^2)[1 - zT(1/z)]. \tag{11d}$$

The auxiliary functions $X(z)$ and $Y(z)$ are given by

$$X(z) = \frac{1}{z - 1}\exp\left[\frac{1}{\pi}\int_0^1 \arg\,\Omega^+(\mu)\frac{d\mu}{\mu - z}\right] \tag{12a}$$

and

$$Y(z) = \exp\left[\frac{1}{\pi}\int_0^1 \arg\,\Lambda^+(\mu)\frac{d\mu}{\mu - z}\right]. \tag{12b}$$

Also

$$a = X(1)Y(-1) + X(-1)Y(1), \tag{13a}$$

$$b = X(1)Y(-1) - X(-1)Y(1), \tag{13b}$$

and

$$c = \frac{X(1)Y(-1) + X(-1)Y(1)}{X(1)Y(-1) - X(-1)Y(1)}. \tag{13c}$$

The half-range orthogonality theorem is stated.

Theorem II: The eigensolutions $\mathbf{\Phi}_+$, $\mathbf{\Phi}_1(\eta, \mu)$, and $\mathbf{\Phi}_2(\eta, \mu)$ have corresponding half-range adjoint solutions, $\mathbf{\Phi}_+^\dagger$, $\mathbf{\Phi}_1^\dagger(\eta, \mu)$, and $\mathbf{\Phi}_2^\dagger(\eta, \mu)$, such that

$$\int_0^1 \tilde{\mathbf{\Phi}}^\dagger(\eta', \mu)\mathbf{\Phi}(\eta, \mu)\,d\mu = 0, \quad \eta \neq \eta', \quad \eta \text{ and } \eta' \geq 0. \tag{14}$$

Here the superscript tilde denotes the transpose operation. Using the notation

$$\langle i \mid j\rangle = \int_0^1 \mathbf{\Phi}_i^\dagger(\eta', \mu)\mathbf{\Phi}_j(\eta, \mu)\,d\mu, \quad \text{for } i, j = +, 1, 2, \tag{15}$$

we list the results found previously[1]:

$$\langle i \mid j\rangle = 0 \quad \text{for} \quad i \neq j, \tag{16a}$$

$$\langle + \mid +\rangle = N_+, \tag{16b}$$

$$\langle 1 \mid 1\rangle = \mathcal{S}_1(\eta)\delta(\eta - \eta'), \tag{16c}$$

$$\langle 2 \mid 2\rangle = \mathcal{S}_2(\eta)\delta(\eta - \eta'), \tag{16d}$$

where

$$N_+ = -\tfrac{2}{3}b \tag{17a}$$

and

$$\mathcal{S}_i(\eta) = \gamma_i(\eta)[\lambda_i^2(\eta) + \tfrac{9}{4}\pi^2\eta^2(1 - \eta^2)^2]. \tag{17b}$$

These functions and their relationships will be useful in solving the problems described in a later section.

III. FULL-RANGE COMPLETENESS

We wish to prove the necessary

Theorem III: The eigensolutions $\mathbf{\Phi}_+$, $\mathbf{\Phi}_1(\eta, \mu)$, $\mathbf{\Phi}_2(\eta, \mu)$, and $\mathbf{\Psi}_-(0, \mu)$ are complete on the full-range, $\mu \in [-1, 1]$, in the sense that an arbitrary two-component vector $\mathbf{\Psi}(\mu)$ defined for $-1 \leq \mu \leq 1$ can be expanded in the form

$$\mathbf{\Psi}(\mu) = A_+\mathbf{\Phi}_+ + A_-\mathbf{\Psi}_-(0, \mu) + \int_{-1}^1 \alpha(\eta)\mathbf{\Phi}_1(\eta, \mu)\,d\eta + \int_{-1}^1 \beta(\eta)\mathbf{\Phi}_2(\eta, \mu)\,d\eta. \tag{18}$$

We first investigate the feasibility of expanding an arbitrary function in terms only of the continuum solutions $\mathbf{\Phi}_1(\eta, \mu)$ and $\mathbf{\Phi}_2(\eta, \mu)$. For the theorem to be true, this procedure should lead to restrictions which can be removed only by adding to the expansion the discrete solutions $\mathbf{\Phi}_+$ and $\mathbf{\Psi}_-(0, \mu)$.[14] Thus, we propose

$$\mathbf{\Psi}'(\mu) = \int_{-1}^1 \alpha(\eta)\mathbf{\Phi}_1(\eta, \mu)\,d\eta + \int_{-1}^1 \beta(\eta)\mathbf{\Phi}_2(\eta, \mu)\,d\eta, \tag{19}$$

[14] Simmons has also developed full-range completeness and orthogonality theorems; however, in his formulation the eigenfunctions were not explicitly available, and thus the determination of the expansion coefficients does not follow as readily as here.

or, expanding in terms of the components, we find

$$\Psi_1'(\mu) = \int_{-1}^{1} \alpha(\eta)\left[\tfrac{3}{2}\eta(1-\mu^2)\frac{P}{\eta-\mu} + \lambda_1(\eta)\delta(\eta-\mu)\right] d\eta + \int_{-1}^{1} \beta(\eta)[-\tfrac{3}{2}\eta(\eta+\mu)]\, d\eta \quad (20)$$

and

$$\Psi_2'(\mu) = \int_{-1}^{1} \beta(\eta)\left[\tfrac{3}{2}\eta(1-\eta^2)\frac{P}{\eta-\mu} + \lambda_2(\eta)\delta(\eta-\mu)\right] d\eta. \quad (21)$$

The procedure will be to solve for $\beta(\eta)$ from Eq. (21). Thus

$$\Psi_2'(\mu) = P\int_{-1}^{1} \tfrac{3}{2}\eta(1-\eta^2)\frac{\beta(\eta)}{\eta-\mu}\, d\eta + \lambda_2(\mu)\beta(\mu). \quad (22)$$

Noting the form of $\lambda_2(\mu)$ in Eq. (7b), we introduce the function

$$\Lambda(z) = 1 + 3(1-z^2)\left[1 - \frac{z}{2}\int_{-1}^{1}\frac{d\eta}{z-\eta}\right] \quad (23)$$

and note that it is analytic in the entire complex plane cut on the real line from -1 to 1. Other important properties of $\Lambda(z)$ that will be used later are

$$\Lambda(1) = \Lambda(-1) = 1,$$
$$\lim_{z\to\infty} \Lambda(z) = 2,$$

and $\Lambda(z)$ has no zeros in the complex plane cut from -1 to 1. Defining the boundary values, $\Lambda^+(z)$ and $\Lambda^-(z)$, to be the values of $\Lambda(z)$ as z approaches the real line on the cut from above and below, respectively, we obtain from Cauchy's theorem

$$\Lambda^{\pm}(\eta) = 1 + 3(1-\eta^2)[1 - \eta T(\eta) \pm \pi i\tfrac{1}{2}\eta]. \quad (24)$$

Thus,

$$\tfrac{1}{2}[\Lambda^+(\eta) + \Lambda^-(\eta)] = \lambda_2(\eta) \quad (25a)$$

and

$$\Lambda^+(\eta) - \Lambda^-(\eta) = 3\pi i\eta(1-\eta^2). \quad (25b)$$

We also define

$$N_2(z) \triangleq \frac{1}{2\pi i}\int_{-1}^{1}\eta(1-\eta^2)\frac{\beta(\eta)}{\eta-z}\, d\eta, \quad (26)$$

and observe that

(a) $N_2(z)$ is analytic in the complex plane cut from -1 to 1, and

(b) $N_2(z) \to -z^{-1}$ as $z \to \infty$.

The boundary values of $N_2(z)$ are

$$N_2^{\pm}(\mu) = \frac{1}{2\pi i}P\int_{-1}^{1}\eta(1-\eta^2)\frac{\beta(\eta)}{\eta-\mu}\, d\eta \pm \tfrac{1}{2}[\mu(1-\mu^2)\beta(\mu)]. \quad (27)$$

Thus,

$$N_2^+(\mu) - N_2^-(\mu) = \mu(1-\mu^2)\beta(\mu), \quad (28a)$$

and

$$N_2^+(\mu) + N_2^-(\mu) = \frac{1}{\pi i}P\int_{-1}^{1}\eta\beta(\eta)(1-\eta^2)\frac{d\eta}{\eta-\mu}. \quad (28b)$$

Substituting Eqs. (25) and (28) into Eq. (22), we arrive at the recognizable inhomogeneous Hilbert problem[15]:

$$N_2^+(\mu)\Lambda^+(\mu) - N_2^-(\mu)\Lambda^-(\mu) = \mu(1-\mu^2)\Psi_2'(\mu), \quad (29)$$

which has the solution

$$N_2(z)\Lambda(z) = \frac{1}{2\pi i}\int_{-1}^{1}\mu(1-\mu^2)\frac{\Psi_2'(\mu)}{\mu-z}\, d\mu + P_k(z), \quad (30)$$

where we have determined $N_2(z)\Lambda(z)$ to within an arbitrary polynomial $P_k(z)$. Noting the properties previously described for $\Lambda(z)$, we find that the function

$$N_2(z) = \frac{1}{2\pi i\Lambda(z)}\int_{-1}^{1}\mu(1-\mu^2)\frac{\Psi_2'(\mu)}{\mu-z}\, d\mu \quad (31)$$

does indeed have the properties derived from its definition in Eq. (26). Thus the polynomial of z required to meet our restrictions on $N_2(z)$ is simply zero.

Having determined $N_2(z)$ in terms of the arbitrary function, $\Psi_2'(\mu)$, we can evaluate $\beta(\eta)$ from Eq. (28a).

Now that $\beta(\eta)$ is known, an attempt is made to determine $\alpha(\eta)$ from Eq. (20). Thus

$$\Psi_1'(\mu) = \int_{-1}^{1}\alpha(\eta)\tfrac{3}{2}\eta(1-\mu^2)\frac{P}{\eta-\mu}\, d\eta + \lambda_1(\mu)\alpha(\mu) - g(\mu), \quad (32)$$

where

$$g(\mu) = \frac{3}{2}\int_{-1}^{1}\eta(\eta+\mu)\beta(\eta)\, d\eta. \quad (33)$$

The dispersion function

$$\Omega(z) = -1 + 3(1-z^2)\left[1 - \frac{z}{2}\int_{-1}^{1}\frac{d\eta}{z-\eta}\right] \quad (34)$$

has the properties:

(a) $\Omega(z)$ is analytic in the complex plane cut from -1 to 1, and

(b) $\Omega(z) \to z^{-2}$ as $z \to \infty$.

The boundary values are given by

$$\Omega^{\pm}(\mu) = -1 + 3(1-\mu^2)\left[1 - \frac{\mu}{2}P\int_{-1}^{1}\frac{d\eta}{\mu-\eta} \pm \pi i\frac{\mu}{2}\right]. \quad (35)$$

Thus,

$$\tfrac{1}{2}[\Omega^+(\mu) + \Omega^-(\mu)] = \lambda_1(\mu), \quad (36a)$$

and

$$\Omega^+(\mu) - \Omega^-(\mu) = 3\pi i\mu(1-\mu^2). \quad (36b)$$

[15] N. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

In the same manner used to solve for $\beta(\eta)$, we define

$$N_1(z) = \frac{1}{2\pi i}\int_{-1}^{1}\alpha(\eta)\frac{\eta}{\eta - z}\,d\eta \tag{37}$$

and note that it is analytic in the complex plane cut from -1 to 1 and vanishes as $1/z$ as $z \to \infty$. Also, the boundary values satisfy

$$N_1^+(\mu) - N_1^-(\mu) = \mu\alpha(\mu) \tag{38a}$$

and

$$N_1^+(\mu) + N_1^-(\mu) = \frac{1}{\pi i}P\int_{-1}^{1}\alpha(\eta)\eta\frac{d\eta}{\eta - \mu}. \tag{38b}$$

Inserting Eqs. (36) and (38) into Eq. (32), we obtain another inhomogeneous Hilbert problem

$$N_1^+(\mu)\Omega^+(\mu) - N_1^-(\mu)\Omega^-(\mu) = \mu[\Psi_1'(\mu) + g(\mu)], \tag{39}$$

which has as its solutions, to within an arbitrary polynomial in z,

$$N_1(z) = \frac{1}{2\pi i\Omega(z)}\int_{-1}^{1}\mu[\Psi_1'(\mu) + g(\mu)]\frac{d\mu}{\mu - z}. \tag{40}$$

In the limit as $z \to \infty$,

$$N_1(z) \to -\frac{z}{2\pi i}\int_{-1}^{1}\tfrac{3}{2}\mu[\Psi_1'(\mu) + g(\mu)] \times\left[1 + \frac{\mu}{z} + \left(\frac{\mu}{z}\right)^2 + \cdots\right]d\mu, \tag{41}$$

so that for $N_1(z)$ to vanish as $1/z$, which it must by its definition, the following must apply:

$$\int_{-1}^{1}\mu[\Psi_1'(\mu) + g(\mu)]\,d\mu = 0 \tag{42a}$$

and

$$\int_{-1}^{1}\mu^2[\Psi_1'(\mu) + g(\mu)]\,d\mu = 0. \tag{42b}$$

Thus we cannot find $\alpha(\eta)$ from the arbitrary function $\Psi_1'(\mu)$ without the restrictions on $\Psi_1'(\mu)$ that are indicated by Eqs. (42). To circumvent these restrictions, we assume that the arbitrary vector $\boldsymbol{\Psi}(\mu)$ can be expanded in terms of the continuum plus both discrete modes; i.e.,

$$\boldsymbol{\Psi}(\mu) = \boldsymbol{\Psi}'(\mu) + A_+\boldsymbol{\Phi}_+ + A_-\boldsymbol{\Psi}_-(0, \mu). \tag{43}$$

Writing Eq. (43) explicitly in terms of the components, we have

$$\Psi_i(\mu) = \Psi_i'(\mu) + A_+ - A_-\mu, \quad i = 1, 2. \tag{44}$$

We can thus make use of Eqs. (42) to evaluate A_+ and A_-. The coefficients are found to be

$$A_+ = \frac{3}{2}\int_{-1}^{1}\mu^2[\Psi_1(\mu) + g(\mu)]\,d\mu \tag{45a}$$

and

$$A_- = -\frac{3}{2}\int_{-1}^{1}\mu[\Psi_1(\mu) + g(\mu)]\,d\mu. \tag{45b}$$

The expansion coefficient $\alpha(\eta)$ may be determined from Eq. (38a) to complete the determination of the expansion coefficients for any given $\boldsymbol{\Psi}(\mu)$.[16] The theorem is therefore proved.

IV. FULL-RANGE ORTHOGONALITY AND NORMALIZATION

The full-range orthogonality is stated as

Theorem IV: The eigenvectors $\boldsymbol{\Phi}_+$, $\boldsymbol{\Phi}_1(\eta, \mu)$, and $\boldsymbol{\Phi}_2(\eta, \mu)$ are orthogonal on the full range with respect to the weight function μ, i.e.,

$$\int_{-1}^{1}\mu\tilde{\boldsymbol{\Phi}}(\eta', \mu)\boldsymbol{\Phi}(\eta, \mu)\,d\mu = 0, \quad \eta \neq \eta'. \tag{46}$$

The proof of this theorem follows in the usual manner; i.e., we write the eigenvalue equation for η and η' as[17]

$$\left(1 - \frac{\mu}{\eta}\right)\boldsymbol{\Phi}(\eta, \mu) = \frac{1}{2}\int_{-1}^{1}\mathbf{K}(\mu, \mu')\boldsymbol{\Phi}(\eta, \mu)\,d\mu \tag{47a}$$

and

$$\left(1 - \frac{\mu}{\eta'}\right)\boldsymbol{\Phi}(\eta', \mu) = \frac{1}{2}\int_{-1}^{1}\mathbf{K}(\mu, \mu')\boldsymbol{\Phi}(\eta', \mu)\,d\mu. \tag{47b}$$

Multiplying Eq. (47a) from the left by $\tilde{\boldsymbol{\Phi}}(\eta', \mu)$, transposing Eq. (47b) and then multiplying it by $\boldsymbol{\Phi}(\eta, \mu)$ from the right, integrating over μ from -1 to 1 and subtracting the two equations proves the theorem immediately; i.e.,

$$\left(\frac{1}{\eta} - \frac{1}{\eta'}\right)\int_{-1}^{1}\mu\tilde{\boldsymbol{\Phi}}(\eta', \mu)\boldsymbol{\Phi}(\eta, \mu)\,d\mu = 0. \tag{48}$$

Here we have made use of the fact that

$$\tilde{\mathbf{K}}(\mu, \mu') = \mathbf{K}(\mu', \mu).$$

In this orthogonality theorem, we might expect that there would be a minor complication introduced by the fact that $\boldsymbol{\Phi}_1(\eta, \mu)$ and $\boldsymbol{\Phi}_2(\eta, \mu)$ are degenerate in the sense that they have the same eigenvalue spectrum. It turns out, however, that this is not the case because

$$\int_{-1}^{1}\mu\tilde{\boldsymbol{\Phi}}_1(\eta', \mu)\boldsymbol{\Phi}_2(\eta, \mu)\,d\mu = 0, \tag{49}$$

as can be easily verified. The adjoint vectors for the full range are thus

$$\boldsymbol{\Phi}^\dagger(\eta, \mu) = \mu\boldsymbol{\Phi}(\eta, \mu). \tag{50}$$

We choose to present only the results for the

[16] It is obvious that in order to determine $\alpha(\eta)$ and $\beta(\eta)$ in terms of $\boldsymbol{\Psi}(\mu)$ rather than $\boldsymbol{\Psi}'(\mu)$, one must make the proper substitutions as indicated in Eq. (43).

[17] The vector $\boldsymbol{\Psi}_-(0, \mu)$ is not included in the set of orthogonal functions since it is not a solution of Eqs. (47).

various normalization integrals, since these calculations, although they are very straightforward, are quite tedious. The Poincaré–Bertrand formula as given by Kuščer, McCormick, and Summerfield has been used to specify the technique for handling the double principal value integrals that were encountered.[18] Defining the full-range scalar product[19]

$$\langle i \mid j \rangle \triangleq \int_{-1}^{1} \tilde{\Phi}_i^{\dagger}(\eta', \mu)\Phi_j(\eta, \mu)\, d\mu, \quad i, j = +, 1, 2, \quad (51)$$

we find

$$\langle i \mid j \rangle = 0; \; i \neq j = +, 1, \text{ or } 2, \quad (52a)$$

$$\langle + \mid + \rangle = 0, \quad (52b)$$

$$\langle 1 \mid 1 \rangle = \eta\Omega^{+}(\eta)\Omega^{-}(\eta)\delta(\eta - \eta'), \quad (52c)$$

and

$$\langle 2 \mid 2 \rangle = \eta\Lambda^{+}(\eta)\Lambda^{-}(\eta)\delta(\eta - \eta'). \quad (52d)$$

In addition to the above results, we will need, in the next section, integrals of the form

$$\langle i \mid - \rangle = \int_{-1}^{1} \tilde{\Phi}_i^{\dagger}(\eta, \mu)\Psi_{-}(0, \mu)\, d\mu. \quad (53)$$

We find

$$\langle + \mid - \rangle = -\tfrac{4}{3} \quad (54a)$$

and

$$\langle i \mid - \rangle = 0, \quad i = -, 1, 2. \quad (54b)$$

With all of the necessary formalism and theorems now established, we proceed to solve in the next section the two problems of interest.

V. THE HALF-SPACE GREEN'S FUNCTION AND THE ALBEDO PROBLEM

Now that the completeness and orthogonality theorems have been established for both the full range and the half range, the solutions for the two problems of interest can be constructed with a minimum of manipulation. The solutions to the homogeneous transport equation have already been found, so that the only remaining task is to find the expansion coefficients such that the boundary conditions are satisfied.

We consider the half-space Green's function, where the transport equation takes the form

$$\mu \frac{\partial}{\partial x}\Psi(x, \mu) + \Psi(x, \mu) = \frac{1}{2}\int_{-1}^{1} \kappa(\mu, \mu')\Psi(x, \mu')\, d\mu' + \mathbf{Q}. \quad (55)$$

Here the source term is given by

$$\mathbf{Q} = \delta(x - x_0)\begin{bmatrix} q_0\delta(\mu - \mu_0) \\ q_1\delta(\mu - \mu_1) \end{bmatrix}. \quad (56)$$

[18] I. Kuščer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. (N.Y.) **30**, 411 (1964).

[19] Note that we have used the same symbols here for the full-range adjoint functions and scalar products as were used in Sec. II for the half range.

In order to have complete generality, one might consider that there are two Green's functions: one corresponding to $q_1 = 0$ and $x_0 = x_0$, and the other to $q_0 = 0$ and $x_0 = x_1$. Thus the solution of the transport equation for any given source term or inhomogenity could be constructed from these two Green's functions. However, the Green's function developed here includes these two cases.

We therefore seek a solution to Eq. (1) subject to the following boundary conditions[3,20]:

(a) The $\lim_{x \to \infty} \Psi_g(x_0, \mu_0, \mu_1 \to x, \mu)$ is to be bounded,

(b)

$$\mu\{\Psi_g(x_0^{+}, \mu_0, \mu_1 \to x, \mu) - \Psi_g(x_0^{-}, \mu_0, \mu_1 \to x, \mu)\} = \begin{bmatrix} q_0\delta(\mu - \mu_0) \\ q_1\delta(\mu - \mu_1) \end{bmatrix},$$

and

(c) $\Psi_g(x_0, \mu_0, \mu_1 \to 0, \mu) = 0$ for $\mu > 0$.

We construct a pseudoinfinite-medium Green's function in the forms

$$\chi(x_0, \mu_0, \mu_1 \to x, \mu) = B_{+}\Phi_{+} + \int_0^1 B_1(\eta)e^{-(x-x_0)/\eta}\Phi_1(\eta, \mu)\, d\eta + \int_0^1 B_2(\eta)e^{-(x-x_0)/\eta}\Phi_2(\eta, \mu)\, d\eta, \quad x > x_0; \quad (57a)$$

and

$$\chi(x_0, \mu_0, \mu_1 \to x, \mu) = -B_{-}\Psi_{-}(x - x_0, \mu) - \int_{-1}^0 B_1(\eta)e^{-(x-x_0)/\eta}\Phi_1(\eta, \mu)\, d\eta - \int_{-1}^0 B_2(\eta)e^{-(x-x_0)/\eta}\Phi_2(\eta, \mu)\, d\eta, \quad x < x_0. \quad (57b)$$

We note that $\chi(x_0, \mu_0, \mu_1 \to x, \mu)$ satisfies boundary condition (a); its divergence as x approaches $-\infty$ is of no concern since we will consider only $x \geq 0$. Applying boundary condition (b), we find that the expansion coefficients are to be determined from

$$\Psi(\mu) \triangleq \frac{1}{\mu}\begin{bmatrix} q_0\delta(\mu - \mu_0) \\ q_1\delta(\mu - \mu_1) \end{bmatrix} = B_{+}\Phi_{+} + B_{-}\Psi_{-}(0, \mu) + \int_{-1}^{1} B_1(\eta)\Phi_1(\eta, \mu)\, d\eta + \int_{-1}^{1} B_2(\eta)\Phi_2(\eta, \mu)\, d\eta, \quad \mu \in [-1, 1]. \quad (58)$$

This is simply a full-range expansion in terms of the normal modes. Theorem III therefore is applicable,

[20] Note that we replace the inhomogenity introduced by Q by the equivalent "jump" condition (b).

and the expansion coefficients can be found by taking full-range scalar products, e.g.,

$$B_1(\eta) = \left[\int_{-1}^{1} \tilde{\Phi}_1^{\dagger}(\eta, \mu)\Psi(\mu)\, d\mu\right] \Big/ [\eta\Omega^{+}(\eta)\Omega^{-}(\eta)]. \quad (59)$$

We find

$$B_1(\eta) = \frac{q_0\{\frac{3}{2}\eta(1-\mu_0^2)[P/(\eta-\mu_0)] + \lambda_1(\eta)\delta(\eta-\mu_0)\}}{\eta\Omega^{+}(\eta)\Omega^{-}(\eta)}, \quad (60a)$$

$$B_2(\eta) = \frac{(-q_0\frac{3}{2}\eta(\eta+\mu_0) + q_1\{\frac{3}{2}\eta(1-\eta^2)[P/(\eta-\mu_1)] + \lambda_2(\eta)\delta(\eta-\mu_1)\})}{\eta\Lambda^{+}(\eta)\Lambda^{-}(\eta)}, \quad (60b)$$

$$B_{+} = \tfrac{3}{4}[q_0\mu_0 + q_1\mu_1] \quad (60c)$$

and

$$B_{-} = -\tfrac{3}{4}(q_0 + q_1). \quad (60d)$$

With these expressions for the expansion coefficients, $\chi(x_0, \mu_0, \mu_1 \to x, \mu)$, as given by Eqs. (57), satisfies the first two boundary conditions. We propose that the half-space Green's function can be written in the form

$$\Psi_g(x_0, \mu_0, \mu_1 \to x, \mu) = \chi(x_0, \mu_0, \mu_1 \to x, \mu) - A_{+}\Phi_{+} - \int_0^1 A_1(\eta)e^{-x/\eta}\Phi_1(\eta, \mu)\, d\eta - \int_0^1 A_2(\eta)e^{-x/\eta}\Phi_2(\eta, \mu)\, d\eta. \quad (61)$$

Stipulating the condition of zero re-entrant radiation, boundary condition (c), we obtain the half-range expansion

$$\chi(x_0, \mu_0, \mu_1 \to 0, \mu) = A_{+}\Phi_{+} + \int_0^1 A_1(\eta)\Phi_1(\eta, \mu)\, d\eta + \int_0^1 A_2(\eta)\Phi_2(\eta, \mu)\, d\eta, \quad \mu \in [0, 1]. \quad (62)$$

Since $\chi(x_0, \mu_0, \mu_1 \to 0, \mu)$ is known and the half-range completeness and orthogonality have been established through Theorems I and II, the coefficients A_{+}, $A_1(\eta)$, and $A_2(\eta)$ are found by taking, this time, *half-range* scalar products. Thus

$$A_{+} = \frac{1}{N_{+}}\int_0^1 \tilde{\Phi}_{+}^{\dagger}\chi(x_0, \mu_0, \mu_1 \to 0, \mu)\, d\mu, \quad (63a)$$

$$A_1(\eta) = \frac{1}{\mathcal{S}_1(\eta)}\int_0^1 \tilde{\Phi}_1^{\dagger}(\eta, \mu)\chi(x_0, \mu_0, \mu_1 \to 0, \mu)\, d\mu, \quad (63b)$$

and

$$A_2(\eta) = \frac{1}{\mathcal{S}_2(\eta)}\int_0^1 \tilde{\Phi}_2^{\dagger}(\eta, \mu)\chi(x_0, \mu_0, \mu_1 \to 0, \mu)\, d\mu. \quad (63c)$$

Although the explicit evaluation of the scalar products indicated above is a tedious task, it is a straightforward one. We illustrate the procedure by further developing the expression for A_{+}.

Inspection of Eq. (57b) shows that for $x = 0$ it can be written as

$$\chi(x_0, \mu_0, \mu_1 \to 0, \mu) = x_0 B_{-}\Phi_{+} - B_{-}\Psi_{-}(0, \mu) - \int_0^1 B_1(-\eta)e^{-x_0/\eta}\Phi_1(-\eta, \mu)\, d\eta - \int_0^1 B_2(-\eta)e^{-x_0/\eta}\Phi_2(-\eta, \mu)\, d\eta. \quad (64)$$

It is at once apparent that upon taking the scalar product of Eq. (64) with $\Phi_{+}^{\dagger}$, we encounter integrals similar to those that we have already discussed in the section on half-range normalization. Therefore, in order to proceed, we must evaluate integrals of the type

$$M_{ij}(\eta', \eta) \triangleq \int_0^1 \tilde{\Phi}_i^{\dagger}(\eta', \mu)\Phi_j(-\eta, \mu)\, d\mu, \quad \eta, \eta' \in [0, 1], \quad i = +1, 2 \text{ and } j = 1, 2. \quad (65)$$

We note that for η and $\mu \in [0, 1]$, $\Phi_j(-\eta, \mu)$ is not singular; the complexity of the above integrals is thereby greatly reduced. We find

$$M_{+1}(\eta', \eta) = \eta X(-\eta), \quad (66a)$$

$$M_{+2}(\eta', \eta) = \eta(b\eta - a)Y(-\eta), \quad (66b)$$

$$M_{11}(\eta', \eta) = \tfrac{3}{2}\eta\eta'[X(-\eta)/(\eta + \eta')] \times [\eta\eta' + c(\eta + \eta') + 1], \quad (66c)$$

$$M_{12}(\eta', \eta) = (15/2b)\eta\eta' Y(-\eta), \quad (66d)$$

$$M_{21}(\eta', \eta) = (3/2b)\eta\eta' X(-\eta), \quad (66e)$$

and

$$M_{22}(\eta', \eta) = \tfrac{3}{2}\eta\eta'[Y(-\eta)/(\eta + \eta')] \times [\eta\eta' - c(\eta + \eta') + 1]. \quad (66f)$$

In addition, the determination of A_{+}, $A_1(\eta)$, and $A_2(\eta)$ necessitates the evaluation of the three integrals

$$M_{i-}(\eta') \triangleq \int_0^1 \tilde{\Phi}_i^{\dagger}(\eta', \mu)\Psi_{-}(0, \mu)\, d\mu, \quad \eta' \in [0, 1] \text{ and } i = +, 1, 2. \quad (67)$$

These same integrals were encountered in the solution to the Milne problem in I. It was found there that

$$M_{+-}(\eta) = -N_{+}z_0 = -N_{+}\left\{c + \tfrac{1}{2}[Y(1) - Y(-1)] + \frac{3}{4}\int_0^1 \frac{\mu^3}{Y(-\mu)}\, d\mu\right\}, \quad (68a)$$

$$M_{1-}(\eta) = -5\eta/b, \quad (68b)$$

and

$$M_{2-}(\eta) = -\eta(\eta - c). \tag{68c}$$

Here z_0 is the Milne problem extrapolation distance.

The solution for the expansion coefficient A_+ can now be expressed in terms of known functions; we find

$$A_+ = B_-(x_0 + z_0) + \frac{3}{2b}\int_0^1 B_1(-\eta)e^{-x_0/\eta}\eta X(-\eta)\,d\eta + \frac{3}{2b}\int_0^1 B_2(-\eta)e^{-x_0/\eta}\eta(b\eta - a)Y(-\eta)\,d\eta. \tag{69}$$

The coefficients $A_1(\eta)$ and $A_2(\eta)$ can be determined in a completely analogous manner. For the sake of brevity we present only the final results. We find

$$\Psi_g(x_0, \mu_0, \mu_1 \to x, \mu) = T_+\Phi_+ + \int_0^1 T_1(\eta)e^{-x/\eta}\Phi_1(\eta, \mu)\,d\eta + \int_0^1 T_2(\eta)e^{-x/\eta}\Phi_2(\eta, \mu)\,d\eta, \quad x > x_0, \tag{70a}$$

where

$$T_+ = \tfrac{3}{4}[q_0\mu_0 + q_1\mu_1 + (q_0 + q_1)(x_0 + z_0)] - \frac{3}{2b}\int_0^1 B_1(-\eta)e^{-x_0/\eta}\eta X(-\eta)\,d\eta - \frac{3}{2b}\int_0^1 B_2(-\eta)e^{-x_0/\eta}\eta(b\eta - a)Y(-\eta)\,d\eta, \tag{70b}$$

$$T_1(\eta) = B_1(\eta)e^{x_0/\eta} - A_1(\eta), \tag{70c}$$

and

$$T_2(\eta) = B_2(\eta)e^{x_0/\eta} - A_2(\eta). \tag{70d}$$

For $x < x_0$, we obtain

$$\Psi_g(x_0, \mu_0, \mu_1 \to x, \mu) = Q_+\Phi_+ - \int_0^1 [B_1(-\eta)e^{-(x_0-x)/\eta}\Phi_1(-\eta, \mu) + A_1(\eta)e^{-x/\eta}\Phi_1(\eta, \mu)]\,d\eta - \int_0^1 [B_2(-\eta)e^{-(x_0-x)/\eta}\Phi_2(-\eta, \mu) + A_2(\eta)e^{-x/\eta}\Phi_2(\eta, \mu)]\,d\eta, \quad x < x_0, \tag{71a}$$

where

$$Q_+ = \tfrac{3}{4}(q_0 + q_1)(x + z_0 - \mu) - \frac{3}{2b}\int_0^1 B_1(-\eta)e^{-x_0/\eta}\eta X(-\eta)\,d\eta - \frac{3}{2b}\int_0^1 B_2(-\eta)e^{-x_0/\eta}\eta(b\eta - a)Y(-\eta)\,d\eta, \tag{71b}$$

$$A_1(\eta) = -\frac{1}{S_1(\eta)}\left[\frac{15\eta}{4b}(q_0 + q_1) + \int_0^1 B_1(-t)e^{-x_0/t}M_{11}(\eta, t)\,dt + \int_0^1 B_2(-t)e^{-x_0/t}M_{12}(\eta, t)\,dt\right], \tag{71c}$$

and

$$A_2(\eta) = -\frac{1}{S_2(\eta)}\left[\tfrac{3}{4}(q_0 + q_1)\eta(\eta - c) + \int_0^1 B_1(-t)e^{-x_0/t}M_{21}(\eta, t)\,dt + \int_0^1 B_2(-t)e^{-x_0/t}M_{22}(\eta, t)\,dt\right]. \tag{71d}$$

The solution to the half-space Green's function corresponding to the source **Q** is thus complete.

The fact that the solution is complicated might have been anticipated, since the class of problems for which it could be used to generate solutions is a very broad one.

Let us now consider the somewhat simpler albedo problem.[21] Here we seek a solution to Eq. (1) subject to the following boundary conditions:

(i) $\Psi_a(0, \mu) = \begin{bmatrix} s_0\delta(\mu - \mu_0) \\ s_1\delta(\mu - \mu_1) \end{bmatrix}$ for μ, μ_0, and $\mu_1 \geq 0$

and

(ii) $\Psi_a(x, \mu)$ must remain finite as x increases without bound.

We immediately write the normal modes that satisfy condition (ii); thus

$$\Psi_a(x, \mu) = A_+\Phi_+ + \int_0^1 \alpha(\eta)e^{-x/\eta}\Phi_1(\eta, \mu)\,d\eta + \int_0^1 \beta(\eta)e^{-x/\eta}\Phi_2(\eta, \mu)\,d\eta. \tag{72}$$

In order to determine the expansion coefficients, we apply condition (i) to obtain

$$\begin{bmatrix} s_0\delta(\mu - \mu_0) \\ s_1\delta(\mu - \mu_1) \end{bmatrix} = A_+\Phi_+ + \int_0^1 \alpha(\eta)\Phi_1(\eta, \mu)\,d\eta + \int_0^1 \beta(\eta)\Phi_2(\eta, \mu)\,d\eta, \quad \mu > 0. \tag{73}$$

Taking half-range scalar products of Eq. (73), we find the following results for the expansion coefficients:

$$A_+ = \frac{3}{2b}[s_1\gamma_2(\mu_1)(a + b\mu_1) - s_0\gamma_1(\mu_0)], \tag{74a}$$

$$\alpha(\eta) = [S_1(\eta)]^{-1}\{s_0\gamma_1(\mu_0)[\tfrac{3}{2}\eta(1 - \eta^2)[P/(\eta - \mu_0)] + \lambda_1(\eta)\delta(\eta - \mu_0) + \tfrac{3}{2}\eta(c + \eta)] + s_1(15\eta/2b)\gamma_2(\mu_1)\}, \tag{74b}$$

and

$$\beta(\eta) = [S_2(\eta)]^{-1}\{s_0(3\eta/2b)\gamma_1(\mu_0) + s_1\gamma_2(\mu_1) \times [\tfrac{3}{2}\eta(1 - \mu_1^2)[P/(\eta - \mu_1)] + \lambda_2(\eta)\delta(\eta - \mu_1) - \tfrac{3}{2}\eta(c + \mu_1)]\}. \tag{74c}$$

With the expansion coefficients thus determined, the solution is completed.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Professor R. L. Murray for suggestions offered during the preparation of the manuscript.

[21] Although the albedo problem was sketched in I, the explicit results were not given there.