JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Generalized Addition Theorem for Spherical Harmonics*

ROGER E. CLAPP
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts

(Received 11 August 1967)

In analogy with the familiar addition formula for Legendre polynomials, a generalized addition theorem is proved. A general spherical harmonic, depending on the two angles θ_1 and ϕ_1, is expressed as an expansion involving spherical-harmonic functions of (θ, ϕ) and of (θ', ϕ'). The six angles are related to each other through the equations

$$\cos\theta_1 = \cos\theta\cos\theta' - \sin\theta\sin\theta'\cos(\phi' - \phi\cos\theta),$$
$$\sin\theta_1\cos(\phi_1 - \phi) = \sin\theta\cos\theta' + \cos\theta\sin\theta'\cos(\phi' - \phi\cos\theta),$$
$$\sin\theta_1\sin(\phi_1 - \phi) = \sin\theta'\sin(\phi' - \phi\cos\theta).$$

The expansion is then used in the proof of an integral theorem for spherical harmonics.

1. INTRODUCTION

The familiar addition formula for Legendre polynomials[1] can be written as

$$P_n(\cos\theta') = P_n(\cos\theta_1)P_n(\cos\theta) + 2\sum_{\mu=1}^{\mu=n}\frac{(n-\mu)!}{(n+\mu)!}P_n^\mu(\cos\theta_1) \times P_n^\mu(\cos\theta)\cos\mu(\phi_1 - \phi), \quad (1.1)$$

where

$$\cos\theta' = \cos\theta_1\cos\theta + \sin\theta_1\sin\theta\cos(\phi_1 - \phi). \quad (1.2)$$

The angle θ' can be interpreted as the included angle between the two vectors $\mathbf{r}_1$ and $\mathbf{r}$, whose directions are specified by (θ_1, ϕ_1) and (θ, ϕ), respectively.

The Legendre polynomials are a subset of the more general spherical harmonics. Equation (1.1) is a relationship between members of such a subset, on the left-hand side, and members of two full sets, on the right-hand side. It is to be expected that other similar relationships exist, in which a full set of spherical harmonics, involving θ' and a suitably defined ϕ', is related to the two full sets of spherical harmonics involving (θ_1, ϕ_1) and (θ, ϕ).

It is first necessary to find a suitable definition for the angle ϕ' which is to accompany θ'. A simple definition is found, and the reason for its selection is explained. An equation of the form (1.1) is then given, in which $P_n(\cos\theta_1)$ is expressed as a summation over spherical harmonics which are functions of $\theta, \phi, \theta', \phi'$. Finally, this equation is generalized so that a general spherical harmonic, written as a function of θ_1 and ϕ_1, is then given as a summation over spherical-harmonic functions involving $\theta, \phi, \theta', \phi'$.

2. DEFINITION OF ϕ'

The vectors $\mathbf{r}$ and $\mathbf{r}_1$, with polar-coordinate components (r, θ, ϕ) and (r_1, θ_1, ϕ_1), are illustrated in Fig. 1. These vectors are shown with respect to a rectangular coordinate system with axes labeled by x, y, and z.

A transformation is needed which will refer the vector $\mathbf{r}_1$ to a new coordinate system whose polar axis lies along $\mathbf{r}$. The transformation to be used is similar to an Euler-angle transformation, but with an important difference which avoids indeterminacies.

Figure 2 shows a new set of coordinate axes, labeled x', y', and z', which are obtained from the first set by two rotations. The first is a rotation about

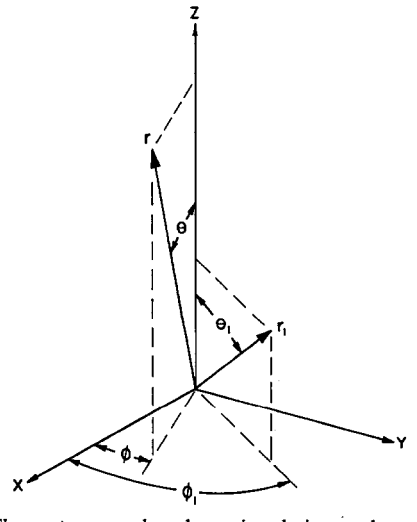


FIG. 1. The vectors $\mathbf{r}$ and $\mathbf{r}_1$ shown in relation to the x, y, z axes.

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., 1953), p. 1274.

the z axis by the angle ϕ; this rotation moves the y direction into the position of the y' axis, which thus must lie in the (x, y) plane. The second rotation is about the y' axis and moves the z direction down by the angle θ, until it lies in the position of the z' axis, which lies along the vector $\mathbf{r}$. In this way, by the two rotations through the angles ϕ and θ, the x, y, z directions are moved into the positions shown as the x', y', z' axes.

Figure 2 thus defines the z' axis unequivocally, since this axis must lie along the direction of the vector $\mathbf{r}$. The y' axis will also be defined uniquely, if θ is greater than zero and less than π. The plane through the origin of coordinates which is perpendicular to $\mathbf{r}$ will then intersect the (x, y) plane in a straight line which contains the y' axis. The direction of this axis can then be obtained, as shown in Fig. 2, through the application of the right-hand rule. If $\mathbf{a}_z$ is a unit vector in the direction of the z axis while $\mathbf{r}/r$ is a unit vector in the direction of the z' axis, then the unit vector in the y' direction is

$$\mathbf{a}_{y'} = \frac{\mathbf{a}_z \times \mathbf{r}}{r \sin \theta}. \tag{2.1}$$

If the vector $\mathbf{r}$ happens to be parallel to the z axis so that $\sin \theta$ is equal to zero and the vector product $\mathbf{a}_z \times \mathbf{r}$ is also equal to zero, then the direction of the y' axis becomes indeterminate. Since any value of ϕ can be used in specifying the direction of $\mathbf{r}$ when θ is equal to zero or to π, it is apparent from Fig. 2 that the y' axis can be directed in any direction within the (x, y) plane.

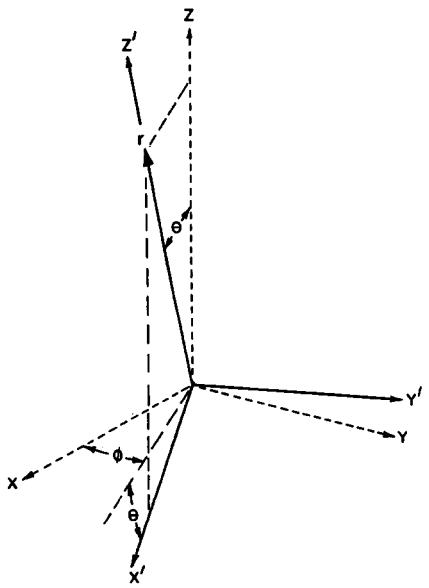


FIG. 2. The relationship between the x', y', z' axes and the x, y, z axes.

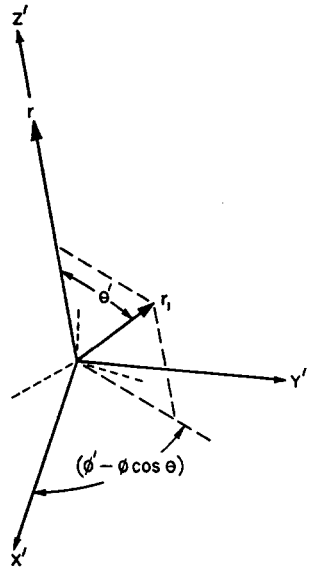


FIG. 3. The relationship between the vector $\mathbf{r}_1$ and the x', y', z' axes.

It is essential that provision be made for this indeterminacy in the orientation of the y' axis (and the corresponding indeterminacy in the orientation of the x' axis) when the angle ϕ' is defined, to accompany the angle θ'.

What is desired is that the definition of ϕ' should permit the specification of the direction of the vector $\mathbf{r}_1$ in terms of the direction of the vector $\mathbf{r}$ and the direction angles θ' and ϕ'. In the ordinary situation, in which $\mathbf{r}$ is not parallel or antiparallel to the z direction, the y' axis is uniquely defined and no problem arises. However, when $\mathbf{r}$ lies in the positive or negative z direction, the angle ϕ is not well defined, and it becomes difficult to give the azimuth angle ϕ_1 for the vector $\mathbf{r}_1$ in terms of the ill-defined azimuth angle ϕ of the vector $\mathbf{r}$.

The solution of this problem of azimuth indeterminacy is shown in Fig. 3. With respect to the x', y', z' axes, the direction of the vector $\mathbf{r}_1$ is specified by the polar angle θ' and by an azimuth angle which is given by the expression $\phi' - \phi \cos \theta$. When $\mathbf{r}$ is not parallel (or antiparallel) to the z axis, then the use of $\phi' - \phi \cos \theta$ instead of ϕ' represents a simple displacement of the azimuth-angle coordinate. However, in the exceptional cases where $\mathbf{r}$ is parallel or antiparallel to the z axis, the choice of $\phi' - \phi \cos \theta$ provides an unequivocal specification of the azimuth angle ϕ_1, even when ϕ itself, and ϕ' accordingly, are both indeterminate.

3. UNIT VECTORS

In terms of the rectangular set of unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ directed along the x, y, z axes, there are three polar-coordinate unit vectors associated with the

vector $\mathbf{r}$; these are

$$\begin{aligned}\mathbf{a}_r &= \mathbf{i}\sin\theta\cos\phi + \mathbf{j}\sin\theta\sin\phi + \mathbf{k}\cos\theta,\\ \mathbf{a}_\theta &= \mathbf{i}\cos\theta\cos\phi + \mathbf{j}\cos\theta\sin\phi - \mathbf{k}\sin\theta,\\ \mathbf{a}_\phi &= -\mathbf{i}\sin\phi + \mathbf{j}\cos\phi.\end{aligned} \tag{3.1}$$

A similar set of unit vectors is associated with the vector $\mathbf{r}_1$:

$$\begin{aligned}\mathbf{a}_{1r} &= \mathbf{i}\sin\theta_1\cos\phi_1 + \mathbf{j}\sin\theta_1\sin\phi_1 + \mathbf{k}\cos\theta_1,\\ \mathbf{a}_{1\theta} &= \mathbf{i}\cos\theta_1\cos\phi_1 + \mathbf{j}\cos\theta_1\sin\phi_1 - \mathbf{k}\sin\theta_1,\\ \mathbf{a}_{1\phi} &= -\mathbf{i}\sin\phi_1 + \mathbf{j}\cos\phi_1.\end{aligned} \tag{3.2}$$

The primed axes in Fig. 2 have been chosen to lie in the directions of the unit vectors (3.1), so that the primed unit vectors $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ are given by

$$\mathbf{i}' = \mathbf{a}_\theta, \quad \mathbf{j}' = \mathbf{a}_\phi, \quad \mathbf{k}' = \mathbf{a}_r. \tag{3.3}$$

The unit vector which is parallel to the vector $\mathbf{r}_1$ can be written as $\mathbf{a}_{1r}$, in the form given in (3.2), but it can also be expressed in terms of the primed axes:

$$\begin{aligned}\mathbf{a}_{1r} = {}& \mathbf{i}'\sin\theta'\cos(\phi' - \phi\cos\theta)\\ &+ \mathbf{j}'\sin\theta'\sin(\phi' - \phi\cos\theta) + \mathbf{k}'\cos\theta'.\end{aligned} \tag{3.4}$$

Equations (3.1)–(3.4) yield the trigonometric transformations

$$\cos\theta_1 = \cos\theta\cos\theta' - \sin\theta\sin\theta'\cos(\phi' - \phi\cos\theta), \tag{3.5a}$$

$$\begin{aligned}&\sin\theta_1\cos(\phi_1 - \phi)\\ &\quad = \sin\theta\cos\theta' + \cos\theta\sin\theta'\cos(\phi' - \phi\cos\theta),\end{aligned} \tag{3.5b}$$

$$\sin\theta_1\sin(\phi_1 - \phi) = \sin\theta'\sin(\phi' - \phi\cos\theta), \tag{3.5c}$$

and the inverse transformations

$$\cos\theta' = \cos\theta_1\cos\theta + \sin\theta_1\sin\theta\cos(\phi_1 - \phi), \tag{3.6a}$$

$$\begin{aligned}&\sin\theta'\cos(\phi' - \phi\cos\theta)\\ &\quad = -\cos\theta_1\sin\theta + \sin\theta_1\cos\theta\cos(\phi_1 - \phi),\end{aligned} \tag{3.6b}$$

$$\sin\theta'\sin(\phi' - \phi\cos\theta) = \sin\theta_1\sin(\phi_1 - \phi). \tag{3.6c}$$

4. ADDITION THEOREM

The addition formula (1.1) is based on (3.6a), but a similar formula can readily be written down, based on (3.5a):

$$\begin{aligned}P_n(\cos\theta_1) = {}& P_n(\cos\theta)P_n(\cos\theta')\\ &+ 2\sum_{\mu=1}^{\mu=n}\frac{(n-\mu)!}{(n+\mu)!}P_n^\mu(\cos\theta)P_n^\mu(\cos\theta')(-1)^\mu\\ &\times\cos(\mu(\phi' - \phi\cos\theta)).\end{aligned} \tag{4.1}$$

It is this formula which will be generalized into an addition theorem applicable to any spherical harmonic $X_n^m(\theta_1, \phi_1)$ and not just to the Legendre polynomial $P_n(\cos\theta_1)$.

In the notation of Morse and Feshbach,[2] a general spherical harmonic can be defined by

$$X_n^m(\theta_1, \phi_1) = e^{im\phi_1}P_n^{|m|}(\cos\theta_1), \tag{4.2}$$

where

$$P_n^{|m|}(\cos\theta_1) = (\sin\theta_1)^{|m|}\frac{d^{|m|}}{d(\cos\theta_1)^{|m|}}P_n(\cos\theta_1). \tag{4.3}$$

It will be assumed that θ, ϕ, θ', ϕ' are independent variables, while θ_1 and ϕ_1 are the dependent functions given by (3.5). Two operators $D_{\theta,\phi}^+$ and $D_{\theta,\phi}^-$ will be defined by

$$D_{\theta,\phi}^\pm = \left[\frac{-1}{\sin\theta}\frac{\partial}{\partial\theta} - \frac{\phi}{\cos\theta}\frac{\partial}{\partial\phi} \mp \frac{i}{\sin^2\theta\cos\theta}\frac{\partial}{\partial\phi}\right]. \tag{4.4}$$

In terms of these operators, the spherical harmonic (4.2) can be written as

$$X_n^m(\theta_1, \phi_1) = e^{im\phi}\sin^{|m|}\theta\,(D_{\theta,\phi})^m P_n(\cos\theta_1), \tag{4.5}$$

in which the operator expression $(D_{\theta,\phi})^m$ has the meaning $(D_{\theta,\phi}^+)^{|m|}$ when m is positive, $(D_{\theta,\phi}^-)^{|m|}$ when m is negative. The proof of (4.5) is based on the easily verified results,

$$D_{\theta,\phi}^+\cos\theta_1 = \frac{e^{i\phi_1}\sin\theta_1}{e^{i\phi}\sin\theta}, \tag{4.6a}$$

$$(D_{\theta,\phi}^+)^2\cos\theta_1 = 0, \tag{4.6b}$$

and the complex conjugate equations involving $D_{\theta,\phi}^-$. It is apparent also that

$$X_n^m(\theta, \phi) = e^{im\phi}\sin^{|m|}\theta\,(D_{\theta,\phi})^m P_n(\cos\theta). \tag{4.7}$$

When $P_n(\cos\theta_1)$ on the right-hand side of (4.5) is replaced by the series expression (4.1), the result is the generalized addition theorem:

$$\begin{aligned}X_n^m(\theta_1, \phi_1) = {}& X_n^m(\theta, \phi)P_n(\cos\theta')\\ &+ 2\sum_{\mu=1}^{\mu=n}\frac{(n-\mu)!}{(n+\mu)!}(-1)^\mu P_n^\mu(\cos\theta')\\ &\times e^{im\phi}\sin^{|m|}\theta\,(D_{\theta,\phi})^m\\ &\times\{P_n^\mu(\cos\theta)\cos[\mu(\phi' - \phi\cos\theta)]\}.\end{aligned} \tag{4.8}$$

Equation (4.8), together with the transformation equations (3.5), constitutes the desired generalization of (1.1) and (1.2). The use of the azimuth-angle

[2] Ref. 1, p. 1898.

expressions $\phi_1 - \phi$ and $\phi' - \phi\cos\theta$ is a necessary complication, without which the important equation (4.5) could not be established. However, once a choice of (n, m) has been made and the operations in (4.8) carried out, the angle $\phi' - \phi\cos\theta$ can be replaced by a redefined ϕ', here and in Fig. 3, if this is desired in a particular application of the formula. The replacement cannot be introduced until after the operations $(D_{\theta,\phi})^m$ have been completed.

5. INTEGRAL THEOREM

The addition theorem (4.8) leads directly to an integral theorem. Equation (4.8) expresses the spherical harmonic $X_n^m(\theta_1, \phi_1)$ as a function of four independent variables θ, ϕ, θ', ϕ'. The dependence upon ϕ' is particularly simple as can be seen when (4.8) is written in the form

$$X_n^m(\theta_1, \phi_1) = X_n^m(\theta, \phi)P_n(\cos\theta') + \sum_{\mu=1}^{\mu=n}[A_c\cos(\mu\phi') + A_s\sin(\mu\phi')], \quad (5.1)$$

where the coefficients A_c and A_s depend upon the variables θ, ϕ, θ' and the parameters n, m, μ, but not upon ϕ'.

Equation (5.1) can now be integrated with respect to ϕ', with the other three variables held constant. Each term in the summation over μ integrates to zero, leaving only

$$\int_{\phi'=0}^{\phi'=2\pi} X_n^m(\theta_1, \phi_1)\,d\phi' = 2\pi X_n^m(\theta, \phi)P_n(\cos\theta'). \quad (5.2)$$

Other related integral theorems are given in a following paper.[3]

ACKNOWLEDGMENTS

The author is grateful to Dr. Luc Huang for an important suggestion early in this analysis, and for a critical reading of the manuscript. The author is also grateful to Prof. A. Erdélyi for guidance to the literature of addition theorems based on Euler-angle transformations.

[3] R. E. Clapp and H. T. Li, J. Math. Phys. **11**, 4 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Six Integral Theorems for Vector Spherical Harmonics*

ROGER E. CLAPP
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts
AND
H. TUNG LI
AirTechnology Corporation, Waltham, Massachusetts
and
College of the Holy Cross, Worcester, Massachusetts

(Received 11 August 1967)

With the aid of a generalized addition theorem for spherical harmonics [J. Math. Phys. **11**, 1 (1970), preceding paper], six integral theorems for vector spherical harmonics are proved. A source-point direction (θ_1, ϕ_1) is first expressed in terms of a field-point direction (θ, ϕ) and a polar-coordinate angle-pair (θ', ϕ'), which has as its polar axis the field-point direction. For a particular choice of n and m, all the components of the vector spherical harmonics for the source, expressed in terms of (θ_1, ϕ_1), are integrated over the relative azimuth angle ϕ' while the field-point direction (θ, ϕ) and the relative polar angle θ' are held fixed. The result in each case is a spherical harmonic or vector spherical harmonic of the field-point direction, with the same n and m but now depending on (θ, ϕ) instead of (θ_1, ϕ_1), multiplied by an explicit function of the relative polar angle θ'.

1. INTRODUCTION

In the preceding article,[1] a generalized addition theorem for spherical harmonics was established. This addition theorem led directly to the proof of an integral theorem for spherical harmonics. In the integration, a vector $\mathbf{r}_1$ is swung about a fixed vector $\mathbf{r}$, with a constant angle θ' separating the two vectors. A suitably defined azimuth angle ϕ' locates the azimuth of $\mathbf{r}_1$ in its motion about $\mathbf{r}$. This integration, over a full circuit, gave the result

$$\int_{\phi'=0}^{\phi'=2\pi} X_n^m(\theta_1, \phi_1)\,d\phi' = 2\pi X_n^m(\theta, \phi)P_n(\cos\theta'), \quad (1.1)$$

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.
[1] R. E. Clapp, J. Math. Phys. **11**, 1 (1970).

expressions $\phi_1 - \phi$ and $\phi' - \phi\cos\theta$ is a necessary complication, without which the important equation (4.5) could not be established. However, once a choice of (n, m) has been made and the operations in (4.8) carried out, the angle $\phi' - \phi\cos\theta$ can be replaced by a redefined ϕ', here and in Fig. 3, if this is desired in a particular application of the formula. The replacement cannot be introduced until after the operations $(D_{\theta,\phi})^m$ have been completed.

5. INTEGRAL THEOREM

The addition theorem (4.8) leads directly to an integral theorem. Equation (4.8) expresses the spherical harmonic $X_n^m(\theta_1, \phi_1)$ as a function of four independent variables θ, ϕ, θ', ϕ'. The dependence upon ϕ' is particularly simple as can be seen when (4.8) is written in the form

$$X_n^m(\theta_1, \phi_1) = X_n^m(\theta, \phi)P_n(\cos\theta') + \sum_{\mu=1}^{\mu=n}[A_c\cos(\mu\phi') + A_s\sin(\mu\phi')], \quad (5.1)$$

where the coefficients A_c and A_s depend upon the variables θ, ϕ, θ' and the parameters n, m, μ, but not upon ϕ'.

Equation (5.1) can now be integrated with respect to ϕ', with the other three variables held constant. Each term in the summation over μ integrates to zero, leaving only

$$\int_{\phi'=0}^{\phi'=2\pi} X_n^m(\theta_1, \phi_1)\,d\phi' = 2\pi X_n^m(\theta, \phi)P_n(\cos\theta'). \quad (5.2)$$

Other related integral theorems are given in a following paper.[3]

ACKNOWLEDGMENTS

The author is grateful to Dr. Luc Huang for an important suggestion early in this analysis, and for a critical reading of the manuscript. The author is also grateful to Prof. A. Erdélyi for guidance to the literature of addition theorems based on Euler-angle transformations.

[3] R. E. Clapp and H. T. Li, J. Math. Phys. **11**, 4 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Six Integral Theorems for Vector Spherical Harmonics*

ROGER E. CLAPP
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts
AND
H. TUNG LI
AirTechnology Corporation, Waltham, Massachusetts
and
College of the Holy Cross, Worcester, Massachusetts

(Received 11 August 1967)

With the aid of a generalized addition theorem for spherical harmonics [J. Math. Phys. **11**, 1 (1970), preceding paper], six integral theorems for vector spherical harmonics are proved. A source-point direction (θ_1, ϕ_1) is first expressed in terms of a field-point direction (θ, ϕ) and a polar-coordinate angle-pair (θ', ϕ'), which has as its polar axis the field-point direction. For a particular choice of n and m, all the components of the vector spherical harmonics for the source, expressed in terms of (θ_1, ϕ_1), are integrated over the relative azimuth angle ϕ' while the field-point direction (θ, ϕ) and the relative polar angle θ' are held fixed. The result in each case is a spherical harmonic or vector spherical harmonic of the field-point direction, with the same n and m but now depending on (θ, ϕ) instead of (θ_1, ϕ_1), multiplied by an explicit function of the relative polar angle θ'.

1. INTRODUCTION

In the preceding article,[1] a generalized addition theorem for spherical harmonics was established. This addition theorem led directly to the proof of an integral theorem for spherical harmonics. In the integration, a vector $\mathbf{r}_1$ is swung about a fixed vector $\mathbf{r}$, with a constant angle θ' separating the two vectors. A suitably defined azimuth angle ϕ' locates the azimuth of $\mathbf{r}_1$ in its motion about $\mathbf{r}$. This integration, over a full circuit, gave the result

$$\int_{\phi'=0}^{\phi'=2\pi} X_n^m(\theta_1, \phi_1)\,d\phi' = 2\pi X_n^m(\theta, \phi)P_n(\cos\theta'), \quad (1.1)$$

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.
[1] R. E. Clapp, J. Math. Phys. **11**, 1 (1970).

where X_n^m is a general spherical harmonic and P_n is a Legendre polynomial.

In this article, the result (1.1) is generalized to apply to problems which arise in the use and application of vector spherical harmonics—vector functions which are closely related to the scalar spherical harmonics X_n^m. Because there are three independent vector spherical harmonics for each choice of (n, m), and each of these has three scalar components, the generalization of (1.1) leads to nine independent scalar equations. However, these nine scalar equations group naturally into three radial-component equations, each of scalar form, and three transverse-component equations, each of which combines two scalar equations into a vector equation with only two independent components rather than three.

Thus, there are only six separate integral theorems to be considered, rather than nine. This article gives the proof of each of these six integral theorems.

2. VECTOR SPHERICAL HARMONICS

The notation for the vector spherical harmonics to be used in this article follows that used by Morse and Feshbach.[2] The scalar spherical harmonics are defined by

$$X_n^m(\theta, \phi) = e^{im\phi} P_n^{|m|}(\cos\theta), \tag{2.1}$$

where

$$P_n^{|m|}(\cos\theta) = (\sin\theta)^{|m|} \frac{d^{|m|}}{d(\cos\theta)^{|m|}} P_n(\cos\theta). \tag{2.2}$$

The vector spherical harmonics separate into three sets:

$$\mathbf{P}_n^m(\theta, \phi) = \mathbf{a}_r X_n^m(\theta, \phi), \tag{2.3}$$

$$\mathbf{B}_n^m(\theta, \phi) = \frac{[n(n+1)]^{\frac{1}{2}}}{(2n+1)\sin\theta}\left[\mathbf{a}_\theta\left(\frac{n-|m|+1}{n+1} X_{n+1}^m - \frac{n+|m|}{n} X_{n-1}^m\right) + \mathbf{a}_\phi \frac{im(2n+1)}{n(n+1)} X_n^m\right], \tag{2.4}$$

$$\mathbf{C}_n^m(\theta, \phi) = \frac{[n(n+1)]^{\frac{1}{2}}}{(2n+1)\sin\theta}\left[\mathbf{a}_\theta \frac{im(2n+1)}{n(n+1)} X_n^m - \mathbf{a}_\phi\left(\frac{n-|m|+1}{n+1} X_{n+1}^m - \frac{n+|m|}{n} X_{n-1}^m\right)\right]. \tag{2.5}$$

In (2.4) and (2.5) the arguments of the scalar harmonics are (θ, ϕ) in each case. Explicit forms for the unit vectors $\mathbf{a}_r$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ are given in Eqs. (3.1) of Ref. 1.

[2] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., 1953), pp. 1898, 1899.

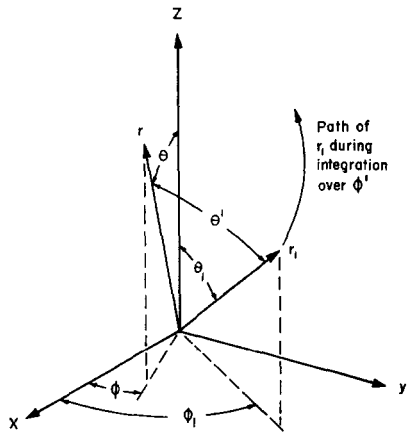


FIG. 1. The relationship between the source-point vector $\mathbf{r}_1$ and the field-point vector $\mathbf{r}$. During the integration over ϕ', the vector $\mathbf{r}_1$ is swung about $\mathbf{r}$ with the relative polar angle θ' held constant.

3. THEOREM 1

The six integral theorems involve integrations over ϕ' of components of the vector spherical harmonics which are expressed as functions of (θ_1, ϕ_1). For example, the vector spherical harmonic $\mathbf{P}_n^m(\theta_1, \phi_1)$ has the form

$$\mathbf{P}_n^m(\theta_1, \phi_1) = \mathbf{a}_{1r} X_n^m(\theta_1, \phi_1), \tag{3.1}$$

obtained from (2.3), where the unit vector $\mathbf{a}_{1r}$ and the orthogonal unit vectors $\mathbf{a}_{1\theta}$ and $\mathbf{a}_{1\phi}$ have been given explicitly in Eqs. (3.2) of Ref. 1. The geometrical configuration of the vectors $\mathbf{r}$ and $\mathbf{r}_1$ is shown in Figs. 1–3 of Ref. 1, while the trigonometric transformations are given in Eqs. (3.5) and (3.6) of Ref. 1.

In the integration over ϕ', the vector $\mathbf{r}$ is held fixed while the vector $\mathbf{r}_1$ is swung in an arc with the angle θ' held constant. The path of integration is shown here in Fig. 1. Of the four independent angular variables θ, ϕ, θ', and ϕ', only ϕ' varies, but both of the angles θ_1 and ϕ_1 will vary during the integration. These latter are treated as dependent variables through the use of Eqs. (3.5) of Ref. 1.

The direction of the vector function (3.1) changes during the integration over ϕ' since $\mathbf{a}_{1r}$ is a function of ϕ'. The vector function (3.1) will therefore be resolved into its components before the integration; the simplest component to be considered is the component which lies in the z' direction, the direction of the unit vector $\mathbf{a}_r$. Accordingly, the projection of (3.1) onto $\mathbf{a}_r$,

$$\mathbf{a}_r \cdot \mathbf{P}_n^m(\theta_1, \phi_1) = \cos\theta' X_n^m(\theta_1, \phi_1), \tag{3.2}$$

will be integrated over ϕ'.

When the generalized addition theorem

$$X_n^m(\theta_1, \phi_1) = X_n^m(\theta, \phi) P_n(\cos\theta') + \sum_{\mu=1}^{n} [A_c \cos(\mu\phi') + A_s \sin(\mu\phi')], \tag{3.3}$$

which was given as Eq. (5.1) of Ref. 1, is substituted in the right-hand side of (3.2) and the resulting expression is integrated term by term over the angle ϕ' with θ, ϕ, and θ' held constant, each term in the summation over μ integrates to zero (since A_c and A_s do not depend upon ϕ'), and the result is the first of the six integral theorems:

Theorem 1:

$$\int_{\phi'=0}^{\phi'=2\pi} \mathbf{a}_r \cdot \mathbf{P}_n^m(\theta_1, \phi_1)\, d\phi' = 2\pi X_n^m(\theta, \phi) \cos\theta' P_n(\cos\theta'). \quad (3.4)$$

4. THEOREMS 2 AND 3

Two more integral theorems are obtained when the addition theorem (3.3) is differentiated with respect to θ' and ϕ'. It can first be verified, from Eqs. (3.1), (3.2), and (3.5) of Ref. 1, that

$$\frac{\partial\theta_1}{\partial\theta'} = -\frac{1}{\sin\theta'}(\mathbf{a}_r \cdot \mathbf{a}_{1\theta}), \quad (4.1a)$$

$$\frac{\partial\phi_1}{\partial\theta'} = \frac{-1}{\sin\theta_1 \sin\theta'}(\mathbf{a}_r \cdot \mathbf{a}_{1\phi}), \quad (4.1b)$$

$$\frac{\partial\theta_1}{\partial\phi'} = (\mathbf{a}_r \cdot \mathbf{a}_{1\phi}), \quad (4.2a)$$

$$\frac{\partial\phi_1}{\partial\phi'} = \frac{-1}{\sin\theta_1}(\mathbf{a}_r \cdot \mathbf{a}_{1\theta}). \quad (4.2b)$$

From the properties of the spherical harmonics (2.1), it can also be verified that

$$\frac{\partial}{\partial\theta_1} X_n^m(\theta_1, \phi_1) = \frac{n(n+1)}{(2n+1)\sin\theta_1} \times \left[\frac{n-|m|+1}{n+1} X_{n+1}^m(\theta_1, \phi_1) - \frac{n+|m|}{n} X_{n-1}^m(\theta_1, \phi_1)\right], \quad (4.3a)$$

$$\frac{\partial}{\partial\phi_1} X_n^m(\theta_1, \phi_1) = imX_n^m(\theta_1, \phi_1). \quad (4.3b)$$

From (4.1)–(4.3) and from the definitions of $\mathbf{B}_n^m$ and $\mathbf{C}_n^m$ in (2.4) and (2.5), it can be seen that

$$\frac{\partial}{\partial\theta'} X_n^m(\theta_1, \phi_1) = -\frac{[n(n+1)]^{\frac{1}{2}}}{\sin\theta'} \mathbf{a}_r \cdot \mathbf{B}_n^m(\theta_1, \phi_1), \quad (4.4)$$

$$\frac{\partial}{\partial\phi'} X_n^m(\theta_1, \phi_1) = -[n(n+1)]^{\frac{1}{2}} \mathbf{a}_r \cdot \mathbf{C}_n^m(\theta_1, \phi_1). \quad (4.5)$$

Differentiation of (3.3) with respect to θ' now gives

$$\mathbf{a}_r \cdot \mathbf{B}_n^m(\theta_1, \phi_1) = [n(n+1)]^{-\frac{1}{2}} \times \left[X_n^m(\theta, \phi) \sin^2\theta' \frac{d}{d(\cos\theta')} P_n(\cos\theta') - \sin\theta' \sum_{\mu=1}^{n} \left(\frac{\partial A_c}{\partial\theta'} \cos(\mu\phi') + \frac{\partial A_s}{\partial\theta'} \sin(\mu\phi')\right)\right], \quad (4.6)$$

while differentiation with respect to ϕ' gives

$$\mathbf{a}_r \cdot \mathbf{C}_n^m(\theta_1, \phi_1) = [n(n+1)]^{-\frac{1}{2}} \times \sum_{\mu=1}^{n} [\mu A_c \sin(\mu\phi') - \mu A_s \cos(\mu\phi')]. \quad (4.7)$$

When the explicit forms for A_c and A_s, obtainable from Eq. (4.8) of Ref. 1, are introduced, then (4.6) and (4.7) become addition theorems in their own right for use with the vector spherical harmonics $\mathbf{B}_n^m$ and $\mathbf{C}_n^m$.

Equations (4.6) and (4.7) can now be integrated with respect to ϕ' to give the second and third integral theorems, respectively:

Theorem 2:

$$\int_{\phi'=0}^{\phi'=2\pi} \mathbf{a}_r \cdot \mathbf{B}_n^m(\theta_1, \phi_1)\, d\phi' = 2\pi[n(n+1)]^{-\frac{1}{2}} X_n^m(\theta, \phi) \sin^2\theta' \frac{d}{d(\cos\theta')} P_n(\cos\theta'); \quad (4.8)$$

Theorem 3:

$$\int_{\phi'=0}^{\phi'=2\pi} \mathbf{a}_r \cdot \mathbf{C}_n^m(\theta_1, \phi_1)\, d\phi' = 0. \quad (4.9)$$

5. RECURSION FORMULAS

The familiar recursion formula for spherical harmonics,[3]

$$\cos\theta\, X_n^m(\theta, \phi) = \frac{n-|m|+1}{2n+1} X_{n+1}^m(\theta, \phi) + \frac{n+|m|}{2n+1} X_{n-1}^m(\theta, \phi), \quad (5.1)$$

leads directly to a similar formula for the vector spherical harmonics $\mathbf{P}_n^m(\theta, \phi)$:

$$\cos\theta\, \mathbf{P}_n^m(\theta, \phi) = \frac{n-|m|+1}{2n+1} \mathbf{P}_{n+1}^m(\theta, \phi) + \frac{n+|m|}{2n+1} \mathbf{P}_{n-1}^m(\theta, \phi). \quad (5.2)$$

[3] Ref. 2, p. 1326.

Corresponding recursion formulas can be established for the vector spherical harmonics $\mathbf{B}_n^m(\theta, \phi)$ and $\mathbf{C}_n^m(\theta, \phi)$, through (5.1) as applied to (2.4) and (2.5), but there is some cross-coupling of the two symmetries:

$$\cos\theta\,\mathbf{B}_n^m(\theta,\phi) = \frac{(n-|m|+1)[n(n+2)]^{\frac{1}{2}}}{(2n+1)(n+1)}\mathbf{B}_{n+1}^m(\theta,\phi) - \frac{im}{n(n+1)}\mathbf{C}_n^m(\theta,\phi) + \frac{(n+|m|)[(n-1)(n+1)]^{\frac{1}{2}}}{(2n+1)n}\mathbf{B}_{n-1}^m(\theta,\phi), \quad (5.3)$$

$$\cos\theta\,\mathbf{C}_n^m(\theta,\phi) = \frac{(n-|m|+1)[n(n+2)]^{\frac{1}{2}}}{(2n+1)(n+1)}\mathbf{C}_{n+1}^m(\theta,\phi) + \frac{im}{n(n+1)}\mathbf{B}_n^m(\theta,\phi) + \frac{(n+|m|)[(n-1)(n+1)]^{\frac{1}{2}}}{(2n+1)n}\mathbf{C}_{n-1}^m(\theta,\phi). \quad (5.4)$$

6. THEOREM 4

From the equations in Sec. 3 of Ref. 1, it can be shown that

$$(\mathbf{a}_\theta \cdot \mathbf{a}_{1r}) = \sin\theta'\cos(\phi' - \phi\cos\theta) = (1/\sin\theta)(\cos\theta\cos\theta' - \cos\theta_1), \quad (6.1)$$

$$(\mathbf{a}_\phi \cdot \mathbf{a}_{1r}) = \sin\theta'\sin(\phi' - \phi\cos\theta) = \frac{1}{\sin\theta}\frac{\partial}{\partial\phi'}(\cos\theta_1). \quad (6.2)$$

From (2.3), (5.1), and (6.1), it is apparent that

$$\mathbf{a}_\theta \cdot \mathbf{P}_n^m(\theta_1,\phi_1) = \frac{-1}{\sin\theta}\left[\frac{n-|m|+1}{2n+1}X_{n+1}^m(\theta_1,\phi_1) - \cos\theta\cos\theta' X_n^m(\theta_1,\phi_1) + \frac{n+|m|}{2n+1}X_{n-1}^m(\theta_1,\phi_1)\right]. \quad (6.3)$$

Each of the scalar spherical harmonics on the right-hand side of (6.3) can be replaced by an expansion of the form (3.3). Integration over ϕ' then gives

$$\int_{\phi'=0}^{\phi'=2\pi}\mathbf{a}_\theta \cdot \mathbf{P}_n^m(\theta_1,\phi_1)\,d\phi' = \frac{-2\pi}{\sin\theta}\left\{\frac{n-|m|+1}{2n+1}X_{n+1}^m(\theta,\phi)P_{n+1}(\cos\theta') - [\cos\theta\,X_n^m(\theta,\phi)][\cos\theta' P_n(\cos\theta')] + \frac{n+|m|}{2n+1}X_{n-1}^m(\theta,\phi)P_{n-1}(\cos\theta')\right\}. \quad (6.4)$$

Equation (6.4) can be simplified with the aid of (5.1) and with the use of (4.3a) in the form it takes with $m = 0$. The result is the first half of the fourth integral theorem:

Theorem 4a:

$$\int_{\phi'=0}^{\phi'=2\pi}\mathbf{a}_\theta \cdot \mathbf{P}_n^m(\theta_1,\phi_1)\,d\phi' = \frac{2\pi}{(2n+1)\sin\theta}\left[\frac{n-|m|+1}{n+1}X_{n+1}^m(\theta,\phi) - \frac{n+|m|}{n}X_{n-1}^m(\theta,\phi)\right]\sin^2\theta'\frac{d}{d(\cos\theta')}P_n(\cos\theta'). \quad (6.5a)$$

From (2.3), (6.2), and (4.5), the relationship

$$\mathbf{a}_\phi \cdot \mathbf{P}_n^m(\theta_1,\phi_1) = \frac{1}{\sin\theta}\frac{\partial}{\partial\phi'}[\cos\theta_1 X_n^m(\theta_1,\phi_1)] + \frac{[n(n+1)]^{\frac{1}{2}}}{\sin\theta}\mathbf{a}_r \cdot [\cos\theta_1 \mathbf{C}_n^m(\theta_1,\phi_1)]$$

can be established, and the recursion formulas (5.1) and (5.4) can then be used to replace the bracketed expressions by linear combinations of scalar or vector harmonics. The integration over ϕ' can then be carried out without difficulty, by methods used in earlier theorems, with the result:

Theorem 4b:

$$\int_{\phi'=0}^{\phi'=2\pi}\mathbf{a}_\phi \cdot \mathbf{P}_n^m(\theta_1,\phi_1)\,d\phi' = \frac{2\pi i m}{n(n+1)\sin\theta}X_n^m(\theta,\phi)\sin^2\theta'\frac{d}{d(\cos\theta')}P_n(\cos\theta'). \quad (6.5b)$$

Theorems (4a) and (4b) deal with the transverse components of $\mathbf{P}_n^m(\theta_1,\phi_1)$, i.e., the components that are perpendicular to the unit vector $\mathbf{a}_r$. A vector combination of just these transverse components is equivalent to the subtraction of the longitudinal component from the original vector:

$$\mathbf{P}_n^m(\theta_1,\phi_1) - \mathbf{a}_r\mathbf{a}_r \cdot \mathbf{P}_n^m(\theta_1,\phi_1) = \mathbf{a}_\theta\mathbf{a}_\theta \cdot \mathbf{P}_n^m(\theta_1,\phi_1) + \mathbf{a}_\phi\mathbf{a}_\phi \cdot \mathbf{P}_n^m(\theta_1,\phi_1). \quad (6.6)$$

When this vector combination of (6.5a) and (6.5b) is made, it is found to have the grouping of terms that appears in (2.4). The result is:

Theorem 4:

$$\int_{\phi'=0}^{\phi'=2\pi}[\mathbf{P}_n^m(\theta_1,\phi_1) - \mathbf{a}_r\mathbf{a}_r \cdot \mathbf{P}_n^m(\theta_1,\phi_1)]\,d\phi' = 2\pi[n(n+1)]^{-\frac{1}{2}}\mathbf{B}_n^m(\theta,\phi)\sin^2\theta'\frac{d}{d(\cos\theta')}P_n(\cos\theta'). \quad (6.7)$$

7. THEOREM 5

The fifth integral theorem deals with the transverse part of the vector spherical harmonic $\mathbf{B}_n^m(\theta_1, \phi_1)$. In analogy with (6.1) and (6.2), four scalar products are needed:

$$(\mathbf{a}_\theta \cdot \mathbf{a}_{1\theta}) = \frac{\sin\theta_1}{\sin\theta} - \frac{\cos\theta \sin\theta'}{\sin\theta}\frac{\partial\theta_1}{\partial\theta'}, \tag{7.1a}$$

$$(\mathbf{a}_\theta \cdot \mathbf{a}_{1\phi}) = -\frac{\sin\theta_1 \cos\theta \sin\theta'}{\sin\theta}\frac{\partial\phi_1}{\partial\theta'}, \tag{7.1b}$$

$$(\mathbf{a}_\phi \cdot \mathbf{a}_{1\theta}) = -\frac{\cos\theta_1}{\sin\theta}\frac{\partial\theta_1}{\partial\phi'}, \tag{7.1c}$$

$$(\mathbf{a}_\phi \cdot \mathbf{a}_{1\phi}) = \frac{\sin\theta_1 \cos\theta'}{\sin\theta} - \frac{\sin\theta_1 \cos\theta_1}{\sin\theta}\frac{\partial\phi_1}{\partial\phi'}, \tag{7.1d}$$

which can all be derived from the equations in Sec. 3 of Ref. 1.

From (2.4), (4.3), and (7.1), it is found that the transverse portion of $\mathbf{B}_n^m(\theta_1, \phi_1)$ is

$$\begin{aligned}
&\mathbf{a}_\theta\mathbf{a}_\theta \cdot \mathbf{B}_n^m(\theta_1, \phi_1) + \mathbf{a}_\phi\mathbf{a}_\phi \cdot \mathbf{B}_n^m(\theta_1, \phi_1) \\
&\quad = -[n(n+1)]^{-\frac{1}{2}}\Big(\mathbf{a}_\theta \frac{\cos\theta \sin\theta'}{\sin\theta}\frac{\partial}{\partial\theta'} X_n^m(\theta_1, \phi_1) \\
&\qquad + \mathbf{a}_\phi \frac{\cos\theta_1}{\sin\theta}\frac{\partial}{\partial\phi'} X_n^m(\theta_1, \phi_1)\Big) \\
&\qquad + \frac{[n(n+1)]^{\frac{1}{2}}}{(2n+1)\sin\theta}\mathbf{a}_\theta\Big(\frac{n - |m| + 1}{n+1} X_{n+1}^m(\theta_1, \phi_1) \\
&\qquad - \frac{n + |m|}{n} X_{n-1}^m(\theta_1, \phi_1)\Big) \\
&\qquad + \frac{im\cos\theta'}{[n(n+1)]^{\frac{1}{2}}\sin\theta}\mathbf{a}_\phi X_n^m(\theta_1, \phi_1).
\end{aligned} \tag{7.2}$$

In the integration of (7.2) over ϕ', there are five terms to be considered. The first two terms can be transformed with the aid of (4.4), (4.5), and (5.4), then integrated through Theorems 2 and 3. The remaining three terms may be integrated directly, by Theorem 1 or its scalar equivalent, Eq. (5.2) of Ref. 1. What is obtained from these integrations is the fifth integral theorem:

Theorem 5:

$$\begin{aligned}
&\int_{\phi'=0}^{\phi'=2\pi} [\mathbf{B}_n^m(\theta_1, \phi_1) - \mathbf{a}_r\mathbf{a}_r \cdot \mathbf{B}_n^m(\theta_1, \phi_1)]\, d\phi' \\
&\qquad = 2\pi\mathbf{B}_n^m(\theta, \phi)\Big(\cos\theta' P_n(\cos\theta') \\
&\qquad\quad + \frac{\sin^2\theta'}{n(n+1)}\frac{d}{d(\cos\theta')} P_n(\cos\theta')\Big).
\end{aligned} \tag{7.3}$$

8. THEOREM 6

Four additional partial derivatives, analogous to those given in (4.1) and (4.2), can be obtained from Eqs. (3.5) of Ref. 1:

$$\begin{aligned}
\frac{\partial\theta_1}{\partial\theta} = \frac{1}{\sin\theta_1}&[\sin\theta\cos\theta' \\
&+ \cos\theta\sin\theta'\cos(\phi' - \phi\cos\theta) \\
&- \phi\sin^2\theta\sin\theta'\sin(\phi' - \phi\cos\theta)],
\end{aligned} \tag{8.1a}$$

$$\frac{\partial\theta_1}{\partial\phi} = \frac{1}{\sin\theta_1}\sin\theta\cos\theta\sin\theta'\sin(\phi' - \phi\cos\theta), \tag{8.1b}$$

$$\begin{aligned}
\frac{\partial\phi_1}{\partial\theta} = \frac{\sin\theta'}{\sin^2\theta_1}&[-\cos\theta\cos\theta'\sin(\phi' - \phi\cos\theta) \\
&+ \sin\theta\sin\theta'\sin(\phi' - \phi\cos\theta) \\
&\times \cos(\phi' - \phi\cos\theta) + \phi\sin\theta\cos\theta\sin\theta' \\
&+ \phi\sin^2\theta\cos\theta'\cos(\phi' - \phi\cos\theta)],
\end{aligned} \tag{8.1c}$$

$$\begin{aligned}
\frac{\partial\phi_1}{\partial\phi} = \frac{\sin\theta}{\sin^2\theta_1}&[\sin\theta + \cos\theta\sin\theta' \\
&\times \cos\theta'\cos(\phi' - \phi\cos\theta) \\
&- \sin\theta\sin^2\theta'\cos^2(\phi' - \phi\cos\theta)].
\end{aligned} \tag{8.1d}$$

In terms of these derivatives, (7.1) can be rewritten as

$$(\mathbf{a}_\theta \cdot \mathbf{a}_{1\theta}) = \frac{\sin\theta_1}{\sin\theta}\frac{\partial\phi_1}{\partial\phi}, \tag{8.2a}$$

$$(\mathbf{a}_\theta \cdot \mathbf{a}_{1\phi}) = \frac{-1}{\sin\theta}\frac{\partial\theta_1}{\partial\phi}, \tag{8.2b}$$

$$(\mathbf{a}_\phi \cdot \mathbf{a}_{1\theta}) = -\sin\theta_1\Big(\frac{\partial\phi_1}{\partial\theta} - \phi\tan\theta + \phi\tan\theta\frac{\partial\phi^1}{\partial\phi}\Big), \tag{8.2c}$$

$$(\mathbf{a}_\phi \cdot \mathbf{a}_{1\phi}) = \frac{\partial\theta_1}{\partial\theta} + \phi\tan\theta\frac{\partial\theta_1}{\partial\phi}. \tag{8.2d}$$

In analogy with (7.2), the transverse portion of the vector spherical harmonic $\mathbf{C}_n^m(\theta_1, \phi_1)$ can be written in the form

$$\begin{aligned}
&\mathbf{a}_\theta\mathbf{a}_\theta \cdot \mathbf{C}_n^m(\theta_1, \phi_1) + \mathbf{a}_\phi\mathbf{a}_\phi \cdot \mathbf{C}_n^m(\theta_1, \phi_1) \\
&= [n(n+1)]^{-\frac{1}{2}}\Big[\mathbf{a}_\theta \frac{1}{\sin\theta}\frac{\partial}{\partial\phi} X_n^m(\theta_1, \phi_1) \\
&\quad - \mathbf{a}_\phi\Big(\frac{\partial}{\partial\theta} - im\phi\tan\theta + \phi\tan\theta\frac{\partial}{\partial\phi}\Big) X_n^m(\theta_1, \phi_1)\Big].
\end{aligned} \tag{8.3}$$

When $X_n^m(\theta_1, \phi_1)$ in (8.3) is replaced by the expansion

(3.3) and the resulting expression integrated over ϕ', the sixth of the integral theorems is obtained:

Theorem 6:

$$\int_{\phi'=0}^{\phi'=2\pi} [\mathbf{C}_n^m(\theta_1, \phi_1) - \mathbf{a}_r\mathbf{a}_r \cdot \mathbf{C}_n^m(\theta_1, \phi_1)]\, d\phi' = 2\pi \mathbf{C}_n^m(\theta, \phi) P_n(\cos\theta'). \quad (8.4)$$

9. APPLICATIONS

These six integral theorems are of particular value when a vector field function is to be obtained from a vector source function. An example is the retarded Hertz vector, as obtained from an electrical source-current distribution.[4] The theorems show that if a source function, depending on the angles (θ_1, ϕ_1), has that dependence characterized by a particular choice of n and m, then the integration over the azimuth angle ϕ' will immediately ensure that the field function, in its dependence upon the angles (θ, ϕ), will be characterized by the same choice of n and m, provided that the Green's function relating field to source does not itself depend upon ϕ'. Whatever dependence the Green's function may have upon the angle θ' (the angle between the vector $\mathbf{r}_1$ to the source point and the vector $\mathbf{r}$ to the field point) is here immaterial, and it cannot affect the "mode separation" of the expansion into vector spherical harmonics.

It should be noted, however, that certain of the six theorems introduce coupling between a $\mathbf{P}_n^m$ source function and a $\mathbf{B}_n^m$ field function, or vice versa. That is, the theorems mix the $\mathbf{P}_n^m$ and $\mathbf{B}_n^m$ symmetries while keeping the $\mathbf{C}_n^m$ symmetry separate from the other two.

The use of these theorems in electromagnetism will be illustrated in a following article.[4]

[4] R. E. Clapp, L. Huang, and H. T. Li, J. Math. Phys. **11**, 9 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Vector Spherical Harmonic Expansion in the Time Domain of the Retarded Hertz Vector for a Distributed, Transient Source-Current Configuration*

ROGER E. CLAPP AND LUC HUANG†
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts
AND
H. TUNG LI
AirTechnology Corporation, Waltham, Massachusetts
and
College of the Holy Cross, Worcester, Massachusetts

(Received 11 August 1967)

An equation giving the retarded Hertz vector $\mathbf{\Pi}(\mathbf{r}, t)$ in terms of a source-current distribution $\mathbf{J}(\mathbf{r}_1, t_1)$ is derived. Both vector functions are written as expansions in vector spherical harmonics, with the expansion coefficients containing the dependence upon radial distance and time, while the angular dependence is kept within the harmonic functions. After integration over two angles, expressions are obtained giving the expansion coefficients for the Hertz vector, which depend upon (r, t), in terms of the corresponding coefficients for the current, which depend upon (r_1, t_1). The original four-dimensional problem is thus reduced to two dimensions, but with the four-dimensional causality requirements satisfied at each step of the analysis.

1. INTRODUCTION

Although the electromagnetic radiation from a general time-dependent source-current distribution is usually analyzed in the frequency domain, for certain problems the time domain is more appropriate. An example is the electromagnetic radiation from the electrical currents generated in the air by an atmospheric nuclear detonation.[1]

In this example the source currents are highly transient, and retardation across the source region plays an important role. Furthermore, part of the current is in the form of relativistic electrons, produced by Compton collisions between gamma rays from the detonation and electrons from air molecules.

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.

† Present address: Department of Physics, University of Vermont, Burlington, Vermont.

[1] A. S. Kompaneets, Zh. Eksp. Teor. Fiz. **35**, 1538 (1959) [Sov. Phys.—JETP **8**, 1076 (1959)].

(3.3) and the resulting expression integrated over ϕ', the sixth of the integral theorems is obtained:

Theorem 6:

$$\int_{\phi'=0}^{\phi'=2\pi} [\mathbf{C}_n^m(\theta_1, \phi_1) - \mathbf{a}_r\mathbf{a}_r \cdot \mathbf{C}_n^m(\theta_1, \phi_1)]\, d\phi' = 2\pi \mathbf{C}_n^m(\theta, \phi) P_n(\cos\theta'). \quad (8.4)$$

9. APPLICATIONS

These six integral theorems are of particular value when a vector field function is to be obtained from a vector source function. An example is the retarded Hertz vector, as obtained from an electrical source-current distribution.[4] The theorems show that if a source function, depending on the angles (θ_1, ϕ_1), has that dependence characterized by a particular choice of n and m, then the integration over the azimuth angle ϕ' will immediately ensure that the field function, in its dependence upon the angles (θ, ϕ), will be characterized by the same choice of n and m, provided that the Green's function relating field to source does not itself depend upon ϕ'. Whatever dependence the Green's function may have upon the angle θ' (the angle between the vector $\mathbf{r}_1$ to the source point and the vector $\mathbf{r}$ to the field point) is here immaterial, and it cannot affect the "mode separation" of the expansion into vector spherical harmonics.

It should be noted, however, that certain of the six theorems introduce coupling between a $\mathbf{P}_n^m$ source function and a $\mathbf{B}_n^m$ field function, or vice versa. That is, the theorems mix the $\mathbf{P}_n^m$ and $\mathbf{B}_n^m$ symmetries while keeping the $\mathbf{C}_n^m$ symmetry separate from the other two.

The use of these theorems in electromagnetism will be illustrated in a following article.[4]

[4] R. E. Clapp, L. Huang, and H. T. Li, J. Math. Phys. **11**, 9 (1970).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Vector Spherical Harmonic Expansion in the Time Domain of the Retarded Hertz Vector for a Distributed, Transient Source-Current Configuration*

ROGER E. CLAPP AND LUC HUANG†
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts
AND
H. TUNG LI
AirTechnology Corporation, Waltham, Massachusetts
and
College of the Holy Cross, Worcester, Massachusetts

(Received 11 August 1967)

An equation giving the retarded Hertz vector $\mathbf{\Pi}(\mathbf{r}, t)$ in terms of a source-current distribution $\mathbf{J}(\mathbf{r}_1, t_1)$ is derived. Both vector functions are written as expansions in vector spherical harmonics, with the expansion coefficients containing the dependence upon radial distance and time, while the angular dependence is kept within the harmonic functions. After integration over two angles, expressions are obtained giving the expansion coefficients for the Hertz vector, which depend upon (r, t), in terms of the corresponding coefficients for the current, which depend upon (r_1, t_1). The original four-dimensional problem is thus reduced to two dimensions, but with the four-dimensional causality requirements satisfied at each step of the analysis.

1. INTRODUCTION

Although the electromagnetic radiation from a general time-dependent source-current distribution is usually analyzed in the frequency domain, for certain problems the time domain is more appropriate. An example is the electromagnetic radiation from the electrical currents generated in the air by an atmospheric nuclear detonation.[1]

In this example the source currents are highly transient, and retardation across the source region plays an important role. Furthermore, part of the current is in the form of relativistic electrons, produced by Compton collisions between gamma rays from the detonation and electrons from air molecules.

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.

† Present address: Department of Physics, University of Vermont, Burlington, Vermont.

[1] A. S. Kompaneets, Zh. Eksp. Teor. Fiz. **35**, 1538 (1959) [Sov. Phys.—JETP **8**, 1076 (1959)].

The fields radiated by relativistic electrons show forward directivity, which is not easily expressed in the frequency domain, but which enters readily into a time-domain formalism.

The analysis is most straightforward if the medium is assumed to be the vacuum. In the example mentioned above,[1] there is a time-dependent conductivity in the source region, but this conductivity is actually formed of electrons and ions whose motion in the local fields can be represented as an addition to the primary source current. Similarly, dielectric polarization of the air molecules can also be represented mathematically as a secondary increment to the source-current distribution. Thus if the secondary currents are all treated explicitly, the vacuum equations will be adequate for the calculation of the fields generated by all the source currents.

The source-current distribution will be expressed as an expansion in vector spherical harmonics, with each coefficient depending in an arbitrary way upon the radial distance and the time. With the aid of theorems established earlier,[2] the angular dependence of the source will be integrated over, leaving the field function for each vector-spherical-harmonic mode expressed as a function of radial distance and time.

The Hertz-vector formalism will be used because of the relative simplicity of the integrations. In later articles[3,4] the magnetic field and electric field will be given explicitly.

2. HERTZ-VECTOR FORMALISM

The electromagnetic fields generated by a time-dependent distribution of charges and currents can be expressed in terms of the vector and scalar potentials, together with an auxiliary condition which limits the charge and current densities to those which satisfy the equation of continuity. It is more convenient, however, to make use of the Hertz-vector formalism.[5] The need for an auxiliary condition is avoided through the use of the free-charge polarization vector $\mathfrak{P}$ to represent the source distribution. As will be shown, the integrals can then be reformulated so that the final equation for the Hertz vector expresses the source in terms of the current density $\mathbf{J}$ without any explicit appearance of the charge density ρ.

The free-charge polarization vector is defined in the equations

$$\mathbf{J} = \frac{\partial}{\partial t}\mathfrak{P}, \quad \rho = -\nabla \cdot \mathfrak{P}. \tag{2.1}$$

For a transient source, therefore,

$$\mathfrak{P} = \int_{t=t_0}^{t=t} \mathbf{J}\, dt, \tag{2.2}$$

where t_0 is a time which precedes any current flow. [Any pre-existing electrostatic field can be represented through a constant of integration $\mathfrak{P}_0$ added to the right-hand side of (2.2).] Because of the form of (2.1), the equation of continuity of charge is automatically satisfied.

The Hertz vector, $\mathbf{\Pi}(\mathbf{r}, t)$, is defined by

$$\mathbf{\Pi}(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \iiint \frac{1}{s}\mathfrak{P}(\mathbf{r}_1, t - s/c)\, dV_1, \tag{2.3}$$

in the MKS units used by Stratton.[5] In (2.3), the volume element at the source point $\mathbf{r}_1$ is dV_1, and s is the distance from the source point to the field point $\mathbf{r}$, that is,

$$s = |\mathbf{r} - \mathbf{r}_1|. \tag{2.4}$$

Equation (2.3) defines the *retarded* Hertz vector, from which the retarded electric and magnetic field vectors can be obtained, by the equations

$$\mathbf{E} = \nabla\nabla \cdot \mathbf{\Pi} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{\Pi}, \tag{2.5}$$

$$\mathbf{H} = \epsilon_0 \nabla \times \frac{\partial}{\partial t}\mathbf{\Pi}. \tag{2.6}$$

These field vectors satisfy Maxwell's equations.

An *advanced* Hertz vector could also be defined, through the use of $\mathfrak{P}(\mathbf{r}_1, t + s/c)$ instead of $\mathfrak{P}(\mathbf{r}_1, t - s/c)$ in the integrand of (2.3). Equations (2.5) and (2.6) would then give advanced electromagnetic field vectors. These would also satisfy Maxwell's equations. However, there is no evidence that these advanced fields play any role in macroscopic electromagnetism. Accordingly, only the retarded Hertz vector (2.3) will be considered further in this article.

3. EXPANSION IN VECTOR SPHERICAL HARMONICS

The scalar and vector spherical harmonics have been defined in Sec. 2 of Ref. 2. The vector functions $\mathbf{J}$, $\mathbf{P}$, and $\mathbf{\Pi}$ will be expanded in terms of the vector spherical harmonics $\mathbf{P}_n^m$, $\mathbf{B}_n^m$, and $\mathbf{C}_n^m$. The Hertz vector is defined at the field point $(\mathbf{r}, t)$ and will be written as the expansion

$$\mathbf{\Pi} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \{\Pi_{r,n}^m(r, t)\mathbf{P}_n^m(\theta, \phi) + \Pi_{B,n}^m(r, t)\mathbf{B}_n^m(\theta, \phi) + \Pi_{C,n}^m(r, t)\mathbf{C}_n^m(\theta, \phi)\}, \tag{3.1}$$

while the free-charge polarization vector is defined at

[2] R. E. Clapp and H. T. Li, J. Math. Phys. **11**, 4 (1970) (preceding paper).
[3] R. E. Clapp and L. Huang, J. Math. Phys. **11**, 14 (1970).
[4] R. E. Clapp, L. Huang, and H. T. Li, J. Math. Phys. **11**, 16 (1970).
[5] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Co., 1941), pp. 28–32, 430, 431.

the source point $(\mathbf{r}_1, t_1)$ and has the expansion

$$\mathfrak{F} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \{\mathfrak{F}^m_{r,n}(r_1, t_1)\mathbf{P}^m_n(\theta_1, \phi_1) + \mathfrak{F}^m_{B,n}(r_1, t_1)\mathbf{B}^m_n(\theta_1, \phi_1) + \mathfrak{F}^m_{C,n}(r_1, t_1)\mathbf{C}^m_n(\theta_1, \phi_1)\}. \quad (3.2)$$

The current density $\mathbf{J}$ has an expansion of just the form (3.2), but involving the expansion coefficients $J^m_{r,n}(r_1, t_1)$, $J^m_{B,n}(r_1, t_1)$, and $J^m_{C,n}(r_1, t_1)$.

From the definitions of the vector spherical harmonics (in Ref. 2), it can be seen that $\mathbf{B}^0_0$ and $\mathbf{C}^0_0$ are both identically zero, so that for $n = 0$ only the terms involving $\mathbf{P}^0_0$ are nonvanishing. The definitions also show that the vector spherical harmonics which differ only in the algebraic sign of m are complex conjugates. Since the vector functions $\mathbf{J}$, $\mathfrak{F}$, and $\mathbf{\Pi}$ are real quantities, not complex quantities (in this time-domain analysis), it consequently follows that the expansion coefficients which differ only in the sign of m are also complex conjugates, as illustrated by

$$\mathfrak{F}^{-m}_{r,n}(r_1, t_1) = [\mathfrak{F}^{+m}_{r,n}(r_1, t_1)]^*. \quad (3.3)$$

4. DEFINITION OF θ' AND ϕ'

It will be convenient to define a polar coordinate system whose polar axis lies along the field-point vector $\mathbf{r}$. In terms of this coordinate frame, the source-point vector $\mathbf{r}_1$ has the polar coordinates (r_1, θ', ϕ'), and the source-point volume element is given by

$$dV_1 = r_1^2 \sin\theta' \, dr_1 \, d\theta' \, d\phi'. \quad (4.1)$$

The polar angle θ' is the angle included between the vectors $\mathbf{r}$ and $\mathbf{r}_1$, so that

$$\mathbf{r} \cdot \mathbf{r}_1 = r r_1 \cos\theta', \quad (4.2)$$

and the azimuth angle ϕ' will be defined as shown in Fig. 3 of Ref. 6. The trigonometric transformation which expresses (θ_1, ϕ_1) in terms of (θ, ϕ) and (θ', ϕ') is given as Eqs. (3.5) of Ref. 6.

In this coordinate system, the Hertz vector (2.3) takes the form

$$\mathbf{\Pi}(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \int_{r_1=0}^{r_1=\infty} r_1^2 \, dr_1 \int_{\theta'=0}^{\theta'=\pi} \frac{1}{s} \sin\theta' \, d\theta' \times \int_{\phi'=0}^{\phi'=2\pi} \mathfrak{F}(\mathbf{r}_1, t - s/c) \, d\phi'. \quad (4.3)$$

5. INTEGRATION OVER ϕ'

In the expression for the Hertz vector [Eq. (4.3)], the integration over the azimuth angle ϕ' will be carried out first. The geometrical configuration is shown in Fig. 1 of Ref. 2. The source-point vector $\mathbf{r}_1$ is maintained at the constant length r_1 and is swung about the field-point vector $\mathbf{r}$ with the angle θ' between these two vectors held constant.

The integration over ϕ' is a full circuit, from $\phi' = 0$ to $\phi' = 2\pi$, and in this integration the distance s, defined in (2.4), remains constant (since θ' is constant). Thus in the integrand of the ϕ' integration, only the source-point angles (θ_1, ϕ_1) will vary. By (3.2) it can be seen that the variation is therefore confined to the vector spherical harmonics themselves and that the expansion coefficients will remain constant.

The theorems of Ref. 2 can now be brought into the analysis and utilized in the ϕ' integration of (4.3). In this way all nine components of the three vector spherical harmonics for a given (n, m) can be integrated over ϕ', giving explicit functions of $\cos\theta'$, multiplying vector spherical harmonics of (θ, ϕ).

6. INTEGRATION OVER θ'

After the integration over ϕ' has been done, the integration over θ' takes the general form

$$F(r, r_1, t) = \int_{\theta'=0}^{\theta'=\pi} \frac{1}{s} g(\cos\theta') f(r_1, t - s/c) \sin\theta' \, d\theta', \quad (6.1)$$

where f is one of the expansion coefficients in (3.2) and g is an explicit function obtained from one of the six integral theorems in Ref. 2.

During the integration (6.1), the distances r and r_1 are held constant, but the distance s changes as θ' changes. By the law of cosines, the relationship (2.4) can be written as

$$s^2 = r^2 + r_1^2 - 2 r r_1 \cos\theta', \quad (6.2)$$

and its differentiation, with r and r_1 held constant, gives

$$s \, ds = r r_1 \sin\theta' \, d\theta'. \quad (6.3)$$

Equation (6.1) can thus be replaced by

$$F(r, r_1, t) = \int_{s=|r-r_1|}^{s=(r+r_1)} \frac{1}{r r_1} g(\zeta) f(r_1, t - s/c) \, ds, \quad (6.4)$$

where the quantity ζ is defined by

$$\zeta = \cos\theta' = \tfrac{1}{2}(r r_1)^{-1}(r^2 + r_1^2 - s^2). \quad (6.5)$$

The integration (6.4) can be rewritten also as an

[6] R. E. Clapp, J. Math. Phys. 11, 1 (1970).

integration over the source-point time variable t_1 by

$$t_1 = t - s/c, \tag{6.6}$$

$$dt_1 = -(1/c)\, ds. \tag{6.7}$$

In this case the quantity ζ is expressed as a function of t_1 having the form ζ_t, where

$$\zeta_t = \tfrac{1}{2}(rr_1)^{-1}[r^2 + r_1^2 - c^2(t - t_1)^2]. \tag{6.8}$$

The integration then has the appearance

$$F(r, r_1, t) = \int_{t_1=t-(1/c)(r+r_1)}^{t_1=t-(1/c)|r-r_1|} \frac{c}{rr_1}\, g(\zeta_t) f(r_1, t_1)\, dt_1. \tag{6.9}$$

The function $g(\zeta)$, obtained from the integral theorems of Ref. 2, is in each case a polynomial in ζ, and by (6.5) it is therefore an even polynomial in s. For each such function, an associated function $G(r, r_1, s)$ can be defined by

$$G(r, r_1, s) = \int_{s=0}^{s=s} g(\zeta)\, ds, \tag{6.10}$$

and this associated function is now an odd polynomial in s.

The integration over θ', as transformed by (6.4) into an integration over s, will be carried out as an integration by parts, with the aid of (6.10). The result is

$$\begin{aligned} F(r, r_1, t) = {} & \left[\frac{1}{rr_1} G(r, r_1, s) f(r_1, t - s/c)\right]_{s=|r-r_1|}^{s=(r+r_1)} \\ & - \int_{s=|r-r_1|}^{s=(r+r_1)} \frac{1}{rr_1} G(r, r_1, s) \\ & \times \left[\frac{\partial}{\partial s} f\left(r_1, t - \frac{s}{c}\right)\right] ds. \end{aligned} \tag{6.11}$$

In each case to be considered, the function $f(r_1, t_1)$ represents an expansion coefficient which has the time dependence of a component of the free-charge polarization vector $\mathfrak{P}$. The time derivative of such a component gives the corresponding component of the source-current density $\mathbf{J}$ as shown by (2.1). Thus the derivative $\partial f/\partial s$ in (6.11) actually represents a current component. An example is given by

$$\frac{\partial}{\partial s} \mathfrak{P}_{r,n}^m\left(r_1, t - \frac{s}{c}\right) = -\frac{1}{c} J_{r,n}^m\left(r_1, t - \frac{s}{c}\right). \tag{6.12}$$

Similarly, where f appears in (6.11) undifferentiated, representing for example the coefficient $\mathfrak{P}_{r,n}^m$, it can be written as a time integral of the coefficient $J_{r,n}^m$:

$$\mathfrak{P}_{r,n}^m(r_1, t - s/c) = \int_{t_1=t_0}^{t_1=t-s/c} J_{r,n}^m(r_1, t_1)\, dt_1, \tag{6.13}$$

where t_0 is to be chosen as a time which precedes any current flow. [As mentioned in connection with Eq. (2.2), any pre-existing static field, formed by an earlier charge displacement which is not included in the transient current that is being analyzed, can be represented by a constant of integration added to the right-hand side of (6.13).]

7. MATRIX REPRESENTATION

After the integration over ϕ' and the transformation from (6.1) to (6.4) but before the integration over s, the Hertz-vector expansion coefficients can be represented compactly through the matrix equation:

$$\begin{aligned} \Pi_{\sigma,n}^m(r, t) = {} & \frac{1}{2\epsilon_0} \int_{r_1=0}^{r_1=\infty} \frac{r_1}{r}\, dr_1 \\ & \times \int_{s=|r-r_1|}^{s=(r+r_1)} g_{\sigma,\lambda}^{(n)}(\zeta) \mathfrak{P}_{\lambda,n}^m\left(r_1, t - \frac{s}{c}\right) ds. \end{aligned} \tag{7.1}$$

In (7.1) each of the Greek indices σ and λ runs through the three values r, B, and C. A repeated Greek index indicates summation over these values. Thus (7.1) represents three equations, each of which may involve three expansion coefficients for the free-charge polarization vector $\mathfrak{P}$.

However, some of the elements of the matrix $g_{\sigma,\lambda}^{(n)}(\zeta)$ are zero, so that the equations are in fact relatively simple. The matrix elements, determined through the use of the theorems in Ref. 2, are

$$g_{r,r}^{(n)}(\zeta) = \zeta P_n(\zeta), \tag{7.2a}$$

$$g_{B,B}^{(n)}(\zeta) = \zeta P_n(\zeta) + \frac{(1 - \zeta^2)}{n(n+1)} \frac{d}{d\zeta} P_n(\zeta), \tag{7.2b}$$

$$g_{C,C}^{(n)}(\zeta) = P_n(\zeta), \tag{7.2c}$$

$$g_{r,B}^{(n)}(\zeta) = g_{B,r}^{(n)}(\zeta) = \frac{(1 - \zeta^2)}{[n(n+1)]^{\frac{1}{2}}} \frac{d}{d\zeta} P_n(\zeta), \tag{7.2d}$$

$$g_{r,C}^{(n)}(\zeta) = g_{C,r}^{(n)}(\zeta) = 0, \tag{7.2e}$$

$$g_{B,C}^{(n)}(\zeta) = g_{C,B}^{(n)}(\zeta) = 0. \tag{7.2f}$$

When the integrals having the form (6.4) are integrated by parts, as illustrated in (6.11), then the Hertz-vector expansion coefficients (7.1) take the

form

$$\Pi^m_{\sigma,n}(r,t) = \frac{1}{2\epsilon_0}\int_{r_1=0}^{r_1=\infty}\frac{r_1}{r}\,dr_1 G^{(n)}_{\sigma,\lambda}(r, r_1, r+r_1) \times \int_{t_1=t_0}^{t_1=t-(r+r_1)/c} J^m_{\lambda,n}(r_1, t_1)\,dt_1 - \frac{1}{2\epsilon_0}\int_{r_1=0}^{r_1=r}\frac{r_1}{r}\,dr_1 G^{(n)}_{\sigma,\lambda}(r, r_1, r-r_1) \times \int_{t_1=t_0}^{t_1=t-(r-r_1)/c} J^m_{\lambda,n}(r_1, t_1)\,dt_1 - \frac{1}{2\epsilon_0}\int_{r_1=r}^{r_1=\infty}\frac{r_1}{r}\,dr_1 G^{(n)}_{\sigma,\lambda}(r, r_1, r_1-r) \times \int_{t_1=t_0}^{t_1=t-(r_1-r)/c} J^m_{\lambda,n}(r_1, t_1)\,dt_1 + \frac{1}{2\epsilon_0}\int_{r_1=0}^{r_1=r}\frac{r_1}{r}\,dr_1 \times \int_{t_1=t-(r+r_1)/c}^{t_1=t-(r-r_1)/c} G^{(n)}_{\sigma,\lambda}(r, r_1, ct-ct_1) \times J^m_{\lambda,n}(r_1, t_1)\,dt_1 + \frac{1}{2\epsilon_0}\int_{r_1=r}^{r_1=\infty}\frac{r_1}{r}\,dr_1 \times \int_{t_1=t-(r+r_1)/c}^{t_1=t-(r_1-r)/c} G^{(n)}_{\sigma,\lambda}(r, r_1, ct-ct_1) \times J^m_{\lambda,n}(r_1, t_1)\,dt_1. \tag{7.3}$$

The matrix elements $G^{(n)}_{\sigma,\lambda}$ are related to the matrix elements $g^{(n)}_{\sigma,\lambda}$ through an equation of the form (6.10):

$$G^{(n)}_{\sigma,\lambda}(r, r_1, s) = \int_{s=0}^{s=s} g^{(n)}_{\sigma,\lambda}(\zeta)\,ds; \tag{7.4}$$

therefore they can be obtained explicitly with the aid of (7.2).

8. SUMMARY

Equation (7.3), together with (3.1) and (3.2), gives the Hertz vector $\mathbf{\Pi}(\mathbf{r}, t)$ in terms of the source-current density function $\mathbf{J}(\mathbf{r}_1, t_1)$. The use of the expansion in terms of vector spherical harmonics, for both the Hertz vector and the source-current density, has provided a mode separation, in which a particular source mode, characterized by (n, m), leads to a field mode which is also characterized by (n, m). A source current with the symmetry of the vector spherical harmonic $\mathbf{C}^m_n$ leads to a field mode with this same symmetry, but there is cross-coupling between source and field modes having the symmetries of the vector spherical harmonics $\mathbf{P}^m_n$ and $\mathbf{B}^m_n$.

The retarded Hertz vector $\mathbf{\Pi}(\mathbf{r}, t)$ is expressed as an integral over the source current $\mathbf{J}(\mathbf{r}_1, t_1)$ within that region of space–time which is consistent with causality requirements. Since the original form of the Hertz vector, given in Eq. (4.3), is in accord with the requirements of causality, and since there is no contamination by "advanced" fields either in (4.3) or in any later stages of the analysis, it can be concluded that the Hertz vector in the form (7.3), though expressed in the r, t plane where causality requirements are not very transparent, will nevertheless remain fully consistent with the physical requirement that the electromagnetic effect of a moving charged particle should travel at the velocity of light if the medium is the vacuum.

The explicit calculation of the electric and magnetic fields from the Hertz vector given here in (7.3) and (3.1), will be carried out in two following articles.[3,4]

Magnetic Field Generated by a Transient Current Distribution*

ROGER E. CLAPP AND LUC HUANG†
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts

(Received 11 August 1967)

The preceding paper [J. Math. Phys. **11**, 9 (1970)] gave the retarded Hertz vector in terms of a general transient source-current distribution. Here, this Hertz vector is differentiated to give the vector potential and the magnetic field vector. All of these vector quantities have been expanded in vector spherical harmonics, and in each case it is the expansion coefficients, which are functions of the radial distance r and the time t, which are expressed in the (r, t) plane as two-dimensional integrals over the source-current expansion coefficients having the corresponding values of n and m, where these are the mode parameters characterizing the vector spherical harmonics $\mathbf{P}_n^m(\theta, \phi)$, $\mathbf{B}_n^m(\theta, \phi)$, and $\mathbf{C}_n^m(\theta, \phi)$.

1. INTRODUCTION

In a previous article,[1] the retarded Hertz vector was obtained for a fully general transient source-current distribution. The source current $\mathbf{J}(\mathbf{r}_1, t_1)$ was expressed as an expansion in vector spherical harmonics, with the dependence upon the angles (θ_1, ϕ_1) contained within these harmonics, while the dependence upon the radial distance r_1 and the source time t_1 was contained within the expansion coefficients. Similarly, the Hertz vector $\mathbf{\Pi}(\mathbf{r}, t)$ was expressed as an expansion in vector spherical harmonics which were functions of the angles (θ, ϕ), with expansion coefficients which were functions of the radial distance r and the time t.

It was found that, for a given (m, n) labeling the vector spherical harmonics, the expansion coefficients for the Hertz vector could be expressed in terms of the expansion coefficients for the source current. There was no coupling between the expansion coefficients for different choices of n or for different choices of m. There was, however, coupling between two of the three vector harmonics—those denoted by $\mathbf{P}_n^m$ and $\mathbf{B}_n^m$—but no cross-coupling between either of these two and the third vector harmonic, denoted by $\mathbf{C}_n^m$.

In the present article, this Hertz vector is substituted into the equation

$$\mathbf{H} = \epsilon_0 \nabla \times \frac{\partial}{\partial t} \mathbf{\Pi}, \tag{1.1}$$

which gives the magnetic field $\mathbf{H}$ in terms of the Hertz vector $\mathbf{\Pi}$ in MKS units for which

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ farad/meter}. \tag{1.2}$$

An alternative formulation will also be given in terms of the vector potential $\mathbf{A}$, defined here by the equation

$$\mathbf{A} = \frac{1}{c^2} \frac{\partial}{\partial t} \mathbf{\Pi}. \tag{1.3}$$

For the quasi-vacuum conditions that have been postulated, the magnetic field $\mathbf{H}$ is then given by

$$\mathbf{H} = \frac{1}{\mu_0} \nabla \times \mathbf{A}, \tag{1.4}$$

where

$$\frac{1}{\mu_0} = \epsilon_0 c^2. \tag{1.5}$$

2. MAGNETIC FIELD EXPANSION

If the Hertz vector and the magnetic field vector are both expressed in polar coordinate components, then the vector equation (1.1) separates into the component equations:

$$H_r = \frac{\epsilon_0}{r \sin\theta}\left[- \frac{\partial}{\partial \phi} \frac{\partial}{\partial t} \Pi_\theta + \frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial t} \Pi_\phi \right) \right], \tag{2.1a}$$

$$H_\theta = \frac{\epsilon_0}{r}\left[\frac{1}{\sin\theta} \frac{\partial}{\partial \phi} \frac{\partial}{\partial t} \Pi_r - \frac{\partial}{\partial r}\left(r \frac{\partial}{\partial t} \Pi_\phi \right) \right], \tag{2.1b}$$

$$H_\phi = \frac{\epsilon_0}{r}\left[- \frac{\partial}{\partial \theta} \frac{\partial}{\partial t} \Pi_r + \frac{\partial}{\partial r}\left(r \frac{\partial}{\partial t} \Pi_\theta \right) \right]. \tag{2.1c}$$

However, it will be more convenient and will lead to more compact field expressions if the magnetic field vector is first expanded in terms of the vector spherical harmonics defined in Sec. 2 of Ref. 2. This expansion can be written in the form

$$\mathbf{H}(\mathbf{r}, t) = \sum_{n=1}^{\infty} \sum_{m=-n}^{+n} \{ H_{r,n}^m(r, t) \mathbf{P}_n^m(\theta, \phi) + H_{B,n}^m(r, t) \mathbf{B}_n^m(\theta, \phi) + H_{C,n}^m(r, t) \mathbf{C}_n^m(\theta, \phi) \}. \tag{2.2}$$

This summation begins with $n = 1$ because the spherically symmetric current component with $n = 0$

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.

† Present address: Department of Physics, University of Vermont, Burlington, Vermont.

[1] R. E. Clapp, L. Huang, and H. T. Li, J. Math. Phys. **11**, 9 (1970) (preceding paper).

[2] R. E. Clapp and H. T. Li, J. Math. Phys. **11**, 4 (1970).

does not give rise to any component of magnetic field, but only to a spherically symmetric, radially-directed electric field.

The expansion (2.2) parallels the similar expansions for the Hertz vector and for the current-density vector, as described in Sec. 3 of Ref. 1. As in the earlier expansions, the reality of the field vector requires that expansion coefficients which differ only in the sign of m should be complex conjugates of each other, as in the example

$$H^{-m}_{r,n}(r, t) = [H^{+m}_{r,n}(r, t)]^*. \tag{2.3}$$

3. TIME DERIVATIVE OF HERTZ VECTOR

When the Hertz vector, in the matrix form in Eq. (7.1) of Ref. 1, is differentiated with respect to the observation time t, the result is

$$\frac{\partial}{\partial t}\Pi^m_{\sigma,n}(r, t) = \frac{1}{2\epsilon_0}\int_{r_1=0}^{r_1=\infty}\frac{r_1}{r}\,dr_1 \times \int_{s=|r-r_1|}^{s=(r+r_1)} g^{(n)}_{\sigma,\lambda}(\zeta)J^m_{\lambda,n}\left(r_1, t-\frac{s}{c}\right)ds. \tag{3.1}$$

The notation here is the same as the notation in Ref. 1, with Greek subscripts running through the three values r, B, C, associated with the three vector spherical harmonics $\mathbf{P}^m_n$, $\mathbf{B}^m_n$, $\mathbf{C}^m_n$. As in (2.2), the dependence upon time and radial distance lies with the expansion coefficients, so that when the time derivative is taken it is only necessary to differentiate these expansion coefficients, as was done here in (3.1).

When the integration over ds in (3.1) is replaced [with the use of (6.4)–(6.9) of Ref. 1] by an integration over dt_1, then (3.1) takes the form

$$\frac{\partial}{\partial t}\Pi^m_{\sigma,n}(r, t) = \frac{c}{2\epsilon_0}\left(\iint\right)\frac{r_1}{r}\,g^{(n)}_{\sigma,\lambda}(\zeta_t)J^m_{\lambda,n}(r_1, t_1)\,dt_1\,dr_1, \tag{3.2}$$

where the abbreviation $(\iint)$ has the following equivalent meanings:

$$\left(\iint\right) = \int_{r_1=0}^{r_1=\infty}\int_{t_1=t-(r+r_1)/c}^{t_1=t-|r-r_1|/c}, \tag{3.3a}$$

$$\left(\iint\right) = \int_{t_1=-\infty}^{t_1=t}\int_{r_1=|r-c(t-t_1)|}^{r_1=r+c(t-t_1)}, \tag{3.3b}$$

$$\left(\iint\right) = \int_{r_1=0}^{r_1=r}\int_{t_1=t-(r+r_1)/c}^{t_1=t-(r-r_1)/c} + \int_{r_1=r}^{r_1=\infty}\int_{t_1=t-(r+r_1)/c}^{t_1=t-(r_1-r)/c}, \tag{3.3c}$$

$$\left(\iint\right) = \int_{t_1=-\infty}^{t_1=t-r/c}\int_{r_1=-r+c(t-t_1)}^{r_1=r+c(t-t_1)} + \int_{t_1=t-r/c}^{t_1=t}\int_{r_1=r-c(t-t_1)}^{r_1=r+c(t-t_1)}, \tag{3.3d}$$

which define the realm of integration in the (r_1, t_1) plane.

4. VECTOR IDENTITIES

While a number of mathematical identities involving vector spherical harmonics have been given by Morse and Feshbach,[3] they are not in the form that is needed here. For the present application, there are three curl identities,

$$\nabla \times \mathbf{P}^m_n(\theta, \phi) = \{[n(n+1)]^{\frac{1}{2}}/r\}\mathbf{C}^m_n(\theta, \phi), \tag{4.1}$$

$$\nabla \times \mathbf{B}^m_n(\theta, \phi) = -(1/r)\mathbf{C}^m_n(\theta, \phi), \tag{4.2}$$

$$\nabla \times \mathbf{C}^m_n(\theta, \phi) = (1/r)\mathbf{B}^m_n(\theta, \phi) + \{[n(n+1)]^{\frac{1}{2}}/r\}\mathbf{P}^m_n(\theta, \phi), \tag{4.3}$$

which can readily be generalized to the situation in which each vector spherical harmonic is multiplied by a scalar function of the form $Q(r, t)$. The generalized identities are

$$\nabla \times [Q\mathbf{P}^m_n] = \{[n(n+1)]^{\frac{1}{2}}/r\}Q\mathbf{C}^m_n, \tag{4.4}$$

$$\nabla \times [Q\mathbf{B}^m_n] = -\frac{1}{r}\frac{\partial}{\partial r}(rQ)\mathbf{C}^m_n, \tag{4.5}$$

$$\nabla \times [Q\mathbf{C}^m_n] = \frac{1}{r}\frac{\partial}{\partial r}(rQ)\mathbf{B}^m_n + \frac{[n(n+1)]^{\frac{1}{2}}}{r}Q\mathbf{P}^m_n. \tag{4.6}$$

5. MAGNETIC FIELD

In the interest of a compact notation, the three vector spherical harmonics will be represented collectively by $\mathbf{V}^m_{\sigma,n}(\theta, \phi)$, where the Greek index σ takes on the three values r, B, C, as in (3.1). The explicit definitions are

$$\mathbf{V}^m_{r,n} = \mathbf{P}^m_n, \tag{5.1a}$$

$$\mathbf{V}^m_{B,n} = \mathbf{B}^m_n, \tag{5.1b}$$

$$\mathbf{V}^m_{C,n} = \mathbf{C}^m_n. \tag{5.1c}$$

In terms of this new notation and with the summation convention for repeated Greek indices, (2.2) can be written compactly as

$$\mathbf{H}(\mathbf{r}, t) = \sum_{n=1}^{\infty}\sum_{m=-n}^{+n} H^m_{\sigma,n}(r, t)\mathbf{V}^m_{\sigma,n}(\theta, \phi), \tag{5.2}$$

with similar expressions for the Hertz vector and its time derivative.

[3] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., 1953), pp. 1898–1901.

In particular, the time derivative of the Hertz vector can be given in the form

$$\frac{\partial}{\partial t}\mathbf{\Pi} = \frac{c}{2\epsilon_0}\sum_{n,m}\mathbf{V}^m_{\sigma,n}\left(\iint\right)\frac{r_1}{r}\, g^{(n)}_{\sigma,\lambda}(\zeta_t)J^m_{\lambda,n}(r_1, t_1)\, dt_1\, dr_1. \tag{5.3}$$

Substituting into (1.1) with the use of (4.4)–(4.6) gives

$$H^m_{r,n}(r, t) = \frac{c}{2r^2}[n(n+1)]^{\frac{1}{2}}\left(\iint\right)P_n(\zeta_t)r_1 J^m_{C,n}(r_1, t_1)\, dt_1\, dr_1, \tag{5.4}$$

$$H^m_{B,n}(r, t) = \frac{c}{2r}\frac{\partial}{\partial r}\left[\left(\iint\right)P_n(\zeta_t)r_1 J^m_{C,n}(r_1, t_1)\, dt_1\, dr_1\right], \tag{5.5}$$

$$H^m_{C,n}(r, t) = \frac{c}{2r}[n(n+1)]^{\frac{1}{2}}\left(\iint\right)P_n(\zeta_t)J^m_{r,n}(r_1, t_1)\, dt_1\, dr_1 - \frac{c}{2r}\left(\iint\right)P_n(\zeta_t)\frac{\partial}{\partial r_1}[r_1 J^m_{B,n}(r_1, t_1)]\, dt_1\, dr_1. \tag{5.6}$$

In Eq. (5.6), the second expression on the right-hand side is obtained through an integration by parts, in which the form (3.3d) is used for the integral operator.

6. DISCUSSION

Equations (5.4)–(5.6) give the magnetic field expansion coefficients, which are to be inserted into (5.2). It can be seen that there is an element of symmetry in the dependence of the three magnetic-field components upon the three current components, but that the symmetry is not as conveniently expressed in matrix form as was the case for the Hertz vector itself, as given in Eq. (7.1) of Ref. 1.

While the forms (5.5) and (5.6) are chosen here because of their compactness, there are other forms for these magnetic-field components which avoid the use of the operators $\partial/\partial r$ and $\partial/\partial r_1$. These other forms can be obtained from (5.5) and (5.6) through integration by parts and by carrying out indicated differentiations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Electric Field Generated by a Transient Current Distribution*

ROGER E. CLAPP AND LUC HUANG†
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts
AND
H. TUNG LI
AirTechnology Corporation, Waltham, Massachusetts
and
College of the Holy Cross, Worcester, Massachusetts

(Received 11 August 1967)

The retarded Hertz vector, obtained earlier [J. Math. Phys. **11**, 9 (1970)], is differentiated to give the scalar potential and the electric-field vector associated with a general transient source-current distribution. The scalar and vector quantities are expanded in terms of scalar and vector spherical harmonics. For each mode, characterized by specific values for n and m in the expansions, the scalar potential and electric-field vector are expressed in the (r, t) plane as two-dimensional integrals over the causally accessible portion of the source-current distribution.

1. INTRODUCTION

In a preceding article,[1] expressions were obtained for the magnetic-field components associated with a general transient current distribution. Now, similar expressions are derived for the electric-field components, and, as before, these are derived from the previously derived expressions for the components of the retarded Hertz vector.[2]

In terms of the retarded Hertz vector $\mathbf{\Pi}$, the electric-field vector is given by the equation

$$\mathbf{E} = \nabla\nabla\cdot\mathbf{\Pi} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{\Pi}. \tag{1.1}$$

In terms of the more familiar scalar and vector potentials, the electric vector is

$$\mathbf{E} = -\nabla\varphi - \frac{\partial}{\partial t}\mathbf{A}. \tag{1.2}$$

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.

† Present address: Department of Physics, College of the Holy Cross, Worcester, Massachusetts.

[1] R. E. Clapp and L. Huang, J. Math. Phys. **11**, 14 (1970).

[2] R. E. Clapp, L. Huang, and H. T. Li, J. Math. Phys. **11**, 9 (1970).

In particular, the time derivative of the Hertz vector can be given in the form

$$\frac{\partial}{\partial t}\mathbf{\Pi} = \frac{c}{2\epsilon_0}\sum_{n,m}\mathbf{V}^m_{\sigma,n}\left(\iint\right)\frac{r_1}{r}\, g^{(n)}_{\sigma,\lambda}(\zeta_t)J^m_{\lambda,n}(r_1, t_1)\, dt_1\, dr_1. \tag{5.3}$$

Substituting into (1.1) with the use of (4.4)–(4.6) gives

$$H^m_{r,n}(r, t) = \frac{c}{2r^2}[n(n+1)]^{\frac{1}{2}}\left(\iint\right)P_n(\zeta_t)r_1 J^m_{C,n}(r_1, t_1)\, dt_1\, dr_1, \tag{5.4}$$

$$H^m_{B,n}(r, t) = \frac{c}{2r}\frac{\partial}{\partial r}\left[\left(\iint\right)P_n(\zeta_t)r_1 J^m_{C,n}(r_1, t_1)\, dt_1\, dr_1\right], \tag{5.5}$$

$$H^m_{C,n}(r, t) = \frac{c}{2r}[n(n+1)]^{\frac{1}{2}}\left(\iint\right)P_n(\zeta_t)J^m_{r,n}(r_1, t_1)\, dt_1\, dr_1 - \frac{c}{2r}\left(\iint\right)P_n(\zeta_t)\frac{\partial}{\partial r_1}[r_1 J^m_{B,n}(r_1, t_1)]\, dt_1\, dr_1. \tag{5.6}$$

In Eq. (5.6), the second expression on the right-hand side is obtained through an integration by parts, in which the form (3.3d) is used for the integral operator.

6. DISCUSSION

Equations (5.4)–(5.6) give the magnetic field expansion coefficients, which are to be inserted into (5.2). It can be seen that there is an element of symmetry in the dependence of the three magnetic-field components upon the three current components, but that the symmetry is not as conveniently expressed in matrix form as was the case for the Hertz vector itself, as given in Eq. (7.1) of Ref. 1.

While the forms (5.5) and (5.6) are chosen here because of their compactness, there are other forms for these magnetic-field components which avoid the use of the operators $\partial/\partial r$ and $\partial/\partial r_1$. These other forms can be obtained from (5.5) and (5.6) through integration by parts and by carrying out indicated differentiations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Electric Field Generated by a Transient Current Distribution*

ROGER E. CLAPP AND LUC HUANG†
AirTechnology Corporation, Waltham, Massachusetts
and
Basic Research Associates, Cambridge, Massachusetts
AND
H. TUNG LI
AirTechnology Corporation, Waltham, Massachusetts
and
College of the Holy Cross, Worcester, Massachusetts

(Received 11 August 1967)

The retarded Hertz vector, obtained earlier [J. Math. Phys. **11**, 9 (1970)], is differentiated to give the scalar potential and the electric-field vector associated with a general transient source-current distribution. The scalar and vector quantities are expanded in terms of scalar and vector spherical harmonics. For each mode, characterized by specific values for n and m in the expansions, the scalar potential and electric-field vector are expressed in the (r, t) plane as two-dimensional integrals over the causally accessible portion of the source-current distribution.

1. INTRODUCTION

In a preceding article,[1] expressions were obtained for the magnetic-field components associated with a general transient current distribution. Now, similar expressions are derived for the electric-field components, and, as before, these are derived from the previously derived expressions for the components of the retarded Hertz vector.[2]

In terms of the retarded Hertz vector $\mathbf{\Pi}$, the electric-field vector is given by the equation

$$\mathbf{E} = \nabla\nabla\cdot\mathbf{\Pi} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{\Pi}. \tag{1.1}$$

In terms of the more familiar scalar and vector potentials, the electric vector is

$$\mathbf{E} = -\nabla\varphi - \frac{\partial}{\partial t}\mathbf{A}. \tag{1.2}$$

* This work was supported by the U.S. Defense Atomic Support Agency, under Contract DA-49-146-XZ-402.

† Present address: Department of Physics, College of the Holy Cross, Worcester, Massachusetts.

[1] R. E. Clapp and L. Huang, J. Math. Phys. **11**, 14 (1970).

[2] R. E. Clapp, L. Huang, and H. T. Li, J. Math. Phys. **11**, 9 (1970).

The identification of the vector potential

$$\mathbf{A} = \frac{1}{c^2}\frac{\partial}{\partial t}\mathbf{\Pi} \tag{1.3}$$

has already been made in (1.3) of Ref. 1. A corresponding identification of the scalar potential, in terms of the Hertz vector, is

$$\varphi = -\nabla \cdot \mathbf{\Pi}. \tag{1.4}$$

2. SCALAR IDENTITIES

In analogy with the vector identities given in Sec. 4 of Ref. 1, there are certain scalar identities which arise from the action of the divergence operator upon the vector spherical harmonics and upon products of a scalar function $Q(r, t)$ and the vector spherical harmonics.

The three basic divergence identities are

$$\nabla \cdot \mathbf{P}_n^m(\theta, \phi) = (2/r)X_n^m(\theta, \phi), \tag{2.1}$$

$$\nabla \cdot \mathbf{B}_n^m(\theta, \phi) = -\{[n(n+1)]^{\frac{1}{2}}/r\}X_n^m(\theta, \phi), \tag{2.2}$$

$$\nabla \cdot \mathbf{C}_n^m(\theta, \phi) = 0. \tag{2.3}$$

The scalar spherical harmonics X_n^m and the vector spherical harmonics $\mathbf{P}_n^m$, $\mathbf{B}_n^m$, and $\mathbf{C}_n^m$ have been defined in Sec. 2 of Ref. 3.

The more general divergence identities, in which each vector spherical harmonic is multiplied by the scalar function $Q(r, t)$, are

$$\nabla \cdot [Q\mathbf{P}_n^m] = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 Q)X_n^m, \tag{2.4}$$

$$\nabla \cdot [Q\mathbf{B}_n^m] = -\{[n(n+1)]^{\frac{1}{2}}/r\}QX_n^m, \tag{2.5}$$

$$\nabla \cdot [Q\mathbf{C}_n^m] = 0. \tag{2.6}$$

As will be seen, it is the vanishing of the divergence in (2.6) that can be associated with the fact that the source-current components involving the vector spherical harmonics $\mathbf{C}_n^m$ do not lead to any charge accumulation; hence these currents do not give any electrostatic contribution to $\mathbf{E}(\mathbf{r}, t)$. These currents are circulatory in character, and make only inductive and radiative-field contributions.

3. DIVERGENCE OF HERTZ VECTOR

In the evaluation of the divergence of the Hertz vector, one of the functions defined in Sec. 7 of Ref. 2 appears repeatedly. This is the function

$$G_{C,C}^{(n)}(r, r_1, s) = \int_{s=0}^{s=s} P_n(\zeta)\, ds, \tag{3.1}$$

where

$$\zeta = (1/2rr_1)(r^2 + r_1^2 - s^2). \tag{3.2}$$

[3] R. E. Clapp and H. T. Li, J. Math. Phys. **11**, 4 (1970).

Special values of this function are

$$G_{C,C}^{(n)}(r, r_1, r + r_1) = (2n+1)^{-1}(r^{n+1}r_1^{-n} + r^{-n}r_1^{n+1}), \tag{3.3a}$$

$$G_{C,C}^{(n)}(r, r_1, r - r_1) = (2n+1)^{-1}(r^{n+1}r_1^{-n} - r^{-n}r_1^{n+1}), \tag{3.3b}$$

$$G_{C,C}^{(n)}(r, r_1, r_1 - r) = (2n+1)^{-1}(-r^{n+1}r_1^{-n} + r^{-n}r_1^{n+1}). \tag{3.3c}$$

Equations (3.3) are established in the Appendix.

The divergence of the Hertz vector does not involve the source currents with the $J_{C,n}^m$ symmetry, as noted earlier in connection with Eq. (2.6). However, both the $J_{r,n}^m$ and the $J_{B,n}^m$ current components make contributions to this divergence. The explicit form for this divergence is

$$\begin{aligned}
\nabla \cdot \mathbf{\Pi} = \Bigg[& -\int_{r_1=0}^{r_1=\infty} r_1^2 \frac{\partial}{\partial r_1}\left(\frac{1}{r_1} G_{C,C}^{(n)}(r, r_1, r + r_1)\right) \\
& \times \int_{t_1=t_0}^{t_1=t-(r+r_1)/c} J_{r,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& + \int_{r_1=0}^{r_1=r} r_1^2 \frac{\partial}{\partial r_1}\left(\frac{1}{r_1} G_{C,C}^{(n)}(r, r_1, r - r_1)\right) \\
& \times \int_{t_1=t_0}^{t_1=t-(r-r_1)/c} J_{r,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& + \int_{r_1=r}^{r_1=\infty} r_1^2 \frac{\partial}{\partial r_1}\left(\frac{1}{r_1} G_{C,C}^{(n)}(r, r_1, r_1 - r)\right) \\
& \times \int_{t_1=t_0}^{t_1=t-(r_1-r)/c} J_{r,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& - \left(\iiint\right) r_1^2 \frac{\partial}{\partial r_1}\left(\frac{1}{r_1} G_{C,C}^{(n)}(r, r_1, ct - ct_1)\right) \\
& \times J_{r,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& - [n(n+1)]^{\frac{1}{2}} \int_{r_1=0}^{r_1=\infty} G_{C,C}^{(n)}(r, r_1, r + r_1) \\
& \times \int_{t_1=t_0}^{t_1=t-(r+r_1)/c} J_{B,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& + [n(n+1)]^{\frac{1}{2}} \int_{r_1=0}^{r_1=r} G_{C,C}^{(n)}(r, r_1, r - r_1) \\
& \times \int_{t_1=t_0}^{t_1=t-(r-r_1)/c} J_{B,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& + [n(n+1)]^{\frac{1}{2}} \int_{r_1=r}^{r_1=\infty} G_{C,C}^{(n)}(r, r_1, r_1 - r) \\
& \times \int_{t_1=t_0}^{t_1=t-(r_1-r)/c} J_{B,n}^m(r_1, t_1)\, dt_1\, dr_1 \\
& - [n(n+1)]^{\frac{1}{2}} \left(\iiint\right) G_{C,C}^{(n)}(r, r_1, ct - ct_1) \\
& \times J_{B,n}^m(r_1, t_1)\, dt_1\, dr_1 \Bigg] \frac{1}{2\epsilon_0 r} X_n^m(\theta, \phi).
\end{aligned} \tag{3.4}$$

The integral symbol ($\iint$) has been defined in Eq. (3.3) of Ref. 1.

4. ELECTRIC FIELD

The electric-field vector is obtained from Eq. (1.1). This includes an operation in which the gradient of the divergence of the Hertz vector is taken, and for this operation it is convenient to make use of the identity

$$\nabla[Q(r,t)X_n^m(\theta,\phi)] = \left(\frac{\partial}{\partial r}Q(r,t)\right)\mathbf{P}_n^m(\theta,\phi) + [n(n+1)]^{\frac{1}{2}}Q(r,t)\mathbf{B}_n^m(\theta,\phi). \tag{4.1}$$

The second time derivative of the Hertz vector also appears, and this is conveniently found from the time derivative of the expression given in Eq. (3.2) of Ref. 1. After an integration by parts, this second time derivative is found to be expressible in the compact matrix form

$$\frac{\partial^2}{\partial t^2}\Pi_{\sigma,n}^m(r,t) = \frac{c}{2\epsilon_0}\left(\iint\right)\frac{r_1}{r}\,g_{\sigma,\lambda}^{(n)}(\zeta_t)\left(\frac{\partial}{\partial t_1}J_{\lambda,n}^m(r_1,t_1)\right)dt_1\,dr_1. \tag{4.2}$$

As before, a repeated Greek index indicates summation over the values r, B, and C, and the functions $g_{\sigma,\lambda}^{(n)}$ are as defined in (7.2) of Ref. 2.

When the operations indicated in Eq. (1.1) are carried out, the resulting expressions for the expansion coefficients for the electric-field vector are

$$\begin{aligned}
E_{r,n}^m(r,t) = {} & -\frac{1}{\epsilon_0}\int_{t_1=t_0}^{t_1=t} J_{r,n}^m(r,t_1)\,dt_1 \\
& + \frac{n(n+1)}{(2n+1)\epsilon_0}\int_{r_1=0}^{r_1=r}\int_{t_1=t_0}^{t_1=t-(r-r_1)/c} r^{-n-2}r_1^{n+1} \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{n(n+1)}{(2n+1)\epsilon_0}\int_{r_1=r}^{r_1=\infty}\int_{t_1=t_0}^{t_1=t-(r_1-r)/c} r^{n-1}r_1^{-n} \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& - \frac{n(n+1)}{2(2n+1)\epsilon_0}\left(\iint\right)(r^{-n-2}r_1^{n+1} + r^{n-1}r_1^{-n}) \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{n(n+1)}{2\epsilon_0 r^2}\left(\iint\right)G_{C,C}^{(n)}(r,r_1,ct-ct_1) \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{(n+1)[n(n+1)]^{\frac{1}{2}}}{(2n+1)\epsilon_0} \\
& \times \int_{r_1=0}^{r_1=r}\int_{t_1=t_0}^{t_1=t-(r-r_1)/c} r^{-n-2}r_1^{n+1} \\
& \times J_{B,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& - \frac{n[n(n+1)]^{\frac{1}{2}}}{(2n+1)\epsilon_0}\int_{r_1=r}^{r_1=\infty}\int_{t_1=t_0}^{t_1=t-(r_1-r)/c} r^{n-1}r_1^{-n} \\
& \times J_{B,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& - \frac{[n(n+1)]^{\frac{1}{2}}}{2(2n+1)\epsilon_0} \\
& \times \left(\iint\right)[(n+1)r^{-n-2}r_1^{n+1} - nr^{n-1}r_1^{-n}] \\
& \times J_{B,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{[n(n+1)]^{\frac{1}{2}}}{2\epsilon_0 r^2} \\
& \times \left(\iint\right)\left[r_1\frac{\partial}{\partial r_1}G_{C,C}^{(n)}(r,r_1,ct-ct_1)\right] \\
& \times J_{B,n}^m(r_1,t_1)\,dt_1\,dr_1,
\end{aligned} \tag{4.3}$$

$$\begin{aligned}
E_{B,n}^m(r,t) = {} & -\frac{n[n(n+1)]^{\frac{1}{2}}}{(2n+1)\epsilon_0}\int_{r_1=0}^{r_1=r}\int_{t_1=t_0}^{t_1=t-(r-r_1)/c} r^{-n-2}r_1^{n+1} \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{(n+1)[n(n+1)]^{\frac{1}{2}}}{(2n+1)\epsilon_0} \\
& \times \int_{r_1=r}^{r_1=\infty}\int_{t_1=t_0}^{t_1=t-(r_1-r)/c} r^{n-1}r_1^{-n} \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{[n(n+1)]^{\frac{1}{2}}}{2(2n+1)\epsilon_0} \\
& \times \left(\iint\right)[nr^{-n-2}r_1^{n+1} - (n+1)r^{n-1}r_1^{-n}] \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& + \frac{[n(n+1)]^{\frac{1}{2}}}{2\epsilon_0 r^2} \\
& \times \left(\iint\right)\left[r\frac{\partial}{\partial r}G_{C,C}^{(n)}(r,r_1,ct-ct_1)\right] \\
& \times J_{r,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& - \frac{1}{2\epsilon_0 cr}\int_{r_1=0}^{r_1=\infty}(-1)^n r_1 \\
& \times J_{B,n}^m\left(r_1,t-\frac{r+r_1}{c}\right)dr_1 \\
& - \frac{1}{2\epsilon_0 cr}\int_{r_1=0}^{r_1=r} r_1 J_{B,n}^m\left(r_1,t-\frac{r-r_1}{c}\right)dr_1 \\
& - \frac{1}{2\epsilon_0 cr}\int_{r_1=r}^{r_1=\infty} r_1 J_{B,n}^m\left(r_1,t-\frac{r_1-r}{c}\right)dr_1 \\
& - \frac{n(n+1)}{(2n+1)\epsilon_0}\int_{r_1=0}^{r_1=r}\int_{t_1=t_0}^{t_1=t-(r-r_1)/c} r^{-n-2}r_1^{n+1} \\
& \times J_{B,n}^m(r_1,t_1)\,dt_1\,dr_1 \\
& - \frac{n(n+1)}{(2n+1)\epsilon_0}\int_{r_1=r}^{r_1=\infty}\int_{t_1=t_0}^{t_1=t-(r_1-r)/c} r^{n-1}r_1^{-n}
\end{aligned}$$

$$\times J^m_{B,n}(r_1, t_1)\, dt_1\, dr_1$$
$$+ \frac{n(n+1)}{2(2n+1)\epsilon_0}\left(\iint\right)(r^{-n-2}r_1^{n+1} + r^{n-1}r_1^{-n})$$
$$\times J^m_{B,n}(r_1, t_1)\, dt_1\, dr_1$$
$$+ \frac{1}{2\epsilon_0 r^2}\left(\iint\right)\left[rr_1 \frac{\partial^2}{\partial r \partial r_1} \times G^{(n)}_{C,C}(r, r_1, ct - ct_1)\right]$$
$$\times J^m_{B,n}(r_1, t_1)\, dt_1\, dr_1, \tag{4.4}$$

$$E^m_{C,n}(r, t) = -\frac{1}{2\epsilon_0 c}\left(\iint\right)\frac{r_1}{r} P_n(\zeta_t) \times \left(\frac{\partial}{\partial t_1} J^m_{C,n}(r_1, t_1)\right) dt_1\, dr_1. \tag{4.5}$$

5. DISCUSSION

Equations (4.3)–(4.5) give the components of the electric field in terms of the components of the source current. These equations are not given in the general matrix form, since their relative complexity precludes for the present a simple matrix representation. However, there are elements of symmetry involved, and a compact matrix representation may eventually be devised.

The electric-field components are here given as explicit integrations over the source-current components. For the inductive and radiative contributions, the integration is limited to the realm specified by the notation $(\iint)$. This realm limits the source currents which are "visible" at the observation point. However, in addition to these immediately-sensed contributions, there are electric-field contributions which are electrostatic and are associated with the earlier flow of current, establishing a distribution of electric charge which generates an electrostatic field. The contributions to the electric field which are associated with this electrostatic dipole-moment distribution are expressed in the integrals which have as their lower limit $t_1 = t_0$, where t_0 is a time preceding any of the transient current flow which contributes to the fields that are being calculated.

The important aspect of Eqs. (4.3)–(4.5) is their expression of the causal relationship between the source current **J** and the electric field **E** which is generated by this source current. When these equations are used as the basis for a numerical solution of Maxwell's equations, the requirements of causality will be met.[4-6]

[4] R. E. Clapp, Proc. IEEE **56**, 329 (1968).
[5] G. H. Peebles and R. E. Clapp, Proc. IEEE **56**, 1365 (1968).
[6] R. E. Clapp, J. Geophys. Res. **73**, 6395 (1968).

APPENDIX

For the derivation of Eqs. (3.3), it is first noted that the Legendre polynomial $P_n(\zeta)$, which can be written in the form

$$P_n(\zeta) = \frac{1}{2^n n!}\frac{d^n}{d\zeta^n}(\zeta^2 - 1)^n, \tag{A1}$$

is an odd polynomial in ζ if n is odd or an even polynomial if n is even, so that

$$P_n(-\zeta) = (-1)^n P_n(\zeta). \tag{A2}$$

When the argument ζ is written out as

$$\zeta = (1/2rr_1)(r^2 + r_1^2 - s^2), \tag{A3}$$

then the Legendre polynomial $P_n(\zeta)$ has the form

$$P_n(\zeta) = [1/(2rr_1)^n] f_n(r^2, r_1^2, s^2), \tag{A4}$$

where the function f_n is a finite polynomial in its three arguments.

It follows that the integral

$$F_n(r, r_1, R) = \int_{s=0}^{s=R} P_n(\zeta)\, ds \tag{A5}$$

is an odd polynomial in R, so that

$$F_n(r, r_1, -R) = -F_n(r, r_1, R). \tag{A6}$$

It can also be shown, from the form of (A4), that

$$F_n(-r, r_1, R) = (-1)^n F_n(r, r_1, R), \tag{A7a}$$
$$F_n(r, -r_1, R) = (-1)^n F_n(r, r_1, R), \tag{A7b}$$
$$F_n(-r, -r_1, R) = F_n(r, r_1, R). \tag{A7c}$$

The three integrals to be evaluated are

$$F_n(r, r_1, r + r_1) = \int_{s=0}^{s=r+r_1} P_n(\zeta)\, ds, \tag{A8a}$$
$$F_n(r, r_1, r - r_1) = \int_{s=0}^{s=r-r_1} P_n(\zeta)\, ds, \tag{A8b}$$
$$F_n(r, r_1, r_1 - r) = \int_{s=0}^{s=r_1-r} P_n(\zeta)\, ds. \tag{A8c}$$

All three are defined here with no restrictions on the relative magnitudes of r and r_1. From (A6) it can be seen that

$$F_n(r, r_1, r - r_1) = -F_n(r, r_1, r_1 - r), \tag{A9a}$$

and from the symmetrical way that r and r_1 enter into ζ in (A3), it is also clear that

$$F_n(r_1, r, r_1 - r) = -F_n(r, r_1, r - r_1), \tag{A9b}$$
$$F_n(r_1, r, r_1 + r) = F_n(r, r_1, r + r_1). \tag{A9c}$$

Equation (A3) can be solved for s, giving

$$s = (r^2 - 2rr_1\zeta + r_1^2)^{\frac{1}{2}}. \tag{A10}$$

When r and r_1 are both held constant, the differentials ds and $d\zeta$ will be related by

$$s\,ds = -rr_1\,d\zeta. \tag{A11}$$

If r_1 is smaller than r, the expansion

$$r/s = P_0(\zeta) + (r_1/r)P_1(\zeta) + (r_1^2/r^2)P_2(\zeta) + \cdots \tag{A12}$$

converges, and it is possible to write

$$\begin{aligned} F_n(r, r_1, r + r_1) - F_n(r, r_1, r - r_1) \\ = \int_{s=r-r_1}^{s=r+r_1} P_n(\zeta)\,ds = -\int_{\zeta=+1}^{\zeta=-1} s^{-1} r r_1 P_n(\zeta)\,d\zeta \\ = r_1 \int_{\zeta=-1}^{\zeta=+1} P_n(\zeta)\frac{r}{s}\,d\zeta = r_1 \frac{r_1^n}{r^n}\frac{2}{(2n+1)}. \end{aligned} \tag{A13}$$

In (A13) the series (A12) has been substituted, and the orthogonality properties of the Legendre polynomials have been utilized.

If, on the other hand, r_1 is greater than r, then the convergent series expansion is

$$r_1/s = P_0(\zeta) + (r/r_1)P_1(\zeta) + (r^2/r_1^2)P_2(\zeta) + \cdots, \tag{A14}$$

and the resulting equation, analogous to (A13), is

$$\begin{aligned} F_n(r, r_1, r + r_1) - F_n(r, r_1, r_1 - r) \\ = r(r^n/r_1^n)2(2n+1)^{-1}. \end{aligned} \tag{A15}$$

From (A4) and (A5), it is apparent that the multiplication of $F_n(r, r_1, R)$ by $(2rr_1)^n$ will cancel the factors in the denominator and give a finite polynomial. In particular, if Eq. (A13) is multiplied on both sides by $(2rr_1)^n$, it becomes a relationship between the sum of two finite polynomials (on the left) and a monomial (on the right). Evidently, all of the terms in the two polynomials must subtract out, except for the terms which add to give the monomial on the right. Moreover, this algebraic relationship involving finite polynomials, while established for r_1 less than r, can be continued analytically to include the regions where r_1 is greater than r. Similarly, Eq. (A15), while established for r_1 greater than r, is a simple algebraic relationship which can be continued analytically to the realm where r_1 is less than r.

When (A9a) is inserted into (A15), the result is

$$\begin{aligned} F_n(r, r_1, r + r_1) + F_n(r, r_1, r - r_1) \\ = r(r^n/r_1^n)2(2n+1)^{-1}. \end{aligned} \tag{A16}$$

Now, from (A13), (A16), and (A9a), the desired results are obtained:

$$F_n(r, r_1, r + r_1) = (2n+1)^{-1}(r^{n+1}r_1^{-n} + r^{-n}r_1^{n+1}), \tag{A17a}$$

$$F_n(r, r_1, r - r_1) = (2n+1)^{-1}(r^{n+1}r_1^{-n} - r^{-n}r_1^{n+1}), \tag{A17b}$$

$$F_n(r, r_1, r_1 - r) = (2n+1)^{-1}(-r^{n+1}r_1^{-n} + r^{-n}r_1^{n+1}). \tag{A17c}$$

These results are equivalent to Eqs. (3.3), since it is apparent from the definition of F_n in (A5) that

$$F_n(r, r_1, R) = G^{(n)}_{C,C}(r, r_1, R). \tag{A18}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Kernel Integral Formulas for the Canonical Commutation Relations of Quantum Fields.* II. Irreducible Representations

GERHARD C. HEGERFELDT†
Institut für Theoretische Physik, Universität Marburg, Germany

(Received 27 December 1968)

Continuing our investigation of the kernel or group integral for the canonical commutation relations introduced by Klauder and McKenna, we prove that a representation fulfilling any sort of kernel integral formula is irreducible. This has been conjectured by Klauder and McKenna. After collecting some auxiliary results, a complete classification of all representations is given for which a kernel integral formula in the form of a limit superior holds. It is shown that these are just the partial tensor-product representations and that the limit superior can be replaced by an ordinary limit over a fixed subsequence, thus allowing the transition from norms to scalar products. Then the basis-independent kernel integral in the form of a sup $\overline{\lim}$ is investigated. Here the supremum is taken over all bases of the test-function space. Under a not-very-strong irreducibility assumption, we show that this can be reduced to the vacuum functional and that there exists a fixed sequence of subspaces of the test-function space such that the sup $\overline{\lim}$ can be replaced by an ordinary limit which again allows a transition to scalar products. The results are strikingly similar to the case of cyclic field. This tempts us to conjecture that a representation fulfilling a kernel integral formula is both irreducible and cyclic with respect to the field just as in the case of finitely many degrees of freedom.

1. INTRODUCTION

In a previous paper,[1] kernel integral formulas for representations of the canonical commutation relations (CCR) of Bose fields have been investigated for the cases where the field is cyclic. This was motivated by the close similarity between the Schrödinger representation for finitely many degrees of freedom and the general form of a representation with cyclic field. A counter example at the end of Part I had shown that not all of these representations can satisfy a kernel integral formula. This suggests that we consider the other important property of the Schrödinger representation—namely, irreducibility. As stressed before, there exist representations with cyclic field which are not irreducible, and vice versa, while the Schrödinger representation is both cyclic with respect to Q, which corresponds to the field, and irreducible.

Let $\mathfrak{V}$ denote the test-function space which is a real and in general incomplete scalar product space. A representation of the CCR is a family of unitary operators $U(f)$, $V(g)$ in a Hilbert space $\mathcal{H}$, f and g in $\mathfrak{V}$, satisfying

$$\begin{aligned} V(g)U(f) &= e^{i(f,g)}U(f)V(g), \\ U(f_1+f_2) &= U(f_1)U(f_2), \\ V(g_1+g_2) &= V(g_1)V(g_2), \end{aligned} \tag{1.1}$$

where (f, g) denotes the scalar product in $\mathfrak{V}$. The unitarity of $U(f)$ and $V(g)$ implies $U(0) = V(0) = \mathbf{1}$, $U(f)^* = U(-f)$, and $V(g)^* = V(-g)$. It is further assumed that $U(\lambda f)$ and $V(\lambda g)$ are strongly continuous in λ for fixed f and g. We put $U(f, g) \equiv U(f)V(g)$.

Let $h_1, h_2, \cdots$ be a complete orthonormal system of $\mathfrak{V}$. Such a system will be called a basis of $\mathfrak{V}$. We denote by W_N the subspace $\{h_1, \cdots, h_N\}$ of all linear combinations of the first N h_i's, and by $\mathfrak{V}_0$ the subspace $\{h_1, h_2, \cdots\}$ of all finite linear combinations of the h_i's. If we deal with several bases simultaneously, they will be indexed by a superscript, such as h_1^β, $h_2^\beta, \cdots$. W_N^β and $\mathfrak{V}_0^\beta$ have corresponding meanings. We denote the Lebesgue measure in W_N induced by the scalar product (f, g) by $d^N f$. For any ψ_1, φ_1, φ_2, $\psi_2 \in \mathcal{H}$ we put

$$\begin{aligned} I_N(\psi_1, \varphi_1, \varphi_2, \psi_2) &\equiv \int_{W_N \times W_N} d^N f\, d^N g\, \frac{1}{(2\pi)^N} \\ &\quad \times \langle \psi_1, U(f,g)\varphi_1\rangle\langle U(f,g)\varphi_2, \psi_2\rangle. \end{aligned} \tag{1.2}$$

As an abbreviation we put

$$\begin{aligned} I_N(\varphi, \varphi_0, \varphi_0, \varphi) &\equiv I_N(\varphi, \varphi_0), \\ I_N(\varphi_0, \varphi_0, \varphi_0, \varphi_0) &\equiv I_N(\varphi_0). \end{aligned} \tag{1.3}$$

In the case of several bases, I_N^β is defined in an analogous way. In Part I a kernel integral formula was said to hold if, for some unit vector $\varphi_0 \in \mathcal{H}$ and all $\varphi_1 \psi \in \mathcal{H}$,

$$\lim_{N\to\infty} I_N(\varphi, \varphi_0, \varphi_0, \psi) = \langle \varphi, \psi\rangle \tag{1.4}$$

or

$$\overline{\lim_{N\to\infty}}\, I_N(\varphi, \varphi_0) = \|\varphi\|^2 \tag{1.5}$$

* Based on a part of the author's *Habilitationsschrift* "Aspekte der kanonischen Vertauschungsrelationen für Quantenfelder," University of Marburg (1968).

† Address from October 1969: Institut für Theoretische Physik, Göttingen.

[1] G. C. Hegerfeldt, J. Math. Phys. **10**, 1681 (1969), hereafter referred to as Part I.

or

$$\sup_{\beta} \overline{\lim_{N\to\infty}} I_N(\varphi, \varphi_0) = \|\varphi\|^2, \tag{1.6}$$

where the supremum is taken over all bases of $\mathfrak{V}$.

In Sec. 2 we are going to prove that any of these formulas implies the irreducibility of the representation $U(f, g)$. This important fact had already been conjectured by Klauder and McKenna.[2] In Sec. 3, some mathematical results are collected which are needed for the following.

A complete classification of all representations fulfilling the kernel integral formula with the limit superior in Eq. (1.5) is obtained in Sec. 4. We will meet the same phenomenon as in Part I, where Eq. (1.5) for a cyclic φ_0 implied the existence of a subsequence n_i such that $\lim_i I_{n_i}(\varphi, \varphi_0, \varphi_0, \psi) = \langle\varphi, \psi\rangle$ for all $\varphi, \psi \in \mathcal{H}$. Here it turns out that one can replace the first φ_0 by any vector φ_1 and the second φ_0 by φ_2, and one obtains as limit $\langle\varphi, \psi\rangle\langle\varphi_2, \varphi_1\rangle$.

In Sec. 5, necessary and sufficient conditions are derived for the kernel integral formula with the supremum over all bases to hold. This is the most general formula and imposes the weakest conditions on the representation. Our results again show a striking similarity to those of Part I. Under the assumption that not only $U(\mathfrak{V}, \mathfrak{V})$ but also $U(\mathfrak{V}_0^\beta, \mathfrak{V}_0^\beta)$ is irreducible, it is shown that the validity of Eq. (1.6) for some φ_0 and $\varphi = \varphi_0$ implies its validity for all φ and all unit vectors φ_0. As in Part I, it turns out that the supremum can be replaced by an ordinary limit over a sequence of bases independent of φ and φ_0. Then the existence of a kind of diagonal sequence $I_{n(\nu)}^{\beta(\nu)}$ is proved such that for any $\psi_1, \varphi_1, \varphi_2, \psi_2 \in \mathcal{H}$, the expression in Eq. (1.2) goes to $\langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle$.

Finally, we discuss our results in the last section. In view of the close similarity of the cyclic and irreducible case, we conjecture that the validity of a kernel integral formula implies that the representation is not only irreducible, but also cyclic with respect to the field. This would be a close analogy to the case of finitely many degrees of freedom.

2. NECESSITY OF IRREDUCIBILITY

In this section it will be shown that irreducibility of the representation is a necessary condition for a kernel integral formula to hold. For the ordinary limit in Eq. (1.4), this has been conjectured by Klauder and McKenna.[2] Up to now it has only been proved for finitely many degrees of freedom by a rather intricate construction.[3] Our derivation is independent of the number of degrees of freedom and fairly simple. It comprises the earlier result as a special case. As it turns out, it even suffices for irreducibility that $\sup_{\beta,n} I_n^\beta(\psi, \varphi_0) = \|\psi\|^2$ for all $\psi \in \mathcal{H}$. This is a weaker condition than Eq. (1.6) since $I_n^\beta(\psi, \varphi_0) \leq \|\psi\|^2$ for all ψ, β, n.

Theorem 2.1: Let $U(f, g)$ be a representation of the CCR with $f, g \in \mathfrak{V}$, and let one of the following conditions hold for all $\psi \in \mathcal{H}$ and some φ_0, $\|\varphi_0\| = 1$:

(a) $\lim_{n\to\infty} I_n^\beta(\psi, \varphi_0) = \|\psi\|^2$ for some basis β;

(b) $\overline{\lim_{n\to\infty}} I_n^\beta(\psi, \varphi_0) = \|\psi\|^2$ for some basis β;

(c) $\sup_\beta \overline{\lim_{n\to\infty}} I_n^\beta(\psi, \varphi_0) = \|\psi\|^2$;

(d) $\sup_{\beta,n} I_n^\beta(\psi, \varphi_0) = \sup_W I_W(\psi, \varphi_0) = \|\psi\|^2$;

where W runs through all finite-dimensional subspaces of $\mathfrak{V}$. *Then* the representation is irreducible. The same result holds for $\dim \mathfrak{V} < \infty$ if one puts $n = \dim \mathfrak{V}$ and omits the limits over n.

The proof is a simple consequence of the following lemma:

Lemma 2.1: Let $U(f, g)$ be a representation of the CCR with f, g. Let

$$\mathcal{H} = \bigoplus_{i=1}^{r} \mathcal{H}_i, \quad \dim \mathcal{H}_i > 0, \quad r > 1, \tag{2.1}$$

and let the $\mathcal{H}_i$ be invariant under $U(f, g)$ for all $f, g \in \mathfrak{V}$. Then for any unit vector $\varphi_0 \in \mathcal{H}$ there exist a $\psi \in \mathcal{H}$ and a $c \geq 0$ such that, for every finite-dimensional subspace W_N of $\mathfrak{V}$,

$$\int_{W_N \times W_N} d^N f\, d^N g\, \frac{1}{(2\pi)^N} |\langle\psi, U(f, g)\varphi_0\rangle|^2 \leq c < \|\psi\|^2, \tag{2.2}$$

where $N = \dim W_N$.

Proof: Let $\varphi_0 = \oplus_i \varphi_i$ be the decomposition of φ_0 with respect to the $\mathcal{H}_i$ in Eq. (2.1). Since $r > 1$, there exists a $\psi = \oplus_i \psi_i$ such that, for all numbers λ,

$$(\|\psi_1\|, \|\psi_2\|, \cdots) \neq \lambda(\|\varphi_1\|, \|\varphi_2\|, \cdots). \tag{2.3}$$

By the invariance of $\mathcal{H}_i$, the left-hand side of Eq. (2.2) becomes

$$\int_{W_N \times W_N} d^N f\, d^N g\, \frac{1}{(2\pi)^N} \left| \sum_{i=1}^{r} \langle\psi_i, U(f, g)\varphi_i\rangle \right|^2$$

$$= \sum_{i,j} \int d^N f\, d^N g\, \frac{1}{(2\pi)^N} \langle\psi_i, U(f, g)\varphi_i\rangle\langle U(f, g)\varphi_j, \psi_j\rangle;$$

[2] J. R. Klauder and J. McKenna, J. Math. Phys. **6**, 68 (1965).
[3] J. McKenna and J. R. Klauder, J. Math. Phys. **5**, 878 (1964).

and by Schwarz's inequality applied to the integral this is

$$\leq \sum_{i,j} \left(\int d^N f\, d^N g\, \frac{1}{(2\pi)^N} |\langle \psi_i, U(f,g)\varphi_i \rangle|^2 \right)^{\frac{1}{2}} \times \left(\int d^N f\, d^N g\, \frac{1}{(2\pi)^N} |\langle \psi_j, U(f,g)\varphi_j \rangle|^2 \right)^{\frac{1}{2}}.$$

Due to Eq. (2.31) of Part I, i.e., $|I_N(\varphi, \varphi_0)|^2 \leq \|\varphi\|^2 \times \|\varphi_0\|^2$, this is,

$$\leq \left(\sum_i \|\psi_i\| \cdot \|\varphi_i\| \right)^2 \equiv c.$$

This number c is independent of W_N. In view of Eq. (2.3), Schwarz's inequality implies

$$c < \left(\sum_i \|\psi_i\|^2 \right) \left(\sum_i \|\varphi_i\|^2 \right) = \|\psi\|^2. \quad \text{Q.E.D.}$$

Proof of Theorem: Assume the representation to be reducible. Due to the unitarity of $U(f,g)$ one can write $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ with nontrivial invariant $\mathcal{H}_1$ and $\mathcal{H}_2$. Then by the above lemma none of the conditions (a)–(d) can hold for all $\psi \in \mathcal{H}$. Q.E.D.

For the case of the ordinary limit and for the limit superior, one immediately obtains as a corollary the following sharpened irreducibility condition:

Corollary 2.1: Let $h_1, h_2, \cdots$ be a basis of $\mathfrak{U}$ for which either condition (a) or (b) of the above theorem holds. Then already $U(\mathfrak{U}_0, \mathfrak{U}_0)$ is irreducible where $\mathfrak{U}_0 = \{h_1, h_2, \cdots\}$.

Proof: Replace $\mathfrak{U}$ by $\mathfrak{U}_0$ in the above theorem. Since $h_1, h_2, \cdots$ is also a basis of $\mathfrak{U}_0$, the statement follows from the theorem.

3. SOME AUXILIARY RESULTS

In this section some results are collected which will be needed later on.

Let $\mathcal{H}$ be the tensor product of two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{3.1}$$

For any vector $\psi \in \mathcal{H}$ one can choose sets of orthonormal vectors $\psi_i^{(1)}$ and $\psi_i^{(2)}$ in $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, $i = 1, \cdots$, such that

$$\psi = \sum_i \lambda_i \psi_i^{(1)} \otimes \psi_i^{(2)}, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq 0. \tag{3.2}$$

These sets can of course be completed to bases of $\mathcal{H}_1$ and $\mathcal{H}_2$. This decomposition is called a *standard diagonal expansion*[4] of ψ with respect to $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Let A be some index set, and let R_α, $\alpha \in A$, be von Neumann algebras of bounded linear operators in a Hilbert space $\mathcal{H}$. Denote by $B(\mathcal{H})$ the set of all bounded linear operators in $\mathcal{H}$ and by

$$R_\alpha \vee R_\beta \equiv \{R_\alpha \cup R_\beta\}'', \tag{3.3}$$

the von Neumann algebra generated by R_α and R_β. Here $\cup$ means set-theoretical union and $''$ double commutant. R_α is a factor if

$$R_\alpha \cap R_\alpha' = \{c \cdot \mathbf{1}\} \tag{3.4}$$

or, equivalently,

$$R_\alpha \vee R_\alpha' = B(\mathcal{H}). \tag{3.5}$$

Definition[5] *3.1:* $(R_\alpha, \alpha \in A)$ is called *factorization* of $B(\mathcal{H})$ if each R_α, $\alpha \in A$, is a factor which commutes with any other R_β and if the von Neumann algebra generated by the R_α equals $B(\mathcal{H})$:

$$\begin{aligned} &R_\alpha \cap R_\alpha' = \{c \cdot \mathbf{1}\}, \\ &R_\alpha \subset R_\beta', \quad \alpha \neq \beta; \\ &\bigvee_{\alpha \in A} R_\alpha = \left\{ \bigcup_{\alpha \in A} R_\alpha \right\}'' = B(\mathcal{H}). \end{aligned} \tag{3.6}$$

If all factors are of type I, $(R_\alpha, \alpha \in A)$ is called a type-I factorization. Murray and von Neumann[6] have shown that any finite type-I factorization $R_1, \cdots, R_n$ is a tensor product factorization, i.e.,

$$\begin{aligned} &\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n, \\ &R_i = B(\mathcal{H}_i) \otimes \left(\bigotimes_{i \neq j} \mathbf{1}_j \right), \quad i = 1, \cdots, n. \end{aligned} \tag{3.7}$$

They have further shown that if (R_1, R_2) is a factorization of $B(\mathcal{H})$ and R_1 is of type I, then also R_2 is of type I.

To apply this to the CCR, consider a representation $U(f,g)$ with $f, g \in \mathfrak{U}$. Let W be a finite-dimensional subspace of $\mathfrak{U}$, dim $W = N$ say. Since the restriction of the representation to $f, g \in W$ just yields a representation of the CCR for N degrees of freedom, one can, according to von Neumann's theorem,[7] write $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ in such a way that for $f, g \in W$

$$U(f,g) = U_1(f,g) \otimes \mathbf{1}_2, \tag{3.8}$$

where $U_1(f,g)$ is the Schrödinger representation for N degrees of freedom in $\mathcal{H}_1$. If one puts

$$R(W) \equiv \{U(f,g); f, g \in W\}'', \tag{3.9}$$

then, by Eq. (3.8), $R(W)$ is a type-I factor.

Now let $h_1, h_2, \cdots$ be a basis of $\mathfrak{U}$, and let $\mathfrak{U}_0 = \{h_1, h_2, \cdots\}$ be defined as in the Introduction. Let

[4] H. Araki and J. Woods, "Complete Boolean Algebras of Type I Factors," University of Maryland Report No. 563 (1966); Publ. Res. Inst. Math. Sci., Ser. A, Kyoto University 2, 157 (1967).

[5] Cf. M. A. Neumark, *Normierte Algebren* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1959), p. 456.

[6] F. J. Murray and J. von Neumann, Ann. Math. 37, 116 (1936).

[7] J. von Neumann, Ann. Math. 32, 191 (1931).

$n_1 < n_2 < \cdots$ be natural numbers; denote by I_{r+1} the set $I_{r+1} \equiv (n_{r+1}, \cdots, n_{r+1})$, and define

$$R_i \equiv \{U(f, g); f, g \in \{h_\alpha; \alpha \in I_i\}\}''. \quad (3.10)$$

Then, according to Eq. (1.1),

$$R_i \subset R_j', \quad i \neq j,$$

$$\bigvee_{i=1}^{\infty} R_i = \left\{\bigcup_1^\infty R_i\right\}'' = \{U(f, g); f, g \in \mathfrak{U}_0\}''. \quad (3.11)$$

Hence, if $U(\mathfrak{U}_0, \mathfrak{V}_0)$ is irreducible, one has

$$\bigvee_i R_i = B(\mathcal{H}) \quad (3.12)$$

and $(R_i, i = 1, 2, \cdots)$ is a type-I factorization. This fact was noted in Ref. 4.

We close this section by proving the following useful lemma which is contained in Ref. 8:

Lemma 3.1: Let $W_1, W_2, \cdots$ be a sequence of finite-dimensional subspaces of $\mathfrak{U}$, dim $W_n = N(n)$ say, and let $U(f, g)$ be a representation of the CCR in $\mathcal{H}$ with $f, g \in \mathfrak{U}$. For each n, let $\mathcal{H}^{(n)}$ be a subspace of $\mathcal{H}$ irreducible under $U(W_n, W_n)$, and denote by P_n the projection operator onto $\mathcal{H}^{(n)}$. Define $I_{N(n)}(\psi_1, \varphi_1, \varphi_2, \psi_2)$ as in Eq. (1.2), the integration now being over $W_n \times W_n$. *Then*, if P_n converges strongly to **1**,

$$\lim_{n\to\infty} I_{N(n)}(\psi_1, \varphi_1, \varphi_2, \psi_2) = \langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle \quad (3.13)$$

for all $\psi_1, \varphi_1, \varphi_2, \psi_2 \in \mathcal{H}$.

Proof: For any $\psi_1^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)}, \psi_2^{(n)} \in \mathcal{H}^{(n)}$, one has

$$I_{N(n)}(\psi_1^{(n)}, \varphi_1^{(n)}, \varphi_2^{(n)}, \psi_2^{(n)}) = \langle\psi_1^{(n)}, \psi_2^{(n)}\rangle\langle\varphi_2^{(n)}, \varphi_1^{(n)}\rangle \quad (3.14)$$

by Lemma 2.1 of Part I. Let $1 \geq \epsilon > 0$, and put $\epsilon' = \epsilon/19$. Then there exists an index n_0 such that $\|\psi_i - P_n\psi_i\| < \epsilon'$, $\|\varphi_i - P_n\varphi_i\| < \epsilon'$ for $n \geq n_0$, $i = 1, 2$. One can assume $\|\psi_i\| = \|\varphi_i\| = 1$, $i = 1, 2$. Then, by Eq. (3.14),

$$\begin{aligned} D_n &\equiv |I_{N(n)}(\psi_1, \varphi_1, \varphi_2, \psi_2) - \langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle| \\ &= |I_{N(n)}(\psi_1, \varphi_1, \varphi_2, \psi_2) \\ &\quad - I_{N(n)}((P_n\psi_1 - \psi_1) + \psi_1, \cdots, \\ &\qquad (P_n\psi_2 - \psi_2) + \psi_2) + \langle P_n\psi_1, P_n\psi_2\rangle \\ &\quad \times \langle P_n\varphi_2, P_n\varphi_1\rangle - \langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle|. \end{aligned} \quad (3.15)$$

Using the linearity of $I_{N(n)}$ in the indicated way, one obtains for the first difference on the right-hand side 15 terms, each of which contains at least once an argument of the form $P_n\varphi - \varphi$, thus being smaller than ϵ' for $n \geq n_0$ by Eq. (2.31) of Part I. The second difference is smaller than $4\epsilon'$. Hence $D_n < 19\epsilon' = \epsilon$ for $n \geq n_0$. Q.E.D.

4. THE KERNEL INTEGRAL AS LIMIT SUPERIOR: COMPLETE SOLUTION

Before classifying all representations fulfilling a kernel integral formula with $\overline{\lim}$ as in Eq. (1.5), we briefly quote a result by Araki and Woods for kernel integrals as an ordinary limit. They make the general assumption that $U(\mathfrak{U}_0, \mathfrak{V}_0)$ is irreducible. In view of our results of Sec. 2, we can write the result of Araki and Woods[4] in the following somewhat sharpened form:

Theorem 4.1: Let $U(f, g)$ be a representation of the CCR in $\mathcal{H}$ with $f, g \in \mathfrak{U}$. Let $h_1, h_2, \cdots$, be a basis of $\mathfrak{U}$, and define I_n as in Eq. (1.2). Then the following conditions are equivalent:

(a) $U(\mathfrak{U}_0, \mathfrak{V}_0)$ is irreducible, and there exists a unit vector $\varphi_0 \in \mathcal{H}$ such that $\lim_{n\to\infty} I_n(\varphi_0) = 1$;

(b) for any $\psi_1, \varphi_1, \varphi_2, \psi_2 \in \mathcal{H}$,

$$\lim_{n\to\infty} I_n(\psi_1, \varphi_1, \varphi_2, \psi_2) = \langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle.$$

Araki and Woods have further shown that $U(f, g)$, when restricted to $f, g \in \mathfrak{U}_0$, is a partial tensor-product representation (PTPR) if (a) or (b) is fulfilled and that it is a direct- or tensor-product representation (TPR) if the correct limits are approached for any ordering of the basis vectors $h_1, h_2, \cdots$. According to the remark on corollary 5.1 of Part I, the converse also holds. Thus we can write the result of Araki and Woods[4] in the following form:

Theorem 4.2: Let $U(f, g)$ with $f, g \in \mathfrak{U}$ be a representation of the CCR, and let $h_1, h_2, \cdots$ be a basis of $\mathfrak{U}$. Then $U(\mathfrak{U}_0, \mathfrak{V}_0)$ is a tensor-product representation with respect to this basis if and only if conditions (a) or (b) of the preceding theorem hold for any ordering of the basis vectors h_i.

This result provides an alternative classification of TPR of the CCR. Now it will be shown that the kernel integral with the limit superior is linked to PTPR. Simultaneously it will turn out that the limit superior can be replaced by an ordinary limit over a subsequence which is independent of φ. This will allow a transition from norms to scalar products. The precise formulation runs as follows:

Theorem 4.3: Let $U(f, g)$ with $f, g \in \mathfrak{U}$ be a representation of the CCR in $\mathcal{H}$, and let $h_1, h_2, \cdots$ be a

[8] J. R. Klauder, J. McKenna, and E. J. Woods, J. Math. Phys. 7, 822 (1966).

basis of $\mathfrak{U}$. Then the following conditions are equivalent:

(a) $U(\mathfrak{U}_0, \mathfrak{U}_0)$ is irreducible, and there exists a unit vector $\varphi_0 \in \mathcal{H}$ such that

$$\overline{\lim_{n\to\infty}} I_n(\varphi_0) = 1; \tag{4.1}$$

(b) There exists a sequence of natural numbers $n_1 < n_2 < \cdots$ such that for any $\psi_1, \varphi_1, \varphi_2, \psi_2 \in \mathcal{H}$,

$$\lim_{i\to\infty} I_{n_i}(\psi_1, \varphi_1, \varphi_2, \psi_2) = \langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle; \tag{4.2}$$

(c) $U(\mathfrak{U}_0, \mathfrak{U}_0)$ is a PTPR with respect to the basis $h_1, h_2, \cdots$.

Proof: We show (a) → (c) → (b) → (a).

Ad (a) → (c): The restriction of $U(f, g)$ to $f, g \in W_n = \{h_1, \cdots, h_n\}$ is a representation of the CCR for n degrees of freedom. According to Eq. (3.8), one can decompose $\mathcal{H}$ as

$$\mathcal{H} = \mathcal{H}_{(n)} \otimes \mathcal{H}'_{(n)} \tag{4.3}$$

in such a way that, for $f, g \in W_n$,

$$U(f, g) = U_n(f, g) \otimes \mathbf{1}'_{(n)}, \tag{4.4}$$

where $U_n(f, g)$ is the Schrödinger representation for n degrees of freedom in $\mathcal{H}_{(n)}$. In the standard diagonal expansion [cf. Eq. (3.2)] of φ_0 with respect to $\mathcal{H}_{(n)} \otimes \mathcal{H}'_{(a)}$,

$$\varphi_0 = \sum_i \lambda_i^{(n)} \psi_i^{(n)} \otimes \psi_i'^{(n)}, \quad \lambda_1^{(n)} \geq \lambda_2^{(n)} \geq \cdots \geq 0, \tag{4.5}$$

one has, due to orthonormality,

$$\sum_i \lambda_i^{(n)2} = 1. \tag{4.6}$$

Inserting Eq. (4.5) into $I_n(\varphi_0)$ and noting that [due to Part I, Eq. (2.31)] I_n is not only linear but also continuous in each argument for fixed n, one can extract the (possibly infinite) sums; the result is

$$\begin{aligned} I_n(\varphi_0) = \sum_{i,j,k,l} \lambda_i^{(n)}\lambda_j^{(n)}\lambda_k^{(n)}\lambda_l^{(n)} \langle\psi_i'^{(n)}, \psi_j'^{(n)}\rangle\langle\psi_k^{(n)}, \psi_l'^{(n)}\rangle \\ \times \int_{W_n\times W_n} d^n f\, d^n g\, \frac{1}{(2\pi)^n} \langle\psi_i^{(n)}, U_n(f, g)\psi_j^{(n)}\rangle \\ \times \langle U_n(f, g)\psi_k^{(n)}, \psi_l^{(n)}\rangle. \end{aligned} \tag{4.7}$$

Using Eq. (2.30) of Part I for the last integral, one gets

$$I_n(\varphi_0) = \sum_i \lambda_i^{(n)4}. \tag{4.8}$$

Equations (4.5) and (4.6) imply

$$I_n(\varphi_0) \leq \lambda_1^{(n)2} \sum_i \lambda_i^{(n)2} = \lambda_1^{(n)2}. \tag{4.9}$$

Since $\lambda_1^{(n)} \leq 1$, Eq. (4.1) yields

$$\overline{\lim_n} \lambda_1^{(n)} = 1. \tag{4.10}$$

Hence there exists a subsequence $\lambda_1^{(n_i)}$ converging to 1. The convergence can be assumed to be so fast that

$$\sum_i (1 - \lambda_1^{(n_i)2}) < \infty. \tag{4.11}$$

Define the factors R_i for these n_i as in Eq. (3.10). Then $(R_i, i = 1, 2, \cdots)$ is a type-I factorization because $U(\mathfrak{U}_0, \mathfrak{U}_0)$ is irreducible by assumption. Now lemmas 2.2 and 4.9 of Ref. 4 are directly applicable. Accordingly, Eq. (4.11) implies that $U(f, g)$ with $f, g \in \mathfrak{U}_0$ is a PTPR with respect to the basis $h_1, h_2, \cdots$ and basis subdivision $(1, \cdots, n_1), (n_1 + 1, \cdots, n_2), \cdots$.

Ad (c) → (b): We use the same notation as in Part I, Sec. 5. According to Lemma 5.1 of that section, one can assume the reference vector ψ_0 of the partial tensor product to be cyclic with respect to $U(\mathfrak{U}_0)$. According to Eq. (5.14) of Part I,

$$\mathcal{H}_{n_i} \equiv \overline{\{U(W_{n_i})\psi_0\}}.$$

is invariant under $V(g)$ for $g \in W_{n_i}$. Hence one has in $\mathcal{H}_{n_i}$ a representation for n_i degrees of freedom which is cyclic with respect to $U(f)$ and thus equivalent to the Schrödinger representation. Moreover, since P_{n_i}, the projection operator onto $\mathcal{H}_{n_i}$, converges strongly to $\mathbf{1}$ by Lemma 4.1 of Part I, (b) is an immediate consequence of Lemma 3.1 above.

Ad (b) → (a): Equation (4.1) follows from

$$|I_n(\varphi, \psi)| \leq \|\varphi\|^2 \|\psi\|^2.$$

The same inequality also implies that condition (b) of Theorem 2.1 is fulfilled. Hence $U(\mathfrak{U}_0, \mathfrak{U}_0)$ is irreducible by Corollary 2.1. Q.E.D.

The above theorem provides a complete classification of all representations for which a kernel integral formula with a limit superior holds because Eq. (4.2) implies $\overline{\lim} I_n(\varphi, \psi) = \|\varphi\|^2 \|\psi\|^2$ for all $\varphi, \psi \in \mathcal{H}$, and this in turn implies Eq. (4.1) and the irreducibility of $U(\mathfrak{U}_0, \mathfrak{U}_0)$. The first part of the above theorem is strikingly similar to Theorem 6.1 of Part I. There it has been shown that if φ_0 is cyclic for $U(\mathfrak{U}_0)$ and $\overline{\lim} I_n(\varphi_0) = 1$, then $\lim I_{n_i}(\varphi, \varphi_0, \varphi_0, \psi) = \langle\varphi, \psi\rangle$ for some fixed subsequence and all $\varphi, \psi \in \mathcal{H}$. Since this implies the irreducibility by Theorem 2.1, one has the following corollary:

Corollary 4.1: Condition (a) in Theorem 4.3 can be replaced by the following:

(a′) There exists a unit vector φ_0 which is cyclic with

respect to $U(\mathfrak{V}_0)$ and which fulfills

$$\overline{\lim_n} I_n(\varphi_0) = 1.$$

Similar to Theorem 6.1 of Part I, the convergence of $I_{n_i}(\varphi_0)$ towards 1 for some subsequence n_i implies the convergence of $I_{n_i}(\psi_1, \varphi_1, \varphi_2, \psi_2)$ towards $\langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle$ if $U(\mathfrak{V}_0, \mathfrak{V}_0)$ is irreducible. This will be shown in a remark following Corollary 5.1.

5. THE BASIS-INDEPENDENT KERNEL INTEGRAL

A definition of the kernel integral as in Eq. (1.6) by a supremum over all bases has the advantage that one deals with basis-independent expressions and that the restrictions on the representation are weaker than in the other two cases. It has, however, the disadvantage that scalar products in $\mathcal{H}$ are not so easily expressible as in the case of an ordinary limit. In this section we therefore pursue a twofold purpose. First we show that under a somewhat stronger irreducibility condition the validity of the basis-independent kernel integral formula for a nonzero vector implies its validity for all vectors of $\mathcal{H}$ and that the supremum can be replaced by a limit over a fixed sequence of bases. Then it will be shown that one can replace this $\lim\overline{\lim}$ by a simple limit which again allows the transition from norms to scalar products.

If $U(f, g), f, g \in \mathfrak{V}$, is an irreducible representation of the CCR and if $\mathfrak{V}_0$ is some linear subspace of $\mathfrak{V}$ which is dense in $\mathfrak{V}$ with respect to the scalar product in $\mathfrak{V}$, then it is probably not true in general that the representation remains irreducible when restricted to $f, g \in \mathfrak{V}_0$. Of course, with appropriate continuity conditions this will follow immediately. If $\mathfrak{V}_0$ consists of all finite linear combinations of a countable set of elements, then the irreducibility and also the cyclicity of $U(\mathfrak{V}_0, \mathfrak{V}_0)$ imply the separability of the Hilbert space $\mathcal{H}$—a fact which is quite welcome for physical applications.

Lemma 5.1: Let $h_1, h_2, \cdots$ be a basis of $\mathfrak{V}$, put $\mathfrak{V}_0 = \{h_1, h_2, \cdots\}$, and let $U(\mathfrak{V}_0, \mathfrak{V}_0)$ be irreducible or, somewhat weaker, let $U(\mathfrak{V}_0, \mathfrak{V}_0)$ be cyclic vector φ_0. Then $\mathcal{H}$ is separable.

Proof: Put $W_n = \{h_1, h_2, \cdots, h_n\}$ and

$$\mathcal{H}_n = \overline{\{U(f, g)\varphi_0; f, g \in W_n\}}.$$

Then, by assumption,

$$\mathcal{H} = \overline{\bigcup_n \mathcal{H}_n}.$$

For $f, g \in W_n$, the representation is continuous with respect to the Euclidean topology of W_n. In W_n the points with rational coordinates are dense and countable. The linear combinations with rational coefficients of the corresponding $U(f, g)\varphi_0$ are dense in $\mathcal{H}_n$. Hence all $\mathcal{H}_n$ and then also $\mathcal{H}$ are separable. Q.E.D.

It is also an open question if, conversely, the separability of $\mathcal{H}$ and irreducibility of $U(\mathfrak{V}, \mathfrak{V})$ imply the irreducibility of $U(\mathfrak{V}_0, \mathfrak{V}_0)$. Therefore, whenever needed, we will have to assume the irreducibility of $U(\mathfrak{V}_0, \mathfrak{V}_0)$. Since, by Theorem 2.1, one knows that at least $U(\mathfrak{V}, \mathfrak{V})$ is irreducible, this assumption seems not very strong. Now we prove the first of the two results mentioned above, where W_n^β, $\mathfrak{V}_0^\beta$, and I_n^β have the same meaning as in the Introduction.

Theorem 5.1: Let $U(f, g), f, g \in \mathfrak{V}$, be a representation of the CCR, and let $U(\mathfrak{V}_0^\beta, \mathfrak{V}_0^\beta)$ be irreducible for any basis. Then the following conditions are equivalent:

(a) There exists a unit vector $\varphi_0 \in \mathcal{H}$ satisfying

$$\sup_\beta \overline{\lim_n} I_n^\beta(\varphi_0) = 1; \tag{5.1}$$

(b) for any $\psi, \varphi \in \mathcal{H}$

$$\sup_\beta \overline{\lim_n} I_n^\beta(\psi, \varphi) = \|\psi\|^2 \|\varphi\|^2; \tag{5.2}$$

(c) there exists a sequence of bases $(h_1^{(\nu)}, h_2^{(\nu)}, \cdots)$, $\nu = 1, 2, \cdots$, such that for any $\psi, \varphi \in \mathcal{H}$

$$\lim_{\nu\to\infty} \overline{\lim_n} I_n^{(\nu)}(\psi, \varphi) = \|\psi\|^2 \|\varphi\|^2. \tag{5.3}$$

The proof is based on the following lemma[9]:

Lemma 5.2: Under the assumptions of the above theorem condition, (a) implies that there exists a sequence of bases $(h_1^{(\nu)}, h_2^{(\nu)}, \cdots)$, $\nu = 1, 2, \cdots$, and that for each ν there are natural numbers $n_i = n_i(\nu)$, $n_1 < n_2 < \cdots$, and projection operators $P_{(\nu, n_i)}$ onto subspaces $\mathcal{H}_{(\nu, n_i)}$ of $\mathcal{H}$ which are irreducible under $U(W_{n_i}^{(\nu)}, W_{n_i}^{(\nu)})$ such that for all $\varphi \in \mathcal{H}$

$$\lim_{\nu\to\infty} \lim_{i\to\infty} \|\varphi - P_{(\nu, n_i)}\varphi\| = 0. \tag{5.4}$$

Proof: If (a) holds, there exists a sequence of bases $(h_1^{(\nu)}, h_2^{(\nu)}, \cdots)$, $\nu = 1, 2, \cdots$, such that

$$\lim_\nu \overline{\lim_n} I_n^{(\nu)}(\varphi_0) = 1. \tag{5.5}$$

For each ν and n one can, as in Eq. (4.3), decompose $\mathcal{H}$ as

$$\mathcal{H} = \mathcal{H}_{(n)}^{(\nu)} \otimes \mathcal{H}_{(n)}^{\prime(\nu)}, \tag{5.6}$$

[9] This is a generalization of Lemma 4.10 of Ref. 4.

such that for $f, g \in W_n^{(\nu)}$

$$U(f, g) = U_n^{(\nu)}(f, g) \otimes \mathbf{1}_{(n)}^{\prime(\nu)}, \tag{5.7}$$

where $U_n^{(\nu)}(f, g)$ is irreducible in $\mathcal{H}_{(n)}^{(\nu)}$. Let the standard diagonal expansion of φ_0 with respect to the decomposition in Eq. (5.6) be

$$\varphi_0 = \sum_i \lambda_i^{(\nu,n)} \psi_i^{(\nu,n)} \otimes \psi_i^{\prime(\nu,n)},$$
$$\lambda_1^{(\nu,n)} \geq \lambda_2^{(\nu,n)} \geq \cdots \geq 0. \tag{5.8}$$

By Eq. (4.9) one then has

$$I_n^{(\nu)}(\varphi_0) \leq \lambda_1^{(\nu,n)^2} \leq 1. \tag{5.9}$$

Define the projection operator $P_{(\nu,n)}$ and the subspace $\mathcal{H}_{(\nu,n)}$ by

$$P_{(\nu,n)} \equiv \mathbf{1}_{(n)}^{(\nu)} \otimes |\psi_1^{\prime(\nu,n)}\rangle\langle\psi_1^{\prime(\nu,n)}|,$$
$$\mathcal{H}_{(\nu,n)} \equiv P_{(\nu,n)}\mathcal{H}. \tag{5.10}$$

According to Eq. (5.7), $U(f, g)$ acts irreducibly in $\mathcal{H}_{(\nu,n)}$ for $f, g \in W_n^{(\nu)}$. Now we put for each ν

$$\overline{\lim_n}\, \lambda_1^{(\nu,n)^2} \equiv 1 - \epsilon_\nu. \tag{5.11}$$

Then, by Eqs. (5.9) and (5.5), $\epsilon_\nu \geq 0$ and

$$\lim_{\nu\to\infty} \epsilon_\nu = 0. \tag{5.12}$$

For each ν, there exists a subsequence $m_i = m_i(\nu)$, $i = 1, 2, \cdots$, such that $\lambda_1^{(\nu,m_i)}$ tends to $1 - \epsilon_\nu$.

The set $\{A; \|A\| \leq 1\}$ of operators in $\mathfrak{S}$ is weakly compact.[10] Therefore, due to the separability of $\mathcal{H}$, there is for each ν a subsequence $P_{(\nu,n_i)}$ of the $P_{(\nu,m_i)}$ which converges weakly to an operator T_ν. We show that $T_\nu = (1 - \epsilon_\nu)\mathbf{1}$. Indeed, let $f, g \in W_n^{(\nu)}$, and let $n_i \geq n$. Then $U(f, g)$ commutes with $P_{(\nu,n_i)}$, hence also with its weak limit T_ν. Thus $U(f, g)$ commutes with T_ν for all $f, g \in \mathfrak{U}_0^{(\nu)}$. Due to the irreducibility of $U(\mathfrak{U}_0^{(\nu)}, \mathfrak{U}_0^{(\nu)})$, T_ν is a multiple of the unit operator, $T_\nu = \alpha_\nu \mathbf{1}$. Since

$$\langle\varphi_0, P_{(\nu,n)}\varphi_0\rangle = 1 - \lambda_1^{(\nu,n)^2},$$

it follows that $\alpha_\nu \doteq 1 - \epsilon_\nu$, and Eq. (5.4) is a consequence of Eq. (5.12). Q.E.D.

Proof of Theorem: Due to the lemma, it suffices to show Eq. (5.4) → (c) → (b) → (a).

Ad Eq. (5.4) → (c): One can assume $\|\varphi\| = \|\psi\| = 1$. Let $1 \geq \epsilon > 0$, and put $\epsilon' \equiv \epsilon/19$. By Eq. (5.4) there exists an index $\nu_0 = \nu_0(\epsilon, \varphi, \psi)$ such that for each $\nu \geq \nu_0$ there is an index $i_0 = i_0(\epsilon, \nu, \varphi, \psi)$ with the property that for $\nu \geq \nu_0$, $i \geq i_0$,

$$\|\psi - P_{(\nu,n_i)}\psi\| \leq \epsilon', \quad \|\varphi - P_{(\nu,n_i)}\varphi\| \leq \epsilon'.$$

Then, as in Eq. (3.15), for $\nu \geq \nu_0$, $i \geq i_0$,

$$|I_{n_i}^{(\nu)}(\psi, \varphi) - I_{n_i}^{(\nu)}(P_{(\nu,n_i)}\psi, P_{(\nu,n_i)}\varphi)| \leq 15\epsilon'. \tag{5.13}$$

Since, for $f, g \in W_{n_i}^{(\nu)}$, $P_{(\nu,n_i)}\mathcal{H}$ is irreducible under $U(f, g)$, the second term becomes $\|P_{(\nu,n_i)}\psi\|^2 \times \|P_{(\nu,n_i)}\varphi\|^2$, according to the kernel integral for n_i degrees of freedom. This product differs from $\|\psi\|^2\,\|\varphi\|^2 = 1$ by at most $4\epsilon'$. Hence for $\nu \geq \nu_0$, $i \geq i_0(\nu)$,

$$1 \geq I_{n_i}^{(\nu)}(\psi, \varphi) \geq 1 - 19\epsilon' = 1 - \epsilon, \tag{5.14}$$

and thus for $\nu \geq \nu_0$

$$1 \geq \overline{\lim_n}\, I_n^{(\nu)}(\psi, \varphi) \geq 1 - \epsilon.$$

This is just Eq. (5.3) for $\|\varphi\| = \|\psi\| = 1$.

Ad (c) → (b) → (a): The first part follows from $|I_n^\beta(\psi, \varphi)| \leq \|\psi\|^2\,\|\varphi\|^2$, and the second part is trivial. Q.E.D.

We note that the irreducibility condition entered in Lemma 5.2 only. Therefore it can be weakened slightly. It suffices that $U(\mathfrak{U}_0^{(\nu)}, \mathfrak{U}_0^{(\nu)})$ is irreducible for $\nu = 1, 2, \cdots$, where the $\mathfrak{U}_0^{(\nu)}$ belong to the bases $(h_1^{(\nu)}, h_2^{(\nu)}, \cdots)$ entering Eq. (5.5). The proof contains the following corollary:

Corollary 5.1: Let $(h_1, h_2^{(\nu)}, \cdots)$ be a sequence of bases, $\nu = 1, 2, \cdots$, and let each $U(\mathfrak{U}_0^{(\nu)}, \mathfrak{U}_0^{(\nu)})$ be irreducible. If there exists a unit vector φ_0 satisfying

$$\lim_{\nu\to\infty} \overline{\lim_{n\to\infty}}\, I_n^{(\nu)}(\varphi_0) = 1, \tag{5.15}$$

then for all $\varphi, \psi \in \mathcal{H}$

$$\lim_{\nu\to\infty} \overline{\lim_{n\to\infty}}\, I_n^{(\nu)}(\psi, \varphi) = \|\psi\|^2\,\|\varphi\|^2. \tag{5.16}$$

With similar methods one can show the corresponding corollary to Theorem 4.3. If $U(\mathfrak{U}_0, \mathfrak{U}_0)$ is irreducible and if for some unit vector φ_0

$$\lim_{i\to\infty} I_{n_i}(\varphi_0) = 1, \tag{5.17}$$

then for all $\psi_1, \varphi_1, \varphi_2, \psi_2 \in \mathcal{H}$

$$\lim I_{n_i}(\psi_1, \varphi_1, \varphi_2, \psi_2) = \langle\psi_1, \psi_2\rangle\langle\varphi_2, \varphi_1\rangle. \tag{5.18}$$

To prove this it suffices to show, by Lemma 3.1, that the projection operators P_{n_i} defined in the same way as in Eq. (5.10) via Eq. (4.5) converge strongly to $\mathbf{1}$. Indeed, since $\|\varphi_0 - P_{n_i}\varphi_0\|^2 = 1 - \lambda_1^{(n_i)^2}$, one has $\|\varphi_0 - P_{n_i}\varphi_0\| \to 0$. Now let $f, g \in W_n$ with $n \leq n_i$. Then P_{n_i} commutes with $U(f, g)$, hence also with all elements of $R(W_n) = \{U(f, g); f, g \in W_n\}''$. If $A \in R(W_n)$ and $\varphi = A\varphi_0$, then $\|\varphi - P_{n_i}\varphi\| \leq \|A\| \cdot \|\varphi_0 - P_{n_i}\varphi_0\| \to 0$. Since the set of these vectors,

[10] J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthiers Villars, Paris, 1957), p. 34.

$n = 1, 2, \cdots$, is dense in $\mathcal{H}$, one has proved that $P_{n_i} \to \mathbf{1}$ strongly. The same reasoning has been used by Araki and Woods[4] to prove (a) → (b) in Theorem 4.1.

Similar to Theorem 6.2 of Part I, it may again not be possible to replace the limit superior in Eq. (5.3) by an ordinary limit over a subsequence which is independent of ψ, φ since one does not know if the limit over i of $I^{(\nu)}_{n_i}(\psi, \varphi)$ in Eq. (5.14) exists. However, similar to Theorem 6.3 of Part I, one can find a kind of diagonal sequence of finite-dimensional subspaces of $\mathfrak{U}$ such that sup $\overline{\lim}$ can be replaced by an ordinary limit which yields the desired value also for scalar products. These subspaces need not be contained in each other.

Theorem 6.2: Let the assumptions be as in the preceding theorem, and let any of the conditions (a), (b), or (c) hold. Then there exists a sequence of finite-dimensional subspaces W_n of $\mathfrak{U}$, dim $W_n = N(n)$ say, such that for all $\psi_1, \varphi_1, \varphi_2, \psi_2 \in \mathcal{H}$

$$\lim_{n\to\infty} \int_{W_n \times W_n} d^N f\, d^N g\, \frac{1}{(2\pi)^N} \times \langle \psi_1, U(f,g)\varphi_1\rangle\langle U(f,g)\varphi_2, \psi_2\rangle = \langle \psi_1, \psi_2\rangle\langle \varphi_2, \varphi_1\rangle. \quad (5.19)$$

Proof: In any strong neighborhood of $\mathbf{1}$, there lies one of the projection operators $P_{(\nu,n_i)}$ of Lemma 5.2, so that $\mathbf{1}$ is a neighboring point of this set of operators. Thus, in view of the separability of $\mathcal{H}$, there exists a subsequence of these operators which converges strongly to the unit operator.[11] Denote the operators of the subsequence by P_n and the associated spaces of the $W^{(\nu)}_{n_i}$ by W_n. Since $P_n\mathcal{H}$ is irreducible under $U(f, g)$ for $f, g \in W_n$, Lemma 3.1 applies. Q.E.D.

It is not clear if, conversely, Eq. (5.19) implies condition (a)—although one would expect this. The difficulty is the following: If $h^\beta_1, h^\beta_2, \cdots$ is a basis of $\mathfrak{U}$, then $I^\beta_n(\varphi_0)$ may alternately increase and decrease for increasing n. The above theorem only states that there exists a sequence of bases and for each of these an $N(\nu)$ such that $I^{(\nu)}_{N(\nu)}(\varphi_0) \to 1$ for $\nu \to \infty$. This means that for appropriate ν, N one has $I^{(\nu)}_N(\varphi_0) \geq 1 - \epsilon$. But $I^{(\nu)}_{N+1}(\varphi_0)$ may already be smaller, and the same may hold for the limit superior over N. In this way one can only conclude that Eq. (5.19) implies

$$\sup_{\beta,n} I^\beta_n(\psi, \varphi) = \|\psi\|^2 \|\varphi\|^2 \quad (5.20)$$

for all $\psi, \varphi \in \mathcal{H}$. This in turn implies the irreducibility of the representation. Choosing any kernel

$$K(f', g'; f'', g'') = \langle U(f', g')\varphi_0, U(f'', g'')\varphi_0\rangle$$

with some unit vector φ_0, one obtains an integral formula for the reproducing kernel if one puts $\varphi_1 = \varphi_2 = \varphi_0$ and $\psi_1 = U(f', g')\varphi_0$ and $\psi_2 = U(f'', g'')\varphi_0$ in Eq. (5.19). This formula in turn implies Eq. (5.19) for all $\psi_1, \psi_2 \in \mathcal{H}$ and for $\varphi_1 = \varphi_2 = \varphi_0$, by the same argument as in Part I, Sec. 2.

Corollary 5.2: Let the assumptions be as in Theorems 5.1 and 5.2. Then, from

$$\sup_\beta \overline{\lim_n} \int_{W_n\beta \times W_n\beta} d^n f\, d^n g\, \frac{1}{(2\pi)^n} |K(0,0;f,g)|^2 = 1,$$

it follows that there exists a sequence of subspaces W_n of $\mathfrak{U}$, dim $W_n = N(n) < \infty$, such that for all f', $g', f'', g'' \in \mathfrak{U}$

$$\lim_{n\to\infty} \int_{W_n \times W_n} d^N f\, d^N g\, \frac{1}{(2\pi)^N} K(f', g'; f, g)K(f, g; f', g') = K(f', g'; f'', g''). \quad (5.21)$$

6. DISCUSSION

Up to now it has been an open question whether or not there exists a kernel integral formula in the form of a supremum over the possible bases of the test function space for *all* irreducible representations. For this case we have obtained partial results only. At least one knows by Theorem 2.1 that reducible representations are excluded. On the other hand, the case of limit superior has been characterized completely by the PTPR. One may ask for which of these the kernel integral in its strongest form, the ordinary limit, holds. This is also not known. Only if the limit is independent of the ordering of the basis vectors of $\mathfrak{U}$ has one a complete description by the TPR.

The similarity between Theorem 6.2 of Part I for representations with cyclic field and Theorem 5.1 for irreducible representations may be an indication that the validity of a kernel formula could imply that the representation is both irreducible and cyclic with respect to the field. Although this has been proved only for the limit and limit superior through the corresponding properties of PTPR, in particular Theorem 6.2 seems to point in this direction. It reminds one somehow of PTPR—however, with the difference that the subspaces W_n appearing there need not be increasing, i.e., ordered with respect to inclusion. In any $\mathcal{H}_n = P_n\mathcal{H}$, the operators $U(f)$, $f \in W_n$, are cyclic since one deals with the Schrödinger representation for finitely many degrees of freedom,

[11] J. Dixmier, Ref. 10, p. 33, Corollary. The existence of such a subsequence is also easily shown directly by the same construction as in the proof of Theorem 6.3 of Part I.

and this may be an indication that $U(f), f \in \bigcup_n W_n$, is cyclic. However, since the $\mathcal{H}_n$ need neither be orthogonal nor contained in each other, it is an open question if one can find a vector

$$\varphi \in \mathfrak{H} = \overline{\bigcup_n \mathcal{H}_n}$$

cyclic for $U(f), f \in \bigcup_n W_n$, and we have not succeeded in constructing such a vector.

If our conjecture about the cyclicity of $U(f)$ were correct, one would have a close analogy to the finite case where one also has irreducibility and cyclicity of $U(f)$. But then, in the infinite case, the kernel integral would not hold for all irreducible representations.

The results of this paper admit of straightforward applications to obtain sufficiency criteria for irreducibility and to obtain necessary and sufficient conditions for PTPR in such a way that only the vacuum functional or the kernels enter. A general sufficient condition of this kind for irreducibility is the validity of Eq. (5.21). Theorems 4.2 and 4.3, as well as Corollary 4.1, can be regarded as necessary and sufficient conditions for TPR and PTPR.

ACKNOWLEDGMENTS

The author's thanks are due to Professor G. Ludwig for his generous encouragement. The author is indebted to Professor J. R. Klauder for several stimulating discussions. He would also like to thank Professor A. Salam and Professor P. Budini for the hospitality extended to him at the International Centre for Theoretical Physics, Trieste, where part of this work was done. Financial support of the Bundesministerium für Wissenschaftliche Forschung, Bonn, is gratefully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Semisimple Subgroups of Semisimple Groups

ANTHONY P. STONE
Mathematics Department, St. Stephen's College, Delhi-7, India

(Received 25 November 1968)

The projection theorem for weights of a representation of a semisimple group G on restriction to a semisimple subgroup is derived, and the existence of a subgroup corresponding to a given projection is discussed. Dynkin's definition of the index of a simple subgroup is extended to the case of G being only semisimple, and the geometrical meaning of the index is given. A method is developed for finding branching rules for both regular and nonregular subgroups. Explicit general formulas for the branching multiplicities are obtained for all cases when G is of rank 2 and for $B_3(R_7) \to G_2$. Applications to the construction of weight diagrams and the "state-labeling" problems for B_2 and G_2 are mentioned.

1. INTRODUCTION

The problem of obtaining branching rules for representations of semisimple Lie groups restricted to a subgroup has recently been treated by various methods.[1,2] The same subgroup can often be realized in different ways with different branching rules. For subgroups of simple groups, Dynkin[3] introduced the index of a simple subgroup to distinguish the different possibilities.

The application to elementary particle theory of some of the $R_3(A_1)$ subgroups of simple groups of rank 2 has been discussed by Behrends *et al.*[4] Another procedure which ultimately rests on branching rules is Racah's[5,6] construction of the infinitesimal operators of $SU_{2n+1}(A_{2n})$, $R_{2n+1}(B_n)$, and G_2 out of tensor operators under R_3, in order to give a group-theoretical classification of atomic and nuclear states in L–S coupling. For j–j coupling, the infinitesimal operators of $Sp_{2n}(C_n)$ can be constructed by the same method.[6] The tensor operators required are those belonging to the irreducible representations of R_3 occurring in a certain branching rule for the regular (adjoint) representation (e) of the desired group. For example,

[1] M. L. Whippman, J. Math. Phys. **6**, 1534 (1965).
[2] V. B. Mandel'tsveig, Sov. Math. **6**, 851 (1965); N. Straumann, Helv. Phys. Acta **38**, 481 (1965); A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967); Nuovo Cimento Suppl. **5**, 963 (1967); B. Gruber, (report of work prior to publication). Two noteworthy papers, inaccessible to the author up to the time of writing, are: A. Mal'cev, Am. Math. Soc. Transl. No. 33 (1950); Shi Sheng-Ming, Chinese Math. **6**, 610 (1965).
[3] E. B. Dynkin, Am. Math. Soc. Transl. II **6**, 111 (1957). For the index see pp. 130, 122.
[4] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[5] G. Racah, "Group Theory and Spectroscopy," lecture notes, Institute for Advanced Study, Princeton, 1951 (unpublished).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, New York, 1963), Chap. 5.

and this may be an indication that $U(f), f \in \bigcup_n W_n$, is cyclic. However, since the $\mathcal{H}_n$ need neither be orthogonal nor contained in each other, it is an open question if one can find a vector

$$\varphi \in \mathfrak{H} = \overline{\bigcup_n \mathcal{H}_n}$$

cyclic for $U(f), f \in \bigcup_n W_n$, and we have not succeeded in constructing such a vector.

If our conjecture about the cyclicity of $U(f)$ were correct, one would have a close analogy to the finite case where one also has irreducibility and cyclicity of $U(f)$. But then, in the infinite case, the kernel integral would not hold for all irreducible representations.

The results of this paper admit of straightforward applications to obtain sufficiency criteria for irreducibility and to obtain necessary and sufficient conditions for PTPR in such a way that only the vacuum functional or the kernels enter. A general sufficient condition of this kind for irreducibility is the validity of Eq. (5.21). Theorems 4.2 and 4.3, as well as Corollary 4.1, can be regarded as necessary and sufficient conditions for TPR and PTPR.

ACKNOWLEDGMENTS

The author's thanks are due to Professor G. Ludwig for his generous encouragement. The author is indebted to Professor J. R. Klauder for several stimulating discussions. He would also like to thank Professor A. Salam and Professor P. Budini for the hospitality extended to him at the International Centre for Theoretical Physics, Trieste, where part of this work was done. Financial support of the Bundesministerium für Wissenschaftliche Forschung, Bonn, is gratefully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Semisimple Subgroups of Semisimple Groups

ANTHONY P. STONE
Mathematics Department, St. Stephen's College, Delhi-7, India

(Received 25 November 1968)

The projection theorem for weights of a representation of a semisimple group G on restriction to a semisimple subgroup is derived, and the existence of a subgroup corresponding to a given projection is discussed. Dynkin's definition of the index of a simple subgroup is extended to the case of G being only semisimple, and the geometrical meaning of the index is given. A method is developed for finding branching rules for both regular and nonregular subgroups. Explicit general formulas for the branching multiplicities are obtained for all cases when G is of rank 2 and for $B_3(R_7) \to G_2$. Applications to the construction of weight diagrams and the "state-labeling" problems for B_2 and G_2 are mentioned.

1. INTRODUCTION

The problem of obtaining branching rules for representations of semisimple Lie groups restricted to a subgroup has recently been treated by various methods.[1,2] The same subgroup can often be realized in different ways with different branching rules. For subgroups of simple groups, Dynkin[3] introduced the index of a simple subgroup to distinguish the different possibilities.

The application to elementary particle theory of some of the $R_3(A_1)$ subgroups of simple groups of rank 2 has been discussed by Behrends *et al.*[4] Another procedure which ultimately rests on branching rules is Racah's[5,6] construction of the infinitesimal operators of $SU_{2n+1}(A_{2n})$, $R_{2n+1}(B_n)$, and G_2 out of tensor operators under R_3, in order to give a group-theoretical classification of atomic and nuclear states in L–S coupling. For j–j coupling, the infinitesimal operators of $Sp_{2n}(C_n)$ can be constructed by the same method.[6] The tensor operators required are those belonging to the irreducible representations of R_3 occurring in a certain branching rule for the regular (adjoint) representation (e) of the desired group. For example,

[1] M. L. Whippman, J. Math. Phys. **6**, 1534 (1965).
[2] V. B. Mandel'tsveig, Sov. Math. **6**, 851 (1965); N. Straumann, Helv. Phys. Acta **38**, 481 (1965); A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967); Nuovo Cimento Suppl. **5**, 963 (1967); B. Gruber, (report of work prior to publication). Two noteworthy papers, inaccessible to the author up to the time of writing, are: A. Mal'cev, Am. Math. Soc. Transl. No. 33 (1950); Shi Sheng-Ming, Chinese Math. **6**, 610 (1965).
[3] E. B. Dynkin, Am. Math. Soc. Transl. II **6**, 111 (1957). For the index see pp. 130, 122.
[4] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[5] G. Racah, "Group Theory and Spectroscopy," lecture notes, Institute for Advanced Study, Princeton, 1951 (unpublished).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, New York, 1963), Chap. 5.

TABLE I. Branching rules for regular representations.

G	G'	(e')	$\Sigma(f)$
A_2	A_1^1	(1)	$2(\frac{1}{2}) + (0)$
	A_1^4	(1)	(2)
$A_1 \times A_1$	A_1^1	(1)	3(0)
	A_1^2	(1)	(1)
$B_2 = C_2$	$A_1^1 \times A_1^1$	$(1) \times (0) + (0) \times (1)$	$(\frac{1}{2}) \times (\frac{1}{2})$
	A_1^1	(1)	$2(\frac{1}{2}) + 3(0)$
	A_1^2	(1)	$2(1) + (0)$
	A_1^{10}	(1)	(3)
G_2	A_2^1	(1, 1)	(1, 0) + (0, 1)
	$A_1^3 \times A_1^1$	$(1) \times (0) + (0) \times (1)$	$(\frac{3}{2}) \times (\frac{1}{2})$
	A_1^1	(1)	$4(\frac{1}{2}) + 3(0)$
	A_1^3	(1)	$2(\frac{3}{2}) + 3(0)$
	A_1^4	(1)	$(2) + 2(1)$
	A_1^{28}	(1)	(5)

in Racah's method, $R_7 \rightarrow R_3$ has

$$(e) \rightarrow (5) + (3) + (1).$$

If, however, R_3 is obtained by restricting rotations in a 7-dimensional space to those in a 3-dimensional subspace, the branching rule is

$$(e) \rightarrow 5(1) + 6(0).$$

Again, $G_2 \rightarrow R_3$ by Racah's method gives

$$(e) \rightarrow (5) + (1),$$

whereas Zorn's[7] construction of G_2 depends upon

$$(1, 0) \rightarrow 2(1) + (0),$$

which corresponds to

$$(e) \rightarrow (2) + 3(1),$$

and there are still other possibilities (see Table I).

In the present paper, the projection theorem[8]—that the weights of a representation project orthogonally into the weights of the induced representations of a subgroup—is proved in Sec. 2. The proof makes use of the one-to-one correspondence between basis vectors and components of tensor operators, and also their unique properties, namely the normalization of the vectors and the existence of Hermitian conjugate operators.[9] The method also brings out the geometrical significance of the index of a simple subgroup. While Dynkin's definition of the index can be applied immediately to semisimple subgroups of a simple group (each simple component having its own index), the case of semisimple containing groups is important and not trivial (R_4 is an example). In Sec. 3, the concept of an index is extended to the general case, and the geometrical meaning is given. Conditions for the existence of a subgroup corresponding to a given projection are discussed in Sec. 4.

Section 5 describes a general method of finding branching rules. The branching multiplicities are obtained in explicit general form in Sec. 6 for all semisimple subgroups of semisimple groups of rank 2 and also for $B_3(R_7) \rightarrow G_2$. Applications to the problem of constructing the weight diagrams for G_2 and B_2 and their state-labeling problems are mentioned in Sec. 7. The Appendix describes the method of obtaining the series expansions needed in Sec. 6.

2. THE PROJECTION THEOREM

First, the notation for infinitesimal generators and tensor operators is explained. Let X_μ be the infinitesimal generators of a semisimple group G of rank N with E_α being the generator corresponding to a nonzero root α and H_r ($r = 1, 2, \cdots, N$) the commuting generators. The basis vectors of a representation (k) of G will be labeled by the set of numbers $\mathbf{k} \equiv k\mathbf{k}_r z$, where the $\mathbf{k}_r$ are the components of the weight of $\mathbf{k}$ and the numbers z distinguish vectors of the same weight. Another vector of the same representation will be labeled by $'\mathbf{k} \equiv k'\mathbf{k}_r z'$. The summation convention will be applied to the symbols $\mathbf{k}$ and to other weight labels.

A tensor operator[10] $T(k)$ belonging to the representation (k) of G has components $T(\mathbf{k})$, where

$$[X_\mu, T(\mathbf{k})] = \langle '\mathbf{k}| X_\mu |\mathbf{k}\rangle T('\mathbf{k}). \tag{1}$$

The generators themselves form a tensor operator $X(e)$ belonging to the regular representation, whose components may be chosen to be

$$X(\alpha) = E_\alpha, \quad X(r) = K^{-\frac{1}{2}} H_r, \tag{2}$$

when the metric tensor is

$$g_{rs} = \sum_\alpha \alpha_r \alpha_s = K\delta_{rs}, \tag{3}$$

and $X_{-\mu}$ is the Hermitian conjugate of X_μ.

Let a semisimple subgroup G' of G, of rank N', have roots a, and generators Y_u, or, in split form, F_a, J_ρ. The branching rule for the representation (k) of G on restriction to G' has the form

$$(k) \rightarrow \sum n(l)(l), \tag{4}$$

[7] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962), p. 142.

[8] E. B. Dynkin, Ref. 3, p. 141.

[9] See G. E. Baird and L. C. Biedenharn, J. Math. Phys. 5, 1730 (1964).

[10] The basic properties of these tensor operators may be found in A. P. Stone [Proc. Cambridge Phil. Soc. 57, 460 (1961)], where the introduction of irreducible tensor operators should have been attributed to E. P. Wigner, *Gruppentheorie* (Frederick Vieweg und Sohn, Braunschweig, 1931), p. 262.

where the numbers $n(l)$ are the branching multiplicities. Then a tensor operator $T(k)$ under G is reduced to tensor operators $S(l)$ under G' by a reducing matrix V:

$$S(\mathbf{l}) = V(\mathbf{l}, \mathbf{k})T(\mathbf{k}). \tag{5}$$

A matrix U which reduces the basis vectors of a representation also reduces the corresponding tensor operator.[10] The correspondence between these reducing matrices is not one-to-one, however. If G' is the direct product of a number of simple groups G_j' with regular representations (f_j), Eq. (1) shows that since the weights of each simple subgroup lie in orthogonal spaces, $Y(f_j)$ and $\lambda_j^{-1}Y(f_j)$ satisfy the same equations, where the λ_j are constants. Choosing the bases to be orthonormal, $X(e)$ is reduced to the set $\lambda_j^{-1}Y(f_j)$, for some values of λ_j, by a unitary matrix U.

Gantmacher[11] proved that one can choose J_ρ not to contain E_α. Suppose the generators of G' are given by

$$F_a = W(a, \mu)X_\mu, \tag{6}$$

$$J_\rho = W(\rho, r)H_r. \tag{7}$$

It will first be shown that F_a does not contain H_r. By substituting Eqs. (5) and (7) in the equation

$$[J_\rho, S(\mathbf{l})] = \mathbf{l}_\rho S(\mathbf{l})$$

and taking the coefficient of $T(\mathbf{k})$, one obtains for each $\mathbf{k}, \mathbf{l}$

$$V(\mathbf{l}, \mathbf{k})[W(\rho, r)\mathbf{k}_r - \mathbf{l}_\rho] = 0. \tag{8}$$

Taking $\mathbf{k}_r = 0$ gives

$$V(\mathbf{l}, \mathbf{k}) = 0 \quad (\mathbf{k}_r = 0, \mathbf{l}_\rho \neq 0).$$

For the regular representations this is

$$W(a, r) = 0.$$

By taking the Hermitian conjugate of Eq. (6),

$$W(-a, -\alpha) = [W(a, \alpha)]^*.$$

Since the operators in Eq. (7) are Hermitian, $W(\rho, r)$ is real. If G is the direct product of simple groups G_i, $W(\rho, r)$ is given in terms of the unitary matrix U by

$$W(\rho(j), r(i)) = \lambda_j(K_j/K_i)^{\frac{1}{2}}U(\rho(j), r(i)), \tag{9}$$

where i, j labels quantities and indices for G_i, G_j', respectively, and Eq. (2) has been used. Writing $|\lambda_j|O$ for $\lambda_j U$, where O is real, shows that O is part of a real orthogonal matrix. Thus the λ_j may be taken to be real and positive, and Eq. (9) holds with U replaced by O.

[11] F. Gantmacher, Mat. Sb. N.S. 5(47), 101 (1939), quoted in Ref. 3, p. 123.

The projection theorem now follows from Eq. (8); if $\mathbf{l}$ comes from $\mathbf{k}$, then

$$\mathbf{l}_{\rho(j)} = \lambda_j(K_j/K_i)^{\frac{1}{2}}O(\rho(j), r(i))\mathbf{k}_{r(i)}. \tag{10}$$

This is an orthogonal projection of the weights of (k) into those of (l) with changes of scale, provided that K_i is the same for all groups G_i whose weights project into the same weight of G'. The simplest condition is to make all K_i the same. If $N' = N$, the orthogonal projection becomes a rotation (with reflections, in general).

3. INDICES OF A SUBGROUP

The definition of an index involves the scalar product for a group. By Eq. (3) the scalar product in the weight space of G_i is

$$(p, q)_i \equiv K_i^{-1}(p, q)_0, \tag{11}$$

where

$$(p, q)_0 \equiv p_r q_r. \tag{12}$$

The scalar product (11) is invariant when the generators H_r of G_i are multiplied by a constant factor, thus preserving the form of Eq. (3). This will be referred to as a renormalization of the operators. If the longest root δ is to have the same length c for several groups, the scalar product has to be multiplied by a factor, giving

$$(p, q) = c^2(p, q)_0/(\delta, \delta)_0. \tag{13}$$

Dynkin[3] defined the index I of a simple subgroup G' of a simple group G as follows: If the scalar products are chosen to make the longest roots of G, G' have equal lengths $c = \sqrt{2}$, and f^* maps the weights of G into those of G', a reverse mapping f is given by

$$(f^*(\mathbf{k}), \mathbf{l}) = (\mathbf{k}, f(\mathbf{l})), \tag{14}$$

where $\mathbf{l}$, $\mathbf{k}$ standing alone represent their weights. Then I is defined by

$$(f(\mathbf{l}), f(\mathbf{l})) = I(\mathbf{l}, \mathbf{l}).$$

The scalar products are those of the appropriate group.

If G' is semisimple, then each simple invariant subgroup G_j' of G' has an index I_j. If d_j is the longest root of G_j', the scalar product for G' becomes

$$(p, q) = \sum_j [c^2/(d_j, d_j)_0]p_{\rho(j)}q_{\rho(j)}. \tag{15}$$

To extend the definition of the indices to semisimple groups G, the scalar product for G has to be taken in the form (13), where δ is a root of G for which $(\delta, \delta)_0$ is greatest, all K_i being equal. The scalar product for G' is still given by Eq. (15). The indices turn out to be independent of the value of c.

The mapping f^* is such that $f^*(\mathbf{k})$ is just $\mathbf{l}$ given by Eq. (10). Equation (14) then shows that

$$f_r(\mathbf{l}) = \sum_j \frac{\lambda_j K_j^{\frac{1}{2}}(\delta, \delta)_0}{K_1^{\frac{1}{2}}(d_j, d_j)_0} O(\rho(j), r)\mathbf{l}_{\rho(j)}, \qquad (16)$$

where K_1 is the common value of the K_i. The index I_j of G_j' is defined by

$$(f(\mathbf{l}), f(\mathbf{l})) = \sum_j I_j(\mathbf{l}, \mathbf{l})_j. \qquad (17)$$

Substituting (16) in (17) gives

$$I_j = \lambda_j^2(\delta, \delta)_1/(d_j, d_j)_j. \qquad (18)$$

Now if a root α of G projects into d_j (apart from the change of scale), the relation between d_j and the projection α_P is, by Eq. (10),

$$d_j = \lambda_j(K_j/K_1)^{\frac{1}{2}}\alpha_P.$$

Since d_j is of length c,

$$\lambda_j^2 = (d_j, d_j)_j/(\alpha_P, \alpha_P)_1. \qquad (19)$$

Substituting this in Eq. (18) leads to

$$I_j = (\delta, \delta)_0/(\alpha_P, \alpha_P)_0.$$

The indices may thus be calculated very easily from the geometry of the projection.

Equation (19) shows that λ_j is invariant under renormalization of the operators. Hence, a normalization of the generators of G' can be found for which W in Eq. (9) becomes identical with O and the projected root diagram of G' is its actual root diagram. With $K_i = K_1$, this will be called a standard projection. The absolute size of the root diagrams is still variable depending upon K_1.

The definition of the indices I_j of a semisimple subgroup G' of a semisimple group G is now as follows: If the root diagrams of the G_i are such that each K_i is the same, if the scalar products are normalized so that the longest roots of G, G_j' are of equal length, where the longest root of G is determined by using the form (12) and the scalar product for G has the form of Eq. (13), and if f^* maps the weights of G into those of G', while the mapping f is given by Eq. (14), the indices I_j of the subgroups G_j' are defined by Eq. (17).

Dynkin[12] proved that for G, G' simple, the index is an integer. Indices are written as superscripts to the symbol for the subgroup.

4. CONDITIONS FOR A SUBGROUP

Dynkin[3] employs certain useful terms which will now be defined. A subgroup is *regular* if all its roots are also roots of the containing group (in a standard projection). Every semisimple group has a *principal* subgroup of type A_1, for which each simple positive root of the group projects into the positive root of A_1. The projection is thus onto the normal to the hyperplane joining the ends of the simple roots. In any case, the positive root of a subgroup A_1, as it occurs in the projection, is termed the *defining vector* of the subgroup.

A necessary and sufficient condition for the existence of a subgroup corresponding to a suitable projection is that coefficients $W(a, \alpha)$ can be found such that the F_a have the correct commutators among themselves. The remaining commutators are guaranteed by the projection.

The general condition arising from $[F_a, F_b]$ for $a + b \neq 0$ is easily found, but it is not needed in the present paper. To deal with the commutator $[F_a, F_{-a}]$, let the generators be normalized to give a standard projection, and let a root α of G project into the root a of G'. Define A by

$$\alpha = a + A,$$

where A is perpendicular to a and has components

$$A_\sigma = O(\sigma, r)\alpha_r.$$

Then $\|O(\rho, r), O(\sigma, r)\|$ is a real orthogonal matrix. Writing

$$L_\sigma \equiv O(\sigma, r)H_r$$

leads to

$$\alpha_r H_r = a_\rho J_\rho + A_\sigma L_\sigma.$$

There are two conditions to be satisfied:

$$\sum_{\alpha,\beta} W(a, \alpha)W(-a, -\beta)c_{\alpha,-\beta}{}^{\gamma} = 0, \quad \gamma \neq 0, \qquad (20)$$

where the c's are structure constants of G and

$$\sum_\alpha |W(a, \alpha)|^2 A_\sigma = 0. \qquad (21)$$

Two particular deductions from Eqs. (20) and (21) are the following:

(i) If F_a is formed out of the two operators E_α, E_β, then $\alpha - \beta$ must not be a root of G, and the vectors α, β, a must be coplanar with a lying between α, β.

(ii) If F_a is formed from one operator E_α, then a must be identical with α, in a standard projection.

Any root of G is the defining vector of a regular subgroup of type A_1. A sufficient condition for the existence of a nonregular subgroup A_1 is the following: If at least two roots α, β satisfying condition (i) project into a, then a is the defining vector of a subgroup A_1.

[12] Ref. 3, p. 131.

For if F_a is formed from E_α and E_β, Eq. (21) and the unitary condition determine $|W(a, \alpha)|$, $|W(a, \beta)|$.

5. BRANCHING RULES

Branching rules may be obtained from the characters of the irreducible representations. A standard projection will be assumed in every case.

Let ϕ^r be the parameters of G and

$$e(q) \equiv \exp(iq_r\phi^r).$$

If the weight q of G projects into the weight m of G', the transformation to the parameters of G' is found from the equation[6]

$$m_\rho\phi'^\rho = q_r\phi^r.$$

It follows that the projected form of $e(q)$ is $e(m)$, with the convention that $e(w)$ is to be calculated with the parameters of the group to which the weight w belongs.

Weyl[13] proved that the character of the irreducible representation (k) is

$$\chi(k) = \xi(k)/\Delta,$$

where

$$\xi(k) \equiv \sum_S \delta_S e(S(k + R)),$$

$$\Delta \equiv \xi(0),$$

R being half the sum of the positive roots and S any element of the Weyl group, of parity δ_S. It is known[14] that

$$\Delta = e(R)\prod(1 - e(-\alpha)), \tag{22}$$

where the product is taken over all the positive roots α. If primes refer to a subgroup, the branching rule (4) corresponds to

$$\xi_P(k)\Delta'/\Delta_P = \sum n(l)\xi'(l), \tag{23}$$

where the suffix P indicates restriction to the subgroup by means of a particular projection.

It will now be proved that Δ_P is divisible by Δ'. The positive part of the weight space will denote the region in which every weight is positive, i.e., its first nonzero Cartesian coordinate is positive. The positive part of the weight space of G' may be chosen to lie in the positive part of the weight space of G. If roots α project into a positive root a, then either α is identical with a or not all the vectors A in Eq. (21) are negative. Hence, at least one positive root of G projects into each positive root of G' and the result follows from Eq. (22).

[13] H. Weyl, Math. Z. **24**, 377 (1926).
[14] N. Jacobson, Ref. 7, p. 252.

The ratio ξ_P/Δ_P becomes indeterminate if a positive root α projects into the null vector. The correct form is easily found in the case $N' = N - 1$. Because of reflections parallel to α the weights occurring in ξ fall into pairs $m \pm p\alpha$, where m is orthogonal to α and p is a number. The indeterminate factor is thus

$$\lim_{\alpha\to 0} \frac{e(p\alpha) - e(-p\alpha)}{1 - e(-\alpha)} = 2p.$$

The correct ratio ξ_P/Δ_P is obtained by omitting the vanishing factor in Δ and replacing ξ_P by

$$2\sum{}' \delta_S pe(m) = \sum \delta_S pe(m), \tag{24}$$

where

$$m + p\alpha = S(k + R),$$

the sum on the left of (24) being over the elements S corresponding to positive values of p and that on the right over all S, p.

To apply Eq. (23), choose the positive roots of G' so that the positive vectors which remain in Δ_P/Δ' may be expressed in terms of certain positive roots $\alpha(s)$ with positive coefficients, and write

$$x_s \equiv e(\alpha(s)).$$

Then using Eq. (22), Δ'/Δ_P can be expanded in a power series containing no positive powers of x_s. If (l), (l') are inequivalent representations of G', then $l + R'$, $S'(l' + R')$ are inequivalent under the Weyl group for G'. Hence, $n(l)$ is the constant term in the expansion of

$$e(-l - R')\xi_P(k)\Delta'/\Delta_P \tag{25}$$

in terms of the x_s. The terms of ξ_P containing a negative power of any x_s do not contribute.

6. APPLICATION

The explicit general branching multiplicities for rank-2 groups and for $B_3 \to G_2$ will now be derived. The existence of the subgroups may be deduced from the conditions of Sec. 4; they have also been tabulated by Dynkin.[3] The branching rules for the regular representations of rank-2 groups in all cases are given in Table I, where

$$(e) \to (e') + \sum (f).$$

The weights and parities occurring in ξ are given in Table II using the coordinates explained below.

The more complicated formulas quoted for the coefficients are derived in the Appendix. It is important to note that all functions are defined to be zero unless all their arguments are nonnegative integers.

TABLE II. Reflected weights.

G	δ_S	$S(k+R)$
A_2	+1	$(\lambda_1+1, \lambda_2+1)$
	−1	$(-\lambda_2-1, -\lambda_1-1)$
	+1	$(\lambda_2+1, -\lambda_1-\lambda_2-2)$
	−1	$(\lambda_1+\lambda_2+2, -\lambda_2-1)$
	+1	$(-\lambda_1-\lambda_2-2, \lambda_1+1)$
	−1	$(-\lambda_1-1, \lambda_1+\lambda_2+2)$
B_2	+1	$\pm(j_1+\frac{3}{2}, j_2+\frac{1}{2})$
	+1	$\pm(j_2+\frac{1}{2}, -j_1-\frac{3}{2})$
	−1	$\pm(j_1+\frac{3}{2}, -j_2-\frac{1}{2})$
	−1	$\pm(j_2+\frac{1}{2}, j_1+\frac{3}{2})$
G_2	+1	$\pm(u_1+2, u_2+1)$
	−1	$\pm(u_2+1, u_1+2)$
	+1	$\pm(u_1+u_2+3, -u_1-2)$
	−1	$\pm(-u_1-2, u_1+u_2+3)$
	+1	$\pm(-u_2-1, u_1+u_2+3)$
	−1	$\pm(u_1+u_2+3, -u_2-1)$

Group A_2

If α_1, α_2 are the simple positive roots, the highest weight of the irreducible representation (λ_1, λ_2), where (1, 0) and (0, 1) are the fundamental representations, is $\lambda_1\delta_1 + \lambda_2\delta_2$, where

$$\delta_1 = \tfrac{1}{3}(2\alpha_1 + \alpha_2), \quad \delta_2 = \tfrac{1}{3}(\alpha_1 + 2\alpha_2). \tag{26}$$

The vectors $\alpha_1, \delta_1, \delta_2, \alpha_2$ form a string of roots for G_2.

Branching $A_2 \to A_1^1$

If α denotes the defining vector of A_1, α may be taken to be $\alpha_1 + \alpha_2$. The projection is then $\alpha_1, \alpha_2 \to \frac{1}{2}\alpha$. By Eqs. (26), a weight (λ_1, λ_2) projects into the weight $\frac{1}{2}(\lambda_1 + \lambda_2)$ of A_1. It is convenient to write

$$x \equiv e(\tfrac{1}{2}\alpha),$$

giving

$$\Delta_P/\Delta' = x(1 - x^{-1})^2.$$

The leading term in $\xi'(j)$ is x^{2j+1}, and (25) leads to

$$n(j) = \varpi(\lambda_1 + \lambda_2 + 1 - 2j) - \varpi(\lambda_1 - 2j) - \varpi(\lambda_2 - 2j),$$

where $\varpi(n)$ is zero unless n is a positive integer, in which case $\varpi(n)$ is equal to n. The different cases are

$$\begin{aligned} n(j) &= \lambda_1 + \lambda_2 + 1 - 2j, && \tfrac{1}{2}(\lambda_1+\lambda_2) \geq j \geq \tfrac{1}{2}\lambda_>, \\ &= \lambda_< + 1, && \tfrac{1}{2}\lambda_> \geq j \geq \tfrac{1}{2}\lambda_<, \\ &= 2j + 1, && \tfrac{1}{2}\lambda_< \geq j \geq 0, \end{aligned}$$

where $\lambda_>$ is the greater and $\lambda_<$ the lesser of λ_1, λ_2. Both integral and half-integral values of j occur, and its maximum value is $\frac{1}{2}(\lambda_1 + \lambda_2)$.

Some particular cases of this branching rule have been given by Behrends *et al.*[4]

Branching $A_2 \to A_1^4$

This is the principal A_1, with the projection $\alpha_1, \alpha_2 \to \alpha$. The method is the same as before and yields

$$n(j) = P_2(\lambda_1 + \lambda_2 - j) - P_2(\lambda_1 - 1 - j) - P_2(\lambda_2 - 1 - j).$$

If n is a nonnegative integer, $P_r(n)$ is the number of partitions of n with no part greater than r. Thus

$$P_2(n) = [\tfrac{1}{2}n] + 1 \quad (n \geq 0),$$

where $[p]$ means the largest integer contained in p. Only integral values of j occur.

Some cases of this branching rule are given by Hamermesh.[15]

Group $A_1 \times A_1$

If the positive roots are called α_1, α_2, weights may be expressed conveniently either as

$$(J_1, J_2) = J_1\alpha_1 + J_2\alpha_2 \tag{27}$$

or as

$$(j_1, j_2) = \tfrac{1}{2}(\alpha_1 + \alpha_2)j_1 + \tfrac{1}{2}(\alpha_2 - \alpha_1)j_2. \tag{28}$$

Any irreducible representation has the form

$$(J_1, J_2) = (J_1) \times (J_2), \tag{29}$$

for which

$$j_1 = J_1 + J_2, \quad j_2 = J_2 - J_1, \quad |j_2| \leq j_1. \tag{30}$$

The branching rules are best obtained by using Eq. (29) and the characters of A_1.

Branching $A_1 \times A_1 \to A_1^1$

This is a regular subgroup. If the projection is $\alpha_1 \to \alpha$, $\alpha_2 \to 0$, then

$$\chi(J_1, J_2) \to (2J_2 + 1)\chi(J_1).$$

Branching $A_1 \times A_1 \to A_1^2$

Here $\alpha_1, \alpha_2 \to \alpha$, giving the principal A_1, and

$$(J_1, J_2) \to (j_1) + (j_1 - 1) + \cdots + (|j_2|). \tag{31}$$

This gives the subgroup R_3 of R_4 obtained by fixing one axis in a 4-dimensional space. The branching rule (31) and the labeling of weights of R_4 by the representations of R_3^2 have been used earlier in connection with Wigner coefficients for R_3 and R_4.[16]

Group B_2

The positive roots may be taken to be those of $A_1 \times A_1$ together with the two vectors in Eq. (28).

[15] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1962), p. 418.

[16] A. P. Stone, Proc. Cambridge Phil. Soc. **52**, 424 (1956); L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).

The simple roots are then

$$\beta_1 = \alpha_1, \quad \beta_2 = \tfrac{1}{2}(\alpha_2 - \alpha_1). \tag{32}$$

The irreducible representation (j_1, j_2) of highest weight given by Eq. (28) has $j_1 \geq j_2 \geq 0, \frac{1}{2}$.

Branching $B_2 \to A_1^1 \times A_1^1$

The positive roots of the subgroup are α_1, α_2. The branching rule is known,[17] but it can also be obtained by the present method. With representations of $A_1 \times A_1$ in the (j_1, j_2) notation,

$$(j_1, j_2) \to \sum (l_1, l_2), \quad j_1 \geq l_1 \geq j_2 \geq l_2 \geq -j_2, \tag{33}$$

where the values of l_1, l_2 shift by unity.

Branching $B_2 \to A_1^1$

Taking α_2 as α, the projection is $\beta_1 \to 0$, $\beta_2 \to \frac{1}{2}\alpha$ and the formula (24) is needed. Using Eq. (32) and the coordinates of Eq. (27), the values of p, m (now a number) follow immediately from the requirement

$$J_1\alpha_1 + J_2\alpha_2 = p\alpha_1 + m\alpha_2.$$

The result is

$$\begin{aligned} n(j) &= (j_1 - j_2 + 1)\varpi(j_1 + j_2 + 1 - 2j) \\ &\quad - (j_1 + j_2 + 2)\varpi(j_1 - j_2 - 2j) \\ &= (j_1 - j_2 + 1)(j_1 + j_2 + 1 - 2j), \\ &\qquad \tfrac{1}{2}(j_1 + j_2) \geq j \geq \tfrac{1}{2}(j_1 - j_2), \\ &= (2j_2 + 1)(2j + 1), \quad \tfrac{1}{2}(j_1 - j_2) \geq j \geq 0. \end{aligned} \tag{34}$$

Both integral and half-integral values of j occur.

Branching $B_2 \to A_1^2$

Here $\beta_1 + \beta_2$ may be taken to be α, and the projection is $\beta_1 \to \alpha$, $\beta_2 \to 0$. Proceeding as in the previous case,

$$\begin{aligned} n(j) &= (2j_2 + 1)\varpi(j_1 + 1 - j) - (2j_1 + 3)\varpi(j_2 - j) \\ &= (2j_2 + 1)(j_1 + 1 - j), \quad j_1 \geq j \geq j_2, \\ &= (j_1 - j_2 + 1)(2j + 1), \quad j_2 \geq j \geq 0, \tfrac{1}{2}. \end{aligned} \tag{35}$$

The values of j are either all integral or all half-integral, as j_1, j_2 are.

Some cases of Eqs. (34), (35) have been given by Behrends *et al.*[4]

Branching $B_2 \to A_1^{10}$

This is the principal A_1. The projection is β_1, $\beta_2 \to \alpha$, and

$$\begin{aligned} n(j) &= P_3(2j_1 + j_2 - j) - P_3(2j_1 - j_2 - 1 - j) \\ &\quad - P_3(j_1 + 2j_2 - 1 - j) \\ &\quad + P_3(|j_1 - 2j_2 + \tfrac{1}{2}| - \tfrac{7}{2} - j). \end{aligned} \tag{36}$$

[17] H. Boerner, *Representations of Groups* (North-Holland Publ. Co., Amsterdam, 1963), p. 251.

The function is given by

$$P_3(n) = \tfrac{1}{12}(n + 1)(n + 5) + (\tfrac{1}{4}, 0)(n) + (\tfrac{1}{3}, 0, 0)(n), \tag{37}$$

where

$$(a_1, a_2, \cdots, a_m)(n) = a_p, \quad n \equiv p - 1 \bmod m.$$

The values of j are either all integral or all half-integral. In the application to the classification of states, only integral values of j occur. Some cases of the branching rule for integral j are given by Judd.[6]

Groups $Sp(4)$, $Sp(2)$

The root diagram of $Sp(4)(C_2)$ is similar to that of B_2, but the irreducible representations of $Sp(4)$ are usually labeled (σ_1, σ_2), where

$$\sigma_1 = j_1 + j_2, \quad \sigma_2 = j_1 - j_2.$$

$Sp(2)$ is an example of A_1, and its irreducible representations are labeled (σ), where

$$\sigma = 2j.$$

Making these substitutions in Eq. (34), we have the branching rule for $Sp(4) \to Sp(2)$ given by Whippman.[1] Other cases may be obtained from Eqs. (35) and (36). The indices of the subgroups are unchanged.

Group G_2

In terms of the vectors defined for A_2, the simple roots of G_2 are

$$\gamma_1 = \alpha_1, \quad \gamma_2 = \delta_1 - \alpha_1.$$

The irreducible representations of G_2 in the notation of Racah[6,18] are of highest weight

$$(u_1, u_2) = u_1\delta_2 + u_2\delta_1, \quad u_1 \geq u_2 \geq 0,$$

where the numbers u_1, u_2 are integers.

Branching $G_2 \to A_2^1$

This is a regular subgroup, the positive roots being α_1, α_2, $\alpha_1 + \alpha_2$. Writing

$$x \equiv e(\delta_1), \quad y \equiv e(\gamma_2)$$

gives

$$\begin{aligned} \Delta'/\Delta &= [xy(1 - x^{-1})(1 - y^{-1})(1 - x^{-1}y^{-1})]^{-1} \\ &= \sum M(n, m)x^{-n-1}y^{-m-1}. \end{aligned} \tag{38}$$

If n, m are both nonnegative integers, then

$$M(n, m) = \min(n, m) + 1.$$

One finds

$$\begin{aligned} n(\lambda_1, \lambda_2) = {} & M(u_1 + u_2 - \lambda_1 - \lambda_2, u_1 - \lambda_2) \\ & - \varpi(u_1 - \lambda_1 - \lambda_2) - \varpi(u_2 - \lambda_2) \\ & + \varpi(u_2 - 1 - \lambda_1 - \lambda_2). \end{aligned}$$

[18] G. Racah, Phys. Rev. 76, 1352 (1949).

This leads to

$$n(\lambda_1, \lambda_2) = n(\lambda_2, \lambda_1).$$

If $u_1 \geq \lambda_1, \lambda_2 \geq u_2$, then

$$\begin{aligned} n(\lambda_1, \lambda_2) &= u_1 + u_2 + 1 - \lambda_1 - \lambda_2, \\ &\qquad u_1 + u_2 \geq \lambda_1 + \lambda_2 \geq u_1, \\ &= u_2 + 1, \quad \lambda_1 + \lambda_2 \leq u_1; \end{aligned}$$

if $\lambda_> \geq u_2, \lambda_< \leq u_2$, then

$$\begin{aligned} n(\lambda_1, \lambda_2) &= u_1 + 1 - \lambda_>, \quad \lambda_1 + \lambda_2 \geq u_1, \\ &= \lambda_< + 1, \qquad \lambda_1 + \lambda_2 \leq u_1; \end{aligned}$$

or if $\lambda_1, \lambda_2 \leq u_2$, then

$$\begin{aligned} n(\lambda_1, \lambda_2) &= u_1 - u_2 + 1, \qquad \lambda_1 + \lambda_2 \geq u_1, \\ &= \lambda_1 + \lambda_2 - u_2 + 1, \quad u_1 \geq \lambda_1 + \lambda_2 \geq u_2, \\ &= 0, \qquad \lambda_1 + \lambda_2 < u_2. \end{aligned} \tag{39}$$

Branching $G_2 \to A_1^3 \times A_1^1$

Here γ_2 will be taken as the defining vector of A_1^3, and $2\gamma_1 + 3\gamma_2$ that of A_1^1. The irreducible representations of the subgroup have highest weights

$$(J_1, J_2) = J_1\gamma_2 + J_2(2\gamma_1 + 3\gamma_2).$$

The result is

$$\begin{aligned} &n((J_1) \times (J_2)) \\ &\quad = A(u_1 + u_2 - 2J_2, 2u_1 + u_2 - J_1 - 3J_2) \\ &\qquad - A(u_1 - 1 - 2J_2, 2u_1 + u_2 - J_1 - 3J_2) \\ &\qquad - A(u_1 + u_2 - 2J_2, u_1 + 2u_2 - 1 - J_1 - 3J_2) \\ &\qquad + A(u_2 - 2 - 2J_2, u_1 + 2u_2 - 1 - J_1 - 3J_2) \\ &\qquad + A(u_1 - 1 - 2J_2, u_1 - u_2 - 4 - J_1 - 3J_2), \end{aligned}$$

where

$$\begin{aligned} A(n, m) &= \tfrac{1}{12}(m + 1)(m + 5) - \tfrac{1}{4}\varpi(m - n) \\ &\qquad \times (m - n + 2) + \tfrac{1}{4}\{\varpi(m - 2n - 1)\}^2 \\ &\qquad + \tfrac{1}{4}\{(1, 0)(m) - (1, 0)(m - n - 1) \\ &\qquad - (1, 0)(m - 2n - 2)\} \\ &\qquad + (\tfrac{1}{3}, 0, 0)(m), \quad m \leq 3n, \\ A(n, m) &= 0, \quad m > 3n. \end{aligned} \tag{40}$$

A simple generating function for $A(n, m)$ is given in the Appendix.

The irreducible representations

$$\begin{gathered} (\tfrac{1}{2}u_1 - \tfrac{1}{2}u_2) \times (\tfrac{1}{2}u_1 + \tfrac{1}{2}u_2), \\ (u_1 + \tfrac{1}{2}u_2) \times (\tfrac{1}{2}u_2), \quad (\tfrac{1}{2}u_1 + u_2) \times (\tfrac{1}{2}u_1) \end{gathered}$$

correspond to the highest weight and two equivalent weights in G_2. The distinct representations in these three are among those which occur just once in the branching rule for (u_1, u_2).

The branching rules for $G_2 \to A_2$, $A_1 \times A_1$ have been given in a slightly different form by Mandel'tsveĭg.[19]

Branching $G_2 \to A_1^1$

If $2\gamma_1 + 3\gamma_2$ is α, the projection is $\gamma_1 \to \frac{1}{2}\alpha$, $\gamma_2 \to 0$, and

$$\begin{aligned} n(j) &= (u_1 - u_2 + 1)C_4(u_1 + u_2 - 2j) \\ &\quad - (u_1 + 2u_2 + 4)C_4(u_1 - 1 - 2j) \\ &\quad + (2u_1 + u_2 + 5)C_4(u_2 - 2 - 2j), \end{aligned}$$

where $C_4(n)$ is the coefficient of x^n in the expansion of $(1 - x)^{-4}$ in ascending powers of x.

Branching $G_2 \to A_1^3$

Here α is taken to be $\gamma_1 + 2\gamma_2$. Hence $\gamma_1 \to 0$, $\gamma_2 \to \frac{1}{2}\alpha$, and

$$\begin{aligned} n(j) &= (u_2 + 1)B(2u_1 + u_2 - 2j) \\ &\quad - (u_1 + 2)B(u_1 + 2u_2 - 1 - 2j) \\ &\quad + (u_1 + u_2 + 3)B(u_1 - u_2 - 4 - 2j), \end{aligned}$$

where

$$\begin{aligned} B(n) &= \tfrac{1}{2}(p + 1)(p + 2)(n + 1 - 2p), \\ p &\equiv [\tfrac{1}{3}n]. \end{aligned} \tag{41}$$

This is the subgroup considered by Behrends *et al.*[4]

Branching $G_2 \to A_1^4$

With α given by $\gamma_1 + \frac{3}{2}\gamma_2$, the projection is $\gamma_1 \to \alpha$, $\gamma_2 \to 0$, and

$$\begin{aligned} n(j) &= (u_1 - u_2 + 1)C(u_1 + u_2 - j) \\ &\quad - (u_1 + 2u_2 + 4)C(u_1 - 1 - j) \\ &\quad + (2u_1 + u_2 + 5)C(u_2 - 2 - j), \end{aligned}$$

where

$$C(n) = \tfrac{1}{24}(n + 1)(n + 3)(2n + 7) + (\tfrac{1}{8}, 0)(n). \tag{42}$$

The subgroups A_1^4, A_1^1 of G_2 are also contained in its subgroup A_2.

Branching $G_2 \to A_1^{28}$

This is the principal A_1 given by $\gamma_1, \gamma_2 \to \alpha$. The present method gives

$$\begin{aligned} n(j) &= P_5(3u_1 + 2u_2 - j) - P_5(3u_1 + u_2 - 1 - j) \\ &\quad - P_5(2u_1 + 3u_2 - 1 - j) \\ &\quad + P_5(u_1 + 3u_2 - 3 - j) \\ &\quad + P_5(2u_1 - u_2 - 5 - j) \\ &\quad - P_5(|u_1 - 2u_2| - 8 - j), \end{aligned}$$

[19] V. B. Mandel'tsveĭg, Yad. Fiz. **1**, 1106 (1965) [Sov. J. Nucl. Phys. **1**, 787 (1965)].

with

$$P_5(n) = (\tfrac{1}{2880})(n+1)(n^3+29n^2+281n+769) + (\tfrac{1}{64})(2n+11)(1,0)(n) + (\tfrac{1}{9},0,0)(n) + \tfrac{1}{4}([\tfrac{1}{4}n]+1) + (\tfrac{1}{5},0,0,0,0)(n). \quad (43)$$

The recurrence relation satisfied by the $P_r(n)$ is given in the Appendix.

In terms of the parameters, this subgroup is given by $\phi^1 \to 3\phi$, $\phi^2 \to 2\phi$.[6] By elementary trigonometry one finds, in terms of characters of A_1,

$$\chi(u_1,u_2) \to \frac{\chi(\frac{1}{2}u_2)\chi(\frac{1}{2}u_1-\frac{1}{2}u_2)\chi(\frac{1}{2}u_1+\frac{1}{2})\chi(\frac{1}{2}u_1+\frac{1}{2}u_2+1)\chi(\frac{1}{2}u_1+u_2+\frac{3}{2})\chi(u_1+\frac{1}{2}u_2+2)}{\chi(\frac{1}{2})\chi(1)\chi(\frac{3}{2})\chi(2)}. \quad (44)$$

The quotient of two characters is found from their multiplication rule. Equation (44) is easy to use in cases where each character in the denominator exactly divides one in the numerator.

This subgroup is the one occurring in spectroscopy.[5,6,18]

Branching $B_3 \to G_2^1$

The positive roots of B_3 are e_r and $e_r \pm e_s$ $(r<s)$, where the e_r are three orthogonal unit vectors. The simple roots are

$$\alpha_1 = e_1 - e_2, \quad \alpha_2 = e_2 - e_3, \quad \alpha_3 = e_3.$$

The irreducible representations have highest weights

$$(w_1,w_2,w_3) = w_1e_1 + w_2e_2 + w_3e_3, \qquad w_1 \geq w_2 \geq w_3 \geq 0, \tfrac{1}{2}.$$

The root diagram of G_2, with positive roots in the positive part of the weight space, is obtained by projecting onto the plane perpendicular to $e_1 - e_2 - e_3$.[20] The projection is

$$\alpha_1, \quad \alpha_3 \to \gamma_2, \quad \alpha_2 \to \gamma_1.$$

Write

$$x \equiv e(\gamma_2), \quad y \equiv e(\gamma_1+\gamma_2).$$

Then Δ'/Δ_I is exactly the same as in Eq. (38). The characters of B_3 may be taken in the determinant form given by Judd,[6] and the final result is

$$\begin{aligned} n(u_1,u_2) &= M(w_1+w_3-u_1, w_1+w_2-u_1-u_2) \\ &\quad - M(w_1-w_3-1-u_1, w_1+w_2-u_1-u_2) \\ &\quad + M(w_2-w_3-2-u_1, w_1+w_2-u_1-u_2) \\ &\quad - M(w_2+w_3-1-u_1, w_1+w_2-u_1-u_2) \\ &\quad + M(w_2+w_3-1-u_1, w_1+w_3-1-u_1-u_2) \\ &\quad + M(w_1-w_2-2-u_1, w_1+w_3-1-u_1-u_2) \\ &\quad - M(w_2-w_3-2-u_1, w_1-w_3-2-u_1-u_2) \\ &\quad - M(w_1-w_2-2-u_1, w_1-w_3-2-u_1-u_2) \\ &\quad - \varpi(w_1+w_3-u_1-u_2) \\ &\quad + \varpi(w_1-w_3-1-u_1-u_2). \end{aligned} \quad (45)$$

A particular case of Eq. (45) is

$$n(0,0) = 1, \quad \text{when} \quad w_1 = w_2 = w_3,$$
$$n(0,0) = 0, \quad \text{otherwise.}$$

By dealing with the whole characters, Judd[6] obtained the branching rule in a different form.

7. WEIGHT DIAGRAMS

There are various methods of determining the weight structure of the irreducible representations of B_2 and G_2.[21,22] The method of using the branching rule for a rank-2 subgroup is discussed here.

For B_2, Eq. (33) gives the branching rule for $B_2 \to A_1 \times A_1$. The irreducible representations (l_1,l_2) of $A_1 \times A_1$ are easily expressed in the $(J_1) \times (J_2)$ form by Eq. (30), and their weights, which are simple, are plotted. The superposition of these weight diagrams is the weight diagram for the irreducible representation (j_1,j_2) of B_2.

Since the irreducible representations of B_2 are simply reducible as representations of $A_1 \times A_1$, the state-labeling problem is completely solved by taking (l_1,l_2) as the numbers z.

To deal with G_2, it is simplest to use its subgroup A_2. The branching rule is given in Eq. (39). The weight structure of all irreducible representations (λ_1,λ_2) of A_2 is known[22,23]: The boundary weights are simple, and the multiplicity increases by one at each step inwards to a polygon with sides parallel to the boundary, until the polygon becomes a triangle; and then the multiplicity remains constant at $\lambda_< + 1$.

The representations (u,u) and $(u,0)$ of G_2 are simply reducible under A_2, and the state-labeling problem reduces to that for A_2, where it has been solved[24] by using the subgroup A_1^1. The labels z are

[20] The geometry of the projection was known to B. R. Judd, private communication (1956).

[21] J. Tarski, J. Math. Phys. **4**, 569 (1963); I. A. Malkin and V. B. Mandel'tsveĭg, Yad. Fiz. **2**, 154 (1965) [Sov. J. Nucl. Phys. **2**, 108 (1966)]; B. Gruber and F. Zaccaria, Nuovo Cimento Suppl. **5**, 914 (1967); D. Radhakrishnan and T. S. Santhanam, J. Math. Phys. **8**, 2206 (1967); B. Gruber and H. J. Weber, Proc. Roy. Irish Acad. **66A**, 31 (1968).

[22] B. Gruber, J. Math. Phys. **7**, 1797 (1966).

[23] A formula for the multiplicity of a weight has been given by B. Gruber and T. S. Santhanam, Nuovo Cimento **45A**, 1046 (1966); B. Gruber, Nuovo Cimento **48A**, 23 (1967).

[24] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

$(\lambda_1, \lambda_2)(j)$. Alternatively, the subgroup $A_1 \times A_1$ may be used for some representations. For example, (2, 1) is simply reducible under $A_1 \times A_1$ but not under A_2. These procedures, however, do not solve the general state-labeling problem for G_2.

APPENDIX

The expansion of Δ'/Δ_I is facilitated by some formulas developed by MacMahon.[25] Let

$$(\mathbf{n}) \equiv 1 - x^n.$$

Then

$$\frac{1}{(\mathbf{1})(\mathbf{2})\cdots(\mathbf{n})} = \sum \frac{N(P)}{(\mathbf{1})^{p_1}(\mathbf{2})^{p_2}\cdots}, \tag{A1}$$

where the sum is over all partitions $P = (1^{p_1}2^{p_2}\cdots)$ of n and

$$N(P) = 1/(1^{p_1}2^{p_2}\cdots p_1!\, p_2!\cdots).$$

Also,

$$\frac{(\mathbf{m}+\mathbf{1})(\mathbf{m}+\mathbf{2})\cdots(\mathbf{m}+\mathbf{n})}{(\mathbf{1})(\mathbf{2})\cdots(\mathbf{n})} = \sum N(P)\left\{\frac{(\mathbf{m}+\mathbf{1})}{(\mathbf{1})}\right\}^{p_1}\left\{\frac{(\mathbf{2m}+\mathbf{2})}{(\mathbf{2})}\right\}^{p_2}\cdots, \tag{A2}$$

summed over the partitions P of n.

The generating function for $P_r(n)$ is

$$\frac{1}{(\mathbf{1})(\mathbf{2})\cdots(\mathbf{r})} = \sum P_r(n)x^n.$$

By repeated use of Eq. (A1) and relations such as

$$\frac{1}{(\mathbf{3})(\mathbf{1})} = \frac{1+x+x^2}{(\mathbf{3})^2},$$

Eqs. (37) and (43) may be obtained. The values of $P_r(n)$ may also be obtained by constructing Euler's table from the recurrence relation[26]

$$P_r(n) = P_{r-1}(n) + P_r(n-r).$$

The function $A(n, m)$ of Eq. (40) arises by expanding in terms of $e(\gamma_1)$, $e(\gamma_2)$. Its generating function is

$$F(y) \equiv \frac{1}{(1-y)(1-yx)(1-yx^2)(1-yx^3)} = \sum A(n, m)y^n x^m.$$

By assuming an expansion of $F(y)$ in powers of y and relating $F(xy)$ to $F(y)$, one obtains

$$F(y) = \frac{(\mathbf{n}+\mathbf{1})(\mathbf{n}+\mathbf{2})(\mathbf{n}+\mathbf{3})}{(\mathbf{1})(\mathbf{2})(\mathbf{3})}\, y^n. \tag{A3}$$

Equation (40) then follows by the use of Eqs. (A2) and (A1). The coefficient of y^n in Eq. (A3) is a simple generating function for $A(n, m)$, expressible as a polynomial in x.

The functions $B(n)$ and $C(n)$ of Eqs. (41) and (42) are easily found from their generating functions,

$$\frac{1}{(\mathbf{1})^2(\mathbf{3})^2} = \sum B(n)x^n,$$

$$\frac{1}{(\mathbf{1})^3(\mathbf{2})} = \sum C(n)x^n.$$

[25] P. A. MacMahon, *Combinatory Analysis* (Cambridge University Press, Cambridge, England, 1916), Vol. II, Sec. VII, Chap. V.

[26] L. E. Dickson, *History of the Theory of Numbers* (Chelsea Publ. Co., New York, 1952), Vol. II, p. 104.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Charge and Pole: Canonical Coordinates without Potentials

EDWARD H. KERNER
Sharp Physics Laboratory, University of Delaware, Newark, Delaware 19711

(Received 27 March 1969)

For particles having both magnetic and electric charge it is shown that (a) in the nonrelativistic many-particle problem where only Coulomb and Biot–Savart fields need be considered and (b) in the one-particle relativistic problem (orbital pole-charge moving around a fixed pole-charge), the well-set classical dynamics can be reduced directly from the equations of motion to Hamiltonian form without the introduction of potentials and Dirac strings. The Lie–Koenigs theorem, which can give Hamiltonian format to any dynamics, is invoked for this. The essential feature is that canonical coordinates cannot be physical particle coordinates. For (a) and (b), suitable canonical variables are explicitly constructed. Using only Bohr–Sommerfeld quantization, the Schwinger charge–pole quantum condition is obtained for pure-charge–pure-pole interactions; but when Coulomb forces are additionally considered, no quantum restriction on charge and pole strength is required.

INTRODUCTION

The development[1] of the theory of magnetic monopoles, from Dirac's demonstration of the charge (e)–pole (g) quantization condition $eg/\hbar c = \frac{1}{2}n$ up to Schwinger's $eg/\hbar c = n$, has relied[2] on the introduction of singular lines or strings emanating from poles. The strings have been conjured up just to allow potentials to be brought in: without a string to prevent it, a closed surface drawn around a pole will have a net flux of **H** through it, so **H** could not be written as curl **A**; with a preventing string, a vector potential can be produced, except on the string. As the only poles observed so far, giving the static force law[3] gg'/r^2, do effectively sit on ends of strings, these must be given due consideration.

But on the basis that poles could be isolated singularities of **H**, strings are plainly a theoretical artifice that, first of all, destroys rotational symmetry; Dirac's quantization condition, in its fashion, restores it,[4] so that it remains a little unclear whether such a quantization condition might not be a consequence of the phrasing of the problem rather than of the problem itself. For that matter, electromagnetic potentials may be counted an artifice to begin with; and in an extreme action-at-a-distance viewpoint (which in recent years[5] has been shown to be entirely tenable for discussing interacting relativistic particles through Lorentz-covariant equations of motion in terms of particle coordinates only) even the fields are dispensable. Quantum theory, however, requires not merely a dynamics but a Hamiltonian formulation of it—or at least an action principle—in order to have a place for the element of action $\hbar$. So, while the inherited structure of electrodynamics ordinarily answers to this requirement through the use of potentials (which then beget difficulties of gauge invariance—difficulties evidently more of the chosen mathematical language than of physics), in the charge–pole problem the usual machinery is additionally embarrassed.

The issue can be sharply drawn at an elementary level, viz., (a) in nonrelativistic limit where static plus Biot–Savart fields are used and the forces on interacting particles (possessing, in general, both electric and magnetic charge) can be written simply in terms of coordinates and velocities and (b) in the relativistic one-particle problem where a charge–pole moves in the field of a fixed charge–pole. In both cases the classical dynamics in their primitive statements are unambiguous and physically consequential; in neither case can they be formulated through an action principle of familiar type without the artifice of strings on top of potentials.

In the present paper, problems (a) and (b) will be shown to have Hamiltonian formulations growing directly out of the fundamental equations of motion, no potentials or strings being used. This is done through the Lie–Koenigs theorem, which can render substantially any dynamics into Hamiltonian form.

---

[1] P. A. M. Dirac, Proc. Roy. Soc. (London) **A133**, 60 (1931); Phys. Rev. **74**, 817 (1948); J. Schwinger, Phys. Rev. **144**, 1087 (1966); Phys. Rev. **173**, 1536 (1968); see E. Goto, H. Kolm, and K. Ford, Phys. Rev. **132**, 387 (1963); see H. Bradner and W. Isbell, Phys. Rev. **114**, 603 (1959) for a more extensive bibliography.

[2] N. Cabibbo and E. Ferrari [Nuovo Cimento **23**, 1147 (1962)] have used Mandelstam's theory of quantum electrodynamics without potentials [Ann. Phys. **19**, 1 (1962)] to obtain the quantization condition; but here there is a dependence of field operators on spacelike integration paths, and restrictions on commutators due to Jacobi's identity have not been considered (Schwinger, 1966).

[3] C. A. de Coulomb, Mem. de. l'Acad. Roy. des. Sci. **560** *et seq.* (1785).

[4] A. Peres, Phys. Rev. **167**, 1449 (1968).

[5] E. H. Kerner, J. Math. Phys. **6**, 1218 (1965); **9**, 222 (1968); R. N. Hill, J. Math. Phys. **8**, 201, 1756 (1967); D. G. Currie, Phys. Rev. **142**, 817 (1966); E. H. Kerner and R. N. Hill, Phys. Rev. Letters **17**, 1156 (1966). In a different vein, H. Van Dam and E. D. Wigner, Phys. Rev. **138**, B1576 (1965); **142**, 838 (1966).

It is then necessary, however, to give up the identity of physical and canonical coordinates.

This may seem rather objectionable, but it goes to illustrate what is really a very general feature of particle dynamics taken at the most primitive level: $\ddot{\mathbf{r}}_i = \mathbf{F}_i(\mathbf{r}_n, \dot{\mathbf{r}}_n)$. Even when such equations of motion are trimmed to the physical requirements of either Galilean or Lorentz[5] covariance, their scope is very much larger than what any Lagrangian $L(\mathbf{r}_n, \dot{\mathbf{r}}_n)$ and subsequent Hamiltonian $H(\mathbf{r}_n, \mathbf{p}_n)$ ($\mathbf{r}_n$ canonical) will countenance.[6] Does this mean that the wider scope of dynamical laws must simply be cast aside so far as quantum theory is concerned? The Lie–Koenigs theorem and the abandoning of canonical coordinates in general as physical coordinates may at least give hints of a significant way around the conventional limits.

We shall postpone the quantal problems to a subsequent paper, and here simply Hamiltonize problems (a) and (b), producing new integrals of motion and placing the peculiarities of the charge–pole problem in a new light.

EQUATIONS OF MOTION

As was very recently pointed out by Zwanziger,[7] the nonrelativistic limit of pole–charge dynamics has a simple formulation. Let m_1 at $\mathbf{r}_1$, with electric and magnetic charges e_1 and g_1, interact with m_2, e_2, g_2 at $\mathbf{r}_2$. The fields to lowest order in v/c, $\mathbf{r}_{12} \equiv \mathbf{r}_1 - \mathbf{r}_2$, are

$$\mathbf{E}\,(\text{at 1 from 2}) = e_2 \frac{\mathbf{r}_{12}}{r_{12}^3} - \frac{g_2}{c}\frac{\mathbf{v}_2 \times \mathbf{r}_{12}}{r_{12}^3},$$

$$\mathbf{H}\,(\text{at 1 from 2}) = g_2 \frac{\mathbf{r}_{12}}{r_{12}^3} + \frac{e_2}{c}\frac{\mathbf{v}_2 \times \mathbf{r}_{12}}{r_{12}^3},$$

so that the equations of motion are

$$m_1\dot{\mathbf{v}}_1 = e_1\left(\mathbf{E} + \frac{\mathbf{v}_1}{c} \times \mathbf{H}\right) + g_1\left(\mathbf{H} - \frac{\mathbf{v}_1}{c} \times \mathbf{E}\right),$$

etc., or, up to terms of order $v_1 v_2/c^2$,

$$m_1\dot{\mathbf{v}}_1 = (e_1e_2 + g_1g_2)\frac{\mathbf{r}_{12}}{r_{12}^3} + \frac{(e_1g_2 - e_2g_1)}{c}\frac{(\mathbf{v}_1 - \mathbf{v}_2) \times \mathbf{r}_{12}}{r_{12}^3} = -m_2\dot{\mathbf{v}}_2.$$

For a system of charge–poles,

$$m_i\dot{\mathbf{v}}_i = \sum_{j\neq i} f_{ij}\frac{\mathbf{r}_{ij}}{r_{ij}^3} + \frac{g_{ij}}{c}\frac{\mathbf{v}_{ij} \times \mathbf{r}_{ij}}{r_{ij}^3}, \tag{1}$$

$$f_{ij} = f_{ji} \equiv e_ie_j + g_ig_j,$$

$$g_{ij} = -g_{ji} \equiv e_ig_j - e_jg_i,$$

$$\mathbf{v}_{ij} \equiv \mathbf{v}_i - \mathbf{v}_j.$$

[6] See P. Havas [Nuovo Cimento Suppl. **5**, 363 (1957)] on the range of Lagrangian formulation of dynamical problems.

[7] D. Zwanziger, Phys. Rev. **176**, 1480, 1489 (1968).

The dynamics is Galilean-invariant and the Newtonian third law holds. Accordingly, there are the usual linear-momentum and center-of-mass integrals

$$m_i\mathbf{v}_i = \mathbf{P}_0,$$

$$m_i\mathbf{r}_i - \mathbf{P}_0 t = \mathbf{K}_0$$

(summation on repeated indices except where otherwise indicated), and additionally the energy integral

$$\tfrac{1}{2}m_i\mathbf{v}_i^2 + \tfrac{1}{2}f_{ij}/r_{ij} = E$$

and the angular-momentum integral

$$\mathbf{r}_i \times m_i\mathbf{v}_i - \tfrac{1}{2}g_{ij}\hat{\mathbf{r}}_{ij}/c = \mathbf{L}.$$

The latter has been known since Poincaré[8] and faithfully contains the field angular momentum $-g\hat{\mathbf{r}}/c$ of each pair.

In the two-particle case, separation of the center-of-mass and relative motions gives

$$M\dot{\mathbf{V}} = 0,$$

$$m\dot{\mathbf{v}} = f\frac{\hat{\mathbf{r}}}{r^2} + \frac{g}{c}\frac{\mathbf{v} \times \mathbf{r}}{r^3}, \tag{2}$$

$$E_m = \tfrac{1}{2}m\mathbf{v}^2 + f/r, \quad \mathbf{L}_m = \mathbf{r} \times m\mathbf{v} - g\hat{\mathbf{r}}/c,$$

with M and m total and reduced masses, $\mathbf{R}$ and $\mathbf{r}$ central and relative coordinates with corresponding velocities $\mathbf{V}$ and $\mathbf{v}$, f and g net Coulomb and Biot–Savart strengths $e_1e_2 + g_1g_2$, $e_1g_2 - e_2g_1$, and E_m and L_m the conserved internal energy and angular momentum.

The rapid motion of a particle m, e_1, g_1 around a fixed center e_2, g_2 at the origin is controlled by

$$\frac{d}{dt}\frac{m\mathbf{v}}{(1 - v^2/c^2)^{\frac{1}{2}}} = f\frac{\hat{\mathbf{r}}}{r^2} + \frac{g}{c}\frac{\mathbf{v} \times \mathbf{r}}{r^3}$$

to all orders in v/c, or ($\mathbf{I}$ being the unit dyadic)

$$m\dot{\mathbf{v}} = \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}\left(\mathbf{I} - \frac{\mathbf{vv}}{c^2}\right) \cdot \left(f\frac{\hat{\mathbf{r}}}{r^3} + \frac{g}{c}\frac{\mathbf{v} \times \mathbf{r}}{r^3}\right), \tag{3}$$

with

$$E' = \frac{mc^2}{(1 - v^2/c^2)^{\frac{1}{2}}} + \frac{f}{r}, \quad L' = \frac{\mathbf{r} \times m\mathbf{v}}{(1 - v^2/c^2)^{\frac{1}{2}}} - \frac{g}{c}\hat{\mathbf{r}}$$

the conserved energy and angular momentum, respectively.

The nonrelativistic (NR) equation (2) and relativistic (R) equation (3) are virtually the same when f is put equal to zero ($e_1 = 0 = g_2$ or $e_2 = 0 = g_1$, pure-pole–pure-charge interaction). In this case an additional simple integral can be found just by asking

[8] H. Poincaré, Compt. Rend. **123**, 530 (1896).

when $R\mathbf{r} + V\mathbf{v} + S\mathbf{s}$ has vanishing time derivative under the motion $\dot{\mathbf{v}} = \lambda\mathbf{s}$, where $\mathbf{s} \equiv \mathbf{r} \times \mathbf{v}$, $\lambda \equiv -g/mc$ (NR) or $-(g/mc)(1 - v^2/c^2)^{\frac{1}{2}}$ (R), with R, V, and S being scalar and hence functions only of $\mathbf{r}^2$, $\mathbf{v}^2$, and $\mathbf{r} \cdot \mathbf{v}$. Differentiating and setting the coefficients of $\mathbf{r}$, $\mathbf{v}$, and $\mathbf{s}$ to zero gives

$$\begin{aligned} \dot{R} + S(\lambda/r^3)\mathbf{v} \cdot \mathbf{r} &= 0, \\ R + \dot{V} - S\lambda/r &= 0, \\ V\lambda/r^3 + \dot{S} &= 0. \end{aligned} \tag{4}$$

Since $|\mathbf{v}|$ and $|\mathbf{s}|$ are constants of motion, R, V, and S can be taken as functions of them and of $\omega = \hat{\mathbf{v}} \cdot \hat{\mathbf{r}}$, whereupon

$$\frac{d}{dt} = \frac{d\omega}{dt}\frac{d}{d\omega} = \frac{v}{r}(1 - \omega^2)\frac{d}{d\omega} = \frac{v^2}{s}(1 - \omega^2)^{\frac{3}{2}}\frac{d}{d\omega}$$

[using $s = rv(1 - \omega^2)^{\frac{1}{2}}$]. Equations (4) become (primes meaning differentiation with respect to ω)

$$R' + S\lambda\frac{v}{s}\frac{\omega}{(1 - \omega^2)^{\frac{1}{2}}} = 0,$$

$$R + \frac{v^2}{s}(1 - \omega^2)^{\frac{3}{2}}V' - S\lambda\frac{v}{s}(1 - \omega^2)^{\frac{3}{2}} = 0,$$

$$V\lambda\frac{v}{s^2} + S' = 0.$$

To decouple, we differentiate the second and introduce R' and S' from the first and third; this yields

$$(1 - \omega^2)V'' - 3\omega V' + V\lambda^2/s^2 = 0. \tag{5}$$

The change of variables $V = F/(1 - \omega^2)^{\frac{1}{2}}$, $\omega = \cos\psi$, brings this to

$$\frac{d^2F}{d\psi^2} + \left(1 + \frac{\lambda^2}{s^2}\right)F = 0,$$

so that

$$V = \begin{Bmatrix}\sin \\ \cos\end{Bmatrix}\left(1 + \frac{\lambda^2}{s^2}\right)^{\frac{1}{2}}\psi/\sin\psi \tag{6}$$

and R and S can be found from quadrature and differentiation from

$$\begin{aligned} \frac{dS}{d\psi} &= V\lambda\frac{v}{s^2}\sin\psi, \\ R &= S\lambda\frac{v}{s}\sin\psi + \frac{dV}{d\psi}\frac{v^2}{s}\sin^2\psi. \end{aligned} \tag{7}$$

The peculiar-looking integral $\mathbf{\Pi} \equiv R\mathbf{r} + V\mathbf{v} + S\mathbf{s}$ just computed has, from (6) and (7), three parameters of integration that are arbitrary functions of s and v. Hence, through a factor s^2 in V, $\mathbf{\Pi}$ can be made regular at $\psi = 0$ or π ($\mathbf{r}$ parallel or opposite to $\mathbf{v}$). But $\mathbf{\Pi}$ is sensitive to an alteration of ψ by a multiple of 2π, i.e., to physically equivalent representations of the relative orientation of $\mathbf{r}$ and $\mathbf{v}$. This oversensitivity is removed when $(1 + \lambda^2/s^2)^{\frac{1}{2}}$ is an integer, i.e.,

$$\lambda^2/s^2 = \nu(\nu + 2), \quad \nu = 0, 1, \cdots,$$

or

$$\frac{g}{c} \equiv \frac{e_1 g_2}{c} = \pm[\nu(\nu + 2)]^{\frac{1}{2}}|\mathbf{r} \times \mathbf{p}_0|, \tag{8}$$

$$\mathbf{p}_0 \equiv m\mathbf{v}(\text{NR}) \quad \text{or} \quad \frac{m\mathbf{v}}{(1 - v^2/c^2)^{\frac{1}{2}}}(\text{R}).$$

This classical eigencondition also assures regular solutions of Eq. (5) as Gegenbauer polynomials,[9] $V = G_\nu(\omega)$, $G_0 = 1$, $G_1 = 2\omega$, $G_2 = 4\omega^2 - 1$, $G_3 = 8\omega^3 - 4\omega$, etc., the close cousins of Legendre polynomials, and then R and S are readily computed in terms of G_ν, $G_{\nu+1}$, $G_{\nu-1}$. With this classical restriction on g/c, the conserved angular momentum has the magnitude

$$|\mathbf{L}| = (\nu + 1)|\mathbf{r} \times \mathbf{p}_0|,$$

allowing $|\mathbf{L}| = 0$, which otherwise is ruled out.

It would seem that $\mathbf{\Pi}$ is a species of Runge–Lenz vector, but its character is unclear in view of the nature of the motion, which quite generally is a geodesic on a cone whose axis is $\mathbf{L}$ with apex angle $\cos^{-1}(g/c\,|\mathbf{L}|)$, the particle winding in on a spiral track to a point s/v from the center of force at the apex and then winding out on a similar track. In particular, a relationship between interaction strength g/c and conserved $|\mathbf{r} \times \mathbf{p}_0|$ appears unaccountable.

LIE–KOENIGS THEORY

This has been discussed previously in the context of instantaneous, action-at-a-distance relativistic particle dynamics,[5] but it will be summarized here for completeness. Conventionally,[10] the starting point is an arbitrary integral-invariant of first order of the differential system $\dot{y}_i = Y_i(y)$ [a generic format for any dynamics $\ddot{\mathbf{r}}_i = \mathbf{F}_i(\mathbf{r}_n, \dot{\mathbf{r}}_n)$ when written as $\dot{\mathbf{r}}_i = \mathbf{v}_i$, $\dot{\mathbf{v}}_i = \mathbf{F}_i(\mathbf{r}_n, \mathbf{v}_n)$]. However, as can be seen below, not just any integral invariant will do.

Let us simply ask for a variational principle

$$\delta\int\{U_i(y)\dot{y}_i - H(y)\}\,dt = 0, \tag{9}$$

whose extremals are to be defined by the prescribed $\dot{y}_i = Y_i$. All y's are to be independently varied,

[9] W. Magnus, F. Oberhettinger, and R. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin, 1966).

[10] E. T. Whittaker, *Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, New York, 1960).

subject to fixed endpoint conditions. The Euler–Lagrange equations are

$$\left(\frac{\partial U_m}{\partial y_i} - \frac{\partial U_i}{\partial y_m}\right)\dot{y}_i = -\frac{\partial H}{\partial y_m}$$

or

$$\Gamma_{mi}\dot{y}_i = -\frac{\partial H}{\partial y_m}, \tag{10}$$

and $\dot{y} = -(\Gamma^{-1})\partial H/\partial y$ is to reproduce the given Y. As a system of partial differential equations for U and H this requirement can always be met at least in the local Cauchy–Kowalewski sense where Y is regular (Γ is antisymmetric and is necessarily singular for odd-order systems, but particle dynamics is always of even order and a local inverse for Γ is then possible).

In vector notation, the action integrand is

$$đ\phi \equiv \mathbf{U}_i \cdot d\mathbf{r}_i + \mathbf{V}_i \cdot d\mathbf{v}_i - H\,dt \tag{11}$$

($\mathbf{U}$, $\mathbf{V}$, and H functions of all $\mathbf{r}$ and $\mathbf{v}$), which is always reducible (Pfaff's problem) to

$$đ\phi \equiv \mathbf{P}_i \cdot d\mathbf{Q}_i - H\,dt$$
$$(\mathbf{P}_i(\mathbf{r}, \mathbf{v});\ \mathbf{Q}_i(\mathbf{r}, \mathbf{v});\ H(\mathbf{r}(P, Q), \mathbf{v}(P, Q))), \tag{12}$$

whereupon the starting dynamics $\ddot{\mathbf{r}}_i = \mathbf{F}_i$ is Hamiltonized.

It may be noted that the entire Hamiltonian apparatus is already inherent in Eq. (11), without solving the often difficult Pfaff's problem down to Eq. (12) and producing explicit $\mathbf{Q}$ and $\mathbf{P}$. For instance, Poisson brackets can be expressed in terms of the primitive coordinates and velocities y through

$$(A, B) = -\frac{\partial A}{\partial y_k}\,\Gamma^{-1}_{kl}\,\frac{\partial B}{\partial y_l};$$

canonical transformation theory works through addition of an exact differential $dW(y)$ to $đ\phi$; the representation of a group of infinitesimal symmetry transformations as canonical transformations may be arranged by requiring $đ\phi$ to change by an exact differential under the transformations; and Noether's theorem giving the conservation laws going with the symmetries is readily obtained.

CHARGE–POLE CANONICAL VARIABLES

Let us develop Eq. (2) on this ground. Then the Hamiltonian formulation of Eqs. (1) and (3) will be straightforward.

A moment's calculation shows first of all that no Lagrangian $L(\mathbf{r}^2, \dot{\mathbf{r}}^2, \mathbf{r}\cdot\dot{\mathbf{r}})$, which would permit $\mathbf{r}$ to be a canonical coordinate, can embrace the equation of motion (2). So we look for an action principle of type (9),

$$đ\phi = \mathbf{U}(\mathbf{r}, \mathbf{v}) \cdot d\mathbf{r} + \mathbf{V}(\mathbf{r}, \mathbf{v}) \cdot d\mathbf{v} - H\,dt.$$

Helpfully, we know what H_m (called E_m above) and the angular momentum $\mathbf{L}_m$ are. The rotational invariance is already incorporated in the designation of $\mathbf{U}$ and $\mathbf{V}$ as vectors and of H_m as a scalar. Accordingly, the application of Noether's theorem gives

$$\mathbf{r}\times\mathbf{U} + \mathbf{v}\times\mathbf{V} = \mathbf{L}_m = -\frac{g}{c}\frac{\mathbf{r}}{r} + m\mathbf{s}.$$

Taking $\mathbf{U}$ most generally as $\mathbf{U} = U_r\mathbf{r} + U_v\mathbf{v} + U_s\mathbf{s}$ and similarly for $\mathbf{V}$ with scalar coefficients U_α and V_α that can depend only on $\mathbf{r}^2$, $\mathbf{v}^2$, and $\mathbf{v}\cdot\mathbf{r}$, the last equation breaks down into

$$U_v - V_r = m,$$
$$U_s\mathbf{v}\cdot\mathbf{r} + V_s\mathbf{v}^2 = -g/cr,$$
$$U_s\mathbf{r}^2 + V_s\mathbf{v}\cdot\mathbf{r} = 0,$$

so that

$$V_s = -\frac{g}{c}\frac{r}{s^2}, \quad U_s = \frac{g}{c}\frac{\mathbf{v}\cdot\mathbf{r}}{s^2},$$

with

$$\mathbf{s}^2 = (\mathbf{r}\times\mathbf{v})^2 = \mathbf{v}^2\mathbf{r}^2 - (\mathbf{v}\cdot\mathbf{r})^2,$$

as before. In addition, it is strongly hinted that it will suffice to take $U_v = m$, $U_r = 0$, $V_r = 0$, and $V_v = 0$, making in all

$$\mathbf{U} = m\mathbf{v} + \frac{g}{c}\frac{\mathbf{v}\cdot\mathbf{r}(\mathbf{r}\times\mathbf{v})}{r(\mathbf{r}\times\mathbf{v})^2},$$

$$\mathbf{V} = -\frac{g}{c}\frac{r(\mathbf{r}\times\mathbf{v})}{(\mathbf{r}\times\mathbf{v})^2}, \tag{13}$$

$$H_m = \tfrac{1}{2}m\mathbf{v}^2 + f/r.$$

It must be seen now whether the surmise yields up the dynamics

$$\dot{\mathbf{r}} = \mathbf{v},$$

$$m\dot{\mathbf{v}} = f\frac{\mathbf{r}}{r^3} - \frac{g}{c}\frac{\mathbf{r}\times\mathbf{v}}{r^3} \tag{14}$$

as the Euler–Lagrange equations (10), which now read

$$\begin{pmatrix} \left(\frac{\partial \mathbf{U}}{\partial \mathbf{r}}\right)^{\mathrm{T}} - \frac{\partial \mathbf{U}}{\partial \mathbf{r}}, & \left(\frac{\partial \mathbf{U}}{\partial \mathbf{v}}\right)^{\mathrm{T}} - \frac{\partial \mathbf{V}}{\partial \mathbf{r}} \\ \left(\frac{\partial \mathbf{V}}{\partial \mathbf{r}}\right)^{\mathrm{T}} - \frac{\partial \mathbf{U}}{\partial \mathbf{v}}, & \left(\frac{\partial \mathbf{V}}{\partial \mathbf{v}}\right)^{\mathrm{T}} - \frac{\partial \mathbf{V}}{\partial \mathbf{v}} \end{pmatrix} \begin{pmatrix} \dot{\mathbf{r}} \\ \dot{\mathbf{v}} \end{pmatrix} = -\begin{pmatrix} \frac{\partial H}{\partial \mathbf{r}} \\ \frac{\partial H}{\partial \mathbf{v}} \end{pmatrix}. \tag{15}$$

Here $\partial\mathbf{U}/\partial\mathbf{r}$, for instance, means the dyadic

$$\frac{\partial\mathbf{U}}{\partial\mathbf{r}} = \frac{\partial}{\partial\mathbf{r}}(U_r\mathbf{r} + U_v\mathbf{v} + U_s\mathbf{s}) = U_r\mathbf{I} + \frac{\partial U_r}{\partial\mathbf{r}}\mathbf{r} + \cdots,$$

and $(\partial\mathbf{U}/\partial\mathbf{r})^{\mathrm{T}}$ its transpose.

A calculation, in which it is useful to write the unit dyadic $\mathbf{I}$ as $\hat{\mathbf{v}}\hat{\mathbf{v}} + \hat{\mathbf{s}}\hat{\mathbf{s}} + \hat{\mathbf{w}}\hat{\mathbf{w}}$ ($\mathbf{w} = \mathbf{v} \times \mathbf{s}$), gives

$$\left(\frac{\partial\mathbf{U}}{\partial\mathbf{r}}\right)^{\mathrm{T}} - \frac{\partial\mathbf{U}}{\partial\mathbf{r}} = \frac{g}{c}\frac{\mathbf{r}\times\mathbf{I}}{r^3},$$

$$\left(\frac{\partial\mathbf{U}}{\partial\mathbf{v}}\right)^{\mathrm{T}} - \frac{\partial\mathbf{V}}{\partial\mathbf{r}} = m\mathbf{I} = -\left[\left(\frac{\partial\mathbf{V}}{\partial\mathbf{r}}\right)^{\mathrm{T}} - \frac{\partial\mathbf{U}}{\partial\mathbf{v}}\right],$$

$$\left(\frac{\partial\mathbf{V}}{\partial\mathbf{v}}\right)^{\mathrm{T}} - \frac{\partial\mathbf{V}}{\partial\mathbf{v}} = 0,$$

so that Eqs. (15) do exactly comprise Eqs. (14).

Now, to construct canonical variables we have to compress $đ\varphi = \mathbf{U}\cdot d\mathbf{r} + \mathbf{V}\cdot d\mathbf{v}$ by half into $\mathbf{P}\cdot d\mathbf{Q}$ (or into $\mathbf{P}\cdot d\mathbf{Q}$ plus an exact differential, the different solutions to Pfaff's problem being canonical transforms of one another). It is natural to group the g parts of $d\varphi$ together, writing

$$đ\varphi = m\mathbf{v}\cdot d\mathbf{r} + (\mathbf{V}\cdot d\mathbf{v} - \mathbf{V}(\mathbf{v}\cdot\mathbf{r}/r^2)\cdot d\mathbf{r}).$$

Noting that

$$\left(\frac{\mathbf{v}\cdot\mathbf{r}}{r^2}\right)d\mathbf{r} = d\left(\mathbf{r}\frac{\mathbf{v}\cdot\mathbf{r}}{r^2}\right) - \mathbf{r}\,d\left(\frac{\mathbf{v}\cdot\mathbf{r}}{r^2}\right)$$

and $\mathbf{V}\cdot\mathbf{r} = 0$, we have

$$đ\varphi = m\mathbf{v}\cdot d\mathbf{r} + \mathbf{V}\cdot d(\mathbf{v} - \mathbf{r}\mathbf{v}\cdot\mathbf{r}/r^2) = m\mathbf{v}\cdot d\mathbf{r} - \mathbf{V}\cdot d\mathbf{t}/r^2,$$

where

$$\mathbf{t} \equiv \mathbf{r}\times(\mathbf{r}\times\mathbf{v}).$$

Place $\mathbf{v} = (\mathbf{r}\,\mathbf{v}\cdot\mathbf{r} - \mathbf{t})/r^2$ into the first member of $đ\varphi$ and use $\mathbf{t}\cdot d\mathbf{r} = -\mathbf{r}\cdot d\mathbf{t}$ and $\mathbf{V}\cdot d\mathbf{t}/r^2 = (\mathbf{V}/r^2)\cdot d\mathbf{t}$, getting

$$đ\varphi = \frac{m\mathbf{v}\cdot\mathbf{r}}{r^2}\mathbf{r}\cdot d\mathbf{r} + \frac{1}{r^2}\left[m\mathbf{r} + \frac{g}{c}\frac{r}{s^2}\mathbf{s}\right]\cdot d\mathbf{t}.$$

The vector in square brackets—call it $\mathbf{u}$—is orthogonal to $\mathbf{t}$, and this simplifies the finding of $\mathbf{P}$ and $\mathbf{Q}$; namely, if $\mathbf{P}$ is set equal to $E\mathbf{t} + F\mathbf{u}$ and $\mathbf{Q}$ to $A\mathbf{t} + B\mathbf{u}$, then

$$\mathbf{P}\cdot d\mathbf{Q} = EA\mathbf{t}\cdot d\mathbf{t} + FB\mathbf{u}\cdot d\mathbf{u} + Et^2\,dA + Fu^2\,dB + EB\mathbf{t}\cdot d\mathbf{u} + FA\mathbf{u}\cdot d\mathbf{t},$$

and this can easily be shaped into $đ\varphi$. A simple choice is $A = r/t = 1/s$, $B = 0$, $E = m\mathbf{v}\cdot\mathbf{r}/r^2 s$, $F = s/r^2$, or

$$\mathbf{Q} = \frac{\mathbf{t}}{s} = \frac{\mathbf{r}\times(\mathbf{r}\times\mathbf{v})}{|\mathbf{r}\times\mathbf{v}|},$$

$$\mathbf{P} = \frac{m\mathbf{v}\cdot\mathbf{r}}{r^2 s}\mathbf{t} + \frac{s}{r^2}\mathbf{u} = \frac{m\mathbf{v}\cdot\mathbf{r}}{r^2|\mathbf{r}\times\mathbf{v}|}\mathbf{r}\times(\mathbf{r}\times\mathbf{v}) + \frac{m|\mathbf{r}\times\mathbf{v}|}{r^2}\mathbf{r} + \frac{g}{c}\frac{1}{r|\mathbf{r}\times\mathbf{v}|}(\mathbf{r}\times\mathbf{v}). \quad (16)$$

The inverse transformation works out as

$$\mathbf{r} = \frac{Q^2 S}{|\mathbf{Q}\times\mathbf{P}|^2}\left(\frac{\mathbf{Q}\times(\mathbf{P}\times\mathbf{Q})}{Q^2} + \frac{g}{c}\frac{\mathbf{P}\times\mathbf{Q}}{QS}\right),$$

$$\mathbf{v} = \frac{\mathbf{P}\cdot Q S}{m|\mathbf{Q}\times\mathbf{P}|^2}\left[\frac{\mathbf{P}\times(\mathbf{P}\times\mathbf{Q})}{\mathbf{P}\cdot\mathbf{Q}} + \frac{g}{c}\frac{\mathbf{P}\times\mathbf{Q}}{QS}\right],$$

$$S \equiv [(\mathbf{Q}\times\mathbf{P})^2 - g^2/c^2]^{\frac{1}{2}}.$$

While every point in $\mathbf{r}$, $\mathbf{v}$ maps into a point in $\mathbf{Q}$, $\mathbf{P}$ space, the converse is not true: the physically relevant part of phase space is restricted by $|\mathbf{Q}\times\mathbf{P}| \geq g/c$ in general. With a classical eigencondition of type (8), such a restriction would not come up.

This set of canonical variables looks even more strange because, for $g = 0$, $\mathbf{Q}$ and $\mathbf{P}$ do not go over into $\mathbf{r}$ and $m\mathbf{v}$. But this is only a consequence of the above particular arrangement for $\mathbf{Q}$, $\mathbf{P}$ and is removable by canonical transformation. In fact, for $g = 0$, $\mathbf{P}\cdot d\mathbf{Q} = m\mathbf{v}\cdot d\mathbf{r}$. Then, quite generally from above,

$$\boldsymbol{\rho} = \frac{\mathbf{Q}\times(\mathbf{P}\times\mathbf{Q})}{|\mathbf{P}\times\mathbf{Q}|},$$

$$\boldsymbol{\pi} = \frac{\mathbf{P}\times(\mathbf{P}\times\mathbf{Q})}{|\mathbf{P}\times\mathbf{Q}|}$$

must be a cononical transformation, which reads, in terms of $\mathbf{r}$ and $\mathbf{v}$,

$$\boldsymbol{\rho} = \frac{1}{(m^2s^2 + g^2/c^2)^{\frac{1}{2}}}\left(ms\mathbf{r} + \frac{g}{c}\frac{r}{s}\mathbf{s}\right),$$

$$\boldsymbol{\pi} = \frac{1}{(m^2s^2 + g^2/c^2)^{\frac{1}{2}}} \times \left(\frac{(m^2s^2 + g^2/c^2)}{s}\mathbf{v} - \frac{g^2}{c^2}\frac{\mathbf{v}\cdot\mathbf{r}}{r^2 s}\mathbf{r} + \frac{g}{c}\frac{m\mathbf{v}\cdot\mathbf{r}}{rs}\mathbf{s}\right)$$

and now $\boldsymbol{\rho}$ and $\boldsymbol{\pi}$ show explicitly how $\mathbf{r}$ and $m\mathbf{v}$ can be recovered as canonical variables when $g \to 0$. The inverse transformation here is

$$\mathbf{r} = \frac{\rho^2 S'}{(\boldsymbol{\rho}\times\boldsymbol{\pi})^2}\left(\boldsymbol{\rho}\frac{|\boldsymbol{\rho}\times\boldsymbol{\pi}|}{\rho^2} - \frac{g}{c}\frac{\boldsymbol{\rho}\times\boldsymbol{\pi}}{\rho S'}\right),$$

$$\mathbf{v} = \frac{\boldsymbol{\pi}\cdot\boldsymbol{\rho}S'}{m(\boldsymbol{\rho}\times\boldsymbol{\pi})^2}\left(\boldsymbol{\pi}\frac{|\boldsymbol{\rho}\times\boldsymbol{\pi}|}{\boldsymbol{\rho}\cdot\boldsymbol{\pi}} - \frac{g}{c}\frac{\boldsymbol{\rho}\times\boldsymbol{\pi}}{\rho S'}\right),$$

$$S' \equiv [(\boldsymbol{\rho}\times\boldsymbol{\pi})^2 - g^2/c^2]^{\frac{1}{2}}.$$

Returning to the simpler $\mathbf{Q}$ and $\mathbf{P}$, it is easy to verify, as must be the case, that $\mathbf{Q} \times \mathbf{P}$ is $m\mathbf{s} - g\hat{\mathbf{r}}/c$, the conserved $\mathbf{L}_m$, and that

$$Q^2 = r^2,$$
$$P^2 = m^2v^2 + \frac{g^2}{c^2}\frac{1}{r^2},$$

whence $H_m(\mathbf{Q}, \mathbf{P})$ is.

$$H_m = \frac{\mathbf{P}^2}{2m} + \frac{f}{Q} - \frac{g^2}{2mc^2}\frac{1}{Q^2}.$$

The motion in $\mathbf{Q}$ space is simply that of a point under Coulomb plus attractive inverse-cube forces, subject, however, to $|\mathbf{Q} \times \mathbf{P}| \geq g/c$.

Quite apart from Case's[11] well-known treatment of the Schrödinger dynamics of singular $1/Q^2$ potentials, the simplicity of Bohr–Sommerfeld quantization is noticeable. For $f = 0$ and circular orbits (the only closed ones)

$$m\frac{\dot{Q}^2}{Q} = \frac{g^2}{mc^2}\frac{1}{Q^3},$$
$$mQ\dot{Q} = n\hbar$$

at energies

$$E_m = \tfrac{1}{2}m\dot{\mathbf{Q}}^2 - \frac{1}{2}\frac{g^2}{mc^2}\frac{1}{Q^2}$$

that are zero for all orbits. Thus we arrive at the Schwinger result painlessly: $g/c = e_1g_2/c$ or $e_2g_1/c = n\hbar$. This reckoning, of course, bypasses the fundamental issue of the standing of canonical commutation rules in the $\mathbf{Q}$, $\mathbf{P}$ language (which, besides being restricted by $|\mathbf{Q} \times \mathbf{P}| \geq g/c$, is indeterminate up to arbitrary canonical transformation). It is only a bit of "quantum statics" (an evaluation of L_m and E_m) and not dynamics (time evolution of state vectors).

Bohr–Sommerfeld statics with an attractive Coulomb potential $f/Q = -f_0/Q$, $f_0 > 0$ gives quantized orbital radii Q_n,

$$n^2\hbar^2 = \frac{g^2}{c^2} + f_0mQ_n,$$

and energies

$$E_m(n) = -\frac{1}{2}\frac{mf_0^2}{n^2\hbar^2 - g^2/c^2},$$

$n^2\hbar^2 > g^2/c^2$. So far as these orbits go, g is unrestricted.

When, therefore, the Biot–Savart force rules the motion, a Dirac-type condition $g/\hbar c = n$ is clearly in the offing. But let additional Coulomb forces (or presumably other binding forces which—at least nonrelativistically—can be fitted into the above framework) act, and such a condition appears less than compelling. For particles which are electromagnetically pure poles but which have nonelectromagnetic interactions as well, the issue of a charge–pole quantization condition would then seem to hinge on the detailed nature of such interactions.

The preceding results transfer over to the one-particle relativistic case, Eq. (3), without trouble now. In Eq. (13) just replace $m\mathbf{v}$ in $\mathbf{U}$ by $m\mathbf{v}/(1 - v^2/c^2)^{\frac{1}{2}}$, with a corresponding adjustment in the evaluation of $\partial\mathbf{U}/\partial\mathbf{v} - \partial\mathbf{V}/\partial\mathbf{r}$ that makes Eq. (15) yield $\dot{\mathbf{r}} = \mathbf{v}$ and Eq. (3) as Euler–Lagrange equations. The solution to Pfaff's problems and the representation of $\mathbf{Q}$, $\mathbf{P}$ in Eq. (16) are as above with m replaced by $m/(1 - v^2/c^2)^{\frac{1}{2}}$. Again $Q^2 = r^2$, but now

$$P^2 = \frac{m^2v^2}{1 - v^2/c^2} + \frac{g^2}{c^2}\frac{1}{r^2}$$

so that the fundamental conserved quantities are

$$H' = \frac{f}{Q} + (m^2c^4 + c^2\mathbf{P}^2 - g^2/\mathbf{Q}^2)^{\frac{1}{2}},$$
$$\mathbf{L}' = \mathbf{Q} \times \mathbf{P}. \tag{17}$$

The equations of motion $\dot{\mathbf{Q}} = \partial H/\partial\mathbf{P}$ and $\dot{\mathbf{P}} = -\partial H/\partial\mathbf{Q}$ are, for $f = 0$,

$$\mathbf{P} = \Omega m\dot{\mathbf{Q}},$$
$$\dot{\mathbf{P}} = -\frac{g^2}{E_0Q^3}\Omega\hat{\mathbf{Q}},$$
$$E_0 \equiv mc^2, \quad \Omega \equiv \left(\frac{1 - g^2/E_0^2Q^2}{1 - \dot{Q}^2/c^2}\right)^{\frac{1}{2}}.$$

In Bohr–Sommerfeld statics, for circular orbits ($\ddot{\mathbf{Q}} = -\dot{Q}^2\hat{\mathbf{Q}}/Q$), we have

$$m\dot{Q}^2/Q = g^2/E_0Q^3\Omega^2,$$
$$mQ\dot{Q}\Omega = n\hbar,$$

giving $g/c = n\hbar$ again, while $H' = E_0\Omega$ remains unfixed. There is no classically apparent reason why $\dot{Q}$ may not exceed c and Q fall below g/E_0.

Including the Coulomb potential $-f_0/Q$ and keeping the sign ambiguity in the square root in Eq. (17), for circular orbits one gets

$$Q_n^2 = a_0^2(n^2 - \beta^2)(n^2 - \alpha^2 - \beta^2),$$
$$\frac{\dot{Q}_n}{c} = \pm\frac{n\alpha}{n^2 - \beta^2},$$
$$\frac{H'_n}{mc^2} = \pm\left(1 - \frac{\alpha^2}{n^2 - \beta^2}\right)^{\frac{1}{2}},$$

[11] K. M. Case, Phys. Rev. **80**, 797 (1950).

where a_0 is the Bohr-like radius $\hbar^2/mf_0$ and α and β are Coulomb and Biot–Savart fine-structure constants $|e_1e_1 + g_1g_1|/\hbar c$ and $|e_1g_2 - e_2g_1|/\hbar c$. There are two formally possible branches of motion: $n \lesssim \beta$ and $n \geq (\alpha^2 + \beta^2)^{\frac{1}{2}} \equiv \gamma$, going with minus and plus signs above, with $\beta < n < \gamma$ forbidden by $Q_n^2 \geq 0$. No basic restrictions on α and β are in sight. The second "normal" branch can begin with $Q_n = 0$ if γ is integral (n'), when central and orbital particles sit on top of one another at zero energy and orbital speed $\dot{Q} = cn'/\alpha$ that can exceed c; beyond that, the orbits expand and the energy approaches mc^2. The first branch consists in a series of shrinking orbits and sinking energies, including another null orbit at infinite speed $\dot{Q}$ and infinite negative energy if β is integral. Since $|\mathbf{Q} \times \mathbf{P}|$ is supposed to be $\geq \beta\hbar, n \geq \beta$, this null orbit alone in the first branch is formally allowed. Under just the restriction $Q_n^2 \geq 0$ (forgetting $|\mathbf{Q} \times \mathbf{P}|$ restrictions), the earlier members $n < \beta$ (if any) of this first branch are also formally allowed. The fuller treatment, considering the extractions of the square root in Eq. (17) and the possibly large value of α and other matters, will be left for another paper.

CONCLUSION

Finally, let us turn to the N-particle nonrelativistic problem (1).

From the action differential

$$đ\phi = \mathbf{V}_i \cdot d\mathbf{r}_i + \mathbf{V}_i \cdot d\mathbf{v}_i - H\,dt = đ\varphi - H\,dt,$$

Noether's theorem will give, for translationally and rotationally invariant $\mathbf{U}$, $\mathbf{V}$, and H,

$$\mathbf{P}_0 = \sum \mathbf{U}_i,$$

$$\mathbf{r}_i \times \mathbf{U}_i + \mathbf{v}_i \times \mathbf{V}_i = \mathbf{L},$$

and from the treatment above of Eq. (2) the step

$$\mathbf{U}_i = m_i\mathbf{v}_i + \sum_{j\neq i} \frac{g_{ij}}{c} \frac{\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{(\mathbf{r}_{ij} \times \mathbf{v}_{ij})^2 r_{ij}} \mathbf{r}_{ij} \times \mathbf{v}_{ij},$$

$$\mathbf{V}_i = -\sum_{j\neq i} \frac{g_{ij}}{c} \frac{r_{ij}}{(\mathbf{r}_{ij} \times \mathbf{v}_{ij})^2} \mathbf{r}_{ij} \times \mathbf{v}_{ij}$$

is visibly in place. The Euler–Lagrange equations that are to come out as Eq. (1) are

$$\left[\left(\frac{\partial \mathbf{U}_i}{\partial \mathbf{r}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{U}_l}{\partial \mathbf{r}_i}\right] \cdot \dot{\mathbf{r}}_l + \left[\left(\frac{\partial \mathbf{U}_i}{\partial \mathbf{v}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{V}_l}{\partial \mathbf{r}_i}\right] \cdot \dot{\mathbf{v}}_l = -\frac{\partial H}{\partial \mathbf{r}_i},$$

$$\left[\left(\frac{\partial \mathbf{V}_i}{\partial \mathbf{r}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{U}_l}{\partial \mathbf{v}_i}\right] \cdot \dot{\mathbf{r}}_l + \left[\left(\frac{\partial \mathbf{V}_i}{\partial \mathbf{v}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{V}_l}{\partial \mathbf{v}_i}\right] \cdot \dot{\mathbf{v}}_l = -\frac{\partial H}{\partial \mathbf{v}_i}.$$

Direct computation gives

$$\left(\frac{\partial \mathbf{U}_i}{\partial \mathbf{r}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{U}_l}{\partial \mathbf{r}_i} = \frac{g_{il}}{c} \frac{\mathbf{r}_{li} \times \mathbf{I}}{r_{li}^3}, \quad l \neq i,$$

$$\left(\frac{\partial \mathbf{U}_i}{\partial \mathbf{r}_i}\right)^{\mathrm{T}} - \left(\frac{\partial \mathbf{V}_i}{\partial \mathbf{r}_i}\right)^{\mathrm{T}} = \sum_{j\neq i} \frac{g_{ij}}{c} \frac{\mathbf{r}_{ij} \times \mathbf{I}}{r_{ij}^3},$$

$$\left(\frac{\partial \mathbf{U}_i}{\partial \mathbf{v}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{V}_l}{\partial \mathbf{r}_i} = m_i\delta_{il}\mathbf{I} = -\left[\left(\frac{\partial \mathbf{V}_i}{\partial \mathbf{r}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{U}_l}{\partial \mathbf{v}_i}\right],$$

$$\left(\frac{\partial \mathbf{V}_i}{\partial \mathbf{v}_l}\right)^{\mathrm{T}} - \frac{\partial \mathbf{V}_l}{\partial \mathbf{v}_i} = 0$$

(no sums on repeated indices here); this does make the expected equations of motion materialize correctly.

Next, $đ\varphi$ can be written

$$đ\varphi = m_l\mathbf{v}_l \cdot d\frac{m_k\mathbf{r}_k}{M} + \frac{1}{2}\frac{m_im_j}{M}\mathbf{v}_{ij} \cdot d\mathbf{r}_{ij} + \frac{1}{2}\frac{g_{ij}}{c}\frac{1}{s_{ij}^2 r_{ij}}\mathbf{s}_{ij} \cdot d\mathbf{t}_{ij},$$

where

$$\mathbf{s}_{ij} \equiv \mathbf{r}_{ij} \times \mathbf{v}_{ij}, \quad \mathbf{t}_{ij} \equiv \mathbf{r}_{ij} \times (\mathbf{r}_{ij} \times \mathbf{v}_{ij}),$$

or, through $\mathbf{v} = (\mathbf{r}\mathbf{v} \cdot \mathbf{r} - \mathbf{t})/r^2$ as before,

$$đ\varphi = \mathbf{P}_0 \cdot d\mathbf{Q}_0 + \frac{1}{2}\frac{m_im_j}{M}\frac{\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{r_{ij}^2}\mathbf{r}_{ij} \cdot d\mathbf{r}_{ij} + \frac{1}{2}\left[\frac{m_im_j}{M}\frac{\mathbf{r}_{ij}}{r_{ij}^2} + \frac{g_{ij}}{c}\frac{\mathbf{s}_{ij}}{s_{ij}^2 r_{ij}}\right] \cdot d\mathbf{t}_{ij},$$

where

$$\mathbf{P}_0 \equiv m_l\mathbf{v}_l, \quad \mathbf{Q}_0 = m_k\mathbf{r}_k/M,$$

and thence to

$$đ\varphi = \mathbf{P}_0 \cdot d\mathbf{Q}_0 + \tfrac{1}{2}\mathbf{P}_{ij} \cdot d\mathbf{Q}_{ij}$$

with

$$\mathbf{Q}_{ij} = \mathbf{r}_{ij} \times (\mathbf{r}_{ij} \times \mathbf{v}_{ij})/s_{ij},$$

$$\mathbf{P}_{ij} = \frac{m_im_j}{M}\frac{\mathbf{r}_{ij} \cdot \mathbf{v}_{ij}}{r_{ij}^2 s_{ij}}\mathbf{r}_{ij} \times (\mathbf{r}_{ij} \times \mathbf{v}_{ij}) + \frac{m_im_j}{M}\frac{s_{ij}}{r_{ij}^2}\mathbf{r}_{ij} + \frac{g_{ij}}{c}\frac{\mathbf{s}_{ij}}{s_{ij}r_{ij}}.$$

The reduction is not quite complete as $đ\varphi$ contains all $N(N-1)/2$ differentials of the "relative canonical coordinates" $\mathbf{Q}_{ij}$. But $\mathbf{P}_{ij}$ and $\mathbf{Q}_{ij}$ are not all independent. If $\mathbf{r}_a$ and $\mathbf{v}_a$ are some selected coordinate–velocity pair, then $\mathbf{Q}_{a\gamma}$ and $\mathbf{P}_{a\gamma}$ ($\gamma = 1, 2, \cdots, a-1, a+1, \cdots, N$) are a complete independent set of relative $\mathbf{Q}$'s and $\mathbf{P}$'s, and all others can be expressed in terms of them; i.e.,

$$\mathbf{r}_{nm}(\mathbf{Q}_{nm}, \mathbf{P}_{nm}) = \mathbf{r}_{am}(\mathbf{Q}_{am}, \mathbf{P}_{am}) - \mathbf{r}_{an}(\mathbf{Q}_{an}, \mathbf{P}_{an}),$$

$$\mathbf{v}_{nm}(\mathbf{Q}_{nm}, \mathbf{P}_{nm}) = \mathbf{v}_{am}(\mathbf{Q}_{am}, \mathbf{P}_{am}) - \mathbf{v}_{an}(\mathbf{Q}_{an}, \mathbf{P}_{an})$$

give

$$\mathbf{Q}_{nm}(\mathbf{Q}_{am}, \mathbf{P}_{am}, \mathbf{Q}_{an}, \mathbf{P}_{an})$$

and

$$\mathbf{P}_{nm}(\mathbf{Q}_{am}, \mathbf{P}_{am}, \mathbf{Q}_{an}, \mathbf{P}_{an}).$$

These can be introduced into $d\varphi$ and then a final reduction to a differential form $\mathbf{P}_0 \cdot d\mathbf{Q}_0 + \bar{\mathbf{P}}_\gamma \cdot d\bar{\mathbf{Q}}_\gamma$ must be effected.

ACKNOWLEDGMENT

I have enjoyed conversations with Dr. R. N. Hill and the partial support of the National Science Foundation in this work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Field-Theoretic Description of Massless Particles with Higher Spin and Definite Parity. I. Integer Spin

T. CUKIERDA AND J. LUKIERSKI
Institute of Theoretical Physics, University of Wroclaw, Wroclaw Cybulskiego 36, Poland

(Received 9 May 1969)

It is known that there exist $S + 1$ equivalent covariant formulations of the theory of free massive particles with integer spin S and definite parity P. From these $S + 1$ equivalent theories one gets by putting $m = 0$ the description of $S + 1$ *different* massless particles, with helicities $(k, -k)$, $k = 0, 1, \cdots, S$. This property is already known in the case $S = 1$. In this paper we demonstrate explicitly how to obtain three kinds of massless particles, with different helicities, from three equivalent formulations of free massive theory with $S = 2$. Further, we outline the general argument for arbitrary integer value of S.

1. INTRODUCTION

It has been demonstrated by Fierz[1] that the free massive particles with spin S can be described by the following set of $S + 1$ covariant tensors:

$$B^{(k,S)}_{\cdots\sigma[\mu\nu]\cdots} \equiv B_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_S\nu_S]}, \quad k = 0, 1, \cdots, S, \quad (1.1)$$

where $[\mu\nu]$ denotes an antisymmetric pair of indices. Furthermore, the tensor (1.1) is symmetric with respect to any permutation (σ_i, σ_j), $i, j \leq k$, and $([\mu_i\nu_i], [\mu_j\nu_j])$.

The notion of free particle with mass m implies that

$$(\Box - m^2)B^{(k,S)} = 0. \quad (1.2)$$

The Lorentz tensor (1.1) satisfying (1.2) describes, therefore, a parity-doubled multiplet of free particles with spins $S, S-1, \cdots, k$ for $k < S$. In order to separate a field operator describing the massive particle with definite parity and spin S, one should impose the following subsidiary conditions:

$$B^{(k,S)\mu}_{\cdots[\mu\nu]\cdots} = 0, \quad (1.3a)$$

$$B^{(k,S)\mu}_{\cdots\mu[\nu\sigma]\cdots} = 0, \quad (1.3b)$$

$$B^{(k,S)}_{\cdots\mu[\nu\sigma]\cdots} + B^{(k,S)}_{\cdots\sigma[\mu\nu]\cdots} + B^{(k,S)}_{\cdots\nu[\sigma\mu]\cdots} = 0, \quad (1.3c)$$

$$\partial^\mu B^{(k,S)}_{\cdots\mu[\nu\sigma]\cdots} = 0, \quad (1.3d)$$

$$\partial_\mu B^{(k,S)}_{\cdots[\nu\sigma]\cdots} + \partial_\sigma B^{(k,S)}_{\cdots[\mu\nu]\cdots} + \partial_\nu B^{(k,S)}_{\cdots[\sigma\mu]\cdots} = 0, \quad k = 0, 1, \cdots, S, \quad (1.3e)$$

and, additionally, if $k = 0$, one should add the following condition:

$$B^{(0,S)\,[\mu\nu]}_{\cdots[\mu\nu]} = 0. \quad (1.3f)$$

It can be shown that the tensor $B^{(k,S)}$, satisfying the subsidiary conditions (1.3a)–(1.3f), has only $2S + 1$ independent components. Particularly, in the case $S = 1$, we have the following two possibilities[2]:

$$(\Box - m^2)B_\mu = 0, \quad \partial^\mu B_\mu = 0, \quad (1.4)$$

and

$$(\Box - m^2)B_{[\mu\nu]} = 0, \quad (1.5a)$$

$$\partial_\mu B_{[\nu\sigma]} + \partial_\sigma B_{[\mu\nu]} + \partial_\nu B_{[\sigma\mu]} = 0 \quad (1.5b)$$

or, introducing

$$G_{[\mu\nu]} = \tfrac{1}{2}\epsilon_{\mu\nu}{}^{\sigma\tau}B_{[\sigma\tau]}, \quad (1.6)$$

one can write (1.5) as follows:

$$(\Box - m^2)G_{[\mu\nu]} = 0, \quad (1.7a)$$

$$\partial^\mu G_{[\mu\nu]} = 0. \quad (1.7b)$$

[1] M. Fierz, Helv. Phys. Acta **12**, 1 (1939).

[2] For the case $S = 1$ and $S = 2$ we shall omit the indices in the upper bracket, because the tensor indices are written explicitly.

give

$$\mathbf{Q}_{nm}(\mathbf{Q}_{am}, \mathbf{P}_{am}, \mathbf{Q}_{an}, \mathbf{P}_{an})$$

and

$$\mathbf{P}_{nm}(\mathbf{Q}_{am}, \mathbf{P}_{am}, \mathbf{Q}_{an}, \mathbf{P}_{an}).$$

These can be introduced into $d\varphi$ and then a final reduction to a differential form $\mathbf{P}_0 \cdot d\mathbf{Q}_0 + \bar{\mathbf{P}}_\gamma \cdot d\bar{\mathbf{Q}}_\gamma$ must be effected.

ACKNOWLEDGMENT

I have enjoyed conversations with Dr. R. N. Hill and the partial support of the National Science Foundation in this work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Field-Theoretic Description of Massless Particles with Higher Spin and Definite Parity. I. Integer Spin

T. CUKIERDA AND J. LUKIERSKI
Institute of Theoretical Physics, University of Wroclaw, Wroclaw Cybulskiego 36, Poland

(Received 9 May 1969)

It is known that there exist $S + 1$ equivalent covariant formulations of the theory of free massive particles with integer spin S and definite parity P. From these $S + 1$ equivalent theories one gets by putting $m = 0$ the description of $S + 1$ *different* massless particles, with helicities $(k, -k)$, $k = 0, 1, \cdots, S$. This property is already known in the case $S = 1$. In this paper we demonstrate explicitly how to obtain three kinds of massless particles, with different helicities, from three equivalent formulations of free massive theory with $S = 2$. Further, we outline the general argument for arbitrary integer value of S.

1. INTRODUCTION

It has been demonstrated by Fierz[1] that the free massive particles with spin S can be described by the following set of $S + 1$ covariant tensors:

$$B^{(k,S)}_{\cdots\sigma[\mu\nu]\cdots} \equiv B_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_S\nu_S]}, \quad k = 0, 1, \cdots, S, \quad (1.1)$$

where $[\mu\nu]$ denotes an antisymmetric pair of indices. Furthermore, the tensor (1.1) is symmetric with respect to any permutation (σ_i, σ_j), $i, j \leq k$, and $([\mu_i\nu_i], [\mu_j\nu_j])$.

The notion of free particle with mass m implies that

$$(\Box - m^2)B^{(k,S)} = 0. \quad (1.2)$$

The Lorentz tensor (1.1) satisfying (1.2) describes, therefore, a parity-doubled multiplet of free particles with spins $S, S-1, \cdots, k$ for $k < S$. In order to separate a field operator describing the massive particle with definite parity and spin S, one should impose the following subsidiary conditions:

$$B^{(k,S)\mu}_{\cdots[\mu\nu]\cdots} = 0, \quad (1.3a)$$

$$B^{(k,S)\mu}_{\cdots\mu[\nu\sigma]\cdots} = 0, \quad (1.3b)$$

$$B^{(k,S)}_{\cdots\mu[\nu\sigma]\cdots} + B^{(k,S)}_{\cdots\sigma[\mu\nu]\cdots} + B^{(k,S)}_{\cdots\nu[\sigma\mu]\cdots} = 0, \quad (1.3c)$$

$$\partial^\mu B^{(k,S)}_{\cdots\mu[\nu\sigma]\cdots} = 0, \quad (1.3d)$$

$$\partial_\mu B^{(k,S)}_{\cdots[\nu\sigma]\cdots} + \partial_\sigma B^{(k,S)}_{\cdots[\mu\nu]\cdots} + \partial_\nu B^{(k,S)}_{\cdots[\sigma\mu]\cdots} = 0, \quad k = 0, 1, \cdots, S, \quad (1.3e)$$

and, additionally, if $k = 0$, one should add the following condition:

$$B^{(0,S)\,[\mu\nu]}_{\cdots[\mu\nu]} = 0. \quad (1.3f)$$

It can be shown that the tensor $B^{(k,S)}$, satisfying the subsidiary conditions (1.3a)–(1.3f), has only $2S + 1$ independent components. Particularly, in the case $S = 1$, we have the following two possibilities[2]:

$$(\Box - m^2)B_\mu = 0, \quad \partial^\mu B_\mu = 0, \quad (1.4)$$

and

$$(\Box - m^2)B_{[\mu\nu]} = 0, \quad (1.5a)$$

$$\partial_\mu B_{[\nu\sigma]} + \partial_\sigma B_{[\mu\nu]} + \partial_\nu B_{[\sigma\mu]} = 0 \quad (1.5b)$$

or, introducing

$$G_{[\mu\nu]} = \tfrac{1}{2}\epsilon_{\mu\nu}{}^{\sigma\tau}B_{[\sigma\tau]}, \quad (1.6)$$

one can write (1.5) as follows:

$$(\Box - m^2)G_{[\mu\nu]} = 0, \quad (1.7a)$$

$$\partial^\mu G_{[\mu\nu]} = 0. \quad (1.7b)$$

[1] M. Fierz, Helv. Phys. Acta **12**, 1 (1939).

[2] For the case $S = 1$ and $S = 2$ we shall omit the indices in the upper bracket, because the tensor indices are written explicitly.

If $m \neq 0$, both theories (1.4) and (1.7) are equivalent.[3-5] If, however, $m = 0$, Eqs. (1.4) and (1.7) allow the introduction of *different* classes of gauge transformations, namely,

$$B_\mu \to B'_\mu = B_\mu - \partial_\mu \Lambda, \quad \Box\Lambda = 0, \tag{1.8}$$

and

$$G_{[\mu\nu]} \to G'_{[\mu\nu]} = G_{[\mu\nu]} - \{\partial_\mu \Lambda_\nu - \partial_\nu \Lambda_\mu\}, \tag{1.9}$$

where, in order to satisfy *both* conditions (1.7), one should assume

$$\Box\Lambda_\mu = 0, \quad \partial^\mu \Lambda_\mu = 0. \tag{1.9'}$$

It can be easily seen[6] that the gauge freedom (1.8) reduces the number of three independent components of B_μ by one and the massless field B'_μ describes *two* degrees of freedom, orthogonal to the three-momentum vector $\mathbf{P}$. One can impose, therefore, the condition

$$\partial^j B'_j = 0, \quad j = 1, 2, 3, \tag{1.10}$$

defining the "physical" radiation gauge in conventional QED.

In the case when the theory of massless particles is described by the set of equations (1.7) with $m = 0$, invariant with respect to the gauge transformations (1.9'), one can reduce the number of three independent components by two.[4] Indeed, if we assume that the four-momentum vector P_μ, characterizing our massless particle, is chosen as $P_\mu = (0, 0, |\mathbf{P}|, |\mathbf{P}|)$, condition (1.7) can be written in momentum space as

$$|\mathbf{P}| \{G_{[\mu 3]}(\mathbf{P}) - G_{[\mu 0]}(\mathbf{P})\} = 0. \tag{1.11}$$

From (1.11) it follows that $G_{[03]}(\mathbf{P}) = 0$ and, choosing

$$\Lambda_r(\mathbf{P}) = |\mathbf{P}|^{-1} G_{[0r]}(\mathbf{P}), \quad r = 1, 2,$$

in our particular coordinate system, one gets finally

$$G'_{[\mu\nu]}(\mathbf{P}) = 0, \quad \text{if} \quad \mu, \nu = 3, 0. \tag{1.12}$$

Finally, the only nonvanishing component is $G'_{[12]}(\mathbf{P})$. The condition (1.12) can be also treated as the result of the reduction of three components $G_{[ij]}(\mathbf{P})$, $i, j = 1, 2, 3$, by means of the following *two* radiation gauge conditions:

$$\partial^i G'_{[ij]}(\mathbf{P}) = 0. \tag{1.13}$$

Introducing

$$F_i = \tfrac{1}{2}\epsilon_{ijk} G'_{[jk]}, \tag{1.14}$$

one can write the condition (1.13) as follows:

$$\nabla \times \mathbf{F} = 0, \tag{1.15}$$

and we get finally that

$$\mathbf{F} = \nabla A. \tag{1.16}$$

We see, therefore, that the theory of massless particles described by the antisymmetric tensor $G_{[\mu\nu]}$ is characterized by the three-dimensional scalar A, which has helicity equal to zero. Such a massless particle has been deduced from the field equations (1.7) in the limit $m = 0$ by Ogievetskiĭ and Polubarinov,[4] and was called a "notoph."

The difference between the massless limits of massive equivalent theories is caused by the fact that *the set of the equations* (1.2), (1.3) *allows k different classes of gauge transformations.* The argument for $S = 1$ can be generalized. In Sec. 2 we present the derivation of three different kinds of massless particles (with helicities $H = 2$, $H = 1$, and $H = 0$) by putting $m = 0$ in the equations and determining equivalent $S = 2$ fields. In Sec. 3 we discuss the general case. In both these sections we shall consider, instead of the Fierz tensor (1.1), the dual ones

$$G^{(k,S)}_{\sigma_1 \cdots \sigma_k [\mu_{k+1}\nu_{k+1}] \cdots [\mu_S \nu_S]} = (\tfrac{1}{2})^{S-k} \epsilon_{\mu_{k+1}\nu_{k+1}}{}^{\rho_{k+1}\tau_{k+1}} \cdots \epsilon_{\mu_S \nu_S}{}^{\rho_S \tau_S} \times B^{(k,S)}_{\sigma_1 \cdots \sigma_k [\rho_{k+1}\tau_{k+1}] \cdots [\rho_S \nu_S]} \tag{1.17}$$

satisfying, after putting $m = 0$, the following set of equations:

$$\begin{aligned} &\Box G^{(k,S)}_{\cdots \sigma[\mu\nu]\cdots} = 0, \quad G^{(k,S)}_{\cdots \sigma[\mu\nu]}{}^{[\mu\nu]\cdots} = 0, \\ &G^{(k,S)\sigma}_{\cdots \sigma[\mu\nu]\cdots} = 0, \quad G^{(k,S)\mu}_{\cdots \ [\mu\nu]\cdots} = 0, \\ &\partial^\sigma G^{(k,S)}_{\cdots \sigma[\mu\nu]\cdots} = 0, \quad \partial^\mu G^{(k,S)}_{\cdots \sigma[\mu\nu]\cdots} = 0, \\ &G^{(k,S)}_{\cdots \sigma[\mu\nu]\cdots} + G^{(k,S)}_{\cdots \nu[\sigma\mu]\cdots} + G^{(k,S)}_{\cdots \mu[\nu\sigma]\cdots} = 0. \end{aligned} \tag{1.18}$$

It will be shown that the set of $S + 1$ tensors $G^{(k,S)}$, $k = 0, \cdots, S$, describes $S + 1$ *different* massless particles, with helicities $H = 0, 1, \cdots, S$. One can say, therefore, that there is an infinite number of ways of expressing the field theory of massless particles with helicity H, because for every $S \geq H$ there exists one Fierz tensor (1.1) which describes such massless particles.

2. THE $S = 2$ CASE

A. $k = 0$

Let us recall briefly the conventional formulation of $S = 2$ particles, based on the symmetric tensor $B_{\mu\nu}$, which leads, if we put $m = 0$, to the helicity

[3] N. Kemmer, Helv. Phys. Acta 33, 829 (1960).
[4] V. I. Ogieveskiĭ and I. V. Polubarinov, Yad. Fiz. 4, 216 (1966) [Sov. J. Nucl. Phys. 4, 156 (1967)].
[5] T. Cukierda and J. Lukierski, Nucl. Phys. B5, 508 (1968).
[6] This property is shown in any textbook on quantum electrodynamics (QED).

$H = 2$ massless particles (gravitations). The subsidiary conditions are the following ones[7]:

$$B_\mu{}^\mu = 0, \tag{2.1}$$

$$\partial^\mu B_{\mu\nu} = 0. \tag{2.2}$$

If we put $m = 0$, the following class of gauge transformation, preserving (2.1) and (2.2) and $\Box B_{\nu\mu} = 0$, is allowed:

$$B_{\mu\nu} \to B'_{\mu\nu} = B_{\mu\nu} - \{\partial_\mu \Lambda_\nu + \partial_\nu \Lambda_\mu\}, \tag{2.3}$$

where

$$\Box \Lambda_\mu = 0, \quad \partial^\mu \Lambda_\mu = 0. \tag{2.4}$$

Putting $P_\mu = (0, 0, |\mathbf{P}|, |\mathbf{P}|)$, one can choose Λ_μ in such a way that

$$B'_{3i}(\mathbf{P}) = 0, \quad i = 1, 2, 3. \tag{2.5}$$

From (2.2) and (2.5) it follows that

$$B'_{0i}(\mathbf{P}) = 0, \quad B'_{00}(\mathbf{P}) = 0. \tag{2.6}$$

Using (2.1) and (2.6), one gets only two independent components $B'_{11}(\mathbf{P}) = -B'_{22}(\mathbf{P})$ and $B'_{12}(\mathbf{P}) = B'_{21}(\mathbf{P})$. Because every symmetric tensor index describing components perpendicular to the three-momentum increases the helicity by one, we get the result that $H = 2$.

B. $k = 1$

One obtains another equivalent theory of massive $S = 2$ free particles by using the field $G_{\mu[\nu\sigma]}$. Besides the field equation

$$\Box G_{\mu[\nu\sigma]} = 0, \tag{2.7}$$

one should assume the following subsidiary conditions:

$$\partial^\mu G_{\mu[\nu\sigma]} = 0, \tag{2.8a}$$

$$\partial^\nu G_{\mu[\nu\sigma]} = 0, \tag{2.8b}$$

$$G^\mu{}_{[\mu\nu]} = 0, \tag{2.8c}$$

$$G_{\mu[\nu\sigma]} + G_{\sigma[\mu\nu]} + G_{\nu[\sigma\mu]} = 0. \tag{2.8d}$$

The set of equations (2.7), (2.8) is invariant under the following class of gauge transformations:

$$G_{\mu[\nu\sigma]} \to G'_{\mu[\nu\sigma]} = G_{\mu[\nu\sigma]} - \{\partial_\nu \Lambda_{\mu\sigma} - \partial_\sigma \Lambda_{\mu\nu}\}, \tag{2.9}$$

where $\Lambda_{\mu\nu} = \Lambda_{\nu\mu}$, and

$$\Box \Lambda_{\mu\nu} = 0, \quad \partial^\mu \Lambda_{\mu\nu} = 0, \tag{2.10a}$$

$$\Lambda_\mu{}^\mu = 0. \tag{2.10b}$$

Let us consider now the coordinate system with $P_\mu = (0, 0, |\mathbf{P}|, |\mathbf{P}|)$. From (2.8a), it follows that

$$G_{3[\nu\sigma]}(\mathbf{P}) = G_{0[\nu\sigma]}(\mathbf{P}), \tag{2.11}$$

and, from (2.8b),

$$G_{\mu[3i]}(\mathbf{P}) = G_{\mu[0i]}(\mathbf{P}), \quad i = 1, 2, \tag{2.12}$$

$$G_{\mu[30]}(\mathbf{P}) = 0. \tag{2.13}$$

The relation (2.8c) implies that

$$G_{i[jk]}(\mathbf{P}) = 0, \quad i, j, k = 1, 2. \tag{2.14}$$

The subsidiary condition (2.8d) for remaining nonvanishing components can be written as follows:

$$G_{3[ij]}(\mathbf{P}) = G_{i[3j]}(\mathbf{P}) - G_{j[3i]}(\mathbf{P}). \tag{2.15}$$

We see, therefore, that the components $G_{i[3j]}$ can be made symmetric with respect to the exchange $i \leftrightarrow j$, because the antisymmetric part is given by $G_{3[ij]}$. We are left with the following five components:

$$G_{3[ij]}, \quad G_{3[i3]}, \quad G_{i[j3]}. \tag{2.16}$$

Four out of these five components can be eliminated by means of the gauge transformations (2.9), (2.10). In our particular coordinate system the following five components of $\Lambda_{\mu\nu}$ can be chosen as independent:

$$\Lambda_{ij}(\mathbf{P}) = \Lambda_{ji}(\mathbf{P}), \quad \Lambda_{3\mu}(\mathbf{P}) = \Lambda_{0\mu}(\mathbf{P}). \tag{2.17}$$

Performing the gauge transformation (2.9), one can easily see that only the components $G_{3[ij]}$ are gauge invariant, and the other four, listed in (2.16), can be put equal to zero by a suitable choice of the gauge transformation. Indeed, we have

$$G'_{3[i3]}(\mathbf{P}) = G_{3[i3]}(\mathbf{P}) - |\mathbf{P}|\,\Lambda_{3i}(\mathbf{P}),$$

$$G'_{i[j3]}(\mathbf{P}) = G_{i[j3]}(\mathbf{P}) - |\mathbf{P}|\,\Lambda_{ij}(\mathbf{P}), \tag{2.18}$$

and we should choose

$$\Lambda_{3i}(\mathbf{P}) = |\mathbf{P}|^{-1} G_{3[i3]}(\mathbf{P}),$$

$$\Lambda_{ij}(\mathbf{P}) = |\mathbf{P}|^{-1} G_{i[j3]}(\mathbf{P}). \tag{2.19}$$

The components $G_{3[ij]}$ define the massless field, described by the field equations (2.7), (2.8). Such a field is characterized fully by one component $G_{3[ij]}$, which describes a massless particle with helicity $H = 0$.

C. $k = 2$

The third formulation of the theory of free massive $S = 2$ particles gives the following set of equations, after putting $m = 0$:

$$\Box G_{[\mu\nu][\sigma\tau]} = 0, \tag{2.20}$$

where $G_{[\mu\nu][\sigma\tau]} = G_{[\sigma\tau][\mu\nu]}$ and

$$\partial^\mu G_{[\mu\nu][\sigma\tau]} = 0, \tag{2.21a}$$

$$G_{[\mu\nu]}{}^{[\mu\nu]} = 0. \tag{2.21b}$$

[7] For $k = 0$, $G_{\mu_1 \cdots \mu_S} \equiv B_{\mu_1 \cdots \mu_S}$ [see Eq. (1.17)].

The set of equations (2.20)–(2.21) is invariant under the following gauge transformation:

$$G_{[\mu\nu][\sigma\tau]} \to G'_{[\mu\nu][\sigma\tau]} = G_{[\mu\nu][\sigma\tau]} - \{\partial_\nu\partial_\tau\Lambda_{\mu\sigma} + \partial_\mu\partial_\sigma\Lambda_{\nu\tau} - \partial_\mu\partial_\tau\Lambda_{\nu\sigma} - \partial_\nu\partial_\sigma\Lambda_{\mu\tau}\}, \quad (2.22)$$

where $\Lambda_{\mu\sigma}$ satisfies the conditions (2.10a).

Let us write the relations (2.21) in our particular coordinate frame with the momentum of the particle $P_\mu = (0, 0, |\mathbf{P}|, |\mathbf{P}|)$. We get from (2.21a) that

$$G_{[3i][\sigma\tau]}(\mathbf{P}) = G_{[0i][\sigma\tau]}(\mathbf{P}), \quad i = 1, 2, \quad (2.23)$$

and

$$G_{[30][\sigma\tau]}(\mathbf{P}) = 0. \quad (2.24)$$

From the symmetry properties it also follows that

$$G_{[3i][3j]}(\mathbf{P}) = G_{[3j][3i]}(\mathbf{P}), \quad (2.25)$$

and from (2.21a) and (2.24–25) it follows that

$$G_{[12][12]}(\mathbf{P}) = 0. \quad (2.26)$$

We see, therefore, that we have the five following independent components:

$$G_{[12][3i]}, \quad G_{[3i][3j]}. \quad (2.27)$$

It is easy to see that in our particular frame the components $G_{[12][3i]}$ are gauge invariant, and

$$G'_{[3i][3j]}(\mathbf{P}) = G_{[3i][3j]}(\mathbf{P}) - |\mathbf{P}|^2\Lambda_{ij}(\mathbf{P}). \quad (2.28)$$

From (2.28) it follows that the components $G_{[3i][3j]}$ can be eliminated. We are left with two components $G_{[12][3i]}$ describing massless particles with helicity $H = 1$.

3. GENERAL CASE: S POSITIVE INTEGER ≥ 3

Now we shall generalize our discussion to the case of arbitrary S. The method is analogous to the one presented in Sec. 2 for the case $S = 2$ theories.

A. $k = S$

The case when a massless particle is described by the symmetric tensor of rank S has been considered by Fierz in Ref. 1. Such a tensor field allows the introduction of gauge transformations, depending on some symmetric tensor $\Lambda^{(s-1)}$ of rank $S - 1$. This gauge transformation reduces the number of components in such a way that we are left only with two independent components (for example $B_{1\ldots11}$, $B_{1\ldots12}$, where there are S indices in each case) describing the massless particle with helicity S.

B. $0 \leq k \leq S - 1$

The free theory for $k \geq 1$ is described by Eqs. (1.18). It can be shown that this set of equations is invariant under the following set of gauge transformations:

$$G'^{(k,S)}_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_S\nu_S]} = G^{(k,S)}_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_S\nu_S]} - \sum_{\text{perm}} (-1)^n \partial_{\mu_{k+1}}\ldots\partial_{\mu_S}\Lambda^{(S)}_{\sigma_1\cdots\sigma_k\nu_{k+1}\cdots\nu_S}, \quad (3.1)$$

where $\sum_{\text{perm}}$ denotes sum over all permutations of pairs (μ_{k+2}, ν_{k+2}) and n describes number of such permutations. Because $G'^{(k,S)}$ should also satisfy Eqs. (1.18), the field $\Lambda^{(S)}$ must satisfy the following conditions:

$$\begin{aligned}
&\Box\Lambda^{(S)} = 0,\\
&\sum_{\mu_{k+1},\nu_{k+1}} \sum_{\text{perm}} (-1)^n \partial^{\mu_{k+1}}\partial_{\mu_{k+1}}\partial_{\mu_{k+2}}\cdots \partial_{\mu_S}\Lambda^{(S)}{}^{\nu_{k+1}}_{\sigma_1\cdots\sigma_k\ \nu_{k+1}\cdots\nu_S} = 0,\\
&\sum_{\text{perm}} (-1)^n \partial_{\mu_{k+1}}\cdots\partial_{\mu_S}\Lambda^{(S)\sigma_1}{}_{\sigma_1\sigma_2\cdots\sigma_k\nu_{k+1}\cdots\nu_S} = 0,\\
&\partial^{\mu_{k+1}} \sum_{\text{perm}} (-1)^n \partial_{\mu_{k+1}}\cdots\partial_{\mu_S}\Lambda^{(S)}_{\sigma_1\cdots\sigma_k\nu_{k+1}\cdots\nu_S} = 0,\\
&\partial^{\sigma_1} \sum_{\text{perm}} (-1)^n \partial_{\mu_{k+1}}\cdots\partial_{\mu_S}\Lambda^{(S)}_{\sigma_1\cdots\sigma_k\nu_{k+1}\cdots\nu_S} = 0.
\end{aligned} \quad (3.2)$$

Let us write now Eqs. (1.18) in momentum space, and choose $P_\mu = (0, 0, |\mathbf{P}|, |\mathbf{P}|)$. Using techniques analogous to the ones presented in Sec. 2, we can restrict the consideration of independent components to the following ones:

$$G^{(k,S)}_{\sigma_1\cdots\sigma_k[i_{k+1}3]\cdots[i_S3]}, \quad (3.3)$$

where $\sigma_i = 0, 1, 2, 3$, $i = 1, 2, \cdots, k$, and

$$G^{(k,S)}_{3\cdots3[12][i_{k+2}3]\cdots[i_S3]}, \quad (3.4)$$

where $i_m = 1, 2$, $m = k + 1, k + 2, \cdots, S$.

The components (3.3) and (3.4) are not all independent, and we should impose the conditions of vanishing traces. Using the property that the fields $G^{(k,S)}$ are determined up to the arbitrary gauge transformation (3.1), it can be shown that all the components (3.3) can be eliminated. Indeed, substituting in (3.1), we obtain

$$\Lambda^{(S)}_{\sigma_1\cdots\sigma_k i_{k+1}\cdots i_S}(\mathbf{P}) = [|\mathbf{P}|^{-1}]^{S-k} G^{(k,S)}_{\sigma_1\cdots\sigma_k[i_{k+1}3]\cdots[i_S3]}(\mathbf{P}); \quad (3.5)$$

we get the result that after the gauge transformation the symmetric part of (3.3), with respect to indices $\sigma_i \leftrightarrow i_m$, $\sigma_i = 1, 2$, $i = 1, \cdots, k$, $m = k + 1, \cdots, S$, vanishes. On the other hand, the antisymmetric part with respect to the above permutation can be expressed by means of the components (3.4) [compare with (2.15) for $S = 2$]. We are left, therefore, with

the components (3.4), and the condition of vanishing trace

$$G^{(k,S)}_{3\cdots 3[12][i_{k+1}3]\cdots[i_S3]}, \tag{3.6}$$

together with symmetry properties, implies that our massless object is described by a symmetric tensor

$$H^{(k;S)}_{i_{k+2}\cdots i_S} = G^{(k,S)}_{3\cdots 3[12][i_{k+2}3]\cdots[i_S3]} \tag{3.7}$$

with vanishing trace. Because, generally, symmetric tensors of nth order in $(m+1)$-dimensional space have $\binom{n+m}{m}$ independent components, we get, after putting $n = S-k-2$ and $m = 1$ that:

(a) the symmetric tensor (3.7) has $S-k-1$ components;

(b) the trace condition eliminates $S-k-3$ components.

We are left, therefore, with the following two components:

$$H^{(k;S)}_{1\cdots 11}, \quad H^{(k;S)}_{1\cdots 12}, \tag{3.8}$$

where again there are $S-k-1$ indices in each component. These two components describe the massless particles with helicity $H = S-k-1, k = 0, 1, \cdots, S-1$.

4. FINAL REMARKS

In this paper it has been shown that, from $S+1$ *equivalent* theories of free massive field with spin S, one can obtain, by putting $m = 0$, $S+1$ *different* theories of massless particles with the helicities $H = 0, 1, 2, \cdots, S$. The essential role of gauge transformations has been pointed out. We have used two different types of gauge transformations: For the theories with $k = S$, the gauge function was represented by a symmetric tensor $\Lambda^{(S-1)}$ of rank $S-1$, and for $k < S$ we were performing the gauge transformation depending on the symmetric tensor $\Lambda^{(S)}$. Because of such a choice the relation between the number k of symmetric indices and the helicity value is described by a formula which treats the case $k = S$ in an exceptional way, namely,

$$k = S \rightarrow H = S,$$
$$k < S \rightarrow H = S-k-1.$$

In this paper we were considering only the field equations for massless particles and we introduced the gauge transformations as ones which leave these field equations invariant. It is possible, however, to introduce as a basis the gauge-invariant Lagrangians and to treat the field equations as a derived result.

The main conclusion of this paper is the following: There are an infinite number of ways of introducing the free theory of massless particles with given helicity H by means of Lorentz-invariant tensors. Particularly, the massless particle with helicity $H = 0$ can be described by any of these following fields:

G,	scalar,
$G_{[\mu\nu]}$,	"notoph,"
$G_{\sigma_1\cdots\sigma_k[\mu\nu]}$,	generalized "notoph,"

where $k = 1, 2, \cdots$.

If we consider *quantized* theory, characterized not only by the field equations, but also by the commutator functions, the limit $m \rightarrow 0$ is much more complicated, because of the presence of the denominators m^{-2k} (see, for example, Ref. 8) in the Green's functions (commutator function, causal propagator). All these problems are here avoided, but in the process of obtaining a complete quantization theory they also should be faced.

ACKNOWLEDGMENT

One of the authors (J. L.) wish to express his thanks to Professor W. I. Ogievetskiĭ for his useful comments.

[8] J. Lukierski, Nuovo Cimento **38**, 1407 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Null Electromagnetic Fields in General Relativity Admitting Timelike or Null Killing Vectors

A. BANERJEE
Department of Physics, Jadavpur University, Calcutta, India

(Received 25 March 1969)

The paper presents a study of stationary null electromagnetic fields coexisting with a dust distribution. Two exact solutions corresponding to pure radiation fields with the propagation vector as the Killing vector are presented.

1. INTRODUCTION

In the literature on general relativity, one finds rather few exact solutions corresponding to null electromagnetic fields. In recent years, however, a type of solutions for null electromagnetic fields has been discovered which is "stationary" in the sense that the space–time admits a group of motions with a timelike generator. In the solutions of this type given by Dutta and Raychaudhuri,[1] the electromagnetic field tensor also is independent of this time variable, and thus the solutions could be interpreted by them as representing superposed static electric and magnetic fields which are orthogonal to each other. In the rotating universes discovered by Ozsvath,[2] the electromagnetic field is, however, time dependent and is therefore naturally considered to be a radiation field. In Ozsvath's solutions, besides the electromagnetic radiation there is a distribution of electrically neutral dust. These solutions are homogeneous and possess the following interesting features:

(1) Neither the velocity vector of the fluid nor the propagation vector for the null electromagnetic field is hypersurface orthogonal. Physically, these correspond to a vorticity of the fluid motion and a curling of the light rays.

(2) The vorticity vector is orthogonal to both the electric and magnetic field vectors and is thus in the same direction as the Poynting vector, with all these vectors being orthogonal to the fluid velocity vector.

(3) The electromagnetic field tensor has an especially simple form.

(4) There exists an additional motion of the space–time with a timelike generator which, unlike the fluid velocity, is hypersurface orthogonal.

In Sec. 2 of the present paper, a general study of the stationary null fields is attempted, and some formulas emerge which allow a clearer understanding of some of the results of Ozsvath, and Dutta and Raychaudhuri. In Sec. 3, we present two null-field solutions which we believe to be new. Both the solutions admit a null translation with the propagation vector of the field as the generator, while one of the solutions is also stationary.

2. GENERAL STUDY OF STATIONARY NULL FIELDS

Let the space–time be stationary; that is, it admits a motion with a timelike generator. Further, in case there is matter, let us assume that this generator is the velocity vector so that the Killing equation,

$$v_{\alpha;\beta} + v_{\beta;\alpha} = 0, \tag{1}$$

is satisfied and the motion of the matter is rigid. We can choose coordinates so that this motion is a translation along the time axis. The $g_{\mu\nu}$'s would then be independent of time, and the line element would assume the form

$$ds^2 = dt^2 + 2g_{0i}\,dt\,dx^i + g_{ik}\,dx^i\,dx^k,$$

where the index "0" signifies the time coordinate. Now from (1) it follows that $v^{\alpha}_{;\alpha} = 0$ and $v^{\alpha}_{;\beta}v^{\beta} = 0$. The field equation can be written in the form

$$R_{\alpha\beta} = -8\pi\rho v_{\alpha}v_{\beta} + (\lambda + 4\pi\rho)g_{\alpha\beta} + 2F_{\alpha\mu}F^{\mu}_{\beta}, \tag{2}$$

where ρ and λ stand for the density of matter (assumed to be in the form of pressureless dust) and the cosmological constant, respectively. $F_{\alpha\mu}$ represents the electromagnetic field tensor. The energy tensor for the null electromagnetic field ($F_{\alpha\beta}F^{\alpha\beta} = 0$, $F^{*}_{\alpha\beta}F^{\alpha\beta} = 0$) may also be written in the form[3]

$$\begin{aligned} T_{\alpha\beta} &= -(4\pi)^{-1}F_{\alpha\mu}F^{\mu}_{\beta} \\ &= (4\pi)^{-1}E^2(2v_{\alpha}v_{\beta} - g_{\alpha\beta}) - \frac{1}{4\pi}(E_{\alpha}E_{\beta} + H_{\alpha}H_{\beta}) \\ &\quad -(4\pi)^{-1}(v_{\alpha}S_{\beta} + v_{\beta}S_{\alpha}), \end{aligned} \tag{3}$$

where S, E, and H represent the Poynting vector, electric field, and magnetic field vectors, respectively,

[1] B. K. Dutta and A. K. Raychaudhuri, J. Math. Phys. 9, 1715 (1968).

[2] I. Ozsvath, in *Perspectives in Geometry and Relativity*, B. Hoffman, Ed. (Indiana University Press, Bloomington, Ind., 1966).

[3] A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (W. A. Benjamin, Inc., New York, 1967).

defined in the following way:

$$S^\mu = \tfrac{1}{2}\eta^{\mu\nu\rho\sigma}E_\nu H_\rho v_\sigma, \quad E^\mu = F^{\alpha\mu}v_\alpha,$$

and

$$H^\mu = \tfrac{1}{2}\eta^{\mu\nu\rho\sigma}F_{\nu\rho}v_\sigma.$$

Thus defined, all three vectors are orthogonal to the velocity vector of matter. Again, as the $g_{\mu\nu}$'s and the vector v_α are independent of time, it follows from Eq. (2) that ρ and the tensor $F_{\alpha\mu}F^\mu_\alpha$ are also independent of time. Hence, $E^2 = (F_{\alpha\mu}F^\mu_\beta)v^\alpha v^\beta$ must also be independent of time so that $E^2_{;\alpha}v^\alpha = 0$. Hence, from (3), it follows that

$$(T^\alpha_\beta v^\beta)_{;\alpha} = -(4\pi)^{-1}S^\alpha_{;\alpha}. \tag{4}$$

Writing the left-hand side in the form

$$(T^\alpha_\beta v^\beta)_{;\alpha} = \tfrac{1}{2}T^{\alpha\beta}v_{(\beta;\alpha)} + (T^{\alpha\beta})_{;\alpha}v^\beta,$$

one finds that both the terms in the right-hand side vanish in view of the relation (1) and the assumption that electromagnetic field is source free. It now follows from (4) that

$$S^\alpha_{;\alpha} = 0. \tag{5}$$

The divergence-free character of the Poynting vector leads to the conclusion that there is no net energy flux through any closed surface.

Next, we show that under such circumstances the motion of the dust cannot be irrotational.

From (2) and (3) we get

$$R_{\alpha\beta}v^\alpha v^\beta = \lambda - 4\pi\rho - 2E^2 \tag{6}$$

and

$$R^\alpha_\beta v^\beta = (\lambda - 4\pi\rho - 2E^2)v^\alpha + 2S^\alpha. \tag{7}$$

Now, using the definition of the Riemann–Christoffel tensor,

$$v_{\alpha;\beta\gamma} - v_{\alpha;\gamma\beta} = -v^\rho R_{\rho\alpha\beta\gamma},$$

one gets after contraction that

$$v^\alpha_{;\alpha\gamma} - v^\alpha_{;\gamma\alpha} = u^\rho R_{\rho\gamma}, \tag{8}$$

and thus,

$$R_{\alpha\beta}v^\alpha v^\beta = (v^\alpha_{;\alpha})_{,\beta}v^\beta - (v^\alpha_{;\beta}v^\beta)_{;\alpha} + v^\alpha_{;\beta}v^\beta_{;\alpha}. \tag{9}$$

Now defining the vorticity tensor in the usual way,

$$\omega_{\alpha\beta} = \tfrac{1}{2}(v_{\alpha;\beta} - v_{\beta;\alpha}),$$

and remembering that the motion is expansionless and geodesic, it is not difficult to obtain from (9) the relation

$$R_{\alpha\beta}v^\alpha v^\beta = -2\omega^2, \tag{10}$$

where $\omega_{\alpha\beta}\omega^{\alpha\beta} = +2\omega^2$. In view of (6), (7), and (10), we have

$$R^\alpha_\beta v^\beta = -2\omega^2 v^\alpha + 2S^\alpha. \tag{11}$$

Let us now use the following identity [cf. Eq. (19) of Ref. 4] in our calculation:

$$\begin{aligned}&\tfrac{2}{3}(v^\beta_{;\beta})_{,\sigma}g^{\sigma\alpha} - R^\alpha_\beta v^\beta \\ &= \varphi^{\rho\alpha}_{;\rho} + \varphi^{\rho\alpha}\dot{v}_\rho + [\dot{v}^\rho_{;\rho} - \tfrac{1}{3}(v^\beta_{;\beta})^2 - \tfrac{1}{3}(v^\beta_{;\beta})_{,\mu}v^\mu + 2\omega^2]v^\alpha \\ &\quad + \eta^{\mu\nu\rho\alpha}(\omega_{\mu,\rho}v_\nu - 2\omega_\mu\dot{v}_\rho v_\nu),\end{aligned} \tag{12}$$

where $\dot{v}^\rho = v^\rho_{;\beta}v^\beta$ stands for the acceleration vector and $\varphi^{\rho\alpha}$ represents the shear tensor defined in the usual manner. ω_μ represents the rotation vector defined by $\omega^\mu = \tfrac{1}{2}\eta^{\mu\alpha\beta\sigma}v_{\alpha;\beta}v_\sigma$.

Combining (11) and (12) and remembering that the motion is rigid, one gets the relation

$$S^\alpha = -\tfrac{1}{2}\eta^{\mu\nu\rho\alpha}\omega_{\mu,\rho}v_\nu. \tag{13}$$

The null field vanishes if the vector S vanishes. Hence, for a nontrivial null-field solution, we must have in view of (13) a nonvanishing vorticity and, further, the curl of the vorticity must not also vanish. In deriving Eq. (13), which does not involve the dust density, use has been made of Eq. (3). When $\rho \neq 0$ and if space–time there admits a motion with a generator other than the velocity, this generator may be hypersurface orthogonal. Such cases actually occur in the Ozsvath solutions and are discussed later on.

We show further that if there is a hypersurface-orthogonal Killing vector l_μ for a pure (i.e., matter-free) radiation field, it will be the null propagation vector.

From the Killing equation

$$l_{\mu;\nu} + l_{\nu;\mu} = 0 \tag{14}$$

and from the condition that l_μ is hypersurface orthogonal, i.e.,

$$(l_{\mu;\nu} - l_{\nu;\mu})l_\alpha + (l_{\nu;\alpha} - l_{\alpha;\nu})l_\mu + (l_{\alpha;\mu} - l_{\mu;\alpha})l_\nu = 0, \tag{15}$$

one gets

$$2l_{\mu;\nu}l^2 = (l^2)_{;\nu}l_\mu - (l^2)_{;\mu}l_\nu, \tag{16}$$

which again, on further differentiation and using (14), yields

$$2l^\mu_{;\nu\mu}l^2 = -2l_{\mu;\nu}(l^2)^{;\mu} - (l^2)^{;\mu}_{;\mu}l_\nu. \tag{17}$$

Again from (16), we have

$$l_{\mu;\nu}(l^2)^{;\mu} = -\tfrac{1}{2}l^{-2}(l^2)_{;\mu}(l^2)^{;\mu}l_\nu,$$

which, after substitution in (17), finally gives

$$R_{\nu\mu}l^\mu = -l^\mu_{;\nu\mu} = \tfrac{1}{2}l^{-2}[(l^2)^{;\mu}_{;\mu} - l^{-2}(l^2)^{;\mu}(l^2)_{;\mu}]l_\nu. \tag{18}$$

Again, from (2) and (22) given below, we have

$$R_{\nu\mu}l^\mu = -2(K_\mu l^\mu)K_\nu + \lambda l_\nu. \tag{19}$$

[4] S. Banerji, Progr. Theoret. Phys. (Kyoto) **39**, 365 (1968).

Comparing (18) and (19), one has to conclude that l_ν is a null vector of the form aK_ν.

Now, as is well known, the null electromagnetic field tensor may be written in the form

$$F_{\mu\nu} = (K_\mu \xi_\nu - K_\nu \xi_\mu), \tag{20}$$

where the null vector K_μ is the propagation vector and ξ_μ is a spacelike vector orthogonal to K_μ. We can further choose ξ_μ, without loss of generality, to be a unit vector. In that case,

$$E^2 = -E_\mu E^\mu = (K_\nu v^\nu)^2, \tag{21}$$

and the electromagnetic energy–momentum tensor can be written in the form

$$T_{\alpha\beta} = (4\pi)^{-1} K_\alpha K_\beta . \tag{22}$$

From Eqs. (3) and (22), we get

$$T_{\alpha\beta} v^\beta = (4\pi)^{-1}[E^2 v_\alpha - S_\alpha] \tag{23}$$

and

$$S^\alpha = E^2 v^\alpha - K^\alpha (K_\beta v^\beta). \tag{24}$$

Let us now introduce the additional assumption that the energy density of the radiation field is constant which could mean, in view of (21), that $K_\beta v^\beta$ is constant. Equation (24) then gives, using Eq. (5),

$$K^\alpha_{;\alpha} = 0. \tag{25}$$

Also $T^{\alpha\beta}_{;\beta} = 0$ owing to the charge–current vector being zero, so that from Eq. (22) using Eq. (25) we have

$$K^\alpha_{;\beta} K^\beta = 0. \tag{26}$$

Thus, Eqs. (25) and (26) show that K^α constitutes a divergence-free null-geodesic congruence. Hence, from Robinson's[5] theorem on null electromagnetic fields, it follows that $K_{(\alpha;\beta)} = 0$. Using this in $(v_\beta K^\beta)_{,\alpha} = 0$ and remembering, in view of (22) and (2), that K^β is an invariant vector for the transformation with the generator v^α such that

$$K_{\alpha;\beta} v^\beta = -v_{\alpha;\beta} K^\beta, \tag{27}$$

we get

$$v_{\alpha;\beta} K^\beta = 0. \tag{28}$$

Using now the definition of the vorticity vector ω^μ, we have

$$\eta_{\alpha\beta\mu\sigma} K^\beta \omega^\mu v^\sigma = 0. \tag{29}$$

In view of relations (24), Eq. (29) can be written as

$$\eta_{\alpha\beta\mu\sigma} S^\beta \omega^\mu v^\sigma = 0. \tag{30}$$

Since S^β and ω^μ are both orthogonal to v^σ, the above shows that we must have a relation of the form $S^\beta = a\omega^\beta$; in other words, the vectors S and ω are in the same direction. However, the vector S is orthogonal to E and H. Hence, the vorticity, the electric, and magnetic field vectors form an orthogonal set of spacelike vectors if the motion of the fluid is rigid and the energy density of radiation is uniform.

Again, since $K^\alpha_{;\alpha} = 0$, $K^\alpha_{;\beta} K^\beta = 0$, and $K_{(\alpha;\beta)} = 0$ in this particular case, we have, using these in the identity (8) with v^α replaced by K^α,

$$R_{\alpha\beta} K^\alpha K^\beta = K^\alpha_{;\beta} K^\beta_{;\alpha}.$$

From (2), substituting for $R_{\alpha\beta}$ and remembering (23), it follows that

$$K^\alpha_{;\beta} K^\beta_{;\alpha} = -8\pi\rho (K^\alpha v_\alpha)^2, \tag{31}$$

that is,

$$K_{[\alpha;\beta]} K^{\alpha;\beta} = 16\pi\rho (K^\alpha v_\alpha)^2, \tag{32}$$

which implies that rays have a nonvanishing curl in presence of matter, and the magnitude of this curl is determined by the product of matter density and radiation density.

Let us now proceed to discuss the form of the electromagnetic field tensor in these cases in the light of their behavior under "duality rotation." Since $R_{\alpha\beta}$, ρ, v_α, and $g_{\alpha\beta}$ are independent of time, so is the electromagnetic stress-energy tensor $T_{\alpha\beta}$ in view of Eq. (2). Now $T_{\alpha\beta}$ determines the electromagnetic field tensor $F_{\alpha\beta}$ up to a "duality rotation." Hence, $F_{\alpha\beta}$ is to be obtained from a time-independent $F_{\alpha\beta}$ by a duality rotation. Writing the time-independent $F_{\alpha\beta}$ as $(F_{\alpha\beta})_0$ and $\omega_{\alpha\beta}$ for $(F_{\alpha\beta} - i{}^*F_{\alpha\beta})$, we have $\omega_{\alpha\beta} = (\omega_{\alpha\beta})_0 e^{i\theta}$ and the Maxwell equation $\omega^{\alpha\beta}_{;\beta} = 0$ gives

$$(\omega^{\alpha\beta})_{0;\beta} + i(\omega^{\alpha\beta})_0 \theta_{,\beta} = 0. \tag{33}$$

Equation (27) shows that there must be a solution with $\theta_{,\beta}$ independent of time or $\theta = at + b$, where a is a constant and b is a function of the space coordinates but not of time. Further, as we shall see, the solution in this case is unique, so that

$$\begin{aligned} F_{\alpha\beta} &= (F_{\alpha\beta})_0 \cos\theta + ({}^*F_{\alpha\beta})_0 \sin\theta \\ &= A_{\alpha\beta} \cos(at + b), \end{aligned} \tag{34}$$

which is the form obtained by Ozsvath. Here it may be noted that, while the coordinate system we have in mind is a comoving one so that the t lines are the world lines of matter, Ozsvath's coordinates are not comoving and the t lines are the paths of hypersurface-orthogonal timelike translation.

We now proceed to discuss the uniqueness of these electromagnetic fields. It has been pointed out by Witten[6] that the electromagnetic field tensor for a

[5] I. Robinson, "Reports on Royamount Conference, 1959."

[6] L. Witten, in *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962).

given metric is, in general, not unique in the null case. However, in the solutions that we are considering, the field is unique. Let $\omega^{\mu\nu}$ ($\equiv F^{\mu\nu} - i^*F^{\mu\nu}$) and $\omega'^{\mu\nu}$ be two different solutions pertaining to the same $T^{\mu\nu}$, both satisfying Maxwell's equation. They must be connected by a duality rotation so that $\omega'^{\mu\nu} = \omega^{\mu\nu}e^{i\alpha}$.

Now the condition that $\omega'^{\mu\nu}_{;\nu} = \omega^{\mu\nu}_{;\nu} = 0$ gives

$$\omega^{\mu\nu}\alpha_{,\nu} = 0. \tag{35}$$

Also, $\omega^{\mu\nu} = K^\mu\zeta^\nu - K^\nu\zeta^\mu$, where ζ is the complex vector $\zeta^\alpha = \xi^\alpha - i\eta^\alpha$ and ξ and η are unit spacelike vectors orthogonal to each other and to K.

Thus

$$K^\mu\zeta^\nu\alpha_{,\nu} - K^\nu\alpha_{,\nu}\zeta^\mu = 0, \tag{36}$$

whence we get

$$K^\nu\alpha_{,\nu} = \xi^\nu\alpha_{,\nu} = \eta^\nu\alpha_{,\nu} = 0. \tag{37}$$

Remembering that K is a null vector, the above requires

$$\alpha_{,\nu} = CK_\nu, \tag{38}$$

where C is a scalar. Equation (38) is integrable if and only if K_ν is hypersurface orthogonal, which is, however, not the case in Ozsvath's solutions, so that there is only the trivial solution, $\alpha = \text{const}$.

Lastly we consider the possibility of the existence of a timelike generator which is hypersurface orthogonal. If we write u^α for this timelike vector, then $u_{\alpha;\beta} = 0$ and hence we get, remembering (8) and using (2) and (22),

$$0 = -8\pi\rho(v_\alpha u^\alpha)v_\beta + (\lambda + 4\pi\rho)u_\beta - 2K_\alpha u^\alpha K_\beta, \tag{39}$$

which yields

$$-8\pi\rho(v_\alpha u^\alpha)^2 + (\lambda + 4\pi\rho)u^2 - 2(K_\alpha u^\alpha)^2 = 0, \tag{40}$$

$$(\lambda + 4\pi\rho)(v^\alpha u_\alpha) - 2(K_\alpha v^\alpha)(K_\beta u^\beta) = 0, \tag{41}$$

$$-8\pi\rho(v_\alpha u^\alpha)(K^\beta v_\beta) + (\lambda + 4\pi\rho)(K^\beta u_\beta) = 0, \tag{42}$$

after contraction with u^β, v^β, and K^β, respectively. Solving the above equations, we get

$$\lambda^2 - 16\pi^2\rho^2 = 16\pi\rho(K^\beta v_\beta)^2,$$
$$(\lambda^2 - 16\pi^2\rho^2)u^2 = 4\lambda(K^\beta u_\beta)^2,$$

and

$$[(\lambda + 4\pi\rho)^2/16\pi\rho\lambda]u^2 = (v^\alpha u_\alpha)^2. \tag{43}$$

It is clear from (43) that u_α can exist only if $\lambda > 4\pi\rho > 0$.

3. TWO SPECIAL NULL-FIELD SOLUTIONS

In this part we present two null-field solutions. We shall start with the condition that the space–time admits a null translation with the propagation vector as the generator, so that the Killing equation

$$K_{\mu;\nu} + K_{\nu;\mu} = 0 \tag{44}$$

is satisfied. Let us choose a coordinate system such that this motion is a translation in the null direction x^1, so that $K^\mu = \delta^\mu_1$. We shall further specialize the line element to the comparatively simple form

$$ds^2 = f(dx^0)^2 + 2a\,dx^0\,dx^1 - m(dx^2)^2 - n(dx^3)^2, \tag{45}$$

where f, m, n, and a are functions of x^0 and x^2. Besides the null translation, the space–time of (45) admits a translation along x^3.

We get the nonvanishing components of $R_{\mu\nu}$ by straightforward calculation for the line-element (45):

$$\begin{aligned}
R_{00} &= \frac{1}{2}\frac{\ddot{m}}{m} + \frac{1}{2}\frac{\ddot{n}}{n} - \frac{1}{4}\left(\frac{\dot{m}}{m}\right)^2 - \frac{1}{4}\left(\frac{\dot{n}}{n}\right)^2 \\
&\quad - \frac{1}{2}\frac{f_{22}}{m} + \frac{1}{4}\frac{f_2 m_2}{(m)^2} + \frac{1}{2}\frac{f_2 a_2}{m\cdot a} \\
&\quad - \frac{1}{2}\frac{f}{m}\left(\frac{a_2}{a}\right)^2 - \frac{1}{2}\frac{\dot{a}}{a}\cdot\frac{\dot{m}}{m} - \frac{1}{2}\frac{\dot{a}}{a}\cdot\frac{\dot{n}}{n} - \frac{1}{4}\frac{f_2 n_2}{m\cdot n}, \\
R_{22} &= \frac{a_{22}}{a} - \frac{1}{2}\left(\frac{a_2}{a}\right)^2 + \frac{1}{2}\frac{n_{22}}{n} - \frac{1}{4}\left(\frac{n_2}{n}\right)^2 \\
&\quad - \frac{1}{2}\left(\frac{m_2}{m}\right)\left(\frac{a_2}{a}\right) - \frac{1}{4}\left(\frac{m_2}{m}\right)\left(\frac{n_2}{n}\right), \\
R_{33} &= \frac{1}{2}\frac{n_{22}}{m} - \frac{1}{4}\frac{m_2 n_2}{(m)^2} - \frac{1}{4}\frac{(n_2)^2}{m\cdot n} + \frac{1}{2}\frac{a_2}{a}\cdot\frac{n_2}{m}, \\
R_{01} &= -\frac{1}{2}\frac{a_{22}}{m} + \frac{1}{4}\frac{a_2 m_2}{(m)^2} - \frac{1}{4}\frac{a_2 n_2}{m\cdot n},
\end{aligned} \tag{46}$$

where the dot indicates differentiation with respect to x^0 and the subscript 2 denotes differentiation with respect to the x^2 coordinate. Assuming that we have a matter-free pure-radiation field and that $\lambda = 0$, Eq. (2) gives

$$R_{\mu\nu} = -2K_\mu K_\nu. \tag{47}$$

Equations (32) and (44) together now give

$$K_{\mu;\nu} = 0. \tag{48}$$

From (47) it follows that $R_{00} = -2a^2$ and $R_{22} = R_{23} = R_{01} = 0$. Relation (48) shows that a is at most a function of x^0 alone. Without any loss of generality, we may, therefore, take a to be a constant. The second and third equations of (46) are actually equivalent and can be written as

$$\frac{n_{22}}{n} - \frac{1}{2}\left(\frac{n_2}{n}\right)^2 - \frac{1}{2}\frac{m_2}{m}\cdot\frac{n_2}{n} = 0, \tag{49}$$

which gives the relation, with $n_2 \neq 0$,

$$n_2/m^{\frac{1}{2}}n^{\frac{1}{2}} = \beta, \tag{50}$$

where β is a function of x^0 alone. The other independent field equation is

$$\frac{1}{2}\frac{\ddot{m}}{m} + \frac{1}{2}\frac{\ddot{n}}{n} - \frac{1}{4}\left(\frac{\dot{m}}{m}\right)^2 - \frac{1}{4}\left(\frac{\dot{n}}{n}\right)^2 - \frac{1}{2}\frac{f_{22}}{m} + \frac{1}{4}\frac{f_2 m_2}{(m)^2} - \frac{1}{4}\frac{f_2 n_2}{m \cdot n} = -2a^2. \tag{51}$$

Equations (50) and (51) provide only two relations between three unknown functions f, m, and n, and thus the system is undetermined. However, the electromagnetic-field tensor is determined up to a duality rotation and is given by

$$F^{12} = m^{-\frac{1}{2}} \sin\left[\tfrac{1}{2}\beta x^3 + \psi(x^0)\right]$$

and

$$F^{13} = n^{-\frac{1}{2}} \cos\left[\tfrac{1}{2}\beta x^3 + \psi(x^0)\right], \tag{52}$$

where $\psi(x^0)$ is an arbitrary function of x^0 and β is given by Eq. (50).

We shall give two specially simple solutions. In both of them, $m = n$. Further, in the first one the x^0 lines are also null lines so that $f = 0$, and from (50) and (51) we then get

$$ds^2 = 2a\, dx^0\, dx^1 - e^{\beta x^2} \cos^2(ax^0)[(dx^2)^2 + (dx^3)^2]; \tag{53}$$

β is constant in this case. The electromagnetic field tensors are, from (53),

$$F^{12} = e^{-\frac{1}{2}\beta x^2} \sec(ax^0) \sin\left[\tfrac{1}{2}\beta x^3 + \psi(x^0)\right],$$
$$F^{13} = e^{-\frac{1}{2}\beta x^2} \sec(ax^0) \cos\left[\tfrac{1}{2}\beta x^3 + \psi(x^0)\right]. \tag{54}$$

The second solution that we shall consider is one in which the space–time admits a further motion along the x^0 lines. The $g_{\mu\nu}$ would then be a function of x^2 alone and, from (46) and (51), we find

$$n_2/m^{\frac{1}{2}}n^{\frac{1}{2}} = \beta \tag{55}$$

and

$$-\frac{1}{2}\frac{f_{22}}{m} + \frac{1}{4}\frac{f_2 m_2}{(m)^2} - \frac{1}{4}\frac{f_2 n_2}{m \cdot n} = -2a^2, \tag{56}$$

where β is now a constant.

A particularly simple solution of the above is given by

$$m = n = e^{\beta x^2}, \quad f = (4a^2/\beta^2)e^{\beta x^2}, \tag{57}$$

so that the line element can be written as

$$ds^2 = e^{\beta x^2}[(dx^0)^2 - (dx^2)^2 - (dx^3)^2] + \beta\, dx^0\, dx^1, \tag{58}$$

and the corresponding electromagnetic field tensors are, from (52),

$$F^{12} = e^{-\frac{1}{2}\beta x^2} \sin\left[\tfrac{1}{2}\beta x^3 + \psi(x^0)\right],$$
$$F^{13} = e^{-\frac{1}{2}\beta x^2} \cos\left[\tfrac{1}{2}\beta x^3 + \psi(x^0)\right]. \tag{59}$$

Here again, $\psi(x^0)$ is an arbitrary function of x^0. The null solutions presented above illustrate beautifully the nonuniqueness of the electromagnetic field tensors. As in this case, the propagation vector is hypersurface orthogonal, any duality rotation through α where $\alpha_{,\nu} = CK_\nu$ is permissible [see Eq. (38)], and as in our case $K_\nu = a\delta^0_\nu$ gives $\alpha = \psi(x^0)$ where $\psi(x^0)$ is an arbitrary function of x^0. The case of the second solution is particularly interesting. With $\psi(x^0) = \text{const}$, this corresponds to a static field, the x^0 coordinate being interpreted as the time coordinate; hence, the field is nonradiating, whereas with $\psi(x^0) = x^0$, the electromagnetic field is oscillatory and simulates a radiation field.

ACKNOWLEDGMENT

The author is indebted to Professor A. K. Raychaudhuri of Presidency College, Calcutta, for his valuable guidance throughout the entire work.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Fourier Expansions of Functions of the Distance between Two Points

A. R. SOUROUR AND A. A. ASHOUR
Mathematics Department, Faculty of Science, Cairo University, Cairo, UAR

(Received 21 November 1968)

The expansion of any function $f(r)$ of the distance r between two points is given as a Fourier series. This is a generalization of results given earlier by Ashour for r^n and $\log r$. The Fourier-series expansion for the product $r^N e^{iM\theta}$ is also given, where now $\mathbf{r} = (r, \theta)$ denotes the sum of $\mathbf{r}_1$ and $\mathbf{r}_2$, the coordinate vectors of the two points. This is further generalized for the product of a function $f(r)$ of r and a circular harmonic. Expansions of similar products involving spherical harmonics have been given earlier by Sack. Special cases when $f(r)$ is a Bessel function or a modified Bessel function are considered.

1. INTRODUCTION

The expansions of arbitrary powers and functions of the distance between two points as series involving Legendre functions have been considered by Sack[1] and earlier by Chapman.[2] Ashour[3] followed Sack's method and obtained an expansion for the general power

$$r^n = (r_1^2 - 2r_1r_2\cos\omega + r^2)^{\frac{1}{2}n}$$

as a Fourier series

$$\sum R_{n,l}(r_1, r_2)\cos l\omega,$$

where r_1, θ_1, ϕ_1 and r_2, θ_2, ϕ_2 are the spherical polar coordinates of the two points and

$$\cos\omega = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\phi_1 - \phi_2).$$

Ashour also gave the Fourier expansion for $\log r$. Clearly, such expansions are appropriate in two-dimensional applications. In the present paper we give a generalization of Ashour's results to any function $f(r)$ of r which can be expressed as a power series in r. This is given in Sec. 2, together with the treatment of certain special cases, namely when $f(r)$ is (i) a Bessel function of zero order, (ii) a modified Bessel function of zero order, and (iii) a Gaussian function.

In a second paper, Sack[4] derived an expansion for the product of a power r^N of r and a surface spherical harmonic $Y_L^M(\theta, \phi)$, where r, θ, ϕ are the coordinates of the point whose position vector is $\mathbf{r}$. Sack's expansion was given in terms of the spherical harmonics of (θ_1, ϕ_1) and (θ_2, ϕ_2), where now $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$ and $\mathbf{r}_1$, $\mathbf{r}_2$ are the position vectors of the points (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2), respectively. In Sec. 3 of the present paper we give the corresponding Fourier expansion for the product of r^N and a circular harmonic $e^{iM\theta}$ where the vector $\mathbf{r} = (r, \theta)$ is the sum of the position vectors of the points (r_1, θ_1) and (r_2, θ_2). This is further generalized in Sec. 4 to products of a circular harmonic and any function $f(r)$ which is representable as a power series in r. Special cases, namely when $f(r)$ is a Bessel function of the first, second, or third kind, or a modified Bessel function of either the first or second kind, are considered in Sec. 5.

For brevity, the notation of Refs. 1, 3, and 4 is followed here.

2. EXPANSION FOR $f(r)$

A. Expansion for Arbitrary Functions $f(r)$

Ashour[3] gave the following expression for r^n:

$$r^n = \sum R_{nl}(r_1, r_2)\cos l\omega, \tag{1}$$

where three different expressions were given for $R_{nl}(r_1, r_2)$, namely,

$$R_{nl}(r_1, r_2) = [(2 - \delta_l^0)/l!](-\tfrac{1}{2}n)r_>^n(r_</r_>)^l \times F(-\tfrac{1}{2}n, l - \tfrac{1}{2}n; l + 1; (r_</r_>)^2), \tag{2a}$$

$$R_{nl}(r_1, r_2) = [(2 - \delta_l^0)/l!](-\tfrac{1}{2}n)_l(r_1r_2)^l(r_1^2 + r_2^2)^{\frac{1}{2}n-l} \times F(\tfrac{1}{2}l - \tfrac{1}{4}n, \tfrac{1}{2}l - \tfrac{1}{4}n + \tfrac{1}{2}; l + 1; 4r_1^2r_2^2/(r_1^2 + r_2^2)^2), \tag{2b}$$

and

$$R_{nl}(r_1, r_2) = [(2 - \delta_l^0)/l!](-\tfrac{1}{2}n)_l(r_1r_2)^l(r_1 + r_2)^{n-2l} \times F(l - \tfrac{1}{2}n, l + \tfrac{1}{2}; 2l + 1; 4r_1r_2/(r_1 + r_2)^2). \tag{2c}$$

Following the method adopted by Sack,[1] each of these expressions will be written in an operational form which, in turn, yields an expansion for any function $f(r)$ which can be expressed as a power series $\sum c_n r^n$. Using the relation

$$(-\tfrac{1}{2}n)_s(-\tfrac{1}{2}n)_{l+s}r_>^{n-2s} = \frac{(-1)^l r_>^{2l}}{2^{l+2s}}\left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l\left[\frac{1}{r_>}\frac{\partial}{\partial r_>}\left(r_>\frac{\partial}{\partial r_>}\right)\right]^s r_>^n, \tag{3}$$

[1] R. A. Sack, J. Math. Phys. **5**, 245 (1964).
[2] S. Chapman, Quart. J. Pure Appl. Math. **185**, 16 (1916).
[3] A. A. Ashour, J. Math. Phys. **6**, 492 (1965).
[4] R. A. Sack, J. Math. Phys. **5**, 252 (1964).

Eq. (2a) transforms into

$$R_{nl}(r_1, r_2) = (2 - \delta_l^0)\left(\frac{r_<}{r_>}\right)^l \sum_{s=0}^{\infty} \frac{(-1)^l r_<^{2s} r_>^{2l}}{2^{l+2s} s!\,(l+s)!} \times \left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l \left[\frac{1}{r_>}\frac{\partial}{\partial r_>}\left(r_>\frac{\partial}{\partial r_>}\right)\right]^s r_>^n. \quad (4)$$

Hence, any function $f(r) = \sum c_n r^n$ can be expanded as

$$f(r) = \sum_{l=0}^{\infty} f_l(r_1, r_2)\cos l\omega, \quad (5)$$

where

$$f_l = (2 - \delta_l^0)(-r_< r_>)^l \left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l \times \sum_{s=0}^{\infty} \frac{r_<^{2s}}{(2s)!!\,(2l+2s)!!}\left[\frac{1}{r_>}\frac{\partial}{\partial r_>} r_> \frac{\partial}{\partial r_>}\right]^s f(r_>)$$

$$= (2 - \delta_l^0)(-r_< r_>)^l \left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l \times \left\{ I_l\left(r_<\left[\frac{1}{r_>}\frac{\partial}{\partial r_>}\left(r_>\frac{\partial}{\partial r_>}\right)\right]^{\frac{1}{2}}\right) \Big/ r_<^l \left[\frac{1}{r_>}\frac{\partial}{\partial r_>}\left(r_>\frac{\partial}{\partial r_>}\right)\right]^{\frac{1}{2}l} \right\} f(r_>), \quad (6)$$

where $n!!$ is defined as in Eq. (43) of Ref. 1, and $I_l(z)$ is the modified Bessel function of the first kind defined in Ref. 5, Sec. (B7.2.2). Similarly, (2b) and (2c) show that

$$f_l(r_1, r_2) = (2 - \delta_l^0)\sum_{s=0}^{\infty} \frac{1}{(2s)!!\,(2l+2s)!!}\left(-\frac{r_1 r_2}{\rho}\frac{\partial}{\partial \rho}\right)^{l+2s} f(\rho)$$

$$= (2 - \delta_l^0) I_l\left(-\frac{r_1 r_2}{\rho}\frac{\partial}{\partial \rho}\right) f(\rho) \quad (7)$$

and

$$f_l(r_1, r_2) = \left(\frac{(2 - \delta_l^0)}{(2l)!!}\right)\left(-\frac{r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right)^l \times \Phi\left(l + \tfrac{1}{2}; 2l + 1; -\frac{2r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right) f(r_+), \quad (8)$$

where $\rho = (r_1^2 + r_2^2)^{\frac{1}{2}}$, $r_+ = r_1 + r_2$, and Φ is the confluent hypergeometric function (B6). In both (7) and (8) the product $r_1 r_2$ is to be treated as a constant on differentiation. The equivalence of (7) and (8) follows from the connection of $\phi(\alpha; 2\alpha; 2z)$ and $I_l(z)$, namely,

$$I_n(z) = [(\tfrac{1}{2}z)^n/\Gamma(n+1)]e^{-z}\Phi(\tfrac{1}{2} + n; 1 + 2n; 2z).$$

[5] A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill Book Co., Inc., New York, 1953). Sections and formulas in this reference are prefixed by the letter B. All references (Bl.m.n.) are to equations unless otherwise specified.

B. Expansion for Certain Functions $f(r)$

The expansions (6), (7), and (8) factorize for certain functions $f(r)$ as follows.

(1) Consider functions satisfying the differential equation

$$\nabla^2 f = \frac{1}{r}\frac{d}{dr}\left(r\frac{d}{dr}\right) f = -k^2 f, \quad (9)$$

i.e., f is a Bessel function of order zero of the first, second, or third kind,

$$J_0(kr), \quad Y_0(kr), \quad H_0^{(1)}(kr), \quad H_0^{(2)}(kr).$$

Using (B7.2.52) and (B7.2.53), we obtain from Eq. (6)

$$f_l = (2 - \delta_l^0)\sum_{s=0}^{\infty} \frac{(-1)^s (kr_<)^{l+2s}}{(2s)!!\,(2l+2s)!!} W_l(kr_>) = (2 - \delta_l^0) J_l(kr_<) W_l(kr_>), \quad (10)$$

where $W = J, Y, H^{(1)}, H^{(2)}$. Thus,

$$W_0(kr) = \sum_l (2 - \delta_l^0) J_l(kr_<) W_l(kr_>)\cos l\omega, \quad (11)$$

which is exactly the addition theorem (B7.15.29).

(2) Consider functions satisfying

$$\nabla^2 f = \frac{1}{r}\frac{d}{dr}\left(r\frac{d}{dr}\right) f = k^2 f, \quad (12)$$

i.e., f is a modified Bessel function of order zero of the first or second kind, $I_0(kr)$ or $K_0(kr)$. Using (B7.11.19)–(B7.11.22) and Eq. (6), we find that

$$I_0(kr) = \sum_{l=0}^{\infty} (-1)^l (2 - \delta_l^0) I_l(kr_<) I_l(kr_>)\cos lw \quad (13a)$$

and

$$K_0(kr) = \sum_{l=0}^{\infty} (2 - \delta_l^0) I_l(kr_<) K_l(kr_>)\cos l\omega, \quad (13b)$$

which are special cases of (B7.15.36) and (B7.15.35), respectively.

(3) If $f(r)$ is a Gaussian function, i.e.,

$$f(r) = e^{-kr^2},$$

$$\left(\frac{1}{r}\frac{d}{dr}\right) f(r) = -2kf(r), \quad (14)$$

then expansion (7) or (8) factorizes, yielding the result

$$\exp(-kr^2) = \sum_{l=0}^{\infty} (2 - \delta_l^0) I_l(2kr_1 r_2) \times \exp[-k(r_1^2 + r_2^2)]\cos l\omega. \quad (15)$$

If we divide by $\exp[-k(r_1^2 + r_2^2)]$, we get

$$\exp(2kr_1 r_2 \cos\omega) = \sum_{l=0}^{\infty} (2 - \delta_l^0) I_l(2kr_1 r_2)\cos l\omega \quad (16)$$

or, by replacing k by ik,

$$\exp(2ikr_1r_2\cos\omega) = \sum_{l=0}^{\infty} i^l(2-\delta_l^0)J_l(2kr_1r_2)\cos l\omega, \tag{17}$$

which is equivalent to (B7.2.27).

3. EXPANSION OF $r^N e^{iM\theta}$

Here $\mathbf{r} = (r, \theta)$ is the sum of the vectors $\mathbf{r}_1 = (r_1, \theta_1)$ and $\mathbf{r}_2 = (r_2, \theta_2)$, and we assume that $r_2 > r_1$. The expansion of $V_{NM} = r^N e^{iM\theta}$ will be given in the form

$$V_{NM} = \sum_{m_1,m_2} R(N, M, m_1, m_2; r_1, r_2)e^{im_1\theta_1}e^{im_2\theta_2}. \tag{18}$$

M is considered to be a positive integer; the case of negative M could be found by taking the complex conjugate of (18). In order to obtain such an expansion, we write V_{NM} in the form

$$V_{NM} = r^n(re^{i\theta})^M, \quad n = N - M. \tag{19}$$

We will give expansions of both r^n and $(re^{i\theta})^M$ in absolutely convergent series; hence, the two series can be multiplied to obtain the required expansion.

The expansion of $r^n = |\mathbf{r}_1 + \mathbf{r}_2|^n$ could be obtained from the expansion of $|\mathbf{r}_1 - \mathbf{r}_2|^n$ given by (1) and (2) upon replacing θ_2 by $\pi + \theta_2$; hence,

$$\begin{aligned} r^n &= \sum_{l=0}^{\infty}(-1)^l R_{nl}(r_1, r_2)\cos[l(\theta_1 - \theta_2)] \\ &= \sum_{l=-\infty}^{+\infty} R'_{nl}(r_1, r_2)e^{il\theta_1}e^{-il\theta_2}, \end{aligned} \tag{20}$$

where

$$R'_{nl}(r_1, r_2) = (-1)^l r_2^n\left(\frac{r_1}{r_2}\right)^{|l|}\sum\frac{(-\frac{1}{2}n)_s(-\frac{1}{2}n)_{s+|l|}}{s!\,(s+|l|)!}\left(\frac{r_1}{r_2}\right)^{2s}. \tag{21}$$

The expansion of $r^M e^{iM\theta} = (r_1e^{i\theta_1} + r_2e^{i\theta_2})^M$ is given, by use of the binomial theorem, as

$$r^M e^{iM\theta} = \sum_{\mu=0}^{M}\binom{M}{\mu}r_1^\mu r_2^{M-\mu}e^{i\mu\theta_1}e^{i(M-\mu)\theta_2}. \tag{22}$$

Multiplying (20) and (22) and rearranging terms, we obtain

$$\begin{aligned} &r^N e^{iM\theta} \\ &= \sum_{m=-\infty}^{\infty}\sum_{\mu=0}^{M}\binom{M}{\mu}r_1^\mu r_2^{M-\mu}R'_{n,m-\mu}(r_1, r_2)e^{im\theta_1}e^{i(M-m)\theta_2}. \end{aligned} \tag{23}$$

Thus, in the expansion (18), the radial function $R(N, M, m_1, m_2; r_1, r_2)$ is given by

$$R(N, M, m_1, m_2; r_1, r_2) = 0, \quad \text{if} \quad m_1 + m_2 \neq M, \tag{24a}$$

$$\begin{aligned} R_N &= R(N, M, m_1, M - m_1; r_1, r_2) \\ &= \sum_{s=0}^{\infty}\sum_{\mu=0}^{M}\binom{M}{\mu}(-1)^{m_1-\mu} \\ &\quad\times\frac{(-\frac{1}{2}n)_s(-\frac{1}{2}n)_{s+|m_1-\mu|}}{s!\,(s+|m_1-\mu|)!}\,r_2^N\left(\frac{r_1}{r_2}\right)^{2s+\mu+|m_1-\mu|}. \end{aligned} \tag{24b}$$

The upper limit in the μ series can be replaced by any number $> M$ since $\binom{M}{\mu} = 0$ for $\mu > M$. This will be done frequently without being mentioned explicitly.

In order to simplify (24b), we consider three cases, namely, $m_1 \geq M$, $0 < m_1 < M$, and $m_1 \leq 0$.

(1) $m_1 \geq M$: In this case $m_1 - \mu > 0$, and (24b) becomes

$$\begin{aligned} R_N &= r_2^N\sum_{s=0}^{\infty}\frac{(-1)^s(-\frac{1}{2}n)_2}{s!}\left(\frac{r_1}{r_2}\right)^{m_1+2s} \\ &\quad\times\left[\sum_{\mu=0}^{m_1+s}\frac{(M;\mu)}{\mu!}\frac{(\frac{1}{2}n; s+m_1-\mu)}{(s+m_1-\mu)!}\right], \end{aligned} \tag{25}$$

where

$$(x;\alpha) = x(x-1)\cdots(x-\alpha+1) = (-1)^\alpha(-x)_\alpha. \tag{26}$$

Using Vandermonde's theorem, namely,

$$\sum_{\alpha=0}^{m}\frac{(x;\alpha)(y;m-\alpha)}{\alpha!\,(m-\alpha)!} = \frac{(x+y;m)}{m!}, \tag{27}$$

the μ series in (25) sums up to

$$(-1)^{s+m}(-\tfrac{1}{2}n - M)_{s+m_1}/(s+m_1)!.$$

Hence,

$$R_N = (-1)^{m_1}r_2^N\left(\frac{r_1}{r_2}\right)^{m_1}\sum_{s=0}^{\infty}\frac{(-\frac{1}{2}n)(-\frac{1}{2}n-M)_{s+m_1}}{s!\,(s+m_1)}\left(\frac{r_1}{r_2}\right)^{2s}. \tag{28}$$

(2) $0 < m_1 < M$: In this case, Eq. (24b) can be written as

$$R_N = \sum_{s=0}^{\infty}\sum_{\mu=0}^{m_1} + \sum_{s=0}^{\infty}\sum_{\mu=m_1+1}^{M} = R_N^{(1)} + R_N^{(2)}, \tag{29}$$

where the missing summands are the same as in (24b). Then

$$\begin{aligned} R_N^{(1)} &= r_2^N\left(\frac{r_1}{r_2}\right)^{m_1}\sum_{s=0}^{\infty}(-1)^s\frac{(-\frac{1}{2}n)_s}{s!}\left(\frac{r_1}{r_2}\right)^{2s} \\ &\quad\times\sum_{\mu=0}^{m_1}\frac{(M;\mu)}{\mu!}\frac{(\frac{1}{2}n; s+m_1-\mu)}{(s+m_1-\mu)!} \end{aligned} \tag{30}$$

and

$$\begin{aligned} R_N^{(2)} &= r_2^N\left(\frac{r_1}{r_2}\right)^{m_1}\sum_{s=0}^{\infty}\sum_{\mu=m_1+1}^{\infty}(-1)^s\frac{(M;\mu)}{\mu!} \\ &\quad\times\frac{(-\frac{1}{2}n; s+m_1+\mu)}{(s+m_1+\mu)!}\frac{(-\frac{1}{2}n)_s}{s!}\left(\frac{r_1}{r_2}\right)^{2s+2\mu-2m_1}. \end{aligned} \tag{31}$$

Rearranging this double series, we obtain

$$R_N^{(2)} = r_2^N\left(\frac{r_1}{r_2}\right)^{m_1}\sum_{j=0}^{\infty}\sum_{\mu=m_1+1}^{m_1+j}(-1)^j\frac{(-\frac{1}{2}n)_j}{j!} \times\frac{(M;\mu)}{\mu!}\frac{(\frac{1}{2}n;j+m_1-\mu)}{(j+m_1-\mu)!}\left(\frac{r_1}{r_2}\right)^{2j}. \quad (32)$$

Adding (30) and (32), we obtain

$$R_N = r_2^N\left(\frac{r_1}{r_2}\right)^{m_1}\sum_{s=0}^{\infty}(-1)^s\frac{(-\frac{1}{2}n)_s}{s!}\left(\frac{r_1}{r_2}\right)^{2s} \times\sum_{\mu=0}^{m_1+s}\frac{(M;\mu)}{\mu!}\frac{(\frac{1}{2}n;s+m_1-\mu)}{(s+m_1-\mu)!}. \quad (33)$$

Applying Vandermonde's theorem to the μ series, we obtain

$$R_N = (-1)^{m_1}r_2^N\left(\frac{r_1}{r_2}\right)^{m_1}\sum_{s=0}^{\infty}\frac{(-\frac{1}{2}n)_s(-\frac{1}{2}n-M)_{s+m_1}}{s!\,(s+m_1)!}\left(\frac{r_1}{r_2}\right)^{2s}, \quad (34)$$

which is the same as Eq. (28) which gives R_N for $m_1 > M$. Thus (34) gives R_N for all positive m.

(3) $m_1 \leq 0$: In this case, (24) becomes

$$R_N = \sum_{s=0}^{\infty}\sum_{\mu=0}^{\infty}\frac{(M;\mu)(\frac{1}{2}n;s)(\frac{1}{2}n;s+\mu-m_1)}{\mu!\,s!\,(s+\mu-m_1)!} \times r_2^N\left(\frac{r_1}{r_2}\right)^{2s+2\mu-m_1}. \quad (35)$$

Rearranging the double series according to ascending powers of (r_1/r_2) and using Vandermonde's theorem, we obtain

$$R_N = (-1)^{m_1}r_1^N\left(\frac{r_1}{r_2}\right)^{-m_1} \times\sum_{s=0}^{\infty}\frac{(-\frac{1}{2}n)_{s-m_1}(-\frac{1}{2}n-M)_s}{(s-m_1)!\,s!}\left(\frac{r_1}{r_2}\right)^{2s}. \quad (36)$$

Summing our results, we have

$$V_{NM} = \sum_{m=0}^{\infty}R_+(N,M,m;r_1,r_2)e^{im\theta_1}e^{i(M-m)\theta_2} + \sum_{m=1}^{\infty}R_-(N,M,m;r_1,r_2)e^{-im\theta_1}e^{i(M+m)\theta_2}, \quad (37)$$

where

$$R_\pm = (-1)^m r_2^N\left(\frac{r_1}{r_2}\right)^m \times\sum_{s=0}^{\infty}\frac{(-\frac{1}{2}N\pm\frac{1}{2}M)_s(-\frac{1}{2}N\mp\frac{1}{2}M)_{s+m}}{s!\,(s+m)}\left(\frac{r_1}{r_2}\right)^{2s} = \frac{(-1)^m}{m!}r_2^N\left(\frac{r_1}{r_2}\right)^m(-\tfrac{1}{2}N\mp\tfrac{1}{2}M)_m F\left(-\tfrac{1}{2}N\pm\tfrac{1}{2}M, -\tfrac{1}{2}N\mp\tfrac{1}{2}M+m; m+1; \frac{r_1^2}{r_2^2}\right). \quad (38)$$

4. AN EXPANSION FOR $f(r)e^{im\theta}$

In order to give an expansion for $f(r)e^{Mi\theta}$, where $f(r)$ is assumed to have a power series expansion $\sum c_n r^n$, the functions R_+ and R_- given by (38) will be written in an operational form. We have

$$(-\tfrac{1}{2}N\pm\tfrac{1}{2}M)_s(-\tfrac{1}{2}N\mp\tfrac{1}{2}M)_{s+m}r_2^{N-m-2s} = \frac{(-1)^m}{2^{m+2s}}r_2^{m\mp M}\left(\frac{1}{r_2}\frac{\partial}{\partial r_2}\right)^m r_2^{\pm M}\left(\frac{\partial^2}{\partial r_2^2}+\frac{1}{r_2}\frac{\partial}{\partial r_2}-\frac{M^2}{r_2^2}\right)^s r_2^N. \quad (39)$$

Thus, Eq. (38) can be written in the form

$$R_\pm = \frac{1}{2^m}\sum_s\frac{r_1^{m+2s}}{2^{2s}s!}\frac{r_2^{m\mp M}}{(m+s)!}\left(\frac{1}{r_2}\frac{\partial}{\partial r_2}\right)^m r_2^{\pm M} \times\left(\frac{\partial^2}{\partial r_2^2}+\frac{1}{r_2}\frac{\partial}{\partial r_2}-\frac{M^2}{r_2^2}\right)^s r_2^N. \quad (40)$$

Hence, the expansion of $f(r)e^{iM\theta}$ is given by

$$f(r)e^{iM\theta} = \sum_{m=0}^{\infty}f_+(N,M,m;r_1,r_2)e^{im\theta_1}e^{i(M-m)\theta_2} + \sum_{m=1}^{\infty}f_-(N,M,m;r_1,r_2)e^{-im\theta_1}e^{i(M+m)\theta_2}, \quad (41)$$

where

$$f_\pm = \frac{1}{2^m}\sum_s\frac{r_1^{m+2s}r_2^{m\mp M}}{(2s)!!\,(2m+2s)!!}\left(\frac{1}{r_2}\frac{d}{dr_2}\right)^m \times r_2^{\pm M}\left[\frac{d^2}{dr_2^2}+\frac{1}{r_2}\frac{d}{dr_2}-\frac{M^2}{r_2^2}\right]^s f(r_2). \quad (42)$$

Alternatively, the powers of $(1/r_2)\,d/dr_2$ can be put last, and we have the result

$$f_\pm = \frac{1}{2^m}\sum_s\frac{r_1^{m+2s}}{(2s)!!\,(2m+2s)!!} \times\left[\frac{d^2}{dr_2^2}+\frac{1}{r_2}\frac{d}{dr_2}-\frac{(m\mp M)^2}{r_2^2}\right]^s \times r_2^{m\mp M}\left(\frac{1}{r_2}\frac{d}{dr_2}\right)^m[r_2^{\pm M}f(r_2)]. \quad (43)$$

Special Cases

As before, the expressions (42) and (43) factorize for special types of functions f.

(1) If $f(r)$ satisfies the differential equation

$$\left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{M^2}{r^2}\right)f(r) = -k^2f(r), \quad (44)$$

i.e., f is a Bessel function in (kr) of order M of the first second, or third kind, then

$$f(r) = W_M(kr), \quad W_M = J_M, Y_M, H_M^{(1)}, H_M^{(2)}. \tag{45}$$

Using (B7.2.52, 53), we obtain from Eq. (42)

$$f_{\pm} = \sum_s \frac{(-1)^{s+m/2\mp m/2}(kr_1)^{m+2s}}{(2s)!!\,(2m+2s)!!} W_{M\mp m}(kr_2)$$
$$= J_{\pm m}(kr_1)W_{M\mp m}(kr_2). \tag{46}$$

Hence,

$$W_M(kr)e^{iM\theta} = \sum_{m=-\infty}^{\infty} J_m(kr_1)W_{M-m}(kr_2)e^{im\theta_1}e^{i(M-m)\theta_2}, \tag{47}$$

which is equivalent to the addition theorems (B7.6.6) and (B7.15.33) and (B7.15.34).

(2) If $f(r)$ satisfies the differential equation

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{M^2}{r^2}\right) f(r) = k^2 f(r), \tag{48}$$

i.e., f is a modified Bessel function of order M, of either the first or second kind, then

$$f(r) = I_M(kr) \quad \text{or} \quad K_M(kr). \tag{49}$$

(i) $f(r) = I_M(kr)$: Using (B7.11.19) and (B7.11.20), we obtain from Eq. (42)

$$f_{\pm} = \sum_s \frac{(kr_1)^{m+2s}}{(2s)!!\,(2m+2s)!!} I_{M\mp m}(kr_2)$$
$$= I_{\pm m}(kr_1)I_{M\mp m}(kr_2). \tag{50}$$

Thus,

$$I_M(kr)e^{iM\theta} = \sum_{m=-\infty}^{+\infty} I_m(kr_1)I_{M-m}(kr_2)e^{im\theta_1}e^{i(M-m)\theta_2}, \tag{51}$$

which is equivalent to (B7.15.36).

(ii) $f(r) = K_M(kr)$: Using (B7.11.21) and (B7.11.22), we obtain from Eq. (42)

$$f_{\pm} = \sum_s \frac{(kr_1)^{m+2s}(-1)^m}{(2s)!!\,(2m+2s)!!} K_{M\mp m}(kr_2)$$
$$= (-1)^m I_{\pm m}(kr_1)K_{M\mp m}(kr_2). \tag{52}$$

Hence,

$$K_M(kr)e^{iM\theta}$$
$$= \sum_{m=-\infty}^{+\infty} (-1)^m I_m(kr_1)K_{M-m}(kr_2)e^{im\theta_1}e^{i(M-m)\theta_2}, \tag{53}$$

which is equivalent to (B7.15.35).

ACKNOWLEDGMENT

We are grateful to Dr. A. A. Fadlalla for stimulating discussions and useful suggestions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Connection between Marchenko Formalism and N/D Equations: Regular Interactions. II*

HENRI CORNILLE
Institut de Physique Nucléaire, Division de Physique Théorique, 91 Orsay, France

(Received 30 December 1966)

In a previous paper we have shown, for S-waves, that the resulting integral equations of the f/f equations (equivalent to the N/D approach) can be obtained from the Marchenko formalism. The potential $V(\mu, r)$ reconstructed from the discontinuity $\mu\Delta(x)$ is $-2(d/dr)[(d\mathfrak{D}/dr)/\mathfrak{D}]$, where $\mathfrak{D}(\mu, r)$ is the Fredholm denominator of the Jost solution and $\mathfrak{D}(\mu, 0)$ that of the resulting integral equation. For "regular discontinuities" we find different classes of $V(\mu, r)$. First, if $\mathfrak{D}(\mu, 0) = 0$, then $V(\mu, r)$ is not "regular at the origin" [in general, we find that $V(\mu, r)$ becomes marginally singular: repulsive and singular like r^{-2}]. Secondly, if $\mathfrak{D}(\mu, 0) \neq 0$, then $V(\mu, r)$ is "regular" at the origin and we obtain the following: (i) If $|\mu|$ is less than the smallest modulus root of $\mathfrak{D}(\mu, 0)$, then $V(\mu, r)$ has no poles for $r \geq 0$. This range of $|\mu|$-values where the iteration series of the resulting integral equations converge is limited by the smallest $|\mu|$ value where a real or complex ghost can appear or where a bound state can appear at zero energy. (ii) For $|\mu|$ larger than this smallest modulus root but μ inside the interval given by the first positive and negative roots, $V(\mu, r)$ has no second-order poles for $r \geq 0$. These results (i) and (ii) are obtained with the restriction that in the considered interval there do not exist (μ, r)-values such that $\mathfrak{D}(\pm\mu, r) = 0$, and from our study we cannot conclude that this is always true. (iii) For μ outside the above interval, $V(\mu, r)$ has poles of the second order for $r > 0$, the "bound states" being, in general, real or complex ghosts, or "bound states" corresponding to badly behaved potentials. We find also that the Jost solutions for energy equal to zero are $\mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r)$. This gives the connection between ghosts and, in general, possible bound states appearing at zero energy; this gives also the relations between poles of $V(\mu, r)$ corresponding to opposite μ values. These results for the Jost function correspond to a normalization at infinity, so we have considered the problem of subtractions with normalization at an arbitrary point. Then the new Fredholm determinant is the product of the old one by the value of the Jost function at the subtracted point. It follows that, if the first μ-greater-than-zero and the first μ-less-than-zero roots of $\mathfrak{D}(\mu, r)$ are not opposite ($r \geq 0$), the $|\mu|$ interval of convergence of the iteration series is enlarged for the subtracted equation.

I. INTRODUCTION

In a previous paper,[1] hereafter called I, we have shown that, in potential scattering for S waves, the resulting integral equation[2] of the f/f equations (equivalent to the N/D approach[3]) can also be obtained from the Marchenko formalism.[4] This fact makes it possible to study the existence of the solutions of the resulting integral equation and to interpret their meaning in terms of potentials. In Sec. II we recall briefly the results obtained in I and also some basic relations obtained in Appendix A. Considering "regular discontinuities" such that the resulting integral equation is of the Fredholm type, we study in the following sections the behavior of the potential reconstructed from the discontinuity, the interpretation of the corresponding states (bound states and ghosts), the Jost solutions and Jost functions, and the usefulness of normalizations and subtractions. We illustrate some of these results by considering the simple case of discontinuities such that the kernels of the Marchenko equations or, equivalently, the kernel of f/f equations are degenerate with only one or two eigenvalues.

II. BASIC RELATIONS

We recall briefly the results (obtained in I), which give the connection between the Marchenko-inversion formalism[4] and dispersion relations in the f/f approach.[2] We consider the Jost solutions

$$f(k, r) \underset{r\to\infty}{\simeq} e^{-ikr}$$

for S waves. In the Marchenko formalism,[4] for Yukawa types of potentials, we have

$$V(r) = \int_m^\infty e^{-\alpha r} C(\alpha)\, d\alpha,$$

$$|C(\alpha)| < \text{const} \times \alpha^{1-\eta}, \quad \eta > 0,$$

$$\int_m^\infty \frac{|C(\alpha)|}{\alpha^2}\, d\alpha < \infty, \tag{1}$$

where $V(r)$ is holomorphic for $\text{Re}\, r > 0$.

* The results of this paper have been presented at the 1966 International Conference on High Energy Physics at Berkeley, Calif.

[1] H. Cornille, J. Math. Phys. **8**, 2268 (1967). Paper presented at the 1966 International Conference on High Energy Physics at Berkeley, Calif.

[2] (a) A. Martin, *Progress in Elementary Particles and Cosmic Ray Physics* (North-Holland Publ. Co., Amsterdam, 1965); Nuovo Cimento **19**, 1257 (1961). (b) V. de Alfaro and T. Regge, Nuovo Cimento **20**, 956 (1961).

[3] (a) G. F. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960); see also other references quoted in Ref. 1. (b) E. H. Nyman, Nuovo Cimento **37**, 429 (1965).

[4] Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon & Breach, Science Publishers, Inc., New York, 1963).

The scattering data $-\mathcal{F}(r)$, defined as

$$-\mathcal{F}(r) = \sum_{j=1}^{P} M_j^2 e^{-|x_j| r} + \frac{1}{2\pi} \int_{-\infty}^{+\infty} [1 - S(k)] e^{ikr}\, dk, \tag{1'}$$

is the inverse Laplace transform of the discontinuity of the S matrix

$$\mathcal{F}(r) = \int_{\frac{1}{2}m}^{\infty} e^{-ru} \Delta(u)\, du. \tag{1''}$$

We have called $\Delta(x)$ the discontinuity. The M_j^2, being greater than 0, are finite normalization constants corresponding to the bound states $-|x_j|^2$ [the bound states, poles of $S(k)$ in $\operatorname{Im} k < 0$, are roots of the Jost function, i.e., all the roots in $\operatorname{Im} k < 0$ are simple and are such that $\operatorname{Re} k = 0$]. If we substitute (1'') in the Marchenko formalism, we get for the Jost solutions an integral equation[1]:

$$F(x, r) = f\,(k = -ix, r),$$
$$F(x, r) = e^{-xr} + \int_{\frac{1}{2}m}^{\infty} \frac{\Delta(y) e^{-r(x+y)}}{x + y} F(y, r)\, dy. \tag{2a}$$

The important point is that, for $r = 0$, Eq. (2a) is the same equation as the resulting integral equation of the f/f formalism. Equation (2a) is an off-the-mass-shell N/D-type equation from which we get results for the on-the-mass-shell N/D equations.

Furthermore, if we consider the Laplace transform

$$F(x, r) = \int_{r}^{\infty} \phi(y, r) e^{-xy}\, dy,$$

we get from (2a)

$$\phi(y, r) = \delta(y - r) + \int_{r}^{\infty} \mathcal{F}(y + t) \phi(t, r)\, dr. \tag{2b}$$

If we put $\phi(y, r) = \delta(y - r) + G(y, r)$, we get the Marchenko fundamental equation giving the potential from the scattering data:

$$G(y, r) = \mathcal{F}(r + y) + \int_{r}^{\infty} \mathcal{F}(t + y) G(t, r)\, dt,$$
$$V(r) = -2 \frac{d}{dr} G(r, r). \tag{2'}$$

We see that (2b) and (2') have exactly the same kernel. This is the main property which makes it possible to connect the f/f equations and the Marchenko formalism. For instance, the eigenvalues [such that nontrivial solutions of the homogeneous equation of (2a), (2b), or (2') exist] are the same and the Fredholm determinants of (2a), (2b), (2') are also the same.[1] Furthermore, the eigenfunctions are connected by a Laplace transform. We can use both kernels in order to study the eigenvalues or, equivalently, the roots of Fredholm determinants.

In fact, for the Yukawa family,[1] we know from the Marchenko results that the only solution of the homogeneous equation (2') is the trivial one and similarly for (2a) and (2b). Now we want to emphasize (always for the Yukawa family)[1] that all these results can be obtained directly from dispersion techniques (without using the Marchenko formalism) by using the well-known analytical properties of the Jost solutions for $r \geq 0$ fixed and the method given by Martin[2(a)] for the Jost functions. We define $g(k, r) = f(k, r) e^{ikr}$; then for Eq. (1), $g(k, r)$ has the following spectral representation, for $r \geq 0$,

$$g(k, r) = 1 - i \int_{\frac{1}{2}m}^{\infty} \frac{R(y, r)}{k - iy}\, dy,$$

where $R(y, r)$ is real. Now as has been shown by Martin,[2(a)] $f(ix \pm \epsilon, r)$ and $f(ix, r)$ are not linearly independent solutions of the Schrödinger equation; thus we get

$$\lim_{\epsilon \to 0} \frac{g(ix + \epsilon, r) - g(ix - \epsilon, r)}{g(-ix, r)}$$
$$= -2i\pi \Delta(x) e^{-2xr} = \frac{-2i\pi R(x, r)}{g(-ix, r)},$$

where $-\Delta(x)$ is the discontinuity of the S matrix. If we substitute this last expression in the spectral representation, we get

$$g(-ix, r) = 1 + \int_{\frac{1}{2}m}^{\infty} \frac{\Delta(y) e^{-2yr} g(-iy, r)}{x + y}\, dy.$$

If we recall that $g(-ix, r) = F(x, r) e^{+xr}$ and substitute in the spectral representation, we get the off-the-mass-shell equation (2a). So, finally, the Marchenko equation can be obtained directly from dispersion methods by taking the Laplace transform of (2a).

We recall that we have assumed a Yukawa-type family of potentials in order to get (2'). In this paper we are interested in the following problem: Find and interpret the solutions of the resulting integral equation of f/f [or (2) for $r = 0$] considered as a linear integral equation. Then we write the discontinuity as $\mu\Delta(x)$, where μ is a real linear parameter. Since $\mu\Delta(x)$ is the input, the scattering data become

$$\mu \int_{\frac{1}{2}m}^{\infty} e^{-ru} \Delta(u)\, du$$

and the S matrix as well as the potentials are μ dependent: That is, $V(\mu, r)$ and $S(k, \mu)$ are μ dependent. We want to interpret the solutions of (2') but, of course, we cannot assume the conditions of finite moments for the potential (we recall that

Marchenko equations and the existence and uniqueness of the corresponding solutions have been established with these conditions of finite moments).

Now we do not assume anything about $\mu\Delta(x)$ and we formally get a set of results from Eq. (2).

The Fredholm determinant of Eqs. (2) and (2′) is

$$\mathfrak{D}(\mu, r) = 1 + \sum_1^\infty \frac{(-\mu)^n}{n!} \times \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n \Delta(u_i) e^{-2ru_i}\right) P_n,$$

$$P_n(u_1, \cdots, u_i, \cdots, u_n) = \begin{vmatrix} \frac{1}{2u_1} & \cdots & \frac{1}{u_1 + u_n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{1}{u_1 + u_n} & \cdots & \frac{1}{2u_n} \end{vmatrix}, \tag{3a}$$

$$\mathfrak{D}(\mu, r) = \exp\left(-\sum_1^\infty \mu^n \frac{A_n(r)}{n}\right),$$

$$A_n(r) = \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \times \frac{\prod_1^n \Delta(u_i) e^{-2ru_i}}{(u_1 + u_2)(u_2 + u_3) \cdots (u_n + u_1)}, \tag{3b}$$

where $A_n(r)$ is the nth trace of the kernels of (2). The potential $V(\mu, r)$ reconstructed from this discontinuity is linked in a very simple way to $\mathfrak{D}(\mu, r)$, as was shown in Ref. 1,

$$V(\mu, r) = -2\frac{\partial}{\partial r}\left(\frac{\partial \mathfrak{D}(\mu, r)}{\partial r} \Big/ \mathfrak{D}(\mu, r)\right). \tag{4}$$

The bound obtained by de Alfaro and Regge for $r = 0$ can be generalized, for $\operatorname{Re} r \geq 0$, to

$$|\mathfrak{D}(\mu, r)| < \exp \int_{\frac{1}{2}m}^\infty \frac{|\mu\Delta(u)|}{2u} e^{-2(\operatorname{Re} r)u}\, du,$$

$$|\mathfrak{D}(\mu, r) - 1| < \exp\left(\int_{\frac{1}{2}m}^\infty \frac{|\mu\Delta(u)|}{2u} e^{-2(\operatorname{Re} r)u}\, du\right) - 1. \tag{5}$$

In Appendix A of the present paper, it is also shown that the Jost solution (2) for zero energy ($x = 0$) can be written

$$F(0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r). \tag{6}$$

In Appendix B it is shown that

$$\left|\frac{d}{dr}\mathfrak{D}(\mu, r)\right| < \int_{\frac{1}{2}m}^\infty |\mu\Delta(u)|\, e^{-2(\operatorname{Re} r)u}\, du \times \exp \int_{\frac{1}{2}m}^\infty \frac{|\mu\Delta(u)|}{2u} e^{-2(\operatorname{Re} r)u}\, du, \quad \operatorname{Re} r > 0. \tag{7}$$

In this paper we restrict our study to "regular discontinuities" $\Delta(x)\theta(x - \frac{1}{2}m)$ such that

$$\int_{\frac{1}{2}m}^\infty \frac{|\Delta(x)|}{x}\, dx < \infty,$$

$$\int_{\frac{1}{2}m}^\infty \int_{\frac{1}{2}m}^\infty \left(\frac{\Delta(y)}{x + y}\right)^2 dx\, dy < \infty. \tag{8}$$

Then Eq. (2) is of the Fredholm type for $r \neq 0$, as well as for $r = 0$. From Eq. (6),

$$|\mathfrak{D}(\mu, r)| \underset{r \to \infty}{\to} 1$$

and Eq. (4) can be written as

$$\exp\left(-\tfrac{1}{2}\int_r^\infty dx \int_x^\infty V(\mu, t)\, dr\right) = \mathfrak{D}(\mu, r), \quad r \geq 0. \tag{9}$$

Because of Eq. (4), $V(\mu, r)$ can have poles of the second order, for $r > 0$, corresponding to roots of $\mathfrak{D}(\mu, r)$. If we write V as a Laplace transform,

$$V(\mu, r) = \int_m^\infty C(\alpha, \mu) e^{-\alpha r}\, d\alpha,$$

or if we consider V and C as a μ series, that is,

$$V = \sum \mu^n V_n(r), \quad C = \sum \mu^n C_n(\alpha),$$

we have also (see I)

$$V_n(r) = \frac{2}{n}\frac{d^2 A_n(r)}{dr^2}, \qquad r > 0,$$

$$\frac{1}{2}\int_m^\infty \frac{C_n(\alpha) e^{-\alpha r}}{\alpha^2}\, d\alpha = A_n(r), \qquad r \geq 0, \tag{10}$$

$$\frac{1}{2}\int_m^\infty e^{-\alpha r}\frac{C(\alpha, \mu)}{\alpha^2}\, d\alpha = -\log \mathfrak{D}(\mu, r) = \sum_1^\infty \mu^n \frac{A_n(r)}{n}, \quad r \geq 0. \tag{11}$$

Now we assume that $\Delta(x)$ has a bad asymptotic behavior (as $x \to \infty$) such that Eq. (8) is not satisfied. We remark that, for such "singular discontinuities," most of the previous results are still valid for $r > 0$. We consider, for instance,

$$\Delta(x) \underset{x \to \infty}{\simeq} \text{const},$$

which is marginally singular, or the case when $\Delta(x) \simeq x^n$, for $n > 0$, which is even more singular, but still with a good behavior of $\Delta(x)$ for x finite; then we see that Eq. (2) is still of the Fredholm type for $r > 0$ (but not for $r = 0$). Roughly speaking, we see from Eqs. (4) and (5) that the main difference for the case we consider, Eq. (8), is that the corresponding potentials $V(\mu, r)$ will be "singular" when r goes to zero for such "singular discontinuities." We do not study these "singular discontinuities" in this paper,

but Eqs. (2), (4), and (9) can be the basis for a further investigation in this field.

In the following, we consider always "regular discontinuities" such that Eq. (8) is satisfied.

III. BEHAVIOR OF THE RECONSTRUCTED POTENTIAL $V(\mu, r)$

It cannot be the object of the present paper to do a complete analysis of the solutions $F(x, r)$ and $V(\mu, r)$ when the conditions of finite moments for the potentials and uniqueness of the solutions are not satisfied—or equivalently when the analytic structure of the $S(k, \mu)$ matrix corresponding to $\mu\Delta(x)$, μ real, does not satisfy the usual conditions: e.g., the location and order of multiplicity of the poles in $\operatorname{Im} k < 0$ or conditions on the normalization constants M_j^2 of these poles [see (1′)]. We want to give only some simple general features. For a $V(\mu, r)$ of the Laplace transform type, we find four different families satisfying two different criteria: namely, (i) $\mathfrak{D}(\mu, r = 0) \neq 0$ or $= 0$, $V(\mu, r)$ "regular" or not regular at the origin, respectively; (ii) if $\mathfrak{D}(\mu, r) \neq 0$ or $= 0$ in $\operatorname{Re} r > 0$, then $V(\mu, r)$ is holomorphic or not in $\operatorname{Re} r > 0$, respectively. Moreover, if $\mathfrak{D}(\mu, r)$ can vanish in $\operatorname{Re} r > 0$, we also have two cases, depending on whether some roots are on the real axis or not. If $V(\mu, r)$ has poles of the second order for $r \geq 0$, then the conditions of finite moments are not satisfied.

Then we have the following families:

(1) $\mathfrak{D}(\mu, r = 0) = 0$ and we find in general that $V(\mu, r)$ is singular and repulsive like r^{-2} at the origin.

(2) $\mathfrak{D}(\mu, r = 0) \neq 0$, but $\mathfrak{D}(\mu, r) = 0$ for some particular positive r values. Then $V(\mu, r)$ has poles of the second order for r real and greater than zero.

(3) $\mathfrak{D}(\mu, r = 0) \neq 0$; $\mathfrak{D}(\mu, r)$ can have roots in $\operatorname{Re} r > 0$ and $V(\mu, r)$ can have poles for $\operatorname{Re} r > 0$, but not for r real and greater than or equal to zero.

(4) $\mathfrak{D}(\mu, r = 0) \neq 0$ and $V(\mu, r)$ holomorphic in $\operatorname{Re} r > 0$ [Yukawa family (1)].

For families (2), (3), (4), $\mathfrak{D}(\mu, r = 0) \neq 0$ and $V(\mu, r)$ is "regular" at the origin (roughly speaking, less singular than the centrifugal potential).

A. An Example: $\Delta(x) = \delta(x - b)$

In order to illustrate these different cases, we consider first the simple case where the discontinuity is replaced by a simple pole leading to the Jost–Bergman potentials[5]

$$V(\mu, r) = 4b\mu e^{-2br}[1 - (\mu/2b)e^{-2br}]^{-2}$$

and

$$\mathfrak{D}(\mu, r) = 1 - (\mu/2b)e^{-2br}.$$

For $\mu = 2b$, we have family (1) and

$$V \underset{r \to 0}{\simeq} 2/r^2.$$

For $\mu \neq 2b$, we have families (2)–(4). For $\mu > 2b$ [family (2)], V has poles of the second order in $\operatorname{Re} r > 0$ and, in particular, for $r > 0$: $\operatorname{Re} r = (1/2b) \log|\mu/2b|$, $\operatorname{Im} r = 2m\pi/b$, $m = 0, 1, 2, \cdots$. For $\mu < -2b$ [family (3)], V has poles in $\operatorname{Re} r > 0$, but not for r real and ≥ 0, that is,

$$\operatorname{Re} r = (1/2b) \log|\mu/2b|, \quad \operatorname{Im} r = \pm(2m + 1)\pi/2b.$$

For $|\mu| < 2b$ [family (4)], V has no poles in $\operatorname{Re} r > 0$.

B. Some Simple Properties for the Roots of $\mathfrak{D}(\mu, r) = 0$

If we assume $\Delta/(x + y)$ to be a real polar kernel,[5] then [from Eq. (8)] (2a) is of Fredholm type with well-known properties for the roots. If we consider (2b), we shall get more information because $\mathcal{F}(x + y)$ is not only real and nondegenerate, but also symmetric. From Eq. (8) we get

$$\int_r^\infty \int_r^\infty [\mathcal{F}(x + y)]^2 \, dx \, dy < \infty, \tag{8′}$$

for $r > 0$ or $r > 0$ going to zero. Then $\mathcal{F}(x + y)$ is a Hilbert–Schmidt kernel[6] defined in $L^2(r, \infty)$, and we can apply the properties of such kernels about the eigenvalues or roots of $\mathfrak{D}(\mu, r)$. The eigenvalues $\mu_{\pm j}(r)$ are real and, for nondegenerate $\mathcal{F}(x + y)$, infinite in number, namely

$$\mu_{+j}(0) > 0, \quad \mu_1(0) < \mu_2(0) \cdots < \mu_j(0) \cdots;$$
$$\mu_{-j}(0) < 0, \quad \mu_1(0) > \mu_{-2}(0) \cdots > \mu_{-j}(0).$$

If Δ is always > 0 or < 0, then we have only $\mu_{+j}(r) > 0$ in one case and $\mu_{-j}(r) < 0$ in the other case.

For $|\mu| < \mu_0$ where

$$\mu_0 = \log 2 \left(\int_{\frac{1}{2}m}^\infty \frac{|\Delta(u)|}{2u} \right)^{-1},$$

we see from the bound (5) that $\mathfrak{D}(\mu, r)$ has no roots for $\operatorname{Re} r \geq 0$ and V has no second-order poles for $\operatorname{Re} r \geq 0$.

On the other hand, the nth trace $A_n(r)$ given by (3b) can also be written as

$$A_n(r) = \int_r^\infty du_1 \cdots \int_r^\infty du_n \times \mathcal{F}(u_1 + u_2) \; (u_2 + u_3) \cdots \mathcal{F}(u_n + u_1).$$

[5] See R. Jost, Helv. Phys. Acta **20**, 256 (1947), and V. Bergman, Rev. Mod. Phys. **21**, 488 (1949).

[6] See the references given by R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I, p. 161. See also a summary of these properties and the references given by K. Meetz, J. Math. Phys. **3**, 690 (1962).

We see that

$$|A_n(r)| \underset{r\to\infty}{\to} 0$$

and

$$|\mu_{\pm j}(r)| \underset{r\to\infty}{\to} \infty,$$

following the expansion of $\mathfrak{D}(\mu, r)$ given by (3b).

For particular cases, we can get more information. For instance, if $\mathcal{F}$ is always > 0, then $A_n(r_1) > A_n(r_2)$ if $r_1 < r_2$. From (3b), we find that the smallest root is $\mu_1(r)$ and $\mu_1(r_1) < \mu_1(r_2)$ if $r_1 < r_2$. We get the same result for the smallest modulus root if $\mathcal{F}$ is always less than zero. [In this case, we can consider the corresponding expansion of (3b) for the Fredholm determinant of the first iterate kernel of $\mathcal{F}(x + y)$; the roots are $\mu^2_{\pm j}(r)$ and only $A_{2n}(r)$ appears in the corresponding expansion.]

Unfortunately, these $\mu_{\pm j}(r)$ roots of $\mathfrak{D}(\mu, r) = 0$ (for r fixed) have an unknown order of multiplicity (with respect to μ). But in our case, Eq. (6) gives us supplementary information about the order of multiplicity in μ. We know[6,7] that the resolvant kernels corresponding to such polar kernels (2a) or symmetric kernels [(2b) and (2′)] have only simple poles in μ. The solutions[7]

$$F(x = 0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r)$$

also have simple poles. From this it follows that, if (μ, r) is a root of $\mathfrak{D}(\mu, r) = 0$ ($r \geq 0$, fixed) but such that $(-\mu, r)$ is not also a root, $\mathfrak{D}(-\mu, r) \neq 0$, the multiplicity is unity, and $\partial\mathfrak{D}(\mu, r)/\partial\mu \neq 0$. In $\mathrm{Re}\, r > 0$, $\mathfrak{D}(\mu, r)$ is analytic in r and is an entire function of μ; in this case, $\mu(r)$ is unique. Near such a root (μ real, and r real and > 0), we can apply the well-known theorems about implicit functions $\mu(r)$ such that $\mathfrak{D}(\mu(r), r) = 0$ (for $r = 0$, we do not have analyticity in r, but only continuity when $r > 0 \to 0$).

C. Are the Solutions of (2) Also Solutions of the Schrödinger Equation?

We recall that in I the integral equation (2) was established with the restriction that the potentials were of the Yukawa type (1) (where V was of the Laplace transform type, holomorphic in $\mathrm{Re}\, r > 0$, and "regular" at the origin). But $\mathfrak{D}(\mu, r)$ can vanish in $\mathrm{Re}\, r > 0$ as well as for $r = 0$; consequently, some assumptions of Yukawa type potentials can be violated, and we must ask ourselves if the solutions of (2) always have a meaning in potential scattering. The study can be made, independently of Marchenko results, by using only the integral equation (2) and the definition (4) of the potential, but this requires a large amount of algebra; so, for simplicity, we shall use Marchenko results.

We seek μ-values such that $V(\mu, r)$ satisfies the conditions of finite moments. For

$$|\mu| < \mu_0 = \log 2\left(\int_{\frac{1}{2}m}^{\infty} \frac{|\Delta(u)|}{2u}\, du\right)^{-1},$$

we know that $\mathfrak{D}(\mu, r) \neq 0$ for $r \geq 0$ and

$$\left|\int_m^{\infty} e^{-\alpha r} \frac{C(\alpha, \mu)}{\alpha^2}\, d\alpha\right| < \infty.$$

We see that in this case V decreases exponentially, is regular at the origin, and has no poles of the second order for $r \geq 0$. But we have no information about $|V(\mu, r)|$ which is sufficient to directly find the conditions of finite moments. Then, following Marchenko,[4] we consider the quantity

$$M = \int_0^{\infty} r\, |\bar{\mathcal{F}}'(r)|\, dr.$$

Marchenko has shown[8] that, if $M < \infty$ and if (2′) has no homogeneous[9] nontrivial solutions [μ-values different from $\mu_{\pm j}(r)$ for all $r \neq 0$ and any j], then

$$A = \int_0^{\infty} r\, |V(r)|\, dr < \infty.$$

First, we consider $|\mu| < \mu_0$ and we shall see that the two Marchenko conditions above are satisfied. On the one hand, we know that in this interval (2a) and (2b) or (2′) have only the trivial (identically zero) solution for the homogeneous integral equations. On the other hand, we wish to determine if $M < \infty$. For this we consider the condition (8) for $\Delta(x)$ and the fact that $\mathcal{F}(r)$ and $\Delta(x)$ are Laplace transforms. For instance, in order to verify (8), we can assume

$$\Delta(u) \underset{u\to\infty}{\simeq} \text{const} \times u^{-\eta}, \quad \eta > 0,$$

and we see that $|\mathcal{F}'(r)|$ decreases exponentially when $r \to \infty$ and is not more singular than $r^{-2+\eta}$ when $r \to 0$. In this case, $M < \infty$, and we can apply the Marchenko results. More generally, using only the condition (8) for $\Delta(x)$, it is shown in Appendix B that $M < \infty$. Then, using the Marchenko results quoted above, $A < \infty$ and (2) is the Jost solution of the Schrödinger equation (at least in this $|\mu|$ range). We will apply these results to $F(0, r)$ and to the corresponding $V(\mu, r)$ given by Eqs. (4)–(6) and write

[7] F. G. Tricomi, *Integral Equations* (Interscience Publishers, Inc., New York, 1958), p. 115.

[8] See Ref. 4, Chap. II, p. 122.

[9] According to our results, this condition is equivalent to the condition that Eqs. (2a) and (2b) have no homogeneous solutions.

them as solutions of the Schrödinger equation:

$$\left[\frac{d^2}{dr^2} - V(\mu, r)\right] F(0, r) \equiv \frac{B(\mu, r)}{(\mathfrak{D}(\mu, r))^2} \equiv 0, \quad (12a)$$

$$B(\mu, r) \equiv \overline{\mathfrak{D}}''\mathfrak{D} + \mathfrak{D}''\overline{\mathfrak{D}} - 2\mathfrak{D}'\overline{\mathfrak{D}}' \equiv 0, \quad \text{for any } r > 0. \quad (12b)$$

In the $|\mu|$ range considered, $\mathfrak{D}(\mu, r) \neq 0$ and we get (12b), where $\overline{\mathfrak{D}} = \mathfrak{D}(-\mu, r)$, $\mathfrak{D} = \mathfrak{D}(\mu, r)$; here the derivatives are taken with respect to r. We substitute the expansions (3a) in $\mathfrak{D}$, $\overline{\mathfrak{D}}$ and we get

$$B(\mu, r) = \sum_1^\infty \mu^{2n} B_{2n}(r),$$

where

$$B_{2n}(r) = \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n \Delta(u_i) e^{-2ru_i}\right) \times h_n(u_1, \cdots, u_i, \cdots, u_n).$$

h_n is independent of r and Δ can be determined from (12b) and (3a). In this $|\mu|$ interval, it follows that $B_{2n} \equiv 0$, for any $r > 0$, from Eqs. (12a) and (12b). We shall use this result in order to show that, in fact, (2) [or (6) for $x = 0$] is really the Jost solution not only inside this restricted range.

Secondly, we consider $|\mu| > \mu_0$. We still have $B_{2n}(r) \equiv 0$ and, consequently, $B(\mu, r) \neq 0$. The only difference with the previous case is that now $\mathfrak{D}(\mu, r)$ can have roots in $r \geq 0$, but we remark that, for μ fixed, $\mathfrak{D}(\mu, r)$ is not $\equiv 0$, so that if we apply the differential operator (12a) to $F(0, r)$, we still find that $F(0, r)$ satisfies the differential Schrödinger equation. Of course, now $F(0, r)$ can have poles for $r > 0$. On the other hand, from (6) we see that $F(0, r) \to 1$ as $r \to \infty$, so that $F(0, r)$ is the Jost solution of the potentials (4). [A similar study can be made for the $F(x, r)$ solution of (2) when x is > 0, using the fact that $\mathfrak{D}(\mu, r)$ is the Fredholm determinant of (2) for $x > 0$, as well as for $x = 0$.]

The fact that Eqs. (2) are really the Jost solutions even when $V(\mu, r)$ given by (4) is badly behaved is of great theoretical importance. We emphasize the great power of the Marchenko formalism and its superiority over the usual techniques in order to get the resulting integral equation of *f/f*. Usually a Yukawa-type family is required and we get only physically acceptable true bound states. From the Marchenko formalism, we see that other potentials lead also to Eq. (2). In particular, there exist badly behaved potentials for $r \geq 0$ such that ghosts will appear.

D. Behavior of $V(\mu, r)$ at the Origin when $\mathfrak{D}(\mu, 0) \neq 0$

We remark that $\mathfrak{D}(\mu, r)$ is the Fredholm determinant of $F(x, r)$. In this case, the Jost solution goes to a constant when $r \to 0$ (this constant, the Jost function, can of course be zero). For instance,

$$F(0, r) \underset{r \to 0}{\to} \mathfrak{D}(-\mu, 0)/\mathfrak{D}(\mu, 0).$$

We recall that for S-waves, this means that the potential is regular at the origin (the behavior near the origin of the solutions of the second-order differential Schrödinger equation is in this case independent of the potential). This remark is the most straightforward proof of the regularity of $V(\mu, r)$ at the origin, but we can obtain more information about the explicit behavior of the potential near the origin if we give more information than (8) about the asymptotic behavior of $\Delta(x)$. For this we use the bounds on $|\mathfrak{D}(\mu, r)|$, $|\mathfrak{D}'(\mu, r)|$ given by (5) and (7) and a similar bound for $|\mathfrak{D}''(\mu, r)|$ [see Eq. (B2)]. We consider V given by (4) and we have to investigate the behavior when $r \to 0$ of Laplace integrals of the type

$$\int_{\frac{1}{2}m}^\infty e^{-2ur} |\Delta(u)| (2u)^m \, du, \quad m = 0, \pm 1,$$

where the asymptotic behavior of $|\Delta(u)|$ for $u \to \infty$ is given. For this we use the well-known behavior of conjugate variables in Laplace transforms ($u \to \infty$, $r \to 0$). For instance, if we assume that $\Delta(u)$ satisfies (8) and also

$$|\Delta(u)| \underset{u \to \infty}{\simeq} \text{const} \times u^{-\epsilon}, \quad \epsilon > 0, \neq 1,$$

we get

$$|\mathfrak{D}(\mu, r)|_{r \to 0} < \text{const} + r^\epsilon \, \text{const},$$

$$|V(\mu, r)|_{r \to 0} < \text{const}_1 \times r^{\epsilon - 2} + \text{const}_2 \times r^{2(\epsilon - 1)},$$

showing explicitly that the potential is less singular than r^{-2}. More generally, it is shown in Appendix B that, if Δ satisfies (8) and $\mathfrak{D}(\mu, r)$ cannot vanish in a finite interval $r \in [0, b]$, then

$$\int_0^b r \, |V(r)| \, dr < \infty.$$

For $|\mu|$ finite and fixed, it will be explained in Secs. III.E and III.F that $\mathfrak{D}(\mu, r)$ has only a finite number of roots in $r \geq 0$ and that such an interval always exists. Since $V(\mu, r)$ is regular near the origin, it follows that if

$$F(x, r) \underset{r \to 0}{\to} 0,$$

then

$$F(x, r) \underset{r \to 0}{\simeq} \text{const} \times r.$$

E. $\mathfrak{D}(\mu, r) = 0$. The Order of Multiplicity in $r \geq 0$ of the Roots

(1) We assume that $\mathfrak{D}(\mu, r_0) = 0$ for $r_0 > 0$. In $\text{Re}\, r > 0$, $\mathfrak{D}(\mu, r)$ is analytic such that

$$\mathfrak{D}(\mu, r) \underset{r \to 0}{\simeq} \frac{(r - r_0)^n}{n!} \mathfrak{D}^{(n)}(\mu, r_0),$$

where n is an integer. Similarly, if $\mathfrak{D}(-\mu, r) = 0$, then the order of multiplicity also is an integer which we call m. We get (n, m) by a self-consistent method because (n, m) gives the singularity of V and from Fuchs theorem the singularities of $F(0, r)$ which must be consistent with the singularities of $F(0, r)$ are obtained directly from (6). We get:

(i) If $n = 0$, then $m = 0$ or 1, and conversely if $m = 0$, then $n = 0$ or 1. We see that, if $\mathfrak{D}(\pm\mu, r)$ does not vanish for the same r value, then the root r is always simple.

(ii) Both $\mathfrak{D}(\pm\mu, r) = 0$. Then the only (n, m) values are $(1, 3)$, $(3, 1)$ $(3, 6)$, $(6, 3) \cdots (\frac{1}{2}p(p-1), \frac{1}{2}p(p+1))$, $(\frac{1}{2}p(p+1), \frac{1}{2}p(p-1))$, where p is an integer.

We remark that, if the possible singularities of the potentials are only poles of the second order, then we have poles of increasing order for the Jost solutions. We remark also that the potentials are repulsive near the poles.

From (6) we see that the product of two $F(0, r_0)$ corresponding to $\pm\mu$ values is unity, so that if one vanishes when $r \to r_0$, then necessarily the other is infinite.

(2) We consider now the more delicate case $r_0 = 0$, where $\mathfrak{D}(\mu, 0) = 0$. The difficulty comes from the fact that $\mathfrak{D}(\mu, r)$ is, in general, nonanalytic in r at the origin.

First, we assume the simple case

$$\Delta(x) = \sum_1^{n_0} \alpha_i \delta(x - b_i),$$

where $b_i > 0$, n_0 arbitrary but finite. In this case, $\mathfrak{D}(\mu, r)$ is analytic at $r = 0$ also. Then the previous analysis made for $r \neq 0$ still holds for $r_0 = 0$. In this case we see that

$$V(\mu, r) \underset{r \to 0}{\simeq} 2n/r^2$$

[where n is the order of multiplicity of the roots of $\mathfrak{D}(\mu, r)$ for $r = 0$] is repulsive and singular like r^{-2}. We can still determine the order of multiplicity as above, whether or not $\mathfrak{D}(-\mu, r)$ vanishes for $r = 0$. It is amusing to remark that, for these special values where $\mathfrak{D}(\pm\mu, 0) = 0$, the reconstructed potential behaves near the origin like a centrifugal potential $p(p+1)/r^2$. If $\mathfrak{D}(-\mu, 0) \neq 0$, then $p = 1$ for $V(\mu, r)$, whereas if $\mathfrak{D}(\pm\mu, 0) = 0$, then for $V(\pm\mu, r)$ we have p or p^{+1}. One of the solutions $F(0, r)$ (corresponding to $\pm\mu, r$) is regular at the origin and the other is singular.

Secondly, we consider the general case where Δ satisfies only (8) and, in particular, the case where $\mathfrak{D}(\mu, 0) = 0$, but $\mathfrak{D}(-\mu, 0) \neq 0$. From the condition that $\mathfrak{D}(-\mu, 0) \neq 0$ and by applying the above results, we see that $V(-\mu, r)$ is regular at the origin, and satisfies (13) so that for $V(-\mu, r)$ the Jost function $F(0, r)$ is

$$\mathfrak{D}(\mu, r)/\mathfrak{D}(-\mu, r) \simeq \text{const} \times r, \quad \text{as} \quad r \to 0.$$

From this it follows that $\mathfrak{D}(\mu, r) \simeq \text{const}\, r$, as $r \to 0$, and for $V(\mu, r)$ the Jost function $\simeq \text{const}/r$, as $r \to 0$. From this behavior of the solution $F(0, r)$ near the origin, it follows that necessarily $V(\mu, r) \simeq 2/r^2$, as $r \to 0$, and we see that near the origin the potential is like a centrifugal potential with $l = 1$. Finally, if both $\mathfrak{D}(\pm\mu, 0) = 0$, we cannot draw any conclusions from our study.

F. Behavior of $V(\mu, r)$ for μ Inside Different Intervals

(1) $|\mu| < \mu_0$ (see Sec. III.C). We recall that $\mathfrak{D}(\pm\mu, r) \neq 0$ and

$$\left| \int_m^\infty \frac{C(\alpha, \pm\mu) e^{-\alpha r}}{\alpha^2} \, d\alpha \right| < \infty$$

for $r \geq 0$. The condition of finite moment is satisfied and the trace

$$\sum \frac{\mu^n A_n(r)}{n}, \quad r \geq 0,$$

converges absolutely.

(2) μ inside $[\mu_{-1}(0), \mu_1(0)]$.

First, we consider the simple case where Δ is always > 0 (or always < 0). For $r \geq 0$, we see that in this case $\mathfrak{D}(\mu, r)$ is always > 0 if μ is < 0 (or $\mu > 0$) [see (3a)]. Then, in this case, $V(\mu, r)$ has no poles for $r \geq 0$ and $\mu < 0$ (or $\mu > 0$), but can have poles in $\operatorname{Re} r > 0$ $\operatorname{Im} r \neq 0$, as was shown in the example given in Sec. III.A, and, consequently, $V(\mu, r)$ is not always holomorphic in $\operatorname{Re} r > 0$. In this case, inside $[-\infty, \mu_1(0)]$ or $[\mu_{-1}(0), +\infty]$, $\mathfrak{D}(\mu, r)$ has no roots for $r \geq 0$.

Secondly, we consider the case where $\mathfrak{D}(\mu, r)$ has no roots corresponding to opposite μ-values. The multiplicity in μ and r is always unity. We want to show that when $|\mu|$ increases, the new roots appear from $r = 0$ and increase with r. When $|\mu|$ increases, the number of roots of $\mathfrak{D}(\mu, r)$ increases also. We consider $\mu^{(1)}$ and $\mu^{(2)}$; $0 < \mu^{(1)} < \mu^{(2)}$ such that $\mathfrak{D}(\mu^{(1)}, r)$ has N roots for $r \geq 0$ and $\mathfrak{D}(\mu^{(2)}, r)$ has $N + 1$ roots for $r \geq 0$. We call $r^{(1)}$, $r^{(2)}$ the smallest roots of $\mathfrak{D}(\mu^{(1)}, r)$ and $\mathfrak{D}(\mu^{(2)}, r)$ [such that $\mathfrak{D}(\mu^{(1)}, r)$ or $\mathfrak{D}(\mu^{(2)}, r)$ has no roots for $r < r^{(1)}$ or $r < r^{(2)}$]. We consider $r^{(3)}$ such that $0 < r^{(3)} < \inf(r^{(1)}, r^{(2)})$. Because $\mathfrak{D}(\mu, r) \to 1$, as $r \to \infty$ for μ finite, and the

roots $\mu_{\pm j}(r)$ are simple, $\mathfrak{D}(\mu^{(1)}, r^{(3)})$ and $\mathfrak{D}(\mu^{(2)}, r^{(3)})$ have opposite signs. Because of the continuity of $\mathfrak{D}(\mu, r)$ in (μ, r), we see that there exists a $\mu^{(3)}$ such that $\mu^{(1)} < \mu^{(3)} < \mu^{(2)}$ and $\mathfrak{D}(\mu^{(3)}, r^{(3)}) = 0$. A similar proof for $\mu < 0$ shows also that the root $|\mu_{-j}(r)|$ increases when r increases. In this case, $\mathfrak{D}(\mu, r)$, inside $[\mu_{-1}(0), \mu_1(0)]$, has no roots for $r \geq 0$, V has no poles, and, following what we have said above in Sec. III.C, V satisfies the condition of finite moments.

Thirdly, we assume that, only for $|\mu| < \sup \times (|\mu_{-1}(0)|, |\mu_1(0)|)$, $\mathfrak{D}(\mu, r)$ has no opposite roots $[\mathfrak{D}(\pm\mu, r) = 0]$. Let us assume that $|\mu_{-1}(0)| > \mu_1(0)$. Then the above type of proof can be used and we still have no roots $\mu_{\pm j}(r)$, $r \geq 0$, inside $[\mu_{-1}(0), \mu_1(0)]$.

This comes from the fact that, on the one hand, $\mathfrak{D}(\mu, r) \to 1$, as $r \to \infty$, and $\mathfrak{D}(\mu, r) \to \text{const} > 0$, as $r \to 0$. On the other hand, the multiplicities of the roots in μ and r are unity, so that $\mathfrak{D}(\mu, r)$ will change sign when r crosses the root. Using the continuity in (μ, r) and the above properties, we can also eliminate more complicated cases where more than one root appears at $r \neq 0$. On the contrary, between $]\mu_1(0), -\mu_{-1}(0)]$, $\mathfrak{D}(\mu, 0)$ can have some roots $\mu_j(0)$, $j = 1, 2, \cdots, p$: We see that the corresponding roots $\mu_j(r)$ first appear at $r = 0$ as μ increases and that $\mu_j(r)$ increases when r increases.

But in the general case, where $\mathfrak{D}(\pm\mu, r)$ can vanish and the multiplicity in r is not always unity,[10] we cannot say from our results that, when $|\mu|$ increases, no new roots appear at $r \neq 0$ or that the $|\mu_{\pm j}(r)|$ are always increasing when r increases. For instance, if $|\mu_{-1}(0)| > \mu_1(0)$, from our above results, we do not know if $V(\mu, r)$ has no poles for $r > 0$ for μ inside $[\mu_{-1}(0), \mu_1(0)]$. Nevertheless, we remark that in the explicit examples to be given later, we shall always find that inside $[\mu_{-1}(0), \mu_1(0)]$ there are no roots of $\mathfrak{D}(\mu, r)$ $(r > 0)$ or that, if $\mathfrak{D}(\pm\mu, r) = 0$ exists, then μ is outside $[\mu_{-1}(0), \mu_1(0)]$. [This will be shown explicitly if there are only one or two eigenvalues $\mu_{\pm j}(r)$.]

We remark also that, if there are no roots for $r > 0$ inside $[\mu_{-1}(0), \mu_1(0)]$, then (following the proof given above in III.C) we have

$$\int_0^\infty r\,|V(r)|\,dr < \infty.$$

In this case it follows that the analytic structure of the S matrix in the physical sheet is the usual one [see (1')]. For instance, the roots of $F(x, 0) = 0$ for $x > 0$ are simple and correspond to true bound states. We have verified these features in our explicit examples.

(3) For μ outside $[\mu_{-1}(0), \mu_1(0)]$, $V(\mu, r)$ has second-order poles for $r \geq 0$, both for $\mu < 0$ and $\mu > 0$.

IV. GHOSTS AND BOUND STATES

When $V(\mu, r)$ has second-order poles for $r > 0$, the corresponding $F(x, r)$ solutions of (2) (solutions also of the Schrödinger equation), in general, have corresponding poles (at least of the first order) and are not normalizable near these poles [$\int F^2(x, r)\,dr$ is not finite]. This remains true even if

$$F(x, r) \underset{r \to 0}{\to} 0$$

(where the $F(x, 0) = 0$ are Jost functions) such that the corresponding "bound-state" wavefunctions are also, in general, nonnormalizable in these cases.

But for $|\mu|$ small enough, $V(\mu, r)$ is regular and has no poles and we can calculate the normalization constant as usual:

$$[N(\mu)]^{-1} = -(4ik^2)^{-1} f(\mu, k) \frac{\partial f}{\partial k}(\mu, -k)\Big|_{k=ix},$$

where

$$f(\mu, k = -ix) = F(x, 0) = 0, \quad x > 0. \tag{13}$$

As long as V is a well-behaved potential, we know that $N(\mu) > 0$. When $V(\mu, r)$ behaves badly, we consider the continuation in μ of $N(\mu)$ given by (13), but now N can be greater or less than zero or even complex, which leads to real or complex ghosts. (We can get complex states like resonances, but localized in the k half-plane corresponding to the physical sheet.) We recall the usual results for well-behaved potentials (see Newton[11]): the Jost functions $f(\mu, -k)$ in $\text{Im}\, k > 0$ have only simple roots localized on the imaginary axis and are interpreted as bound states; these results come from the fact that the states are normalizable, $N > 0$, and the potentials satisfy the condition of finite moments. When these necessary regularity conditions on V are violated, then the corresponding results for $N > 0$ and the location and order of the roots for the Jost functions do not, in general, remain true. We can even find different states for the same μ-value: some with $N > 0$ and others with $N < 0$ so that it is difficult to consider such types of interaction as physically acceptable (an illustration of these results will be given in Sec. VI). We want to study how ghosts and bound states appear on the physical sheet when $|\mu|$ increases [see also Ref. 3b].

[10] When we restrict Δ to have only a finite number of sign changes, we have a mathematical problem similar to one treated by K. Meetz [J. Math. Phys. **3**, 690 (1962)]. [But see the counterexample given by M. Ruby and J. R. Mines, J. Nucl. Phys. **54**, 28 (1964).]

[11] R. G. Newton, J. Math. Phys. **1**, 319 (1960).

(1) We call $\mu_c \geq \mu_0$ the largest $|\mu|$ value such that, for any $|\mu| < \mu_c$, $\mathfrak{D}(\mu, r) \neq 0$ for $r \geq 0$. We recall that, if there are no $\mathfrak{D}(\pm\mu, r)$ roots for $r > 0$ and for $\mu \in [\mu_{-1}(0), \mu_1(0)]$, then

$$\mu_c = \inf(|\mu_{-1}(0)|, \mu_1(0)).$$

By taking into account the fact that V is well behaved, we show in Appendix B that there are neither ghosts nor bound states for $|\mu| \leq \mu_c$.

(2) In the following section, we study the roots $\mu(x)$, $x > 0$, of $F(x, r = 0)\,\mathfrak{D}(\mu, r = 0) = 0$ and we find two different cases. On the one hand, if a root $\mu_{\pm j}(0)$ is such that $-\mu_{\pm j}(0)$ is not also a root of $\mathfrak{D}(\mu, r = 0)$, then a root of $F(x, r = 0)$ appears at $x = \infty$ for that value $\mu_{\pm j}(0)$. In this case, we have a real ghost appearing in the physical sheet at infinity. On the other hand, for roots $\mu_{\pm j}(0)$ such that $\mathfrak{D}(\pm\mu_{\pm j}(0), 0) = 0$, roots do not necessarily appear for $F(x, r = 0)$ at $x = \infty$. (We can find complex ghosts.) In this last case, V is not well behaved for both $\pm\mu$, at least for $r = 0$. But from $F(x = 0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r)$, we see that, for one of the two $\pm\mu$-values, the Jost function $F(x = 0, r = 0) \to 0$, as $r \to 0$, vanishes at $x = 0$.

(a) First, we consider the case $|\mu_{-1}(0)| \neq \mu_1(0)$; for example, $|\mu_{-1}(0)| > \mu_1(0)$. While for μ_1 $(r = 0)$ a real ghost appears at $x = \infty$, corresponding to $V \simeq 2/r^2$ as $r \to 0$, for $-\mu_1$ $(r = 0)$ we see from the relation $F(x = 0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r) \simeq$ const, as $r \to 0$, that a bound state appears at $x = 0$. For μ inside $[\mu_1(0), -\mu_{-1}(0)]$, other roots $\mu_2(0), \cdots, \mu_k(0)$ of $\mathfrak{D}(\mu, 0) = 0$ can exist. For the values $\mu_2(0), \cdots, \mu_k(0)$, real ghosts appear at $x = \infty$. For the values $-\mu_2(0), \cdots, -\mu_k(0)$, on the other hand, the Jost function vanishes at $x = 0$. This corresponds to well-behaved V at $r = 0$; if $\mathfrak{D}(\mu, r)$ has no $\pm\mu$-roots inside $|\mu| < |\mu_{-1}(0)|$ for $r > 0$, then V is also well behaved for $r > 0$ and the corresponding states for the $-\mu_j(0)$, $j = 1, \cdots, k$, are true bound states.

What can we say for μ outside $[\mu_{-1}(0), \mu_1(0)]$? On the one hand, for the values $\mu_{\pm j}(0)$ [but where $-\mu_{\pm j}(0)$ are not roots], real ghosts appear at ∞, and for $-\mu_{\pm j}(0)$ bound states appear at $x = 0$ corresponding to badly behaved $V(\mu, r)$ having second-order poles in $r > 0$. On the other hand, for the $\mu_{\pm j}(0)$, such that $\mathfrak{D}(\pm\mu_{\pm j}(0), 0) = 0$, real ghost states do not necessarily appear at $x = \infty$, but complex ghosts can appear as well as bound states at $x = 0$, corresponding to badly behaved $V(\mu, r)$, both for $r > 0$ and $r = 0$. Of course, the multiplicity in x of the roots $F(x, r = 0) = 0$ for the Jost functions is not necessarily unity, but can be 2, 3, $\cdots$, and complex states can appear in the physical sheet.

(b) Secondly, for the case $\mu_{-1}(0) = -\mu_1(0)$, we do not necessarily have real ghosts appearing at infinity for the two values $\pm\mu_1(0)$ (we can have complex ghosts), but still a "bound state" corresponding to a singular repulsive potential at the origin appears at $x = 0$, for one of the two values

$$F(x = 0, r) \to 0, \quad \text{as} \quad r \to 0.$$

Concerning the usual approximation where the discontinuity is reduced to its first Born term, let us consider, for instance, the potential λV of the Yukawa family and $\Delta(x, \lambda) = \sum \lambda^n \Delta_n(x)$, where Δ_n is the nth Born discontinuity. For the whole discontinuity, the Fredholm determinant, $r = 0$,equals

$$-\lambda \int_{\frac{1}{2}m}^{\infty} \frac{C(\alpha)}{\alpha^2}\, d\alpha \neq 0,$$

and the states on the physical sheet are true bound states. On the contrary, $\mathfrak{D}(\lambda)$ corresponding to the first Born term can vanish and we get ghosts as $|\lambda|$ increases. In this last case, instead of studying the problem with $\lambda V(r)$, we use, in fact, a virtual potential $V(\lambda, r)$ given by Eqs. (3) and (4) [where $\mu\Delta(x)$ is $\lambda\Delta_1(x)$ (first Born discontinuity)].

V. USEFULNESS OF SUBTRACTIONS

The solution $F(x, r = 0)$ by successive substitutions of Eq. (2) exists only for $|\mu| < \inf(|\mu_{-1}(0)|, \mu_1(0))$. This interval is limited by the smallest $|\mu|$ value for which a bound state of zero energy can appear. We consider the case $|\mu_{-1}(0)| \neq \mu_1(0)$. We remark that this result recalls another well-known result: The perturbative expansion (Born series) of the physical solution of the Schrödinger equation for S-waves, corresponding to well-behaved potentials $\lambda V(r) > 0$, has its circle of convergence given by the smallest $|\lambda|$ value for which a bound state of zero energy can appear (but in that case the series for the Jost function given by the successive substitution exists for any λ). In our case, $F(x, 0)$ is also the Jost function, so we can hope to formulate the problem in such a way that the available $|\mu|$ interval can increase. In dispersion-relation theory, the subtraction method is generally considered as a tool in order to get a better convergence of the solution (3b) or a better asymptotic behavior. Here we have assumed (8) for $\Delta(x)$, so we are not interested in the asymptotic behavior, but only try to find a larger $|\mu|$ interval where the successive substitution series converge. We write

$$F(x, r = 0) = F(x).$$

The solution of (2), $r = 0$, can be written

$$F(x) = 1 + \left(\mu \int_{\frac{1}{2}m}^{\infty} N(x, y, \mu)\, dy\right) \Big/ \mathfrak{D}(\mu), \quad (14)$$

where $\mu N(x, y, \mu)$ and $\mathfrak{D}(\mu) = \mathfrak{D}(\mu, r = 0)$ are the Fredholm numerator and denominator corresponding to the kernel $\Delta(y)/(x + y)$. The validity of this formulation of the solution, if (8) is satisfied, follows from the bounds obtained by de Alfaro and Regge.[2b]

We subtract the equation for $x = x_0 > 0$ from Eq. (2), with $r = 0$, and obtain

$$F(x) = F(x_0) + \mu \int_{\frac{1}{2}m}^{\infty} K_{x_0}(x, y) F(y)\, dy,$$

$$K_{x_0}(x, y) = (x_0 - x)\Delta(y)/(x_0 + y)(x + y). \quad (15)$$

We remark that $F(x_0)$ is a constant depending on μ. We see that, if $F(x)$ is such that $F(x_0) = 0$, for particular μ-values, then μ and $F(x)$ are eigenvalues and eigenfunctions of $K_{x_0}(x, y)$.

The Fredholm solution of (15) is

$$F(x) = F(x_0)\left[1 + \left(\mu \int_{\frac{1}{2}m}^{\infty} N_{x_0}(x, y, \mu)\, dy\right) \Big/ \mathfrak{D}_{x_0}(\mu)\right], \quad (16)$$

where $N_{x_0}(x, y, \mu)$ and $\mathfrak{D}_{x_0}(\mu)$ are the Fredholm numerator and denominator corresponding to $K_{x_0}(x, y)$.

In Appendix C it is shown that

$$\mathfrak{D}_{x_0}(\mu) = \mathfrak{D}(\mu) F(x_0) = \mathfrak{D}(\mu) + \mu \int_{\frac{1}{2}m}^{\infty} N(x_0, y, \mu)\, dy. \quad (17)$$

We see that the numerator and the denominator in (16) have the same factor $F(x_0)$. The smallest modulus root of $\mathfrak{D}_{x_0}(\mu)$ gives the $|\mu|$ interval of convergence for the substitution series of the solution of (15). If we identify the two formulations of $F(x)$ given by (14) and (16), we get

$$\int_{\frac{1}{2}m}^{\infty} (N(x, y, \mu) - N(x_0, y, \mu))\, dy = \int_{\frac{1}{2}m}^{\infty} N_{x_0}(x, y, \mu)\, dy. \quad (18)$$

Conversely, (18) can be proved directly by investigating the Fredholm numerator of $\Delta(y)/(x + y)$ and $K_{x_0}(x, y)$, so that (18) and (17) can be used to show explicitly that the solutions $F(x)$ given by (2) and by the subtracted equation (15) are identical and consequently independent of the subtraction point. Since the unknown $F(x_0)$ appears explicitly in the integral equation (15), any approximation will include this constant. But we remark that (2), for $r = 0$, is the Jost function normalized to unity when $x \to \infty$, that is,

$$F(x) \to 1, \quad \text{as} \quad x \to \infty.$$

Since the S matrix is the ratio of the two Jost functions, we are free to normalize to unity at an arbitrary point; we define $g(x) = F(x)/F(x_0)$, $g(x_0) = 1$ and

$$g(x) = 1 + \mu \int_{\frac{1}{2}m}^{\infty} K_{x_0}(x, y) g(y)\, dy, \quad (19)$$

$$g(x) = 1 + \left[\mu \int_{\frac{1}{2}m}^{\infty} N_{x_0}(x, y, \mu)\, dy\right] \Big/ \mathfrak{D}_{x_0}(\mu). \quad (20)$$

We study the $|\mu|$ interval where the substitution series of (19) converges. In order to do this, we study the roots of $\mathfrak{D}_{x_0}(\mu)$.

First, we remark that $\mathfrak{D}_{x_0}(\mu)$ is also the Fredholm determinant of

$$G_{x_0}(x, y) = (x_0 - y)\Delta(y)/(x_0 + y)(x + y).$$

This can be shown directly from the explicit form of the determinants corresponding to the kernels G_{x_0} and K_{x_0}, or we can see that the homogeneous equations for the two kernels lead to the same μ-values. If $\Delta/(x + y)$ is a polar kernel,[5,6] the same property holds for G_{x_0}. Then the roots $\bar{\mu}_{\pm j}(x_0)$, for $x_0 > 0$ fixed, are still real, with an infinite number of values greater and less than zero.

Secondly, we remark that, when $x_0 \to \infty$, $K_{x_0}(x, y) \to \Delta(y)/(x + y)$ such that $\bar{\mu}_{\pm j}$ $(x_0 = \infty) = \mu_{\pm j}$, $r = 0$. This also follows from the fact that when $x_0 \to \infty$, $F(x_0) \to 1$ and, according to (17),

$$\mathfrak{D}_{x_0}(\mu) \to \mathfrak{D}(\mu).$$

Thirdly, from (17) and (6) (for $r = 0$) we get $\mathfrak{D}_{x_0=0}(\mu) = \mathfrak{D}(-\mu)$. Thus, we see that, for $x_0 = 0$ or ∞, we have the same interval of convergence for the corresponding substitution series

$$\inf(|\mu_{-1}\ (r = 0)|, \mu_1\ (r = 0)).$$

For this reason we consider $x_0 > 0$, fixed and different from both 0 and ∞. The $|\mu|$ interval of convergence is now limited by $\inf |\bar{\mu}_{\pm j}(x_0)|$, for all j. Previously, we have seen that for

$$|\mu| < \inf |\mu_{-1}(0), \mu_1(0)|,$$

if $\mathfrak{D}(\mu, r)$ has no roots corresponding to $\pm$ μ-values for $r > 0$, then V is well behaved and there are neither ghosts nor bound states. In this case, it follows that the $|\bar{\mu}_{\pm j}(x_0)|$ are outside this interval and, consequently, for x_0 fixed, the interval of convergence $|\mu| < \inf |\bar{\mu}_{\pm j}(x_0)|$, for all j, is larger than

$$\inf(|\mu_{-1}(0)|, \mu_1(0)).$$

This is true only if the $\bar{\mu}_{\pm j}(x_0)$ are really functions of

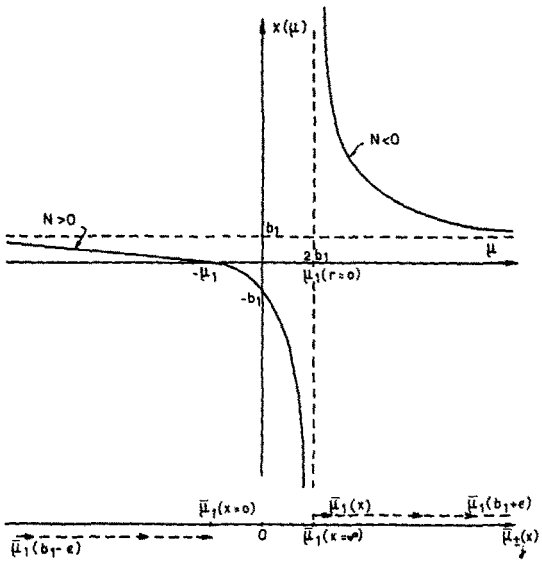


FIG. 1. $\Delta(x) = \delta(x - b_1)$.

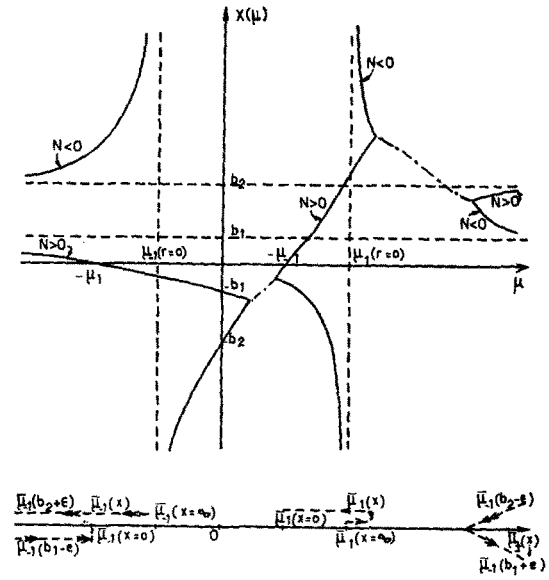


FIG. 3. $\Delta(x) = \delta(x - b_1) + a\delta(x - b_2)$, where $0 < b_1 < b_2$, $a < 0$, and $|\mu_{-1}(r = 0)| \neq \mu_1(r = 0)$.

x_0. In fact, it can happen that for special values of j, $\bar{\mu}_{\pm j}(x_0)$ is independent of x_0: $\bar{\mu}_{\pm j}(x_0) \equiv \bar{\mu}_{\pm j}(x_0 = \infty)$. For such a $\bar{\mu}$ value, $\mathfrak{D}_{x_0}(\mu)$ is identically zero, independent of x_0. We assume now that there exist such particular values. Then, because $\mathfrak{D}_{x_0}(\mu)$ reduces to $\mathfrak{D}(\pm\mu)$ when $x_0 \to \infty$ or 0, we have that these particular values are roots of both $\mathfrak{D}(\pm\mu) = 0$ and we see that this can appear only for opposite roots of $\mathfrak{D}(\mu)$. If this happens for the first root $\mu_1(r = 0) = -\mu_1(r = 0) = \bar{\mu}_1(x_0)$, for any x_0, [or $= \bar{\mu}_{-1}(x_0)$, for any x_0], we see that the interval of convergence of the substitution series cannot change by a subtraction. [We remark that this is also the case where $V(\mu, r)$ cannot be well behaved outside inf $(\mu_1(r = 0), |\mu_{-1}(r = 0)|)$.] On the contrary, if this happens for $j > 1$, that is, $j = 2, 3, \cdots$, a subtraction can still give a better interval of convergence for the substitution series.

In conclusion, for a given $x_0 \geq 0$, the $|\mu|$ interval of convergence of $g(x)$, $x > 0$, is given by the smallest $|\bar{\mu}_{\pm j}(x_0)|$ ghost or bound state which exists.

VI. SOME SIMPLE EXAMPLES

In order to illustrate the general results, we consider simple cases where $\mu\Delta(x)$ is such that (2) or (2′) has only one or two eigenvalues $\mu_{\pm j}(r)$. In Figs. 1–4, we represent the curves $x(\mu)$ and the roots of the

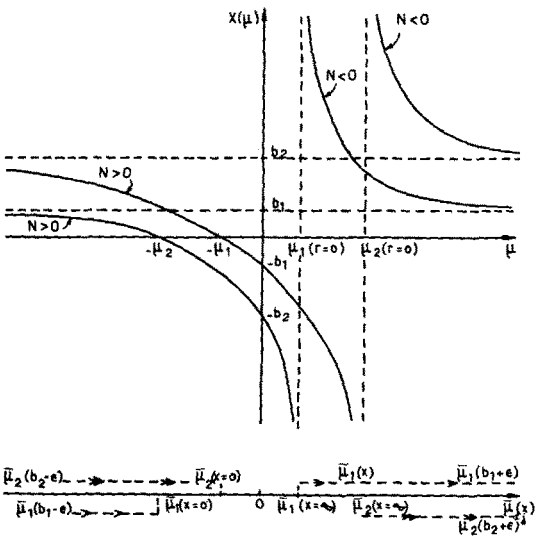


FIG. 2. $\Delta(x) = \delta(x - b_1) + a\delta(x - b_2)$, where $0 < b_1 < b_2$ and $a > 0$.

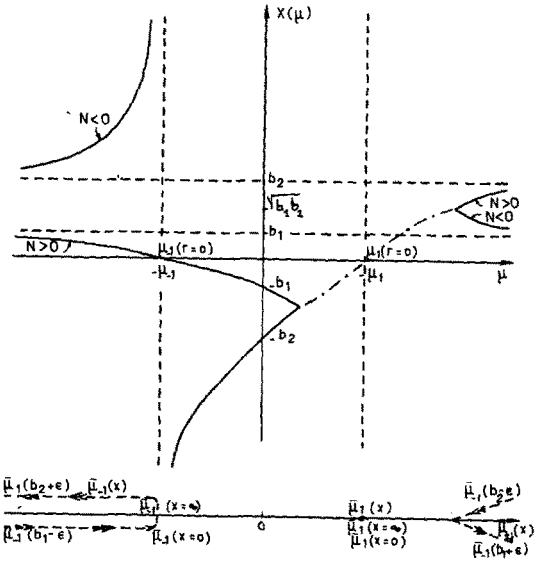


FIG. 4. $\Delta(x) = \delta(x - b_1) - (b_2/b_1)\delta(x - b_1)$, where $0 < b_1 < b_2$ and $|\mu_{-1}(r = 0)| = \mu_1(r = 0)$.

Jost functions $F(x) = 0$ when x is real and Re x when x is complex. We recall that $x(\mu)$ real and greater than zero corresponds to bound states or real ghosts. We give the trajectories $\bar{\mu}_{\pm j}(x_0)$, the roots of $\mathfrak{D}_{x_0}(\mu) = 0$, and the sign of $N(\mu)$.

A. $\Delta(x) = \delta(x - b_1), b_1 > 0; \mu_1(r) = 2b_1 e^{2b_1 r} \geq \mu_1(0)$

One real ghost appears when μ crosses $\mu_1(0)$ and one true bound state at $-\mu_1(0)$.

B. $\Delta(x) = \delta(x - b_1) + a\delta(x - b_2), 0 < b_1 < b_2$
$\mathfrak{D}(\mu, r) = 1 - \mu(e^{-2b_1 r}/2b_1 + ae^{-2b_2 r}/2b_2)$
$+ \mu^2 a e^{-2(b_1+b_2)r}(b_1 - b_2)^2/4b_1 b_2(b_1 + b_2)^2$

We have two roots $\mu_{\pm j}(r)$, both > 0 if $a > 0$, but with different sign $\mu_{\pm 1}(r)$ if $a < 0$.

(i) For $a > 0$, $\Delta > 0$, we get $\mu_1(r) > 0$, $\mu_2(r) > 0$, and $\mu_1(r) + \mu_2(r) \neq 0$; thus, the multiplicity in r of the roots is always unity and, for $r > 0$, $\mu_j(r) > \mu_j(0)$. As in the one-pole case, because there is no $\mu_{-j}(r)$, the interval $[\mu_{-1}(0), \mu_1(0)]$ in which V is well behaved is, in fact, $[-\infty, \mu_1(0)]$. The states are true bound states appearing at $-\mu_1(0)$ and $-\mu_2(0)$ for $\mu < 0$, whereas for $\mu \geq \mu_1$ we get real ghosts appearing at $\mu_1(0)$ and $\mu_2(0)$.

(ii) For $a < 0$, $a \neq -b_2/b_1$, $\mu_1(r) + \mu_{-1}(r)$ *can* vanish (but not necessarily) for one $r > 0$ value. For such $\mathfrak{D}(\pm\mu^{(1)}, r^{(1)})$, the multiplicities in r of $\pm\mu^{(1)}$, $r^{(1)}$ are 1, 3 or 3, 1, respectively, because $\partial\mathfrak{D}(\mu, r)/\partial r$ has only one root and the multiplicity in μ is unity because $\partial\mathfrak{D}(\mu, r)/\partial\mu \neq 0$ for $(\pm\mu^{(1)}, r^{(1)})$. Then inside $[\mu_{-1}(0), \mu_1(0)]$ no such root can occur, because $\mathfrak{D} \to 1$, as $r \to \infty$, and $\mathfrak{D} > 0$, as $r \to 0$. V has no poles; however, outside this interval V has poles like $2/(r - r_0)^2$, except perhaps for one exceptional value where $V \simeq 6/(r - r_0)^2$. In Fig. 3, where $|\mu_{-1}(0)| < \mu_1(0)$, we have a true bound state appearing at $-\mu_{-1}(0)$ inside $[\mu_{-1}(0), \mu_1(0)]$; for $\mu < \mu_{-1}(0)$, we have both a ghost appearing at ∞ and a bound state appearing at $-\mu_1(0)$ corresponding to a badly behaved V. For $\mu > \mu_1(0)$, the situation is still more confused because we find a real ghost appearing first at ∞ and we also find a complex state in the physical sheet of the S matrix. We find also second-order roots of the Jost functions or second-order poles for the S matrix.

(iii) For $a < 0$, $a = -b_2/b_1$, $\mu_1(0) = -\mu_{-1}(0)$, where the previous $(\pm\mu^{(1)}, r^{(1)})$ roots exist for $r = 0$. For $\mu_{-1}(0), \mu_1(0)$, the multiplicities in r are 1, 3, respectively; thus, as $r \to 0$, $V \simeq 2/r^2$, for $\mu_{-1}(0)$, and $V \simeq 6/r^2$, for $\mu_1(0)$. When μ crosses $\mu_{-1}(0)$, a real ghost appears at infinity as well as a bound state corresponding to this marginally singular potential [that is, $F(0, r) \simeq$ const $\times\, r^2$]. When μ crosses $\mu_1(0)$, complex states appear on the physical sheet; when $x(\mu) = \sqrt{(b_1 b_2)}$, they become real.

We note also that this is the only case among the given examples where $\mathfrak{D}(\pm\mu_1(0), 0) = 0$ and, consequently, this is the only case where we find a fixed root $\bar{\mu}_1(x) = \mu_1$ $(r = 0)$ for the trajectories $\bar{\mu}_{\pm j}$ $(x > 0)$. It follows that this is the only case where we cannot enlarge the $|\mu|$ interval of convergence for the substitution series by using subtractions.

C. $\Delta = \delta'(x - b)$

We get

$$\mu_{\pm 1}(r) = (2b^2)e^{2br}\{1 + 2br \pm \sqrt{[(1 + 2br)^2 + 1]}\}$$

such that

$$\mu_{\pm 1}(r)/\mu_{\pm 1}(0) > 0.$$

Then V has no second-order pole inside $[\mu_{-1}(0), \mu_1(0)]$ for $r > 0$, and the bound state appearing for $-\mu_{-1}(0)$ is a true bound state. On the contrary, outside this interval V has a second-order pole like $2/(r - r_0)^2$ only because $\mu_{+1}(r) + \mu_{-1}(r)$ is always different from zero.

D. $\mathcal{F}(r + y) = \sum_i \dfrac{\phi_i(r)\phi_i(y)}{\mu_i(0)}$

In fact, the above examples [where, following (1'), we have $\mathcal{F}(r) = e^{-b_1 r}, e^{-b_1 r} + ae^{-b_2 r}, -re^{-br}$] are particular examples of a more general case when $\mathcal{F}(x + y)$ is a degenerate $L^2(0, \infty)$ kernel with a finite number of eigenvalues $\mu_i(0)$ and orthonormalized eigenfunctions $\phi_i(y)$ such that

$$\int_0^\infty \phi_i(y)\phi_j(y)\, dy = \delta_{ij}\,.$$

[Here $\mu_i(0)$ are positive or negative.] We consider two cases as follows:

$i = 1$:

$$\mathfrak{D}(\mu, r) = 1 - \mu\int_r^\infty \frac{\phi^2(y)\, dy}{\mu(0)}$$

such that

$$\frac{\mu(r)}{\mu(0)} = \left[\int_r^\infty \phi^2(y)\, dy\right]^{-1} > 1.$$

$i = 2$: This case is studied in Appendix D. First, if $\mu_1(0)\mu_2(0) > 0$, then the $\mu_i(r)$ always have the same sign and $\mu_1(r) + \mu_2(r) \neq 0$. We can apply the results of Secs. III and IV for V here as well as for the ghosts and bound states.

Secondly, if $\mu_1(0)\mu_2(0) < 0$ [or $\mu_{\pm 1}(0) < 0$], we find in all cases that the $\mu_i(r)$ for $r > 0$ are outside $[\mu_{-1}(0), \mu_1(0)]$. This remains true if there exist roots of $\mathfrak{D}(\pm\mu, r) = 0$, because these roots are not inside the above interval. Consequently, V is well behaved

inside $[\mu_{-1}(0), \mu_1(0)]$ but, in general, V has poles like $2/(r - r_0)^2$ outside this interval; only for exceptional values can the poles be like $6/(r - r_0)^2$. If $\mu_{-1}(0) + \mu_1(0) \neq 0$, then a true bound state appears inside the above interval; outside this interval, however, real or complex ghosts and bound states corresponding to badly behaved potentials appear in all cases, as in the examples quoted in VI.B and VI.C.

E. $\mu \to \mu_{\pm j}(0)$

In this case, we must be careful, in general, about the validity of the location of the states obtained from $\lim x(\mu)$, as $\mu \to \mu_{\pm j}(0)$. For $\mu \neq \mu_{\pm j}(0)$, $V(\mu, r)$ [see (6)] is well behaved at the origin and the corresponding $F(x, 0)$ is really the Jost function. On the contrary,

$$V(\mu_{\pm j}(0), r) \underset{r \to 0}{\simeq} p(p+1)/r^2$$

and the corresponding $F(x, 0)$ is such that its Fredholm determinant vanishes. The true Jost function corresponding to $V(\mu_{\pm j}(0), r)$ is $\tilde{F}(x, \mu_{\pm j}(0)) = \lim r^p F(x,r)$, as $r \to 0$, where $F(x, r)$ is the solution of (2a) for $\mu = \mu_{\pm j}(0)$ fixed and $r > 0$, and the true roots of the Jost function for $V(\mu_{\pm j}(0), r)$ are the roots of $\tilde{F}(x, \mu_{\pm j}(0)) = 0$. Because these roots $\tilde{x}(\mu_{\pm j}(0))$ (of $\tilde{F} = 0$) and $\lim x(\mu)$, as $\mu \to \mu_{\pm j}(0)$, are obtained by different limiting processes, it does not follow that we shall always get the same values. For instance, in the previous cases VI.A and VI.B for $a > 0$ and $a < 0$, $a \neq -b_2/b_1$, or in the previously considered case, $a = -b_2/b_1$, $\mu = \mu_{-1}(0)$, we find the same roots; but for $a = -b_2/b_1$, $\mu = \mu_1(0)$, if we investigate the true Jost solution, then it is easy to see that

$$\lim_{\mu \to \mu_1(0)} x(\mu) = \pm i(b_1 b_2)^{\frac{1}{2}} \neq \tilde{x},$$

corresponding to $V(\mu_1(0), r)$. Nevertheless, in the case of opposite roots $\mathfrak{D}(\pm\mu_j(0), 0) = 0$, because of the relation

$$F(x = 0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r),$$

we see that, for one of the two values $\pm\mu_j(0)$ such that $F(x = 0, r) \to 0$, as $r \to 0$ (see Sec. III), $x = 0$ is necessarily a root of the true Jost function corresponding to one of the two $V(\pm\mu_j(0), r)$.

VII. CONCLUSION

We have seen in this paper[12] that, for "regular" discontinuities $\mu\Delta(x)$, there always exists a restricted interval including $\mu = 0$ [although one of the limits of this interval can be $+\infty$ (or $-\infty$) if the eigenvalues of the Marchenko equations or, equivalently, those of the resulting integral equation of f/f have always the same sign] such that the $V(\mu, r)$ are well behaved. We find a slight generalization of the Yukawa-type family (Laplace transform, regular at the origin, not always holomorphic in $\mathrm{Re}\, r > 0$, but with no poles for $r > 0$, and $|V|$ satisfying the condition of finite moments). Consequently, inside this interval the analytic structure of the S matrix (concerning order and location of the poles) is the usual one corresponding to well-behaved potentials.

Also, there always exists an integral including the origin such that, if μ is outside this interval, $V(\mu, r)$ is badly behaved with second-order poles for $r \geq 0$. Consequently, some usual features of the $S(k)$ matrix in the k half-plane corresponding to the physical sheet are not conserved. The Jost functions can have roots, not always of the first order, and these roots are not always located on the imaginary k axis. We find real or complex ghosts or real or complex bound states corresponding to these badly behaved potentials.

In these papers we have not considered the case of singular $\Delta(x)$ or the case of regular $\Delta(x)$ with $l \neq 0$, but we hope to be able to extend the present results to these cases. Concerning the problem $l \neq 0$, we remark that our previous argument (see Ref. 1, Introduction), which was at the origin of the investigation of Fredholm determinant, is still valid. If we consider a Yukawa-type family $\lambda V(r)$, it is also true that the Jost function $f_l(k) = D_l(k^2)$ exists for all λ, so that even if the Fredholm determinant still has an infinity of roots corresponding to singular values, it must also exhibit in a way (that we hope is simple) the property that it cannot vanish for any λ.

Perhaps the main result of these papers is the following:

(1) In potential scattering, if we consider the interaction as given by a $V(r)$ which is regular, marginally singular, or singular, then λV has, in general, the same feature. (For a rough analogy with Q.F.T., we can speak also of similar features for super-renormalizable, renormalizable, or nonrenormalizable interactions.)

(2) On the contrary, in dispersion relations, if we consider the interaction as given by a regular discontinuity $\Delta(x)$, then $\mu\Delta(x)$ (interpreted in terms of a potential or interpreted by the corresponding states, bound states and ghosts) can give two types of different interactions: regular or marginally singular. But the μ-intervals where we find the two types of different interactions are separated.

[12] *Note Added in Proof:* We want to point out that many generalizations of the results presented in this paper are now published: H. Cornille, Nucl. Phys. **B3**, 655 (1967); H. Cornille and G. Rubinstein, J. Math. Phys. **9**,1501 (1968); Nuovo Cimento **56**, 867 (1968); H. Cornille, J. Math. Phys. **11**, 79 (1970).

APPENDIX A

We study the solution $F(x, r)$ for $x = 0$ of the integral equation

$$F(x, r) = e^{-xr} + \mu \int_{\frac{1}{2}m}^{\infty} \frac{\Delta(y) e^{-(x+y)r}}{(x + y)} F(y, r)\, dy. \quad \text{(A1)}$$

We want to show that the Fredholm solution $F(0, r)$ of (A1) can be written

$$F(0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r), \quad \text{(A2)}$$

where $\mathfrak{D}(\mu, r)$ is the Fredholm denominator determinant of (A1) given by (3a).

The Fredholm solution of (A1) is

$$F(0, r) = 1 + \mathcal{N}(\mu, r)/\mathfrak{D}(\mu, r), \quad \text{(A3)}$$

where

$$\mathcal{N}(\mu, r) = \sum_{1}^{\infty} \frac{\mu^n (-1)^{n-1}}{(n-1)!} \times \int_{\frac{1}{2}m}^{\infty} du_1 \cdots \int_{\frac{1}{2}m}^{\infty} du_n \left(\prod_{1}^{n} \Delta(u_i) e^{-2ru_i} \right) B_n,$$

$$B_n(u_1, \cdots, u_n) = \begin{vmatrix} \frac{1}{u_1} & \frac{1}{u_2} & \cdots & \frac{1}{u_n} \\ \frac{1}{u_1 + u_2} & \frac{1}{2u_2} & \cdots & \frac{1}{u_2 + u_n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{1}{u_1 + u_n} & \frac{1}{u_2 + u_n} & \cdots & \frac{1}{2u_n} \end{vmatrix}.$$

B_n is the same determinant as P_n given in (3a), but the first row is replaced by $1/u_1, \cdots, 1/u_n$.

$\mathfrak{D}(\mu, r)$ can be written

$$\mathfrak{D}(\mu, r) = \sum_{0}^{\infty} \frac{(-\mu)^n}{n!} \times \int_{\frac{1}{2}m}^{\infty} du_1 \cdots \int_{\frac{1}{2}m}^{\infty} du_n \left(\prod_{1}^{n} \Delta(u_i) e^{-2ru_i} \right) A_n, \quad \text{(A4)}$$

where

$$A_n(u_1, \cdots, u_n) = \begin{vmatrix} 1 & \frac{1}{u_1} & \frac{1}{u_2} & \cdots & \frac{1}{u_n} \\ 0 & \frac{1}{2u_1} & \frac{1}{u_1 + u_2} & \cdots & \frac{1}{u_1 + u_n} \\ 0 & \frac{1}{u_1 + u_2} & & \cdots & \frac{1}{u_2 + u_n} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 0 & \frac{1}{u_1 + u_n} & & \cdots & \frac{1}{2u_n} \end{vmatrix}.$$

We note that in A_n the minor corresponding to the element $(A_n)_{11}$ is P_n.

Now we consider C_n, given by

$$C_n = \begin{vmatrix} 0 & \frac{1}{u_1} & \cdots & \frac{1}{u_k} & \cdots & \frac{1}{u_n} \\ -1 & \frac{1}{2u_1} & \cdots & \frac{1}{u_1 + u_k} & \cdots & \frac{1}{u_1 + u_n} \\ -1 & \frac{1}{u_1 + u_2} & & & & \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ -1 & \frac{1}{u_1 + u_n} & & & \cdots & \frac{1}{2u_n} \end{vmatrix}.$$

In C_n, the minor corresponding to $(C_n)_{11}$ is still P_n. We develop C_n by following the elements of the first column, where we call $C_{k,n}$ the minor corresponding to the $(k + 1)$th element of the first column. We remark that $C_{1,n} = B_n$: We want to show that, for any $k = 1, 2, \cdots, n$,

$$\int_{\frac{1}{2}m}^{\infty} du_1 \cdots \int_{\frac{1}{2}m}^{\infty} du_n \left(\prod_{1}^{n} \Delta(u_i) e^{-2ru_i} \right) \times [C_{k,n}(u_i) + (-1)^k C_{1,n}(u_i)] = 0. \quad \text{(A5)}$$

For this, in $C_{k,n}$ we make the following substitutions: column 1 $\rightarrow$ column 2, column 2 $\rightarrow$ column 3, $\cdots$, column $k - 1 \rightarrow$ column k, and column $k \rightarrow$ column 1. Then we get a new determinant $\bar{C}_{k,n} = (-1)^{k-1} C_{k,n}$. If in $\bar{C}_{k,n}$ we put $u_1 = u_2$, $u_2 = u_3, \cdots, u_{k-1} = u_k$, $u_k = u_1$, we remark that $\bar{C}_{k,n}$ becomes $C_{1,n} = B_n$. Then, using the fact that the factor

$$\prod_i \Delta(u_i)\, e^{-2ru_i}$$

is symmetric with respect to all the variables u_i, we see that with the previous change of variables we get the relation (A5). From (A5) it follows that

$$\int_{\frac{1}{2}m}^{\infty} du_1 \cdots \int_{\frac{1}{2}m}^{\infty} du_n \left(\prod_{1}^{n} \Delta(u_i) e^{-2ru_i} \right) (C_n - nB_n) = 0. \quad \text{(A6)}$$

Taking (A6) into account in (A3), we get

$$\mathfrak{D}(\mu, r) + \mathcal{N}(\mu, r) = \sum_{0}^{\infty} \frac{(-\mu)^n}{n!} \int_{\frac{1}{2}m}^{\infty} du_1 \cdots \int_{\frac{1}{2}m}^{\infty} du_n \times \left(\prod_{1}^{n} \Delta(u_i) e^{-2ru_i} \right) (A_n - C_n). \quad \text{(A7)}$$

With some algebra, we get $A_n - C_n = -(1)^n P_n$; so the right-hand side of (A7) is $\mathfrak{D}(-\mu, r)$ and, finally, the result (A2) follows.

APPENDIX B

We want to find a bound for $d\mathfrak{D}(\mu, r)/dr$ when $\operatorname{Re} r > 0$. From (3a) we get

$$\frac{d}{dr}\mathfrak{D}(\mu, r) = \sum_1^\infty \frac{(-\mu)^n}{n!}\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \times \left[\prod_1^n \Delta(u_i)e^{-2ru_i}\right] P_n(u_1, \cdots, u_n).$$

But de Alfaro and Regge have obtained the bound $\prod_1^n 1/2u_i$ for P_n. Then

$$|P_n \textstyle\sum 2u_i| < \sum_i \left[\prod_{j=1, j\neq i}^n 2u_j\right]^{-1}.$$

Taking into account the fact that the factor

$$\prod_1^n \Delta(u_i)e^{-2\operatorname{Re} ru_i}$$

is symmetric with all the variables u_i, we get

$$\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n |\Delta(u_i)e^{-2\operatorname{Re} ru_i|}\right) |P_n \textstyle\sum 2u_i| < n\int_{\frac{1}{2}m}^\infty |\Delta(u)|\, e^{-2\operatorname{Re} ru}\, du \times \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_{n-1}\left(\prod_1^{n-1} \frac{|\Delta(u_i)|}{2u_i}\, e^{-2\operatorname{Re} ru_i}\right)$$

and

$$\left|\frac{d}{dr}\mathfrak{D}(\mu, r)\right| < \sum_1^\infty \frac{|\mu|^n}{(n-1)!}\int_{\frac{1}{2}m}^\infty |\Delta(u)e^{-2\operatorname{Re} u}|\, du \times \left[\int_{\frac{1}{2}m}^\infty \frac{|\Delta(v)|}{2v}\, e^{-2\operatorname{Re} v}\, dv\right]^{n-1}$$

or

$$\left|\frac{d}{dr}\mathfrak{D}(\mu, r)\right| < \int_{\frac{1}{2}m}^\infty |\mu|\, |\Delta(u)e^{-2\operatorname{Re} ru}|\, du \times \left[\exp\int_{\frac{1}{2}m}^\infty \frac{|\mu|\, |\Delta(u)|}{2u}\, e^{-2\operatorname{Re} u}\, du\right].$$

We want to find a bound for $d^2\mathfrak{D}(\mu, r)/dr^2$ when $\operatorname{Re} r > 0$. From (3a) we get

$$\frac{d^2}{dr^2}\mathfrak{D}(\mu, r) = \sum_1^\infty \frac{(-\mu)^n}{n!}\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \times \left[\prod_1^n \Delta(u_i)e^{-2ru_i}\right] P_n\left(\sum_i 2u_i\right)^2. \quad \text{(B1)}$$

Using the same considerations for $|P_n \sum 2u_i|$ as we used above for $\mathfrak{D}'(\mu, r)$, we get

$$\left|\frac{d^2}{dr^2}\mathfrak{D}(\mu, r)\right| < \sum_1^\infty \frac{|\mu|^n}{(n-1)!}\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \times \left[\prod_1^n |\Delta(u_i)|\, e^{-2\operatorname{Re} ru_i}\right]\left(\sum_1^n 2u_j\right)\left(\prod_{k=1}^{n-1} \frac{1}{2u_k}\right).$$

We write

$$\sum_1^n 2u_j = 2u_n + \sum_1^{n-1} 2u_j$$

and get two terms for the bounds. The first is

$$\sum_1^\infty \frac{|\mu|^n}{(n-1)!}\left(\int_{\frac{1}{2}m}^\infty du_n\, |\Delta(u_n)|\, e^{-2\operatorname{Re} ru_n}\right) \times \left(\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_{n-1} \prod_1^{n-1} \frac{|\Delta(u_i)|}{2u_i}\, e^{-2\operatorname{Re} ru_i}\right).$$

For the second term a factor

$$\left(\sum_1^{n-1} 2u_j\right)\prod_{k=1}^{n-1} \frac{1}{2u_k}$$

appears. Using the symmetry properties of

$$\prod_1^{n-1} |\Delta(u_i)|\, e^{-2\operatorname{Re} ru_i},$$

we get the following bound for the second term:

$$\sum_1^\infty \frac{|\mu|^n}{(n-2)!}\left(\int_{\frac{1}{2}m}^\infty du_n\, |\Delta(u_n)|\, e^{-2\operatorname{Re} ru_n}\right) \times \left(\int_{\frac{1}{2}m}^\infty du_{n-1}\, |\Delta(u_{n-1})|\, e^{-2\operatorname{Re} ru_{n-1}}\right) \times \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_{n-2} \prod_2^{n-2} \frac{|\Delta(u_j)|}{2u_j}\, e^{-2\operatorname{Re} u_j}.$$

Finally, for $|\mathfrak{D}''|$ we get

$$|\mathfrak{D}''| < \left(\exp\int_{\frac{1}{2}m}^\infty |\mu|\, \frac{|\Delta(u)|}{2u}\, e^{-2\operatorname{Re} ru}\, du\right) \times \left[\int_{\frac{1}{2}m}^\infty |\mu|\, 2u\, |\Delta(u)|\, e^{-2\operatorname{Re} ru}\, du + \left(\int_{\frac{1}{2}m}^\infty |\mu|\, |\Delta(u)|\, e^{-2\operatorname{Re} ru}\, du\right)^2\right]. \quad \text{(B2)}$$

We want to show that, if (8) is satisfied, that is,

$$\int_{\frac{1}{2}m}^\infty \frac{|\Delta(u)|}{u}\, du < \infty,$$

then

$$M = \int_0^\infty r\, |\mathcal{F}'(r)|\, dr < \infty,$$

where $\mathcal{F}(r)$ is the scattering data (1″). For this we define

$$M(\epsilon) = \int_\epsilon^\infty r\, |\mathcal{F}'(r)|\, dr,$$

where ϵ is greater than zero and small. We remark that, if $0 < \epsilon_2 < \epsilon_1$, then $0 < M(\epsilon_1) < M(\epsilon_2)$. We want to show that $\lim M(\epsilon)$, as $\epsilon \to 0$, exists and is

bounded. We have, for ϵ finite,

$$M(\epsilon) \leq |\mu| \int_{\epsilon}^{\infty} dr \left(\int_{\frac{1}{2}m}^{\infty} e^{-ru} u \, |\Delta(u)| \, du \right) dr. \quad \text{(B3)}$$

We integrate first in r and we get

$$M(\epsilon) \leq |\mu| \, [\sum M_1(\epsilon) + M_2(\epsilon)],$$

where

$$M_1(\epsilon) = |u| \int_{\frac{1}{2}m}^{\infty} e^{-\epsilon u} \left| \frac{\Delta(u)}{u} \right| du$$

$$< |\mu| \int_{\frac{1}{2}m}^{\infty} \left| \frac{\Delta(u)}{u} \right| du < \infty$$

and

$$M_2(\epsilon) = \epsilon \int_{\frac{1}{2}m}^{\infty} u e^{-\epsilon u} \left| \frac{\Delta(u)}{u} \right| du.$$

We have

$$u e^{-\epsilon u} \leq (\epsilon e)^{-1}.$$

Thus

$$M_2(\epsilon) < e^{-1} \int_{\frac{1}{2}m}^{\infty} \left| \frac{\Delta(u)}{u} \right| du < \infty,$$

as $\epsilon \to 0$, and $\lim M(\epsilon)$ exists and is bounded.

We want to show that, if (8) is satisfied and if $\mathfrak{D}(\mu, r)$ does not vanish in a small interval $r \in [0, b]$ (b small but finite), then

$$\int_0^{\infty} r \, |V(r)| \, dr < \infty.$$

We recall that $V = 2(\mathfrak{D}'^2 - \mathfrak{D}\mathfrak{D}'')/\mathfrak{D}^2$, so that, with the bounds (B1) for $\mathfrak{D}'$ and (B2) for $\mathfrak{D}''$, we get

$$\int_0^b r \, |V(r)| \, dr < \text{const}_1 \times \int_0^b r \left(\int_{m/2}^{\infty} u \, |\Delta(u)| \, e^{-2ur} du \right) dr$$

$$+ \text{const}_2 \times \int_0^b r \left(\int_{m/2}^{\infty} |\Delta(u)| \, e^{-2ur} \, du \right)^2 dr.$$

But the first term on the right of the inequality is bounded by a constant because we have shown above that $\lim M(\epsilon) < \text{const}$ as $\epsilon \to 0$ [see (B3)]. For the second term we apply the Schwarz relation and we get

$$\int_0^b r \, |V(r)| \, dr$$

$$< \text{const} + \text{const} \times \int_0^b r \left[\int_{\frac{1}{2}m}^{\infty} |\Delta(u)| \, e^{-2ur} u \, du \right.$$

$$\left. \times \int_{\frac{1}{2}m}^{\infty} \left| \frac{\Delta(v)}{2v} \right| e^{-2vr} \, dv \right] dr < \infty.$$

We want to show that, if the $\mathfrak{D}(\pm\mu, r)$ have no roots for $r \geq 0$ and $|\mu| < \mu_C$, then $F(x, 0)$ for the same interval $|\mu| < \mu_C$ cannot vanish for $x > 0$, so that there are no bound states. For this we will show that, if $F(x = 0, r)$ has no root for $r \geq 0$, then this is also true for $x > 0$ [$F(x, r) \neq 0$, $x > 0$, $r \geq 0$]. We recall that $F(x, r) \simeq e^{-xr}$, as $r \to \infty$. In this $|\mu|$ range, V is well behaved, $F(0, r) = \mathfrak{D}(-\mu, r)/\mathfrak{D}(\mu, r) > 0$ for $r \geq 0$, and $F(0, r) \to 1$ as $r \to \infty$. By the usual combination of the Schrödinger equation for $F(0, r)$ and $F(x, r)$, we get

$$F'(0, r_0)F(x, r_0) - F'(x, r_0)F(0, r_0)$$

$$= \int_{r_0}^{\infty} x^2 F(x, r) F(0, r) \, dr.$$

We assume that $r_0 \geq 0$ is the first root of $F(x, r)$ for $r \in [r_0, \infty]$. Then $F'(x, r_0) \geq 0$ and the left-hand side is less than zero. But the right-hand side is greater than zero and we get a contradiction. Thus $F(x, r) \neq 0$ for $r \geq 0$.

APPENDIX C

We study the Fredholm solutions (14)–(16). We have

$$F(x) - F(x_0)$$

$$= \left[\mu \int_{\frac{1}{2}m}^{\infty} (N(x, y, \mu) - N(x_0, y, \mu)) \, dy \right] [\mathfrak{D}(\mu)]^{-1}$$

$$= \left[\mu \int_{\frac{1}{2}m}^{\infty} N_{x_0}(x, y, \mu) \, dy \right] [\mathfrak{D}_{x_0}(\mu)]^{-1} F(x_0),$$

where $\mu N(x, y, \mu)$ and $\mathfrak{D}(\mu)$ are the Fredholm numerator and denominator of (2a) corresponding to the kernel $\Delta(y)/(x + y)$. Similarly μN_{x_0} and $\mathfrak{D}_{x_0}$ are the Fredholm numerator and denominator of (15) corresponding to the kernel

$$\Delta(y)(x_0 - x)/(x + y)(x_0 + y).$$

The three relations

$$\mathfrak{D}_{x_0}(\mu) = F(x_0)\mathfrak{D}(\mu), \quad \text{(C1)}$$

$$\mu \int_{\frac{1}{2}m}^{\infty} N_{x_0}(x_0, y, \mu) \, dy$$

$$= \mu \int_{\frac{1}{2}m}^{\infty} (N(x, y, \mu) - N(x_0, y, \mu)) \, dy, \quad \text{(C2)}$$

$$\mathfrak{D}_{x_0}(\mu) = \mathfrak{D}(\mu) + \mu \int_{\frac{1}{2}m}^{\infty} N(x_0, y, \mu) \, dy \quad \text{(C3)}$$

are equivalent. We now want to prove (C3). We have

$$\mathfrak{D}_{x_0}(\mu) = 1 + \sum_1^{\infty} \frac{(-\mu)^n}{n!} \int_{\frac{1}{2}m}^{\infty} du_1 \cdots \int_{\frac{1}{2}m}^{\infty} du_n$$

$$\times \left(\prod_1^n \Delta(u_i) \right) D_n, \quad \text{(C4)}$$

where

$$D_n(x_0, u_1, \cdots, u_n)$$

$$= \begin{vmatrix} \frac{1}{2u_1} - \frac{1}{x_0 + u_1} & \cdots & \frac{1}{u_1 + u_n} - \frac{1}{x_0 + u_n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{1}{u_1 + u_n} - \frac{1}{x_0 + u_1} & \cdots & \frac{1}{2u_n} - \frac{1}{x_0 + u_n} \end{vmatrix}.$$

The first term on the right-hand side of (C3) is

$$\mathfrak{D}(\mu) = \sum_0^\infty \frac{(-\mu)^n}{n!} \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n \Delta(u_i) \right) \times L_n(x_0, u_1, \cdots, u_n) \quad \text{(C5)}$$

where

$$L_n = P_n = \begin{vmatrix} 1 & \frac{1}{x_0+u_1} & \frac{1}{x_0+u_2} & \cdots & \frac{1}{x_0+u_n} \\ 0 & \frac{1}{2u_1} & \frac{1}{u_1+u_2} & \cdots & \frac{1}{u_1+u_n} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 0 & \frac{1}{u_1+u_n} & & \cdots & \frac{1}{2u_n} \end{vmatrix}.$$

The second term is

$$\mu \int_{\frac{1}{2}m}^\infty N(x_0, y, \mu)\, dy = \int_1^\infty \frac{\mu^n(-1)^{n-1}}{(n-1)!} \int_{m/2}^\infty du_1 \cdots \int_{m/2}^\infty du_n \left(\prod_1^n \Delta(u_i) \right) E_n, \quad \text{(C6)}$$

where

$$E_n = \begin{vmatrix} \frac{1}{x_0+u_1} & \frac{1}{x_0+u_2} & \cdots & \frac{1}{x_0+u_n} \\ \frac{1}{u_1+u_2} & \frac{1}{2u_2} & \cdots & \frac{1}{u_2+u_n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \frac{1}{u_1+u_n} & & \cdots & \frac{1}{2u_n} \end{vmatrix}.$$

Now we consider M_n, given by

$$M_n = \begin{vmatrix} 0 & \frac{1}{x_0+u_1} & \cdots & \frac{1}{x_0+u_k} & \cdots & \frac{1}{x_0+u_n} \\ -1 & \frac{1}{2u_1} & \cdots & \frac{1}{u_1+u_k} & \cdots & \frac{1}{u_1+u_n} \\ -1 & & & & & \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ -1 & \frac{1}{u_1+u_n} & & \cdots & & \frac{1}{2u_n} \end{vmatrix}. \quad \text{(C7)}$$

Note that the minor corresponding to $(L_n)_{11}$ or $(M_n)_{11}$ is P_n. We remark that M_n is the same determinant as C_n given in Appendix A, except for the elements of the first row. We develop M_n by following the elements of the first column, where we call $M_{k,n}$ the minor corresponding to the $(k+1)$th element of the first column. We remark that $M_{1,n} = E_n$ and also that $M_{k,n}$ is the same determinant as $C_{k,n}$, except for the elements of the first row. We want to show that

$$\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n \Delta(u_i) \right) \times [M_{k,n} + (-1)^k M_{1,n}] = 0. \quad \text{(C8)}$$

We follow the same method as in Appendix A. In $M_{k,n}$ we make the following substitutions: column $1 \to$ column 2, column $2 \to$ column 3, $\cdots$, column $k-1 \to$ column k, column $k \to$ column 1. We get a new determinant $\bar{M}_{k,n}$ such that $\bar{M}_{k,n} = (-1)^{k-1} M_{k,n}$; $\bar{M}_{k,n}$ is the same determinant as $\bar{C}_{k,n}$ except for the elements of the first row. If in $\bar{M}_{k,n}$ we put $u_1 = u_2$, $u_2 = u_3, \cdots$, $u_{k-1} = u_k$, $u_k = u_1$, we remark that $\bar{M}_{k,n}$ becomes $\bar{M}_{1,n} = E_n$. Then, taking into account the fact that the factor $\prod_j^n \Delta(u_i)$ is symmetric with respect to all the variables u_i, we get the relation (C8).

It follows that

$$\int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n \Delta(u_i) \right) (E_n - nM_n) = 0$$

and

$$\mathfrak{D}(\mu) + \mu \int_{\frac{1}{2}m}^\infty N(x_0, y)\, dy = \sum_0^\infty \frac{(-\mu)^n}{n!} \int_{\frac{1}{2}m}^\infty du_1 \cdots \int_{\frac{1}{2}m}^\infty du_n \left(\prod_1^n \Delta(u_i) \right) (L_n - M_n).$$

With some algebra, we get $L_n - M_n = D_n$, and so the relation (C3) follows.

APPENDIX D

We assume that the scattering data [see (1″)]

$$\mathcal{F}(y+r) = \sum_{i=1}^2 \frac{\phi_i(y)\phi_i(r)}{\mu_i(0)}$$

kernel of (2b) has only two eigenvalues $\mu_i(0)$ and two orthonormalized eigenfunctions

$$\int_0^\infty \phi_i(y)\phi_j(y)\, dy = \delta_{i,j}$$

for $r = 0$. We call

$$\eta_i = \int_r^\infty \phi_i^2(y)\, dy \leq 1$$

and

$$\zeta = \int_r^\infty \phi_1(y)\phi_2(y)\, dy.$$

Then

$$\mathfrak{D}(\mu, r) = 1 - \mu A_1(r) + \mu^2 A_2(r),$$

where

$$A_1(r) = \frac{\eta_1}{\mu_1(0)} + \frac{\eta_2}{\mu_2(0)}, \quad A_2(r) = \frac{\eta_1\eta_2 - \zeta^2}{\mu_1(0)\mu_2(0)},$$

and $\eta_1\eta_2 - \zeta^2 > 0$ from the Schwarz inequality. We want to show that the $\mu_i(r)$ roots of $\mathfrak{D}(\mu, r)$ are outside $[\mu_{-1}(0), \mu_1(0)]$.

1. We assume $\mu_1(0)\mu_2(0) > 0$. Then $\mathfrak{D}(\mu, r)$ has no roots ($\mu_1\mu_2 > 0$) corresponding to $\pm\mu$ values and we can apply the results of Sec. III.

2. $\mu_1(0)\mu_2(0) < 0$. Then $\mu_1(r) + \mu_2(r)$ can vanish (we call $\mu_2 = \mu_{-1}$) if $A_1(r) = 0$ for some r_j value. If this happens for two values r_1, r_2 because A_2 is reduced to

$$\int_r^\infty \int_r^\infty (\mathcal{F}(y + r))^2 \, dy \, dr,$$

then the corresponding μ-values are different.

Further, because $\partial\mathfrak{D}(\mu, r)/\partial r = 0$ has only one root $\mu = -A_1'/A_2'$, only one of $\pm\mu(r_j)$ can be of multiplicity 3 in r, whereas the other is simple. Finally, such μ value cannot be inside $[\mu_{-1}(0), \mu_1(0)]$ because $\mathfrak{D} \to 1$ ($r \to \infty$), $\mathfrak{D} > 0$ ($r \to 0$), and only one root with multiplicity odd is possible.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Existence and Uniqueness of Crossing Symmetric N/D-Type Equations Corresponding to the Klein–Gordon Equation

H. CORNILLE*
Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay
BP. n° 2-91-GIF-sur-YVETTE, France

(Received 24 February 1969)

N/D-type equations satisfying crossing symmetry are established inside the relativistic wave-mechanics formalism (Klein–Gordon equations). We show that, for a superposition of exponential potentials with finite zeroth moment, the N/D-type equations have unique solutions. Furthermore, for sufficiently weak couplings the solutions are physically available.

I. INTRODUCTION

One of the most interesting problems in the dynamics of strong interactions is to understand the possibility of reconstructing the partial waves from the corresponding left-hand cut discontinuities. Generally, one uses the N/D method. The full problem presents many difficulties so that it appears convenient to consider models successively close to the actual case. In the most simple model, the nonrelativistic potential, the existence and uniqueness of N/D equations corresponding to Yukawa family are characterized by a reduction formula of the Fredholm determinant.[1] Other models are the Klein–Gordon formalism, the Bethe–Salpeter equation, or the elastic unitarity approximation. In this paper we consider relativistic potentials of the exponential type corresponding to Klein–Gordon[2–5] equations where the following important features prevent us from applying the procedure used in the nonrelativistic case:

(i) The phase shift does not tend to zero when the energy E goes to infinity. As a consequence, besides the usual dynamical cuts due to the potentials, there exist supplementary branch points in the k momentum plane $[E = (k^2 + m^2)^{\frac{1}{2}}]$. The Jost functions have cuts in both lower and upper half-k planes. There is no simple factorization that leads with only cuts in one half-plane.

(ii) Due to the occurrence of the two-valued energy quantities $E = \pm(k^2 + m^2)^{\frac{1}{2}}$ and the fact that energy and potentials appear like an EV product, the particle and antiparticle scattering are related by a simple crossing relation.[4]

(iii) We use property (ii) in order to solve the difficulty explained in (i). With the two discontinuities of the particle and antiparticle scattering, we get N/D-type equations for two functions extracted from the Jost function, but which have only dynamical cuts in one half-plane. The advantage is that we have the possibility of reconstructing the partial wave for both particle and antiparticle scattering.

We study first the on-the-mass-shell equations. We obtain the relations which give the possibility of reconstructing the discontinuities from the potentials (and conversely). We show that the eigenvalues of the N/D equations must correspond to the particular values of the coupling strength for which the phase shift at infinity goes like $(2m + 1)\pi/2$.

Secondly, we consider the off-the-mass-shell equations (radial coordinate different from zero). For nonrelativistic potentials we recall[1] that the Marchenko inversion formalism was used to understand N/D formalism. In the Klein–Gordon case we get the extension of Marchenko equations with the possibility of reconstructing the potentials from the S-matrix discontinuities. We have also obtained integral equations for the Jost solutions with kernels proportional to the discontinuities; for these equations we have investigated the roots of the Fredholm determinants. From this off-the-mass-shell formalism we determine explicitly the reduction formula of the Fredholm determinant corresponding to the on-the-mass-shell equations:

$$\mathfrak{D}(\lambda_1) = \cos \lambda_1 \int_0^\infty \tfrac{1}{2} V_1 \, dr.$$

So we are able to prove explicitly that there exists for $\int_0^\infty \lambda_1 V \, dr$ finite a unique solution of the N/D equations which can be used in order to reconstruct the partial wave for both particle and antiparticle scattering.

---

* Centre National de la Recherche Scientifique.

[1] H. Cornille, J. Math. Phys. **8**, 2268 (1967); **11**, 61 (1970) (preceding article).

[2] For general properties see H. Feshbach and F. Villars, Rev. Mod. Phys. **30**, 24 (1958); E. Corinaldesi and F. Strocchi, *Relativistic Wave Mechanics* (North-Holland Publ. Co., Amsterdam, 1965).

[3] For the inversion problem see E. Corinaldesi, Nuovo Cimento **5**, 468 (1954); M. Verde, Nucl. Phys. **9**, 255 (1958); V. De Alfaro, Nuovo Cimento **4**, 675 (1958).

[4] For the use of crossing relations see R. Oehme, Nuovo Cimento **25**, 183 (1962).

[5] For off-shell N/D method see M. Petrãs, Nucl. Phys. **87**, 141 (1966). Note that in the present paper our aim is to get solutions with cuts in only one half-k plane.

Thirdly, contrary to the nonrelativistic case, regular properties of the coordinate behavior of the potentials are not sufficient to obtain only available physical states. Indeed, complex frequencies[6] (zero-norm pathological states) can occur. We show that for sufficiently weak coupling constants (of the potentials), our unique N/D-type solutions do not lead to such difficulties.

We consider only the S-wave Klein–Gordon equation and we begin by a brief report of the properties for the nonrelativistic potential case.

II. BASIC RESULTS FOR THE EXISTENCE OF N/D EQUATIONS IN NONRELATIVISTIC POTENTIAL THEORY

We recall briefly the results for the S wave for Yukawa family:

$$\lambda V(r) = \lambda \int_{\mu}^{\infty} e^{-\alpha r} C(\alpha)\, d\alpha, \quad \int_{\mu}^{\infty} \frac{C(\alpha)}{\alpha^2}\, d\alpha < \infty. \tag{1}$$

The Jost function $f(k)$ satisfies the equation

$$[F(x) = f\ (k = -ix)];$$

$$F(x) = 1 + \int_{\frac{1}{2}\mu}^{\infty} \frac{\Delta(y, \lambda) F(y)}{x + y}\, dy, \tag{2}$$

$\Delta(x, \lambda)$ being the discontinuity of the S matrix corresponding to $\lambda V(r)$. The existence of the solutions of (2) corresponding to (1) can be shown in two ways:

(i) The Fredholm determinant of (2) is

$$\mathfrak{D}(\lambda) = \exp - \int_{m}^{\infty} \lambda \frac{C(\alpha)}{\alpha^2}\, d\alpha.$$

For this result the inversion formalism was useful. For the Jost solutions,

$$f(k, r) = e^{-ikr} + \int_{r}^{\infty} K(r, t) e^{-ikt}\, dt,$$

we know that

$$K(r, r) = \int_{r}^{\infty} \frac{\lambda V(r)}{2}\, dr.$$

Then $F(x, r)$ satisfies off-the-mass-shell integral equations corresponding to (2) and the Fredholm determinant $\mathfrak{D}(\lambda, r)$ satisfies $(\partial \mathfrak{D}/\partial r)/\mathfrak{D} = K(r, r)$. Finally, $\mathfrak{D}(\lambda, 0) = \mathfrak{D}(\lambda)$.

(ii) On the one hand, if we perform a subtraction in (2),

$$F(x) = F(x_0) + \int_{\frac{1}{2}\mu}^{\infty} K_{x_0}(x, y) F(y)\, dy,$$

where

$$K_{x_0} = [\Delta/(x + y)][(x_0 - x)/(x_0 + y)].$$

If $F(x_0) = 0$, then $F(x)$ is an eigenfunction of K_{x_0}. If this happens when $x_0 \to \infty$, then

$$F(x) \xrightarrow[|x| \to \infty]{} 0;$$

and since

$$\lim_{|x_0| \to \infty} K_{x_0} = \Delta/(x + y),$$

we see that the eigenvalues of (1) correspond to the Jost function vanishing at infinity. On the other hand, we know for family (1) that

$$f(k) \xrightarrow[|k| \to \infty]{} 1$$

(in the k complex plane outside the cut) and so $\mathfrak{D}(\lambda)$ cannot vanish for (1). Note that the converse is not true in general: If

$$f(k) \xrightarrow[|k| \to \infty]{} 1,$$

the Fredholm determinant of (2) is different from zero but the reconstructed potentials from (2) are not always of type (1). It is necessary to analyze the states corresponding to the roots of the solutions of (2).

III. ON-THE-MASS-SHELL N/D-TYPE EQUATIONS FOR RELATIVISTIC POTENTIAL THEORY

We consider a slight generalization of the Klein–Gordon equation for S waves,

$$f'' + k^2 f = [\lambda_1 E V_1 + \lambda_2 V_2] f, \tag{3}$$

with real potentials $V_i(r)$:

$$V_i(r) = \int_{\mu_i}^{\infty} e^{-\alpha r} C_i(\alpha)\, d\alpha, \quad \int_{\mu_1}^{\infty} \frac{|C_1(\alpha)|}{\alpha}\, d\alpha < \infty,$$

$$\int_{\mu_2}^{\infty} \frac{|C_2(\alpha)|}{\alpha^2}\, d\alpha < \infty, \quad |\lambda_1 C_1(\alpha)| < \lambda_1 \alpha^{-\eta_1}, \tag{4}$$

$$|\lambda_2 C_2(\alpha)| < |\lambda_2|\, \alpha^{1-\eta_2}, \quad \eta_i > 0.$$

For the true Klein–Gordon case

$$\lambda_2 V_2 = -\tfrac{1}{4}(\lambda_1 V_1)^2. \tag{5}$$

Roughly speaking, when $r \to 0$, $V_1(r)$ must be less singular than const/r and $V_2(r)$ less than const/r^2. Furthermore, we restrict $V_i(r)$ to be a superposition of exponential-type potentials in order to have the usual location of the dynamical cuts in the k plane. The energy $E = (k^2 + m^2)^{\frac{1}{2}}$ is a two-valued function of k. The positive branch $E = +(k^2 + m^2)^{\frac{1}{2}}$ for k real corresponds to the particle scattering, and the negative one $[E = -(k^2 + m^2)^{\frac{1}{2}}$ for k real] to the antiparticle scattering. In a similar way we can consider $V_1 \to -V_1$ in (3). The Jost solutions

$$f_{\epsilon}(k, r) \underset{r \to \infty}{\simeq} e^{-ikr} \quad (\epsilon = \pm 1, E = \pm\sqrt{\ })$$

[6] L. I. Schiff, H. Snyder, and J. Weunberg, Phys. Rev. **15**, 315 (1940); see also Ref. 3.

satisfy

$$f_\epsilon(k, r) = e^{-ikr} + \int_r^\infty \frac{\sin k(s-r)}{k} \times (\lambda_1 \epsilon E V_1 + \lambda_2 V_2) f_\epsilon(k, s)\, ds. \quad (6a)$$

The analytical properties of $f_\epsilon(k, r)$ in the k complex plane are easily studied with the help of the two current methods in potential scattering. From the radial integral equation (6a) we follow Regge[7] and investigate first in a strip of the k complex plane; then we can rotate the strip using the analytical properties in $\mathrm{Re}\, r >$ const of the considered potentials (4). On the other hand, following Martin,[8] we take great advantage of the Laplace transform properties of (6a). We put

$$e^{ikr} f_\epsilon(k, r) = 1 + \int_\mu^\infty \rho_{k,\epsilon}(\alpha) e^{-\alpha r} \quad (\mu = \inf(\mu_1, \mu_2)),$$

and the problem is reduced to the study of the integral equation for $\rho_{k,\epsilon}(\alpha)$:

$$\rho_{k,\epsilon}(\alpha)\alpha(\alpha + 2ik) = \mathcal{C}_\epsilon(\alpha, k, \lambda_1, \lambda_2) + \int_\mu^\alpha \mathcal{C}_\epsilon(\alpha - \beta, k, \lambda_1, \lambda_2)\rho_{k,\epsilon}(\beta)\, d\beta,$$

$$\mathcal{C}_\epsilon = \epsilon E \lambda_1 C_1(\alpha)\theta(\alpha - \mu_1) + \lambda_2 C_2(\alpha)\theta(\alpha - \mu_2). \quad (6b)$$

From (6b) the singularities of $f_\epsilon(k, r)e^{ikr}$ in the k complex plane are apparent: the two branch points $k = \pm im$ coming from E and the dynamical cut $k = i[\mu/2, \infty]$ coming from $V_i(r)$. In Appendix A Corinaldesi's result[3] in $\mathrm{Im}\, k \leq 0$ is extended to the whole complex k plane outside the cuts

$$f_\epsilon(k, r) \underset{|k| \to \infty}{\simeq} \exp\left[-ikr + i\epsilon \frac{E}{k} \lambda_1 \int_r^\infty \frac{V_1(r')}{2}\, dr'\right]. \quad (7)$$

This result is very different from the nonrelativistic one because the asymptotic $|k|$ behavior depends always explicitly on the interaction. For instance, the phase shift

$$\delta_\epsilon(k) \xrightarrow[k \to +\infty]{} \epsilon \int_0^\infty \lambda_1 \frac{V_1(r)}{2}\, dr \quad (8)$$

is not, in general, an integral multiple of π. The usual Levinson theorem is modified and consequently we have a supplementary branch point at $k = -im$ in the lower half-plane. We define the Jost functions and the S matrix:

$$f_\epsilon(k, E) = f_\epsilon(k, 0),$$
$$S_\epsilon(k, E) = \frac{f_\epsilon(k, E)}{f_\epsilon(-k, E)}. \quad (9)$$

Similarly, $f_\epsilon(k, 0)$ has the dynamical cuts $i[\mu/2, \infty]$ and the supplementary branch points $\pm im$ for which we introduce the cuts $\pm i[m, \infty]$. In the full relativistic case the N/D-type equations for the Jost functions were first written by Omnès[9] if there exist cuts in only one half-k plane. Ciulli *et al.*[10] have considered the marginally singular left-hand cut case [$\delta(\infty)$ not an integer multiple of π] where cuts are present in both half-planes. These authors[10] try to write their Jost functions as a product of two functions—one a simple factor having the two cuts and the other only one cut.

Here we remark first that the $\pm im$ branch points are not connected with the use of too-singular interactions [marginally singular potentials like a pure Yukawa resulting from a vector meson interaction lead to $\delta(\infty) = \infty$]. It is a general feature of the Klein–Gordon or Dirac formalism and so is due to the model. Secondly, if we write (like Ciulli *et al.*)

$$f_\epsilon(k, 0) = \left(\exp i\epsilon \frac{E}{k} \delta(\infty)\right) g_\epsilon(k),$$

it is easy to verify, from (6) for $r = 0$, that g_ϵ has still the two branch points $\pm im$. No simple factorization of the Jost functions leads to the cuts in only one half-plane. Therefore we must use another method.

We can similarly (as Oehme did with the Coulomb potential[4]) define crossing relations for the family (4). From (3) $V_1 \to -V_1$ we get

$$S_{-1}(k, E) = S_{+1}(k, -E). \quad (10)$$

In a simple way for the Klein–Gordon case (10) exhibits the link between S_ϵ for $\epsilon = \pm 1$. However, in the following, we keep always the index ϵ in order to emphasize that we investigate both particle and antiparticle scattering. From (6b) and the choice of the cuts for the branch points $(\rho_{-k^*}(\alpha)^*) = \rho_k(\alpha)$,

$$[f_\epsilon(-k^*, 0)]^* = f_\epsilon(k). \quad (11)$$

From Eqs. (10)–(11) for E and k real,

$$S_{-1}(k, E) = (S_{+1}(-k, -E))^*. \quad (12)$$

Our method is the following: From the two S_ϵ and the two $f_\epsilon(k, 0)$ we define two $S^\pm$ and two f_i even in E, and so, without the branch points $k = \pm im$, we have

$$2S^+(k, E^2) = S_{+1}(k, E) + S_{-1}(k, E),$$
$$f_\epsilon(k, E) = f_1(k, E^2) + \epsilon E f_2(k, E^2), \quad (13)$$
$$2S^-(k, E) = (S_{+1}(k, E) - S_{-1}(k, E))/E.$$

[7] T. Regge, M. Bottino, and G. Langoni, Nuovo Cimento **23**, 954 (1962).

[8] A. Martin, *Progress in Elementary Particles and Cosmic Ray Physics* (North-Holland Publ. Co., Amsterdam, 1969), Vol. VIII.

[9] T. Omnès, Nuovo Cimento **21**, 524 (1961).

[10] S. Ciulli, G. Ghika, M. Stihi, and M. Visnescu, Phys. Rev. **154**, 1344 (1967).

Because $f_i(k, E^2)$ has only the dynamical cut in the upper half-k plane, our problem is thus reduced to finding coupled integral equations for $f_i(k)$ with kernels related to the discontinuities of $S^{\pm}$ (or S_ϵ). In order to study f_i we put

$$f_1(k, E^2) = 1 + \int_{\bar{\mu}_1}^{\infty} \tau_{1,k}(\alpha)\, d\alpha,$$
$$\rho_{\epsilon,k}(\alpha) = \tau_{1,k}(\alpha) + \epsilon E \tau_{2,k}(\alpha), \qquad (14)$$
$$f_2(k, E^2) = \int_{\bar{\mu}_2}^{\infty} \tau_{2,k}(\alpha)\, d\alpha,$$

where τ_i satisfies the coupled integral equation

$$\frac{1}{\alpha(\alpha + 2iK)} \begin{pmatrix} \tau_{1,k}(\alpha) \\ \tau_{2,k}(\alpha) \end{pmatrix} = \begin{pmatrix} \lambda_2 C_2(\alpha)\theta(\alpha - \mu_2) \\ \lambda_1 C_1(\alpha)\theta(\alpha - \mu_1) \end{pmatrix} + \begin{pmatrix} \int_{\bar{\mu}_1}^{\alpha-\mu_2} \lambda_2 C_2(\alpha - \beta), E^2 \int_{\mu_1}^{\alpha-\mu_1} \lambda_1 C_1(\alpha - \beta) \\ \int_{\bar{\mu}_1}^{\alpha-\mu_1} \lambda_1 C_1(\alpha - \beta), \int_{\mu_1}^{\alpha-\mu_2} \lambda_2 C_2(\alpha - \beta) \end{pmatrix} \times \begin{pmatrix} \tau_{1,k}(\beta) \\ \tau_{2,k}(\beta) \end{pmatrix} d\beta. \qquad (15)$$

From (15) we see that f_1 is analytic in the cut k-plane with a dynamical cut coming from $V_i(r)$ along $i[(\bar{\mu}_1)/2, \infty]$ with $\bar{\mu}_1 = \inf(\mu_2, 2\mu_1)$. Similarly, f_2 is analytic with a cut $i[(\bar{\mu}_2)/2, \infty]$, where $\bar{\mu}_2 = \mu_1$. Further, $f_i(-k^*, E^{2*})^* = f_i(k)$. On the other hand, outside the cuts,

$$\left| f_1(k, E) - \cos \int_0^{\infty} \frac{\lambda_1 V_1}{2}\, dr \right| \xrightarrow[|k| \to \infty]{} 0$$

and

$$|f_2(k, E)| \xrightarrow[|k| \to \infty]{} 0$$

sufficiently quickly so that the following spectral representations hold:

$$f_1(k, E^2) = \cos \int_0^{\infty} \frac{\lambda_1 V_1}{2}\, dr - i \int_{\bar{\mu}_1}^{\infty} \frac{R_1(y)}{k - iy}\, dy,$$
$$f_2(k, E^2) = -i \int_{\bar{\mu}_2}^{\infty} \frac{R_2(y)}{k - iy}\, dy. \qquad (16)$$

It remains now to find the link between the weight function R_i and the discontinuities of $S^{\pm}$ in the upper half-k plane. We define

$$\Delta_\epsilon(x) = \lim_{\epsilon' \to 0} S_\epsilon(ix - \epsilon') - S_\epsilon(ix + \epsilon')/2i\pi,$$
$$\Delta^{\pm}(x) = \lim_{\epsilon' \to 0} S^{\pm}(ix - \epsilon', -x^2 + m^2) - S^{\pm}(ix + \epsilon', -x^2 + m^2)/2i\pi. \qquad (17)$$

If in (17) we substitute the f_i given by (16), we get

$$\begin{pmatrix} R_1(x) \\ R_2(x) \end{pmatrix} = \begin{pmatrix} \Delta^+(x), & (m^2 - x^2)\Delta^-(x) \\ \Delta^-(x), & \Delta^+(x) \end{pmatrix} \times \begin{pmatrix} f_1(-ix, -x^2 + m^2) \\ f_2(-ix, -x^2 + m^2) \end{pmatrix}. \qquad (18)$$

We now write

$$F_i(x) = f_i(-ix, -x^2 + m^2), \qquad (19)$$

and from Eqs. (16)–(18) we get, finally,

$$F_i(x) = F_i(\infty) + \sum_j \int_0^{\infty} (\Delta_{i,j}(x, y)) F_j(y)\, dy, \qquad i, j = 1, 2,$$
$$(x + y)\Delta_{i,j} = \begin{pmatrix} \Delta^+(y)\theta(y - \mu_+), & (m^2 - y^2)\Delta^-(y)\theta(y - \mu_-) \\ \Delta^-(y)\theta(y - \mu_-), & \Delta^+(y)\theta(y - \mu_+) \end{pmatrix}, \qquad (20a)$$
$$\mu_- = \tfrac{1}{2}(\mu_1), \quad \mu_+ = \inf(\tfrac{1}{2}(\mu_2), \mu_1),$$
$$F_1(\infty) = \cos \int_0^{\infty} \frac{\lambda_1 V_1}{2}\, dr, \quad F_2(\infty) = 0.$$

It is the aim of this paper to study the existence of the solutions of these coupled integral equations which for the Klein–Gordon formalism is equivalent to the resulting integral equation for D in an N/D formalism. Note that from a solution $F_i(x)$, $i = 1, 2$, according to Eq. (13), we get both $F_\epsilon(k, 0)$; thus we can reconstruct both $S_\epsilon(k, E)$ for the particle and antiparticle scattering.

First we consider (20a) as really a linear equation and study the Fredholm formulation. For this we substitute the kernel $\nu(x + y)\Delta_{i,j}$, where ν is the linear parameter. It is shown in Appendix B that the Fredholm determinant has the following bound[11]:

$$|D(\nu)| < \exp\left[|\nu| \int_{\mu^+}^{\infty} \frac{|\Delta^+|}{x}\, dx + |\nu|^2 \int_{\mu^-}^{\infty} \int_{\mu^-}^{\infty} \frac{|\Delta^-(y)(m^2 - y^2)\Delta^-(x)|}{(x + y)^2}\, dx\, dy \right].$$

Let us assume

$$|\Delta^+| < \frac{\text{const}}{x^{\eta}}, \quad |\Delta^-| < \frac{\text{const}}{x^{1+\eta}}, \quad \eta > 0. \qquad (20b)$$

From the results of Appendix B, $D(\nu)$ is an entire function of ν and bounded. Furthermore,[11] $D(\nu)$ has

[11] These results are the extension of those obtained by De Alfaro and Regge in nonrelativistic potential theory [Nuovo Cimento **20**, 956 (1961)].

no roots for

$$|\nu|\int_{\mu^+}^{\infty}\frac{|\Delta^+|}{x} + |\nu|^2\int_{\mu^-}^{\infty}\int_{\mu^-}^{\infty}\frac{|\Delta^-(y)(m^2-y^2)\Delta^-(x)|}{(x+y)^2}\,dx\,dy < \log 2. \tag{20c}$$

This result being interesting for the existence of a perturbative solution, a similar analysis of the Fredholm numerators can be made, but it requires tedious calculations. So we want to transform the equation in such a way that the free terms, as well as the four elements of the kernels, are square integrable without modifying the Fredholm determinant. We put $F_1 - F_1(\infty) = g_1(x)x^{\alpha_1}$ and $F_2 = g_2(x)/x^{\alpha_2}$, where α_1, α_2 are positive; we get

$$g_i(x) = \nu F_1(\infty)\tilde{g}_1(x) + \sum_j \int_0^{\infty} \nu\Delta_{i,j}^{(\alpha_1,\alpha_2)}(x, y)g_j(y)\,dy, \tag{20d}$$

with

$$\tilde{g}_1(x) = x^{-\alpha_1}\int_{\mu^+}^{\infty}\frac{\Delta^+(y)}{x+y}\,dy, \quad \tilde{g}_2(x) = x^{\alpha_2}\int_{\mu^-}^{\infty}\frac{\Delta^-(y)}{x+y}\,dy,$$

with

$$(x+y)\Delta_{ij}^{(\alpha_1,\alpha_2)} = \begin{pmatrix} \Delta^+(y)\dfrac{y^{\alpha_1}}{x^{\alpha_1}}\theta(y-\mu_+) & (m^2-y^2)\Delta^-(y)x^{-\alpha_1}y^{-\alpha_2}\theta(y-\mu_-) \\ \Delta^-(y)y^{\alpha_1}x^{\alpha_2}\theta(y-\mu_-) & \Delta^+(y)x^{\alpha_2}y^{-\alpha_2}\theta(y-\mu_+) \end{pmatrix}.$$

Taking into account the bounds (20b) for $\Delta^{\pm}$, and by a suitable choice of α_1, α_2, $\alpha_1 = \alpha_2 = 0$ if $\eta > 1$, $\alpha_1 = \frac{1}{2}$ and $\alpha_2 = \frac{1}{2} - \eta/2$ if $0 < \eta \leq 1$, it follows that

$$\int_0^{\infty}dx\int_0^{\infty}dy\ \mathrm{Tr}\,(\Delta_{ij}^{(\alpha_1,\alpha_2)}(x, y))^T(\Delta_{ij}^{(\alpha_1,\alpha_2)}(x, y)) < \infty,$$

and the free terms of (20d) are also square integrable. Finally, a Fredholm formulation of (20d) exists and we remark that both the Fredholm determinants of (20a) and (20d) are the same. In fact, it is shown in Appendix B that if $\mathfrak{D}(\nu) \neq 0$, the solution is unique, not only for (20d), but also for our original Eq. (20a).

Secondly, we want to study the Fredholm solution of (20a) or (20d) for $\nu = 1$ and $\Delta^{\pm}$, corresponding to the potentials $\lambda_i C_i(\alpha)$ considered in (4). Now $\Delta^{\pm}$ depends on the linear parameters of the potentials λ_i, but $\Delta^{\pm}$ as well as (20a) are, of course, not linear in these parameters. We must find the asymptotic behavior of $\Delta^{\pm}(x)$ and see if the bounds (20b) hold. For this we shall establish the link between the discontinuities $\Delta^{\pm}$ (or $\Delta_{\pm}$) and the potentials $\lambda_i C_i(\alpha)$, as was done by Martin[8] in the nonrelativistic case.

We want to evaluate explicitly the discontinuities across the cuts. From the solutions of (15) let us define for $k = ix \pm \epsilon'/2$ two coupled spectral functions $\tau_{j,ix}^{\pm}(\alpha)$, where $j = 1, 2$ and $\epsilon' > 0$ is arbitrarily small. From (15) we get the following:

(i) for $\alpha < 2x - \epsilon'$:

$$\tau_{j,ix}^{+}(\alpha) = \tau_{j,ix}^{-}(\alpha), \quad j = 1, 2;$$

(ii) for $2x - \epsilon' < \alpha < 2x + \epsilon'$:

$$\begin{pmatrix} \tau_{1,ix}^{+}(\alpha) - \tau_{1,ix}^{-}(\alpha) \\ \tau_{2,ix}^{+}(\alpha) - \tau_{2,ix}^{+}(\alpha) \end{pmatrix} = -2i\pi\delta(\alpha - 2x)\left[\begin{pmatrix} \lambda_2 C_2(\alpha) \\ \lambda_1 C_1(\alpha) \end{pmatrix} + \begin{pmatrix} \displaystyle\int_{\tilde{\mu}_1}^{\alpha-\mu_2}\lambda_2 C_2(\alpha-\beta) & (m^2-x^2)\displaystyle\int_{\mu_1}^{\alpha-\mu_1}\lambda_1 C_1(\alpha-\beta) \\ \displaystyle\int_{\tilde{\mu}_1}^{\alpha-\mu_1}\lambda_1 C_1(\alpha-\beta) & \displaystyle\int_{\mu_1}^{\alpha-\mu_2}\lambda_2 C_2(\alpha-\beta) \end{pmatrix}\begin{pmatrix} \tau_{1,ix}(\beta) \\ \tau_{2,ix}(\beta) \end{pmatrix} d\beta\right]. \tag{21}$$

Because of (i), the right-hand side is well defined.

(iii) for $2x + \epsilon' < \alpha < 2x + \inf(\mu_1, \mu_2) - \epsilon'$:

$$\tau_{j,ix}^{+}(\alpha) = \tau_{j,ix}^{-}(\alpha).$$

These $\tau_{j,ix}^{\pm}(\alpha)$ will be used for the discontinuities of the $f_\epsilon(k, r)$ along the $\mathrm{Im}\,k > 0$ axis. As for the Jost functions, let us separate in the Jost solutions the even and odd E parts of the solutions of (6a):

$$\begin{aligned} f_\epsilon(k, r) &= f_1(k, E^2, r) + \epsilon E f_2(k, E^2, r), \\ f_i(k, E^2, 0) &= f_i(k, E^2), \quad i = 1, 2, \\ f_1(k, E^2, r) &= e^{-ikr}\left[1 + \int_{\tilde{\mu}_1}^{\infty}\tau_{1,k}(\alpha)e^{-\alpha r}\,d\alpha\right], \\ f_2(k, E^2, r) &= e^{-ikr}\int_{\tilde{\mu}_2}^{\infty}\tau_{2,k}(\alpha)e^{-\alpha r}\,d\alpha. \end{aligned} \tag{22}$$

The three functions [(ϵ fixed), $\lim_{\epsilon'\to 0} f_\epsilon(ix \pm \epsilon', r)$, and $f_\epsilon(-ix, r)$], being solutions of the same Schrödinger equation, are not independent. Due to their asymptotic behavior when $r \to \infty$ ($|x|$ finite) and $r \to 0$, it turns out that they satisfy

$$\lim_{\epsilon'\to 0} [f_\epsilon(ix + \epsilon', r) - f_\epsilon(ix - \epsilon', r)] = -2i\pi\Delta_\epsilon f_\epsilon(-ix, r). \quad (23a)$$

From this it follows that

$$-2i\pi\Delta_\epsilon(x) = \lim_{r\to\infty}\left\{e^{xr}\lim_{\epsilon'\to 0}[f_\epsilon(ix + \epsilon', r) - f_\epsilon(ix - \epsilon', r)]\right\}. \quad (23b)$$

Now from the relations $\Delta^+ = \frac{1}{2}(\Delta_+ + \Delta_-)$, $\Delta^- = (\Delta_+ - \Delta_-)/2(m^2 - x^2)^{\frac{1}{2}}$, and the above results obtained for the spectral functions $\tau^\pm(\alpha), j = 1, 2$, finally, from (23), we get

$$2x\begin{pmatrix}\Delta^+(x)\\ \Delta^-(x)\end{pmatrix} = \begin{pmatrix}\lambda_2 C_2(2x)\\ \lambda_1 C_1(2x)\end{pmatrix} + \begin{pmatrix}\int_{\bar\mu_1}^{2x-\mu_2}\lambda_2 C_2(2x-\alpha) & (m^2-x^2)\int_{\mu_1}^{2x-\mu_1}\lambda_1 C_1(2x-\alpha)\\ \int_{\bar\mu_1}^{2x-\mu_1}\lambda_1 C_1(2x-\alpha) & \int_{\mu_1}^{2x-\mu_2}\lambda_2 C_2(2x-\alpha)\end{pmatrix}\begin{pmatrix}\tau_{1,ix}(\alpha)\\ \tau_{2,ix}(\alpha)\end{pmatrix} d\alpha. \quad (24)$$

The two relations (15) and (24) give the possibility of constructing the discontinuities $\Delta^\pm(x)$ [or $\Delta_\epsilon(x)$] from the knowledge of the potentials $\lambda_i C_i(\alpha)$ and inversely. Indeed, as was done by Martin[8] in the nonrelativistic case, it is easy to see that from (15) and (24) we can also reconstruct, step by step, the potentials $C_i(\alpha)$ from the discontinuities $\Delta^\pm(x)$ [or $\Delta_\epsilon(x)$].

In principle, from (15) and (24) it is also possible to determine the Fredholm determinant of (20a) in terms of the $C_i(\alpha)$. For instance, for the pure Klein–Gordon case

$$C_2(\alpha) = \int_{\mu_1}^{\alpha} C_1(\alpha - \beta)C_1(\beta)\, d\beta,$$

up to the fourth order in λ_1 we get

$$\mathfrak{D}(\lambda_1) = 1 - \frac{\lambda_1^2}{2}\left[\int_{\mu_1}^{\infty}\frac{C_1(\alpha)}{2\alpha}\right]^2 + \lambda_1^4\cdots. \quad (25)$$

Unfortunately, the method is very cumbersome and, as with nonrelativistic potentials,[1] it appears that the combined use of (15) and (24) is not very practical to determine explicitly the Fredholm determinant of (20a).

In Appendix A the behavior for large α of $\tau_{j,ix}(\alpha)$, $\alpha < 2x$, is investigated; if we substitute it in (24), for $\Delta^\pm(x)$ and x large we get

$$|\Delta^+(x)| < \frac{C_+}{x^\eta}\theta(x - \mu_+),$$

$$|\Delta^-(x)| < \frac{C_-}{x^{1+\eta}}\theta(x - \mu_-), \quad (26)$$

where $\eta = \inf(\eta_1, \eta_2)$ and $C_\pm$ are positive λ_i-dependent constants but finite for $|\lambda_i|$ finite. Moreover, $C_\pm \to 0$ when both $|\lambda_i| \to 0$.

Furthermore, $\Delta^\pm$ are entire functions of λ_i and satisfy the bounds (20a). So, from the results of Appendix B, the Fredholm determinant considered as a function of λ_i, $\mathfrak{D}(\lambda_1, \lambda_2)$ is also an entire function of λ_i. Moreover, if both $|\lambda_i|$ are sufficiently small [see (26)], we can always satisfy the bound (20c) (so that $\mathfrak{D} \neq 0$), and the Neumann series of (20d) exists. But we are interested in the existence of the solutions of (20a) and (20d), not only for small $|\lambda_i|$, but for any finite $|\lambda_i|$ and so we must look at the roots of $\mathfrak{D}(\lambda_1, \lambda_2)$.

In order to understand the meaning of the eigenvalues corresponding to the kernel Δ_{ij}, we make a subtraction in (20):

$$F_i(x) = F_i(x_0) + \sum_j\int\left(\frac{x_0 - x}{x_0 + y}\right)\Delta_{ij}(x, y)F_j(y)\, dy.$$

Let us assume that $F_i(x_0) = 0$ ($i = 1$ and 2). Then, for the subtracted kernel, $F_i(x)$ is an eigenfunction and the corresponding Fredholm determinant vanishes. Let us assume now that this happens when $x_0 \to \infty$ [$F_i(x_0 = \infty) = 0, i = 1, 2$]; then the Fredholm determinant corresponding to the kernel

$$\lim_{x_0\to\infty}\frac{(x_0 - x)}{x_0 + y}\frac{\Delta_{ij}(x, y)}{x + y} = \frac{\Delta_{ij}(x, y)}{x + y} \quad (27)$$

vanishes. In fact, from (20) it was clear that the roots of

$$\cos\lambda_1\int_0^\infty\frac{V_1}{2}\, dr$$

for which $F_1(\infty) = 0$ correspond to the homogeneous equations and so are roots of $\mathfrak{D}$. Inversely, if λ_i is not such a root, we expect that the Fredholm solution (24) of (20a) must exist. Finally, the Fredholm determinant $\mathfrak{D}$ of (20a), when the discontinuities $\Delta^\pm(x)$ are expressed by (15) and (24) in terms of λ_1 and λ_2, is a very complicated function of (λ_1, λ_2),

but we can expect that it has the form

$$\mathfrak{D}(\lambda_1, \lambda_2) = \cos \lambda_1 \int_0^\infty \frac{V_1}{2} dr \tilde{\mathfrak{D}}(\lambda_1, \lambda_2), \qquad (28)$$

where from the bounds (26), $\tilde{\mathfrak{D}}(\lambda_1, \lambda_2)$ is entire in (λ_1, λ_2), well defined for

$$\lambda_1 \int_0^\infty \frac{V}{2} dr \neq (2m + 1)\pi/2,$$

and cannot vanish for other λ_i values. We cannot say more because we do not know anything about the order of multiplicity of the roots. Note that this kind of connection between the eigenvalues or the roots of the Fredholm determinant, on the one hand, and the asymptotic behavior, on the other hand, can be applied very generally in dispersion relation. By using the off-the-mass-shell N/D-type equations, we shall explicitly investigate how this relation (28) arises.

IV. OFF-THE-MASS-SHELL FORMALISM

We recall that in nonrelativistic potential scattering[1,12] for $l = 0$ as well as for $l \neq 0$, the off-the-mass-shell N/D-type equations were very useful in understanding many problems connected with the on-the-mass-shell equations (existence and uniqueness of the solutions, ghosts, threshold behavior, subtracted equations, self-damping, Regge poles or Regge cuts, etc.).

The main reason is the following: The off-the-mass-shell equations are the solutions for the Jost solutions with kernels still proportional to the S-matrix discontinuities, whereas the on-the-mass-shell equations are limits of the previous ones when $r \to 0$ (Jost functions). Furthermore, the Laplace transform of these off-the-mass-shell equations are indeed the Marchenko[13] equations arising in the inversion problem for $l = 0$, giving the possibility of reconstructing the potential. In other words, from the spectrum (eigenvalues) of the off-the-mass-shell equations (connected with the reconstruction of the potential) we get information when $r \to 0$ about the spectrum (eigenvalues) of the resulting integral N/D equations (which are connected with the reconstruction of the partial waves). Also with these off-the-mass-shell equations we have the straightforward connection between the two formalisms: perturbation expansion and dispersion relation. Note that the connection between the inverse Laplace transform of the potentials and discontinuities made in the previous section was a first step in off-the-mass-shell formalism, but not sufficient—as we have said.

Unfortunately, the extension of the inversion problem "à la Marchenko" has not been made (at least to the knowledge of the author) for Klein–Gordon or Dirac formalism. Nevertheless, it was realized[1] that for exponential-type potentials, pure dispersive methods can give the same information as the Marchenko equation. So, without treating the complete relativistic inversion problem (this is not the object of the present paper), we shall establish only the sufficient part for our connection: namely, the link between the potentials and the eigenvalues of the on-the-mass-shell equations (20a) or (20d).

A. Integral Equations for the Jost Solutions

We define the corresponding off-the-mass-shell functions of the preceding section,

$$S_\epsilon(k, E, r) = f_\epsilon(k, E, r)/f_\epsilon(-k, E, r),$$

as well as the even associated energy functions

$$2S^\pm(k, E^2, r) = S_{+1}(k, E, r) \pm S_{-1}(k, E, r)/(1 \pm 1 + E(1 \mp 1)). \qquad (29)$$

The $f_i(k, E^2, r)$, even E functions, deduced from $f_\epsilon(k, E, r)$, satisfy

$$\begin{pmatrix} f_1(k, E^2, r) \\ f_2(k, E^2, r) \end{pmatrix} = \begin{pmatrix} e^{-ikr} \\ 0 \end{pmatrix} + \int_r^\infty ds \frac{\sin k(s - r)}{k} \times \begin{pmatrix} \lambda_2 V_2(s), \lambda_1 E^2 V_1(s) \\ \lambda_1 V_1(s), \lambda_2 V_2(s) \end{pmatrix} \begin{pmatrix} f_1(k, E^2, s) \\ f_2(k, E^2, s) \end{pmatrix}. \qquad (30)$$

From (30) or from the Laplace transform equation (15), the analytical properties of $f_i(k, E^2, r)$ in the k complex plane can be studied directly, but these properties follow also from the ones of $f_\epsilon(k, r)$ and the asymptotic behavior (7) (see Appendix A). The following spectral representations hold:

$$f_i(k, E^2, r)e^{ikr} = f_i(\infty, r) - i \int_{\bar{\mu}_i}^\infty \frac{R_i(y, r)}{k - iy} dy, \qquad (31)$$

where

$$f_1(\infty, r) = \cos \int_r^\infty \frac{V_1 \, dr}{2} \quad \text{and} \quad f_2(\infty, r) = 0.$$

If we define the discontinuities

$$\Delta^\pm(x, r) = \frac{1}{2i\pi} [S^\pm(ix - \epsilon', m^2 - x^2, r) - S^\pm(ix + \epsilon', m^2 - x^2, r)]$$

and similarly, $\Delta_\epsilon(x, r)$ for $S_\epsilon(k, E, r)$, it turns out that, owing to (23b), $\Delta^\pm(x, r) = \Delta^\pm(x)$ and $\Delta_\epsilon(x, r) = \Delta_\epsilon(x)$. If we substitute in $\Delta^\pm(x, r)$ the $f_i(k, E^2, r)$

[12] H. Cornille and G. Rubinstein, J. Math. Phys. **9**, 1501 (1968); Nuovo Cimento **56**, 867 (1968).

[13] Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, Science Publishers, Inc., New York, 1963).

given by (31), we get

$$\begin{pmatrix} e^{xr}R_1(x,r) \\ e^{xr}R_2(x,r) \end{pmatrix} = \begin{pmatrix} \Delta^+(x), (m^2-x^2)\Delta^-(x) \\ \Delta^-(x), \Delta^+(x) \end{pmatrix} \times \begin{pmatrix} f_1(-ix, m^2-x^2, r) \\ f_2(-ix, m^2-x^2, r) \end{pmatrix}.$$

We put $f_i(-ix, m^2-x^2, r) = F_i(x, r)$ and finally obtain

$$F_i(x,r) = F_i(\infty, r)e^{-xr} + \sum_{j=1}^{j=2} \int_0^\infty \Delta_{ij}(x,y)e^{-(x+y)r}F_j(y,r)\,dy, \quad (32)$$

where Δ_{ij} has been defined in (20a),

$$F_1(\infty, r) = \cos \int_r^\infty \frac{\lambda_1 V_1\,dx}{2},$$

and $F_2(\infty, r) = 0$. First, for $r = 0$ the off-the-mass-shell equations reduce to the on-the-mass-shell (20a). Secondly, for $r > 0$, because of the exponential damping factor $e^{-(x+y)r}$ in (32), all the four terms of the kernel, as well as the free terms, are now square integrable. So if we linearize (32) and consider a kernel $\nu\Delta_{ij}(x,y)e^{-(x+y)r}$, the Fredholm formulation of the solution exists. If for $r \neq 0$ we apply the same method as is done for $r = 0$ in Appendix B, then, for the Fredholm determinant $\mathfrak{D}(\nu, r)$, we get

$$|\mathfrak{D}(\nu,r)| < \exp\left[|\nu| \int_{\mu^+}^\infty \frac{|\Delta^+(x)|\,e^{-2xr}}{x} + |\nu|^2 \int_{\mu^-}^\infty \int_{\mu^-}^\infty \frac{|\Delta^-(x)\Delta^-(y)(m^2-y^2)|\,e^{-2(x+y)r}}{(x+y)^2}\,dx\,dy\right].$$

It follows that if (20c) is satisfied, then $\mathfrak{D}(\nu, r)$ has no roots for $r \geq 0$. Now we put $\nu = 1$ and consider that the λ_i are the parameters. From the bounds for $\Delta^\pm$ obtained in the previous section and the results of Appendix B for $\nu \neq 0$, we see that the Fredholm determinants $\mathfrak{D}(\lambda_1, \lambda_2, r)$ are entire functions of λ_1 and λ_2. Furthermore, if $|\lambda_i|$ are sufficiently small so that (20c) is satisfied, then $\mathfrak{D}(\lambda_1, \lambda_2, r) \neq 0$ for $r \geq 0$.

Finally, we note that from the solutions (32) (obtained from the discontinuities) the whole Jost solutions $f_\epsilon(k, r)$ can be reconstructed.

B. Extensions of the Marchenko Equations

Similarly to the nonrelativistic case, we put

$$F_1(x,r) = e^{-xr}\Phi(r) + \int_r^\infty K_1(r,y)e^{-xy}\,dy,$$
$$F_2(x,r) = \int_r^\infty K_2(r,y)e^{-xy}\,dy, \quad (33)$$
$$\Phi(r) = \cos\lambda_1 \int_r^\infty \frac{V_1\,du}{2}.$$

From the integral equation (30) it is easy to see that such a representation is valid. If we substitute in (33) or take the Laplace transform, we get coupled integral equations for $K_i(r, y)$:

$$K_i(r,y) = \bar{K}_i(r,y) + \sum_j \int_r^\infty \mathcal{F}_{i,j}(y+t)K_j(r,t)\,dt, \quad (34)$$

where

$$\mathcal{F}_{1,1}(z) = \mathcal{F}_{2,2}(z) = \int_{\mu^+}^\infty \Delta^+(u)e^{-zu}\,du,$$
$$\mathcal{F}_{1,2}(z) = \int_{\mu^-}^\infty \Delta^-(u)e^{-zu}(m^2-u^2)\,du,$$
$$\mathcal{F}_{2,1}(z) = \int_{\mu^-}^\infty \Delta^-(u)e^{-zu}\,du,$$
$$\bar{K}_i(r,y) = \mathcal{F}_{i,1}(r+y)\cos\int_r^\infty \frac{\lambda_1 V_1(u)}{2}\,du.$$

In order to study (34) it is convenient to put

$$K_{1,1}(r,y) = K_1(r,y)\Big/\cos\int_r^\infty \frac{\lambda_1 V_1}{2}\,du,$$
$$K_{2,1}(r,y) = K_2(r,y)\Big/\cos\int_r^\infty \frac{\lambda_1 V_1}{2}\,du, \quad (35)$$

and to introduce two other components,[14] solutions of (34) defined by new free terms, $K_{i,j}(r, y)$ being now a 2×2 matrix solution of

$$K_{i,j}(r,y) = \mathcal{F}_{i,j}(r+y) + \sum_l \int_r^\infty \mathcal{F}_{i,l}(y+t)K_{l,j}(r,t)\,dt. \quad (36)$$

Equation (36) is the extension to the Klein–Gordon formalism of the dispersive form[1] of the Marchenko equation, which gives the possibility of reconstructing $K_{i,j}(r, y)$ [also the potentials $V_i(r)$; see below] from the discontinuities $\Delta^\pm(x)$. In Appendix C we linearize (36) by introducing a linear parameter $\mathcal{F}_{ij} \to \nu\mathcal{F}_{ij}$, and the extension to (36) of the fundamental property[1] of the corresponding nonrelativistic equation is obtained:

$$K_{11}(r,r) = K_{22}(r,r) = \frac{1}{2}\left(\frac{(\partial/\partial_r)\mathfrak{D}(\nu,r)}{\mathfrak{D}(\nu,r)}\right). \quad (37)$$

$\mathfrak{D}(\nu, r)$ is the Fredholm determinant of both (36) and (32). In order to get (37), putting for instance $\nu = 1$, note that we use only $\mathcal{F}_{11} = \mathcal{F}_{22}$ and do not take into account the other relations among $\mathcal{F}_{ij}$, which come from the fact that the $\mathcal{F}_{ij}$ are all obtained following (24) from $\lambda_1 C_1$ and $\lambda_2 C_2$.

[14] Note that also for the extension of Gel'fand–Levitan formalism to the Klein–Gordon case it is convenient to use a 2×2 matrix formalism (Verde, Ref. 3, and De Alfaro, Ref. 3).

So (37) is an intrinsic property of (36), the corresponding Marchenko equation for the Klein–Gordon case. Note that $\mathfrak{D}(\nu, r)$ is also the Fredholm determinant of our off-the-mass-shell N/D-type equation (with $\Delta^{\pm} \to \nu\Delta^{\pm}$) for the Jost solution (32) and so for $r \to 0$, $\mathfrak{D}(\nu, r)$ reduces to the Fredholm determinant of the on-the-mass-shell equations (20a) and (20d). Finally,

$$\mathfrak{D}(\nu, r) = \exp\left(-2\int_r^\infty K_{11}(t, t)\, dt\right) = \exp\left(-2\int_r^\infty \frac{K_1(t, t)\, dt}{\cos\frac{1}{2}\lambda_1 \int_t^\infty V_1(u)\, du}\right), \quad r \geq 0. \tag{38a}$$

Coming back to the case $\nu = 1$ where λ_i are the parameters, we get

$$\mathfrak{D}(\lambda_1, \lambda_2) = \exp\left(-2\int_0^\infty \frac{K_1(t, t, \lambda_1, \lambda_2)\, dt}{\cos\frac{1}{2}\lambda_1 \int_t^\infty V_1(u)\, du}\right), \tag{38b}$$

where we have written $K_1(t, t)$ as $K_1(t, t, \lambda_1, \lambda_2)$ in order to recall that it is in fact a function of λ_1, λ_2. It remains now to exhibit the relations between $K_i(r, r)$ and the two potentials $\lambda_i V_i(r)$.

C. Meaning of $K_i(r, r)$ and Relations with $\lambda_i V_i(r)$

We recall that in nonrelativistic theory[13] $K(r, r)$ is the primitive of the potential and so is linear in the coupling strength of the potential. On the contrary, for the Klein–Gordon case we shall see that $K_i(r, r)$ are given by series in λ_i, even if (5) is satisfied. At the beginning let us assume only that $\lambda_i V_i$ satisfy sufficient conditions, such as that the integral equations (6a) and (30) for $f_i(k, E^2, r)$ exist. In Appendix D it is shown that if we insert the representation (33) of f_i into the integral equations (30) and eliminate the trigonometric functions, then we get for $K_i(r, t)$ coupled integral equations from the $V_i(r)$. These equations could be the starting point of a study of $K_i(r, t)$. So, as in potential,[13] it is possible to seek sufficient conditions for $V_i(r)$ such as Φ and $K_i(r, t)$ exist. But this is outside the scope of the present paper. We are mainly interested in $K_i(r, r)$ and we find (see Appendix D)

$$\Phi(r) = \cos\int_r^\infty \frac{\lambda_1 V_1(u)}{2}\, du, \tag{39a}$$

$$K_2(r, r) = \sin\int_r^\infty \frac{\lambda_1 V_1(u)}{2}\, du, \tag{39b}$$

$$K_1(r, r) - \frac{\lambda_1 V_1(r)}{4} K_2(r, r) = \tfrac{1}{2}\int_r^\infty \Phi(s)\left[\lambda_2 V_2(s) + \frac{\lambda_1^2 V_1^2}{4}(s)\right] - \tfrac{1}{4}\int_r^\infty \lambda_1\, dr V_1(s) \int_s^\infty [V_1(u) K_1(u, u)\lambda_1 + \lambda_2 V_2(u) K_2(u, u)]\, du. \tag{39c}$$

First, we note that the dominant parts of the transforms of f_i are given [(39a) and (39b)] only by the potentials $\lambda_1 V_1(r)$. This is of course connected to the Corinaldesi's result[3] that the asymptotic $|k|$ behavior is given entirely by $\lambda_1 V_1$. Only for $K_1(r, r)$ must both V_1 and V_2 be taken into account. For the family of potentials (4) and (5) that we consider we must distinguish two cases:

(*1*) *Pure Klein–Gordon case:* $\lambda_2 V_2 = \frac{1}{4}(-\lambda_1^2 V_1^2)$. From (39c) we get

$$K_1(r, r) = -\frac{\lambda_1 V_1}{4} \sin\frac{\lambda_1}{2}\int_r^\infty \frac{V_1(u)}{2}\, du. \tag{40}$$

The off-the-mass-shell Fredholm determinant of both (32) and (36) is

$$\mathfrak{D}(\lambda_1, r) = \cos\int_r^\infty \frac{\lambda_1 V_1(u)}{2}\, du,$$

$$\lambda_1 V_1(r) = \frac{2\partial/\partial r(\mathfrak{D}(r))}{[1 - \mathfrak{D}^2(r)]^{\frac{1}{2}}}, \tag{41}$$

whereas the Fredholm determinant of our on-the-mass-shell (20a) reduced N/D integral equation is

$$\mathfrak{D}(\lambda_1) = \cos\frac{\lambda_1}{2}\int_0^\infty V_1\, du = F_1(\infty). \tag{42}$$

This confirms the results given in the previous section. The eigenvalues of (20a) must correspond to the special λ_1 values such as

$$\frac{\lambda_1}{2}\int_0^\infty V_1(u)\, du = (2m + 1)(\pi/2), \qquad m = 0 \pm 1, \pm 2, \cdots. \tag{43}$$

(*2*) *General case:* $\lambda_i V_i$ satisfying the conditions (4), but without particular relation between $\lambda_1 V_1$ and $\lambda_2 V_2$. In this general case, from (39c) we see that K_1 [and consequently $\mathfrak{D}(\lambda_1, \lambda_2)$] has no special reduction, but in Appendix D the following results are obtained. Let us define

$$K_1(r, r) = -\frac{\lambda_1 V_1(r)}{4}\left(\sin\frac{\lambda_1}{2}\int_r^\infty V_1(u)\, du\right) + \tilde{K}_1(r, r, \lambda_1, \lambda_2).$$

Then

$$\mathfrak{D}(\lambda_1, \lambda_2) = \left(\cos \lambda_1 \int_0^\infty \frac{V_1}{2}\, du\right) \times \exp\left(-\int_0^\infty \frac{\tilde{K}_1(r, r, \lambda_1, \lambda_2)}{\cos \int_r^\infty \frac{1}{2}(\lambda_1 V_1)\, du}\, dr\right). \quad (44)$$

Both $|K(r, r)|$ and $|\tilde{K}|$ can be bounded by functions of the type Ce^{-ar}/r^b, where $a > 0$ and $0 < b < 1$, C being a constant greater than 0, finite for $|\lambda_1|$ and $|\lambda_2|$ finite. Furthermore,

$$\int_0^\infty dr\, |K_1(r, r)|\, dr \quad \text{and} \quad \int_0^\infty dr\, |\tilde{K}(r, r)|$$

are also bounded by finite constants for $|\lambda_1|$ and $|\lambda_2|$ finite. If we look at (38) and (44), we see that the only singularity can come in the integration from the end point (if

$$\cos \lambda_i \int_r^\infty \frac{V_1\, du}{2}$$

vanishes at $r = 0$). Finally, in both cases

$$(\lambda_1, \lambda_2) = \cos \frac{\lambda_1}{2} \int_0^\infty V_1\, du \tilde{\mathfrak{D}}(\lambda_1, \lambda_2)$$

where $\mathfrak{D} \neq 0$ if $\lambda_1 \int_0^\infty V_1\, du \neq (2m + 1)\pi/2$.

Note that if $\lambda_1 V_1 \equiv 0$, then $K_2 \equiv 0$ and

$$K_1(r, r) = \frac{\lambda_2}{2} \int_r^\infty V_2(r')\, dr',$$

leading to

$$\mathfrak{D}(\lambda_1 = 0, \lambda_2) = \exp \int \frac{\lambda_2 C_2(\alpha)}{\alpha^2}\, d\alpha,$$

which is, of course, the result of the nonrelativistic case.[1]

D. Existence and Uniqueness of the Solution of (20) when $\mathfrak{D}(\lambda_1, \lambda_2) \neq 0$

We consider (20d) with $\nu = 1$ and λ_i being the parameters. The Fredholm formulation of the solution can be written

$$g_1(x)x^{\alpha_1} = \frac{N_1(x, \lambda_1, \lambda_2)F_1(\infty)}{\mathfrak{D}(\lambda_1, \lambda_2)},$$

$$\frac{g_2(x)}{x^{\alpha_2}} = \frac{N_2(x, \lambda_1, \lambda_2)F_2(\infty)}{(\lambda_1, \lambda_2)},$$

and the solution is unique [$\mathfrak{D}(\lambda_1, \lambda_2) \neq 0$], as we have seen, if $F_1(\infty) \neq 0$. Consequently, we can write the solution of (20a) as

$$F_1(x) = F_1(\infty) + \frac{N_1(x, \lambda_1, \lambda_2)F_1(\infty)}{\mathfrak{D}(\lambda_1, \lambda_2)},$$

$$F_2(x) = \frac{N_2(x, \lambda_1, \lambda_2)F_1(\infty)}{\mathfrak{D}(\lambda_1, \lambda_2)}, \quad (45)$$

which is still unique if $F_1(\infty) \neq 0$.

In the pure Klein–Gordon case (5), the solution (45) is reduced to a very simple form because, from (42), we know that $\mathfrak{D}(\lambda_1) = F_1(\infty)$ in this case and we get

$$F_1(x) = N_1(\lambda_1, x) + \cos \int_0^\infty \frac{\lambda_1 V_1\, du}{2},$$

$$F_2(x) = N_2(\lambda_1, x), \quad (45')$$

where N_1 and N_2 are still obtained from the Fredholm numerators of (20d).

On the other hand, we know from the analysis of Sec. III and Appendix A that F_1 and F_2 are entire functions of both λ_1 and λ_2. [See (14) and (15), where F_i are defined from $\tau_{i,k}(\alpha)$, and $\tau_{i,k}(\alpha)$ satisfy Volterra equations where λ_i appears as the parameters.] Now we report this result in (45) and (45'). We recall that also both $F_1(\infty)$ and $\mathfrak{D}(\lambda_1, \lambda_2)$ are integer functions of λ_1 and λ_2, and $\mathfrak{D}(\lambda_1, \lambda_2)$ factorizes the roots of $F_1(\infty)$. So, finally, $N_i(x, \lambda_1, \lambda_2)$ and $N_i(\lambda_1, x)$ in (45) and (45') ($i = 1, 2$) are entire functions of λ_1, λ_2.

E. Pure Klein–Gordon Case when $(\lambda_1/2) \int_0^\infty V_1\, du = (2m + 1)\pi/2$

From (45') we get in this case

$$F_1(x) = N_1(\lambda_1, x),$$

$$F_2(x) = N_2(\lambda_2, x),$$

where the right-hand sides coming from the Fredholm numerator of (20d) are well defined, not identically zero for any x, because the left-hand sides [see (14) and (15)] are not identically zero for these special λ_1 values. Finally, the solution is still unique. In the general case (4), because we do not know the order of multiplicity of the roots of $\mathfrak{D}(\lambda_1, \lambda_2) = 0$ for these special λ_1 values, we cannot conclude with only the above results.

V. WEAK COUPLINGS

Up to now we have only discussed the problem of the existence and uniqueness for our N/D-type equations (20a). We are concerned now with a second problem: Do we know if the states corresponding to these equations [roots of $F_1(x) + \epsilon(m^2 - x^2)^{\frac{1}{2}}F_2(x)$] are always physically available? In other words, can we show that complex binding energies or bad states like ghosts or antiresonances do not occur? We still consider this problem in the case where (20a) is not really a linear equation; we put $\nu = 1$ and the parameters are λ_i the coupling strengths of the potentials (4) and (5).

Note that for nonrelativistic potentials the corresponding answer is known because Yukawa family (1) leads only to true bound states on the physical sheet. This comes from the following results[15]: The Jost function for regular interactions have in $\operatorname{Im} k < 0$ only roots on the imaginary axis (no complex k^2 roots). Furthermore, the roots are simple. We recall that these results hold also for the Dirac formalism,[16] for instance, if the potentials satisfy the same conditions as V_1 in (4). For the nonrelativistic case[1,12] the existence of bad states in N/D equations is not due to special λ values of the potential $\lambda V(r)$, but to a bad r behavior—for instance, second-order poles in $r_0 \geq 0$. So for these bad states the corresponding wavefunctions in general have poles at r_0 (at least of the first order and which are not square integrable).

For the Klein–Gordon formalism real binding energies[3] $-m < E < m$ are usually required in order to imply the existence of antiparticles ($E < 0$) or particles ($E > 0$) bound states. But in that case there exist[6] simple examples of regular interactions where complex states occur when the coupling strengths become "too strong." If this happens for the family (4) that we consider, we can note the difference from the above-mentioned ghosts in nonrelativistic potential because now the spatial behavior of the wavefunctions is quite correct. This difficulty is not due to the dispersive approach. It is present in the perturbative approach [(3) or (6)]. So we investigate the properties of the roots of the Jost function directly from (3). We study the roots $f_\epsilon(-k) = 0$ in $\operatorname{Im} k \geq 0$. For $\epsilon\lambda_1 V_1(r)$ satisfying (4), we define

$$\varphi_\epsilon(k, r) \underset{r\to 0}{\simeq} r,$$

solutions of (3), and we get, as usual,

$$\varphi_\epsilon(k, r) = (1/2ik)[f_\epsilon(-k)f_\epsilon(k, r) - f_\epsilon(k)f_\epsilon(-k, r)]. \tag{46}$$

As in nonrelativistic potential, from (46), (11), and the boundary condition of $\varphi_\epsilon(k, r)$, we see that $f_\epsilon(k)$ cannot have roots for k real $\neq 0$ [$f_\epsilon(k) = 0$ implies $f_\epsilon(-k) = 0$ and $\varphi_\epsilon(k, r) \equiv 0$]. We have the usual relation with the Wronskian $f_\epsilon(-k) = W[f_\epsilon(-k, r), \varphi_\epsilon(k, r)]$. As in nonrelativistic potential, if $f_\epsilon(-k) = 0$ in $\operatorname{Im} k > 0$, then, from the Wronskian relation $\varphi_\epsilon(k, r) = Gf_\epsilon(k, r)$ and from the boundary conditions for the solutions, $G \equiv 0$ even for $k = 0$. If k is inside the analyticity domain of $f_\epsilon(k)$, then $G = -f_\epsilon(k)/2ik$. The great difference in $\operatorname{Im} k > 0$ is for the location as well as the multiplicity of the roots. From the usual transformation we get

$$(k^2 - k^{2*})\int_0^\infty |\varphi_\epsilon|^2\, dr = \epsilon[(m^2 + k^2)^{\frac{1}{2}} - (m^2 + k^2)^{\frac{1}{2}}]^* \times \int_0^\infty \lambda_1 V_1 |\varphi_\epsilon|^2\, dr. \tag{47}$$

Here we do not necessarily have $\operatorname{Re} k = 0$ in $\operatorname{Im} k > 0$ [except in the case $\lambda_1 V_1 = 0$ and V_2 not linked to V_1, where the right-hand side of (47) vanishes and we recover the nonrelativistic result]. The supplementary term on the right-hand side is a consequence of the quasi-orthogonality relation[3] satisfied by the eigenfunctions of the Klein–Gordon equation (3), but this term, being linear in λ_1, is small with $|\lambda_1|$ small. We want to show, in the neighborhood of $k = 0$ ($|k| \ll m$) in $\operatorname{Im} k > 0$, that the root is on the imaginary axis if $|\lambda_1|$ is sufficiently small. From (47),

$$(k^2 - k^{2*})\left[\int_0^\infty |\varphi_\epsilon|^2\, dr - \frac{\lambda_1}{2}\int_0^\infty V_1 |\varphi_\epsilon|^2\, dr + O(|k^2 - k^{2*}|^2)\right] = 0. \tag{48}$$

Whereas the first term remains finite when $|\lambda_1| \to 0$, the second one vanishes; so we can find $|\lambda_1|$ sufficiently small such as the bracket is greater than 0. The remainder factor $(k^2 - k^{2*}) = 4i \operatorname{Im} k \operatorname{Re} k$ implies $\operatorname{Re} k = 0$ for $\operatorname{Im} k > 0$.

The possibility of complex roots in $\operatorname{Im} k > 0$ [at k and $-k^*$ following (11)] is a consequence of the fact that the roots along the imaginary axis are not necessarily simple. But we shall show that the multiplicity is one if $k = 0$ or if $|\lambda_1|$ is sufficiently small for k along a finite interval of the imaginary axis. Taking into account $f_\epsilon(-k) = 0$, we get

$$\partial f_\epsilon/\partial k = GW[\partial f_\epsilon/\partial k, f_\epsilon] + G^{-1}W[\varphi_\epsilon, \partial\varphi_\epsilon/\partial k],$$

and with the usual combination of the solutions of (3) for $(f_\epsilon, \partial f_\epsilon/\partial k)$ and $(\varphi_\epsilon, \partial\varphi_\epsilon/\partial k)$ we get

$$\frac{\partial f_\epsilon}{\partial k} = G\left[-2k\int_0^\infty f_\epsilon^2\, dr + \frac{\epsilon k\lambda_1}{(k^2 + m^2)^{\frac{1}{2}}}\int_0^\infty f_\epsilon^2 V_1\, dr\right].$$

Whereas the first term does not vanish when $\lambda_1 \to 0$, the second does. So this second term for $|\lambda_1|$ small is negligible. We use this result for $k = i\chi$ ($0 < \chi < m$).

For family (4) we can write $f_\epsilon(-i\chi, r) = e^{-\chi r} + g_\epsilon(\chi, r)$, where $|g_\epsilon| < \text{const} \exp[-\chi r - \text{const}\, r]$. Then

$$\frac{\partial f_\epsilon}{\partial\chi}(-i\chi) = G\chi\left[\int_0^\infty f_\epsilon^2\, dr + \frac{\epsilon\lambda_1}{(m^2 - \chi^2)^{\frac{1}{2}}}\int_0^\infty V_1 f_\epsilon^2\, dr\right] \tag{49}$$

$$= G\left[1 + 2\chi\int_0^\infty [g_\epsilon^2 + 2e^{-\chi r}g_\epsilon]\, dr + \frac{\epsilon\chi\lambda_1}{(m^2 - \chi^2)^{\frac{1}{2}}}\int_0^\infty f_\epsilon^2 V_\eta\, dr\right]. \tag{49'}$$

[15] R. G. Newton, J. Math. Phys. **1**, 319 (1960).
[16] M. C. Barthelemy, Ann. Inst. H. Poincaré **7**, 115 (1967).

In (49) for $|\lambda_1|$ small the first term in the bracket is dominant. Furthermore, in (49′) the second term vanishes if $\chi \to 0$ and the third term vanishes if $|\lambda_1| \to 0$ or $\chi \to 0$. So for $|\lambda_1|$ sufficiently small we can always choose χ_0 such that, for $0 \leq \chi \leq \chi_0$, the sum of the modulus of the second and the third terms is less than one and that $\partial f_\epsilon(-k)/\partial\chi$ does not vanish in the interval. Finally, we can take the limit $\chi > 0$ going to zero in (49′) because the integrals go to finite limits; we get $(\partial f_\epsilon(-i\chi)/\partial\chi)_{\chi=0} = G \neq 0$, without restriction about λ_1.

In conclusion, for $|\lambda_1|$ small there exists always a finite strip $0 \leq \operatorname{Im} k \leq \chi_0$ where the roots of $f_\epsilon(-k) = 0$, being simple and on the imaginary axis, can be associated with bound states.

So, concerning the possibility of bound states for regular interactions, we note the difference between Schrödinger or Dirac formalism on the one hand and Klein–Gordon on the other. This possibility is clearly connected with the appearance of EV_1 in (3), from which it follows that in the normalization, as well as the orthogonality relation, V_1 cannot disappear. Now the confluence of a pair of real E roots or the possibility of conjugate E pair of complex frequencies (studied in Ref. 6, where the difference with the Dirac case is emphasized) has been rejected by physicists because it leads to difficulties for the Hamiltonian. We can explain qualitatively how this happened.[6] Let us assume that for $\lambda_1 < \lambda_0$ there exist two roots of $f_\epsilon(-k) = 0$ (one for $\epsilon = 1$ and one for $\epsilon = -1$) along the $\operatorname{Im} k > 0$ axis $(0 < \operatorname{Im} k < m)$. As λ_1 increases, one of the two roots can go through the cut in the other sheet of the complex k plane where it meets the other root; subsequently they can go off the imaginary axis (k and $-k^*$). If we compare (47) for the Klein–Gordon equation and the corresponding ones for Dirac[16] and Schrödinger formalisms, we see that the difference concerning the location of the roots along the $\operatorname{Im} k > 0$ axis is due to the existence (or not) of positivity properties in the integration. We recall that the same kind of difference occurs for the fourth component of the associate currents.

VI. CONCLUSION

In this paper we have mainly tried to understand whether or not the N/D equations corresponding to the Klein–Gordon formalism have a unique solution when the dependence of the potential is taken into account explicitly in the discontinuities. Because the Klein–Gordon formalism has not been as completely investigated as the Schrödinger one, we must extend some results of the nonrelativistic potential theory: analytical properties in the k plane, the Martin inversion relation, the dispersive Marchenko equation, the De Alfaro and Regge bound, etc. Because people working in the inversion problem use both phase shifts for positive and negative energies, we take into account both the corresponding discontinuities.

Consequently, we get some complications in the N/D equations, but we have the advantage of obtaining solutions for both positive and negative energies. The main difference with Schrödinger theory (besides the cuts in the two half-planes for the Jost functions) is the fact that the asymptotic behavior depends explicitly on the interaction. So the Fredholm determinant of the N/D equations exhibits this result. This is the key for the problem of the existence of the solutions. On the other hand, the necessity of going to weak coupling limits in order to avoid complex frequencies is certainly a great restriction of the theory. With this restriction in mind, we note that the Klein–Gordon formalism is as convenient a model as the Schrödinger formalism for studying some aspects of strong interaction dynamics. We remark that many other aspects[1,11] of the N/D equations not considered in this paper can be investigated as well. Similarly, a complete formalism of the inversion problem "à la Marchenko" is certainly possible. The extension of these results to Dirac formalism might also be useful.

ACKNOWLEDGMENTS

I thank Dr. J. D. Bessis, Dr. G. Mahoux, Dr. H Navelet, and Professor C. Bloch for discussions. It is also a great pleasure for me to thank Dr. J. Bros, Dr. M. Froissart, and Dr. R. Stora for very useful discussions, suggestions, and help in this work. I thank also Professor C. Bloch and Dr. M. Froissart for the warm hospitality at the Theoretical Division.

APPENDIX A

We want to obtain some asymptotic properties for both the Jost solutions and the discontinuities $\Delta^{\pm}(x)$ using the Laplace transform properties of the solutions. Then the properties are the extension to the Klein–Gordon case of results obtained by Martin[8] for the Schrödinger one.

I. First, for $k \neq \pm im$ and k outside the cuts $k = i[\mu/2, \infty]$, we want to show that

$$f_\epsilon(k, r)e^{ikr} \underset{|k|\to\infty}{\simeq} \exp\left(-i\epsilon \frac{E}{k}\lambda_1 \int_r^\infty \frac{V_1(u)}{2}\,du\right), \quad \text{(A1)}$$

where

$$f_\epsilon(k, r)e^{ikr} = 1 + \sum_{n=1}^{\infty} \int_\mu^\infty \rho_{k,\epsilon}^{(n)}(\alpha)e^{-\alpha r}\,d\alpha, \quad \text{(A2)}$$

$\rho_{k,\epsilon}^{(n)}(\alpha)$ being the nth iterate term of the solution of (6b).

Using the Martin's bounds[8] outside the cuts, $|\alpha + 2ik| > \alpha \sin \epsilon'$ and $|\alpha + 2ik| > 2|k| \sin \epsilon'$, from (6b) we get

$$|\rho_{k,\epsilon}(\alpha)| < \frac{a_1}{\alpha^{1+\eta}} + \frac{a_1}{\alpha} \times \int_{\mu}^{\alpha-\sup(\mu_1,\mu_2)} \left(\frac{1}{(\alpha-\beta)^{\eta_1}} + \frac{a_2}{(\alpha-\beta)^{\eta_2}} \right) \times |\rho_{k,\epsilon}(\beta)|\, d\beta, \quad \text{(A3)}$$

$\eta = \inf(\eta_1, \eta_2)$ and where a_1 and a_2 are positive k-independent constants such that

$$\left| \frac{C_\epsilon(\alpha)}{\alpha(\alpha+2ik)} \right| < \frac{|\lambda_1|\,|E|}{2k \sin\epsilon' \alpha^{1+\eta_1}} + \frac{|\lambda_2|}{\alpha^{1+\eta_2} \sin\epsilon'} < \frac{a_1}{\alpha^{1+\eta}},$$

$$\left| \frac{C_\epsilon(\alpha-\beta)}{\alpha(\alpha+2ik)} \right| < \frac{a_1}{\alpha(\alpha-\beta)^{\eta_1}} + \frac{a_1 a_2}{\alpha(\alpha-\beta)^{\eta_2}}.$$

From (A3), $|\rho_{k,\epsilon}(\alpha)| < B(\alpha)$, where

$$B(\alpha) = F(\alpha)\left[1 + \int_{\mu}^{\alpha-\sup(\mu_1,\mu_2)} M(\alpha,\beta) B(\beta)\, d\beta\right] \quad \text{(A4)}$$

and

$$F(\alpha) = a_1/\alpha^{1+\eta},$$
$$M(\alpha,\beta) = \alpha^{\eta} a_1 [(\alpha-\beta)^{-\eta_1} + a_2(\alpha-\beta)^{-\eta_2}].$$

Note that $B(\alpha)$ is a majorant k independent. If we iterate (A4), we find

$$\int_{\mu}^{\infty} |\rho_{k,\epsilon}^{(n)}(\alpha)|\, e^{-\alpha r}\, d\alpha < \int_{\mu}^{\infty} B^{(n)}(\alpha) e^{-\alpha r}\, d\alpha < \int_{\mu}^{\infty} d\alpha F(\alpha) e^{-\alpha r} \frac{1}{n!}\left[\int_{\mu}^{\alpha-\sup(\mu_1,\mu_2)} M(\alpha,\beta) F(\beta)\, d\beta\right]^n < \int_{\mu}^{\infty} d\alpha F(\alpha) e^{-\alpha r} \cdot \exp \text{const} < \text{const},$$

because, as we shall see later,

$$\int_{\mu}^{\alpha-\sup(\mu_1,\mu_2)} MF\, d\beta < \text{const}.$$

So the series in (A2) are absolutely and uniformly convergent in the whole k complex plane for k outside the cuts. In order to get the limit $|k| \to \infty$ in (A2), we can, for each term, take the limit $|k| \to \infty$. But

$$\lim_{|k|\to\infty} \int_{\mu}^{\infty} \rho_{k,\epsilon}^{(n)}(\alpha) e^{-\alpha r}\, d\alpha = \int_{\mu}^{\infty} \rho^{(n)}(\alpha) e^{-\alpha r}\, d\alpha,$$

where $\rho^{(n)}$ is the nth iteration of

$$\rho(\alpha) = \lim_{|k|\to\infty}\left[\frac{E\lambda_1 \epsilon C_1(\alpha)}{2ik\alpha} + \frac{E\lambda_1\epsilon}{2ik\alpha} \times \int_{\mu_1}^{\alpha-\mu_1} C_1(\alpha-\beta)\rho(\beta)\, d\beta\right].$$

Then

$$\rho^{(n)}(\alpha) = \lim_{|k|\to\infty}\left[\frac{E\lambda_1\epsilon}{2ik}\right]^n \times \int_{\mu_1}^{\alpha-\mu_1} \frac{C_1(\alpha-\alpha_1)}{\alpha} \int_{\mu_1}^{\alpha_1-\mu_2} \frac{C_1(\alpha_1-\alpha_2)}{\alpha_1} \cdots \times \int_{\mu_1}^{\alpha_{n-2}-\mu_1} C_1(\alpha_{n-2}-\alpha_{n-1}) \frac{C_1(\alpha_{n-1})}{\alpha_{n-1}}$$

and

$$\lim_{|k|\to\infty} \int_{\mu}^{\infty} \rho_{k,\epsilon}^{(n)}(\alpha) e^{-\alpha r}\, d\alpha = \lim_{|k|\to\infty} \left(\frac{E}{ik}\right)^n \frac{1}{n!}\left[\int_r^{\infty} \frac{\lambda_1 \epsilon V_1(u)\, du}{2}\right]^n.$$

The result (A1) follows.

II. Secondly, we want to show that the solutions (15) for $k = ix$ and $\alpha < 2x - \sup(\mu_1, \mu_2)$ satisfy the following bounds:

$$\begin{pmatrix} |\tau_{1,ix}(\alpha)| \\ |\tau_{2,ix}(\alpha)| \end{pmatrix} < \begin{pmatrix} \dfrac{x\,|\lambda_2|\, a_1}{\alpha^{\eta+1}(2x-\alpha)} \\ \dfrac{|\lambda_1|\, a_2}{\alpha^{\eta+1}(2x-\alpha)} \end{pmatrix}, \quad \alpha < 2x - \sup(\mu_1,\mu_2), \quad \text{(A5)}$$

a_1 and a_2 being positive constants, λ_i-dependent but finite for $|\lambda_i|$ finite. Furthermore, a_i are α and x independent. Moreover, $\eta = \inf(\eta_1, \eta_2)$.

Using the bounds (4) for C_i, from (15) we get, for $k = ix$, an inequality integral equation which we write in the following manner:

$$\tau_x(\alpha) < F_x(\alpha)\left[\Lambda + \int_{\mu\inf}^{\alpha-\mu\sup} K_x(\alpha,\beta)\tau_x(\beta)\, d\beta\right], \quad \text{(A6)}$$

where

$$\tau_x(\alpha) = \begin{pmatrix} |\tau_{1,ix}(\alpha)| \\ |\tau_{2,ix}(\alpha)| \end{pmatrix}, \quad \Lambda = \begin{pmatrix} |\lambda_1| \\ |\lambda_2| \end{pmatrix},$$

$$F_x(\alpha) = \begin{pmatrix} \dfrac{x \text{ const}}{\alpha^{1+\eta}(2x-\alpha)} & 0 \\ 0 & \dfrac{\text{const}}{\alpha^{\eta+1}(2x-\alpha)} \end{pmatrix},$$

and

$$K_x(\alpha,\beta) = \begin{pmatrix} \alpha^{\eta}|\lambda_2|(\alpha-\beta)^{1-\eta_2}x^{-1} & (m^2-x^2)|\lambda_1|\alpha^{\eta}(\alpha-\beta)^{-\eta_1} \\ |\lambda_1|\alpha^{\eta}(\alpha-\beta)^{-\eta_1} & |\lambda_2|(\alpha-\beta)^{1-\eta_2}\alpha^{\eta} \end{pmatrix}.$$

If we iterate (A6), we get

$$\tau_x(\alpha) < F_x(\alpha) \sum_0^{\infty} \frac{1}{n!}\left[\int_{\mu\inf}^{\alpha-\mu\sup} K_x(\alpha,\beta) F_x(\beta)\right] \cdot \Lambda, \quad \text{(A7)}$$

where

$$\int K_x F_x\, d\beta = \begin{pmatrix} T_{1,1}(x,\alpha) & T_{1,2}(x,\alpha) \\ T_{2,1}(x,\alpha) & T_{2,2}(x,\alpha) \end{pmatrix}.$$

Let us define

$$I(\alpha, \eta, \eta_i, \eta_j) = \alpha^\eta \int_a^{\alpha-b} (\alpha - \beta)^{-\eta_i} \beta^{-(1+\eta_j)}\, d\beta,$$

where a and b are arbitrary constants > 0. Then it is easy to find

$$I < \text{const} \quad (\text{for any } \eta_i, \eta_j). \tag{A8}$$

We have

$$T_{1,1}(x, \alpha) = |\lambda|\, \alpha^\eta \int_a^{\alpha-b} \frac{(\alpha - \beta)^{1-\eta_2}}{\beta^{1+\eta}(2x - \beta)} < |\lambda_2|\, I$$

using $2x - \beta > \alpha - \beta$,

$$T_{1,2}(x, \alpha) = \frac{|m^2 - x^2|}{x} |\lambda| \times \int_a^{\alpha-b} \alpha^\eta \frac{(\alpha - \beta)^{-\eta_1}}{\beta^{1+\eta}(2x - \beta)}\, d\beta < |\lambda_1|\, I\, \text{const},$$

using

$$\frac{1}{\beta(2x - \beta)} = \frac{1}{2x}\left(\frac{1}{\beta} + \frac{1}{2x - \beta}\right) < \frac{1}{2x}\left(\frac{1}{\beta} + \frac{1}{\alpha - \beta}\right),$$

and

$$T_{2,1} = |\lambda_1|\, \alpha^\eta \int_a^{\alpha-b} \frac{d\beta}{(\alpha - \beta)^{\eta_1} \beta^{1+\eta}(2x - \beta)} < |\lambda_1|\, I\, \text{const},$$

using the same decomposition as $T_{1,2}$. Also,

$$T_{2,2} = |\lambda|\, \alpha^\eta \int_a^{\alpha-b} \frac{(\alpha - \beta)^{1-\eta_2}}{\beta^{1+\eta}(2x - \beta)} < |\lambda_1|\, I\, \text{const},$$

using $2x - \beta > \alpha - \beta$. Finally, if we substitute in (A7),

$$\tau_x(\alpha) < F_x(\alpha) \sum_0^\infty \frac{1}{n!} \begin{pmatrix} C_{11}\,|\lambda_2| & C_{12}\,|\lambda_1| \\ C_{21}\,|\lambda_1| & C_{22}\,|\lambda_2| \end{pmatrix}^n \Lambda, \tag{A9}$$

where C_{ij} are positive constants independent of λ_1, λ_2.

III. Thirdly, if we substitute in (24) $|\tau_{j,ix}(\alpha)|$ by the bounds (A9), we get bounds for $|\Delta^{\pm}(x)|$ when x is large. From (24),

$$|\Delta^+(x)| < \frac{|\lambda_2|}{x^{\eta_2}} \text{const} + |\lambda_2| \frac{I(2x, \eta, \eta_1, \eta)\, \text{const}\,(|\lambda_i|)}{x^\eta} + \frac{|m^2 - x^2|}{2x} \int_\mu^{2x-\mu_1} \left(\frac{\text{const}\,(|\lambda_i|)}{\alpha^\eta (2x - \alpha)^{\eta_1} 2x}\right) \times \left(\frac{1}{\alpha} + \frac{1}{2x - \alpha}\right) d\alpha,$$

$$|\Delta^-(x)| < \frac{|\lambda_1|\, \text{const}}{x^{1+\eta_1}} + \text{const}\,(|\lambda_i|) \times \int_{\mu_1}^{2x-\mu_1} d\alpha \left(\frac{|\lambda_1|^{d\alpha}}{\alpha^{\eta_1}(2x - \alpha)^{\eta_1} 2x}\right) \left(\frac{1}{\alpha} + \frac{1}{2x - \alpha}\right) + \frac{\text{const}}{x^{1+\eta}} (|\lambda_i|)\, |\lambda_2|\, I(2x, \eta, \eta_2, \eta).$$

Finally,

$$\begin{pmatrix} |\Delta^+(x)| \\ |\Delta^-(x)| \end{pmatrix} < \begin{pmatrix} \dfrac{C^+}{x^\eta} \\ \dfrac{C^-}{x^{1+\eta}} \end{pmatrix},$$

$C^{\pm}$ being $|\lambda_i|$-dependent constants but finite for $|\lambda_1|$ and $|\lambda_2|$ finite.

APPENDIX B

I. We investigate the Fredholm determinant of $\nu\Delta_{ij}$ given in (20a):

$$\nu\Delta_{ij}(x, y) = \begin{pmatrix} \dfrac{\nu}{x + y} & \Delta^+(y)\theta(y - \mu_+) & \Delta^{\sim}(y)\theta(y - \mu_-) \\ & \Delta^-(y)\theta(y - \mu_-) & \Delta^+(y)\theta(y - \mu_+) \end{pmatrix},$$

where we have written $\Delta^-(y)(m^2 - y^2) = \Delta^{\sim}(y)$. We have

$$\mathfrak{D}(\nu) = 1 + \sum_1^\infty \frac{(-\nu)^n}{n!} \mathfrak{D}^{(n)},$$

$$\mathfrak{D}^{(n)} = \sum_{i_1, i_2, \cdots, i_n} \int_0^\infty dy_1 \cdots \int_0^\infty dy_n \begin{vmatrix} \Delta_{i_1,i_1}(y_1, y_1) & \Delta_{i_1,i_2}(y_1, y_2) & \cdots & \Delta_{i_1,i_n}(y_1, y_n) \\ \Delta_{i_2,i_1}(y_2, y_1) & \Delta_{i_2,i_2}(y_2, y_2) & \cdots & \Delta_{i_2,i_n}(y_2, y_n) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \Delta_{i_n,i_1}(y_n, y_1) & \cdots\cdots\cdots & & \Delta_{i_n,i_n}(y_n, y_n) \end{vmatrix}. \tag{B1}$$

First, using mainly the symmetry of the diagonal elements $\Delta_{11} = \Delta_{22}$, we get

$$\sum_{i_1, i_2, \cdots, i_n} \int dy_1 \cdots \int dy_n = 2 \sum_{i_2, \cdots, i_n} \int dy_1 \cdots \int dy_n.$$

Secondly, still using $\Delta_{11} = \Delta_{22}$ and symmetry properties in the integration of y_i, it can be shown that $\mathfrak{D}^{(n)}$ can be written as

$$\frac{\mathfrak{D}^{(n)}}{2} = \sum_{p=0}^{p_{\max}} \gamma_n^p \mathfrak{D}_p^{(n)}. \tag{B2}$$

If n is odd, $p_{\max} = \frac{1}{2}(n - 1)$ and $\gamma_n^p = C_n^p = n!/p!\,(n - p)!$. If n is even, $p_{\max} = \frac{1}{2}n$, $\gamma_n^p = C_n^p$ for

$p < p_{\max}$, and $\gamma_n^{n/2} = C_{n-1}^{(n/2)-1}$.

$$\mathfrak{D}_p^{(n)} = \int_0^\infty dy_1 \cdots \int_0^\infty dy_n$$

$i \backslash j$	1	p	n
1	$\dfrac{\Delta^+(y_j)}{y_i + y_j}$	$\dfrac{\Delta^{\sim}(y_j)}{y_i + y_j}$	
p	$\dfrac{\Delta^-(y_j)}{y_i + y_j}$	$\dfrac{\Delta^+(y_j)}{y_i + y_j}$	,
n			

where for simplicity the $\theta(y - \mu_\pm)$ have been omitted.

This decomposition is trivial for $n = 2$ and easy to get from symmetry properties for $n = 3$:

$$\tfrac{1}{2}\mathfrak{D}^{(2)} = \mathfrak{D}_0^{(2)} + \mathfrak{D}_1^{(2)}, \quad \tfrac{1}{2}\mathfrak{D}^{(3)} = \mathfrak{D}_0^{(3)} + 3\mathfrak{D}_1^{(3)}.$$

Note that $\Delta_{i_1=1,i_n}$ is either $\Delta^+/y_1 + y_n$ or $\Delta^{\sim}/y_1 + y_n$ accordingly as $i_n = 1$ or 2. For $i_1 \neq i_n$ note also that Δ_{i_1,i_n} corresponds to $\Delta^{\sim}$ (or Δ^-) and Δ_{i_m,i_1} to Δ^- (or $\Delta^{\sim}$). Note also that Δ_{ii} the diagonal elements correspond to Δ^+. Assuming (B2) for $n - 1$, adding Δ_{1,i_n}, and taking account of symmetry properties, it can be shown true for n. Thirdly, for each $\mathfrak{D}_p^{(n)}$, still taking into account symmetry properties, we get

$$\mathfrak{D}_p^{(n)} = \sum_{q=0}^{q=p} C_p^q \mathfrak{D}_{p,q}^{(n)}, \tag{B3}$$

$$\mathfrak{D}_{p,q}^{(n)} = \int_0^\infty dy_1 \cdots \int_0^\infty dy_n$$

$i \backslash j$	1	$p-q$	p	n
1	$\dfrac{\Delta^+(y_j)}{y_i + y_j}$	0	$\dfrac{\Delta^{\sim}(y_j)}{y_i + y_j}$	
p	0	$\dfrac{\Delta^-(y_j)}{y_i + y_j}$	$\dfrac{\Delta^+(y_j)}{y_i + y_j}$	,
n				

where now only q columns with Δ^- exist. This decomposition is obtained in the following way: In the first column $\mathfrak{D}_p^{(n)}$ we put all the elements Δ^+ equal to zero and after the elements with Δ^- equal to zero. Making the same operation with both determinants in the second, $\cdots$, pth column, we get 2^p determinants. Taking into account the symmetry properties, we group them following (B3). $\mathfrak{D}_{p,q}^{(n)}$ can be calculated explicitly, taking into account the symmetries in the integrations.

We give briefly the method. First we factorize $\Delta^+(y_j)$ in the first $(p - q)$ columns and factorize $(y_1 + y_i)^{-1}$ in the first p rows. Then the p first elements of the first column are equal to one. We subtract from all elements of the ith row $(i = 2, p)$ the corresponding elements of the first row. We are thus led to a new determinant, where the first column and the first row of the first one are absent. We continue this operation up to a determinant where the $(p - q)$ first rows and $(p - q)$ first columns of the original one are absent:

$i \backslash j$	$p-q+1$	p	n
$p-q+1$	0	$\dfrac{\Delta^{\sim}(y_j)}{y_i + y_j} \prod_{m=1}^{p-q} \left(\dfrac{y_m - y_j}{y_m + y_j}\right)$	
p	$\dfrac{\Delta^-(y_j)}{y_i + y_j}$	$\dfrac{\Delta^+(y_j)}{y_i + y_j}$	.
n			

We develop the determinant following the $(n - p)$ elements $\neq 0$ of the first row and remark (with symmetry properties) that they give the same contribution. So we consider the minor corresponding to the element of the first row and last column ($i = p - q + 1$ and $j = n$). Similarly, we develop this new determinant following the $(n - p - 1)$ elements $\neq 0$ of the first row which still give the same contribution. Still taking the minor corresponding to the first row and the last column, we continue the operation up to a

determinant (where $\Delta^{\sim}$ has disappeared) with $n-p$ rows and columns:

i \ j	$p+1-q$	p	$n-q$
$p+1$	$\dfrac{\Delta^{-}(y_j)}{y_i+y_j}$	$\dfrac{\Delta^{+}(y_j)}{y_i+y_j}$	
n			

In this determinant we factorize $\Delta^{-}(y_j)$ and $\Delta^{+}(y_j)$; the remaining determinant with elements $(y_i+y_j)^{-1}$ $(i=p+1,\cdots,n;\ j=p+1-q,\cdots,n-q)$ can be easily calculated:

$$\mathfrak{D}^{(n)}_{p,q}=\int_{\mu^+}^{\infty}dy_1\cdots\int_{\mu^+}^{\infty}dy_{p-q}\int_{\mu^+}^{\infty}dy_{p+1}\cdots\int_{\mu^+}^{\infty}dy_{n-q}\int_{\mu^-}^{\infty}dy_{p-q+1}\cdots\int_{\mu^-}^{\infty}dy_p\int_{\mu^-}^{\infty}dy_{n-q+1}\cdots\int_{\mu^-}^{\infty}dy_n M^{(n)}_{p,q}(y_1,\cdots,y_n),$$

$$M^{(n)}_{p,q}=\prod_{k=1}^{k=q}\left[\frac{\Delta^{\sim}(y_{n-k+1})\Delta^{-}(y_{p-q+k})}{(y_{n-k+1}+y_{p-q+k})^2}\prod_{i=p+1}^{i=n-k}\frac{(y_{n-k+1}-y_i)}{(y_{p-q+k}+y_i)}\prod_{i=k+p+1-q}^{n-q}\frac{(y_{p-q+k}-y_i)}{(y_{n-k+1}+y_i)}\right]\frac{(n-p)!}{(n-p-q)!}$$

$$\times\prod_1^{p-q}\frac{\Delta^{+}(y_j)}{2y_j}\prod_{i=1}^{p-q}\left[\prod_{j=i+1}^{p-q}\frac{(y_i-y_j)}{(y_i+y_j)}\prod_{j=i+1}^{p}\frac{(y_i-y_j)}{(y_i+y_j)}\prod_{j=n-q+1}^{n}\frac{(y_i-y_j)}{(y_i+y_j)}\right]\prod_{p+1}^{n-q}\frac{\Delta^{+}(y_i)}{2y_i}\prod_{\substack{i=p+1\\ j=p+1}}^{\substack{i<j\\ j=n-q}}\left(\frac{y_i-y_j}{y_i+y_j}\right)^2.$$

We put

$$\int_{\mu^+}^{\infty}\frac{|\Delta^{+}(x)|}{x}\,dx=a,$$

$$\int_{\mu^-}^{\infty}\int_{\mu^-}^{\infty}dx\,dy\,\frac{|\Delta^{-}(x)|\,|\Delta^{\sim}(y)|}{(x+y)^2}=b,$$

and we get

$$|\mathfrak{D}^{(n)}_{p,q}|<\frac{1}{2^{n-2q}}a^{n-2q}b^q\frac{(n-p)!}{(n-p-q)!}. \qquad \text{(B4)}$$

From (B1)–(B4) we obtain finally

$$|\mathfrak{D}(\nu)|<\exp\left[|\nu|\,a+|\nu|^2\,b\right],$$

$$|\mathfrak{D}(\nu)-1|<\exp\left[|\nu|\,a+|\nu|^2\,b\right]-1.$$

II. For the kernel $\nu\Delta_{ij}(x,y)e^{-(x+y)r}$, (B1) becomes

$$\mathfrak{D}(\nu,r)=1+\sum_1^{\infty}\frac{(-\nu)^n}{n!}\mathfrak{D}^{(n)}(r),$$

$$\mathfrak{D}^{(n)}(r)=\sum_{i_1,i_2,\cdots,i_n}\int_0^{\infty}dy_1\cdots\int_0^{\infty}dy_n\left(\prod_1^n e^{-2y_ir}\right)\times\begin{vmatrix}\Delta_{i_1,i_1}(2y_1) & \cdots & \Delta_{i_1,i_n}(y_1,y_n)\\ \cdot & & \cdot\\ \cdot & & \cdot\\ \cdot & & \cdot\\ \Delta_{i_n,i_1}(y_n,y_1) & \cdots & \Delta_{i_n,i_n}(y_n,y_n)\end{vmatrix}. \qquad \text{(B5)}$$

The only change is the product $\prod e^{-2y_ir}$, and finally we get the bound

$$|\mathfrak{D}(\nu,r)|<\exp\left[|\nu|\int_{\mu^+}^{\infty}e^{-2xr}\frac{|\Delta^{+}(x)|}{x}\,dx+|\nu|^2\int_{\mu^-}^{\infty}\int_{\mu^-}^{\infty}dx\,dy e^{-2(x+y)r}\frac{\Delta^{-}(x)\Delta^{\sim}(y)}{(x+y)^2}\right].$$

III. For the kernel $\nu\mathcal{F}_{i,l}(y+t)$ the same method as was used in Ref. 1 shows that it leads to the same Fredholm determinant $\mathfrak{D}(\nu,r)$ given in (B5).

IV. We want to show that if $\mathfrak{D}(\nu_0)\neq 0$, then (20d) has a unique solution in L^2 and (20a) has also a unique solution in a well-defined space. We put

$$F=\begin{pmatrix}F_1(x)\\ F_2(x)\end{pmatrix}\quad\text{and}\quad F\in\mathcal{E}\quad\text{if}\quad\begin{pmatrix}|F_1(x)|<Cste\\ |F_2(x)|<Cste/x\end{pmatrix}$$

$$\text{for}\quad x\in[c,\infty],\quad 0<c<\inf(\mu_+,\mu_-).$$

We write Eq. (20a) as

$$F=F(\infty)+\nu\Delta\cdot F. \qquad \text{(B6)}$$

Let us assume that $F\in\mathcal{E}$ and F is a solution of Eqs. (20a) or (B6); we define $\bar{F}=F-F(\infty)$, $\bar{F}\in\bar{\mathcal{E}}$, and $\bar{\mathcal{E}}\subset\mathcal{E}$. Because $\bar{F}=\nu\Delta\cdot F$, if we use the bounds (20b),

$$|\Delta^{+}|<\text{const}/x^{\eta},\quad |\Delta^{-}|<\text{const}/x^{1+\eta},\quad\text{and}\quad \eta>0,$$

we get

$$\bar{F}=\begin{pmatrix}\bar{F}_1(x)\\ \bar{F}_2(x)\end{pmatrix},$$

where

$$\begin{pmatrix}|\bar{F}_1(x)|<\text{const}/\eta',\quad 0<\eta'<\eta\quad\text{if}\quad 0<\eta\leq 1,\\ \qquad\qquad |\bar{F}_1(x)|<\text{const}/x\quad\text{if}\quad \eta>1\\ |\bar{F}_2(x)|<\text{const}/x\end{pmatrix}.$$

For each $\bar{F}\in\bar{\mathcal{E}}$ [or $F\in\mathcal{E}$ and a solution of (B6)] we define

$$G^{(\alpha_1,\alpha_2)}=\begin{pmatrix}\dfrac{F_1(x)-F_1(\infty)}{x^{\alpha_1}}=\dfrac{\bar{F}_1(x)}{x^{\alpha_1}}\\ F_2(x)x^{\alpha_2}=\bar{F}_2(x)x^{\alpha_2}\end{pmatrix}=\begin{pmatrix}G_1(x)\\ G_2(x)\end{pmatrix}, \qquad \text{(B7)}$$

and we write Eq. (20d) as

$$G^{(\alpha_1,\alpha_2)} = G_0^{(\alpha_1,\alpha_2)}F_1(\infty) + \nu\Delta^{(\alpha_1,\alpha_2)}G^{(\alpha_1,\alpha_2)}. \quad \text{(B8)}$$

Note that

$$F \Leftrightarrow G^{(\alpha_1,\alpha_2)}.$$

If we choose $\alpha_1 = \alpha_2 = 0$ for $\eta > 1$ and $\alpha_1 = \frac{1}{2}$, $\alpha_2 = (\frac{1}{2}) - (\eta/2)$ for $0 < \eta \leq 1$, then $G^{(\alpha_1,\alpha_2)}$ as defined by (B7) belongs to L^2. With these choices of α_1 and α_2, and still using the bounds (20b) for $\Delta^{\pm}$, it is easy to see that the two elements of the free terms of (20d), as well as the four terms of the kernel $\Delta^{(\alpha_1,\alpha_2)}$, are square integrable. So $G^{(\alpha_1,\alpha_2)} \in L^2$ and $\Delta^{(\alpha_1,\alpha_2)} \in L^2$. Finally, our problem is reduced to solving a Fredholm equation (B8) or (20d) in L^2. Note that the Fredholm determinants of (B6) and (B8) are the same formally and so are well defined for (B8) from the theory of integral equation in L^2. If for a special value ν_0 one has $\mathfrak{D}(\nu_0) \neq 0$, then there exists a unique solution $G^{(\alpha_1,\alpha_2)}$ of (B8) in L^2 and the corresponding $F_{\nu_0} \in \bar{\mathfrak{F}}$ [or $F \in \mathfrak{F}$ and F a solution of (B6)] defined by (B7) is the unique solution of (B6).

APPENDIX C

We study Eq. (36), the extension of Marchenko equation to the Klein–Gordon case.

I. First[17] we want to show that the diagonal elements in the indices K_{ii} and diagonal in the variables $r = y$ are equal:

$$K_{11}(r, r) = K_{22}(r, r). \quad \text{(C1)}$$

Note that this is true for the free term because $\mathcal{F}_{11}(r, r) = \mathcal{F}_{22}(r, r)$. We write the kernel

$$\begin{pmatrix} \mathcal{F}_{11}(y+t) & \mathcal{F}_{12}(y+t) \\ \mathcal{F}_{21}(y+t) & \mathcal{F}_{11}(y+t) \end{pmatrix}$$

of (36) in an abstract form: $\mathcal{F} = F_{11} \otimes \mathbb{1} + F_{12} \otimes \sigma_+ + F_{21} \otimes \sigma_-$, where $\mathbb{1}$ and $\sigma_\pm$ are the identity and usual Pauli matrices acting on the indices (i, j), whereas F_{11}, F_{12}, and F_{21} are symmetric operators acting on the variables r and q. Now

$$\begin{aligned}\mathcal{F}^2 = {} & [F_{11}^2 + \tfrac{1}{2}(F_{21}F_{12} + F_{12}F_{21})] \otimes \mathbb{1} \\ & + \tfrac{1}{2}(F_{12}F_{21} - F_{21}F_{12}) \otimes \sigma_3 \\ & + (F_{11}F_{12} + F_{12}F_{11}) \otimes \sigma_+ \\ & + (F_{11}F_{21} + F_{21}F_{11}) \otimes \sigma_-.\end{aligned}$$

If we consider the diagonal elements in the indices, we look at $\mathbb{1}$ and σ_3. Because the factor of $\mathbb{1}$ is symmetric in the variable space, whereas that of σ_3 is antisymmetric, we see that the diagonal elements of $\mathcal{F}^2$ in both indices and variables space are equal.

Let us assume for $\mathcal{F}^n$ a structure similar to $\mathcal{F}^2$:

$$\mathcal{F}^n = S \otimes \mathbb{1} + S_+ \otimes \sigma_+ + S_- \otimes \sigma_- + A\sigma_3,$$

where $S_\pm$, S are symmetric, whereas A is antisymmetric. We have only to verify that $\mathcal{F}^{n+1}$ has the same structure:

$$\begin{aligned}\mathcal{F}^{(n+1)} &= \tfrac{1}{2}[\mathcal{F}\mathcal{F}^n + \mathcal{F}^n\mathcal{F}] \\ &= (AF_{12} - F_{12}A)\sigma_+ + (F_{21}A - AF_{21})\sigma_- \\ &\quad + [SF_{11} + F_{11}S + \tfrac{1}{2}(S_+F_{12} + F_{21}S_- \\ &\quad + S_-F_{21} + F_{12}S_+)] \otimes \mathbb{1} \\ &\quad + [AF_{11} + F_{11}A + (S_+F_{12} + F_{21}S_- \\ &\quad - S_-F_{21} - F_{12}S_+)] \otimes \sigma_3.\end{aligned}$$

We see that the factor of $\mathbb{1}$ is symmetric, whereas the factor of σ_3 is antisymmetric.

II. Secondly, we introduce a linear parameter ν in (36):

$$\begin{aligned}K_{i,j}(r, y) &= \nu\mathcal{F}_{i,j}(y + r) \\ &\quad + \nu \sum_l \int_r^\infty \mathcal{F}_{i,l}(y + t)K_{l,j}(r, t)\, dt. \quad \text{(C2)}\end{aligned}$$

We want to show that the trace of $K_{i,j}$ is linked to the Fredholm determinant of (C2):

$$K_{1,1}(r, r) + K_{2,2}(r, r) = \frac{(\partial/\partial r)\mathfrak{D}(\nu, r)}{\mathfrak{D}(\nu, r)}. \quad \text{(C3)}$$

The Fredholm-type solution of (C2) for $y = r$ can be written as

$$\begin{aligned}&\mathfrak{D}(\nu, r)\left[\sum_i K_{i,i}(r, r)\right] \\ &\quad = \nu\mathfrak{D}(\nu, r) \sum_i \mathcal{F}_{i,i}(2r) \\ &\quad\quad + \nu^2 \int_r^\infty \sum_{i,j} \mathcal{N}_{i,j}(r, t)\mathcal{F}_{j,i}(r + t)\, dt, \quad \text{(C4)}\end{aligned}$$

where

$$\mathfrak{D}(\nu, r) = 1 + \sum \frac{(-\nu)^n}{n!} \mathfrak{D}^{(n)}(r),$$

$$\mathcal{N}_{i,j}(r, t) = \sum \frac{(-\nu)^n}{n!} \mathcal{N}_{i,j}^{(n)}(r, t). \quad \text{(C5)}$$

In order to prove (C3), due to (C4) and (C5), we have only to show that

$$\begin{aligned}&\frac{1}{n+2}\frac{\partial}{\partial r}\mathfrak{D}^{(n+2)}(r) \\ &\quad = -\mathfrak{D}^{(n+1)}(r) \sum_i \mathcal{F}_{i,i}(2r) + (n+1) \\ &\quad\quad \times \sum_{i_1,i_2} \int_r^\infty \mathcal{N}_{i_1,i_2}^{(n)}(r, t_1)\mathcal{F}_{i_1,i_2}(r + t_1)\, dt. \quad \text{(C6)}\end{aligned}$$

From Fredholm's theory we get

$$\mathfrak{D}^{(n)}(r) = \int_r^\infty dt_1 \cdots \int_r^\infty dt_n E_n(t_1, t_2, \cdots, t_n),$$

$$\mathcal{N}_{i_1,i_2}^{(n)}(r, t_1) = \int_r^\infty dt_2 \cdots \int_r^\infty dt_{n+1} H_{i_1,i_2}^{(n)}(r, t_1, \cdots, t_{n+1}),$$

[17] I thank R. Stora for this proof.

where

$$E_n(t_1, t_2, \cdots, t_n) = \sum_{i_1, i_2, \cdots, i_n} \begin{vmatrix} \mathcal{F}_{i_1,i_1}(2t_1) & \mathcal{F}_{i_1,i_2}(t_1+t_2) & \cdots & \mathcal{F}_{i_1,i_n}(t_1+t_n) \\ \mathcal{F}_{i_2,i_1}(t_2+t_1) & \mathcal{F}_{i_2,i_2}(2t_2) & \cdots & \mathcal{F}_{i_2,i_n}(t_2+t_n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathcal{F}_{i_n,i_1}(t_n+t_1) & \mathcal{F}_{i_n,i_2}(t_n+t_2) & \cdots & \mathcal{F}_{i_n,i_n}(2t_n) \end{vmatrix},$$

$$H^{(n)}_{i_1,i_2}(r, t_1, t_2, \cdots, t_{n+1}) = \sum_{i_3, i_4, \cdots, i_{n+2}} \begin{vmatrix} \mathcal{F}_{i_1,i_2}(r+t_1) & \mathcal{F}_{i_1,i_3}(r+t_2) & \cdots & \mathcal{F}_{i_1,i_{n+2}}(r+t_{n+1}) \\ \mathcal{F}_{i_3,i_2}(t_2+t_1) & \mathcal{F}_{i_3,i_3}(2t_2) & \cdots & \mathcal{F}_{i_3,i_{n+2}}(t_2+t_{n+1}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathcal{F}_{i_{n+2},i_2}(t_{n+1}+t_1) & \mathcal{F}_{i_{n+2},i_3}(t_2+t_{n+1}) & \cdots & \mathcal{F}_{i_{n+2},i_{n+2}}(2t_{n+1}) \end{vmatrix}.$$

We get

$$\frac{1}{n+2}\frac{\partial}{\partial r}\mathcal{D}^{(n+2)}(r) = -\frac{1}{n+2}\sum_{j=1}^{n+2}\int_r^\infty dt_1 \cdots \int_r^\infty dt_{j-1}\int_r^\infty dt_{j+1}\cdots\int_r^\infty dt_{n+2}E_{n+2}(t_1, \cdots, t_{j-1}, r, t_{j+1}, \cdots, t_{n+2})$$

$$= -\int_r^\infty dt_1 \cdots \int_r^\infty dt_{n+1}E_{n+2}(r, t_1, \cdots, t_{n+1}),$$

where we have used symmetry properties in the exchange of $t_i \rightleftarrows t_j$:

$$E_{n+2}(r, t_1, \cdots, t_{n+1}) = \sum_{i_1, i_2, \cdots, i_{n+2}} \begin{vmatrix} \mathcal{F}_{i_1,i_1}(2r) & \mathcal{F}_{i_1,i_2}(r+t_1) & \cdots & \mathcal{F}_{i_1,i_{n+2}}(r+t_{n+1}) \\ \mathcal{F}_{i_2,i_1}(t_1+r) & \mathcal{F}_{i_2,i_2}(2t_1) & \cdots & \mathcal{F}_{i_2,i_{n+2}}(t_1+t_{n+1}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathcal{F}_{i_{n+2},i_1}(t_{n+1}+r) & \mathcal{F}_{i_{n+2},i_2}(t_{n+1}+t_1) & \cdots & \mathcal{F}_{i_{n+2},i_{n+2}}(2t_{n+1}) \end{vmatrix}.$$

We develop E_{n+2} following the elements of the first column and we write

$$E_{n+2}(r, t_1, \cdots, t_{n+1}) = \sum_{i_1}\mathcal{F}_{i_1,i_1}(2r)E_{n+1}(t_1, t_2, \cdots, t_{n+1}) + \sum_{j=0}^{n+1}(-1)^j\sum_{i_{j+1},i_1}\mathcal{F}_{i_{j+1},i_1}(t_j+r)M^{(n+1,j)}_{i_1,i_{j+1}}(r, t_1, \cdots, t_{n+1}),$$

where

$$M^{(n+1,1)}_{i_1,i_2}(r, t_1, \cdots, t_{n+1}) = H^{(n)}_{i_1,i_2}(r, t_1, \cdots, t_{n+1})$$

and

$$M^{(n+1,j)}_{i_1,i_{j+1}}(r, t_1, \cdots, t_{n+1}) = \sum_{i_2, i_3, \cdots, i_j, i_{j+2}, \cdots, i_{n+2}} \begin{vmatrix} \mathcal{F}_{i_1,i_2}(r+t_1) & \cdots & \mathcal{F}_{i_1,i_{n+2}}(r+t_{n+1}) \\ \mathcal{F}_{i_2,i_2}(2t_1) & \cdots & \mathcal{F}_{i_2,i_{n+2}}(t_1+t_{n+1}) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \mathcal{F}_{i_j,i_2}(t_{j-1}+t_1) & \cdots & \mathcal{F}_{i_j,i_{n+2}}(t_{j-1}+t_{n+1}) \\ \mathcal{F}_{i_{j+2},i_2}(t_{j+1}+t_1) & \cdots & \mathcal{F}_{i_{j+2},i_{n+2}}(t_{j+1}+t_{n+1}) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \mathcal{F}_{i_{n+2},i_2}(t_{n+1}+t_1) & \cdots & \mathcal{F}_{i_{n+2},i_{n+2}}(2t_{n+1}) \end{vmatrix}.$$

Similarly, as was done for the nonrelativistic case,[1–12] it is easy to get the identity

$$\int_r^\infty dt_1 \cdots \int^\infty dt_{n+1} \sum_{i_{j+1},i_1} \mathcal{F}_{i_{j+1},i_1}(t_j + r) M^{(n+1,j)}_{i_1,i_{j+1}}(r, t_1, \cdots, t_{n+1})$$
$$= \int_r^\infty dt_1 \cdots \int_r^\infty dt_{n+1} (-1)^{j-1} \sum_{i_2,i_1} \mathcal{F}_{i_2,i_1}(t_1 + r) M^{(n+1,1)}_{i_1,i_2}(r, t_1, \cdots, t_{n+1}).$$

Finally, we get

$$\frac{1}{n+2} \frac{\partial}{\partial r} \mathfrak{D}^{(n+2)}(r) = -\sum_i \mathcal{F}_{ii}(r) \int_r^\infty dt_1 \cdots \int_r^\infty dt_{n+1} E_{n+1}(t_1, t_2, \cdots, t_{n+1})$$
$$+ (n+1) \int_r^\infty dt_1 \cdots \int_r^\infty dt_{n+1} \sum_{i_1,i_2} \mathcal{F}_{i_1,i_2}(t_1 + r) H^{(n)}_{i_1,i_2}(r, t_1, \cdots, t_{n+1}),$$

which is the relation (C6) or (C3).

APPENDIX D

We want to find the relations between K_i, Φ transforms of f_i:

$$f_1 = e^{-ikr}\Phi(r) + \int_r^\infty K_1(r,t) e^{-ikt}\, dt,$$
$$f_2 = \int_r^\infty K_2(r,t) e^{-ikt}\, dt, \qquad \text{Im } k < 0, \quad \text{(D1)}$$

and the potentials $\lambda_i V_i$ using the fact that the f_i satisfy the coupled integral equations (30). First, in

$$f_2(k, E^2, r) = \int_r^\infty \frac{\sin k(s-r)}{k} [f_1 \lambda_1 V_1 + f_2 \lambda_2 V_2]\, ds, \quad \text{(D2)}$$

we insert the f_i given by the representations (D1). After eliminating the trigonometric functions, we get

$$K_2(r,t) = \tfrac{1}{2} \int_{\frac{1}{2}(t+r)}^\infty \lambda_1 V_1(s) \Phi(s)\, ds$$
$$+ \tfrac{1}{2} \int_r^\infty ds \int_s^\infty du \left[\sum_i \lambda_i V_i(s) K_i(s,u) \right]$$
$$\times \left[\theta(t+s-r-u) - \theta\left(\frac{t+r}{2} - s\right) \right.$$
$$\left. \times\, \theta(t+r-s-u) \right]. \quad \text{(D3)}$$

Secondly, in the relation given by the first row in (30) we insert the expression (D2) of the second row

$$f_1 = e^{-ikr} + \int_r^\infty \sin \frac{k(s-r)}{k} \lambda_2 f_1 V_2\, ds$$
$$+ \int_r^\infty ds \sin k(s-r) \frac{(k^2+m^2)}{k^2} \lambda_1 V_1(s)$$
$$\times \int_s^\infty du \sin k(u-s) [f_1 \lambda_1 V_1 + \lambda_2 V_2 f_2]. \quad \text{(D4)}$$

Similarly in (D4), we insert (D1) and eliminate the trigonometric functions:

$$\Phi(r) = 1 - \tfrac{1}{4} \int_r^\infty \lambda_1 V_1(s) \int_1^\infty \lambda_1 V_1(u) \Phi(u)\, du,$$
$$\Phi = \cos \int_r^\infty \lambda_1 \frac{V_1(u)}{2}\, du, \quad \text{(D5)}$$

$$K_1(r,t) = \frac{\lambda_2}{2} \int_{\frac{1}{2}(t+r)}^\infty V_2(s) \Phi(s)\, ds + \tfrac{1}{2} \int_r^\infty ds \int_s^\infty du [m^2 \lambda_1 V_1(s) K_2(s,u) + \lambda_2 V_2(s) K_1(s,u)]$$
$$\times \left[\theta(t+s-r-u) - \theta\left(\frac{t+r}{2} - s\right) \theta(t+r-s-u) \right]$$
$$+ \frac{\lambda_1^2}{8} \int_r^\infty V_1(s) \left[-\left(V_1\left(\frac{t+r}{2}\right) \right) \left(\Phi\left(\frac{t+r}{2}\right) \Phi\left(\frac{t+r}{2} - s\right) - \Phi(s) \theta\left(s - \left(\frac{t+r}{2}\right) \right) \right) \right.$$
$$\left. + V_1\left(\frac{t-r}{2} + s\right) \Phi\left(s + \left(\frac{t-r}{2}\right) \right) \right]$$
$$- \tfrac{1}{4} \int_r^\infty ds \int_s^\infty du \lambda_1 V_1(s) \left[\sum_i \lambda_i V_i(u) \left(K_i(u, t+u-r) - K_i(u, t+2s-u-r) \right.\right.$$
$$\times\, \theta\left(s + \frac{t-r}{2} - u \right) - \theta\left(\frac{t+r}{2} - s\right) \left(K_i(u, t+r+u-2s) \right.$$
$$\left.\left. \left. - \theta\left(\frac{t+r}{2} - u\right) K_i(u, t+r-u) \right) \right) \right]. \quad \text{(D6)}$$

From (D3)–(D6) for $t = r$ we get

$$K_2(r, r) = \frac{\lambda_1}{2}\int_r^\infty V_1(s)\Phi(s)\,ds = \sin\left(\frac{\lambda_1}{2}\int_r^\infty V_1(u)\,du\right),$$

$$K_1(r, r) = \lambda_1 \frac{V_1}{4} K_2 + \frac{1}{2}\int_r^\infty ds\Phi\left[\lambda_2 V_2 + \frac{\lambda_1^2}{4} V_1^2\right] - \frac{1}{4}\int_r^\infty ds\lambda_1 V_1(s)\int_s^\infty \sum_i \lambda_i V_i(u) K_i(u, u)\,du. \tag{D7}$$

Let us define $K^{\sim} = K_1 - \frac{1}{4}(\lambda_1 V_1)K_2$:

$$K^{\sim}(r, r) = \frac{1}{2}\int_r^\infty \Phi\left[\lambda_2 V_2 + \frac{\lambda_1^2}{4} V_1\right] ds - \frac{\lambda_1}{4}\int_r^\infty V_1(s)\,ds \times \int_s^\infty \left[K_2\left(\lambda_2 V_2 + \frac{\lambda_1^2 V_1^2}{4}\right) + \lambda_1 V_1 K^{\sim}\right] du. \tag{D8}$$

Note that $|\Phi| < 1$ and $|K_2(r, r)| < 1$. From the conditions (4) about V_i there exist positive constants a_i, b_i, and c_i such that

$$|V_1(r)| < \frac{c_1 e^{-a_1 r}}{r^{b_1}}, \quad |V_2| < \frac{c_2 e^{-a_2 r}}{r^{b_2+1}}, \quad 0 < b_i < 1.$$

From (D7) and (D8) we get

$$|K_1(r, r)| < \frac{e^{-ar}}{r^b}\left[D_1 + D_2\int_r^\infty ds\int_s^\infty \frac{e^{-au}}{u^b}|K_1(u, u)|\,du\right],$$

$$|K^{\sim}(r, r)| < \frac{e^{-ar}}{r^b}\left[D_1^{\sim} + D_2^{\sim}\int_r^\infty ds\int_s^\infty \frac{e^{-au}}{u^b}|K^{\sim}(u, u)|\,du\right],$$

where $a > 0$ and $0 < b < 1$ are $|\lambda_i|$ independent but D_i and $D_i^{\sim}$ are constants greater than zero, $|\lambda_i|$ dependent but finite for $|\lambda_i|$ finite. It follows that

$$|K_1(r, r)| < \frac{D_1 e^{-ar}}{r^b}\exp D_2\int_r^\infty ds\int_s^\infty \frac{e^{-au}}{u^b}\,du < \frac{D_1 e^{-ar}}{r^b} D_3$$

and

$$\int_0^\infty |K_1(r, r)|\,dr < D_4.$$

We get also similar bounds for $|K^{\sim}|$ and $\int_0^\infty |K^{\sim}|\,dr$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Diffraction of Waves by a Conducting Cylinder Coated with a Moving Plasma Sheath*

C. YEH
Department of Engineering, University of California, Los Angeles, California

(Received 4 March 1969)

The scattering of plane electromagnetic waves by a perfectly conducting cylinder coated with a moving dielectric or plasma sheath is investigated theoretically. The homogeneous sheath is assumed to be moving in the axial direction with a uniform velocity v_z with respect to the conducting cylinder. Solutions of this problem are obtained by making use of the special theory of relativity, the covariance of Maxwell's equations, and the Lorentz transformations. Results are given in terms of the radiation patterns of the scattered fields. A rather unique feature concerning mode coupling between the incident wave and the scattered wave is found. Even at normal incidence for $v_z \neq 0$, an incident E wave or H wave will produce a scattered wave which contains both E and H waves. Detailed discussions are presented.

I. INTRODUCTION

In an attempt to understand the problem of the interaction of electromagnetic waves with moving penetrable medium, a great deal of work on the reflection and refraction of waves by various moving penetrable media has been carried out in recent years.[1-3] Many interesting and sometimes unexpected results are obtained. However, the problem of the diffraction of waves by a finite (resonant) size obstacle containing moving medium has not been considered. The purpose of this paper is to treat this problem. Specifically, the problem of the scattering of electromagnetic waves by a conducting cylinder coated with a dielectric or plasma sheath which is moving axially with a uniform velocity v_z is solved. This problem is not only of interest from a theoretical point of view but also has an important application, i.e., the understanding of the re-entry problem. It is well known that the plasma surrounding a re-entry vehicle streams pass the conducting core and that the vehicle is moving with respect to an observer.

Solutions of this problem are obtained by making use of the special theory of relativity, the covariance of Maxwell's equations, and the Lorentz transformations.[4] Several interesting features concerning the radiation patterns and the magnitude of the scattered waves as a function of the velocity of the moving medium are discussed.

II. FORMULATION OF THE PROBLEM

The geometry of this problem is shown in Fig. 1. It is assumed that an infinite, perfectly conducting cylinder of radius a, surrounded by a homogeneous moving plasma sheath of thickness $(b - a)$, is immersed in free-space (ϵ_0, μ_0). The plasma sheath is moving in the axial direction with respect to the conducting cylinder at a uniform velocity v_z. The incident wave in the free-space region is assumed to be plane with a harmonic time dependence. The case for an incident E wave is analyzed in detail.

In the observer's system S, which is stationary with respect to the conducting cylinder, the axial components of the incident plane wave in free space takes the form

$$E_z^{(i)} = E_0 \cos\theta_0 \exp(-ik_0 \cos\theta_0 y + ik_0 \sin\theta_0 z) \times \exp(-i\omega t), \quad (1)$$

$$H_z^{(i)} = 0, \quad (2)$$

where E_0 and ω are, respectively, the amplitude and the frequency of the incident wave and $k_0 = \omega(\mu_0\epsilon_0)^{\frac{1}{2}}$. θ_0 is the angle between the propagation vector and the positive y axis in the y–z plane.

In the moving system S', which is stationary with respect to the uniformly moving plasma sheath, the

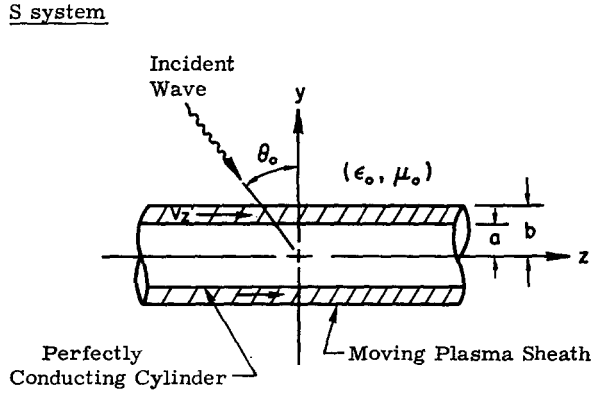


FIG. 1. The geometry of the problem.

* Supported by the National Science Foundation.

[1] H. Fujioka, F. Nihei, and N. Kumagai, J. Appl. Phys. **39**, 2161 (1968).
[2] C. Yeh, J. Appl. Phys. **38**, 5194 (1967).
[3] V. P. Pyati, J. Appl. Phys. **38**, 652 (1967).
[4] C. Møller, *The Theory of Relativity* (Oxford University Press, London, 1957).

incident plane wave takes the form

$$E_z^{(i)'} = E_0' \cos\theta' \exp(-ik_0' \cos\theta' y' + ik_0' \sin\theta' z') \times \exp(-i\omega' t')$$
$$= F_E' \sum_{n=-\infty}^{\infty} (-1)^n J_n(k_0' r' \cos\theta') e^{in\phi'}, \quad (3)$$

$$H_z^{(i)'} = 0, \quad (4)$$

where

$$\omega' = \gamma_z \omega(1 - \beta_z \sin\theta_0), \quad (5a)$$

$$\gamma_z = (1 - \beta_z^2)^{-\frac{1}{2}}, \quad \beta_z = v_z/c,$$

c = speed of light in vacuum,

$$k_0' \cos\theta' = k_0 \cos\theta_0, \quad (5b)$$

$$\sin\theta' = (\sin\theta_0 - \beta_z)/(1 - \beta_z \sin\theta_0), \quad (5c)$$

$$k_0' = \omega'(\mu_0\epsilon_0)^{\frac{1}{2}} = \gamma_z k_0(1 - \beta_z \sin\theta_0), \quad (5d)$$

$$E_0' = \gamma_z E_0(1 - \beta_z \sin\theta_0), \quad (5e)$$

$$F_E' = E_0' \cos\theta' \exp(ik_0' \sin\theta' z' - i\omega' t')$$
$$= E_0 \cos\theta_0 \exp(ik_0 \sin\theta_0 z - i\omega t). \quad (5f)$$

The above expansions are obtained by making use of the principle of phase invariance of plane waves, the Lorentz transformations, and the covariance of Maxwell's equations. $J_n(p)$ is the Bessel function of order n and argument p. A polar coordinate system (r', ϕ', z') is introduced. The scattered wave and the penetrated wave in the sheath must have the form[5]

$$E_{z'}^{(s)'} = F_E' \sum_{n=-\infty}^{\infty} (-1)^n A_n' H_n^{(1)}(k_0' r' \cos\theta') e^{in\phi'}, \quad (6)$$

$$H_{z'}^{(s)'} = F_E' \sum_{n=-\infty}^{\infty} (-1)^n B_n' i(\epsilon_0/\mu_0)^{\frac{1}{2}} H_n^{(1)}(k_0' r' \cos\theta') e^{in\phi'} \quad (7)$$

and

$$E_{z'}^{(p)'} = F_E' \sum_{n=-\infty}^{\infty} (-1)^n C_n' P_n(\lambda' r') e^{in\phi'}, \quad (8)$$

$$H_{z'}^{(p)'} = F_E' \sum_{n=-\infty}^{\infty} (-1)^n D_n' i(\epsilon_0/\mu_0)^{\frac{1}{2}} Q_n(\lambda' r') e^{in\phi'}, \quad (9)$$

respectively, with

$$P_n(\lambda' r') = J_n(\lambda' r') - [J_n(\lambda' a)/N_n(\lambda' a)] N_n(\lambda, r'), \quad (10)$$

$$Q_n(\lambda' r') = J_n(\lambda' r') - \left(\frac{dJ_n(\lambda' a)}{d(\lambda' a)} \Big/ \frac{dN_n(\lambda' a)}{d(\lambda' a)}\right) N_n(\lambda' r'), \quad (11)$$

$$\lambda' = k_0'(\mu_1\epsilon_1/\mu_0\epsilon_0 - \sin^2\theta')^{\frac{1}{2}}, \quad (12)$$

[5] J. R. Wait, *Electromagnetic Radiation from Cylindrical Structures* (Pergamon Press, Inc., New York, 1959).

where (μ_1, ϵ_1) characterizes the electromagnetic property of the sheath in the S' system and $H_n^{(1)}(k_0' r' \cos\theta')$ is the Hankel function. A_n', B_n', C_n', and D_n' are as yet unknown arbitrary constants to be determined according to the appropriate boundary conditions.

III. FORMAL SOLUTIONS

Satisfying the boundary conditions in the S' system, which requires the continuity of the tangential electric and magnetic fields at the boundary surface $r' = b$, gives the following equation from which the unknown coefficients A_n', B_n', C_n', and D_n' can be obtained:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} A_n' \\ B_n' \\ C_n' \\ D_n' \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}, \quad (13)$$

where

$$\begin{aligned}
a_{11} &= H_n^{(1)}(k_0' b \cos\theta'), \\
a_{12} &= 0, \\
a_{13} &= -P_n(\lambda' b), \\
a_{14} &= 0, \\
a_{21} &= 0, \\
a_{22} &= H_n^{(1)}(k_0' b \cos\theta'), \\
a_{23} &= 0, \\
a_{24} &= -Q_n(\lambda' b), \\
a_{31} &= -n \sin\theta' H_n^{(1)}(k_0' b \cos\theta'), \\
a_{32} &= k_0' b \cos\theta' H_n^{(1)'}(k_0' b \cos\theta'), \\
a_{33} &= (k_0'/\lambda')^2 \cos^2\theta' n P_n(\lambda' b) \sin\theta', \\
a_{34} &= -(k_0'/\lambda')^2 \cos^2\theta' \lambda' b Q_n(\lambda' b)(\mu_1/\mu_0), \\
a_{41} &= k_0' b \cos\theta' H_n^{(1)'}(k_0' b \cos\theta'), \\
a_{42} &= -\sin\theta' n H_n^{(1)}(k_0' b \cos\theta'), \\
a_{43} &= -(k_0'/\lambda')^2 \cos^2\theta'(\epsilon_1/\epsilon_0)\lambda' b P_n'(\lambda' b), \\
a_{44} &= (k_0'/\lambda')^2 \cos^2\theta' \sin\theta' n Q_n(\lambda' b),
\end{aligned} \quad (14)$$

and

$$\begin{aligned}
b_1 &= -J_n(k_0' b \cos\theta'), \\
b_2 &= 0, \\
b_3 &= \sin\theta' n J_n(k_0' b \cos\theta'), \\
b_4 &= -k_0' b \cos\theta' J_n'(k_0' b \cos\theta').
\end{aligned}$$

This is the formal solution for the problem of the scattering of a stationary dielectric coated cylinder by an obliquely incident plane E wave in the S' system. It is noted that the scattered wave as well as the penetrated wave contain both E and H waves, although only an E wave is incident upon the coated cylinder. If the incident wave is an H wave, the above results are still applicable provided that we replace

E' by H' and H' by $-E'$, ϵ by μ and μ by ϵ, Q_n by P_n and P_n by Q_n, throughout.

In the observer's system S, the field components of the scattered wave are

$$E_z^{(s)} = E_{z'}^{(s)'}, \tag{15}$$

$$H_z^{(s)} = H_{z'}^{(s)'}, \tag{16}$$

$$E_\phi^{(s)} = \gamma_z(E_{\phi'}^{(s)'} - v_z\mu_0 H_{r'}^{(s)'}), \tag{17}$$

$$H_\phi^{(s)} = \gamma_z(H_{\phi'}^{(s')} + v_z\epsilon_0 E_{r'}^{(s)'}), \tag{18}$$

$$E_r^{(s)} = \gamma_z(E_{r'}^{(s)'} + v_z\epsilon_0 H_{\phi'}^{(s)'}), \tag{19}$$

$$H_r^{(s)} = \gamma_z(H_{r'}^{(s)'} - v_z\mu_0 E_{\phi'}^{(s)'}). \tag{20}$$

Upon inspection of the above expressions, one notes that, even at normal incidence ($\theta_0 = 0°$) when $v_z \neq 0$, an incident E or H wave will produce a scattered wave which contains both E and H waves. This is a rather unique feature concerning the coupling between the incident wave and the scattered wave, which is only present when the sheath is moving. It is also worthwhile to point out that the above results are equally valid when the perfectly conducting center core is moving with respect to the plasma sheath or when the perfectly conducting center core is stationary with respect to the plasma sheath. This is because the boundary conditions remain unchanged and time independent whether the perfectly conducting cylinder is moving or not, so long as the movement is parallel to the interface.

At large distances from the cylinder, the asymptotic expression for the Hankel function

$$H_n^{(1)}(k_0 r\cos\theta_0) \to \left(\frac{2}{\pi k_0 r\cos\theta_0}\right)^{\frac{1}{2}} e^{i[k_0 r\cos\theta_0 - \frac{1}{4}(2n+1)\pi]}$$

is applicable provided that $k_0 r\cos\theta_0 \gg 1$ and $k_0 r\cos\theta_0 \gg n$. Using the above equation, we obtain the following expressions for the far-zone scattered fields in the S system:

for incident E wave,

$$\left|\frac{E_z^{(s)}}{E_0}\right|_{E\text{ wave}} \sim \left|\sum_{n=-\infty}^{\infty}(-1)^n A_n' e^{in(\phi-\frac{1}{2}\pi)}\right|, \tag{21}$$

$$\left|\frac{H_z^{(s)}}{E_0(\epsilon_0/\mu_0)^{\frac{1}{2}}}\right|_{E\text{ wave}} \sim \left|\sum_{n=-\infty}^{\infty}(-1)^n B_n' e^{in(\phi-\frac{1}{2}\pi)}\right|; \tag{22}$$

for incident H wave,

$$\left|\frac{H_z^{(s)}}{H_0}\right|_{H\text{ wave}} \sim \left|\sum_{n=-\infty}^{\infty}(-1)^n A_n' e^{in(\phi-\frac{1}{2}\pi)}\right|, \tag{23}$$

$$\left|\frac{E_z^{(s)}}{H_0(\mu_0/\epsilon_0)^{\frac{1}{2}}}\right|_{H\text{ ewav}} \sim \left|\sum_{n=-\infty}^{\infty}(-1)^n B_n' e^{in(\phi-\frac{1}{2}\pi)}\right|. \tag{24}$$

IV. DISCUSSION OF THE RESULTS

To have a qualitative idea of how the scattered fields behave as a function of the velocity of the moving medium, numerical computations are carried out for the moving-plasma-sheath case. The permittivity and permeability of a cold plasma medium in the S' system are, respectively,

$$\epsilon_1/\epsilon_0 = 1 - \omega_p^2/\omega'^2,$$
$$\mu_1/\mu_0 = 1. \tag{25}$$

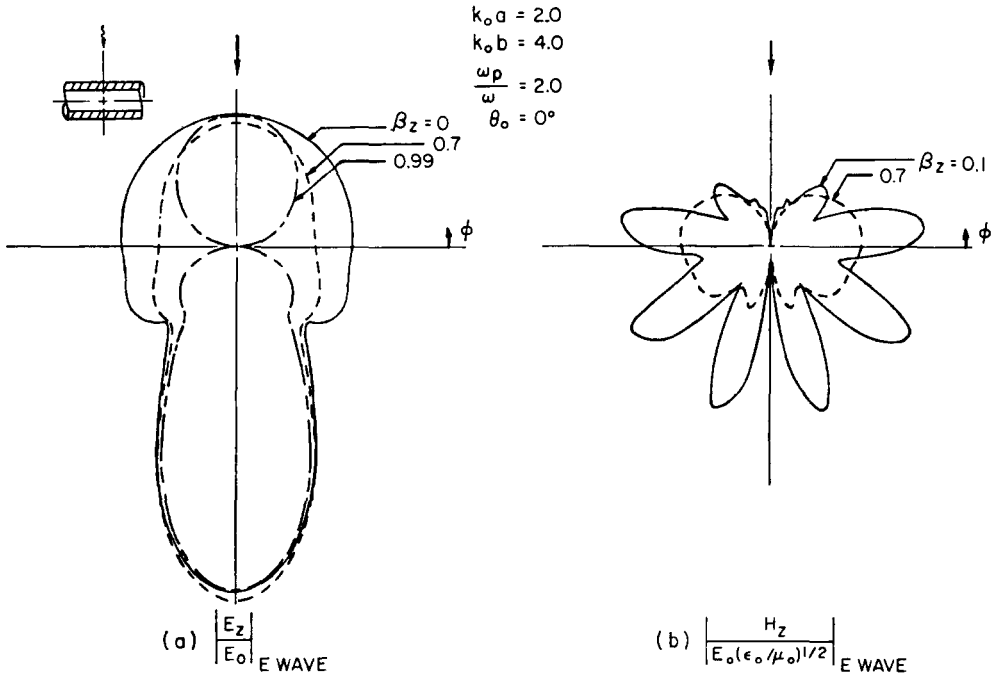


FIG. 2. Radiation patterns of the scattered waves for an incident E wave with $\theta_0 = 0°$.

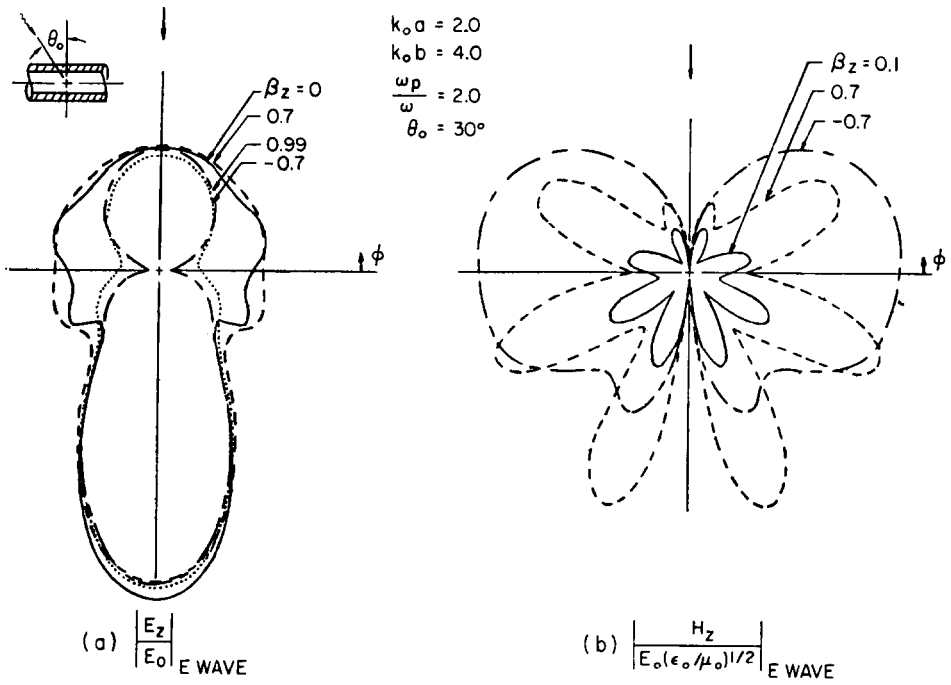


FIG. 3. Radiation patterns of the scattered waves for an incident E wave with $\theta_0 = 30°$.

In the S system, they are

$$\epsilon_1/\epsilon_0 = 1 - (\omega_p^2/\omega^2)[\gamma_z^2(1 - \beta_z \sin\theta_0)^2]^{-1},$$
$$\mu_1/\mu_0 = 1, \tag{26}$$

respectively. Substituting these expressions into Eq. (12) gives

$$\lambda' = k_0(\cos^2\theta_0 - \omega_p^2/\omega^2)^{\frac{1}{2}}, \tag{27}$$

which is independent of the movement of the sheath. Radiation patterns of the scattered waves [i.e., Eqs. (21)–(24)] are obtained for various values of θ_0 and v_z/c with $k_0a = 2.0$, $k_0b = 4.0$, and $\omega_p/\omega = 2.0$. In Figs. 2 and 3, the radiation patterns are plotted for various values of β_z and θ_0. Two angles of incidence, $\theta_0 = 0°$, 30°, are considered. It is noted that only representative patterns were shown in these figures. As can be seen from Fig. 2(a), the forward main lobe for the radiation patterns of $|E_z/E_0|_{E\ \text{wave}}$ remain relatively unchanged as $|\beta_z|$ increases. On the other hand, $|E_z/E_0|_{E\ \text{wave}}$ changes quite significantly in other directions; as β_z increases from 0, nulls appear in the $\phi = 0°$ and 180° directions. The fact that the movement of the sheath introduces coupling between an incident E wave with the scattered H wave even at normal incidence can best be seen from Fig. 2(b). As β_z increases from 0, a multilobe radiation pattern for $|H_z/E_0(\epsilon_0/\mu_0)^{\frac{1}{2}}|$ is produced; for higher values of β_z, the radiation pattern becomes basically a two-lobe structure. For all values of β_z, there exist two nulls in the forward and backward directions for

$$|H_z/E_0(\epsilon_0/\mu_0)^{\frac{1}{2}}|_{E\ \text{wave}},$$

while two main lobes exist for $|E_z/E_0|_{E\ \text{wave}}$ in the forward and backward directions. Similar radiation patterns are obtained for the $\theta_0 = 30°$ case as shown in Fig. 3. The general behavior of these patterns as a function of β is very similar to that for the $\theta_0 = 0°$ case. The only major difference is that at $\beta_z = 0$ the pattern for $|H_z/E_0(\epsilon_0/\mu_0)^{\frac{1}{2}}|_{E\ \text{wave}}$ is not zero.

Not only are significant variations for the radiation

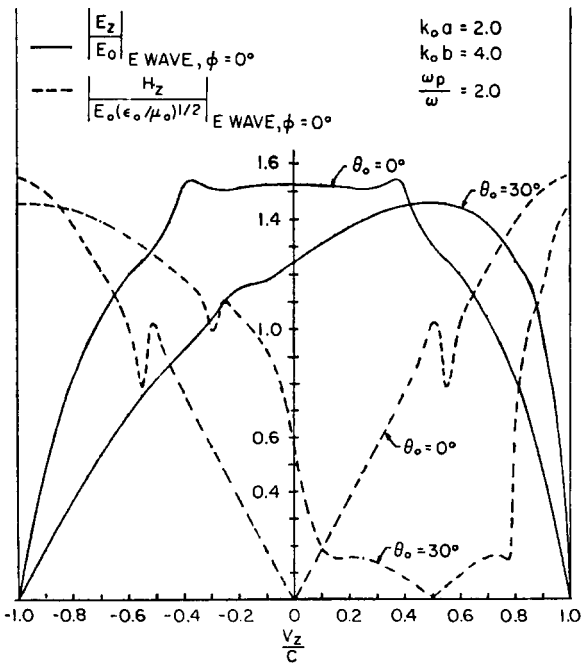


FIG. 4. Magnitude of the scattered waves as a function of the velocity of the plasma sheath for an incident E wave.

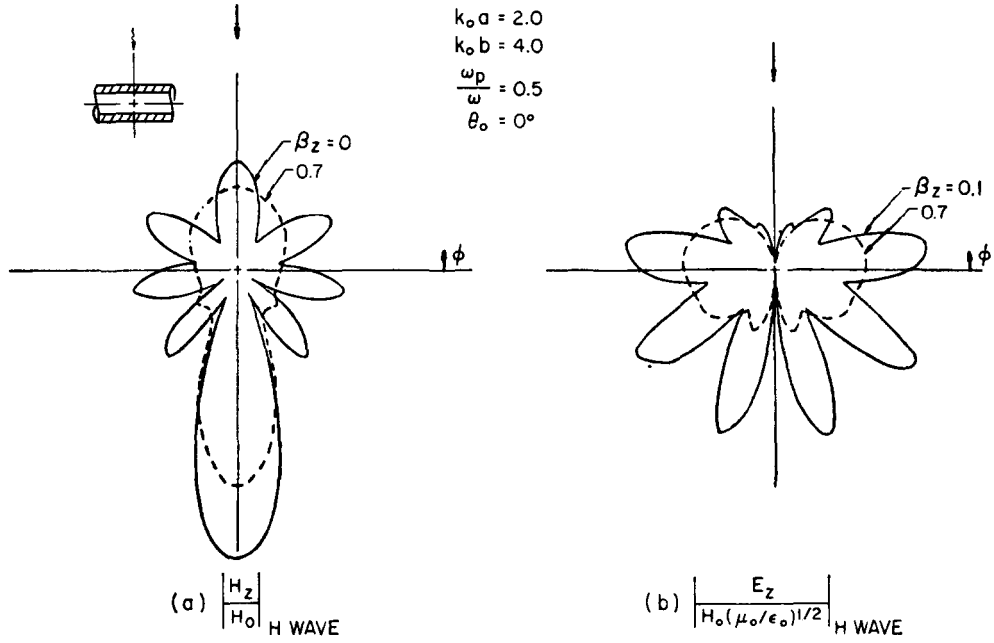


FIG. 5. Radiation patterns of the scattered waves for an incident H wave with $\theta_0 = 0°$.

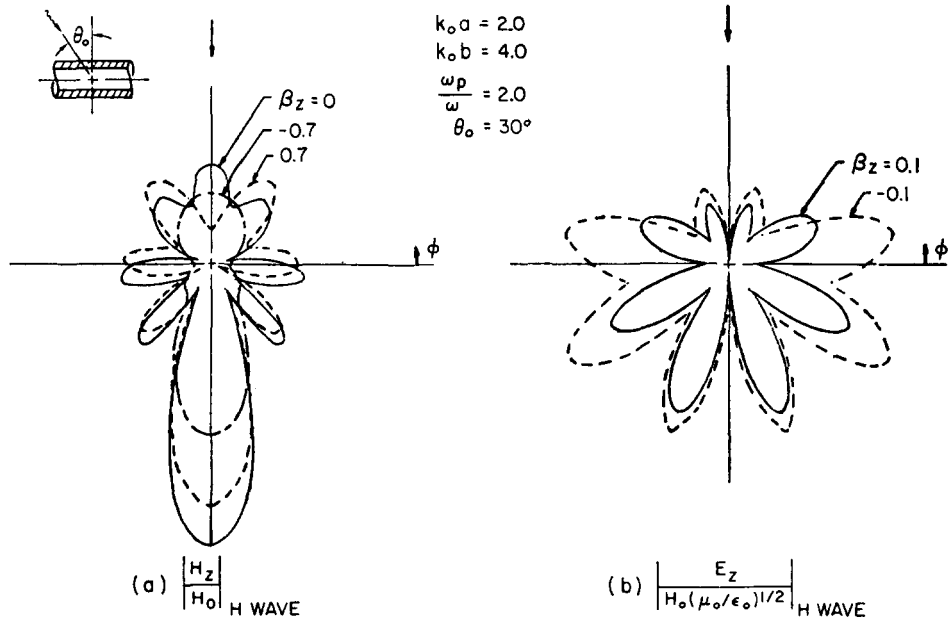


FIG. 6. Radiation patterns of the scattered waves for an incident H wave with $\theta_0 = 30°$.

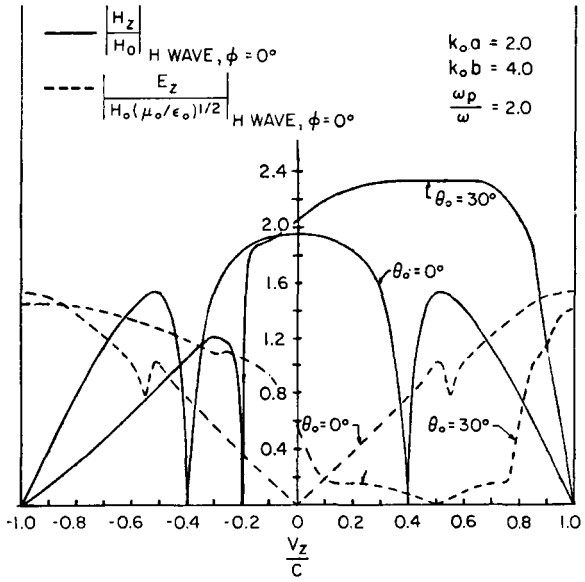


FIG. 7. Magnitude of the scattered waves as a function of the velocity of the plasma sheath for an incident H wave.

patterns observed for the scattered wave as β_z and θ_0 vary, but also for the magnitude of the scattered wave. Figure 4 is introduced to indicate the variation of $|E_z/E_0|_{E \text{ wave}}$ and $|H_z/E_0(\epsilon_0/\mu_0)^{\frac{1}{2}}|_{E \text{ wave}}$ at $\phi = 0°$ as a function of the velocity of the moving sheath. As $|v_z| \to c$, the magnitude of the scattered E wave at $\phi = 0°$ approaches zero and the magnitude of the scattered H wave at $\phi = 0°$ approaches a certain constant value. So the mode-coupling phenomenon appears to be most prominent at $\phi = 0°$ as $|\beta_z| \to 1$. As expected, the magnitude as well as the radiation patterns for the scattered wave are not symmetric with respect to β_z for $\theta_0 \neq 0°$.

Similar computations were carried out for an incident H wave. Results are shown in Figs. 5–7. It appears that the radiation patterns for an incident H wave are affected more dominantly by the movement of the sheath than those for an incident E wave. Again, coupling exists between an incident H wave with the scattered E wave at normal incidence. Computations were also carried out for $\omega_p/\omega = 0.5$, the under-dense case. Similar results as those discussed above were obtained. Since the plasma medium at $\omega_p/\omega = 0.5$ is rather transparent to the incident wave, the scattered fields $|E_z/E_0|_{E \text{ wave}}$ or $|H_z/H_0|_{H \text{ wave}}$ are not very sensitive to the movement of the plasma medium. Hence, the results for the $\omega_p/\omega = 0.5$ case are not included here.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Symmetrized Tensor Geometry

DONALD R. TOMPKINS
Department of Physics, University of Georgia, Athens, Georgia 30601

(Received 25 November 1968)

For Weyl symmetrized tensors, the inner product structure has previously been proven to be identical within equivalent multiplets. Here the inner product structure *between* a set of equivalent multiplets is proven to be essentially the same as that *within* the equivalent multiplets. This property is important because it separates the problem of orthogonalization within equivalent multiplets from the problem of orthogonalization between equivalent multiplets. Thus, full orthogonalization can be achieved by, first, identically orthogonalizing each of a set of equivalent multiplets and, then, recoupling the equivalent multiplets with coefficients which do not depend on individual vectors within the multiplets. These statements apply to symmetrized tensor multiplets of both the permutation group and an underlying group (such as U_n).

1. INTRODUCTION

The invariant subspaces (e.g., invariant under U_n) of a symmetrized tensor are composed of a direct sum of multiplets (of the index permutation group S_r for a tensor of rank r) belonging to distinct sets of index values. If the underlying space is orthogonal, then multiplets belonging to distinct sets of index values are mutually orthogonal. However, if Young symmetrizers[1] are used (Weyl symmetrization[2]), then not only are the vectors within each multiplet of S_r nonorthogonal, but also the bases of equivalent representations, belonging to a common set of index values, are not mutually orthogonal. This has long been considered to be a fundamental defect of Weyl symmetrized tensors.[3] Recently, it was shown that the inner product structures (IPS) within bases of equivalent representations obtained by Weyl symmetrization were identical, and it was asserted that the IPS between independent bases of equivalent representations were essentially the same (except for a constant factor) as those common to the bases themselves.[4] The author has recently discovered the proof of this last assertion to be circular, and the purpose here is to present an actual proof. To make this paper self-contained, a few results in Ref. 4 are repeated.

2. INNER-PRODUCT STRUCTURES

Bases of equivalent representations of S_r can be written[4] as

$$\begin{aligned}
\{A_i^\mu &\equiv (PQ)_i^\mu S_{i1} T_{i_1 \cdots i_r} = S_{ia}(PQ)_a^\mu S_{a1} T_{i_1 \cdots i_r}\},\\
\{B_i^\mu &\equiv (PQ)_i^\mu S_{i2} T_{i_1 \cdots i_r} = S_{ia}(PQ)_a^\mu S_{a2} T_{i_1 \cdots i_r}\},\\
&\cdots\\
\{D_i^\mu &\equiv (PQ)_i^\mu S_{im} T_{i_1 \cdots i_r} = S_{ia}(PQ)_a^\mu S_{am} T_{i_1 \cdots i_r}\},
\end{aligned} \tag{1}$$

where PQ denotes a Young symmetrizer and

$$(PQ)_i^\mu = S_{ik}(PQ)_k^\mu S_{ki}.$$

At present we will restrict our attention to a set of symmetrized Cartesian tensors with distinct index values. Young symmetrization will be assumed for all bases of S_r, and we shall use $\{A\}$ and $\{B\}$ as models for a typical pair of bases of equivalent representations. The full Cartesian tensor space (not restricted to distinct index values) is assumed to have an inner product for which it is orthonormal, i.e.,

$$\langle T_{i_1 \cdots i_r} \mid T_{j_1 \cdots j_r} \rangle = \delta_{i_1 j_1} \cdots \delta_{i_r j_r}.$$

Lemma 1: All bases of equivalent representations of S_r have identical IPS.

Proof: The right column of Eqs. (1) shows that bases of equivalent representations differ only by a common permutation of indices as a function of symmetrized tensor position. Since all index positions in a Cartesian tensor are *similar*, then the bases of equivalent representations are structurally *similar* and thus have identical IPS.

Lemma 2: For any pair of bases of equivalent representations,

$$\langle A_i \mid B_j \rangle = \langle B_i \mid A_j \rangle. \tag{2}$$

Proof: Define

$$\alpha_i \equiv A_i + B_i,$$
$$\beta_i \equiv A_i - B_i.$$

[1] (a) For a discussion of tensor spaces and Young tableau lemmas, see H. Boerner, *Representations of Groups* (North-Holland Publ. Co., Amsterdam, 1963). (b) For a general discussion of representations of the permutation group, see D. E. Rutherford, *Substitutional Analysis* (Edinburgh University Press, Edinburgh, Scotland, 1948).
[2] H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, N.J., 1946).
[3] G. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[4] D. R. Tompkins, J. Math. Phys. **8**, 1502 (1967).

Since $\{\alpha\}$ and $\{\beta\}$ are identical linear forms in

$$\{T_{i_1\cdots i_r} + T_{i_{p1}\cdots i_{pr}}\} \quad \text{and} \quad \{T_{i_1\cdots i_r} - T_{i_{p1}\cdots i_{pr}}\},$$

respectively,[5] and are mutually orthogonal, then for all i and j,

$$\langle \alpha_i \mid \beta_j \rangle = 0.$$

Thus,

$$\langle \alpha_i + \beta_i \mid \alpha_j - \beta_j \rangle = \langle \alpha_i \mid \alpha_j \rangle - \langle \beta_i \mid \beta_j \rangle,$$

which becomes

$$\langle A_i \mid B_j \rangle = \langle B_i \mid A_j \rangle.$$

Lemma 3: For any pair of equivalent representation bases,

$$2\langle A_i \mid B_j \rangle + 2\langle B_i \mid B_j \rangle = \langle \alpha_i \mid \alpha_j \rangle \tag{3}$$

and

$$-2\langle B_i \mid A_j \rangle + 2\langle A_i \mid A_j \rangle = \langle \beta_i \mid \beta_j \rangle. \tag{4}$$

Proof: Simply write out

$$\langle \alpha_i \mid \alpha_j - \beta_j \rangle = \langle \alpha_i \mid \alpha_j \rangle$$

and

$$\langle \beta_i \mid \beta_j - \alpha_j \rangle = \langle \beta_i \mid \beta_j \rangle$$

and use Lemma 1.

Lemma 4: The matrices $\langle \alpha_i \mid \alpha_j \rangle$ and $\langle \beta_i \mid \beta_j \rangle$ can be diagonalized simultaneously by transforming $\{A\}$ and $\{B\}$ identically.

Proof: By their construction, if $\langle \alpha_i \mid \alpha_j \rangle$ and $\langle \beta_i \mid \beta_j \rangle$ are diagonalized by a single transformation, then the transformation must transform $\{A\}$ and $\{B\}$ identically. Because the Young symmetrizers are real and the underlying basis is orthogonal, then $\langle A_i \mid A_j \rangle$ and $\langle B_i \mid B_j \rangle$ are real symmetric forms. By Lemma 2, $\langle A_i \mid B_j \rangle$ is also (real) symmetric and, hence, $\langle \alpha_i \mid \alpha_j \rangle$ and $\langle \beta_i \mid \beta_j \rangle$ are real symmetric. Writing

$$\alpha'_i \equiv A'_i + B'_i,$$
$$\beta'_i \equiv A'_i - B'_i,$$

where $\{A'\}$ and $\{B'\}$ are obtained by identical diagonalizations (to unit matrices) of $\{A\}$ and $\{B\}$, respectively, we have

$$\langle \alpha'_i \mid \alpha'_j \rangle = 2\langle A'_i \mid B'_j \rangle + 2\langle B'_i \mid B'_j \rangle,$$
$$\langle \beta'_i \mid \beta'_j \rangle = -2\langle A'_i \mid B'_j \rangle + 2\langle A'_i \mid A'_j \rangle,$$

where Lemmas 1 and 2 were used. (Clearly these lemmas apply, as the diagonalizations of $\{A\}$ and $\{B\}$ were identical.) Now, we need only to ask if matrices $\langle \alpha'_i \mid \alpha'_j \rangle$ and $\langle \beta'_i \mid \beta'_j \rangle$ commute, i.e.,

$$\sum_j \langle \alpha'_i \mid \alpha'_j \rangle \langle \beta'_j \mid \beta'_k \rangle \stackrel{?}{=} \sum_j \langle \beta'_i \mid \beta'_j \rangle \langle \alpha'_j \mid \alpha'_k \rangle.$$

[5] $\mathfrak{P} T_{i_1\cdots i_r} \equiv T_{i_{p1}\cdots i_{pr}}$ where $\mathfrak{P}$ is the permutation of indices (as a function of position) relating equivalent bases.

Using Lemmas 1 and 2, this becomes the question

$$\sum_j [\langle A'_i \mid B'_j \rangle + \langle B'_i \mid B'_j \rangle][-\langle A'_j \mid B'_k \rangle + \langle A'_j \mid A'_k \rangle]$$
$$\stackrel{?}{=} \sum_j [-\langle A'_i \mid B'_j \rangle + \langle A'_i \mid A'_j \rangle][\langle A'_j \mid B'_k \rangle + \langle B'_j \mid B'_k \rangle].$$

Canceling obviously identical terms gives

$$\sum_j \langle A'_i \mid B'_j \rangle \langle A'_j \mid A'_k \rangle - \sum_j \langle B'_i \mid B'_j \rangle \langle A'_j \mid B'_k \rangle$$
$$\stackrel{?}{=} -\sum_j \langle A'_i \mid B'_j \rangle \langle B'_j \mid B'_k \rangle + \sum_j \langle A'_i \mid A'_j \rangle \langle A'_j \mid B'_k \rangle.$$

By Lemma 1, the left side of this question can be written as

$$\sum_j \langle A'_i \mid B'_j \rangle \langle B'_j \mid B'_k \rangle - \sum_j \langle A'_i \mid A'_j \rangle \langle A'_j \mid B'_k \rangle,$$

which becomes

$$\langle A'_i \mid B'_k \rangle - \langle A'_i \mid B'_k \rangle \equiv 0,$$

because

$$\sum_j |B'_j\rangle \langle B'_j \mid B'_k \rangle = |B'_k\rangle$$

and

$$\sum_j \langle A'_i \mid A'_j \rangle \langle A'_j| = \langle A'_i|$$

when bases $\{A'\}$ and $\{B'\}$ are simultaneously normalized. The right side of the above question similarly vanishes to establish commutation. Since commutation is a sufficient condition for simultaneous diagonalization of real symmetric matrices, the lemma is proven.

Lemma 5: All vectors of equivalent representation bases are of equal length.

Proof: The left column of Eqs. (1) shows that for each vector of a basis, the relationship between index values in the Young pattern and in the Cartesian tensor is the same. Since all Cartesian tensor index positions are similar, the vectors do not differ in a way affecting length. Since equivalent representation bases have identical IPS (Lemma 1), then in fact *all* vectors in Eqs. (1) have equal lengths.

Lemma 6: All vectors of $\{\alpha\}$ are of equal length as also are all vectors of $\{\beta\}$.

Proof: For $i = 1, 2, \cdots, m$, the quantities

$$\langle A_i \mid B_i \rangle = \langle S_{ia}(PQ)^{\mu}_{a} S_{a1} T_{i_1\cdots i_r} \mid S_{ia}(PQ)^{\mu}_{a} S_{a2} T_{i_1\cdots i_r} \rangle$$

do not differ in the relationship between tableau and Cartesian tensor index values (i.e., index values are only rearranged in a fixed tableau–tensor mosaic). Thus, the values of these quantities are independent of i. This plus Lemma 5 completes the proof.

Lemma 7: The IPS of $\{\alpha\}$ is essentially the same (i.e., it differs by, at most, a common factor) as that of $\{\beta\}$.

Proof: The matrices $\langle \alpha_i \mid \alpha_j \rangle$ and $\langle \beta_i \mid \beta_j \rangle$ can be diagonalized to yield $\langle \alpha_i' \mid \alpha_j' \rangle$ and $\langle \beta_i' \mid \beta_j' \rangle$, respectively. Using the orthogonality of the diagonal form bases, and considering only diagonal entries of $\langle \alpha_i \mid \alpha_j \rangle$ and $\langle \beta_i \mid \beta_j \rangle$, we have[6]

$$\langle \alpha_i \mid \alpha_i \rangle = \sum_j R_{ij}^{\mathrm{T}} \langle \alpha_j' \mid \alpha_j' \rangle R_{ji}, \tag{5}$$

$$\langle \beta_i \mid \beta_i \rangle = \sum_j R_{ij}^{\mathrm{T}} \langle \beta_j' \mid \beta_j' \rangle R_{ji}, \tag{6}$$

where R^{T} denotes the transpose of R.

These equations are linear in the diagonal matrix elements, but because the elements of R appear quadratically, we may not get an independent equation for each value of i. Because the relationship between $\{\langle \alpha_i \mid \alpha_i \rangle\}$ and $\{\langle \alpha_j' \mid \alpha_j' \rangle\}$ is identical to that between $\{\langle \beta_i \mid \beta_i \rangle\}$ and $\{\langle \beta_j' \mid \beta_j' \rangle\}$, we need only show that there always exists a simultaneous diagonalization for which the amplitudes $\{\langle \alpha_i \mid \alpha_i \rangle\}(\{\langle \beta_i \mid \beta_i \rangle\})$ are independent functions of $\{\langle \alpha_j' \mid \alpha_j' \rangle\}(\{\langle \beta_j' \mid \beta_j' \rangle\})$. We call this the nondegenerate case. Degeneracy occurs between, for example, $\langle \alpha_a \mid \alpha_a \rangle$ and $\langle \alpha_b \mid \alpha_b \rangle$ when, for all j,

$$R_{aj}^{\mathrm{T}} R_{ja} = R_{bj}^{\mathrm{T}} R_{jb},$$

and is due to the equality of angles squared, i.e.,

$$(\angle \alpha_j', \alpha_a)^2 = (\angle \alpha_j', \alpha_b)^2.$$

In an m-dimensional space, for $\langle \alpha_a \mid \alpha_a \rangle$ and $\langle \alpha_b \mid \alpha_b \rangle$ to be degenerate, it is necessary that the orthogonal basis vectors $|\alpha_a\rangle$ and $|\alpha_b\rangle$ lie in the $(m-1)$-dimensional "symmetric" hyperplanes generated by the intersections among the family of all pairs of identical intersecting hypercones centered about α_a and α_b, $-\alpha_a$ and α_b, α_a and $-\alpha_b$, and $-\alpha_a$ and $-\alpha_b$, and having apexes at the origin.[7] Considering all pairs of nonorthogonal basis vectors, it is seen that there is at most $2[(m-1)+(m-2)+\cdots+1]$ distinct such hyperplanes. For an m-dimensional space, there does not exist any finite set of $(m-1)$-dimensional hyperplanes having the property that no orthogonal basis cannot lie completely outside of them.[8] Thus, there exist orthogonal basis vectors whose amplitude expressions are not degenerate relative to any pair of initial basis vectors. Hence, there exist simultaneous diagonalizations in which the amplitudes

$$\{\langle \alpha_i \mid \alpha_i \rangle\}(\{\langle \beta_i \mid \beta_i \rangle\})$$

are all independent functions of diagonal basis amplitudes.

We now select a simultaneous diagonalization for which $\{\langle \alpha_i \mid \alpha_i \rangle\}$ and $\{\langle \beta_i \mid \beta_i \rangle\}$ appear as independent functions of diagonal basis amplitudes. With R given and (by Lemma 6)

$$\langle \alpha_i \mid \alpha_i \rangle = \lambda,$$
$$\langle \beta_i \mid \beta_i \rangle = \mu,$$

then Eqs. (5) and (6) present m linearly independent equations for the m matrix elements $\{\langle \alpha_j' \mid \alpha_j' \rangle\}$ and m similar equations for the m matrix elements $\{\langle \beta_j' \mid \beta_j' \rangle\}$. The elements $\{\langle \alpha_j' \mid \alpha_j' \rangle\}$ and $\{\langle \beta_j' \mid \beta_j' \rangle\}$ are then uniquely determined, so we must have

$$\lambda/\mu = \langle \alpha_j' \mid \alpha_j' \rangle / \langle \beta_j' \mid \beta_j' \rangle.$$

Then, using the inverse transformation to recover $\langle \alpha_i \mid \alpha_j \rangle$ and $\langle \beta_i \mid \beta_j \rangle$ proves the lemma.

Corollary: The bases $\{\alpha'\}$ and $\{\beta'\}$ can be simultaneously brought to the forms

$$|\alpha_i'\rangle\langle\alpha_j'| = \lambda \delta_{ij}$$

and

$$|\beta_i'\rangle\langle\beta_j'| = \mu \delta_{ij}.$$

Theorem 1: The IPS of $\langle A_i \mid B_j \rangle$ is essentially the same as that of $\langle A_i \mid A_j \rangle(\langle B_i \mid B_j \rangle)$.

Proof: Adding (subtracting) Eqs. (3) and (4) shows that $\langle A_i \mid A_j \rangle$ and $\langle B_i \mid B_j \rangle(\langle A_i \mid B_j \rangle$ and $\langle B_i \mid A_j \rangle)$ have essentially the same IPS as that essentially common (by Lemma 7) to $\{\alpha\}$ and $\{\beta\}$. [It is assumed that equivalent representation bases are constructed with the state correspondence displayed in Eq. (1).]

We now consider the situation for tensors carrying a set of index values with repeated entries. With repeated index values, the basis $\{A\}$ becomes what is denoted as basis $\{\mathcal{A}\}$, etc. Let C denote an operation which yields symmetric subsets of indices by combining states of $\{A\}$. As far as inner-product structure is concerned, such symmetric subsets of indices act

[6] G. D. Mostow, J. H. Sampson, and J. P. Meyer, *Fundamental Structures of Algebra* (McGraw-Hill Book Co., New York, 1963), p. 449.

[7] An equivalent viewpoint is that, for degeneracy, it is necessary that all vectors of the orthogonal bases lie in the $m-1$ dimensional subspaces orthogonal to $-\alpha_a+\alpha_b$ and $\alpha_a+\alpha_b$.

[8] The $(m-1)$-dimensional hyperplanes orthogonal to each vector of an orthogonal basis define a set of m orthogonal basis hyperplanes. There exists an infinite number of orientations of any initial orthogonal basis yielding an infinite number of sets of m orthogonal basis hyperplanes. Because the set of "symmetric" hyperplanes is finite, there exist infinitely many sets of m orthogonal basis hyperplanes not containing any "symmetric" hyperplane. Suppose now we are in an orthogonal coordinate system whose set of m orthogonal basis hyperplanes contains no "symmetric" hyperplanes but that some vectors of the orthogonal basis lie in "symmetric" hyperplanes. Because the set of "symmetric" planes is finite and distinct from the orthogonal basis hyperplanes of this selected basis, we can perform an infinitesimal re-orientation to an orthogonal coordinate system for which no basis vectors lie in "symmetric" hyperplanes. This follows because the vectors lying in a finite set of $(m-1)$-dimensional hyperplanes are not dense on an m-dimensional space.

the same as repeated indices, so inner product results derived with such states will also apply to corresponding repeated index-value states. Because some states of $\{A\}$ may become identical or vanish under such an operation, C is generally rectangular.[9] If an operation F makes the same indices of $\{B\}$ symmetric, then because $\{B\}$ and $\{A\}$ carry identical representations of S_r, we see $F = C$. Writing matrix $\langle A_i \mid A_j\rangle$ as $\langle A \mid A\rangle$, etc., we have

$$C^{\mathrm{T}}\langle A \mid A\rangle C = \langle \mathcal{A} \mid \mathcal{A}\rangle,$$

$$C^{\mathrm{T}}\langle B \mid B\rangle C = \langle \mathcal{B} \mid \mathcal{B}\rangle,$$

$$C^{\mathrm{T}}\langle A \mid B\rangle C = \langle \mathcal{A} \mid \mathcal{B}\rangle,$$

$$C^{\mathrm{T}}\langle B \mid A\rangle C = \langle \mathcal{B} \mid \mathcal{A}\rangle,$$

and because

$$\langle A \mid A\rangle = \langle B \mid B\rangle,$$

we see that

$$\langle \mathcal{A} \mid \mathcal{A}\rangle = \langle \mathcal{B} \mid \mathcal{B}\rangle,$$

which shows that Lemma 1 remains valid with repeated indices.

If

$$\langle A \mid B\rangle = \lambda\langle A \mid A\rangle,$$

then

$$C^{\mathrm{T}}\langle A \mid B\rangle C = \lambda C^{\mathrm{T}}\langle A \mid A\rangle C = \lambda\langle \mathcal{A} \mid \mathcal{A}\rangle$$

and

$$C^{\mathrm{T}}\langle B \mid A\rangle C = \lambda C^{\mathrm{T}}\langle B \mid B\rangle C = \lambda\langle \mathcal{B} \mid \mathcal{B}\rangle,$$

so that

$$\langle \mathcal{A} \mid \mathcal{B}\rangle = \lambda\langle \mathcal{A} \mid \mathcal{A}\rangle = \lambda\langle \mathcal{B} \mid \mathcal{B}\rangle = \langle \mathcal{B} \mid \mathcal{A}\rangle,$$

which shows that Theorem 1 remains valid with repeated indices,[10] and also shows that the value of λ does not depend on the index complexion.[11] Equivalent representation bases of S_r belonging to different sets of index values are mutually orthogonal. Thus, we have proved the following:

Theorem 2: The IPS is identical for independent bases of equivalent symmetrized tensor representations of U_n, and the IPS between independent bases of equivalent symmetrized tensor representatives of U_n is essentially the same as that common to the equivalent representations themselves. [It is assumed that the equivalent multiplets of U_n are constructed with the state correspondence shown in Eqs. (1) and also with identical ordering of index-value set entries.]

Bases $\{A'\}$, $\{A''\}, \cdots$ constructed on Cartesian tensors $T'_{i_1\cdots i_r}$, $T''_{i_1\cdots i_r}, \cdots$ either coincide or else do not intersect. If $\{A'\}$ and $\{A''\}$ (each of dimension m) intersect but do not coincide, then $\{A', A''\}$ would span an invariant subspace of dimension greater than m but less than $2m$, which is impossible, because only equivalent representations of S_r can appear. With this start we can proceed, as shown in the note added in the proof in Ref. 4, to extend Theorem 2 to symmetrized non-Cartesian tensors.

3. DISCUSSION

Theorem 2 is useful because it shows that with a single transformation acting on *each* maximum multiplicity multiplet of S_r plus the "derived" transformations[12] acting on all lesser multiplicity multiplets, one can identically diagonalize all equivalent (equivalences due to S_r) symmetrized tensor multiplets of U_n, and such a set of transformations diagonalizes the IPS between the bases of equivalent representations with *all* resulting diagonals among equivalent representation bases having essentially the same structure. This shows that equivalent symmetrized tensor multiplets of U_n can be made mutually orthogonal by being recoupled with state-independent coefficients (hence recoupled without disturbing the IPS within each of the equivalent multiplets). This shows how to decouple the inner (interior to each multiplet) and outer (between equivalent multiplets) orthogonality problems.

---

[9] To illustrate the operation C, we consider

$$\{A\} \equiv \{(PQ)_{ab,c}T_{abc}, (PQ)_{ac,b}T_{acb}\}$$

and

$$C\{A\} \equiv A_1 + A_2 = (T_{abc} + T_{acb}) + (T_{bac} + T_{cab}) - (T_{cba} + T_{bca}) - (T_{cab} + T_{bac}).$$

Here $C = [1\ 1]$, and for the inner-product structure, the state $A_1 + A_2$ acts essentially the same as the state $[(PQ)_{ab,c}T_{abc}]_{b=c}$.

[10] If $\{A\}$ and $\{B\}$ are equivalent distinct index-value bases of S_r obtained with unitary representation Wigner symmetrizers (i.e., Wigner "projection" operators based on unitary representations of S_r, see Footnote 10 of Ref. 4), and $\{\mathcal{A}\}$ and $\{\mathcal{B}\}$ are the corresponding bases belonging to a set of index values with symmetric entries (simulating repeated entries), then the Wigner operator algebra imposes that

$$\langle A_i \mid A_j\rangle = \langle B_i \mid B_j\rangle = \delta_{ij},$$
$$\langle \mathcal{A}_a \mid \mathcal{A}_b\rangle = \langle \mathcal{B}_a \mid \mathcal{B}_b\rangle = \delta_{ab},$$

and that $\langle A_i \mid B_j\rangle$ and $\langle \mathcal{A}_a \mid \mathcal{B}_b\rangle$ are diagonal. If $|A\rangle C = |\mathcal{A}\rangle$ and $|B\rangle F = |\mathcal{B}\rangle$, then

$$C^{\mathrm{T}}\langle A \mid A\rangle C = \langle \mathcal{A} \mid \mathcal{A}\rangle,$$
$$F^{\mathrm{T}}\langle B \mid B\rangle F = \langle \mathcal{B} \mid \mathcal{B}\rangle.$$

If the underlying space is orthogonal, then

$$\langle A \mid B\rangle = 0,$$

and thus

$$C^{\mathrm{T}}\langle A \mid B\rangle F = \langle \mathcal{A} \mid \mathcal{B}\rangle = 0.$$

This easily gained result, plus the note added in proof of D. R. Tompkins, J. Math. Phys. **9**, 1 (1968), establishes the full orthogonality displayed in Eq. (3.2) of Ref. 4.

[11] By complexion we mean the essential partition of index values; e.g., *aabb*, *abab*, and *aacc* are all the same complexion.

[12] The derived transformations are what the transformation for distinct index-value multiplets become when repeated index values are entered.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Relativistic Continuum Theory for the Interaction of Electromagnetic Fields with Deformable Bodies*

R. A. GROT†
Technological Institute, Northwestern University, Evanston, Illinois 60201

(Received 25 April 1969)

A variational principle is formulated which yields the balance laws and constitutive equations of a nonconducting, charge-free elastic solid interacting with electromagnetic fields. It is found that the form of the total energy–momentum tensor and the constitutive equations that follow from a Lagrangian action which depends arbitrarily on the inverse deformation gradients and the electromagnetic field tensor are identical to those obtained by formulating a constitutive theory of a nondissipative material based on the basic mechanical, thermodynamical, and electromagnetic balance laws of a continuum.

1. INTRODUCTION

In a recent series of articles Grot and Eringen[1,2] have proposed a continuum theory of the interaction of electromagnetic fields with a deformable body. Their basic approach was to formulate the fundamental balance laws of a continuum in a relativistically invariant manner and to construct relativistically invariant constitutive equations. For the case of a nondissipative solid, one would expect that these equations are the result of a variational principle. In this article we present a slightly modified form of the theory presented by Grot and Eringen and then show that these equations follow from a variational principle. By a Legendre transformation we produce a Lagrangian action that is a relativistic generalization of the Lagrangian introduced by Toupin.[3,4]

Throughout this article gravitational effects are neglected; thus, we restrict our considerations to the special theory of relativity. The signature of the Lorentz metric $\gamma^{\alpha\beta}$ is chosen to be $(+++-)$, i.e., $\gamma^{11} = \gamma^{22} = \gamma^{33} = 1, \gamma^{44} = -1$, and all other $\gamma^{\alpha\beta} = 0$. The Greek subscripts and superscripts assume the values 1, 2, 3, 4, and are raised and lowered by the metric $\gamma^{\alpha\beta}$. The small italic subscripts and superscripts assume the values 1, 2, 3, and will always denote the spatial coordinates of the space–time of events. The large italic subscripts or superscripts assume the values 1, 2, 3, and denote the coordinates of the reference state. For convenience we also set the speed of light equal to unity, i.e., $c = 1$, and choose the dielectric constant of free space $\epsilon_0 = 1$; thus, the permeability of free space $\mu_0 = 1$. Parentheses around a set of indices denote symmetrization and brackets denote alternation. We define

$$\epsilon_{\alpha\beta\gamma\delta} \equiv e_{\alpha\beta\gamma\delta},$$

where $e_{\alpha\beta\gamma\delta}$ is the permutation symbol and define

$$\epsilon^{\alpha\beta\gamma\delta} = \gamma^{\alpha\alpha_1}\gamma^{\beta\beta_1}\gamma^{\gamma\gamma_1}\gamma^{\delta\delta_1}\epsilon_{\alpha_1\beta_1\gamma_1\delta_1} = -e^{\alpha\beta\gamma\delta}.$$

2. BASIC BALANCE LAWS

The basic continuum balance laws of a material body in the context of special relativity[1,2] consist of the conservation of particle number, balance of energy–momentum, balance of moment of energy–momentum, conservation of charge, conservation of magnetic flux, and Ampere's and Gauss's laws. We briefly list the differential form of these equations.

Conservation of Particle Number:

$$n^{\alpha}{}_{,\alpha} = 0, \tag{2.1}$$

where n^{α} has the form

$$n^{\alpha} = n_0 u^{\alpha}, \quad u^{\alpha}u_{\alpha} = -1, \tag{2.2}$$

and

$$n_0 = n(1 - \mathbf{v}^2)^{\frac{1}{2}}. \tag{2.3}$$

Here u^{α} denotes the world velocity, n the density of particles, and $\mathbf{v}$ the velocity vector.

Balance of Energy–Momentum:

$$T^{\alpha\beta}{}_{,\beta} = \pi^{\beta\gamma}\phi^{\alpha}{}_{\gamma,\beta} + \sigma^{\beta}\phi^{\alpha}{}_{\beta} + f^{\alpha}, \tag{2.4}$$

where $T^{\alpha\beta}$ is the energy–momentum tensor of the body, $\pi^{\beta\alpha}$ the polarization–magnetization tensor, $\phi_{\alpha\gamma}$ the magnetic flux tensor, and f^{α} the body-force four-vector.

Balance of Moment of Energy–Momentum:

$$T^{[\alpha\beta]} = \pi_{\gamma}{}^{[\alpha}\phi^{\beta]\gamma}. \tag{2.5}$$

Conservation of Charge:

$$\sigma^{\alpha}{}_{,\alpha} = 0, \tag{2.6}$$

where σ^{α} is the charge–current four-vector.

* This work was supported by the Advanced Research Project Agency of the Department of Defense through the Northwestern University Materials Research Center.

† Present address: Department of Aerospace and Mechanical Sciences, Princeton University, Princeton, N.J.

[1] R. A. Grot and A. C. Eringen, Intern. J. Engr. Sci. **4**, 611 (1966).
[2] R. A. Grot and A. C. Eringen, Intern. J. Engr. Sci. **4**, 638 (1966).
[3] R. A. Toupin, J. Ratl. Mech. Anal. **5**, 849 (1956).
[4] R. A. Toupin, Arch. Ratl. Mech. Anal. **5**, 440 (1960).

Conservation of Magnetic Flux:

$$\phi_{\alpha\gamma,\beta} + \phi_{\gamma\beta,\alpha} + \phi_{\beta\alpha,\gamma} = 0, \quad \phi_{\alpha\beta} = -\phi_{\beta\alpha}. \quad (2.7)$$

Ampere's and Gauss' Law:

$$\mathcal{G}^{\alpha\beta}{}_{,\beta} = \sigma^{\alpha}, \quad \mathcal{G}^{\alpha\beta} = -\mathcal{G}^{\beta\alpha}, \quad (2.8)$$

where

$$\mathcal{G}^{\alpha\beta} = \phi^{\alpha\beta} - \pi^{\alpha\beta}. \quad (2.9)$$

In terms of the classical electromagnetic fields,

$$\begin{aligned} \phi_{\alpha\beta} &= [\text{dual } \mathbf{B}, \mathbf{E}], \\ \mathcal{G}^{\alpha\beta} &= [\text{dual } \mathbf{H}, -\mathbf{D}], \\ \pi^{\alpha\beta} &= [\text{dual } (\mathbf{M} + \mathbf{v} \times \mathbf{P}), \mathbf{P}], \end{aligned} \quad (2.10)$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ the magnetic flux density, $\mathbf{H}$ the magnetic field, $\mathbf{D}$ the electric displacement, $\mathbf{P}$ the polarization, and $\mathbf{M}$ the magnetization. The forms (2.10) will not be needed in this article; they are listed for concreteness.

One can write[1,2] $T^{\alpha\beta}$, $\phi_{\alpha\beta}$, $\mathcal{G}^{\alpha\beta}$, and $\pi^{\alpha\beta}$ in the form

$$T^{\alpha\beta} = n_0 e u^{\alpha}u^{\beta} + u^{\alpha}q^{\beta} + p^{\alpha}u^{\beta} - t^{\alpha\beta}, \quad (2.11)$$

$$\phi_{\alpha\beta} = \mathcal{E}_{\beta}u_{\alpha} - \mathcal{E}_{\alpha}u_{\beta} + \epsilon_{\alpha\beta\gamma\delta}\mathcal{B}^{\gamma}u^{\delta}, \quad (2.12)$$

$$\mathcal{G}^{\alpha\beta} = \mathcal{D}^{\beta}u^{\alpha} - \mathcal{D}^{\alpha}u^{\beta} + \epsilon^{\alpha\beta\gamma\delta}\mathcal{H}_{\gamma}u_{\delta}, \quad (2.13)$$

$$\pi^{\alpha\beta} = \mathcal{P}^{\alpha}u^{\beta} - \mathcal{P}^{\beta}u^{\alpha} + \epsilon^{\alpha\beta\gamma\delta}\mathcal{M}_{\gamma}u_{\delta}, \quad (2.14)$$

where

$$n_0 e = T^{\alpha\beta}u_{\alpha}u_{\beta}, \quad (2.15)$$

$$q^{\beta} = -S^{\beta}{}_{\gamma}T^{\alpha\gamma}u_{\alpha}, \quad (2.16)$$

$$p^{\alpha} = -S^{\alpha}{}_{\gamma}T^{\gamma\beta}u_{\beta}, \quad (2.17)$$

$$t^{\alpha\beta} = -S^{\alpha}{}_{\gamma}S^{\beta}{}_{\delta}T^{\gamma\delta}, \quad (2.18)$$

$$\mathcal{E}_{\beta} = \phi_{\beta\alpha}u^{\alpha}, \quad (2.19)$$

$$\mathcal{B}^{\alpha} = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}\phi_{\beta\gamma}u_{\delta}, \quad (2.20)$$

$$\mathcal{D}^{\beta} = \mathcal{G}^{\beta\alpha}u_{\alpha}, \quad (2.21)$$

$$\mathcal{H}_{\alpha} = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}\mathcal{G}^{\beta\gamma}u^{\delta}, \quad (2.22)$$

$$\mathcal{D}^{\alpha} = \mathcal{E}^{\alpha} + \mathcal{P}^{\alpha}, \quad (2.23)$$

$$\mathcal{H}^{\alpha} = \mathcal{B}^{\alpha} - \mathcal{M}^{\alpha}. \quad (2.24)$$

We have defined $S^{\alpha}{}_{\beta}$ as

$$S^{\alpha}{}_{\beta} = \delta^{\alpha}{}_{\beta} + u^{\alpha}u_{\beta}. \quad (2.25)$$

To complete the above system of equations, one must formulate constitutive equations which are invariant under the proper orthochronous Lorentz group and which satisfy the Clausius–Duhem inequality,

$$n_0 D\eta_{00} + \left(\frac{q^{\beta}}{\theta}\right)_{,\beta} + \frac{h_0}{\theta} \geq 0, \quad (2.26)$$

where η_{00} is the entropy density, θ the temperature, and h_0 is given by

$$h_0 = f^{\alpha}u_{\alpha} \quad (2.27)$$

and $D\eta_{00} = \eta_{00,\alpha}u^{\alpha}$.

The choice of constitutive variables is, in a certain sense, arbitrary. In our previous publication,[2] we considered e, η_{00}, p^{α}, q^{α}, $t^{\alpha\beta}$, $\mathcal{P}^{\alpha}$, and $\mathcal{M}^{\alpha}$ as functions of θ, $X^{K}{}_{,\alpha}$, $\mathcal{E}^{\alpha}$, $\mathcal{B}^{\alpha}$, $\theta_{,\alpha}$, where $X^{K}(x^{\alpha})$ is the inverse deformation. For the purpose of this article, we consider e, θ, p^{α}, $t^{\alpha\beta}$, $\mathcal{P}^{\alpha}$, and $\mathcal{M}^{\alpha}$ as functions of $\eta_{00}, X^{K}{}_{,\alpha}$, $\mathcal{E}^{\alpha}$, $\mathcal{B}^{\alpha}$. We will neglect dissipation and consider a nonconducting, charge-free material ($q^{\alpha} = 0$, $\sigma^{\alpha} = 0$). Repeating the reasoning of Ref. 2, we obtain

$$t_{\alpha}{}^{\beta} = -n_0 \frac{\partial e}{\partial X^{K}{}_{,\beta}} X^{K}{}_{,\alpha}, \quad (2.28)$$

$$\mathcal{P}^{\alpha} = -n_0 \frac{\partial e}{\partial \mathcal{E}_{\alpha}}, \quad (2.29)$$

$$\mathcal{M}^{\alpha} = -n_0 \frac{\partial e}{\partial \mathcal{B}_{\alpha}}, \quad (2.30)$$

$$p^{\alpha} = -\epsilon^{\alpha\beta\gamma\delta}\mathcal{P}_{\beta}\mathcal{B}_{\gamma}u_{\delta} - \epsilon^{\alpha\beta\gamma\delta}\mathcal{E}_{\beta}\mathcal{M}_{\gamma}u_{\delta}, \quad (2.31)$$

$$\theta = \frac{\partial e}{\partial \eta_{00}}, \quad (2.32)$$

where

$$e = e(\eta_{00}, X^{K}{}_{,\alpha}, \mathcal{E}_{\alpha}, \mathcal{B}_{\alpha}). \quad (2.33)$$

The balance of energy–momentum (2.4) can be expressed as

$$\hat{T}^{\alpha\beta}{}_{,\beta} = f^{\alpha}, \quad (2.34)$$

where $\hat{T}^{\alpha\beta}$ is the total energy–momentum tensor[5]

$$\hat{T}^{\alpha\beta} = T^{\alpha\beta} - \phi^{\alpha}{}_{\gamma}\mathcal{G}^{\gamma\beta} + \tfrac{1}{4}\phi^{\delta\gamma}\phi_{\gamma\delta}\gamma^{\alpha\beta}, \quad (2.35)$$

$$\hat{T}^{[\alpha\beta]} = 0. \quad (2.36)$$

By using (2.28)–(2.31), $\hat{T}^{\alpha\beta}$ can be written as

$$\hat{T}_{\alpha}{}^{\beta} = n_0 e u_{\alpha}u^{\beta} + p_{\alpha}u^{\beta} + n_0 \frac{\partial e}{\partial X^{K}{}_{,\beta}} X^{K}{}_{,\alpha} - \phi_{\alpha\gamma}\mathcal{G}^{\gamma\beta} + \tfrac{1}{4}\phi^{\delta\gamma}\phi_{\gamma\delta}\delta_{\alpha}{}^{\beta}, \quad (2.37)$$

where e has the functional form (2.33). By using the invariance of $e(\eta_{00}, X^{K}{}_{,\alpha}, \mathcal{E}_{\alpha}, \mathcal{B}_{\alpha})$ under the proper orthochronous Lorentz group, it is not difficult to show that $\hat{T}^{[\alpha\beta]} = 0$ is satisfied.

3. A VARIATION PRINCIPLE FOR ELECTROMAGNETIC SOLIDS

In this section we show that the equations [(2.29), (2.30), and (2.37)] proposed in Ref. 2 for the total

[5] There is an error of a minus sign in Eq. (5.28) of Ref. 2.

energy–momentum tensor and the constitutive equations of a nonconducting, charge-free solid follow from a variational principle. We consider the Lagrangian action $\mathfrak{J}$ of a material tube[6] τ:

$$\mathfrak{J} = -\int_\tau n_0 \tilde{e}(\eta_{00}, X^K{}_{,\alpha}, \phi_{\alpha\beta})\, dv^4 + \tfrac{1}{4}\int_{V_4} \phi_{\alpha\beta}\phi^{\beta\alpha}\, dv^4, \tag{3.1}$$

where dv^4 is the four-dimensional volume element of space–time and V_4 is all space–time.

The form (3.1) of the Lagrangian action is very general. Within the material tube we have assumed only that the Lagrangian density is a function of η_{00}, $X^K{}_{,\alpha}$, $\phi_{\alpha\beta}$ (that is, a function of the entropy, deformation gradients, the velocity $v^i = -x^i{}_{,K}X^K{}_{,4}$, and the electric and magnetic fields). Since $n_0 \neq 0$, we can also assume that the Lagrangian density has the form $n_0\tilde{e}$. There is no loss of generality in subtracting off the free-space Lagrangian density of the electromagnetic fields.

The physical interpretation of the terms appearing in (3.1) is the following: By using (2.3) and an expansion in v^2 in the first term of (3.1), one obtains the usual Lagrangian of a deformable-body kinetic energy minus the stored internal energy, with the mass, as expected in a relativistic theory, depending on the internal energy of the material. The second term is the usual Lagrangian action of the electromagnetic field in free space.

The function $\tilde{e}(\eta_{00}, X^K{}_{,\alpha}, \phi_{\alpha\beta})$ is related to the function $e(\eta_{00}, X^K{}_{,\alpha}, \mathcal{E}_\alpha, \mathcal{B}_\alpha)$ by

$$e(\eta_{00}, X^K{}_{,\alpha}, \mathcal{E}_\alpha, \mathcal{B}_\alpha) = \tilde{e}(\eta_{00}, X^K{}_{,\alpha}, \phi_{\alpha\beta}), \tag{3.2}$$

where the decomposition (2.12) is used. (The world velocity u^α can be expressed as a function of $X^K{}_{,\alpha}$.)

We introduce the variations δx^α of the space–time coordinates,

$$\bar{x}^\alpha = x^\alpha + \epsilon\delta x^\alpha, \tag{3.3}$$

and the partial variations of the magnetic-flux tensor $\phi'_{\alpha\beta}$, of the inverse deformation X'^K, and of the entropy η'_{00}:

$$\begin{aligned} \bar{\phi}_{\alpha\beta}(x^\gamma) &= \phi_{\alpha\beta}(x^\gamma) + \epsilon\phi'_{\alpha\beta}(x^\gamma), \\ \bar{X}^K(x^\gamma) &= X^K(x^\gamma) + \epsilon X'^K(x^\gamma), \\ \bar{\eta}_{00}(x^\gamma) &= \eta_{00}(x^\gamma) + \epsilon\eta'_{00}(x^\gamma), \end{aligned} \tag{3.4}$$

where ϵ is a small parameter. The complete variations $\delta\phi_{\alpha\beta}$, δX^K, and $\delta\eta_{00}$ are defined as

$$\begin{aligned} \bar{\phi}_{\alpha\beta}(\bar{x}^\gamma) &= \phi_{\alpha\beta}(x^\gamma) + \epsilon\delta\phi_{\alpha\beta}(x^\gamma), \\ \bar{X}^K(\bar{x}^\gamma) &= X^K(x^\gamma) + \epsilon\delta X^K(x^\gamma), \\ \bar{\eta}_{00}(\bar{x}^\gamma) &= \eta_{00}(x^\gamma) + \epsilon\delta\eta_{00}(x^\gamma). \end{aligned} \tag{3.5}$$

[6] A material tube is the four-dimensional volume swept out by the material body as it undergoes deformation. See Ref. 1.

The total variation $\delta\mathfrak{J}$ of the Lagrangian action $\mathfrak{J}$ is given by

$$\bar{\mathfrak{J}} = \mathfrak{J} + \epsilon\delta\mathfrak{J}, \tag{3.6}$$

where

$$\bar{\mathfrak{J}} = -\int_{\bar{\tau}} \bar{n}_0 \tilde{e}(\bar{\eta}_{00}, \overline{X^K}_{,\alpha}, \bar{\phi}_{\alpha\beta})\, dv^4 + \tfrac{1}{4}\int_{\bar{V}_4} \bar{\phi}_{\alpha\gamma}\bar{\phi}^{\gamma\alpha}\, dv^4. \tag{3.7}$$

The variational principle for a conducting, charge-free solid ($q^\alpha = 0$, $\sigma^\alpha = 0$) is

$$\delta\mathfrak{J} = -\int_\tau f_\alpha \delta x^\alpha\, dv^4 + \int_{\partial\tau} T_\alpha \delta x^\alpha\, ds^3 \tag{3.8}$$

under the conditions

$$\delta X^K = 0 \quad \text{in } \tau, \tag{3.9}$$

$$\delta\eta_{00} = 0 \quad \text{in } \tau, \tag{3.10}$$

$$\delta\int_S \phi_{\alpha\beta}\, dx^\alpha \wedge dx^\beta = \delta\int_{\partial S} A_\alpha\, dx^\alpha, \tag{3.11}$$

where S is an arbitrary two-dimensional surface. In Eq. (3.8), f_α is the applied body-force four-vector and T_α is the applied surface-traction four-vector. The restriction (3.9) preserves the identity of the material particles. This condition is equivalent to the variation usually used in classical elasticity theories,

$$\delta x^k = \frac{d}{d\lambda} x^k(X^K, t, \lambda)|_{\lambda=0}.$$

Condition (3.10) holds the entropy of the material particles constant. The restriction (3.11) preserves the conservation of magnetic flux (2.7). The four-vector A_α can be interpreted as the four-potential. This procedure is equivalent to the gauge-invariant variations of Weiss[7] when the Lagrangian action is considered a function of the four-potential A_α.

The restriction (3.9) implies that

$$X'^K + X^K{}_{,\alpha}\delta x^\alpha = 0.$$

Thus we obtain

$$\delta(X^K{}_{,\alpha}) = -X^K{}_{,\beta}(\delta x^\beta)_{,\alpha}. \tag{3.12}$$

By performing the variation (3.11), we see that

$$\int_S \delta\phi_{\alpha\beta}\, dx^\alpha \wedge dx^\beta + \int_S \phi_{\alpha\beta}\, d(\delta x^\alpha) \wedge dx^\beta + \int_S \phi_{\alpha\beta}\, dx^\alpha \wedge d(\delta x^\beta) = \int_{\partial S} \delta A_\alpha\, dx^\alpha + \int_{\partial S} A_\alpha\, d(\delta x^\alpha).$$

[7] P. Weiss, Proc. Roy. Soc. (London) **A169**, 119 (1939).

This implies

$$\int_S [\delta\phi_{\alpha\beta} + \phi_{\gamma\beta}(\delta x^\gamma)_{,\alpha} - \phi_{\gamma\alpha}(\delta x^\gamma)_{,\beta}]\, dx^\alpha \wedge dx^\beta = \int_{\partial S} \hat{\delta} A_\alpha\, dx^\alpha, \quad (3.13)$$

where we have used the theorem of Stokes and introduced the variation

$$\hat{\delta} A_\alpha = \delta A_\alpha - A_{\gamma,\alpha}\delta x^\gamma = A'_\alpha + (A_{\alpha,\gamma} - A_{\gamma,\alpha})\delta x^\gamma. \quad (3.14)$$

$\hat{\delta} A_\alpha$ is the gauge-invariant variation of Weiss.[7] Since (3.13) is valid for an arbitrary two-dimensional surface S, we obtain

$$\delta\phi_{\alpha\beta} = (\hat{\delta} A_\beta)_{,\alpha} - (\hat{\delta} A_\alpha)_{,\beta} - \phi_{\gamma\beta}(\delta x^\gamma)_{,\alpha} + \phi_{\gamma\alpha}(\delta x^\gamma)_{,\beta}. \quad (3.15)$$

From the kinematical relations,[1]

$$v^i = -X^K{}_{,4} x^i{}_{,K},$$
$$n_0 = n_{00} \tfrac{1}{6} \epsilon^{\alpha\beta\gamma\delta} \epsilon_{KLM} X^K{}_{,\alpha} X^L{}_{,\beta} X^M{}_{,\gamma} u_\delta, \quad n_{00} = n_{00}(X^K),$$

it can be shown that

$$\delta u^\alpha = S^\alpha{}_\beta (\delta x^\beta)_{,\gamma} u^\gamma,$$
$$\delta n_0 = -n_0 S^\beta{}_\alpha (\delta x^\alpha)_{,\beta}, \quad (3.16)$$

where $S^\alpha{}_\beta$ is given by (2.25).

Using (3.10), (3.12), (3.15), and (3.16), the variation $\delta\mathfrak{J}$ can be expressed as[8]

$$\delta\mathfrak{J} = \int_\tau [\hat{T}_\alpha{}^\beta (\delta x^\alpha)_{,\beta} + \mathfrak{G}^{\alpha\beta}(\hat{\delta} A_\alpha)_{,\beta}]\, dv^4 + \int_{V_4 - \tau} [\hat{T}_\alpha{}^\beta (\delta x^\alpha)_{,\beta} + \mathfrak{G}^{\alpha\beta}(\hat{\delta} A_\alpha)_{,\beta}]\, dv^4, \quad (3.17)$$

where we have defined the energy–momentum tensor $\hat{T}_\alpha{}^\beta$,

$$\hat{T}_\alpha{}^\beta = n_0 \tilde{e} u_\alpha u^\beta + n_0 \frac{\partial \tilde{e}}{\partial X^K{}_{,\beta}} X^K{}_{,\alpha} - \phi_{\alpha\gamma}\mathfrak{G}^{\gamma\beta} + \tfrac{1}{4}\phi_{\delta\gamma}\phi^{\gamma\delta}\delta_\alpha{}^\beta \quad \text{in } \tau,$$
$$\hat{T}_\alpha{}^\beta = -\phi_{\alpha\gamma}\phi^{\gamma\beta} + \tfrac{1}{4}\phi_{\delta\gamma}\phi^{\gamma\delta}\delta_\alpha{}^\beta \quad \text{in } V_4 - \tau, \quad (3.18)$$

and $\mathfrak{G}^{\alpha\beta}$ is defined by

$$\mathfrak{G}^{\alpha\beta} = \phi^{\alpha\beta} + n_0 \frac{\partial \tilde{e}}{\partial \phi_{\alpha\beta}} \quad \text{in } \tau,$$
$$\mathfrak{G}^{\alpha\beta} = \phi^{\alpha\beta} \quad \text{in } V_4 - \tau. \quad (3.19)$$

Integrating (3.17) by parts and setting the surface integrals at infinity equal to zero, the variational principle (3.8) reduces to

$$\int_\tau (-\hat{T}_\alpha{}^\beta{}_{,\beta}\delta x^\alpha - \mathfrak{G}^{\alpha\beta}{}_{,\beta}\hat{\delta} A_\alpha)\, dv^4 + \int_{V_4 - \tau} (-\hat{T}_\alpha{}^\beta{}_{,\beta}\delta x^\alpha - \mathfrak{G}^{\alpha\beta}{}_{,\beta}\hat{\delta} A_\alpha) - \int_{\partial\tau} ([\hat{T}_\alpha{}^\beta] N_\beta \delta x^\alpha + [\mathfrak{G}^{\alpha\beta}] N_\beta \hat{\delta} A_\alpha)\, ds^3 = -\int_\tau f_\alpha \delta x^\alpha\, dv^4 + \int_{\partial\tau} T_\alpha \delta x^\alpha\, ds^3. \quad (3.20)$$

Here N_β is the outward normal of $\partial\tau$ and by $[h]$ we denote $h_+ - h_-$, where h_+ is the value on the positive side of $\partial\tau$ of the function h and h_- its value on the negative side of $\partial\tau$.

Since (3.20) is valid for all variations δx^α, $\hat{\delta} A_\alpha$, we obtain

$$\hat{T}_\alpha{}^\beta{}_{,\beta} = f_\alpha \quad \text{in } \tau, \quad (3.21)$$
$$\mathfrak{G}^{\alpha\beta}{}_{,\beta} = 0 \quad \text{in } \tau, \quad (3.22)$$
$$\hat{T}_\alpha{}^\beta{}_{,\beta} = 0 \quad \text{in } V_4 - \tau, \quad (3.23)$$
$$\mathfrak{G}^{\alpha\beta}{}_{,\beta} = 0 \quad \text{in } V_4 - \tau, \quad (3.24)$$
$$[\hat{T}_\alpha{}^\beta] N_\beta = -T_\alpha \quad \text{on } \partial\tau, \quad (3.25)$$
$$[\mathfrak{G}_\alpha{}^\beta] N_\beta = 0 \quad \text{on } \partial\tau. \quad (3.26)$$

It is not difficult to show that (2.7) and (3.24) imply that (3.23) is identically satisfied. Equations (2.7) and (3.24) are Maxwell's equations in free space. From (2.9) and (3.19) we see that $\pi^{\alpha\beta}$ is given by

$$\pi^{\alpha\beta} = -n_0 \frac{\partial \tilde{e}}{\partial \phi_{\alpha\beta}}. \quad (3.27)$$

This is the constitutive relation for the polarization–magnetization tensor in τ. Equations (2.7) and (3.22) are Maxwell's equations in the material body. Equation (3.21) is the balance of energy–momentum in the material.

To demonstrate that the form of the total energy–momentum tensor given by (3.18) and the electromagnetic constitutive equations (3.27) is equivalent to Eqs. (2.29), (2.30), and (2.37), we use the identification (3.2) along with (2.12) and[9]

$$\frac{\partial u^\alpha}{\partial X^K{}_{,\beta}} = -u^\beta x^\alpha{}_K, \quad (3.28)$$

where

$$x^\alpha{}_K = S^\alpha{}_i x^i{}_{,K},$$
$$x^\alpha{}_K X^K{}_{,\beta} = S^\alpha{}_\beta,$$
$$x^\alpha{}_L X^K{}_{,\alpha} = \delta^K{}_L.$$

[8] The independent variables are $\phi_{\alpha\beta}$ ($\alpha < \beta$); however, it is found convenient to sum over all α, β. This is the reason for the presence of various factors of 2 in Eq. (3.17) and others that follow.

[9] This follows from the relation listed after (3.15).

From (3.2), (2.12), and (3.27) we have

$$n_0 \frac{\partial e}{\partial \mathcal{E}_\gamma} = \tfrac{1}{2} n_0 \frac{\partial \tilde{e}}{\partial \phi_{\alpha\beta}} \frac{\partial \phi_{\alpha\beta}}{\partial \mathcal{E}_\gamma} = -\tfrac{1}{2}\pi^{\alpha\beta}(\delta^\gamma{}_\beta u_\alpha - \delta_\alpha{}^\gamma u_\beta)$$
$$= -\pi^{\alpha\gamma} u_\alpha = -\mathcal{P}^\gamma, \tag{3.29}$$

where we have employed (2.14). Similarly,

$$n_0 \frac{\partial e}{\partial \mathcal{B}_\gamma} = \tfrac{1}{2} n_0 \frac{\partial \tilde{e}}{\partial \phi_{\alpha\beta}} \frac{\partial \phi_{\alpha\beta}}{\partial \mathcal{B}_\gamma} = -\tfrac{1}{2}\pi^{\alpha\beta}\epsilon_{\alpha\beta}{}^\gamma{}_\delta u^\delta = -\mathcal{M}^\gamma. \tag{3.30}$$

Also from (3.27) and (3.28),

$$n_0 \frac{\partial e}{\partial X^K{}_{,\beta}}$$
$$= n_0 \frac{\partial \tilde{e}}{\partial X^K{}_{,\beta}} + \tfrac{1}{2} n_0 \frac{\partial \tilde{e}}{\partial \phi^{\alpha\gamma}} \frac{\partial \phi^{\alpha\gamma}}{\partial u^\delta} \frac{\partial u^\delta}{\partial X^K{}_{,\beta}}$$
$$= n_0 \frac{\partial \tilde{e}}{\partial X^K{}_{,\beta}} + \tfrac{1}{2}\pi_{\alpha\gamma}(\mathcal{E}^\gamma \delta_\delta{}^\alpha - \mathcal{E}^\alpha \delta_\delta{}^\gamma + \epsilon^{\alpha\gamma\sigma}{}_\delta \mathcal{B}_\sigma) u^\beta x^\delta{}_K$$
$$= n_0 \frac{\partial \tilde{e}}{\partial X^K{}_{,\beta}} - p_\gamma u^\beta x^\gamma{}_K, \tag{3.31}$$

where we have used (2.31). Thus, $\hat{T}_\alpha{}^\beta$ can be written as

$$\hat{T}_\alpha{}^\beta = n_0 e u_\alpha u^\beta + n_0 \frac{\partial e}{\partial X^K{}_{,\beta}} X^K{}_{,\alpha} + p_\alpha u^\beta$$
$$- \phi_{\alpha\gamma}\mathcal{G}^{\gamma\beta} + \tfrac{1}{4}\phi_{\delta\gamma}\phi^{\gamma\delta}\delta_\alpha{}^\beta \quad \text{in } \tau, \tag{3.32}$$

which is identical to (2.37).

The requirement that e is a scalar invariant under the proper orthochronous Lorentz group yields[2]

$$e = e(C^{-1KL}, \mathcal{E}^K, \mathcal{B}^K), \tag{3.33}$$

where

$$C^{-1KL} = \gamma^{\alpha\beta} X^K{}_{,\alpha} X^L{}_{,\beta},$$
$$\mathcal{E}^K = X^K{}_{,\alpha}\mathcal{E}^\alpha, \tag{3.34}$$
$$\mathcal{B}^K = X^K{}_{,\alpha}\mathcal{B}^\alpha.$$

Using (3.29) and (3.30), we obtain

$$\mathcal{P}_\alpha = -n_0 \frac{\partial e}{\partial \mathcal{E}^K} X^K{}_{,\alpha},$$
$$\mathcal{M}_\alpha = -n_0 \frac{\partial e}{\partial \mathcal{B}^K} X^K{}_{,\alpha}, \tag{3.35}$$

and from (3.32) we have

$$\hat{T}_{\alpha\beta} = n_0 e u_\alpha u_\beta + 2n_0 \frac{\partial e}{\partial C^{-1KL}} X^K{}_{,\alpha} X^L{}_{,\beta} + p_\alpha u^\beta - \mathcal{P}_\alpha \mathcal{E}_\beta$$
$$- \mathcal{M}_\alpha \mathcal{B}_\beta - \phi_{\alpha\gamma}\mathcal{G}^\gamma{}_\beta + \tfrac{1}{4}\phi_{\delta\gamma}\phi^{\gamma\delta}\gamma_{\alpha\beta}, \tag{3.36}$$

where e has the functional form (3.33).

4. AN ALTERNATE FORMULATION

The theory presented in the previous sections treated $\mathcal{E}_\alpha$ and $\mathcal{B}_\alpha$ as the independent electromagnetic constitutive variables. At times it is convenient to consider another set. In this section we shall show how this can be done when $\mathcal{P}_\alpha$ and $\mathcal{M}_\alpha$ are chosen as the independent electromagnetic variables. Thus, we consider e, θ, p^α, $t^{\alpha\beta}$, $\mathcal{E}^\alpha$, and $\mathcal{B}^\alpha$ functions of η_{00}, $X^K{}_{,\alpha}$, $\mathcal{P}^\alpha$, and $\mathcal{M}^\alpha$. By using Eq. (5.24) of Ref. 2 in the second law (2.26) and repeating the reasoning of Ref. 2, it is not difficult to demonstrate that

$$t_\alpha{}^\beta = -n_0 \frac{\partial \hat{e}}{\partial X^K{}_{,\beta}} X^K{}_{,\alpha} - (\mathcal{P}^\gamma \mathcal{E}_\gamma + \mathcal{M}^\gamma \mathcal{B}_\gamma) S_\alpha{}^\beta, \tag{4.1}$$

$$\mathcal{E}^\alpha = n_0 \frac{\partial \hat{e}}{\partial \mathcal{P}_\alpha}, \tag{4.2}$$

$$\mathcal{B}^\alpha = n_0 \frac{\partial \hat{e}}{\partial \mathcal{M}_\alpha}, \tag{4.3}$$

$$\theta = \frac{\partial \hat{e}}{\partial \eta_{00}}, \tag{4.4}$$

where

$$\hat{e} \equiv e + \frac{\mathcal{P}^\alpha \mathcal{E}_\alpha}{n_0} + \frac{\mathcal{M}^\alpha \mathcal{B}_\alpha}{n_0}. \tag{4.5}$$

Here $\hat{e}$ has the functional form

$$\hat{e} = \hat{e}(\eta_{00}, X^K{}_{,\alpha}, \mathcal{P}^\alpha, \mathcal{M}^\alpha). \tag{4.6}$$

The nonmechanical momentum p_α is still given by (2.31). The total energy–momentum tensor has the form

$$\hat{T}_\alpha{}^\beta = n_0 \hat{e} u_\alpha u^\beta + p_\alpha u^\beta + n_0 \frac{\partial \hat{e}}{\partial X^K{}_{,\beta}} X^K{}_{,\alpha} - \tfrac{1}{2}\pi^{\delta\gamma}\phi_{\gamma\delta}\delta_\alpha{}^\beta$$
$$- \phi_{\alpha\gamma}\mathcal{G}^{\gamma\beta} + \tfrac{1}{4}\phi_{\gamma\delta}\phi^{\delta\gamma}\delta_\alpha{}^\beta. \tag{4.7}$$

The Lagrangian action $\mathfrak{I}^*$ that yields the above theory follows from (3.1) by a Legendre transformation,

$$\mathfrak{I}^* = -\int_\tau n_0 \bar{e}(\eta_{00}, X^K{}_{,\alpha}, \pi_{\alpha\beta})\, dv^4$$
$$+ \tfrac{1}{4}\int_{V_4} \phi_{\alpha\gamma}\phi^{\gamma\alpha}\, dv^4 - \tfrac{1}{2}\int_\tau \pi^{\alpha\beta}\phi_{\beta\alpha}\, dv^4. \tag{4.8}$$

The function $\bar{e}(\eta_{00}, X^K{}_{,\alpha}, \pi_{\alpha\beta})$ is related to the function $\hat{e}(\eta_{00}, X^K{}_{,\alpha}, \mathcal{P}_\alpha, \mathcal{M}_\alpha)$ by

$$\hat{e}(\eta_{00}, X^K{}_{,\alpha}, \mathcal{P}_\alpha, \mathcal{M}_\alpha) = \bar{e}(\eta_{00}, X^K{}_{,\alpha}, \pi_{\alpha\beta}), \tag{4.9}$$

where the decomposition (2.14) is used. If one neglects the magnetization, the Lagrangian $\mathfrak{I}^*$ is the relativistic generalization of the Lagrangian employed by Toupin[3,4] for a static elastic dielectric. By introducing the independent variations $\delta\pi_{\alpha\beta}$ and using restrictions (3.9)–(3.11), it is not difficult to show that (3.21)–(3.27) follow from (4.8) with $\mathcal{E}^\alpha$, B^α, and $\hat{T}_\alpha{}^\beta$ given by (4.2), (4.3), and (4.7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Legendre Polynomial Expansions of Hypergeometric Functions with Applications*†

JONAS T. HOLDEMAN, JR.‡
University of California Los Alamos Scientific Laboratory, Los Alamos, New Mexico

(Received 10 June 1966)

The expansion of a class of hypergeometric functions in a series of Legendre polynomials is derived. The range of validity and the meaning to be attached to the sums is investigated. Several applications to the problem of the scattering of charged particles are presented.

1. INTRODUCTION

Occasionally one encounters in a scattering situation the problem of solving a Schrödinger (or Dirac) equation which contains a potential λr^{-n} in the presence of an αr^{-1} (point Coulomb) potential. A standard technique is to treat the point Coulomb potential exactly and the potential λr^{-n} in the Born (or distorted wave Born) approximation (BA). The scattering amplitude can be expressed as a sum of the Coulomb amplitude and a residual amplitude. To first order in the parameter λ, this residual amplitude contains a sum over the BA phase shifts times the Legendre polynomials weighted by the Coulomb S-matrix elements. When the BA phase shifts can be expressed as a certain ratio of gamma functions, the expansions we shall give are helpful.

2. THE EXPANSIONS

The notation we will use for the hypergeometric function (HF) is standard and all properties of these functions needed in the following discussion can be found in Ref. 1, Chap. 2. We expand the function ${}_2F_1(a, b; c; \frac{1}{2}(1 + x))$ as

$$ {}_2F_1(a, b; c; \tfrac{1}{2}(1 + x)) = \sum_l \alpha_l P_l(x), \tag{2.1} $$

where $P_l(x)$ is the Legendre polynomial. The expansion coefficients α_l are given by

$$ \alpha_l = (l + \tfrac{1}{2}) \int_{-1}^{1} {}_2F_1(a, b; c; \tfrac{1}{2}(1 + x)) P_l(x)\, dx, \tag{2.2} $$

provided Re $(a + b - c) < \frac{3}{4}$, although the existence of Eq. (2.2) merely requires Re $(a + b - c) < 1$.

If we carry out the indicated integration (using either the integral representation of the HF or the series representation and integrating term by term), we find

$$ \alpha_l = \frac{(2l + 1)(a)_l(b)_l l!}{(c)_l \Gamma(2l + 2)} {}_3F_2\left[\begin{matrix} a + l, b + l, 1 + l; \\ c + l, 2l + 2 \end{matrix} 1\right], \tag{2.3} $$

where ${}_3F_2[\ ]$ is a generalized hypergeometric series (GHS) of unit argument. For the special case $c = 1$, the GHS reduces to an HF and can be expressed in terms of gamma functions

$$ \begin{aligned} {}_3F_2\left[\begin{matrix} a + l, b + l, 1 + l; \\ c + l, 2l + 2 \end{matrix} 1\right]_{c=1} &= {}_2F_1(a + l, b + l, 2l + 2; 1) \\ &= \frac{\Gamma(2l + 2)\Gamma(2 - a - b)}{\Gamma(l + 2 - a)\Gamma(l + 2 - b)}. \end{aligned} \tag{2.4} $$

Inserting Eq. (2.4) into Eq. (2.3) yields

$$ \alpha_l = \Gamma(2 - a - b) \frac{(2l + 1)(a)_l(b)_l}{\Gamma(2 - a + l)\Gamma(2 - b + l)}, \tag{2.5} $$

where

$$ (a)_l = a(a + 1) \cdots (a + l - 1) = \Gamma(a + l)/\Gamma(a). \tag{2.6} $$

We note that $\alpha_l/\Gamma(2 - a - b)$ is an analytic function of a and b for all a and b. Our principal result is then

$$ \begin{aligned} \sum_{l=0}^{\infty} (2l + 1) \frac{(a)_l}{\Gamma(l + 2 - a)} \frac{(b)_l}{\Gamma(l + 2 - b)} P_l(x) \\ = \frac{1}{\Gamma(2 - a - b)} {}_2F_1(a, b; 1; \tfrac{1}{2}(1 + x)), \end{aligned} \tag{2.7} $$

where for the moment we have the restriction that Re $(a + b) < \frac{7}{4}$. If, in addition, Re $(a + b) < 1$, the series on the left of Eq. (2.7) converges uniformly to the function on the right in the closed interval $[-1, 1]$. For the case $b = 1$, Eq. (2.7) becomes

$$ \sum_{l=0}^{\infty} (2l + 1) \frac{(a)_l}{\Gamma(l + 2 - a)} P_l(x) = \frac{1}{\Gamma(1 - a)} \left(\frac{1 - x}{2}\right)^{-a}. \tag{2.8} $$

This is our second result. Differentiating Eq. (2.7) and

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

† Part of this work was done while the author was at Case Institute of Technology.

‡ Present address: Department of Physics, Michigan State University, East Lansing, Mich. 48823.

[1] A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill Book Co., Inc., New York, 1953), Vol. 1.

using the relation

$$P_l^1(x) = (1 - x^2)^{\frac{1}{2}} \frac{d}{dx} P_l(x), \tag{2.9}$$

where $P_l^1(x)$ is the first associated Legendre polynomial, we have

$$\sum_{l=1}^{\infty} (2l + 1) \frac{(a)_l}{\Gamma(l + 2 - a)} \frac{\Gamma(l + b)}{\Gamma(l + 2 - b)} P_l^1(x) = \frac{a}{2} \frac{\Gamma(b + 1)}{\Gamma(2 - a - b)} (1 - x^2)^{\frac{1}{2}} \times {}_2F_1(a + 1, b + 1; 2; \tfrac{1}{2}(1 + x)). \tag{2.10}$$

For Re $(a + b) < 0$, the series on the left of Eq. (2.10) converges uniformly to the function on the right in the closed interval $[-1, 1]$. If we now set $b = 0$ in Eq. (2.10) and continue the resulting HF analytically, we find our third result

$$\sum_{l=1}^{\infty} \frac{2l + 1}{l(l + 1)} \frac{(a)_l}{\Gamma(l + 2 - a)} P_l^1(x) = \frac{1}{\Gamma(2 - a)} \left(\frac{1 - x}{1 + x}\right)^{\frac{1}{2}} \left[\left(\frac{1 - x}{2}\right)^{-a} - 1\right]. \tag{2.11}$$

The restrictions placed on the parameter a and b in the course of these derivations are such as to exclude the application of Eqs. (2.7), (2.8), and (2.11) to some interesting physical problems. We have already implied that it is possible to loosen these restrictions but this will require a reinterpretation of the sums in the equations. We note that the integral defining α_l [Eq. (2.2)] does not exist for Re $(a + b) > 2$, but we can use Eq. (2.5) (as an analytic continuation) to define the quantity $\alpha_l/\Gamma(2 - a - b)$ for all a and b.[2a] When Re $(a + b) > 2$, the sums in Eqs. (2.7), (2.8), and (2.11) do not even converge pointwise, but meaning can be attached to them in several ways. First, the HF on the right is an analytic continuation of the sum for fixed x into regions of a and b where the series does not converge; second, the series are summable by summation methods such as described in Ref. 2b; and third, these equations can be regularized to define continuous linear functionals (CLF) or generalized functions.

We define the set of functions G such that a function $g_\lambda(x)$ belongs to G if, on the closed segment $[-1, 1]$, $g_\lambda(x)$ is defined and has the properties that (i) it is piecewise continuous, (ii) its Legendre polynomial expansion coefficients γ_l exist in the usual sense (and hence $g_\lambda(x)$ is integrable on $[-1, 1]$) and $\gamma_l = O(l^{-2\mu+1})$ where $\mu > 0$, and (iii) in the neighborhood of $x = 1$, $g_\lambda(x) = O([\frac{1}{2}(1 - x)]^{-1+\nu})$, where $\nu > 0$ and λ is given by $\lambda = \min(\mu, \nu)$. We let G_{λ_0} be a subset of G such that $g_\lambda(x) \in G_{\lambda_0}$ if and only if $\lambda \geq \lambda_0$.

For some fixed (but arbitrary) λ_0 we consider the integral

$$[\Gamma(2 - a - b)]^{-1} \int_{-1}^{1} {}_2F_1(a, b; 1; \tfrac{1}{2}(1 + x)) \times g_\lambda(x)\, dx = f_1(a, b), \tag{2.12}$$

for all $g_\lambda(x) \in G_{\lambda_0}$. For all values of a and b such that Re $(a + b) < \lambda_0 + 1$, the integrand is integrable on $[-1, 1]$ and is an analytic function of a and b on the interval $(-1, 1)$. Thus, for all a and b such that Re $(a + b) < \lambda_0 + 1$, the integral in Eq. (2.12) is uniformly continuous in a and b and defines an analytic function $f_1(a, b)$ and a CLF on G_{λ_0}.

For the same λ_0 and subset G_{λ_0}, consider the integral

$$\int_{-1}^{1} \sum_{l=0}^{\infty} (2l + 1) \frac{(a)_l}{\Gamma(2 - a + l)} \frac{(b)_l}{\Gamma(2 - b + l)} P_l(x) g_\lambda(x)\, dx = \sum_{l=0}^{\infty} 2 \frac{(a)_l}{\Gamma(2 - a + l)} \frac{(b)_l}{\Gamma(2 - b + l)} \gamma_l = f_2(a, b) \tag{2.13}$$

(where by the integral of the sum when Re $(a + b) > \frac{7}{4}$ we mean implicitly term by term integration).[3] For all a and b such that Re $(a + b) < \lambda_0 + 1$, the second sum in Eq. (2.13) is uniformly convergent and defines an analytic function $f_2(a, b)$ and also a CLF on G_{λ_0}.

From the uniform convergence of the sum in Eq. (2.7) we see that $f_1(a, b) = f_2(a, b) = f(a, b)$ for Re $(a + b) < 1$. For these values of a and b, Eqs. (2.12) and (2.13) define the same CLF on G. But Eqs. (2.12) and (2.13) define analytic continuations of $f(a, b)$ into the region $1 \leq$ Re $(a + b) < \lambda_0 + 1$. Since the analytic continuation of $f(a, b)$ is unique, we must have $f_1(a, b) = f_2(a, b)$ in this region. Thus, Eqs. (2.12) and (2.13) define the same CLF on G_{λ_0} for all a and b such that Re $(a + b) < \lambda_0 + 1$. Thus we see that Eq. (2.7) is valid in the sense of a CLF on G_{λ_0}, for any $\lambda_0 > 0$ and Re $(a + b) < \lambda_0 + 1$, as is Eq. (2.8) for Re $a < \lambda_0$. Equation (2.11) is defined in a similar manner for Re $a < \lambda_0 + 1$.

The preceding discussion only gives sufficient conditions under which the equations in question are defined. These results can be extended to cases where

[2] (a) For the case where $a + b - 1$ is a positive integer, the continuation of Eq. (2.5) yields sums which over some set of functions G_λ yet to be defined are functionally equivalent to zero (FEZ). The existence of sums FEZ has an important bearing on the uniqueness of Eq. (2.7) and following. If in Eq. (2.8) we set a equal to a positive integer, the resulting sums (which are FEZ) can be recognized as expansions of the delta functions and its derivatives. Further interesting results can be obtained in a manner similar to that described in Footnote 4. (b) J. T. Holdeman, Math. Comp. 23, 275 (1969).

[3] The interchange of the order of summation and integration when Re $(a + b) \leq \frac{7}{4}$ is allowed because $g_\lambda(x)$ is absolutely integrable and the sum of Eq. (2.7) converges uniformly.

$a + b - 1$ is equal to a positive integer[4] and to a slightly wider class of functions than G.[5]

3. APPLICATIONS

The Eqs. (2.7), (2.8), and (2.11) are of value in that they provide formulas for evaluating sums of the type mentioned in the introduction. As a simple example, if in Eq. (2.8) we set $a = 1 + i\eta$, we find

$$(2ik)^{-1}\sum_{l=0}^{\infty}(2l+1)e^{2i\sigma_l}P_l(x) = -(\eta/2k)e^{2i\sigma_0}[\tfrac{1}{2}(1-x)]^{-1-i\eta}, \quad (3.1)$$

where

$$e^{2i\sigma_l} = \Gamma(l+1+i\eta)/\Gamma(l+1-i\eta). \quad (3.2)$$

Although a delta function is conventionally included, we see that Eq. (3.1) is just the expansion of the nonrelativistic Coulomb scattering amplitude.

Now suppose that the Hamiltonian contains a term λr^{-n}, where $n > 1$. Then in BA the contribution δ_l of this potential to the total phase shift is given by[6]

$$\tan\delta_l = \lambda A_n \Gamma(l+1-\tfrac{1}{2}(n-1))/\Gamma(l+1+\tfrac{1}{2}(n-1)), \quad (3.3)$$

where

$$A_n = -\pi k^n 2^{-n}(n-2)!/E[\Gamma(\tfrac{1}{2}n)]^2. \quad (3.4)$$

To first order in the parameter λ, the contribution to the scattering amplitude (the residual amplitude) of the potential λr^{-n} in BA is

$$\begin{aligned} f'(\theta) &= \lambda A_n(k)^{-1}\sum_{l=0}^{\infty}(2l+1) \\ &\quad\times\frac{\Gamma(l+1+i\eta)}{\Gamma(l+1-i\eta)}\frac{\Gamma(l+1-\frac{1}{2}(n-1))}{\Gamma(l+1+\frac{1}{2}(n-1))}P_l(x) \\ &= \lambda A_n(k)^{-1}\frac{\Gamma(1+i\eta)\Gamma(\frac{1}{2}(3-n))}{\Gamma(\frac{1}{2}n-\frac{1}{2}-i\eta)} \\ &\quad\times {}_2F_1(1+i\eta, \tfrac{3}{2}-\tfrac{1}{2}n; 1; \tfrac{1}{2}(1+x)). \end{aligned} \quad (3.5)$$

For n even, this gives the residual amplitude in terms of known functions. For n odd and greater than unity, Eq. (3.5) is undefined. It is just the terms given incorrectly by BA that are responsible for this difficulty.

The case in which n is an odd integer and greater than unity can be treated in the following way: For $l \leq \frac{1}{2}(n-3)$, let $\tilde{\delta}_l$ be phase shifts obtained in some manner other than by BA. We replace n by $n + \epsilon$, subtract off the terms with BA phase shifts, and add terms with $\tilde{\delta}_l$ for $l \leq \frac{1}{2}(n-3)$. We then take the limit as $\epsilon \to 0$. This yields

$$\begin{aligned} f'(x) &= f_0'(x) - \frac{\lambda 2^n k^{-1}}{2E}\frac{(n-2)!}{(n-1)!^2} \\ &\quad\times\frac{\Gamma(\frac{1}{2}n+\frac{1}{2}+i\eta)}{\Gamma(\frac{1}{2}n-\frac{1}{2}-i\eta)}[\tfrac{1}{2}(1+x)]^{\frac{1}{2}(n-1)} \\ &\quad\times {}_3F_2\left[\begin{matrix}\frac{1}{2}(n+1)+i\eta, 1, 1\\ \frac{1}{2}(n+1), \frac{1}{2}(n+1)\end{matrix}; \tfrac{1}{2}(1+x)\right], \end{aligned} \quad (3.6)$$

where

$$f_0'(x) = (2ik)^{-1}\sum_{l=0}^{l=\frac{1}{2}(n-3)}(2l+1)e^{2i\sigma_l}(e^{2i\tilde{\delta}_l}-1)P_l(x). \quad (3.7)$$

As a third example, suppose we wish to include the effects of the interaction of the magnetic moment of a charged particle with a point Coulomb field. This interaction is described by a term in the Hamiltonian

$$H' = (\mu_p - \tfrac{1}{2})\frac{\hbar Z e^2}{2m^2c^2}r^{-3}\boldsymbol{\sigma}\cdot\mathbf{L}. \quad (3.8)$$

If the phase shift due to H' is calculated in BA, then, to first order in the strength of H', the spin-dependent scattering amplitude $h(\theta)$ is given by

$$h(\theta) = i(\mu_p - \tfrac{1}{2})\left(\frac{2E}{mc^2}\right)\eta k^{-1}\sum_{l=1}^{\infty}\frac{(2l+1)}{l(l+1)}e^{2i\sigma_l}P_l^1(x). \quad (3.9)$$

If in Eq. (2.11) we set $a = 1 + i\eta$, we find[7]

$$\begin{aligned} h(\theta) &= 2i(\mu_p - \tfrac{1}{2})\left(\frac{E}{mc^2}\right)\left(\frac{\eta}{2k}\right) \\ &\quad\times e^{2i\sigma_0}\left(\frac{1-x}{1+x}\right)^{\frac{1}{2}}\{[\tfrac{1}{2}(1-x)]^{-1-i\eta}-1\}. \end{aligned} \quad (3.10)$$

4. CONCLUSIONS

We have shown that, with certain general restrictions on the parameters, the sums and functions of Eqs. (2.7), (2.8), and (2.11) are equal in the sense of an analytic continuation and that they define the same functional on sets of functions $G_{\lambda_0} \subseteq G$, even when the sums do not converge in the ordinary sense. The conditions imposed on the functions contained in these sets were rather weak and we did not find the most general set on which these equations define functionals.

[4] For instance, if in Eq. (2.8) we take $a = 1 + \epsilon$ and $a = 1 - \epsilon$, subtract the two equations, and take the limit $\epsilon \to 0$, we find

$$\sum_l (2l+1)\psi(l+1)P_l(x) = (x-1)^{-1}.$$

[5] For a (nonrigorous) discussion of the case of Eq. (2.8) when $\mathrm{Re}\,(a) = \lambda_0 = 1$ see J. T. Holdeman and R. M. Thaler, Phys. Rev. **139**, B1186 (1965).

[6] For $l \leq \frac{1}{2}(n-3)$ the phase shifts as given by BA are qualitatively incorrect. These terms can always be subtracted off as in Eq. (3.6).

[7] This result is discussed in the author's Ph.D. thesis. See Nuclear Physics Laboratory, Case Institute of Technology, Cleveland, Ohio, technical report No. COO-1573-1.

One might conjecture that, operationally, the following criterion can be used: If both sides of Eqs. (2.7), (2.8), or (2.11) are multiplied by some function and integrated over some interval contained in $[-1, 1]$ and if both integrals exist, then they will be equal.

Though we gave several examples of applications of the expansions derived in Sec. 2, we have obviously not exhausted all of the possibilities. Several cases of Eqs. (2.7) and (2.8) have appeared in the literature for special values of the parameters a and b,[8] but the equations given here seem to sum up most of the earlier results.

[8] See, for example, N. F. Mott, Proc. Roy. Soc. (London) **A124**, 425 (1929); J. H. Bartlett and R. E. Watson, Proc. Am. Acad. Arts Sci. **74**, 53 (1940).

ACKNOWLEDGMENTS

The author wishes to thank Professor R. M. Thaler for his guidance and assistance and Dr. L. Heller for several helpful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

New Method for Finding Eigenvalues

J. T. ANDERSON*
Physics Department, Queens College of the City University of New York, Flushing, New York

(Received 25 March 1969)

The eigenvalue equation of a function of a single complex variable is shown to be equivalent to the Cauchy–Riemann equations so that the eigenvalue problem is reducible to the problem of finding the analytic regions of the function. The eigenvalues of a function of several complex variables are also found from the analytic regions on each plane. A self-consistent treatment of two relativistic fields is developed and applied to the interaction of spinor and scalar fields.

INTRODUCTION

The analytic properties of complex functions are of interest in the theory of scattering as well as other fields. For example, certain classes of integral equations, which include those reducible to a Dyson equation, can be solved exactly for the renormalized particle spectra[1]; the analytic regions on the momentum and energy planes are defined by solutions of eigenvalue equations which generate the Cauchy–Riemann equations in the two variables. Sufficient conditions for analyticity on both planes have also been given.[2] In the case of angular momentum, the irreducible representations of the complex rotation groups define both the real (particle) and complex (resonance) eigenvalues of the S matrix,[3] and the groups of rotation operators acting on functions of one and two complex variables have been considered by several authors.[4–6]

In this paper we show that, for functions having a complex basis and satisfying either classical or quantum bracket identities, eigenvalues are easily found from solutions of a divergence equation, which is a form of the Cauchy–Riemann equations, and thus provide a necessary condition for analyticity. In this case the eigenvalue problem is reduced to that of finding the analytic regions of the function.

1. EIGENVALUE EQUATION

We consider a function of $P(z) = u(p, q) + iv(p, q)$ of a single complex variable $z = p + iq$. The Poisson brackets read

$$[z, p] = [p, z^*] = i, \tag{1}$$

$$[z, q] = -[q, z^*] = -1, \tag{2}$$

$$[q, P] = \frac{\partial P}{\partial p}, \quad [p, P] = -\frac{\partial P}{\partial q}, \tag{3}$$

so that

$$[z, P] = \frac{\partial P}{\partial q} + i\frac{\partial P}{\partial p} = i\nabla P = 2i\frac{\partial P}{\partial z^*}, \tag{4}$$

where

$$\nabla = \frac{\partial}{\partial p} + i\frac{\partial}{\partial q}.$$

Also

$$[z^*, P^*] = -2i\frac{\partial P^*}{\partial z}. \tag{5}$$

* Permanent address: Box 94, La Jolla, California.

[1] J. T. Anderson, J. Math. Phys. **8**, 998 (1967).
[2] J. T. Anderson, Nuovo Cimento **52**, 375 (1967).
[3] V. S. Popov and E. I. Dolinsky, Zh. Eksp. Teor. Fiz. **46**, 970 1829 (1963) [Sov. Phys.—JETP **19**, 661, 1232 (1964)].
[4] E. G. Beltrami and G. Luzzato, Nuovo Cimento **51**, 147 (1967).
[5] J. Gunson, J. Math. Phys. **6**, 852 (1965).
[6] V. N. Gribov and I. Ya. Pomeranchuk, Phys. Rev. Letters **8**, 343 (1962).

One might conjecture that, operationally, the following criterion can be used: If both sides of Eqs. (2.7), (2.8), or (2.11) are multiplied by some function and integrated over some interval contained in $[-1, 1]$ and if both integrals exist, then they will be equal.

Though we gave several examples of applications of the expansions derived in Sec. 2, we have obviously not exhausted all of the possibilities. Several cases of Eqs. (2.7) and (2.8) have appeared in the literature for special values of the parameters a and b,[8] but the equations given here seem to sum up most of the earlier results.

[8] See, for example, N. F. Mott, Proc. Roy. Soc. (London) **A124**, 425 (1929); J. H. Bartlett and R. E. Watson, Proc. Am. Acad. Arts Sci. **74**, 53 (1940).

ACKNOWLEDGMENTS

The author wishes to thank Professor R. M. Thaler for his guidance and assistance and Dr. L. Heller for several helpful discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

New Method for Finding Eigenvalues

J. T. ANDERSON*
Physics Department, Queens College of the City University of New York, Flushing, New York

(Received 25 March 1969)

The eigenvalue equation of a function of a single complex variable is shown to be equivalent to the Cauchy–Riemann equations so that the eigenvalue problem is reducible to the problem of finding the analytic regions of the function. The eigenvalues of a function of several complex variables are also found from the analytic regions on each plane. A self-consistent treatment of two relativistic fields is developed and applied to the interaction of spinor and scalar fields.

INTRODUCTION

The analytic properties of complex functions are of interest in the theory of scattering as well as other fields. For example, certain classes of integral equations, which include those reducible to a Dyson equation, can be solved exactly for the renormalized particle spectra[1]; the analytic regions on the momentum and energy planes are defined by solutions of eigenvalue equations which generate the Cauchy–Riemann equations in the two variables. Sufficient conditions for analyticity on both planes have also been given.[2] In the case of angular momentum, the irreducible representations of the complex rotation groups define both the real (particle) and complex (resonance) eigenvalues of the S matrix,[3] and the groups of rotation operators acting on functions of one and two complex variables have been considered by several authors.[4–6]

In this paper we show that, for functions having a complex basis and satisfying either classical or quantum bracket identities, eigenvalues are easily found from solutions of a divergence equation, which is a form of the Cauchy–Riemann equations, and thus provide a necessary condition for analyticity. In this case the eigenvalue problem is reduced to that of finding the analytic regions of the function.

1. EIGENVALUE EQUATION

We consider a function of $P(z) = u(p, q) + iv(p, q)$ of a single complex variable $z = p + iq$. The Poisson brackets read

$$[z, p] = [p, z^*] = i, \tag{1}$$

$$[z, q] = -[q, z^*] = -1, \tag{2}$$

$$[q, P] = \frac{\partial P}{\partial p}, \quad [p, P] = -\frac{\partial P}{\partial q}, \tag{3}$$

so that

$$[z, P] = \frac{\partial P}{\partial q} + i\frac{\partial P}{\partial p} = i\nabla P = 2i\frac{\partial P}{\partial z^*}, \tag{4}$$

where

$$\nabla = \frac{\partial}{\partial p} + i\frac{\partial}{\partial q}.$$

Also

$$[z^*, P^*] = -2i\frac{\partial P^*}{\partial z}. \tag{5}$$

* Permanent address: Box 94, La Jolla, California.

[1] J. T. Anderson, J. Math. Phys. **8**, 998 (1967).
[2] J. T. Anderson, Nuovo Cimento **52**, 375 (1967).
[3] V. S. Popov and E. I. Dolinsky, Zh. Eksp. Teor. Fiz. **46**, 970 1829 (1963) [Sov. Phys.—JETP **19**, 661, 1232 (1964)].
[4] E. G. Beltrami and G. Luzzato, Nuovo Cimento **51**, 147 (1967).
[5] J. Gunson, J. Math. Phys. **6**, 852 (1965).
[6] V. N. Gribov and I. Ya. Pomeranchuk, Phys. Rev. Letters **8**, 343 (1962).

Thus, for complex functions, the fundamental brackets can be replaced by a simple variational function. The transition to quantum systems is straightforward. With

$$[z, P]_{\mathrm{cl}} = (-i/h)[z, P]_{\mathrm{qm}}, \tag{6}$$

the quantum brackets (qm) give the same equations multiplied by a constant as do the classical brackets (cl); we write for both classical and quantum commutators,

$$[z, z^*] = \sigma_z. \tag{7}$$

Note that the quantum equivalent of (4) is obtained in momentum representation by replacing z by $2ih\partial/\partial z^*$. Also

$$[z, z] = [z^*, z^*] = 0. \tag{8}$$

Now (4) equates the commutator of z with P to the variation of P with respect to the conjugate variable, so that, for example, the constants of motion of a dynamical system, or more generally, the eigenvalues belonging to P, are found from the solutions of the differential equation $\partial P(z)/\partial z^* = 0$. But this equation is just the Cauchy–Riemann equations, so that, provided P is differentiable, the eigenvalues of (4) are defined by the analytic regions of P on the z plane. If we take for P the Hamiltonian of a dynamical system, then $\partial H/\partial z^* = 0$ resembles the minimization condition in the familiar variational method for real variables. The variational method, however, yields an approximation to the eigenvalues, whereas for complex variables (4) is exact.

From these considerations we obtain the following:

Theorem 1: If the z' are solutions of

$$\lim_{z \to z'} \frac{\partial P}{\partial z^*} \equiv 0 \tag{9}$$

and $P(z)$ is differentiable and real on the real axis, then $P(z')$ is analytic at z', and z' are eigenvalues of P satisfying

$$[z', P(z')] = 0. \tag{10}$$

The converse gives the following corollary:

Corollary 1: If there exist z' satisfying (10) and $P(z')$ is differentiable and real on the real axis, then (9) holds and the z' define an analytic region of $P(z)$.

It can now be shown that if P is analytic in z, then P^* is necessarily analytic. Thus, we obtain a second corollary:

Corollary 2: If Theorem 1 holds for $P(z)$, then

$$\lim_{z \to z'} \frac{\partial P^*}{\partial z} \equiv 0, \tag{11}$$

and $P^*(z')$ is analytic in a region including z', and $z^{*\prime}$ are eigenvalues satisfying

$$[z^{*\prime}, P^*(z')] = 0. \tag{12}$$

Proof: With operators such as those in (4) and (5), one easily obtains

$$\frac{\partial P}{\partial z^*} = \frac{\partial P^*}{\partial z} + 2i \operatorname{Im} \frac{\partial P}{\partial z^*}. \tag{13}$$

If there are z' for which (13) vanishes and Theorem 1 holds, then the rhs vanishes and $\partial P^*(z')/\partial z' = 0$ as well. Q.E.D.

Equation (9) gives an exact relation between the eigenvalues z' and the analytic regions of P, so that, in effect, the eigenvalue problem is reducible to that of finding the analytic regions, or of finding the solutions of the differential equation (9). In general there will be a finite number of solutions representing points around which P may be expanded in powers of $z' - z$, i.e., in a Laurent series. The set of such points is the analytic region on the z plane, and it is defined by the solutions of (9) and the requirement that P be differentiable.

2. UNITARY TRANSFORMATIONS

We can now utilize Theorem 1 to obtain the condition that a transformation be unitary on the complex plane. Consider the transformation

$$P^\dagger = e^z P e^{-z} = P + \sigma_z \frac{\partial P}{\partial z^*} + \sigma_z^2 \frac{\partial^2 P}{\partial z^{*2}} + \cdots. \tag{14}$$

With Theorem 1 we now obtain the following:

Theorem 2: If $P(z)$ satisfies Theorem 1, then

$$P^\dagger = e^z P e^{-z} \tag{15}$$

is a unitary transformation on the analytic region including z'.

Proof: If P is analytic, the second-order bracket on the rhs of (14) becomes

$$[z, [z, P]] = \sigma_z \left[z, \frac{\partial P}{\partial z^*} \right]$$

$$= \sigma_z \left[z \frac{\partial P}{\partial z^*} - \frac{\partial}{\partial z^*} zP \right] \tag{16}$$

$$\equiv 0. \tag{17}$$

Hence, also

$$[z, [z, [z, P]]] = 0, \tag{18}$$

and all higher-order commutators vanish identically.

A similar relation

$$P^{*\dagger} = e^{z^*} P^* e^{-z^*} \tag{19}$$

can be proved for P^*. However, the two remaining possible transformations of P on the z^* semiplane and P^* on the z semiplane cannot be unitary, unless $P =$ constant; i.e., if $\partial P/\partial z^* = 0$, then $\partial P/\partial z \neq 0$, and if $\partial P^*/\partial z = 0$, then $\partial P^*/\partial z^* \neq 0$. Hence the only unitary transformations occur with P on the upper semiplane and with P^* on the lower semiplane.

3. SEVERAL COMPLEX VARIABLES

Theorems 1 and 2 may be extended to include P as a function of several complex variables by everywhere replacing the partial derivatives with total derivatives. Then (4) becomes for $P(x, y, z, \cdots)$

$$[z, P] = \sigma_z \frac{d}{dz^*} P(x, y, z, \cdots)$$
$$= \sigma_z \left(\frac{\partial P}{\partial z^*} + \frac{\partial P}{\partial y^*} \frac{dy^*}{dz^*} + \frac{\partial P}{\partial x^*} \frac{dx^*}{dz^*} + \cdots \right), \tag{20}$$

and there is a similar relation replacing (5). Thus (20) vanishes only if P satisfies the Cauchy–Riemann equations in every variable, and then all higher-order brackets vanish as well. Similarly, (14) becomes

$$e^{x+y+z} P e^{-x-y-z} = P + \sigma_x \frac{dP}{dx^*} + \sigma_y \frac{dP}{dy^*} + \sigma_z \frac{dP}{dz^*}$$
$$+ \sigma_y \sigma_z \frac{d}{dy^*} \frac{dP}{dz^*} + \cdots$$
$$+ \sigma_x \sigma_y \sigma_z \frac{d}{dx^*} \frac{d}{dy^*} \frac{dP}{dz^*} + \cdots . \tag{21}$$

Theorem 2 holds also for any number of complex variables, regardless of whether the $x, y, z, \cdots$ are independent or not. This may be verified by writing out the higher order terms in (21) and noting that all are of the form (17) and (18) with one of the z variables replaced by x or y; after expanding the second-order terms, one obtains a sum of commutators involving only products of scalars and derivatives of scalars, and since the scalars commute, the commutators all vanish identically. Thus we obtain the following theorem:

Theorem 3: If $P(x, y, z, \cdots)$ is an analytic function of $x, y, z, \cdots$, then

$$P^\dagger = e^{x+y+z\cdots} P e^{-x-y-z\cdots}$$
$$= P \tag{22}$$

is a unitary transformation on every plane.

Proof: One second-order term for two variables reads

$$\lim_{y \to y'} \lim_{x \to x'} [y, [z, P]] = [y', 0]$$
$$\equiv 0, \tag{23}$$

and similar relations are obtained for the remaining variables. From these relations, one obtains

$$\frac{\partial y'}{\partial z^{*\prime}} = \frac{\partial x'}{\partial z^{*\prime}} = \cdots = 0, \tag{24}$$

so that if P is analytic in $x, y, z, \cdots$, then $x, y, z, \cdots$ are analytic functions of each other.

An alternative proof may be given by using the sufficient condition for analyticity; consider

$$[y', [z', P]] = [z', [y', P]], \tag{25}$$

which holds because of the interchangeability of operators in an analytic region. Expanding both sides of (25) gives

$$P[z', y'] + [y', z']P = [P, [z', y']]$$
$$= -\sigma \frac{\partial P}{\partial (dy^{*\prime}/dz')}$$
$$\equiv 0, \tag{26}$$

which is the requirement that P satisfy the sufficient condition for analyticity in y and z. Thus (23) satisfies the sufficient condition as well.

4. ANALYTICITY OF HIGHER-ORDER FUNCTIONS

In this section we utilize Theorems 1–3 to prove that ordinary algebraic functionals of analytic functions are also analytic. Let $Q, R, S, \cdots$ be functions of z; then we obtain the following theorem:

Theorem 4: If $P = QRS \cdots$ and $Q, R, S, \cdots$ are analytic functions of z, then P is also analytic.

Proof: The necessary condition that P be analytic is

$$\frac{\partial P}{\partial z^*} \equiv QR \frac{\partial S}{\partial z^*} + SQ \frac{\partial R}{\partial z^*} + RS \frac{\partial Q}{\partial z^*} + \cdots$$
$$= 0. \tag{27}$$

By Theorem 1, if

$$\frac{\partial Q}{\partial z^*} = \frac{\partial R}{\partial z^*} = \frac{\partial S}{\partial z^*} = \cdots = 0, \tag{28}$$

and if $Q, R, S, \cdots$ are differentiable, then $Q, R, S, \cdots$ are analytic, and it follows that P is analytic as well.

The proofs of functions of the form

$$P = \frac{Q}{R}\frac{S}{T}\cdots,$$

$$P = Q + R + S + \cdots, \tag{29}$$

and

$$P = Q^R$$

follow in the same way. Thus, any function expressible as a convergent power series in z is analytic, and the inverse of an analytic function is also analytic. Thus, if $\partial P/\partial z^* = 0$, then also $\partial z^*/\partial P = 0$, except when all partial derivatives vanish identically, i.e., when $P =$ const. In this way, it can be shown that some logarithmic functions are also analytic. For example, let

$$Q = e^z. \tag{30}$$

Then

$$\frac{\partial \log Q}{\partial z^*} = 0. \tag{31}$$

Thus, the usual illustration of e^{iz} as a nonanalytic function is not valid if z is a complex variable; i.e., the many-valuedness of e^{iz} requires that z be real, whereas for complex z, e^{iz} describes a spiral rather than a circular function.

5. ANALYTICITY AND CONSERVATION LAWS

From the preceding sections, it is evident that the eigenvalues of an operator are related to the analytic regions of the operator on the complex plane. Hence we expect that, for a complex variable representation, the analytic regions correspond to conserved or constant regions in a real variable representation; that is, the analyticity requirement corresponds to a conservation law. This result is well known in potential theory and in fluid dynamics problems involving constant sources; in the former case analyticity corresponds to charge conservation for electrostatic fields, and in the latter case to conservation of mass. Evidently, then, any conserved function in a real variable representation corresponds to an analytic function in a complex variable representation.

Now, define the minimal electromagnetic current j_α by

$$j_\alpha = \frac{\delta \mathfrak{L}}{\delta \partial \bar{\varphi}_\alpha}, \tag{32}$$

where $\varphi_\alpha = \varphi(q_\alpha)$ is a function of the four-momentum q_α, and $\mathfrak{L} = \mathfrak{L}(\varphi, \partial\varphi)$ is defined on the upper semiplanes. Now writing

$$\mathfrak{L}^\dagger = e^{q_\alpha}\mathfrak{L}e^{-q_\alpha}$$

$$= \mathfrak{L} + [q_\alpha, \mathfrak{L}] + \cdots, \tag{33}$$

where q_α is a sum of normalized complex scalars, it can be shown that $\mathfrak{L}$ is invariant under the transformation, provided the Lagrangian is analytic in q_α, that is, provided energy and momentum are conserved, and that if $\mathfrak{L}$ is analytic, then the Lagrangian is also invariant under a rotation through π on each semiplane carrying $\mathfrak{L}$ into the continued function $\mathfrak{L}^\dagger$ on the lower semiplanes. The complete proof is given in a later paragraph.

If, in the case of electromagnetic interactions, $\mathfrak{L}$ is invariant under a gauge transformation of the first kind, then

$$\frac{\partial \mathfrak{L}}{\partial \bar{\varphi}_\alpha} = \partial j_\alpha = 0, \tag{34}$$

which is equivalent to conservation of charge. Evidently this conservation law can be related to the requirement that the Lagrangian be an analytic function by means of the unitary transformation

$$\mathfrak{L}^\dagger = e^{\varphi_\alpha}\mathfrak{L}e^{-\varphi_\alpha}$$

$$= \mathfrak{L} + \sigma_\varphi \frac{\partial \mathfrak{L}}{\partial \bar{\varphi}_\alpha} + \cdots$$

$$= \mathfrak{L}, \tag{35}$$

so that charge conservation is equivalent to analyticity in φ_α. Conversely, if the current is not conserved, then the splitting is given exactly by the second term on the rhs in (35) since

$$\frac{\partial \mathfrak{L}}{\partial \bar{\varphi}_\alpha} = -\delta m^2 \varphi_\alpha$$

$$= (1/\sigma_\varphi)\partial j_\alpha. \tag{36}$$

This procedure may be generalized to include Lagrangian functions of an arbitrary number of variables and with a conservation law corresponding to analyticity in each variable. Thus, if $\varphi_{\alpha\beta}$ are complex bases satisfying the usual bracket relations, then

$$e^{\varphi_{\alpha\beta}}\mathfrak{L}e^{-\varphi_{\alpha\beta}} = \mathfrak{L} + \sigma_{\alpha\beta}\frac{\partial \mathfrak{L}}{\partial \bar{\varphi}_{\alpha\beta}} + \cdots$$

$$= \mathfrak{L} + \sigma_{\alpha\beta}\partial j_{\alpha\beta} + \cdots. \tag{37}$$

The higher-order terms are related to the mass splittings by

$$\sigma_{\alpha\beta}\frac{\partial \mathfrak{L}}{\partial \bar{\varphi}_{\alpha\beta}} = -\delta m_\beta^2 \varphi_{\alpha\beta}. \tag{38}$$

Conversely, if $\mathfrak{L}$ is analytic in a variable $\varphi_{\alpha\beta}$, then all commutators involving $\varphi_{\alpha\beta}$ vanish identically.

We now prove that analyticity is equivalent to rotational invariance. Consider the transformation

$$\mathfrak{L}^\dagger = \bar{\varphi}[\varphi, \mathfrak{L}]$$

$$= \sigma_\varphi \bar{\varphi}\frac{\partial \mathfrak{L}}{\partial \bar{\varphi}}$$

$$= \sigma_\varphi[S_\varphi + iL_\varphi]\mathfrak{L}, \tag{39}$$

and with $\varphi = x + iy$,

$$L_\varphi = \mathrm{Im}\left(\bar{\varphi}\frac{\partial}{\partial\bar{\varphi}}\right) = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \tag{40}$$

is the usual rotation operator, and

$$S_\varphi = \mathrm{Re}\left(\bar{\varphi}\frac{\partial}{\partial\bar{\varphi}}\right) = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} \tag{41}$$

measures any radial deformation under the rotation. (We omit the indices of the dimensionality but a sum of terms is implied.) If one makes the usual assumption of rigid body rotations, then (39) gives the familiar rotation operator acting on $\mathfrak{L}$. If $L_\varphi\mathfrak{L}$ vanishes, $\mathfrak{L}$ is invariant under rotations in the φ plane, and it is evident that this condition and the rigid body requirement are both satisfied, provided $\mathfrak{L}$ is analytic in φ, or

$$\frac{\partial\mathfrak{L}}{\partial\bar{\varphi}} = 0, \tag{42}$$

since both real and imaginary parts of (39) vanish under this condition. This can be shown by writing out the components of (39), with $\mathfrak{L} = u + iv$,

$$\bar{\varphi}\frac{\partial\mathfrak{L}}{\partial\bar{\varphi}} = \frac{1}{2}\left\{x\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + y\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) + i\left[x\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) - y\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)\right]\right\} \tag{43}$$

$$= \tfrac{1}{2}[S_\varphi u - L_\varphi v + i(L_\varphi u + S_\varphi v)], \tag{44}$$

which vanishes identically, provided the Cauchy–Riemann equations are satisfied.

6. RELATIVISTIC SELF-CONSISTENT FIELDS

We now apply the methods developed in the previous sections to the solution of the coupled equations of motion of spinor and scalar fields. We consider a Lagrangian $\mathfrak{L}(\varphi, \psi, \nabla\varphi, \nabla\psi)$ with equations of motion

$$\frac{\partial\mathfrak{L}}{\partial\varphi} = 0 \quad \text{and} \quad \frac{\partial\mathfrak{L}}{\partial(\overline{\nabla\psi})} = 0. \tag{45}$$

After the Fourier transformation, consider the analytic continuations of (45) and $\mathfrak{L}$. The requirement that $\mathfrak{L}$ conserve energy and momentum, as was shown in the previous section, is equivalent to the requirement that $\mathfrak{L}$ be invariant under rotations on the four-momentum planes. Thus we require

$$\lim_{q_\alpha \to q_{\alpha'}} [q_\alpha, \mathfrak{L}] = \sigma_{q_{\alpha'}}\frac{d\mathfrak{L}}{d\bar{q}'_\alpha} = \sigma_{q_{\alpha'}}\left[\frac{\partial\mathfrak{L}}{\partial\bar{\varphi}}\frac{d\bar{\varphi}}{d\bar{q}'_\alpha} + \frac{\partial\mathfrak{L}}{\partial\bar{\psi}}\frac{d\bar{\psi}}{d\bar{q}'_\alpha}\right] + \text{terms which vanish identically by (45)} = 0. \tag{46}$$

If (46) holds, then all higher-order terms in (33) vanish according to (15) and (17). Hence the solutions of (46) determine the eigenvalues q'_α for which $\mathfrak{L}$ is a constant. As will become evident, Eqs. (45) and (46) are the only possible conditions for which one obtains nontrivial solutions.

The rhs of (46) must vanish independently for each variation with respect to q_α since, in general, the variations with respect to the four-momentum do not vanish. Now unless $\mathfrak{L}$ is a constant, it is evident from (45) that $\partial\mathfrak{L}/\partial\bar{\varphi} \neq 0$, so we obtain the condition

$$\left(\frac{\overline{d\varphi}}{dq'_\alpha}\right) = 0. \tag{47}$$

Then either

$$\frac{\partial\mathfrak{L}}{\partial\bar{\psi}} = 0 \quad \text{or} \quad \frac{d\bar{\psi}}{\partial\bar{q}'_\alpha} = 0. \tag{48}$$

From general considerations of the equation of motion of the Lagrangian, one can show that the first of Eqs. (48) must hold. Thus, from

$$\frac{\partial\mathfrak{L}}{\partial(\overline{\nabla\psi})} = \int dx\,\frac{\partial\mathfrak{L}}{\partial\bar{\psi}} \tag{49}$$

and (45), one obtains the first of Eqs. (48). Hence the conditions that (46) hold are given by (47) and

$$\frac{\partial\mathfrak{L}}{\partial\bar{\psi}} = 0. \tag{50}$$

We now treat the case of scalar and spinor fields. Consider the Lagrangian

$$\mathfrak{L} = \tfrac{1}{2}(\nabla\varphi)^2 - \tfrac{1}{2}m^2\varphi^2 + \bar{\psi}\nabla\psi - \bar{\psi}M\psi + g\bar{\psi}\varphi\psi. \tag{51}$$

The variations of (51) give

$$(\Box - m^2)\varphi = -g\bar{\psi}\psi \tag{52}$$

and

$$(\nabla - M + g\varphi)\psi = 0. \tag{53}$$

In momentum representation, these equations read

$$\varphi = -ig\Gamma\bar{\psi}\psi + \text{const}, \tag{54}$$

where $\Gamma = (q^2 - m^2)^{-1}$, and with $k = p + q$, we have

$$(\not{k} - M + g\varphi)\psi = 0 \tag{55}$$

since φ must be independent of or a constant function of q'_α by (47). Now (47) gives

$$\frac{d\varphi}{dq'_\alpha} = -g\Gamma\bar{\psi}_\alpha\left(-2q'_\alpha\Gamma\psi_\alpha + \frac{d\psi_\alpha}{dq'_\alpha}\right) = 0, \tag{56}$$

since (47) requires also that $d\varphi/dq'_\alpha = 0$. Taking q'_α as a column vector, solutions of (56) read

$$\psi_\alpha = \frac{\Gamma_0}{\Gamma}\psi_{0\alpha} = \frac{q'^2 - m^2}{M^2 - m^2}\psi_{0\alpha}, \tag{57}$$

and $\Gamma_0 = (p^2 - m^2)^{-1}$ for a free field. In obtaining (57), the integrations have been carried out from $q_\alpha = p_\alpha = 0$, $\omega' = \omega_0$, to q' in an analytic region, so that the path need not be specified; that ψ is analytic in q_α follows from $\partial\bar{\psi}/\partial\bar{q}'_\alpha \neq 0$ [see (48) and (50)], and the integrand over dq_α is analytic by Theorem 4 since the component functions are all analytic, as can be verified by writing the Cauchy–Riemann equations for each.

Setting the constant in (54) to zero, we find

$$\varphi = -ig\frac{\bar{q}'^2 - m^2}{(M^2 - m^2)^2}|\psi_0|^2, \tag{58}$$

where

$$|\psi_0|^2 = \sum_\alpha \bar{\psi}_{0\alpha}\psi_{0\alpha}.$$

Equation (55) now reads

$$[\not{k} - M - \rho(\bar{q}'^2 - m^2)]\psi = 0, \tag{59}$$

where

$$\rho = ig^2|\psi_0|^2/(M^2 - m^2)^2. \tag{60}$$

Solving (59) for motion in the xy plane (see Appendix A), one obtains

$$\bar{q}'^2 = m^2 + M(M^2 - m^2)^2/ig^2|\psi_0|^2. \tag{61}$$

The sign of the interaction term in (61) is fixed by convergence requirements [see (68) and (70)]. It is easily verified that (61) is an analytic function of the four-momentum and of energy and space momentum, separately. With $|\psi_0|^2 = 2M$, we find q'^2 is a constant depending only upon the masses and the coupling strength.

The mass splitting in (61) is symmetric (there is another component with negative sign for the upper semiplane), and the mass difference is a function only of the coupling strength and the initial masses so that the coupling strength can be found provided one knows the mass difference. Writing $(\Delta m)^2$ for the third term in (61),

$$g^2 = (M^2 - m^2)^2/2(\Delta m)^2. \tag{62}$$

Finally, (57) and (61) yield

$$\psi_\alpha = [(M^2 - m^2)/2ig^2]\psi_{0\alpha} \tag{63}$$

and

$$\psi = -M/g. \tag{64}$$

The relation between the total wavefunction $|\psi|^2$ and φ is

$$|\psi|^2 = -[(M^2 - m^2)^2/2g^3]\varphi, \tag{65}$$

so that φ can be regarded as an effective field or average over the particle field components, multiplied by a coupling factor. The fact that φ is a constant shows a strong correlation among the particle fields so that the latter may, in fact, be written as the product of φ, $\psi_{0\alpha}$, and an effective coupling strength:

$$\psi_\alpha = (i\varphi/2gM)(M^2 - m^2)\psi_{0\alpha}. \tag{66}$$

Thus, the effective scalar field φ has no structure or short-range properties and is, in fact, an infinite-range field proportional to the average particle field.

The Lagrangian of (64) and (66) is a function only of the masses and the coupling strength

$$\mathfrak{L} = \frac{(M^2 - m^2)^3}{4Mg^4}\left\{iM + \frac{1}{M}\bar{\psi}_{0\alpha}(\not{k} - 2M)\psi_{0\alpha}\right\}, \tag{67}$$

and so is completely determined by the self-consistent solutions of the equations of motion.

In the limit $\psi \to \psi_0$, consistent limits of (61) and (63) are obtained if $g^2 \to (M^2 - m^2)/2i$ as the coupling is reduced to zero. Then (64) becomes

$$\lim_{m^2 \to M^2}\varphi = \sum_{\pm}\frac{\pm(2i)^{\frac{1}{2}}}{(M^2 - m^2)^{\frac{1}{2}}} = 0. \tag{68}$$

Equation (61) may be written in terms of a single complex variable as follows: With $z = u + iv$ and resolving (61) with $q_\alpha = u_\alpha + iv_\alpha$, provided g is real,

$$u'^2 - v'^2 = \sum_\alpha (-1)^\beta(u'^2_\alpha - v'^2_\alpha) = m^2, \quad \begin{matrix}\beta = 0, & \alpha = 1,\\ = 1, & \alpha = 2, 3,\end{matrix} \tag{69}$$

and

$$2u'v' = 2\sum_\alpha (-1)^\beta u'_\alpha v'_\alpha = \frac{(M^2 - m^2)^2}{2g^2}, \tag{70}$$

or

$$u'^2 = \tfrac{1}{2}m^2\{1 + [1 + (\Delta m/m)^4]^{\frac{1}{2}}\} \tag{71}$$

and

$$v'^2 = -\tfrac{1}{2}m^2\{1 - [1 + (\Delta m/m)^4]^{\frac{1}{2}}\}. \tag{72}$$

With $q_1 = E = \omega + i\gamma$, $q_\alpha = \pi_\alpha + i\kappa_\alpha$ ($\alpha = 2, 3$), it is not difficult to show that

$$\omega'^2 - \sum_\alpha \pi_\alpha'^2 = u'^2, \quad \gamma'^2 - \sum_\alpha \kappa_\alpha'^2 = v'^2, \tag{73}$$

and the requirement that $m^2 + (\Delta m)^2$ be analytic in the four-momentum gives the first equality in

$$\frac{\omega'^2}{\gamma'^2} = \frac{\sum \pi_\alpha'^2}{\sum \kappa_\alpha'^2} = \frac{u'^2}{v'^2}, \tag{74}$$

and the second is then obtained from (73). The energy and decay time can now be obtained from

$$\omega'^2 = \sum \pi_\alpha'^2 + \tfrac{1}{2}m^2\{1 + [1 + (\Delta m/m)^4]^{\frac{1}{2}}\} \tag{75}$$

and

$$\tau'^2 = \frac{1}{\sum \kappa_\alpha'^2 - \tfrac{1}{2}m^2\{1 - [1 + (\Delta m/m)^4]^{\frac{1}{2}}\}}, \tag{76}$$

and the interaction range can be obtained from

$$\sum r_\alpha'^2 = -\frac{1}{\sum \pi_\alpha'^2}\frac{1 + [1 + (\Delta m/m)^4]^{\frac{1}{2}}}{1 - [1 + (\Delta m/m)^4]^{\frac{1}{2}}}. \tag{77}$$

For $\Delta m/m \ll 1$, $r' \sim 2^{\frac{1}{2}}(m/\Delta m)^2 r_0$ so that the interaction range varies inversely with the mass splitting; but for $\Delta m/m > 1$, r' approaches a constant r_0 in the limit.

The zero coupling limits of (75)–(77) give $\pi_\alpha' = p_\alpha$, $m = M$ and $\omega' = \omega_0$, with $\gamma_\alpha' = 0 = \kappa_\alpha'$ and reduce to $\omega' = \omega_0 = M$ at rest. The initial values obtained in this limit represent a particle and field of identical mass and velocity either moving or at rest, so that the coupling is turned on only when the particle emits the field, i.e., when differential motion begins. We find then that the renormalized field momentum becomes

$$\pi_\alpha' = -\frac{u'}{u' - 2M}p_\alpha, \tag{78}$$

by means of which ω', τ', and κ' may all be determined as functions of the initial masses and the coupling strength.

APPENDIX A: SOLUTION OF THE DIRAC EQUATION

For motion in the xy plane, Eq. (59) reads

$$\begin{vmatrix} \mathcal{E}'-[M+\rho(\bar{q}'^2-m^2)] & 0 & k_2'+ik_3' & 0 \\ 0 & \mathcal{E}'-[M+\rho(\bar{q}'^2-m^2)] & k_2'-ik_3' & 0 \\ -k_2'+ik_3' & 0 & -\mathcal{E}'-[M+\rho(\bar{q}'^2-m^2)] & 0 \\ 0 & -k_2'+ik_3' & 0 & -\mathcal{E}'-[M+\rho(\bar{q}'^2-m^2)] \end{vmatrix} \begin{vmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{vmatrix} = 0, \tag{A1}$$

which gives

$$M + \rho(\bar{q}'^2 - m^2)]^2 = \mathcal{E}'^2 - \sum_\alpha k_\alpha'^2. \tag{A2}$$

Writing $\mathcal{M}^2$ for the rhs, we find

$$\operatorname{Re}\mathcal{M}^2 = -\sum(p_\alpha + \pi_\alpha')^2 + (\omega_0 + \omega')^2 - \gamma'^2 + \sum \kappa_\alpha'^2, \tag{A3}$$

$$\operatorname{Im}\mathcal{M}^2 = -2\sum(p_\alpha + \pi_\alpha')\kappa_\alpha' + 2\gamma'(\omega_0 + \omega'). \tag{A4}$$

We require that (A2) be an analytic function of the momentum, energy, and four-momentum, since the lhs of (A2) is obviously analytic in those variables. Thus, for

$$\sum \frac{\partial \operatorname{Re}\mathcal{M}^2}{\partial \pi_\alpha'} = -2\sum(p_\alpha + \pi_\alpha') - 2\gamma'\sum\frac{\partial \gamma'}{\partial \pi_\alpha'} + 2(\omega_0 + \omega')\sum\frac{\partial \omega'}{\partial \pi_\alpha'} \tag{A5}$$

and

$$\sum \frac{\partial \operatorname{Im}\mathcal{M}^2}{\partial \kappa_\alpha'} = -2\sum(p_\alpha + \pi_\alpha') + 2\gamma'\sum\frac{\partial \omega'}{\partial \kappa_\alpha'} + 2(\omega_0 + \omega')\sum\frac{\partial \gamma'}{\partial \kappa_\alpha'}, \tag{A6}$$

we require (A5) = (A6), and

$$\sum \frac{\partial \operatorname{Re}\mathcal{M}^2}{\partial \kappa_\alpha'} = 2\sum \kappa_\alpha' - 2\gamma'\sum\frac{\partial \gamma'}{\partial \kappa_\alpha'} + 2(\omega_0 + \omega')\sum\frac{\partial \omega'}{\partial \kappa_\alpha'} \tag{A7}$$

and

$$\sum \frac{\partial \operatorname{Im}\mathcal{M}^2}{\partial \pi_\alpha'} = -2\sum \kappa_\alpha' + 2\gamma'\sum\frac{\partial \omega'}{\partial \pi_\alpha'} + 2(\omega_0 + \omega')\sum\frac{\partial \gamma'}{\partial \pi_\alpha'} \tag{A8}$$

must give (A7) = −(A8). Then (A5)–(A8) give

$$\frac{\gamma'}{\omega_0 + \omega'} = \sum\left(\frac{\partial \omega'}{\partial \pi_\alpha'} - \frac{\partial \gamma'}{\partial \kappa_\alpha'}\right)\Big/\sum\left(\frac{\partial \gamma'}{\partial \pi_\alpha'} + \frac{\partial \omega'}{\partial \kappa_\alpha'}\right) = \sum\left(\frac{\partial \gamma'}{\partial \pi_\alpha'} + \frac{\partial \omega'}{\partial \kappa_\alpha'}\right)\Big/\sum\left(\frac{\partial \gamma'}{\partial \kappa_\alpha'} - \frac{\partial \omega'}{\partial \pi_\alpha'}\right), \tag{A9}$$

which is the necessary condition that $\mathcal{M}^2$ be analytic

in the momentum. For $\mathcal{M}^2$ to be analytic in the energy,

$$\sum \frac{\partial \operatorname{Re} \mathcal{M}^2}{\partial \gamma'} = -2 \sum (p_\alpha + \pi'_\alpha) \frac{\partial \pi'_\alpha}{\partial \gamma'} + 2 \sum \kappa'_\alpha \frac{\partial \kappa'_\alpha}{\partial \gamma'} - 2\gamma' \quad \text{(A10)}$$

and

$$\sum \frac{\partial \operatorname{Im} \mathcal{M}^2}{\partial \omega'} = -2 \sum (p_\alpha + \pi'_\alpha) \frac{\partial \kappa'_\alpha}{\partial \omega'} - 2 \sum \kappa'_\alpha \frac{\partial \pi'_\alpha}{\partial \omega'} + 2\gamma' \quad \text{(A11)}$$

must give (A10) = −(A11), and

$$\sum \frac{\partial \operatorname{Re} \mathcal{M}^2}{\partial \omega'} = -2 \sum (p_\alpha + \pi'_\alpha) \frac{\partial \pi'_\alpha}{\partial \omega'} + 2(\omega_0 + \omega') + 2 \sum \kappa'_\alpha \frac{\partial \kappa'_\alpha}{\partial \omega'} \quad \text{(A12)}$$

and

$$\sum \frac{\partial \operatorname{Im} \mathcal{M}^2}{\partial \gamma'} = -2 \sum (p_\alpha + \pi'_\alpha) \frac{\partial \kappa'_\alpha}{\partial \gamma'} + 2(\omega_0 + \omega') - 2 \sum \kappa'_\alpha \frac{\partial \pi'_\alpha}{\partial \gamma'} \quad \text{(A13)}$$

must give (A12) = +(A13). Then (A10)–(A13) give

$$\frac{\kappa'_\alpha}{p_\alpha + \pi'_\alpha} = \left(\frac{\partial \pi'_\alpha}{\partial \omega'} - \frac{\partial \kappa'_\alpha}{\partial \gamma'}\right) \Big/ \left(\frac{\partial \kappa'_\alpha}{\partial \omega'} + \frac{\partial \pi'_\alpha}{\partial \gamma'}\right) = \left(\frac{\partial \pi'_\alpha}{\partial \gamma'} + \frac{\partial \kappa'_\alpha}{\partial \omega'}\right) \Big/ \left(\frac{\partial \kappa'_\alpha}{\partial \gamma'} - \frac{\partial \pi'_\alpha}{\partial \omega'}\right), \quad \text{(A14)}$$

since the sum of all cross products vanishes.

It is a simple matter now to show that $\mathcal{M}^2$ is analytic in the four-momentum, provided

$$\sum \frac{p_\alpha + \pi'_\alpha}{\omega_0 + \omega'} = \sum \frac{\kappa'_\alpha}{\gamma'}. \quad \text{(A15)}$$

The same result is also obtained by requiring that $\operatorname{Re} \mathcal{M}^2$ and $\operatorname{Im} \mathcal{M}^2$ be consistent; that is, $\operatorname{Re} \mathcal{M}^2 = U^2 - V^2$, $\operatorname{Im} \mathcal{M}^2 = 2UV$ gives

$$U^2V^2 = [\gamma'(\omega_0 + \omega') - \sum (p_\alpha + \pi'_\alpha)\kappa'_\alpha]^2 = |(\omega_0 + \omega')^2 + \sum (p_\alpha + \pi'_\alpha)^2][\gamma'^2_1 + \sum \kappa'^2_\alpha], \quad \text{(A16)}$$

which gives (A15) since all sums of products of the form $x_\alpha y_\beta$ must vanish.

With (A15) we get

$$\operatorname{Re} \mathcal{M}^2 = \sum \left(1 - \frac{\kappa'^2_\alpha}{\gamma'^2}\right)[(\omega_0 + \omega')^2 - \gamma'^2] \quad \text{(A17)}$$

and

$$\operatorname{Im} \mathcal{M}^2 = 2 \sum \gamma'(\omega_0 + \omega')\left(1 - \frac{\kappa'^2_\alpha}{\gamma'^2}\right), \quad \text{(A18)}$$

so that, with $\mathcal{E}' = \omega_0 + \omega' + i\gamma'$,

$$\mathcal{M}^2 = \mathcal{E}'^2 - \sum \mathcal{E}'^2 \frac{\kappa'^2_\alpha}{\gamma'^2} = \mathcal{E}'^2 - \sum (p_\alpha + \pi'_\alpha)^2. \quad \text{(A19)}$$

The rest mass of the excited state can be determined in the limit $\pi'_\alpha = p_\alpha = 0$, $\omega' = \omega_0$, which gives $\mathcal{M}^2 = 4M^2$, and the same result is obtained for the dynamical case $\pi'_\alpha = p_\alpha$, so that, finally, (A2) gives with (A19)

$$\bar{q}'^2 - m^2 = M/\rho. \quad \text{(A20)}$$

The behavior of π'_α in the zero coupling limit can be studied by using (74), which follows directly from the requirement that q'^2 be an analytic function of the four-momentum. With (73), (74), and (A15), we obtain

$$\sum (p_\alpha + \pi'_\alpha)^2 = \frac{\mathcal{M}^2}{u'^2} \sum \pi'^2_\alpha, \quad \text{(A21)}$$

which gives

$$\pi'_\alpha = -\frac{u'}{u' - 2M} p_\alpha. \quad \text{(A22)}$$

In the limit $g^2 \to (M^2 - m^2) \to 0$, $u \to m \to M$ and

$$\lim_{g^2 \to 0} \pi'_\alpha = p_\alpha. \quad \text{(A23)}$$

Then (75) gives $\omega' = \omega_0$ in agreement with the integration limit in (57), and (59) reduces to the Dirac equation for a noninteracting particle.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Degeneracy of the Dirac Equation with Electric and Magnetic Coulomb Potentials*

MANUEL BERRONDO† AND HAROLD V. MCINTOSH‡
Quantum Chemistry Group, Uppsala University, Box 518, 751-20 *Uppsala, Sweden*

(Received 10 February 1969)

Investigation is made of the symmetry and degeneracy of the Dirac equation for a Coulomb potential with a fixed center bearing both electric and magnetic charge. Seen from the viewpoint of classical mechanics, relativistic precession removes the accidental degeneracy of the nonrelativistic potential, and may be so severe as to lead to spiral rather than precessing elliptic orbits. The degeneracy may be restored by the introduction of a vector potential which combats the precession and leads to closed relativistic orbits. An angular momentum and a Runge vector are found for the "symmetric" potential for arbitrary values of electric and magnetic nuclear charges. A related symmetric Hamiltonian and constants of the motion may be constructed for the Dirac equation, which reduce to those of Biedenharn and Swamy in the absence of magnetic charge. Magnetic charge must be quantized—a requirement seen from the angular part of the wavefunction exactly as in the nonrelativistic problem. The Dirac Hamiltonian is singular for the lowest admissible angular momentum state, corresponding to the spiral orbits, when the magnetic charge is nonzero. The remaining states show an accidental doubling of degeneracy, whose presence may be deduced from an operator which reduces to that of Johnson and Lippman, or the algebra of Malkin and Manko, without the magnetic charge.

I. INTRODUCTION

In the past few years there have been numerous investigations into the symmetry and group-theoretical structure of quantum-mechanical problems. The hydrogen atom and the harmonic oscillator have always received the most attention, probably because of the intrinsic importance of these two systems to theoretical physics, and also because of the fact that they are the only two systems of such importance showing a high degree of accidental degeneracy. An explanation for their accidental degeneracy has been known for quite some time,[1] and the reason is that the phase space of Hamiltonian mechanics accommodates rather more symmetry than may be found solely in the configuration space, which is rarely more than the spherical symmetry enjoyed by central-force potentials. The recently increased activity has been concerned with establishing the precise extent of the occurrence of hidden symmetry of this kind, on the one hand, and with "noninvariance," on the other. Noninvariance is concerned with determining the shape and not merely the degeneracy of the eigenvalue spectrum from group-theoretical information.

The study of the degeneracy and symmetry of relativistic wave equations has probably been inhibited by the low degree of accidental degeneracy found in the Dirac equation for the Coulomb potential, not to mention that the relativistic hydrogen atom is just about the only system whose Dirac equation leads to any bound states at all. Although a number of magnetic-field configurations lead to confined motion, they are not described by the adjective "bound" in quite the sense that one would think of for an attractive center; in any event, the Coulomb potential seems to be the only electrostatic potential expressible as a power series in r and $1/r$ which leads to bound-state solutions.[2] All the other potentials have only oscillatory solutions which extend to infinity with nonvanishing amplitude. Phenomena of this category comprise the Klein paradox.[3]

Although Dirac chose the form of his Hamiltonian deliberately, so that the wave equation would be a set of coupled linear first-order differential equations and not a second-order equation such as the nonrelativistic Schrödinger equation, the very fact that the wave equation is such a set of first-order equations obscures their symmetry to a considerable extent.

The most important symmetry property of the Dirac equation for a spherically symmetric potential is, of course, that very same spherical symmetry and the concomitant conservation of angular momentum. Just as a system of first-order equations cannot

* Sponsored in part by the Swedish National Research Council and in part by the Air Force Office of Scientific Research through the European Office of Aerospace Research, U.S. Air Force, Grant EOOAR-69-0043.

† Instituto Nacional de la Investigación Científica Fellow (*México*).

‡ Permanent address: Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional, México 14, D.F., México. "Profesor Becario de la Comisión de Operación y Fomento de Actividades Academicas del Instituto Politecnico Nacional."

[1] V. Fock, Z. Physik **98**, 145 (1935); V. Bargmann, *ibid.* **99**, 576 (1936); J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940). A survey of the earliest papers on accidental degeneracy may be found in H. V. McIntosh, Am. J. Phys. **27**, 620 (1959). Among the many recent treatments, the hydrogen atom is especially well discussed in G. Györgyi, Nuovo Cimento **53A**, 717 (1968); M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966).

[2] M. S. Plesset, Phys. Rev. **41**, 278 (1932).

[3] O. Klein, Z. Physik **53**, 157 (1929).

possess spherical symmetry as written, it is found that the "orbital" angular momentum $\mathbf{r} \times \mathbf{p}$, which generates infinitesimal rotations of the coordinates and momenta, is not conserved. One can attribute this failure of the canonical angular momentum to generate a symmetry operation to the fact that one deals with a system of equations, and thus with a vector field rather than a simple wavefunction. It is a familiar property of fields that a transformation of the coordinates with respect to which they are defined causes simultaneous mixing of the field components, which must be separated again before one can discuss the possibility of a symmetry.[4] This resolution of the components can be effected by a matrix and gives rise to the "spin" component of the angular momentum. The combination of orbital and spin angular momentum, which leads to a symmetry operation, is called the total angular momentum, and was found by Dirac[5] to be a constant of the motion for systems with spherically symmetric potentials.

The study of the reflective symmetry of the Dirac equation is similarly complicated by the fact that a spatial reflection must be combined with a spin adjustment, and even the experimental consequences of reflective symmetry were not appreciated for twenty years, until the parity-conservation experiments were performed.

Above and beyond any rotational or reflective symmetries common to all central-force potentials, there are some potentials which exhibit the higher and supposedly accidental degeneracy, similar to the instances which we have already mentioned in the nonrelativistic theory. The additional symmetry has been explained in terms of "hidden" symmetries from the phase space of the problem, the degeneracy in the hydrogen atom generally being attributed to the existence of a second conserved vector, the Runge vector. This is a polar vector in contrast to the first vector constant, the angular momentum, which is an axial vector. In classical mechanics, the Runge vector points from the attracting center to the perihelion of the elliptical Keplerian orbit, whose constancy reflects the independence of the orbital energy from the eccentricity of the orbit. Instead, the energy of any orbit depends only on its semimajor axis. One might expect to find a residue of the nonrelativistic symmetry in the relativistic problem.

Even in classical mechanics, however, the symmetry implied by the constancy of the Runge vector is lost in the relativistic problem. The variation of the particle's mass with its velocity results in an orbital precession, so that the perihelial vector no longer remains constant. It would therefore seem unlikely that a quantized version of the motion would show any greater degree of symmetry. Such expectations for the Dirac equation of the hydrogen atom are confirmed with one slight exception: One finds double the degeneracy which one expects from spherical symmetry and the behavior of the irreducible representations of the rotation group. The hydrogen atom consequently *does* show an accidental degeneracy in this doubling of the rotational degeneracy.

Historically, this doubling was quite important because its incompatibility with the known degeneracies due to the spherical symmetry first led to the hypothesis of electron spin, and eventually to the Dirac equation itself. Ironically, it is the symmetry aspect of the Dirac equation which has resisted a group-theoretical explanation for the longest time.

In 1949, Sáenz[6] reported an unsuccessful search for a hidden symmetry group which would explain this degeneracy, although he did succeed in showing that no operator similar to the Runge vector could be constant. Almost simultaneously, Johnson and Lippman[7] reported an operator which, by its anticommutation properties, could account for the accidental degeneracy. At about the same time, Johnson and Lippman[8] also investigated the motion of a Dirac particle in a uniform magnetic field, finding some new constants of the motion, their commutation rules, and that a similar doubling of energy levels could be attributed to the indifference of the energy to the relative orientation of the spin and the orbital angular momentum.

The accidental degeneracy of the hydrogen atom could certainly be explained in similar terms, but only after one had solved Dirac's equation and obtained explicit formulas for the wavefunctions and energy levels. Biedenharn[9] in 1962 first established a direct connection between the Johnson–Lippman constant of the motion and the "helicity independence" of the Dirac Hamiltonian.

If one understands the loss of symmetry in passing from the nonrelativistic to the relativistic hydrogen atom in terms of the precession caused by the relativistic mass variation, it is certainly tempting to try to restore the symmetry by counteracting the precession in some way. Biedenharn and Swamy[10] were able to identify a term in the Dirac Hamiltonian

[4] D. L. Pursey, Ann. Phys. (N.Y.) **32**, 157 (1965).
[5] P. A. M. Dirac, Proc. Roy. Soc. (London) **A117**, 610 (1928).
[6] A. W. Sáenz, Phys. Rev. **79**, 1004 (1950).
[7] M. H. Johnson and B. A. Lippman, Phys. Rev. **78A**, 329 (1950).
[8] M. H. Johnson and B. A. Lippman, Phys. Rev. **76**, 828 (1949); **77**, 701 (1950).
[9] L. C. Biedenharn, Phys. Rev. **126**, 845 (1962).
[10] L. C. Biedenharn and N. V. V. J. Swamy, Phys. Rev. **133B**, 1353 (1964).

which could be held responsible for the loss of degeneracy, and whose removal led to a highly degenerate system. Christened as their "symmetric" Hamiltonian, it was shown to have two vector constants of the motion, whose commutation rules led to identification of R_4 as a hidden symmetry group. However, their symmetric Hamiltonian still showed a doubling, this time of R_4 degenerate levels, and it was found that the two vector constants of the motion were no longer orthogonal, as they are for the nonrelativistic hydrogen atom.

By taking the Foldy–Wouthuysen limit of the "symmetric" Hamiltonian, Sheth[11] has shown that it differs from the Coulombic Hamiltonian by including a vector potential whose influence is just adequate to counteract the relativistic precession.

Probably the most satisfactory accounting for the degeneracy doubling in the Dirac equation is found in a recent series of articles by Malkin and Manko.[12] It may still be an open question as to whether the phenomenon appears in non-Coulombic problems, although it is readily enough seen for free particles and for uniform magnetic fields. Some insight may be gained from the "two-dimensional" Dirac equation of Ionesco–Pallas,[13] or the multidimensional theory of Coulson and Joseph,[14] but it does not seem that one can look for analogies among other potentials, since they do not generally yield bound states.

A system closely related to the relativistic hydrogen atom is the charged magnetic monopole,[15] which does not seem to have been previously treated in the Dirac theory. Harish-Chandra[16] investigated the Dirac equation for a charged particle moving in the field of an uncharged monopole to see whether the intrinsic magnetic moment of the electron could lead to bound states. It did not. Banderet[17] treated scattering from an uncharged monopole according to the Dirac theory. Malkus[18] studied the influence of the presence of a magnetic monopole in an atomic nucleus on its electronic energy levels, as well as the motion of a nucleus in the field of the monopole in the Pauli approximation. Eliezer and Roy,[19] as well as McIntosh and Cisneros,[20] have used the nonrelativistic Schrödinger equation to study motion in the field of a combination of magnetic and electric charge. There are numerous nonrelativistic accounts of motion in the field of an uncharged monopole, but there are no bound states in such systems. Hurst[21] has made a fundamental investigation of the requirements of quantization of the magnetic charge imposed by integrability conditions for both the Schrödinger and Dirac equations for the monopole.

There are some significant differences between the Coulombic hydrogen atom and the charged magnetic monopole, whose mathematical origin lies in the vector potential of the monopole's magnetic field. The angular momentum is no longer conserved, there being a certain amount of angular momentum resident in the magnetic field. Again, there is a conserved total angular momentum, originally discovered by Poincaré,[22] but whose minimal value is nonzero and depends on the magnetic charge of the monopole.

The existence of such a lower limit affects the wavefunctions in such a way that there are never s-states, nor even some other states of low angular momentum, depending upon the magnitude of the magnetic charge.

Since the Dirac equation demands the inclusion of spin angular momentum in the conserved total angular momentum of a spherically symmetric system, it is worthwhile to see how the threefold combination of spin, field, and orbital angular momentum is to be made in the relativistic theory of the monopole. It is a consideration which is not entirely academic, in the light of proposals which have been entertained from time to time,[23] that particles are composites of magnetic monopoles. Schwinger[23a] has recently proposed a possible justification of the "quark" model following this idea. Also, an attempt has been made to explain the great mass and anomalous magnetic moment of the neutron in this way.[23b] Even though quite different explanations in terms of meson theories are generally given to these phenomena, the nonrelativistic theory of the monopole contains some interesting possibilities for speculation. Classically, two charged monopoles would not be expected to collide because of the magnetic mirror effect. This result is retained quantum mechanically, and may be attributed to the angular momentum of the magnetic field, so that even in the

[11] C. V. Sheth, Nuovo Cimento **54A**, 549 (1968).

[12] (a) I. A. Malkin and V. I. Manko, Yad. Phys. **8**, 627 (1968), [Sov. J. Nucl. Phys. **8**, 363 (1969)]; (b) I. A. Malkin and V. I. Manko, ZhETE Pis. Red. **7**, 105 (1968) [JETP Lett. **7**, 79 (1968)].

[13] N. J. Ionesco-Pallas, Rev. Roumaine de Phys. **12**, 327 (1967).

[14] A. Joseph, Rev. Mod. Phys. **39**, 829 (1967); C. A. Coulson and A. Joseph, Rev. Mod. Phys. **39**, 838 (1967).

[15] (a) P. A. M. Dirac, Proc. Roy. Soc. (London) **A133**, 60 (1931); (b) Ig. Tamm, Z. Physik **71**, 141 (1931).

[16] Harish-Chandra, Phys. Rev. **74**, 883 (1948).

[17] P. P. Banderet, Helv. Phys. Acta **19**, 503 (1946).

[18] W. V. R. Malkus, Phys. Rev. **83**, 899 (1951).

[19] C. J. Eliezer and S. K. Roy, Proc. Cambridge Phil. Soc. **58**, 401 (1962).

[20] H. V. McIntosh and A. Cisneros, Bull. Am. Phys. Soc. **13A**, 909 (1968); J. Math. Phys. (to be published).

[21] C. A. Hurst, Ann. Phys. (N.Y.) **50**, 51 (1968).

[22] H. Poincaré, Compt. Rend. **123**, 530 (1896).

[23] (a) J. Schwinger, Phys. Rev. **173**, 1536 (1968); (b) M. N. Saha, Indian J. Math. **10**, 141 (1936); (c) R. F. Palmer and J. G. Taylor, Nature **219L**, 1033 (1968).

ground states, the repulsive centrifugal potential occurring in the radial equation is nonzero, and the particles would be excluded from the origin of the center-of-mass system.

Such reasoning makes one think that there might be a state of quite high binding energy, due to the strength of magnetic interaction between two oppositely charged monopoles, and of relatively long lifetime, since the probability of coincidence of the two particles would be greatly reduced. Such expectations are quashed in the relativistic Dirac theory, at least to the extent that one can obtain information by studying the motion of a spin-$\frac{1}{2}$ particle in the field of a fixed, spinless particle. The reason is that the spin angular momentum can oppose the residual angular momentum responsible for excluding the particle from the origin in the nonrelativistic theory; in fact, when the effects of relativistic precession are also included, one obtains a singular potential for the ground state.

This singular potential is familiar in the Dirac theory, and has a classical relativistic analog in the spiral orbits which arise for sufficiently low angular momentum when the relativistic precession completely overwhelms the orbital motion. A very thorough account of the singular case was given by Frenkel and Rojansky[24] in 1938, but for the relativistic Coulomb potential one does not encounter such difficulties for nuclei with $Z \leq 137$, due to the small size of the fine structure constant. While the possible occurrence of spiral motion is then of no practical importance for the study of atomic nuclei, we find that it occurs already from $Z = 1$ when the nucleus carries a magnetic charge.

Aside from the occurrence of a singular potential, we may note that the degeneracy of the nonrelativistic Coulomb potential, which is lost to the charged monopole, may be regained by considering a fictitious problem in which there is a repulsive centrifugal potential proportional to the square of the magnetic pole strength. At the same time, the possibility of the occurrence of the singular potential would be avoided, and we would have a system bearing the same relation to a relativistic physical potential as Biedenharn and Swamy's symmetric Hamiltonian has to the relativistic hydrogen atom.

II. CLASSICAL TREATMENT

The classical relativistic Hamiltonian[25] for the hydrogenlike atom is

$$\mathcal{H} = (p^2c^2 + m^2c^4)^{\frac{1}{2}} - Ze^2/r, \tag{1}$$

where Z is the nuclear charge. Squaring this Hamiltonian leads one to a Hamilton–Jacobi equation which is separable in polar spherical coordinates

$$\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \theta}\right)^2 + \frac{1}{r^2 \sin^2\theta}\left(\frac{\partial S}{\partial \varphi}\right)^2 = \left(\frac{E}{c} + \frac{\alpha Z}{r}\right)^2 - m^2c^2, \tag{2}$$

where

$$\alpha = e^2/c.$$

Introducing separation constants μ (z projection of the angular momentum), ℓ^2 (square of the angular momentum), and E (energy), we obtain their definitions as follows:

$$\begin{aligned} \mu &= \frac{\partial S}{\partial \varphi}, \\ \ell^2 &= \left(\frac{\partial S}{\partial \theta}\right)^2 + \frac{\mu^2}{\sin^2\theta}. \end{aligned} \tag{3}$$

Using these separation constants, we may introduce the action-angle variables

$$\begin{aligned} J_\varphi &\equiv \oint \frac{\partial S}{\partial \varphi}\, d\varphi = 2\pi\mu, \\ J_\theta &\equiv \oint \frac{\partial S}{\partial \theta}\, d\theta = 2\pi(\ell - \mu), \\ J_r &\equiv \oint \frac{\partial S}{\partial r}\, dr \\ &= -2\pi(\ell^2 - \alpha^2 Z^2)^{\frac{1}{2}} + \frac{\alpha ZE}{(m^2c^2 - E^2/c^2)^{\frac{1}{2}}}. \end{aligned} \tag{4}$$

This formula for the radial action is only valid for sufficiently large angular momentum, viz., $\ell > \alpha Z$. The restriction which must be laid upon the range of angular momentum arises from the fact that the Hamiltonian–Jacobi equation contains the term in $1/r^2$, which is effectively an attractive potential with the same radial dependence as the "centrifugal" potential which appears on separation of the radial variables. Both of these are singular potentials in the sense that the action integral diverges if the range of integration extends to the origin. If the Coulomb potential is attractive and the centrifugal potential repulsive, the innermost classical turning point will lie at some distance from the origin, producing the expression for the radial action given above. However, if both potentials are attractive, the particle may reach the origin, the action integral diverges, and the formulation of Hamiltonian mechanics in terms of action-angle variables is not applicable.

Upon solving for the energy, we obtain

$$E = mc^2\left(1 + \frac{\alpha^2 Z^2}{\{n_r + [(n_\theta + n_\varphi)^2 - \alpha^2 Z^2]^{\frac{1}{2}}\}^2}\right)^{-\frac{1}{2}}, \tag{5}$$

where

$$n_i = J_i/2\pi.$$

[24] J. Frenkel and J. Rojansky, Physik. Z. Sowjetunion **13**, 181 (1938).

[25] A. Sommerfeld, Ann. Physik **51**, 1, 125 (1916); M. Born, *The Mechanics of the Atom* (Fredrick Ungar Publ. Co., New York, 1960).

Using the old quantization rules, this is seen to be nothing but Sommerfeld's expression for the energy levels,[25] provided that n_r, n_θ, and n_φ are integers,[26] which gives account of the fine structure of the levels. The equation for the orbit is that of a forward precessing ellipse, with a frequency $(1 + \alpha^2Z^2/\ell^2)^{\frac{1}{2}}$. An explicit solution was found by Copel[27] and also by Frenkel and Rojansky.[24]

In addition to the appearance of the rest-mass energy and the double sign, two main features can be noticed in comparing Eq. (2) with the nonrelativistic Hamiltonian–Jacobi equation. One of them is the inclusion of an increment to the centrifugal potential α^2Z^2/r^2, while the other one is a factor $2E/c^2$ in the Coulomb potential term $-e^2Z/r$. The former causes the precession of the orbital ellipse due to the instantaneous mass change with the velocity, while the latter one is a screening factor due to the average mass change. It is the modification to the centrifugal potential which destroys the familiar R_4 symmetry of the nonrelativistic hydrogen atom. On account of this term, the Runge vector will no longer be a constant vector, but in actuality it precesses with a constant velocity.

The potential responsible for the orbital precession is attractive and results in an advance, rather than a retardation, of the perihelion with each cycle of the orbit. It has the same radial dependence as the centrifugal barrier, namely, $1/r^2$, and,as we have seen, the nature of the motion changes drastically when this term changes sign. As long as the centrifugal terms in $1/r^2$ are repulsive, the particle can never penetrate to the origin, it oscillates back and forth between a maximum and minimum radius, and a radial action variable may be defined. When the term is attractive, the particle may reach the origin, with unpleasant effects on the action integral, which then diverges. That the particle should pass through the origin would probably be of minor physical consequence were it not for the fact that Newton's laws do not seem to be adequate to describe the passage through the origin. Wave equations seem to suffer a similar inability.

The difficulty is that, although the radial velocity of the particle is nearly constant as it approaches the origin, it loses potential energy so rapidly that it must accelerate tremendously in order to maintain its angular momentum. It therefore spirals around and around the origin, and loses its orientation on passing through the origin. Such behavior normally does not occur in a Coulomb potential, but the relativistic mass variation prevents a particle from undergoing an adequate acceleration, and its maximum velocity is limited by the velocity of light.

One might recall the considerable difficulty which the old quantum mechanics encountered with orbits through the nucleus, and take note of the fact that even wave mechanics is not free of delicacies in the treatment of states of zero angular momentum.[28] Fortunately for the relativistic treatment of atomic spectra, singular behavior in the Dirac equation does not occur until the critical nuclear charge of $Z = 137$, well beyond the heaviest stable nucleus now known.

It is most interesting to construct a relativistic Hamiltonian which restores the broken symmetry, and look at this as an approximation to the hydrogenlike problem. This is achieved by introducing a vector potential whose magnitude is eZ/r and forcing it to be orthogonal to $\mathbf{p}$, the canonical momentum vector, i.e., in the relativistic Hamiltonian

$$\mathcal{H} = \{[\mathbf{p} - (e/c)\mathcal{A}]^2c^2 + m^2c^4\}^{\frac{1}{2}} + e\phi, \qquad (6)$$

we take $\phi = -Ze/r$ and $\mathcal{A} = (Ze/r)\hat{\boldsymbol{\lambda}}$, where $\hat{\boldsymbol{\lambda}}$ is a unit vector such that $\mathbf{p} \cdot \hat{\boldsymbol{\lambda}} = 0$. The Hamilton–Jacobi equation for the squared Hamiltonian,

$$\mathcal{H}^2 = p^2c^2 + m^2c^4 - (2\alpha Z/r)\mathcal{H}, \qquad (7)$$

now reads

$$\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \theta}\right)^2 + \frac{1}{r^2\sin^2\theta}\left(\frac{\partial S}{\partial \varphi}\right)^2 = \frac{E^2}{c^2} + \frac{2\alpha ZE}{cr} - m^2c^2, \qquad (8)$$

where E is the total energy of the system.

Although the "screening factor" remains, the centrifugal potential α^2Z^2/r^2 is no longer present. The separation constants are again

$$\mu = \frac{\partial S}{\partial \varphi},$$

$$\ell^2 = \left(\frac{\partial S}{\partial \theta}\right)^2 + \frac{\mu^2}{\sin^2\theta},$$

and the energy E. For the action-angle variables, one finds

$$\begin{aligned} J_\varphi &= 2\pi\mu, \\ J_\theta &= 2\pi(\ell - \mu), \\ J_r &= -2\pi[(\ell^2)^{\frac{1}{2}} + \alpha ZE/(m^2c^2 - E^2/c^2)^{\frac{1}{2}}]; \end{aligned} \qquad (9)$$

thus, choosing the positive square root for ℓ^2, the energy is given as

$$E = mc^2(1 + \alpha^2Z^2/N^2)^{-\frac{1}{2}}. \qquad (10)$$

[26] We will use $\hbar = 1$ throughout.

[27] P. Copel, J. Phys. Radium **4**, 636 (1933); see also R. G. Cawley, J. Math. Phys. **8**, 2092 (1967).

[28] T. Tits, Sov.Phys.—JETP **3**, 777 (1956).

This is exactly the same as the one found by Frenkel and Rojansky[24] for the critical case in which the two potentials just cancel.

The expression for N is

$$N = (J_r + J_\theta + J_\varphi)/2\pi, \tag{11}$$

which, together with the energy expression (10), shows that the motion is doubly degenerate, as the radial, azimuthal, and zenithal frequencies are the same, namely,

$$\nu = \frac{\partial \mathcal{H}}{\partial J_r} = \frac{\partial \mathcal{H}}{\partial J_\theta} = \frac{\partial \mathcal{H}}{\partial J_\varphi} = \frac{mc^2\alpha^2Z^2}{2\pi}(N^2 + \alpha^2Z^2)^{-\frac{3}{2}}. \tag{12}$$

The equation of the orbit is readily found to be that of an ellipse, as expected, but with a semimajor axis

$$a = (Ze^2/E)[(m^2c^4/E^2) - 1]^{-1} \tag{13}$$

instead of the usual $Ze^2/2E$.

The canonical equations of motion for our Hamiltonian

$$\mathcal{H} = (p^2c^2 + m^2c^4 - \alpha^2Z^2c^2/r^2)^{\frac{1}{2}} - \alpha Zc/r \tag{14}$$

are given as

$$\begin{aligned} \dot{\mathbf{r}} &= \nabla_p \mathcal{H} = \mathbf{p}/M_0, \\ \dot{\mathbf{p}} &= -\nabla_r \mathcal{H} = (\alpha^2Z^2/M_0r^4)\mathbf{r} - (\alpha Zc/r^3)\mathbf{r}, \end{aligned} \tag{15}$$

where

$$M_0 = (1/c^2)[p^2c^2 + m^2c^4 + (\alpha^2Z^2/r^2)c^2]^{\frac{1}{2}} \tag{16}$$

plays the role of the mass.

Using these equations, we immediately see that the canonical angular momentum is a constant of the motion, for

$$\begin{aligned} \dot{\boldsymbol{\ell}} &= \dot{\mathbf{r}} \times \mathbf{p} + \mathbf{r} \times \dot{\mathbf{p}} \\ &= (\mathbf{p}/M_0) \times \mathbf{p} + \mathbf{r} \times [(\alpha^2Z^2/M_0r^4) - \alpha Zc/r^3]\mathbf{r} = 0 \end{aligned} \tag{17}$$

and as, by construction, it is orthogonal to $\mathbf{p}$, we can identify $\hat{\boldsymbol{\lambda}}$ with $\hat{\boldsymbol{\ell}}$ and notice that $\mathbf{p}$ and $\mathbf{r}$ lie in a plane, the plane of the orbit. Furthermore, the separation constants ℓ and μ on Eq. (9) are, respectively, the total canonical angular momentum and its projection on the z axis. It is worth noticing the unusual fact that it is the canonical, rather than the mechanical angular momentum which lies on the orbit plane, a result due to the fact that we are dealing with a velocity-dependent vector potential. A word of warning should also be spoken regarding the Lagrangian function and the Lorentz force, quantities which have to be rederived from the Hamiltonian (14), rather than from the potentials themselves.

As the orientation of the ellipse is fixed for this problem, one expects the Runge vector to be also a constant of the motion. This is true provided one replaces the Hamiltonian function over c^2 for the mass in the usual formula. That is, we define it as

$$\mathbf{A}' = \boldsymbol{\ell} \times \mathbf{p} + (\alpha Z/c)\mathcal{H}\hat{\mathbf{r}}. \tag{18}$$

Using the fact that

$$\dot{\hat{\boldsymbol{\gamma}}} = [\mathbf{r} \times (\dot{\mathbf{r}} \times \mathbf{r})]/r^3, \tag{19}$$

recalling that $\mathcal{H}$ and $\boldsymbol{\ell}$ are constants, and recalling the expansion for a double vector product, one finds

$$\dot{\mathbf{A}}' = \boldsymbol{\ell} \times \dot{\mathbf{p}} + (\alpha Z/c)\mathcal{H}\dot{\hat{\mathbf{r}}} = 0, \tag{20}$$

where we have used the identity

$$\mathcal{H} = M_0c^2 - \alpha Zc/r. \tag{21}$$

It is convenient to normalize $\mathbf{A}'$ as

$$\mathbf{A} = c(m^2c^4 - \mathcal{H}^2)^{-\frac{1}{2}}(\boldsymbol{\ell} \times \mathbf{p} + (\alpha Z/c)\mathcal{H}\hat{\mathbf{r}}), \tag{22}$$

a vector which fulfills the two identities

$$\begin{aligned} \boldsymbol{\ell} \cdot \mathbf{A} &= 0, \\ \ell^2 + A^2 &= [\alpha^2Z^2/(m^2c^4 - \mathcal{H}^2)]\mathcal{H}^2. \end{aligned} \tag{23}$$

Thus, from the last one, we can express the Hamiltonian in terms of the magnitudes of the two constant vectors $\boldsymbol{\ell}$ and $\mathbf{A}$:

$$\mathcal{H} = mc[1 + \alpha^2Z^2/(A^2 + \ell^2)]^{-\frac{1}{2}}. \tag{24}$$

In addition to the Hamiltonian $\mathcal{H}$, we have thus found six constants of the motion—the components of $\boldsymbol{\ell}$ and $\mathbf{A}$—which are not independent of each other as we can see from (23). By direct computation of the Poisson brackets, it is easily shown that

$$\begin{aligned} \{\ell_i, \ell_j\} &= \epsilon_{ijk}\ell_k, \\ \{\ell_i, A_j\} &= \epsilon_{ijk}A_k, \\ \{A_i, A_j\} &= \epsilon_{ijk}\ell_k, \end{aligned}$$

for $i, j, k = 1, 2, 3$.

The relations show that these components are the generators of a rotation group R_4 in four dimensions.

III. SYMMETRY OF THE RELATIVISTIC HYDROGEN ATOM

Separation of the Hamiltonian–Jacobi equation for the relativistic hydrogen atom in polar coordinates yields an expression for the energy which can be quantized according to the rules of the old quantum

mechanics. Such was the origin of the Sommerfeld formula for the fine structure of the hydrogen spectrum, which may have been one of the most successful applications of the old quantum mechanics. Nevertheless, there had been experimental indications that the behavior of electrons was more complicated than could be simply extrapolated from classical considerations, and this eventually led to the hypothesis of electron spin. Moreover, the attempt to treat the hydrogen spectrum by a relativistic version of wave mechanics led to unequivocally wrong predictions of the fine structure, and thereby discredited the wave equation which has eventually become known as the Klein–Gordon equation. By its automatic inclusion of the electron spin and an accurate placement of the hydrogenic levels, Dirac's relativistic wave equation gained immediate acceptance.

As is well known, this equation is a system of four simultaneous first-order partial differential equations

$$(\rho_1 \boldsymbol{\sigma} \cdot \mathbf{p}c + \rho_3 mc^2 + V)\Psi = E\Psi, \tag{25}$$

which is most readily solved by iteration after being written in spherical coordinates. Its most important symmetry property, when the vector potential $\mathcal{A}$ is zero and $V(r)$ is a central force potential, depends upon its behavior with respect to a rotation of coordinates. As Dirac showed when he first introduced the equation, a change of spatial coordinates always results in the mixing of the components of a multicomponent wavefunction. So it was that the orbital angular momentum

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}$$

alone was not conserved, but rather, the total angular momentum

$$\mathbf{J} = \mathbf{L} + \tfrac{1}{2}\boldsymbol{\sigma}. \tag{26}$$

Indeed, this was the automatic way in which the Dirac equation incorporated the electron's "spin" $\boldsymbol{\sigma}$. Since $\mathbf{J}$ is a vector constant of the motion, its magnitude J^2 and one of its components, say J_z, may be used to define quantum numbers for the wavefunctions. As the components of $\mathbf{J}$ satisfy the angular-momentum commutation rules

$$[J_\alpha, J_\beta] = i\epsilon_{\alpha\beta\gamma}J_\gamma, \tag{27}$$

the familiar algebra of angular momentum is applicable. It is only necessary to recall that $\mathbf{J}$ is a *total* angular momentum, incorporating a simultaneous charge of coordinates and mixing of wave components.

These two operations are possible individually, and for the classification of the wavefunctions, it is often preferable to consider $\mathbf{L}$ and $\mathbf{J}$ as separately diagonal. In such a representation the Dirac Hamiltonian will not be diagonal, but its nearness to being diagonal often makes the product representation much more useful. In fact, if we consider the Lie algebra generated by $\mathbf{L}$ and $\boldsymbol{\sigma}$ separately, the rotational symmetry group of the Dirac equation is the Kronecker product of these two Lie groups, generated by $\mathbf{J}$. The Casimir operators of the direct product are L^2, σ^2, and $\boldsymbol{\sigma} \cdot \mathbf{L}$, while J^2 is the Casimir operator of the Kronecker product.

It will be found that these Casimir operators are not themselves constants of the Dirac equation, but the closely related operator

$$\mathcal{K} = \rho_3(\boldsymbol{\sigma} \cdot \mathbf{L} + 1) \tag{28}$$

is such a constant. Its square is not independent of J^2 since

$$J^2 = \mathcal{K}^2 - \tfrac{1}{4} \tag{29}$$

so its eigenvalues $(-\kappa)$ fulfill

$$\kappa^2 = (j + \tfrac{1}{2})^2 \tag{30}$$

and have an important interpretation. Namely, they measure whether the spin is parallel or antiparallel to the orbital angular momentum, and distinguish whether a state of given j is formed as $l + \frac{1}{2}$ or $l - \frac{1}{2}$. The eigenvalues $-\kappa$ of $\mathcal{K}$ have opposite signs according to these two possibilities: If $\kappa = +j + \frac{1}{2}$, then $\kappa = 1$; while if $\kappa = -(j + \frac{1}{2})$, then $\kappa = -(l + 1)$, recalling that

$$L^2 = \mathcal{K}^2 - \rho_3\mathcal{K}. \tag{31}$$

When the Dirac Hamiltonian is written in polar coordinates, the operator $\mathcal{K}$ makes its appearance in the role of an angular momentum operator. We find the equation taking the form

$$\left[\rho_1 \boldsymbol{\sigma} \cdot \hat{r}\left(\frac{1}{i}\frac{\partial}{\partial r} + i\frac{\rho_3\mathcal{K} - 1}{r}\right)c + \rho_3 mc^2 + V(r) - E\right]\Psi = 0. \tag{32}$$

There are various ways to solve the Dirac equation, particularly for the Coulomb potential, but they share the common feature that $\mathcal{K}$ be diagonalized. The same spherical harmonics arising from Pauli's nonrelativistic spin equation suffices for the purpose, throwing the main burden of the solution on a determination of the radial wavefunctions. Trial solutions of the form of power series may be used, according to the time-honored methods of solving linear differential equations. Yet operational methods are more elegant and often are much more instructive.

The earliest operational solution of the Dirac equation for the hydrogen atom seems to have been

published by Temple[29] in 1930. Working with the Dirac equation itself, even after the angular variables have been separated, leads to four coupled radial equations, so that an effort is generally made to reduce this number to two by iterating the Dirac equation to form a second-order equation analogous to the Klein–Gordon equation. This is best done by reinterpreting the Dirac equation as a projection operator for positive masses. If we define

$$\mathcal{O}_+ = (1/2mc^2)[i\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}c + mc^2 - \rho_3(E + \alpha Zc/r)], \tag{33a}$$

and similarly,

$$\mathcal{O}_- = (1/2mc^2)[i\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}c - mc^2 - \rho_3(E + \alpha Zc/r)], \tag{33b}$$

it is readily verified that

$$\begin{aligned} \mathcal{O}_+^2 &= \mathcal{O}_+, \\ \mathcal{O}_-^2 &= \mathcal{O}_-, \\ \mathcal{O}_+\mathcal{O}_- &= \mathcal{O}_-\mathcal{O}_+. \end{aligned} \tag{34}$$

As a consequence, any solution of the Dirac equation

$$\mathcal{O}_+\Psi = 0 \tag{35}$$

also satisfies the iterated equation

$$\mathcal{O}_-\mathcal{O}_+\Psi = 0. \tag{36}$$

Conversely, we may manufacture a solution of Dirac's equation by projecting any solution of the iterated equation. We need only read this iterated equation in the form $\mathcal{O}_+(\mathcal{O}_-\Phi) = 0$, and define $\Psi = \mathcal{O}_-\Phi$.

Written in spherical coordinates, the iterated equation has the form

$$\left[\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} - \frac{\Gamma(\Gamma-1)}{r^2} - \frac{2\alpha ZE}{rc} + k^2\right]\Phi = 0, \tag{37}$$

wherein one has defined

$$\begin{aligned} \Gamma &= \rho_3\mathcal{K} + i\alpha Z\rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \\ k^2 &= (E^2/c^2) - m^2c^2, \end{aligned} \tag{38}$$

and exploited the identities

$$\{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \rho_3\mathcal{K}\} \equiv \{\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \boldsymbol{\sigma}\cdot\mathbf{L} + 1\} = 0, \tag{39a}$$

$$\{\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}, \rho_3\alpha Z/r\} = \alpha Z\rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}/r^2, \tag{39b}$$

$$\Gamma^2 = \mathcal{K}^2 - \alpha^2Z^2. \tag{39c}$$

Over three decades elapsed before the suggestion was made that Γ itself be diagonalized, rather than $\mathcal{K}$. In 1962, Biedenharn[9] showed that, in the coordinate system S, defined by

$$S = \exp\,[-\tfrac{1}{2}\rho_2\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\,\operatorname{arc\,tanh}\,(\alpha Z/\mathcal{K})], \tag{40}$$

the diagonal Γ is

$$S\Gamma S^{-1} = \rho_3\mathcal{K}(1 - \alpha^2Z^2/\mathcal{K}^2)^{\frac{1}{2}}, \tag{41}$$

and the equation for the radial wavefunction also became simpler. As predicted, Fradkin[30] found that S_1 defined a coordinate system in which the scattering solutions for the hydrogen atom were more tractable.

The eigenvalues of Γ also enter into the energy eigenvalues of the Dirac Hamiltonian in a way which is best seen in the work of Coulson and Joseph.[14] They showed that the Hamiltonian had the operational form[31]

$$\mathcal{H} = mc^2[1 + \alpha^2Z^2/(R + \mathcal{K})^2]^{-\frac{1}{2}}, \tag{42}$$

where the operator R is defined by

$$R = (m^2c^2 - \mathcal{H}^2/c^2)^{-\frac{1}{2}}[-i(\mathcal{K}\rho_1/c)(\mathcal{H} - \rho_3mc^2) + \alpha Zc\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}]. \tag{43}$$

This form of the Hamiltonian leads, with the help of (38) and (39c), to the Sommerfeld fine structure formula for the hydrogen atom:

$$E = mc^2\left(1 + \frac{\alpha^2Z^2}{[n + (|\kappa|^2 - \alpha^2Z^2)^{\frac{1}{2}}]^2}\right)^{-\frac{1}{2}}, \tag{44}$$

where n is the radial quantum number and $-\kappa$ the eigenvalue of $\mathcal{K}$. It will be seen that this formula requires $\kappa \neq 0$, which is always satisfied on account of the allowable range of j, and that $\alpha Z \leq 1$, since one is the least admissible value of $|\kappa|$. This latter requirement is satisfied for all the natural elements, for which $Z \leq 137$, and is seen to be a quantum-mechanical manifestation of the requirements for the nonoccurrence of spiral orbits which we encountered in the classical analysis of the previous section.

If these restrictions are violated, it by no means implies that the Dirac equation has no solution, although it does indeed become singular and a more specialized treatment becomes necessary. In any event, the validity of the energy formula is not assured for the spiral orbits, and one should not simply assume that there are complex energy eigenvalues. Since the cutoff for spiral orbits depends on the angular momentum, we see that for $0 \leq \alpha Z \leq 1$ the Sommerfeld formula is completely valid, that for $1 \leq \alpha Z \leq 2$ it only applies in unmodified form to states with $|\kappa| \geq 2$, and that, as αZ is steadily increased, one loses one more angular-momentum state each time that it passes an integer value.

[29] G. Temple, Proc. Roy. Soc. (London) **A127**, 349 (1930); and also F. Sauter, Z. Physik **63**, 803 (1930); **64**, 295 (1930).

[30] D. M. Fradkin, Phys. Rev. **135B**, 1085 (1964).

[31] There is a misprint in their formulas (5.8) and (5.9), in which a factor H^2 is missing.

In analyzing the symmetry of the hydrogen atom and the degeneracy arising therefrom, it is important to notice that formula (44) is independent of J_z, as one would expect from the "spherical symmetry" of the system. Dependence upon the total angular momentum **J** means that the accidental degeneracy of the nonrelativistic hydrogen atom is no longer present. This might have been foreseen from the relativistic precession of the hydrogen orbits since that would mean that the Runge vector, to whose constancy the nonrelativistic degeneracy was due, no longer plays that role. However, the fact that the absolute value of κ and not j itself appears in the energy formula means that the relativistic hydrogen atom still has a twofold accidental degeneracy above and beyond the degeneracy due to its spherical symmetry. An indication that it is due to spin effects is given by the fact that the degeneracy indicates an insensitivity of the energy to the relative orientation of spin and angular momenta, according to whether they are parallel or antiparallel.

A constant of the motion explaining the sign degeneracy of $\mathcal{K}$ in the relativistic hydrogen atom was first announced by Johnson and Lippman,[7] but was given an explicit derivation by Biedenharn,[9] who used his coordinate system S to construct the Johnson–Lippman constant explicitly as an operator that reversed the sign of κ and thus inverted the "helicity" (which he called the "Coulomb helicity") of the electron. Coulson and Joseph[14] generalized the analysis to a multidimensional hydrogen atom and interpreted the degeneracy in terms of self-adjoint ladder operators. This is the operator R in (43).

The Johnson–Lippman operator produces degeneracy in a somewhat unconventional manner, since it *commutes* with the Dirac Hamiltonian and anticommutes with $\mathcal{K}$. By constructing normalized operators whose eigenvalues are ± 1, Malkin and Manko[12] formed a set of quaternionic operators which account for the degeneracy doubling in terms of commutation rules, as group theory is usually applied to the study of degeneracy.

Just as it is clear that the fine structure of the hydrogen spectrum can be attributed to the nondegeneracy of the operator Γ, one can see the temptation to construct an approximation to the Coulomb–Dirac Hamiltoniam in which the nonrelativistic degeneracy would be restored. In terms of classical mechanics, we have already done so by introducing the vector potential $Ze\hat{\ell}/r$ in Eq. (6). Due to the importance of spin (i.e., multicomponent) effects in the Dirac equation, it does not seem that such a simple expedient will produce a degenerate relativistic system, but the complications are of the same type. It is necessary to modify the Dirac Hamiltonian in a way which will produce a desired alteration in the iterated equation.

Biedenharn[32] found a way to restore the nonrelativistic degeneracy which he and Swamy expounded in an article introducing their "symmetric" Hamiltonian.[10] A number of features of their restoration merit comment, so we shall review the construction. Their guiding principle was to retain the eigenfunctions of Γ while modifying its eigenvalues to become $\pm\kappa$, a procedure motivated by the fact that Γ reduces to $\rho_3\mathcal{K}$ in the absence of any nuclear charge. The result is to produce a matrix G

$$G = \rho_3\mathcal{K}(1 + \alpha^2Z^2/\mathcal{K}^2)^{\frac{1}{2}} + i\alpha Z\rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} \tag{45}$$

instead of Γ. Now, there is a transformation S_1

$$S_1 = \exp\left[-\tfrac{1}{2}\rho_2\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\ \text{arc sinh}\ (\alpha Z/\mathcal{K})\right] \tag{46}$$

for which

$$S_1GS_1^{-1} = \rho_3\mathcal{K}(1 + \alpha^2Z^2/\mathcal{K}^2)^{\frac{1}{2}}, \tag{47}$$

and we notice that $G^2 = \mathcal{K}^2$.

In the coordinate system, the iterated equation S_1 takes the form

$$(\nabla^2 - 2\alpha ZE/rc + k^2)\Phi = 0, \tag{48}$$

so that the nonrelativistic degeneracy has been restored in such a way that the relativistic screening of the nuclear charge has been retained, but the orbital precession is lost. The energy levels have the form

$$E = -mc^2(1 + \alpha^2Z^2/N^2)^{-\frac{1}{2}}, \tag{49}$$

for $N = 1, 2, 3, \cdots$.

If we ask for the Dirac Hamiltonian which produces G in place of Γ on iteration, we find

$$\mathcal{H}_s = \rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}c\left\{\frac{1}{i}\frac{\partial}{\partial r} + \frac{i}{r}\left[\rho_3\mathcal{K}\left(1 + \frac{\alpha^2Z^2}{\mathcal{K}^2}\right)^{\frac{1}{2}} - 1\right] + \rho_3mc^2 - \alpha Zc/r, \tag{50}$$

or

$$\mathcal{H}_s = \mathcal{H} + \mathcal{H}_{fs},$$

where $\mathcal{H}$ is the ordinary Coulomb–Dirac Hamiltonian, and

$$\mathcal{H}_{fs} = \rho_2(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}/\mathbf{r})c\mathcal{K}[(1 + \alpha^2Z^2/\mathcal{K}^2)^{\frac{1}{2}} - 1] \tag{51}$$

has been called the "fine structure interaction" by Biedenharn and Swamy. Our foregoing discussion should clarify the extent to which it deserves this terminology, but the analogy may be completed by referring to a result by Sheth.[11] He shows that, in the

[32] L. C. Biedenharn, Bull. Am. Phys. Soc. 7A, 314 (1962).

Foldy–Wouthuysen limit, $\mathcal{H}_{fs}$ passes over into a vector potential of the general form $Ze\hat{\ell}/r$. It is interesting to notice that $\mathcal{H}_{fs}$ is invariant under the S_1 transformation.

The term $\mathcal{H}_{fs}$ in the Dirac Hamiltonian which results in the operator G, whose eigenvalues coincide with those of $\mathcal{K}$, may be deduced straightforwardly from assuming that it has the form $c\rho_2(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}/r)f(\mathcal{K})$; it should be stressed, however, that its presence removes some relativistic effects (orbital precession) but not others (nuclear screening). Biedenharn and Swamy have emphasized how the simultaneous presence of both effects in the Coulomb–Dirac Hamiltonian have complicated power series expansions such as the one made by Sommerfeld and Maue[33] in terms of the fine structure constant.

Once the effects of orbital precession have been canceled from the hydrogen atom Hamiltonian, one might hope that the Runge vector would be restored as a constant of the motion. In fact, Biedenharn's iterated symmetric equation coincides with the Klein–Gordon equation for the degenerate classical problem, for which one finds a conserved Runge vector

$$\mathbf{A} = [c/(m^2c^4 - E^2)^{\frac{1}{2}}][(\alpha Z/c)E\hat{\mathbf{r}} + \tfrac{1}{2}(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L})], \tag{52}$$

where E is the energy solution of the stationary degenerate Klein–Gordon equation.

It is to be noted that the energy replaces the mass if this is compared to the classical formula as happens in Eq. (18).

The usual Klein–Gordon equation is a one-component second-order equation, whereas the iterated equation (48) applies to a four-component wavefunction. Grouping the components in pairs according to their magnitudes for positive and negative energies, we still find that there is a complete freedom of spin transformation. Translated into terms of degeneracies, we find that the symmetric Hamiltonian has a $2N^2$-fold degeneracy in the state of principal quantum number N, while the non-relativistic level is N^2-fold degenerate. Again, the doubling due to the spin orientation is present but in a perverse way. If we count the number of states of positive κ in this degenerate level, the number is $N(N-1)$, while belonging to negative κ the number is $N(N+1)$. The discrepancy is due to the fact that κ runs from 1 to $N-1$ when $\kappa = l$, and from -1 to $-N$ when $\kappa = -(l+1)$.

However, the discrepancy means that if R_4 is the symmetry group, the representations which occur are *not* the ones of dimension N^2 that are realizable in terms of hyperspherical harmonics.

The Runge vector (52) itself is not a constant of the motion for the symmetric Hamiltonian since operators which commute with the iterated operator do not necessarily commute with the Hamiltonian itself. However, Biedenharn and Swamy found a vector constant of the motion **K** (see Appendix) as well as a "Coulomb helicity" operator constant of the motion defined as

$$\Lambda = (i\kappa/|\kappa|)(N^2 - \mathcal{K}^2)^{-\frac{1}{2}}(m^2c^2 - \mathcal{H}_B^2)^{-\frac{1}{2}} \times (\rho_3\alpha Z\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\mathcal{H}_B - i\mathcal{K}\boldsymbol{\sigma}\cdot\mathbf{p}c), \tag{53}$$

where $\mathcal{H}_B = S_1\mathcal{H}_sS_1^{-1}$ is the symmetric Hamiltonian in the original coordinate system. This operator is proportional to $\rho_3\boldsymbol{\sigma}\cdot\mathbf{A}'$, where $\mathbf{A}'$ is given by Eq. (52) with $\mathcal{H}_B$ replacing E, and changes the sign of κ when acting on an eigenfunction of $\mathcal{K}$, i.e., it anticommutes with $\mathcal{K}$ and, furthermore, commutes with **J**. Finally, upon defining

$$\mathcal{N} = \Lambda' + \mathcal{K},$$
$$\Lambda' = [m^2c^4 - \mathcal{H}_B^2]^{-\frac{1}{2}}[\rho_3\alpha Z\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\mathcal{H}_B - i\mathcal{K}\boldsymbol{\sigma}\cdot\mathbf{p}c], \tag{54}$$

they proved that

$$\mathbf{K}\times\mathbf{K} = i\mathbf{J}, \tag{55a}$$
$$2\mathbf{J}\cdot\mathbf{K} = \mathcal{N}^2 - \tfrac{1}{2}, \tag{55b}$$

which, with the fact that **K** is a vector and

$$\mathbf{J}\times\mathbf{J} = i\mathbf{J}, \tag{56}$$

define **J** and **K** as the generators of an R_4 group. They also proved that

$$(\mathbf{J}+\mathbf{K})^2 + 1 = [\alpha^2Z^2/(m^2c^4 - \mathcal{H}_B^2)]\mathcal{H}_B^2, \tag{57}$$

in analogy to the classical (23).

Equation (55b) confirms the fact that the representations of this group are not of the same dimension (N^2).

Malkin and Manko [Ref. 12(a)] recently defined three normalized operators

$$\begin{aligned} M_1 &= [(\mathcal{H}_B^2 - m^2c^4) + \alpha^2Z^2\mathcal{H}_B^2/\mathcal{K}^2]^{-\frac{1}{2}} \times [\boldsymbol{\sigma}\cdot\mathbf{p}c - i\alpha Z(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\rho_3\mathcal{K}/\mathcal{K}^2)\mathcal{H}_B], \\ M_2 &= [(\mathcal{H}_B^2 - m^2c^4)\mathcal{K}^2 + \alpha^2Z^2\mathcal{H}_B^2]^{-\frac{1}{2}} \times (\rho_3\alpha Z\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\mathcal{H}_B - i\mathcal{K}\boldsymbol{\sigma}\cdot\mathbf{p}c), \\ M_3 &= |\mathcal{K}|^{-1}\mathcal{K}, \end{aligned} \tag{58}$$

which fulfilled the following relations:

$$\begin{aligned} &[\mathcal{H}_B, M_i] = 0, \\ &[M_i, M_j] = 2_i\epsilon_{ijk}M_k, \\ &\{M_i, M_j\} = 2\delta_{ij}. \end{aligned} \tag{59}$$

[33] A. Sommerfeld and A. W. Maue, Ann. Physik **22**, 629 (1935).

Such relationships can be used in two different ways. The commutation rules are those which define the Lie algebra of the unitary unimodular group SU_2, while the anticommutation rules ensure that the representation which they realize is the two-dimensional one. On the other hand, the anticommutation rules alone show that they form a representation of the quaternions, and since it is the non-Abelian representation, it is again two-dimensional. They have furthermore shown[12b] that similar quaternionic constants of the motion can be constructed both for the planewave and the ordinary Coulomb–Dirac problems. Except for normalization, we see that M_2 and M_3 are just the anticommutating Λ and $\mathcal{K}$, while M_1 is essentially their product which reduces to the chirality $\boldsymbol{\sigma} \cdot \mathbf{p}/|\mathbf{p}|$ when $Z \to 0$.

In other words, the symmetry group for the discrete part of the spectrum is just $d^{(2)}(SU_2) \otimes R_4$; the result which could be expected from the iterated equation (48), for which $\boldsymbol{\sigma}$, $\mathbf{L}$, and $\mathbf{A}$ are constants of the motion.

Some light might be shed by noticing that the states with positive κ define an irreducible representation of R_4 of dimension $N(N-1)$ on one hand, and those with negative κ an irreducible representation of dimension $N(N+1)$ on the other hand; but by adjoining the "Coulomb helicity" inversion operator Λ to this R_4 group, we mix the two representations, and it is the whole set of $2N^2$ states with fixed N which remains irreducible, rather than either of them separately.

IV. RELATIVISTIC SYMMETRY OF THE MAGNETIC MONOPOLE

Aside from the practical interest which exists in the knowledge of the solution of Dirac's equation for the magnetic monopole which is required in order to establish the experimentally observable properties which it might possess, the equation is of theoretical interest for the high symmetry which it possesses. A vector potential is required for its study by Hamiltonian methods, whose presence complicates the symmetry properties, since it may be required to couple a transformation of coordinates with a compensating change of gauge. In fact, as Poincaré[22] noticed as long ago as 1896, the canonical angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ alone is not conserved in the spherically symmetrical field of the monopole, but rather a total angular momentum

$$\mathbf{D} = \mathbf{M} - \epsilon \hat{\mathbf{r}}, \tag{60}$$

where $\mathbf{M}$ is the mechanical angular momentum $\mathbf{M} = \mathbf{r} \times \boldsymbol{\pi}$ and ϵ is the monopole strength.

On account of the relation $\mathbf{D} \cdot \hat{\mathbf{r}} = -\epsilon$, the trajectory of an electrically charged particle moving in the magnetic field due to the monopole will be confined to a cone whose axis is $\mathbf{D}$ and whose half-angle is determined by the pole strength ϵ. In the limiting case of zero pole strength, we recover the accustomed configuration in which a particle, whose canonical angular momentum is conserved, moves in a plane orthogonal to the angular-momentum vector $\mathbf{L}$.

Quantum-mechanical treatment of the motion of a charge in a monopole field originated with Dirac's[15a] speculations in 1928, and Tamm's[15b] solution of its Schrödinger equation. Relativistic treatments began with Banderet's paper[17] of 1946, and Harish-Chandra's[16] of 1948, but few people have treated a nucleus bearing both a magnetic and an electric charge. Malkus[18] considered such a possibility in the Pauli approximation and made an experimental attempt to detect monopoles. Eliezer and Roy[19] studied the shift in energy levels of a hydrogenic atom incorporating a monopole, but apparently did not observe some of the restrictions on quantum numbers arising from the particular properties of the total angular momentum $\mathbf{D}$, which were found by Fierz[34] and later by McIntosh and Cisneros.[20]

The precise importance of these restrictions is that the total angular momentum is bounded below by the value of the quantized pole strength, which is $\epsilon = \frac{1}{2}$ according to Dirac's original speculations, or twice that amount according to more recent field-theoretic arguments of Schwinger.[35] In either event, no s-states of orbital angular momentum are possible so that a particle moving in such a field would not spend an appreciable part of its time in the immediate neighborhood of the origin. One assuredly expects similar results from the Dirac theory, but with the added complication that the spin angular momentum must be taken into account as well.

In the gauge used by McIntosh and Cisneros, and Schwinger, which is not the one used by the other authors cited, the vector potential for the monopole field is

$$\mathcal{A} = [c\epsilon Z/er(x^2 + y^2)](y, -x, 0). \tag{61}$$

The Dirac Hamiltonian would have to be written in the form incorporating the vector potential,

$$\mathcal{H} = \rho_1 \boldsymbol{\sigma} \cdot (\mathbf{p}c - c\mathbf{A}) + \rho_3 mc^2 + V(r), \tag{62}$$

but it is the most enlightening to write both the vector

[34] M. Fierz, Helv. Phys. Acta 17, 27 (1944).
[35] J. Schwinger, Phys. Rev. 144, 1087 (1966).

potential and the Dirac equation in polar coordinates:

$$\mathcal{A}_r = 0, \quad \mathcal{A}_0 = 0, \quad \mathcal{A}_\varphi = (c\epsilon/er)\cot\theta, \tag{63}$$

$$\mathcal{H} = c\rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\left(\frac{1}{i}\frac{\partial}{\partial r} + i\frac{\boldsymbol{\sigma}\cdot\mathcal{M}}{r}\right) + \rho_3 mc^2 + V(r). \tag{64}$$

As to the spherical symmetry of the Hamiltonian, we start from Poincaré's conserved total angular momentum (60) to find its commutator with the Hamiltonian (62):

$$[\mathcal{H}, \mathbf{D}] = -i\rho_1\boldsymbol{\sigma}\times\boldsymbol{\pi}c. \tag{65}$$

The commutator of the monopole Hamiltonian with the spin is the negative value of (65),

$$[\mathcal{H}, \tfrac{1}{2}\boldsymbol{\sigma}] = i\rho_1\boldsymbol{\sigma}\times\boldsymbol{\pi}c, \tag{66}$$

so that the conserved total angular momentum for the Dirac monopole Hamiltonian **J** incorporates the spin angular momentum $\frac{1}{2}\boldsymbol{\sigma}$, the mechanical angular momentum $\mathbf{M} = \mathbf{r}\times\boldsymbol{\pi}$, and the contribution from the magnetic field $-\epsilon\hat{\mathbf{r}}$:

$$\mathbf{J} = \mathbf{D} + \tfrac{1}{2}\boldsymbol{\sigma}. \tag{67}$$

The components of the total angular momentum obey the commutation rules of angular momentum operators,

$$\mathbf{J}\times\mathbf{J} = i\mathbf{J}, \tag{68}$$

and the Casimir operator J^2 is also a constant of the motion with eigenvalues $j(j+1)$. One would be tempted by analogy to the nonmagnetic Hamiltonian to think that $\rho_3(\boldsymbol{\sigma}\cdot\mathbf{D} + 1)$ would be a constant of the motion, but calculation shows that the gods entertained other plans for the monopole. Instead, one should define an operator

$$\mathcal{K} = \rho_3(\boldsymbol{\sigma}\cdot\mathbf{M} + 1) \tag{69}$$

in terms of the *mechanical* angular momentum. From the definition of **D** (69) we find that $\mathbf{M} = \mathbf{D} + \epsilon\hat{\mathbf{r}}$, and that the commutation rules for the components of these two vectors are

$$\mathbf{D}\times\mathbf{D} = i\mathbf{D},$$
$$\mathbf{M}\times\mathbf{M} = i(\mathbf{M} + \epsilon\hat{\mathbf{r}}). \tag{70}$$

With respect to the eventual separation of the Dirac equation in spherical coordinates, we should note that $\mathcal{K}$ differs from its version in which **D** replaces **M** by a term of the form $\epsilon\rho_3\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}$, which has only radial components in the σ space. We may therefore expect to classify the wavefunctions with the eigenvalues of D^2 and σ^2 playing the roles usually held by L^2 and σ^2. There are further relations between $\mathcal{K}^2$ and J^2. We have

$$\begin{aligned} D^2 &= M^2 + \epsilon^2, \\ J^2 &= D^2 + \boldsymbol{\sigma}\cdot\mathbf{D} + \tfrac{3}{4}, \\ \mathcal{K}^2 &= M^2 + \rho_3\mathcal{K} - \epsilon\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}, \end{aligned} \tag{71}$$

so that

$$J^2 = \mathcal{K}^2 + \epsilon^2 - \tfrac{1}{4}. \tag{72}$$

The charged monopole is defined by the electric and magnetic fields

$$\mathcal{E} = -(Ze/r^2)\hat{\mathbf{r}}, \quad \mathbf{B} = -(c\epsilon/er^2)\hat{\mathbf{r}}. \tag{73}$$

To form the iterated equation, we introduce the projection operators

$$\begin{aligned} \mathcal{O}_+ &= (1/2mc^2)[i\rho_2\boldsymbol{\sigma}\cdot\boldsymbol{\pi}c + mc^2 - \rho_3(E + c\alpha Z/r)], \\ \mathcal{O}_- &= (1/2mc^2)[i\rho_2\boldsymbol{\sigma}\cdot\boldsymbol{\pi}c - mc^2 - \rho_3(E + c\alpha Z/r)], \end{aligned} \tag{74}$$

to obtain

$$\begin{aligned} 4m^2c^2\mathcal{O}_+\mathcal{O}_- = {} & \pi_r^2 - M^2/r^2 - (e/c)\boldsymbol{\sigma}\cdot \\ & - i(e/c)\rho_1\boldsymbol{\sigma}\cdot\boldsymbol{\epsilon} + k^2 + \alpha^2Z^2/r^2 - 2\alpha ZE/rc, \end{aligned} \tag{75}$$

where

$$k^2 = E^2/c^2 - m^2c^2.$$

Again it is convenient to introduce Temple's operator

$$\Gamma = \rho_3\mathcal{K} + i\alpha Z\rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} \tag{76}$$

which incorporates the term $(e/c)\boldsymbol{\sigma}\cdot\mathbf{B}$ in its square,

$$\Gamma^2 = M^2 + \rho_3\mathcal{K} - \epsilon\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} - \alpha^2Z^2, \tag{77}$$

and using the fact that $\rho_3\mathcal{K}$ anticommutes with $\rho_1\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}$, the iterated equation reads

$$(p_r^2 - \Gamma(\Gamma - 1)/r^2 - 2\alpha ZE/cr + k^2)\Phi = 0. \tag{78}$$

This expression is exactly the one we encountered in the absence of the monopole, with the unique distinction that $\mathcal{K}$ is defined by Eq. (69) rather than Eq. (28). The solution of the iterated equation would then be effected by diagonalizing Γ, solving the resulting uncoupled radial wave equations, and projecting the result.

In terms of **J**, we get from Eqs. (71) and (72) that

$$\Gamma^2 = J^2 + \tfrac{1}{4} - \epsilon^2 - \alpha^2Z^2. \tag{79}$$

Its eigenvalues are then given by

$$\begin{aligned} j^2 &= j(j+1) + \tfrac{1}{4} - \epsilon^2 - \alpha^2Z^2 \\ &= (j + \tfrac{1}{2})^2 - \epsilon^2 - \alpha^2Z^2. \end{aligned} \tag{80}$$

Diagonalization in ρ space may be effected by the Foldy–Wouthuysen-type transformation introduced by Biedenharn:

$$S = \exp\left[-\tfrac{1}{2}\rho_2 \boldsymbol{\sigma}\cdot\hat{\mathbf{r}} \operatorname{arc\,tanh}(\alpha Z/\mathcal{K})\right]. \quad (81)$$

Then $S\Gamma S^{-1}$ is free of odd operators, and now takes the form

$$S\Gamma S^{-1} = \rho_3 \mathcal{K}(1 - \alpha^2 Z^2/\mathcal{K}^2)^{\frac{1}{2}}. \quad (82)$$

It will be completely diagonal when we have diagonalized $\mathcal{K}$, whose eigenvalues are $\pm[(j+\frac{1}{2})^2 - \epsilon^2]^{\frac{1}{2}}$. Since $\mathcal{K}$ is a Hermitian matrix, we know that these eigenvalues must be real, which imposes the condition

$$j + \tfrac{1}{2} \geq |\epsilon|. \quad (83)$$

However, $\mathcal{K}$ will not be invertible when $j + \frac{1}{2} = |\epsilon|$, a critical case which has to be treated very carefully since it results in a negative value of γ^2 [Eq. (80)]. It leads to spiral orbits for every nonzero value of Z while without the magnetic charge such complications did not arise until Z surpassed 137. Since j is a semi-integer or integer as the eigenvalue of a Casimir operator of R_3, this limitation (83) to its values is probably the most important consequence of the presence of the magnetic charge.

Substituting the above expression for the eigenvalues of $\mathcal{K}$ and thence Γ into the iterated equation (78), we finally obtain an expression for the admissible discrete energy eigenvalues:

$$E = mc^2\left(1 + \frac{\alpha^2 Z^2}{\{n + [(j+\frac{1}{2})^2 - \epsilon^2 - \alpha^2 Z^2]^{\frac{1}{2}}\}^2}\right)^{-\frac{1}{2}} \quad (84)$$

for $j + \frac{1}{2} > |\epsilon|$. This formulation should be compared with the Sommerfeld fine-structure formula, and we are to find that the fine structure αZ due to relativistic precession and the fine structure ϵ due to the confinement of the monopole orbits to a conic surface combine according to the Pythagorean relation for right triangles.

The previously known quantization of the magnetic pole strength for the Schrödinger equation for the monopole is again confirmed here because one encounters the same angular wave equation as in the other problems, and so no new arguments may be contributed to this phenomenon.

The Dirac equation for the monopole exhibits its doublet degeneracy which is so characteristic, and we may readily adapt the Lippman–Johnson constant to apply here as well. It is formally the same:

$$R = \boldsymbol{\sigma}\cdot\hat{\mathbf{r}} - i\mathcal{K}\rho_1(\mathcal{H} - \rho_3 mc^2)/\alpha Z mc^2, \quad (85)$$

but with the definition (69) for $\mathcal{K}$ appropriate to the monopole. As R and $\mathcal{K}$ anticommute while R and $\mathcal{H}$ commute, we are assured that the eigenvalues of R as well as those of $\mathcal{K}$ occur in negative pairs, belonging in each case to the same energy, and the same technique of Malkin and Manko can be employed to construct the chiral algebra, with the appropriate definition of $\mathcal{K}$.

In order to obtain an algebra of quaternions, it is necessary to divide the operators $\mathcal{K}$ and R by their absolute values so as to obtain operators with eigenvalues ± 1, which assumes that zero is not an eigenvalue of $\mathcal{K}$. This condition was met for the purely electrostatic potential of the hydrogen atom, but not for the lowest angular-momentum state of the monopole. One may still obtain a quaternion algebra if the states of zero eigenvalue of $\mathcal{K}$ are excluded, but it must be remembered that the conclusions regarding their doubling will no longer apply, and that only the remaining states will show accidental degeneracy.

In view of the close resemblance of the Dirac equations for a charged nucleus both with and without a magnetic charge, it is hardly surprising that they can both be degenerated by the methods of Biedenharn and Swamy, although there are slight differences in the dimensionalities of the R_4 representations which occur in the two cases. Even the classical relativistic treatment are closely similar. Although McIntosh and Cisneros degenerated the monopole problem by introducing a repulsive centrifugal potential ϵ^2/r^2 to counteract the attractive tendencies arising from the separation of the variables, to make such an alteration here would introduce inverse fourth-power terms in the iterated Dirac equation. So, we are forced to resort to a vector rather than an electrostatic potential, precisely the same strategem by which we handled the relativistic precession. The required classical Hamiltonian is

$$\mathcal{H} = \left[\left(\boldsymbol{\pi} - \frac{(\alpha^2 Z^2 + \epsilon^2)^{\frac{1}{2}}}{r^2}\hat{M}\right)^2 c^2 + m^2 c^4\right]^{\frac{1}{2}} - \frac{Ze^2}{r}. \quad (86)$$

The perturbing vector potential is now of magnitude $(\alpha^2 Z^2 + \epsilon^2)^{\frac{1}{2}}/r^2$, the total rate of precession, and parallel to the mechanical angular momentum **M**, rather than to the canonical angular momentum **L**.

V. SUMMARY

We have studied the symmetry and degeneracy properties of relativistic Coulombic system in which a fixed attracting center may bear both electric and magnetic charge, in the fields of which a charged

particle, such as the electron, is supposed to move. Such a system may be treated classically—its Hamilton–Jacobi equation being soluble by separation in spherical polar coordinates. It is found that the high symmetry and degeneracy usually associated with the electrostatic Coulomb potential is lost, and that only the degeneracy due to the spherical symmetry remains. The removal of the accidental degeneracy can be attributed to a term in the Hamilton–Jacobi equation, quadratic in $1/r$, which consists of two parts. One is due to the relativistic precession of the orbits coming from the variation of the electron mass with velocity, and the other is due to the fact that in a monopole field, the particle moves on the surface of a cone rather than a plane, the result of which difference in geometry is also an apparent orbital precession.

Since the lack of degeneracy can be attributed so clearly to a term of the nature of a potential energy, classically it is a simple matter to alter the potential and re-establish a degenerate system, admittedly distinct from the original one. Since the alteration must be made through the introduction of some potential directly into the relativistic Hamiltonian and not into an equivalent form of the Hamilton–Jacobi equation, a supplementary (velocity dependent, as it turns out) vector potential can be introduced, with the desired result. It is therefore possible to study some "symmetric" relativistic systems which, in spite of their being relativistic and containing a magnetically charged nucleus, have bounded closed orbits. There should be a constant vector such as the Runge vector, which we have also found to be the case, and the necessary modifications to the familiar vector seem perfectly natural to the new context. The mass must be replaced by its relativistically covariant equivalent, the Hamiltonian itself; also, the angular momentum of the magnetic field must always be incorporated in the angular-momentum vector.

There is relatively little effort required to establish similar results for the Klein–Gordon equation, where again the accidental degeneracy of the hydrogen atom is lost and can be restored by considering the alternative system. Likewise, the classical constants can be established in operator form, and the degeneracy of the "symmetric" variant is deduced as the consequence of a hidden symmetry. But, since the Klein–Gordon equation is not applicable to particles such as electrons, the Dirac equation must be considered, and with it the complications of a wavefunction with several components.

A comprehensive survey of the symmetry properties of the Dirac equation has not yet been given. Indeed, there are relatively few systems for which it has even been solved explicitly; Stanciu[36] has enumerated them and recently added a few new ones. With the exception of the Coulomb potential, the soluble fields are all electromagnetic fields of some especially high degree of symmetry, such as a uniform electric[37] or magnetic field,[8,38] or the field of a monochromatic plane wave.[39] Plesset[2] showed in 1930 that Klein's paradox[3] would be exhibited by all electrostatic potentials expressible as a finite power series in r or $1/r$, there being no bound states in any of them with the sole exception of the Coulomb potential. Nikolsky[40] and Postepska[41] confirmed this behavior of the harmonic oscillator potential, which is the other highly degenerate potential in nonrelativistic theory. Titchmarsh[42] devoted quite a few of his last papers to the effect and the completeness properties of solutions of Dirac's equation.

In those Dirac equations for which solutions are known, the spin doubling is a pervasive feature. We have seen that for the Coulombic equations, including the monopole field, it was forced by the existence of an anticommuting pair of operators, one of them the Johnson–Lippman constant and the other being Dirac's $\mathcal{K}$. Every spherical potential is bound to have $\mathbf{J}$ as a constant of the motion, generating rotations and ensuring the presence of a spin angular momentum. But, since we are always dealing with the direct product of a configuration and a spin space, the representations of $\mathbf{J}$ will always be reducible. The Kronecker product will always contain a constituent of angular momentum $j = l + \frac{1}{2}$, and one of angular momentum $j = l - \frac{1}{2}$ (l being greater than zero). The constant of the motion $\mathcal{K}$ owes its existence to the possibility of always distinguishing these two constituents and the Hamiltonian containing no matrix elements connecting them. This does not imply that they should have the same energy, however, and therein lies the accidental degeneracy of the Dirac equation for the hydrogen atom. The Johnson–Lippman operator establishes such a degeneracy through a constant of the motion, while the chiral algebra of Malkin and Manko[12] formalizes the degeneracy in group-theoretical terms.

As far as we know, it is still an open question as to

[36] G. N. Stanciu, Phys. Letters **23**, 232 (1966); J. Math. Phys. **8**, 2043 (1967).
[37] F. Sauter, Z. Physik **69**, 742 (1931).
[38] I. I. Rabi, Z. Physik **49**, 507 (1928).
[39] D. M. Wolkow, Z. Physik **94**, 250 (1935); N. D. Sengupta, Bull. Calcutta Math. Soc. **39**, 147 (1947); A. H. Taub, Rev. Mod. Phys. **21**, 388 (1949).
[40] K. Nikolsky, Z. Physik **62**, 677 (1930).
[41] I. Postepska, Acta Phys. Polon. **4**, 269 (1935).
[42] E. C. Titchmarsh, Quart. J. Math. (Oxford) **15**, 193 (1964); **14**, 65, 147 (1963); **13**, 181, 255 (1962); **12**, 227 (1961); Proc. Roy. Soc. (London) **A262**, 489 (1961); **A266**, 33 (1962); **A245**, 147 (1958); Quart. J. Math. (Oxford) **13**, 1 (1942). See also related work of W. D. Evans, *ibid.* **17**, 211, 345 (1966).

whether the existence of the Malkin–Manko algebra extends beyond Coulombic systems, since the exact form of the Coulomb potential has to be used in verifying the commutation rules of the Johnson–Lippman constant, the Hamiltonian, and the other members of the algebra. The degeneracy at least appears to be more an attribute of the spin than the Coulomb potential, and a similar chiral algebra can be formed for a plane wave (a limiting case of the hydrogen atom as the nuclear charge vanishes).

Turning now to the degenerated versions of the Coulomb–Dirac equation, we have seen that a slightly different treatment is needed than simply introducing the vector potential which served for the classical relativistic systems, even though the perturbing potential approximates the latter in the Foldy–Wouthuysen limit. But the presence of the spin creates some unique complications. It is one thing to restore the missing degeneracy, but quite another to recover the same symmetry, since attention has to be paid to the way in which the degenerate levels are distributed between the parallel and antiparallel chiral states. Here the splitting tends to be unsymmetrical, because the total number of states in the antiparallel set ($\kappa > 0$) is always lower than the total number of states in the parallel set ($\kappa < 0$).

This asymmetrical grouping of the degenerate states implies that the Runge vector will not be translated into a constant of the Dirac equation unless it incorporates spin operators in some way. If our intuition is correct that a change of spatial coordinates must be coupled with a mixing of spin components, we have only to think of the infinitesimal transformations generated by the Runge vector to see that more may be required than simply replacing $\mathbf{L}$ by $\mathbf{J}$ in the classical formula for the Runge vector.

In fact, Biedenharn and Swamy's construction of the second vector constant for the symmetric Hamiltonian left us completely without any concrete physical interpretation for it or its components at all. One knows that there are several ways of constructing a constant of the motion for the Dirac equation from known constants of the Klein–Gordon equation. Fortunately for the symmetric Hamiltonian, the iterated Dirac operator is actually the Klein–Gordon operator, so that the latter is available as a starting point. For example, the mass projection $\mathcal{O}_{-}A\mathcal{O}_{-}$ of the Runge vector must be a constant of the Dirac Hamiltonian, but there is no assurance that its components still satisfy the same commutation rules as the original Runge vector for the Klein–Gordon equation. In any event, the relation $\mathbf{L} \cdot \mathbf{A} = 0$ cannot be preserved, for this is the quantity which determines the rectangularity of the dimensions of the irreducible representation of R_4. The dimension is $(2m + 1) \times (2n + 1)$ with m and n two integers and $m - n = 2\mathbf{L} \cdot \mathbf{A}$, a quantity which we know changes when going from the Klein–Gordon to the Dirac equation.

Anyhow, mass projection was not the approach adopted by Biedenharn and Swamy, but it is still reassuring to know that the existence of constants for the Klein–Gordon equation implies constants for the Dirac equation as well.

Latent in all discussions of the symmetry of the Dirac equation is the problem of its singularity for certain ranges of its parameters. When one is only interested in obtaining the energy levels and wavefunctions of the naturally occurring nuclei, the problem can be put aside because it requires a superheavy nucleus, whose atomic number lies beyond 137, to make the Dirac equation singular. Even then, one frequently encounters the reasoning that the singularity owes its origin to the assumption of a point nucleus, and that if one were actually interested in superheavy nuclei, one also ought to take into account that the nuclear size corresponds to the extent of the lowest wavefunction and abandon the assumption of a point nucleus.

Reasonable as such an attitude might be from the purely pragmatic point of view, the issue of a singular differential operator poses a legitimate mathematical problem which, were it satisfactorially resolved, would still allow us to retain the convenient idealization of a point charge. Singular Hamiltonians are known in other contexts, one of the simplest being that of the motion of a particle in an attractive potential of the form q/r^2. It was treated by Shortley[43] in 1931, and some similar potentials by Jaffé[44] in 1930. The classical motion in such a potential is spiral in nature and is noteworthy for its dilational symmetry. Its Schrödinger equation possesses a square-integrable eigenfunction for every negative energy,[45] but not every pair of such eigenfunctions is orthogonal.

It was once thought that such motion could not be quantized, and indeed in the old quantum mechanics, one finds that the action integral diverges, preventing quantization by the Bohr–Sommerfeld rules. However, the philosophy of quantization is satisfied if we can pick out a complete orthonormal set of eigenfunctions, so that our only conclusion is that the criterion of square-integrability is not sufficient for quantization when dealing with singular potentials. Case[46] has

[43] G. H. Shortley, Phys. Rev. **38**, 120 (1931).
[44] G. Jaffé, Z. Physik **66**, 748 (1930).
[45] E. A. Guggenheim, Proc. Phys. Soc. (London) **89**, 491 (1966).
[46] K. M. Case, Phys. Rev. **80**, 797 (1950).

proposed one scheme, wherein he has shown that the selection of wavefunctions with a fixed phase at the origin will yield an orthonormal set, and he has shown that this procedure is consistent with the more straightforward cutoff procedures for the potential. Scarf[47] has proposed another technique which is more elegant from the point of view of complex variable theory, and claims to obtain eigenvalues similar to those obtained by the cutoff approach. From the physical point of view, one must be prepared to deal with complex eigenvalues in order to use his technique.

The fact that supplementary information to the requirement of square integrability is required to quantize the bound states of the singular potentials has its classical counterpart, since Newton's equations do not completely define the motion of the particle as it passes through the singular point of the potential. Such behavior is probably representative of the difficulties encountered in relativistic motion as well, where the Hamiltonian equations do not apply at the attracting center itself.

The results we have obtained show that there is some considerable variation in the range of parameters over which the Dirac equation will be singular. Without a magnetic charge on the nucleus, singularity only occurs for superheavy nuclei, $Z > 137$ (for the Klein–Gordon equation, the singularity first occurs at half this atomic number, $Z > 68$), while at multiples of this critical atomic number, additional angular-momentum levels become singular. When there is a magnetic charge at the nucleus, in the Dirac equation, the lowest angular-momentum level is singular. The angular momentum used is the combination of orbital and magnetic field angular momentum **D**, whose value determines the lowest possible value of the total angular momentum j. The singularity occurs when the spin and minimum **D** are antiparallel, leaving the relativistic precession as the dominant effect. In this context, the spin is responsible for the singularity, which does not occur in the nonrelativistic treatment where the angular momentum in the magnetic field is always sufficient to repel the particle from the origin, even when the orbital angular momentum is zero. Nor does one find the singularity in the Klein–Gordon equation, where spin effects are also absent, at least for low atomic numbers.

The singularity would still be very mild were it not for the relativistic precession, so it is not surprising that the singularities disappear from the degenerated Hamiltonians, both with and without magnetic charge. Even with a magnetic charge one finds the full spin doubling, in contrast to the "natural" Hamiltonians, for which the singular states fail to show the accidental degeneracy due to the spin doubling.

We have paid little attention to the scattering states of a charged monopole, partly because the bound states are more amenable to analysis in terms of symmetry and degeneracy. However, the formulas which we have derived admit positive as well as negative energies and, in principle, all energy values are included. The principal difference lies in the commutation rules for the R_4 symmetry group, which must be changed to those of the Lorentz group for scattering states. We shall not belabor the point further.

[47] F. L. Scarf, Phys. Rev. **109**, 2170 (1958).

ACKNOWLEDGMENTS

We gratefully acknowledge the hospitality and stimulating atmosphere of the Quantum Chemistry Group, for which we express our sincere thanks to Professor Per-Olov Löwdin. Dr. N. V. V. J. Swamy has helpfully provided us with some further details concerning his joint paper with Professor Biedenharn, and Drs. I. A. Malkin and V. I. Manko have sent preprints of several of their papers, which kindnesses have benefited this work.

APPENDIX

Given a constant of the motion for the iterated equation, we construct a constant of the motion for the Dirac equation in the following way:

Assume

$$[\mathcal{O}_+\mathcal{O}_-, T] = 0. \tag{A1}$$

As $[\mathcal{O}_+, \mathcal{O}_-] = 0$, we find

$$\mathcal{O}_-[\mathcal{O}_+, T] + [\mathcal{O}_-, T]\mathcal{O}_+ = 0,$$
$$\mathcal{O}_+[\mathcal{O}_-, T] + [\mathcal{O}_+, T]\mathcal{O}_- = 0. \tag{A2}$$

Furthermore, as $\mathcal{O}_+$ and $\mathcal{O}_-$ differ by unity, we have

$$[\mathcal{O}_+, T] = [\mathcal{O}_-, T], \tag{A3}$$

so adding both Eqs. (88), we get

$$\{[\mathcal{O}_+, T], \mathcal{O}_+ + \mathcal{O}_-\} = 0. \tag{A4}$$

In order to find $\mathcal{O}_+$ and $\mathcal{O}_-$, it is convenient to transform to the S_1 reference frame, so we define

$$\mathcal{K}_B = S_1\mathcal{K}_S S_1^{-1}, \tag{A5}$$

which reduces to

$$\mathcal{K}_B = [(1 + \alpha^2 Z^2/\mathcal{K}^2)^{\frac{1}{2}} - \rho_2\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\alpha Z/\mathcal{K}] \times (\rho_1\boldsymbol{\sigma}\cdot\mathbf{p}c - \rho_3 mc^2)$$
$$= S_1^2(\rho_1\boldsymbol{\sigma}\cdot\mathbf{p}c - \rho_3 mc^2), \tag{A6}$$

and, in this frame, Dirac's equation can be expressed as

$$(i\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}c - mc^2 - \rho_3 E S_1^{-2})\Psi = 0, \quad \text{(A7)}$$

and one gets

$$\begin{aligned}\mathcal{O}_+ &= (1/2mc^2)\rho_3 S_1^{-2}(\mathcal{H}_B - E)\\ &= (1/2mc^2)(i\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}c - mc^2 - \rho_3 E S_1^{-2}),\end{aligned} \quad \text{(A8)}$$

and similarly for $\mathcal{O}_-$. We notice that $\mathcal{O}_+ + \mathcal{O}_-$ contains a term proportional to ρ_2 and a term proportional to ρ_3, so it anticommutes with ρ_1 and, from Eq. (90), we get

$$[\rho_1[\mathcal{O}_+, T], \mathcal{O}_+ + \mathcal{O}_-] = 0, \quad \text{(A9)}$$

or, explicitly,

$$[\rho_1[\rho_3 S_1^{-2}(\mathcal{H}_B - E), T], \rho_3 S_1^{-2}(\mathcal{H}_B - E)] = 0, \quad \text{(A10)}$$

where we have used the fact that $\mathcal{O}_+ + \mathcal{O}_- = 2\mathcal{O}_+ + 1$. When applied to an eigenfunction Ψ, we get, using Dirac equation,

$$[\rho_1\rho_3 S_1^{-2}[\mathcal{H}_B, T], \mathcal{H}_B]\Psi = 0 \quad \text{(A11)}$$

independently of E, and since the eigenfunctions form a complete set, we have proven that $\rho_2 S_1^{-2}[\mathcal{H}_B, T]$ is a constant of the motion, since it commutes with $\mathcal{H}_B$, in the S_1 frame.

Regarding the choice of the Klein–Gordon equation's constant T with which we start, we quote Biedenharn and Swamy's argument, taking $\mathbf{T} = \boldsymbol{\sigma} \times \mathbf{L}$, as this is orthogonal to $\mathbf{J}$, and when applied to a function of definite κ, it will produce that with $-\kappa - 1$; so, with the help of Λ [Eq. (53)], which reverses the sign of κ, one can construct ladder operators for the eigenvalue κ, which, with appropriate coefficients, yield a constant vector $\mathbf{K}$ as the sum of three parts—one for the lowering operator, one for the raising operator, and $\mathbf{J}$ which satisfies the Runge vector commutation rules (see Ref. 10 for details).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

High-Temperature Properties of Random Spin Systems

G. GALLAVOTTI
The Rockefeller University, New York, New York 10021

(Received 19 March 1969)

The free energy and the correlation functions are proved to be analytic functions of the inverse temperature β, the external magnetic field H, and the impurity probability p, provided the temperature is high enough and the spin–spin interaction has finite range. In particular, this excludes spontaneous magnetization at high temperatures.

1. INTRODUCTION

The aim of this paper is to show how some general techniques developed to analyze the high-temperature properties of lattice gases and magnets[1,2] can be applied to random spin systems.

The random spin systems that will be considered here are of the type recently studied in Ref. 3; their precise description follows.

Consider a γ-dimensional lattice Z and a finite cubic portion of it K, and suppose that spins are located on the sites of a randomly chosen sublattice $D \subset K$; let $D = (x_1, \cdots, x_{N(D)})$ and suppose that the energy of a spin configuration $(s_1, \cdots, s_{N(D)})$—$s_i = \pm 1$ and the spin s_i is located at x_i—is given by

$$U(s_1, \cdots, s_{N(D)}) = H\sum_i s_i + \sum_{i<j}\varphi(x_i - x_j)s_i s_j, \quad (1)$$

where H is the external magnetic field and the spin–spin potential φ is restricted to be a finite-range potential.

Furthermore, we suppose that the probability of the random set $D \subset K$ is given by

$$P_K(D) = p^{N(D)}(1-p)^{N(K)-N(D)};$$

i.e., interpreting the sites outside of D as points occupied by impurities, we suppose that the probabilities for a given point in K to be occupied by a spin or by an impurity are, respectively, p and $(1-p)$.

The free energy can be defined as the average of the

[1] R. L. Dobrushin (unpublished); G. Gallavotti, S. Miracle-Sole, and D. Robinson, Phys. Letters **25A**, 693 (1967).
[2] J. Ginibre, Commun. Math. Phys. **10**, 140 (1968); G. Gallavotti, S. Miracle-Sole, and D. Robinson, Commun. Math. Phys. **10**, 311 (1968); W. Greenberg, Commun. Math. Phys. **11**, 314, 1969.
[3] R. B. Griffiths and J. L. Lebowitz, J. Math. Phys. **9**, 1284 (1968).

and, in this frame, Dirac's equation can be expressed as

$$(i\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}c - mc^2 - \rho_3 E S_1^{-2})\Psi = 0, \quad \text{(A7)}$$

and one gets

$$\begin{aligned}\mathcal{O}_+ &= (1/2mc^2)\rho_3 S_1^{-2}(\mathcal{H}_B - E)\\ &= (1/2mc^2)(i\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}c - mc^2 - \rho_3 E S_1^{-2}),\end{aligned} \quad \text{(A8)}$$

and similarly for $\mathcal{O}_-$. We notice that $\mathcal{O}_+ + \mathcal{O}_-$ contains a term proportional to ρ_2 and a term proportional to ρ_3, so it anticommutes with ρ_1 and, from Eq. (90), we get

$$[\rho_1[\mathcal{O}_+, T], \mathcal{O}_+ + \mathcal{O}_-] = 0, \quad \text{(A9)}$$

or, explicitly,

$$[\rho_1[\rho_3 S_1^{-2}(\mathcal{H}_B - E), T], \rho_3 S_1^{-2}(\mathcal{H}_B - E)] = 0, \quad \text{(A10)}$$

where we have used the fact that $\mathcal{O}_+ + \mathcal{O}_- = 2\mathcal{O}_+ + 1$. When applied to an eigenfunction Ψ, we get, using Dirac equation,

$$[\rho_1\rho_3 S_1^{-2}[\mathcal{H}_B, T], \mathcal{H}_B]\Psi = 0 \quad \text{(A11)}$$

independently of E, and since the eigenfunctions form a complete set, we have proven that $\rho_2 S_1^{-2}[\mathcal{H}_B, T]$ is a constant of the motion, since it commutes with $\mathcal{H}_B$, in the S_1 frame.

Regarding the choice of the Klein–Gordon equation's constant T with which we start, we quote Biedenharn and Swamy's argument, taking $\mathbf{T} = \boldsymbol{\sigma} \times \mathbf{L}$, as this is orthogonal to $\mathbf{J}$, and when applied to a function of definite κ, it will produce that with $-\kappa - 1$; so, with the help of Λ [Eq. (53)], which reverses the sign of κ, one can construct ladder operators for the eigenvalue κ, which, with appropriate coefficients, yield a constant vector $\mathbf{K}$ as the sum of three parts—one for the lowering operator, one for the raising operator, and $\mathbf{J}$ which satisfies the Runge vector commutation rules (see Ref. 10 for details).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

High-Temperature Properties of Random Spin Systems

G. GALLAVOTTI
The Rockefeller University, New York, New York 10021

(Received 19 March 1969)

The free energy and the correlation functions are proved to be analytic functions of the inverse temperature β, the external magnetic field H, and the impurity probability p, provided the temperature is high enough and the spin–spin interaction has finite range. In particular, this excludes spontaneous magnetization at high temperatures.

1. INTRODUCTION

The aim of this paper is to show how some general techniques developed to analyze the high-temperature properties of lattice gases and magnets[1,2] can be applied to random spin systems.

The random spin systems that will be considered here are of the type recently studied in Ref. 3; their precise description follows.

Consider a γ-dimensional lattice Z and a finite cubic portion of it K, and suppose that spins are located on the sites of a randomly chosen sublattice $D \subset K$; let $D = (x_1, \cdots, x_{N(D)})$ and suppose that the energy of a spin configuration $(s_1, \cdots, s_{N(D)})$—$s_i = \pm 1$ and the spin s_i is located at x_i—is given by

$$U(s_1, \cdots, s_{N(D)}) = H\sum_i s_i + \sum_{i<j}\varphi(x_i - x_j)s_i s_j, \quad (1)$$

where H is the external magnetic field and the spin–spin potential φ is restricted to be a finite-range potential.

Furthermore, we suppose that the probability of the random set $D \subset K$ is given by

$$P_K(D) = p^{N(D)}(1-p)^{N(K)-N(D)};$$

i.e., interpreting the sites outside of D as points occupied by impurities, we suppose that the probabilities for a given point in K to be occupied by a spin or by an impurity are, respectively, p and $(1-p)$.

The free energy can be defined as the average of the

[1] R. L. Dobrushin (unpublished); G. Gallavotti, S. Miracle-Sole, and D. Robinson, Phys. Letters **25A**, 693 (1967).
[2] J. Ginibre, Commun. Math. Phys. **10**, 140 (1968); G. Gallavotti, S. Miracle-Sole, and D. Robinson, Commun. Math. Phys. **10**, 311 (1968); W. Greenberg, Commun. Math. Phys. **11**, 314, 1969.
[3] R. B. Griffiths and J. L. Lebowitz, J. Math. Phys. **9**, 1284 (1968).

free energies that correspond to the various impurity distributions; i.e., introducing $F(D)$ as

$$-\beta F(D) = \log \sum_{(s_1,\cdots,s_{N(D)})} \exp\left[-\beta U(s_1,\cdots,s_{N(D)})\right], \tag{2}$$

the free energy of our random spin system (when it is confined in the region K) is

$$F_K = \sum_{D\subset K} P_K(D)F(D) = \sum_{D\subset K} p^{N(D)}(1-p)^{N(K)-N(D)}F(D), \tag{3}$$

and this formula concludes the description of the model. (For a discussion of the physical relevance of the model, we refer the reader to Ref. 3.)

If β denotes (as above) the inverse temperature, p the impurity probability, and $\zeta = \exp(-2\beta H)$, then our main results can be summarized in the following theorem:

Theorem 1: Let $f(\beta, \zeta, p)$ denote the mean free energy defined for $\beta > 0$, $\zeta > 0$, $0 \le p \le 1$ as

$$f(\beta,\zeta,p) = \lim_{K\to\infty} N(K)^{-1}F_K; \tag{4}$$

then $\beta f(\beta, \zeta, p)$ can be extended to an analytic function of β, ζ, p in the region Q_f defined by:

(i) β is contained in a sufficiently small neighborhood of $\operatorname{Im}\beta = 0$, $|\operatorname{Re}\beta| \le \beta_0$, with β_0 a suitable constant;

(ii) ζ is in the complement of a (closed) circle lying in the half-plane $\operatorname{Re}\zeta < 0$ and centered on the real axis; furthermore, $\zeta \neq 0$;

(iii) p is inside the circle $|p| < 1$.

The same analyticity properties hold for the correlation functions, but with an analyticity domain Q obtained from Q_f by suppressing the condition $\zeta \neq 0$.

Item (ii) leaves the possibility of a singularity at $\zeta = 0$. This singularity indeed exists, but it is a trivial nonthermodynamic singularity. The proof of these results is divided in two steps: The first is the transformation of the problem into a lattice gas problem and the application of already-known high-temperature techniques to this random lattice gas to obtain expressions for the correlation functions in the limit $K \to \infty$ (Secs. 2 and 3); the second step (Sec. 4) is the study of the analytic properties of the series expansions obtained in Sec. 3. As a by-product of the proof of Sec. 4, we also find an upper bound on the critical temperature (Sec. 5). The proofs in Secs. 3 and 4 are rather technical, but the key point is the remark after formula (26).

2. THE EQUIVALENT LATTICE-GAS PROBLEM

The random spin system can be easily transformed into a lattice gas whose particles can occupy sites in a subset $D \subset K$ chosen at random with probability $P_K(D) = p^{N(D)}(1-p)^{N(K)-N(D)}$. Since the passage from a random spin system to a random lattice gas is accomplished through the simple and well-known algebraic transformations used for ordinary spin systems, we skip the detailed derivation which, anyway, is outlined in Ref. 3; the result is that the equivalent random lattice gas is described by a pair potential Φ and a position- and impurity-dependent chemical potential $\nu_x(D)$ defined, respectively, as

$$\Phi(x-y) = 4\varphi(x-y), \tag{5}$$

$$\nu_x(D) = -2H + \tfrac{1}{2}\sum_{y\in D, y\neq x}\Phi(x-y) = -2H + A_x(D), \tag{6}$$

where $A_x(D)$, which will be needed later, is implicitly defined by (6).

The potential energy of a configuration $Y = (y_1,\cdots,y_{N(Y)})$ is then given by

$$U(Y;D) = \sum_{y\in Y}\nu_y(D) + \sum_{i<j}\Phi(y_i - y_j). \tag{7}$$

The problem of studying the analytic properties of the free energy f and of the spin-correlation functions becomes, now, the problem of studying the analytic properties of the pressure Π and the correlation functions ρ of the "random" gas defined through the formulas below. In these formulas the symbols Y, T, $\cdots$ denote configurations $(y_1,\cdots,y_{N(Y)})$, $(t_1,\cdots,t_{N(T)})$, and the definitions (5), (6), and (7) are used:

$$\Pi = \lim_{K\to\infty}\Pi_K, \tag{8}$$

$$\Pi_K = N(K)^{-1}\sum_{D\subset K}P_K(D)\log\sum_{Y\subset D}\zeta^{N(Y)} \times \exp\left[-\beta U(Y;D)\right], \tag{9}$$

$$\rho(Y) = \lim_{K\to\infty}\sum_{D\subset K}P_K(D)\rho_D(Y), \tag{10}$$

$$\rho_D(Y) = 0, \quad \text{if } Y \not\subset D,$$
$$= \sum_{Y\subset T\subset D}\zeta^{N(T)}\exp\left[-\beta U(T;D)\right] \Big/ \sum_{T\subset D}\zeta^{N(T)}\exp\left[-\beta U(T;D)\right], \quad \text{if } Y \subset D. \tag{11}$$

As an example of the relation between the thermodynamic functions of the spin system and the lattice gas, we give the expression of f in terms of Π (see Ref. 3):

$$\beta f(\beta,\zeta,p) = -p\log\zeta - \tfrac{1}{8}p^2\beta\sum_{y\neq 0}\Phi(y) - \beta\Pi. \tag{12}$$

So we see that if Π extends to analytic function without singularities in ζ at $\zeta = 0$, then f has a logarithmic singularity at $\zeta = 0$; for the correlation functions this phenomenon does not occur and in fact it turns out that the spin correlation functions have the same analyticity properties in β, ζ, and p as the gas correlation functions.

We shall first prove that the functions $\rho(Y)$ defined through (10) and (11) are analytic functions in a region Q of the form described in Theorem 1; then we shall obtain the same results for $\beta\Pi$ by showing the validity (for $\beta > 0$, $\zeta > 0$, $0 \leq p \leq 1$) of the following thermodynamic relation between the pressure Π and the density (one-point correlation function) $\rho = \rho(Y)$, $N(Y) = 1$:

$$\frac{\partial \beta \Pi}{\partial \log \zeta} = \rho. \tag{13}$$

By translating these results into the spin-system language [using (12) and the remark following it], Theorem 1 will follow.

3. THE CORRELATION FUNCTIONS OF THE RANDOM LATTICE GAS

In this section we derive some explicit expressions [see (22) below] for the functions $\rho_D(Y)$ defined in (11), which will be used in the next section to prove the analyticity properties of $\rho(Y)$ and $\beta\Pi$.

It has been shown in Refs. 1 and 4 that the functions $\rho_D(Y)$, regarded as elements of the space E of the functions defined over the *nonempty* configurations Y, satisfy the following equations:

$$\rho_D = \chi_D \alpha_D + \chi_D \mathfrak{K}_D \rho_D, \tag{14}$$

where α_D is an element of E and χ_D and $\mathfrak{K}_D$ are operators on E defined by

$$\alpha_D(Y) = 0, \quad \text{if} \quad N(Y) > 1,$$

$$\alpha_D(Y) = (\zeta \exp \beta A_y(D))(1 + \zeta \exp \beta A_y(D))^{-1}, \tag{15}$$

$$(\chi_D \psi)(Y) = 0, \quad \text{if} \quad Y \not\subset D,$$

$$(\chi_D \psi)(Y) = \psi(Y), \quad \text{if } Y \subset D, \text{ for all } \psi \in E, \tag{16}$$

$$\begin{aligned}(\mathfrak{K}_D \psi)(Y) &= [\zeta \exp \beta A_y(D) - \beta U_y(Y)] \\ &\quad \times [1 + \zeta \exp \beta A_y(D) - \beta U_y(D)]^{-1} \\ &\quad \times \Bigg[\delta_{N(Y)>1} \psi(Y^{(1)}) \\ &\quad + \sum_{\substack{T \cap Y = \varnothing, \\ T \neq \varnothing}} K_y(T)(\psi(Y^{(1)} \cup T) - \psi(Y \cup T)) \Bigg],\end{aligned} \tag{17}$$

where y is a point in Y chosen with a well-defined criterion (for instance, the first point in lexicographic order), $Y^{(1)}$ denotes what is left of the set Y when the point y is substracted, $\delta_{N(Y)>1}$ is defined to be one if $N(Y) > 1$, and zero if $N(Y) = 1$. Finally, we have

$$U_y(Y) = \sum_{y' \in Y, y' \neq y} \Phi(y - y'), \tag{18}$$

$$K_y(T) = \prod_{t \in T} (\{\exp [-\beta \Phi(t - y)]\} - 1). \tag{19}$$

The following remark is important: If we call $(\chi_D \mathfrak{K}_D)(Y/Y')$ the matrix elements of the operator $\chi_D \mathfrak{K}_D$, we see from (17) that $|(\chi_D \mathfrak{K}_D)(Y/Y')|$ is majorized by the matrix elements $K(Y/Y')$ of the operator K defined as

$$\begin{aligned}&(K\psi)(Y) \\ &= \left(\max_{\lambda} |(\zeta e^{\lambda\beta}(1 + \zeta e^{\lambda\beta})^{-1}| \right) \\ &\quad \times \Bigg(\delta_{N(Y)>1} \psi(Y^{(1)}) \\ &\quad + \sum_{\substack{T \cap Y = \varnothing, \\ T \neq \varnothing}} |K_y(T)| \, [\psi(Y^{(1)} \cup T) + \psi(Y \cup T)],\end{aligned} \tag{20}$$

where the max has to be taken as λ varies between $-\frac{3}{2}A$ and $\frac{3}{2}A$, with $A = \sum_{y \neq 0} |\Phi(y)|$.

Now if b is any positive number and if we call

$$I(\beta) = \sum_{y \neq 0} |\{\exp [-\beta \Phi(y)]\} - 1|$$

and

$$m = \max |\zeta e^{\lambda\beta}(1 + \zeta e^{\lambda\beta})^{-1}|$$

when $-\frac{3}{2}A \leq \lambda \leq \frac{3}{2}A$, we have the following estimate of the "magnitude" of $\chi_D \mathfrak{K}_D$:

$$\begin{aligned}&\sup_{Y \neq \varnothing} b^{-N(Y)} \sum_{Y'} K(Y/Y') b^{N(Y')} \\ &\leq m \sup_{Y \neq \varnothing} b^{-N(Y)} \Bigg[\delta_{N(Y)>1} b^{N(Y)-1} \\ &\quad + \sum_{\substack{T \cap Y = \varnothing, \\ T \neq \varnothing}} |K_y(T)| \, (b^{N(Y)-1+N(T)} + b^{N(Y)+N(T)}) \Bigg] \\ &\leq m\{b^{-1} \exp bI(\beta) + [\exp bI(\beta) - 1]\} \\ &= k(b; \beta, \zeta),\end{aligned} \tag{21}$$

where the quantity $k(b; \beta, \zeta)$ is implicitly defined by the last step of (21).

We can make the following use of inequality (21): If there exists $b > 0$ such that $k(b; \beta, \zeta) < 1$, then the correlation functions $\rho_D(Y)$ will be given by the iterative solution of (14):

$$\rho_D(Y) = \sum_{n=0}^{\infty} [(\chi_D \mathfrak{K}_D)^n (\chi_D \alpha_D)](Y). \tag{22}$$

[4] G. Gallavotti and S. Miracle-Sole, Commun. Math. Phys. 7, 274 (1968).

This solution exists [and is the unique solution of (14) if, as it is always the case in this paper, D is a finite set] since the series in (22) is absolutely convergent because

$$b^{-N(Y)}\sum_{n=0}^{\infty}|((\chi_D\mathfrak{K}_D)^n(\chi_D\alpha_D))(Y)|$$
$$\leq\sum_{n=0}^{\infty}\sum_{Y_1,\cdots,Y_n} b^{-N(Y)}K(Y/Y_1)b^{N(Y_1)}b^{-N(Y_1)}$$
$$\times K(Y_1/Y_2)\cdots b^{-N(Y_{n-1})}$$
$$\times K(Y_{n-1}/Y_n)b^{N(Y_n)}b^{-N(Y_n)}|(\chi_D\alpha_D)(Y_n)|$$
$$\times\left(\max_Y b^{-N(Y)}|(\chi_D\alpha_D)(Y)|\right)\cdot\sum_{n=0}k(b;\beta,\zeta)^n$$
$$= b^{-1}(1-k(b;\beta,\zeta))^{-1} < +\infty, \tag{23}$$

where we have used (21) and (15), and $|\chi_D\mathfrak{K}_D(Y/Y')|\leq K(Y/Y')$.

The correlation functions defined in (10) will then be given by

$$\rho(Y)=\lim_{K\to\infty}\sum_{n=0}^{\infty}\sum_{D\subset K}P_K(D)((\chi_D\mathfrak{K}_D)^n(\chi_D\alpha_D))(Y). \tag{24}$$

4. PROOF OF ANALYTICITY

The nth term in (24) can be written as

$$\rho_K^{(n)}(Y)=\sum_{Y_1,\cdots,Y_n}\sum_{D\subset K}(\chi_D\mathfrak{K}_D)(Y/Y_1)\cdots(\chi_D\mathfrak{K}_D)$$
$$\times(Y_{n-1}/Y_n)(\chi_D\alpha_D)(Y_n)P_K(D), \tag{25}$$

and we observe that in the sum (25) only the terms with $Y, Y_1, \cdots, Y_n$ such that $Y\cup Y_1\cup\cdots\cup Y_n\subset D$ are different from zero because of the operators χ_D.

Another crucial remark is that the quantity

$$(\chi_D\mathfrak{K}_D)(Y/Y_1)\cdots(\chi_D\mathfrak{K}_D)(Y_{n-1}/Y_n)(\chi_D\alpha_D)(Y_n) \tag{26}$$

does not depend too much on D, provided $Y\cup Y_1\cup\cdots\cup Y_n\subset D$; more precisely, denote y_i and $S_r(y_i)$, respectively, the privileged points in Y_i as they have been defined after (17), and a sphere centered at y_i with radius r equal to the range of the interaction potential (which we are supposing finite); then it is immediately seen from the definition (17) of $\mathfrak{K}_D$ that (26) depends on D only through the part of D defined as

$$D\cap\left[Y\cup Y_1\cup\cdots\cup Y_n\cup\bigcup_{i=0}^{n}S_r(y_i)\right]$$

[where y_0 denotes the privileged point of Y]. This last property is a consequence of the fact that the matrix elements of $\mathfrak{K}_D$ depend on D only through $A_{Y_i}(D)$ [see (17)].

Using these remarks, the summation over D in (25) can be partially performed because we can decompose any D which gives a nonvanishing contribution to (25) in two parts D_1 and D_2, with D_1 such that

$$(Y\cup Y_1\cup\cdots\cup Y_n)$$
$$\subset D_1\subset\left(Y\cup Y_1\cup\cdots\cup Y_n\cup\bigcup_{i=0}^{n}S_r(y_i)\right),$$

and D_2 is contained in the complement of

$$\left(Y\cup Y_1\cup\cdots\cup Y_n\cup\bigcup_{i=0}^{n}S_r(y_i)\right);$$

the summation over D_2 can be immediately done using the relation

$$\sum_{D_2\subset K_2}P_{K_1\cup K_2}(D_1\cup D_2)=P_{K_1}(D_1)$$
$$\text{for all}\quad K_1, K_2, D_1\subset K_1,$$

and the fact that the summand in (25) depends on D only through the above defined D_1. Applying the last formula to (25) and taking into account the above remarks, we obtain that 25 can be written as

$$\rho_K^{(n)}(Y)$$
$$=\sum_{Y_1,\cdots,Y_N}\sum_{\substack{D\supset Y\cup\cdots\cup Y_N\\ D\subset K\cap R}}(\chi_D\mathfrak{K}_D)(Y/Y_1)\cdots(\chi_D\mathfrak{K}_D)(Y_n)$$
$$\times P_{R\cap K}(D), \tag{27}$$

where R abbreviates

$$Y\cup Y_1\cup\cdots\cup Y_n\cup\bigcup_{i=0}^{n}S_r(y_i).$$

We remark also that (27) has the same analyticity properties as the matrix elements $(\chi_D\mathfrak{K}_D)(Y/Y')$ and the functions $(\chi_D\alpha_D)(Y)$ and $P_K(D)$, since it is a finite sum of products of such matrix elements and functions; it is also clear from the definitions of $P_K(D)$, $\chi_D\mathfrak{K}_D$, and $\chi_D\alpha_D$ that $P_K(D)$ is a polynomial in the impurity probability p and that the functions $(\chi_D\alpha_D)(Y)$ and $(\chi_D\mathfrak{K}_D)(Y/Y')$ are analytic in β and ζ, provided

$$\max|\zeta e^{\lambda\beta}(1+\zeta e^{\lambda\beta})^{-1}| < +\infty,$$

where the maximum has to be taken for $(-\frac{3}{2})A\leq\lambda\leq(\frac{3}{2})A$.

Now, if N_0 denotes the number of points in $S_r(y_i)$, and if we take into account (15) and the inequality

$$\sum_{\substack{D\supset Y\cup Y_1\cup\cdots\cup Y_n\\ D\subset K\cap R}}|P_{K\cap R}(D)|$$
$$\leq(|p|+|1-p|)^{N_0 n}|p|^{N(Y\cup Y_1\cup\cdots\cup Y_n)},$$

it follows that (27) can be majorized by

$$|\rho_K^{(n)}(Y)| \leq \sum_{\substack{Y_1,\cdots,Y_n \\ N(Y_n)=1}} |p|^{N(Y\cup Y_1\cup\cdots\cup Y_n)} K(Y/Y_1)\cdots K(Y_{n-1}|Y_n) \times \left(\max_{D,Y}|\chi_D\alpha_D(Y)|\right)\cdot(|p|+|1-p|)^{N_0 n}, \tag{28}$$

where the kernel $K(\cdot/\cdot)$ has been introduced and studied in (20) and (21), and the condition $N(Y_n)=1$ is due to the fact that $\alpha_D(Y_n)$ vanishes unless $N(Y_n)=1$ [see definition (15)].

Inequality (28), together with (21), allows us to show, through a chain of inequalities of the type (23), that the sum of the moduli of the addends in (27) is majorized by

$$\bar\rho^{(n)}(Y) = \left(\max_{D,Y}|\alpha_D(Y)|\right)\cdot[(|p|+|1-p|)^{N_0 n} \times k(b;\beta,\zeta)]^n \cdot b^{N(Y)}, \tag{29}$$

provided $|p|\leq 1$; in (29) b is any positive number.

Now if $(|p|+|1-p|)^{N_0}\cdot k(b;\beta,\zeta)<1$, it is clear from (28) and (29) that (27) has a limit as $K\to\infty$, which is equal to

$$\rho^{(n)}(Y) = \sum_{Y_1,\cdots,Y_n}\ \sum_{\substack{D\supset Y\cup Y_1\cup\cdots\cup Y_n \\ D\subset R}} (\chi_D\mathfrak{K}_D)(Y/Y_1)\cdots (\chi_D\mathfrak{K}_D)(Y_{n-1}/Y_n)\cdot P_R(D), \tag{30}$$

where R abbreviates, as above,

$$Y\cup Y_1\cup\cdots\cup Y_n\cup\bigcup_{i=0}^{n} S_r(y_i);$$

furthermore, $|\rho_K^{(n)}(Y)|\leq\bar\rho^{(n)}(Y)$. Therefore it follows that the limit (24) exists [cf. (29)] and is equal to

$$\rho(Y)=\sum_{n=0}^{\infty}\rho^{(n)}(Y). \tag{31}$$

It is also clear that, as a consequence of the uniform bounds on $|\rho_K^{(n)}(Y)|$ obtained in (27), (28), and (29), the analyticity properties of (27) become analyticity properties of $\rho(Y)$ in the region where

$$(|p|+|1-p|)^{N_0}\cdot\inf_b k(b;\beta,\zeta)<1,\quad |p|\leq 1, \tag{32}$$

and it is straightforward to check that the analyticity region so defined contains a region Q of the form described in Theorem 1 (for details see Appendix).

The discussion above proves the part of Theorem 1 concerning the correlation functions; to prove the promised results on the free energy we need to prove that the function $\beta\Pi$, defined for $0\leq p\leq 1$, $\zeta>0$, $\beta>0$, by the limit (8), can be extended to an analytic function in the region Q where the correlation functions have been shown to be analytic. In fact, we shall obtain the analyticity of the pressure by proving (13) through the following lemma [although it could be easily deduced from (22) and the inequalities derived in (23), we omit the proof since it is essentially identical to the analogous proof first given in Ref. 5 (see also Ref. 6)]:

Lemma: Suppose that $\inf_b k(b;\beta,\zeta)<1$, and denote $S_d(y)$ as a sphere of radius d around y. Given a configuration Y and $y\in Y$, we have

$$|\rho_D(Y+a)-\rho_{D\cap S_d(y+a)}(Y+a)|\xrightarrow[d\to\infty]{}0 \tag{33}$$

uniformly in D and a. Furthermore, the functions $\rho_D(Y+a)$ are uniformly bounded in a and D (at fixed Y).

In other terms this lemma says that the correlation functions $\rho_D(Y)$ and their translations in space $\rho_D(Y+a)$ do not depend too much on the points of D located outside a sphere with large radius and centered at a point of $Y+a$; furthermore, the "small dependence" on the far points is uniform in D and a.

Now suppose β, ζ, and p are real and positive ($p\leq 1$), and consider the identity

$$\frac{\partial\beta\Pi_K}{\partial\log\zeta}=N(K)^{-1}\sum_{a\in K}\sum_{D\subset K}P_K(D)\rho_K(a), \tag{34}$$

which follows from the definition (11) of ρ_D and by differentiation of the definition (9) of Π_K. Using the convexity of $\beta\Pi$ as a function of $\log\zeta$ (the existence of the limit defining Π has been proved in Ref. 3) and taking into account (33) and (10), we find

$$\begin{aligned}\frac{\partial\beta\Pi}{\partial\log\zeta}&=\lim_{K\to\infty}\frac{\partial\beta\Pi_K}{\partial\log\zeta}\\&=\lim_{K\to\infty}N(K)^{-1}\sum_{a\in K}\sum_{D\subset K}P_K(D)\rho_D(a)\\&=\lim_{K\to\infty}\sum_{D\subset K}P_K(D)\rho_D(a)=\rho(a)=\rho,\end{aligned} \tag{35}$$

which proves that $\beta\Pi$ can be extended into an analytic function in the region Q where ρ is analytic, and this completes the proof of Theorem 1. We remark that the fact that to obtain Π one has to integrate over ζ the function ρ/ζ, does not introduce any singularity at $\zeta=0$ because it can be easily verified [from (22) and (31)] that ρ is proportional to ζ at small ζ (i.e.,

[5] D. Ruelle, Ann. Phys. (N.Y.) **25**, 109 (1963).
[6] See Ref. 4, Proposition 2, and proof of Theorem 1.

$\rho = \zeta + c\zeta^2 + \cdots$); however, one cannot guarantee that the function $\beta\Pi$ obtained integrating ρ is single valued in the nonsimply connected domain Q.[7]

5. CONCLUDING REMARKS

The results obtained so far are not complete in the sense that essential use is made of the fact that the potential has finite range and also because, if we suppose $p > 0$, the analyticity region in the (β, ζ) space implicitly obtained in the preceding proofs does not become larger and larger as $p \to 0$, as it can be expected from the fact that the critical temperature tends to zero when $p \to 0$.[3]

On the contrary, one realizes that the analyticity region in (β, ζ) remains unchanged as p varies, provided it stays positive.

In particular, one finds that, if $0 \leq p \leq 1$, the free energy and the correlation functions are analytic functions in (β, ζ) if $0 \leq \beta \leq \beta_c$, where β_c is defined by the equation

$$\inf_b k(b; \beta_c, \zeta) < 1 \quad \text{for all} \quad \zeta > 0,$$

which gives the numerical lower bound β_c on the critical inverse temperature: $\beta_c \sum_{y \neq 0} |\Phi(y)| = 0.15$. This last result can probably be improved, perhaps by replacing 0.15 by 0.40 and by using more refined techniques which take into account the spin reversal symmetry.[1]

It would be interesting to find a further enlargement of the analyticity regions to get bounds on the critical temperature which tend to zero as $p \to 0$. (Doubts on this point have, however, been aroused in Ref. 8.)

Note added in proof: The impossibility that the critical temperature, defined as the first temperature where a nonanalyticity appears, tends to zero as $p \to 0$, has been recently proved by Griffiths[9] (in the case of the Ising model).

APPENDIX

The region defined by (32) is the set of p, β, ζ, such that

$$(|p| + |1 - p|)^{N_0} \cdot \inf_b m \cdot [b^{-1} e^{bI(\beta)} + (e^{bI(\beta)} - 1)] < 1, \quad |p| < 1, \quad \text{(A1)}$$

where we have used the definition (22) of $k(b; \beta, \zeta)$.

To study the region defined by (A1) put $b = 3^{N_0+1}$ (say) and suppose $|\beta| < \beta_0$ with β_0 so small that

$$(|p| + |1 - p|)^{N_0}(b^{-1} e^{bI(\beta)} + (1 + e^{bI(\beta)}) < \tfrac{1}{2}. \quad \text{(A2)}$$

Then clearly the analyticity region will contain the circle $|p| < 1$ and the (β, ζ) space region defined by

$$m = \max |\zeta e^{\lambda\beta}(1 + \zeta e^{\lambda\beta})| < 2, \quad -\tfrac{3}{2}A \leq \lambda \leq \tfrac{3}{2}A, \quad |\beta| < \beta_0. \quad \text{(A3)}$$

The set $|\zeta e^{\lambda\beta}(1 + \zeta e^{\lambda\beta})| < 2$ is, for $\lambda\beta$ real, the exterior of a circle centered at the point on the negative real axis with coordinate $\xi = (-\frac{4}{3})e^{-\lambda\beta}$, and with radius $(\frac{2}{3})e^{-\lambda\beta}$, and so, for β real, the set $m < 2$ contains in the ζ space the complement of a circle centered at the middle of the interval $((-\frac{6}{3})e^{(\frac{3}{2})\beta A}, (-\frac{2}{3})e^{-(\frac{3}{2})\beta A})$, passing through its extreme points.

By continuity arguments, one realizes that the above discussion implies that the analyticity region in the (p, β, ζ) space contains a domain of the form Q of Theorem 1. However, it is also clear that the analyticity region is in fact quite larger.

ACKNOWLEDGMENTS

I am indebted to J. L. Lebowitz and R. B. Griffiths for suggesting the problem and for useful comments.

[7] D. Ruelle, *Statistical Mechanics* (W. A. Benjamin, Inc., New York, 1969), Chap. 5; 2.1.

[8] M. A. Mikulinskii, Zh. Eksp. Teor. Fiz. **10**, [Sov. Phys.—JETP **26**, 637, 1968].

[9] R. B. Griffiths, Phys. Rev. Letters **23**, 17 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1969

Absence of Phase Transitions in Hard-Core One-Dimensional Systems with Long-Range Interactions

G. GALLAVOTTI
Rockefeller University, New York City, New York 10021
AND
S. MIRACLE-SOLE
Institut des Hautes Etudes Scientifiques, 91 *Bures-sur-Yvette, France*

(Received 22 November 1968)

We prove the impossibility of phase transitions for a class of infinite-range potentials extending recent analogous results. We prove also a cluster property for the equilibrium state $\bar{\gamma}$ and apply some collateral results to describe, in the case of finite-range interactions, the state $\bar{\gamma}$ in terms of a family of density distributions, and to verify a general variational property of $\bar{\gamma}$.

1. INTRODUCTION

It has been conjectured that even in one-dimensional systems a phase transition can occur if the range of the potential in infinite (in the sense that the first moment of the potential diverges) and the potential is attractive.[1] [*Note added in proof:* A proof of this statement has been given by F. J. Dyson in Commun. Math. Phys. **12,** 91 (1969).]

On the contrary, if the potential has infinite range but the first moment is convergent, it is commonly accepted that, at least if the potential is sufficiently regular, no phase transitions occur in one dimension[2]; in this case one can also conjecture that the pressure has some analyticity properties with respect to the chemical potential and the temperature, and also that the correlation functions have some cluster property.

Most of these conjectures have been proved to be true by Ruelle in the case of lattice gases[3]; using his technique, it has been possible to obtain similar results in the continuous hard-core case.[4] In fact, the possibility of phase transitions has been excluded for the class of continuous bounded-pair potentials, bounded in absolute value by a decreasing function $\varphi(x)$ such that

$$\int_a^{+\infty} \varphi(x) x \, dx < +\infty, \tag{1}$$

and this extends considerably the well-known results of Van Hove.[2]

In this paper we present an extension of these results to more general unbounded and not necessarily continuous potentials involving two or more bodies; we also prove a cluster property for the correlation functions and, using collateral results, we give a description of the equilibrium state in the case of finite-range forces in terms of a family of density distribution, and we verify an extremum property of the equilibrium state.

2. DESCRIPTION OF THE SYSTEM; RESULTS

Let us consider a one-dimensional system of hard rods. A configuration will be represented by the set X of the points of the real axis R occupied by the centers of the rods. If $a > 0$ is the length of the rods, then X is restricted to verify the condition $|x - x'| \geq a$ if $x, x' \in X$, $x \neq x'$.

We say that a sequence $\{X_\alpha\}$ of configurations tends as $\alpha \to \infty$ to the limit configuration X_0 if, for every bounded interval (a, b) such that $a \notin X_0$, $b \notin X_0$, the configuration $X_\alpha \cap (a, b)$ tends point by point to $X_0 \cap (a, b)$.

With this definition of convergence the set K of all the configurations (finite or not) becomes a compact space. It will also be useful to consider the compact subsets K_+, $K_- \subset K$ built up, respectively, with the configurations of K contained in $R_- = (-\infty, 0]$ and in $R_+ = [0, +\infty)$. The symbols $C(K)$, $C(K_+)$, and $C(K_-)$ will denote the continuous functions, respectively, over K, K_+, and K_-; $C_{[0,b)}$ $(b > 0)$ will denote the set of continuous functions in $C(K_+)$, with the property that $f(Y) = f(Y \cap [0, b))$ for all $Y \in K_+$. It can be shown that the set $\bigcup_{b>0} C_{[0,b)}$ is dense in $C(K_+)$ [in the sup norm on $C(K_+)$].

Suppose the rods interact through symmetric translationally invariant many-body potentials $\Phi^{(k)}(x_1, \cdots, x_k)$ and consider these as a function Φ on the configurations $X \in K$ defined as $\Phi(X) = \Phi^{(k)}(x_1, \cdots, x_k)$ if $X = \{x_1, \cdots, x_k\}$ $0 < k < \infty$ and $\Phi(X) = 0$ if $k = 0, +\infty$; the one-particle potential $\Phi^{(1)}$ has to be interpreted as minus the chemical potential, so the

[1] M. E. Fisher, Physics 3, 255(1967); M. Kac, "Mathematical Mechanism of Phase Transitions" (to be published).
[2] L. Van Hove, Physica **16,** 137 (1950).
[3] D. Ruelle, Commun. Math. Phys. **9,** 267 (1968).
[4] G. Gallavotti, S. Miracle-Sole, and D. Ruelle, Phys. Letters **26A,** 350 (1968).

energy of a configuration is

$$U(X) = \sum_{S \subset X} \Phi(S). \tag{2}$$

We want to allow the potential Φ to be of the form

$$\Phi = \tilde{\Phi}_0 + \tilde{\Phi}, \tag{3}$$

where $\tilde{\Phi}_0$ is a nonnegative measurable finite-range pair potential (not necessarily bounded), while $\tilde{\Phi}$ is supposed to belong to the real space $\mathcal{B}$ defined as the closure in the norm (4) below of the set $\mathcal{B}_0$ of the real finite-range bounded measurable potentials with the "continuity" property that if $0 \leq h \leq k-1$ and $\bar{x}_1, \cdots, \bar{x}_h$ are fixed, then the function $\Phi^{(k)}(x_1, \cdots, x_h, x_{h+1}, \cdots, x_k)$ is continuous in the variables $x_1, \cdots, x_h$ at the point $\bar{x}_1, \cdots, \bar{x}_h$ for almost all $x_{h+1}, \cdots, x_k$ (which is a further restriction on $\tilde{\Phi}^{(k)}$ only for $k \geq 3$). The norm with respect to which the closure has to be taken is

$$\|\tilde{\Phi}\| = \sup_{X \in P} \sum_{\substack{T \subset X \\ T \cap R_+ \neq \emptyset \neq T \cap R_-}} |\Phi(T)|. \tag{4}$$

One can convince himself that the condition that $\tilde{\Phi}$ should be in some closure of finite-range potentials can be interpreted as a condition of decrease at large distances and the fact that this closure has to be taken with respect to the norm (4) is a condition similar to that of having a finite first moment. In fact, one can see that the requirement that $\Phi \in \mathcal{B}$ is in general less restrictive than the requirement that $\tilde{\Phi}$ verifies simultaneously a condition of type (1) and a decreasing condition in the sense usually found in the literature.[5] The "continuity" condition imposed on the potentials is probably unnecessary and simple measurability should be sufficient. In view of technical difficulties we shall not deal with this point.

The thermodynamics of the system is described by the partition function (grand canonical ensemble):

$$Z_{(b_1,b_2)}(\Phi) = \int_{X \subset (b_1,b_2)} e^{-U(X)}\, dX, \tag{5}$$

where

$$\int \cdot\, dX = \sum_{n \geq 0} \int \cdot \frac{dx_1 \cdots dx_n}{n!}$$

and the inverse temperature factor β has been included in the interaction energy.

The main result of this paper is the theorem below, which will be proved in Sec. 4, using the preliminary lemmas of Sec. 3, following the scheme used in Ref. 3 to prove the analogous results in the lattice case.

[5] M. E. Fisher, Arch. Ratl. Mech. Anal. **17**, 377 (1964).

Theorem: Let the interaction potential Φ be given by $\Phi = \tilde{\Phi}_0 + \tilde{\Phi}$, where $\tilde{\Phi}_0$ is a nonnegative finite-range pair potential and $\tilde{\Phi} \in \mathcal{B}$. Then:

(i) The function

$$P(\Phi) = \lim_{|b_2 - b_1| \to \infty} |b_2 - b_1|^{-1} \lg Z_{(b_1,b_2)}(\Phi)$$

has the property that, given $\tilde{\Phi}_1, \cdots, \tilde{\Phi}_n \in \mathcal{B}$, then $P(\sum_{i=0}^{n} \lambda_i \tilde{\Phi}_i)$ is continuously differentiable in $\lambda_0 \in [0, \infty)$ and $\lambda_1, \cdots, \lambda_n \in (-\infty, +\infty)$.

(ii) If $\bar{\gamma}$ denotes the probability measure (on the set K of all the configurations) which represents the equilibrium state relative to the potential Φ, and if $A_1, A_2, \cdots, A_n \in C(K)$ and τ_x denotes the translation operator on $C(K)$ [defined as $(\tau_x A)(Y) = A(Y + x)$ for $x \in (-\infty, +\infty)$, $A \in C(K)$], then $\bar{\gamma}$ is translationally invariant and

$$\lim_{\substack{|x_i - x_j| \to \infty \\ i \neq j}} \bar{\gamma}(\tau_{x_1} A_i \cdot \tau_{x_2} A_2 \cdot \cdots \cdot \tau_{x_n} A_n) = \bar{\gamma}(A) \cdots \bar{\gamma}(A_n). \tag{6}$$

This theorem excludes the possibility not only of first-order phase transitions, but also those of higher order in the sense that (i) and (ii) imply not only that the density is continuous as a function of the temperature and the chemical potential, but also that the equilibrium state is unique (given the interaction), and that all the correlation functions are continuous with respect to the interaction potentials and do not exhibit long-range order.

3. STUDY OF A SEMI-INFINITE SYSTEM

In this section the symbols m, n, r, and k will denote nonnegative integers and the symbol τ_x will mean the translation operator on the set of configurations $X \in K$ defined by $\tau_x X = X + x$.

Given Φ as discussed in the preceding section, we define a family of operators which map $C(K_+)$ into itself linearly and continuously with respect to the sup norm with which we suppose $C(K_+)$ to be equipped. These operators are defined for all $x \geq 0$ as

$$(\mathfrak{L}_x f)(Y) = \int_{X \subset [0,x)} e^{-U(X \mid \tau_x Y)} f(X \cup \tau_x Y)\, dX, \quad \text{for } Y \in K_+, \tag{7}$$

where $U(R \mid W)$ is defined for R and W in K as

$$\begin{aligned} U(R \mid W) &= \sum_{\substack{\emptyset \neq S \subset R \\ T \subset W}} \Phi(S \cup T) \\ &= \sum_{\substack{\emptyset \neq S \subset R \\ \emptyset \neq T \subset W}} \Phi(S \cup T) + \sum_{\emptyset \neq S \subset R} \Phi(S) \\ &= I(R \mid W) + U(R), \end{aligned} \tag{8}$$

and where $\Phi(S \cup T)$ has to be taken $+\infty$ if $S \cup T \notin K$. One verifies, as a consequence of the assumed continuity properties of Φ, that $\mathfrak{L}_x$ maps $C(K_+)$ into itself and $\|\mathfrak{L}_x\| \leq e^{\|\tilde{\Phi}\| Z_{[0,x)}}$.

The adjoint operators to $\mathfrak{L}_x$ are operators on the space $C(K_+)^* = \{\text{space of bounded measures on } K_+\} = \{\text{dual of } C(K_+)\}$:

$$(\mathfrak{L}_x^* \mu)(dX\, d\tau_x Y) = e^{-U(X \mid \tau_x Y)}\, dX \mu(dY), \quad X \subset [0, x), \quad Y \in K_+. \tag{9}$$

A simple calculation based on (7) shows that

$$\mathfrak{L}_x \mathfrak{L}_y = \mathfrak{L}_{x+y}, \quad \mathfrak{L}_x^* \mathfrak{L}_y^* = \mathfrak{L}_{x+y}^*, \quad x, y \geq 0. \tag{10}$$

Let us call $\bar{\gamma}_+$ the probability measure on K_+ which describes the equilibrium state of a semi-infinite system (contained in R_+); then, formally, one expects that $\mathfrak{L}_x^* \bar{\gamma}_+ = e^{xP} \bar{\gamma}_+$, where P is the thermodynamic pressure and also (if there are no phase transitions) that $\bar{\gamma}_+$ is unique. Two steps towards the proof of such a property are lemmas 1 and 2:

Lemma 1: There exists a probability measure $\nu \in C(K_+)^*$ and $\lambda \geq 1$ such that

$$\mathfrak{L}_a^* \nu = \lambda \nu, \tag{11}$$

where a denotes the hard rod length.

In fact, let E be the set of probability measures on K_+; then, since $\mathfrak{L}_a^* \mu(1) \geq 1$ if $\mu \in E$, the mapping of E into itself defined by $\mu \to [\mathfrak{L}_a^* \mu(1)]^{-1}\, \mathfrak{L}_a^* \mu$ is unambiguously defined and weakly continuous. The set E being convex and weakly compact, the Schauder–Tychonoff theorem[6] applies to give a fixed point $\nu \in E$ which verifies (11).

Lemma 2: There exists $C_0 > 0$ such that, for all integers $n \geq 0$,

$$C_0 e^{-\|\tilde{\Phi}\|} \leq \lambda^n Z_{[0,na)}^{-1} \leq e^{\|\tilde{\Phi}\|}; \tag{12}$$

hence it follows that $\lambda = e^{aP}$, where P is the thermodynamic pressure.

We have, if 1 denotes the function of $C(K_+)$ identically equal to unity,

$$(\mathfrak{L}_{na} 1)(Y) \leq \int_{X \subset [0,na)} e^{-U(X)} e^{\|\tilde{\Phi}\|}\, dX = e^{\|\tilde{\Phi}\|} Z_{[0,na)}. \tag{13}$$

We have also, if ra is greater than the range of $\tilde{\Phi}_0$,

$$(\mathfrak{L}_{na} 1)(Y) \geq \int_{X \subset [0,na)} e^{-U(X)} e^{-\|\tilde{\Phi}\|}\, dX \bar{f}_r(Y) = \bar{f}_r(Y) e^{-\|\tilde{\Phi}\|} Z_{[0,na)}, \tag{14}$$

where the function $\bar{f}_r(Y)$ is defined to be 1 if $Y \subset [2ra, +\infty)$ and zero otherwise. This function is a Borel function, so we can integrate (13) and (14) with respect to ν and obtain (12) with $C_0 = \nu(\bar{f}_r)$. To prove that $C_0 > 0$ let m be an integer, $m > 2r$; then

$$(\lambda^{-m} \mathfrak{L}_{ma} \bar{f}_r)(Y) = \lambda^{-m} \int_{[0,ma)} e^{-U(X \mid \tau_{ma} Y)} \bar{f}_r(X)\, dX \geq \lambda^{-m} \int_{[ra,2ra)} e^{-U(X)}\, dX e^{-\|\tilde{\Phi}\|}. \tag{15}$$

Hence,

$$\nu(\bar{f}_r) = \nu(\lambda^{-m} \mathfrak{L}_{ma} \bar{f}_r) \geq \lambda^{-m} e^{-\|\tilde{\Phi}\|} Z_{[ra,2ra)} > 0.$$

We now want to show that ν is the unique solution of the eigenvalue problem $\mathfrak{L}_a^* \nu = e^{aP} \nu$. We observe that the "matrix elements" of $\mathfrak{L}_a^*$ are all nonnegative, so it is tempting to try to obtain unicity on the same lines of the proof of the Frobenius theorem for finite matrices.[7] The key for that theorem is the study of the adjoint eigenvalue problem, which in our case would be $\mathfrak{L}_a h = \lambda h$; this problem is solved by means of the lemmas below.

Lemma 3: If $f \in C(K_+)$ and $\nu(|f|) = 0$, then $\|f\| = 0$. Hence, if $\{f_n\}$ is a conditionally compact sequence of elements of $C(K_+)$, from the limit $\lim \nu(|f_n|) = 0$ as $n \to \infty$, we can deduce that $\lim \|f_n\| = 0$ as $n \to \infty$.

Let $f \geq 0$ and $\nu(f) = 0$. Suppose $f \not\equiv 0$; thus there exists $\bar{Y} \in K_+$ and $f(\bar{Y}) > 0$. Given $\epsilon < \frac{1}{2} f(\bar{Y})$, one can find, because of the continuity of f and the nature of the topology on K_+, a $k > 0$ and a nonempty open set $G \subset K_+$ such that

$$Y \in G \leftrightarrow Y \cap [0, ka) \in G,$$
$$Y \in G \Rightarrow |f(Y) - f(\bar{Y})| < \epsilon.$$

If χ_G is the characteristic function of G, we have

$$\nu(f) \geq \nu(f \chi_G) \geq \epsilon \nu(\chi_G). \tag{16}$$

To prove that $\nu(\chi_G) > 0$, let n be an integer greater than $(k + r)$ where r is chosen such that ra is greater than the range of $\tilde{\Phi}_0$. Then

$$\lambda^{-n} \mathfrak{L}_{na} \chi_G(Y) = \lambda^{-n} \int_{X \subset [0,na)} e^{-U(X \mid \tau_{na} Y)} \chi_G(X)\, dX \geq \lambda^{-n} e^{-\|\tilde{\Phi}\|} \int_{X \subset [0,ka)} e^{-U(X)} \chi_G(X)\, dX > 0,$$

[6] N. Dunford and J. Schwartz, *Linear Operators* (Interscience Publishers, Inc., N.Y., 1958), Vol. I, Chap. V, Sec. 10, item 5.

[7] D. Ruelle, *Statistical Mechanics. Rigorous Results* (W. A. Benjamin, Inc., N.Y., 1969), Chap. IV, Sec. 6, item 3.

so that

$$\nu(\chi_G) = \nu(\lambda^{-n}\mathcal{L}_{na}\chi_G) \geq \lambda^{-n}e^{-\|\tilde{\Phi}\|}\int_{[0,ka)} e^{-U(X)}\chi_G(X)\,dX > 0.$$

Lemma 4: If r is so chosen that ra is greater than the range of $\tilde{\Phi}_0$ and if $g \in C_{[0,ma)}$, there exists $A > 0$ such that

$$\mathcal{L}_{na}g(Y)/\mathcal{L}_{na}g(Y') \leq A, \quad \text{for all} \quad Y,\ Y' \in K_+ \text{ and } n > m + 2r. \tag{17}$$

In fact, if $n > m + 2r$, using Lemma 2, we have

$$\frac{\mathcal{L}_{na}g(Y)}{\mathcal{L}_{na}g(Y')} = \left(\int_{X\subset[0,na)} dX e^{-U(X)}g(X)e^{-I(X|\tau_{na}Y)}\right)(Y \to Y')^{-1}$$

$$= \left(\int_{X_1\subset[0,ma)} e^{-U(X_1)}g(X_1)\,dX_1 \int_{X_2\subset[ma,na)} e^{-U(X_2)}\,dX_2 \exp\left[-I(X_1 \mid X_2 \cup \tau_{na}Y)\right] \times e^{-I(X_2|\tau_{na}Y)}\right)(Y \to Y')^{-1}$$

$$\leq \sup_{X_1\subset[0,ma)} \left(\int_{X_2\subset[ma,na)} dX_2 e^{-U(X_2)} \exp\left[-I(X_1 \mid X_2 \cup \tau_{na}Y)\right] \times e^{-I(X_2|\tau_{na}Y)}\right)(Y \to Y')^{-1}$$

$$\leq \frac{e^{2\|\tilde{\Phi}\|}\int_{X_2\subset[ma,na)} dX_2 e^{-U(X_2)}}{e^{-2\|\tilde{\Phi}\|}\int_{X_2\subset[(m+r)a,(n-r)a)} dX_2 e^{-U(X_2)}} \leq e^{4\|\tilde{\Phi}\|}\frac{Z_{[0,(n-m)a)}}{Z_{[0,(n-m-2r)a)}} \leq \frac{e^{6\|\tilde{\Phi}\|}}{C_0}\lambda^{2r},$$

where $Y \to Y'$ means the same term as in the numerator with Y replaced by Y'.

Lemma 5: If $f \in C_{[0,ma)}$ and $\nu(f) = 0$, then, if r denotes an integer such that ra is greater than the range of Φ_0, we have

$$\nu(|\lambda^{-n}\mathcal{L}_{na}f|) \leq (1 - A^{-1})\nu(|f|), \quad n > m + 2r. \tag{18}$$

In fact, if f' is any positive function in $C_{[0,ma)}$, then, using the preceding lemma, we find, for $n > m + 2r$,

$$\lambda^{-n}\mathcal{L}_{na}f' \geq \inf \lambda^{-n}\mathcal{L}_{na}f' \geq A^{-1}\sup \lambda^{-n}\mathcal{L}_{na}f' \geq A^{-1}\nu(\lambda^{-n}\mathcal{L}_{na}f') = A^{-1}\nu(f'). \tag{19}$$

Now let $f \in C_{[0,ma)}$ $\nu(f) = 0$ and let f_+, f_- be, respectively, the positive and negative parts of f [i.e., $f_+ = \sup(f, 0), f_- = \sup(-f, 0)$]; then, using $\nu(f_+) = \nu(f_-)$, $\nu(|f|) = \nu(f_+) + \nu(f_-)$, and Eq. (19), we find

$$\begin{aligned}
&\nu(|\lambda^{-n}\mathcal{L}_{na}f|) \\
&= \nu(|\lambda^{-n}\mathcal{L}_{na}f_+ - \lambda^{-n}\mathcal{L}_{na}f_-|) \\
&= (\nu|\lambda^{-n}\mathcal{L}_{na}f_+ - A^{-1}\nu(f_+)) - (\lambda^{-n}\mathcal{L}_{na}f_- - A^{-1}\nu(f_-))| \\
&\leq \nu(\lambda^{-n}\mathcal{L}_{na}f_+ - A^{-1}\nu(f_+)) + \nu(\lambda^{-n}\mathcal{L}_{na}f_- - A^{-1}\nu(f_-)) \\
&= (1 - A^{-1})\nu(|f|).
\end{aligned}$$

Lemma 6: If $f \in C(K_+)$, then, given a positive integer N, there exists an integer $m(N)$ such that all the functions $\lambda^{-n}\mathcal{L}_{na}f$ with $n \geq m(N)$ can be approximated within $1/N!$ by functions $f_n \in C_{[0,m(N)a)}$; i.e., for $n > m(N)$, there exists $f_n \in C_{[0,m(N)a)}$ and

$$\|\lambda^{-n}\mathcal{L}_{na}f - f_n\| < 1/N!. \tag{20}$$

Since $\|\lambda^{-n}\mathcal{L}_{na}\| \leq e^{2\|\tilde{\Phi}\|}/C_0$ (Lemma 2), it is sufficient to prove Lemma 6 for $f \in C_{[0,ma)}$. Let $f \in C_{[0,ma)}$; then, if $n > m$,

$$\lambda^{-n}\mathcal{L}_{na}f(Y) = \lambda^{-n}\int_{[0,na)} e^{-U(X)}e^{-I(X|\tau_{na}Y)}f(X)\,dX.$$

Now denote Φ_k the potential obtained from Φ by replacing $\tilde{\Phi}$ with a potential $\tilde{\Phi}_k \in \mathcal{B}$ with range ka, and such that $\|\tilde{\Phi} - \tilde{\Phi}_k\| \to 0$ as $k \to \infty$ and $\|\tilde{\Phi}_k\| \leq \|\tilde{\Phi}\|$. If r denotes, as everywhere in the paper, an integer such that ra is greater than the range of $\tilde{\Phi}_0$, let m (N) be an integer greater than $2r$ and m and such that

$$\|\tilde{\Phi} - \tilde{\Phi}_{m(N)}\| < \frac{1}{N!}C_0 e^{-2\|\tilde{\Phi}\|}\,\|f\|^{-1}.$$

Then for $n > m(N)$ define

$$f_n(Y) = \lambda^{-n}\int_{[0,na)} e^{-U(X)}f(X)\,dX \times \exp\left[-I(X \mid \tau_{na}(Y \cap [0, m(N)a))\right].$$

We now have $f_n \in C_{[0,m(N)a)}$ and

$$\begin{aligned}
&|\lambda^{-n}\mathcal{L}_{na}f(Y) - f_n(Y)| \\
&\leq \lambda^{-n}\,\|f\|\int_{[0,na)} e^{-U(X)}\,dX \\
&\quad \times |\exp\left[-I(X \mid \tau_{na}(Y \cap [0, m(N)a))\right] - e^{-I(X|\tau_{na}Y)}| \\
&\leq \lambda^{-n}\,\|f\|\int_{[0,na)} e^{-U(X)}\,dX e^{\|\tilde{\Phi}\|}\,\|\tilde{\Phi} - \tilde{\Phi}_{m(N)}\| \\
&\leq \lambda^{-n}Z_{[0,na)}\,\|f\|\,e^{\|\tilde{\Phi}\|}\,\|\tilde{\Phi} - \tilde{\Phi}_{m(N)}\| \leq \frac{1}{N!}.
\end{aligned}$$

Lemma 7: If $f \in C(K_+)$ and $\nu(f) = 0$, then

$$\lim_{n\to\infty} \nu(|\lambda^{-n}\mathcal{L}_{na}f|) = 0. \tag{21}$$

Let $m(N)$ be the number defined in the preceding lemma and k be an arbitrary integer. Let

$$n > 2\sum_{i=1}^{k}(m(N+i)+2r).$$

One has $n - m(N+k) - 2r > m(N+k) + 2r$, so using Lemma 6, we can approximate within $1/(N+k)!$ the function $\lambda^{-(n-m(N+k)-2r)}\mathfrak{L}_{(n-m(N+k)-2r)a}f$ with a function $f_k \in C_{[0,m(N+k)a)}$; then, using Lemma 5 and the bound $\|\lambda^{-s}\mathfrak{L}_{sa}\| \leq e^{2\|\tilde{\Phi}\|}/C_0$ (consequence of Lemma 2), we have

$$\begin{aligned}
&\nu(|\lambda^{-n}\mathfrak{L}_{na}f|)\\
&\quad = \nu(|\lambda^{-(m(N+k)+2r)}\mathfrak{L}_{(m(N+k)+2r)a}\lambda^{-(n-m(N+k)-2r)}\\
&\qquad \times \mathfrak{L}_{(n-m(N+k)-2r)a}f|)\\
&\qquad \leq \frac{e^{2\|\tilde{\Phi}\|}}{C_0}\frac{1}{(N+k)!} + (1-A^{-1})\nu(|f_k|)\\
&\qquad \leq \frac{1}{(N+k)!}\left(\frac{e^{2\|\tilde{\Phi}\|}}{C_0} + (1-A^{-1})\right)\\
&\qquad + (1-A^{-1})\nu(|\lambda^{-(n-m(N+k)-2r)}\mathfrak{L}_{(n-m(N+k)-2r)a}f|);
\end{aligned}$$

and since $n - m(N+k) - 2r > 2\sum_{i=1}^{k-1}(m(N+i) + 2k)$, we can iterate the above procedure and find

$$\begin{aligned}
\nu(|\lambda^{-n}\mathfrak{L}_{na}f|) \leq 1\left(\frac{e^{2\|\tilde{\Phi}\|}}{C_0} + (1-A^{-1})\right)\sum_{i=0}^{k-1}\frac{(1-A^{-1})^i}{(N+k-i)!}\\
+ (1-A^{-1})^k\frac{e^{2\|\tilde{\Phi}\|}}{C_0}\|f\|.
\end{aligned}$$

This proves the lemma since N and k were arbitrary.

Lemma 8: If $f \in C(K_+)$, the following limit holds uniformly in n:

$$\lim_{Y' \to Y}\lambda^{-n}\mathfrak{L}_{na}f(Y') = \lambda^{-n}\mathfrak{L}_{na}f(Y). \tag{22}$$

Since $\|\lambda^{-n}\mathfrak{L}_{na}\| \leq e^{2\|\tilde{\Phi}\|}C_0^{-1}$, it will be sufficient to prove this lemma in the case $f \in C_{[0,ma)}$. Suppose $f \in C_{[0,ma)}$ and $n > m$; let Φ_k, $\tilde{\Phi}_k$ be the potentials introduced in the proof of Lemma 6. For fixed $k > r$ (where r is such that ra is greater than the range of $\tilde{\Phi}_0$), we have

$$\begin{aligned}
&|\lambda^{-n}\mathfrak{L}_{na}f(Y) - \lambda^{-n}\mathfrak{L}_{na}f(Y')|\\
&\leq \|f\|\lambda^{-n}\int_{[0,na)} e^{-U(X)}\,dX\\
&\quad \times |e^{-I(X\,|\,\tau_{na}Y)} - e^{-I(X\,|\,\tau_{na}Y')}|\\
&\leq \lambda^{-n}\|f\|\int_{[0,na)} e^{-U(X)}\,dX\\
&\quad \times |\exp[-I_\Phi(X\,|\,\tau_{na}Y)] - \exp[-I_{\Phi_k}(X\,|\,\tau_{na}Y)]|\\
&\quad + \lambda^{-n}\|f\|\int_{[0,na)} dX e^{-U(X)}\\
&\quad \times |\exp[-I_{\Phi_k}(X\,|\,\tau_{na}Y)] - \exp[-I_{\Phi_k}(X\,|\,\tau_{na}Y')]|\\
&\leq 2\|f\|\lambda^{-n}Z_{[0,na)}e^{\|\tilde{\Phi}\|}\|\tilde{\Phi} - \tilde{\Phi}_k\|\\
&\quad + \lambda^{-(n-k)}Z_{[0,(n-k)a)}e^{\|\tilde{\Phi}\|}\|f\|\\
&\quad \times \lambda^{-k}\int_{[(n-k)a,na)} |\exp[-I_{\Phi_k}(X\,|\,\tau_{na}Y)]\\
&\quad - \exp[-I_{\Phi_k}(X\,|\,\tau_{na}Y')]|\,e^{-U(X)}\,dX,
\end{aligned}$$

when $\lambda^{-(n-k)}Z_{[0,(n-k)a)}$ has to be taken equal to one if $n - k \leq 0$; now using Lemma 2, the chain of inequalities ends as

$$\begin{aligned}
&\leq 2\|f\|\frac{e^{2\|\tilde{\Phi}\|}}{C_0}\|\tilde{\Phi} - \tilde{\Phi}_k\|\\
&\quad + \|f\|\frac{e^{2\|\tilde{\Phi}\|}}{C_0}\lambda^{-k}\int_{[0,ka)} |\exp[-I_{\Phi_k}(X\,|\,\tau_{ka}Y)]\\
&\quad - \exp[-I_{\Phi_k}(X\,|\,\tau_{ka}Y')]|\,e^{-U(X)}\,dX, \qquad (22')
\end{aligned}$$

and this proves the lemma because the function of Y, Y' appearing in the last inequality tends to zero as $Y' \to Y$ (consequence of the continuity properties of the potentials) and because $\lim \|\tilde{\Phi} - \tilde{\Phi}_k\| = 0$.

Remark: This lemma implies that the set of functions $\{\lambda^{-n}\mathfrak{L}_{na}f\}$ is equicontinuous and (since $\|\lambda^{-n}\mathfrak{L}_{na}\| \leq e^{2\|\tilde{\Phi}\|}C_0^{-1}$) norm bounded, so it is conditionally compact[8]; in particular, there exists a subsequence $\{n_i\}$ and a function $h \in C(K_+)$ (depending on f and $\{n_i\}$) such that

$$\lim_{i\to\infty}\|\lambda^{-n_i}\mathfrak{L}_{n_ia}f - h\| = 0. \tag{23}$$

Lemma 9: There exists $h \in C(K_+)$, $h \geq 0$, $\nu(h) = 1$, and

(i) $$\lambda^{-1}\mathfrak{L}_a h = h, \tag{24}$$

(ii) $$\lim_{n\to\infty}\|\lambda^{-n}\mathfrak{L}_{na}f - \nu(f)h\| = 0, \quad f \in C(K_+), \tag{25}$$

(iii) $$\lim_{n\to\infty}\lambda^{-n}\mathfrak{L}^*_{na}\mu - \mu(h)\nu = 0, \quad \mu \in C(K_+)^*, \tag{26}$$

where the limit holds in the weak sense.

In fact, consider the function $g = 1 - \lambda^{-1}\mathfrak{L}_a 1$; we have $\nu(g) = 0$. Thus, using Lemmas 7 and 3, we find

$$\lim_{n\to\infty}\|\lambda^{-n}\mathfrak{L}_{na}(1 - \lambda^{-1}\mathfrak{L}_a 1)\| = 0, \tag{27}$$

and then (27) and (23) imply $\|h - \lambda^{-1}\mathfrak{L}_a h\| = 0$. Since clearly $\nu(h) = 1$, $h \geq 0$, (i) is proved.

[8] N. Dunford and J. Schwartz, *Linear Operators* (Interscience Publishers, Inc., N.Y., 1958), Vol. I, Chap. IV, Sec. 6, item 7.

To prove (ii) consider a function $f \in C(K_+)$ and define $\tilde{g} \in C(K_+)$ as $\tilde{g} = f - \nu(f)h$; clearly $\nu(\tilde{g}) = 0$. Thus, using Lemmas 7 and 3,

$$0 = \lim_{n\to\infty} \|\lambda^{-n}\mathcal{L}_{na}g\| = \lim_{n\to\infty} \|\lambda^{-n}\mathcal{L}_{na}f - \nu(f)h\|.$$

Part (iii) is simply the dual statement of (ii).

Remark 1: (ii) implies that h is the unique solution of the equation $e^{+aP}h = \mathcal{L}_a h$. Since the commutativity of the operators $\mathcal{L}_x$ implies that also $\mathcal{L}_x h$ is a solution of the same equation, we must have $\mathcal{L}_x h = \lambda(x)h$, and (10) implies $\lambda(x)\lambda(y) = \lambda(x+y)$ $x, y \geq 0$; we have also that $\lambda(x)$ is a finite-valued continuous function of x because $(\mathcal{L}_x f)(Y)$, as is easily verified, is a finite-valued continuous function of x at fixed Y. Thus[9] $\lambda(x) = e^{xP}$.

An analogous argument holds for ν which turns out to be the unique solution of the equation $\mathcal{L}_x^* \nu = e^{xP}\nu$.

Remark 2: one can drop in (ii) and (iii) the condition that n is an integer. This is a consequence of the following inequality, holding for n integer and $0 < x < a$:

$$\|e^{-naP}\mathcal{L}_{na}e^{-xP}\mathcal{L}_x f - \nu(f)h\| = \|e^{-xP}\mathcal{L}_x(\mathcal{L}_{na}e^{-anP}f - \nu(f)h)\| \leq e^{\|\tilde{\Phi}\|}Z_{[0,a)}\,\|\mathcal{L}_{na}e^{-naP}f - \nu(f)h\|.$$

4. PROOF OF THEOREM 1

To prove differentiability of $P(\Phi)$ it is necessary to study the dependence of ν and h on Φ. Let $\tilde{\Phi}'$ be either $\tilde{\Phi}_0$ or an arbitrary element in $\mathcal{B}$. Let us consider the potential $\Phi + z\tilde{\Phi}'$. Let us also emphasize the dependence of $\mathcal{L}_\alpha$, h, ν, and λ on Φ by writing $\mathcal{L}_{a,\Phi}$, h_Φ, ν_Φ, and λ_Φ. From the proof of Lemma 8 and from the continuity of $\lambda_{\Phi+z\tilde{\Phi}'}$ in z (which follows[9] from the convexity properties of λ_Φ as a function of Φ), it follows easily that the limit

$$\lim_{Y'\to Y} \lambda_{\Phi+z\tilde{\Phi}'}^{-n}\mathcal{L}_{na,\Phi+z\tilde{\Phi}'}f(Y') = \lambda_{\Phi+z\tilde{\Phi}'}^{-n}\mathcal{L}_{na,\Phi+z\tilde{\Phi}'}f(Y) \tag{28}$$

holds uniformly not only in n, but also in z for z in a bounded interval. This implies that (Lemma 9) the limit

$$\lim_{Y'\to Y} \nu_{\Phi+z\tilde{\Phi}'}(f)h_{\Phi+z\tilde{\Phi}'}(Y') = \nu_{\Phi+z\tilde{\Phi}'}(f)h_{\Phi+z\tilde{\Phi}'}(Y) \tag{29}$$

holds uniformly for z in a bounded interval. From this equicontinuity property it follows that if $\{z_n\}$ is a sequence $z_n \xrightarrow[n\to\infty]{} z_0$, there exists a subsequence $\{n_i\}$ such that the limits

$$\lim_{i\to\infty} \nu_{\Phi+z_{n_i}\tilde{\Phi}'}(f), \tag{30}$$

$$\lim_{i\to\infty} \nu_{\Phi+z_{n_i}\tilde{\Phi}'}(f)h_{\Phi+z_{n_i}\tilde{\Phi}'} \tag{31}$$

exist (the second in the norm sense). Since $\mathcal{L}_{\Phi+z\tilde{\Phi}'}$ and $\lambda_{\Phi+z\tilde{\Phi}'}$ depend continuously (in the operator norm sense) on z, the limit (31) has to be an eigenvalue of $\mathcal{L}_{\Phi+z_0\tilde{\Phi}'}$ and so it must be proportional to $h_{\Phi+z_0\tilde{\Phi}'}$. From this it follows that

$$\lim_{z\to z_0} \nu_{\Phi+z\tilde{\Phi}'}(f) = \nu_{\Phi+z_0\tilde{\Phi}'}(f), \quad f \in C(K_+), \tag{32}$$

$$\lim_{z\to z_0} h_{\Phi+z\tilde{\Phi}'} = h_{\Phi+z_0\tilde{\Phi}'}, \tag{33}$$

i.e., $h_{\Phi+z\tilde{\Phi}'}$, $\nu_{\Phi+z\tilde{\Phi}'}$ are, respectively, norm and weakly continuous in z.

Now observe that the operator defined by

$$\left(\frac{\partial\mathcal{L}_{a,\Phi}}{\partial\tilde{\Phi}'}f\right)(Y) = -\int_{X\subset[0,a)} e^{-U_\Phi(X\,|\,\tau_a Y)} \times U_{\tilde{\Phi}'}(X\,|\,\tau_a Y)f(x \cup \tau_a Y)\,dX \tag{34}$$

is such that

$$\left\|\frac{\mathcal{L}_{a,\Phi+z\tilde{\Phi}'} - \mathcal{L}_{a,\Phi}}{z} - \frac{\partial\mathcal{L}_{a,\Phi}}{\partial\tilde{\Phi}'}\right\| \xrightarrow[z\to 0]{} 0. \tag{35}$$

So, using the just-proved continuity properties of ν, h, and λ and the identity

$$\frac{\lambda_{\Phi+z\tilde{\Phi}'} - \lambda_\Phi}{z} = \nu_\Phi\left(\frac{\mathcal{L}_{a,\Phi+z\tilde{\Phi}'} - \mathcal{L}_{a,\Phi}}{z}\,\frac{h_{\Phi+z\tilde{\Phi}'}}{\nu_\Phi(h_{\Phi+z\tilde{\Phi}'})}\right), \tag{36}$$

we find that

$$\lim_{z\to 0} \frac{\lambda_{\Phi+z\tilde{\Phi}'} - \lambda_\Phi}{z} = \nu_\Phi\left(\frac{\partial\mathcal{L}_{a,\Phi}}{\partial\tilde{\Phi}'}\,h_\Phi\right), \tag{37}$$

which proves that

$$\frac{d}{dz}\,\lambda_{\Phi+z\tilde{\Phi}'} = \nu_{\Phi+z\tilde{\Phi}'}\left(\frac{\partial\mathcal{L}_{a,\Phi+z\tilde{\Phi}'}}{\partial z}\,h_{\Phi+z\tilde{\Phi}'}\right)$$

and also that this derivative is a continuous function of z (since $\nu_{\Phi+z\tilde{\Phi}'}$ is weakly continuous in z and $\partial\mathcal{L}_{a,\Phi+z\tilde{\Phi}'}/\partial z$ and $h_{\Phi+z\tilde{\Phi}'}$ are norm continuous in z). Now part (i) of Theorem 1 follows easily.

To prove the cluster property we use a procedure essentially contained in Ref. 2 and used there to prove some ergodicity properties of the equilibrium state in lattice systems.

We have first to construct the state $\bar{\gamma}$ which corresponds to the equilibrium state of the system when it occupies all R.

[9] N. Dunford and J. Schwartz, *Linear Operators* (Interscience Publishers, Inc., N.Y., 1958), Vol. I, Chap. VIII, Sec. 1, item 2.

In the remainder of this section the letters b and b' with or without indices will denote finite real numbers. Since the sets $C_{(b,b')} \subset C(K)$ of functions $A \in C(K)$ such that $A(X) = A(X \cap (b, b'))$ are dense in $C(K)$, it is sufficient, in order to determine the probability measure $\bar{\gamma}$ on K, to find $\bar{\gamma}(A)$ for all $A \in C_{(b,b')}$ with $b < b'$ arbitrary.

We remark that if $A \in C_{(b,b')}$, then $\tau_b A$ (translate of A by a length b) can be identified with an element $(\tau_b A)_+ \in C(K_+)$ defined for $X \in K_+$ as

$$(\tau_b A)_+(X) = (\tau_b A)(X \cap (0, b' - b)) = A(X \cap (0, b' - b) + b). \quad (38)$$

With this notation and denoting $\bar{\gamma}_{(c,d)}$ the normalized Gibbs measure on (c, d), i.e., the measure

$$\bar{\gamma}_{(c,d)}(dX) = [e^{-U(X)}\, dX/Z_{(c,d)}], \quad X \subset (c, d),$$

the value of $\bar{\gamma}(A)$ can be defined as

$$\begin{aligned}\bar{\gamma}(A) &= \lim_{\substack{y \to +\infty \\ y' \to +\infty}} \bar{\gamma}_{(-y,y')}(A) \\ &= \lim_{\substack{y \to +\infty \\ y' \to +\infty}} \bar{\gamma}_{(0,y+y')}(\tau_y A) \\ &= \lim_{\substack{y_1 \to +\infty \\ y_2 \to +\infty}} \bar{\gamma}_{(0,y_1+y_2)}(\tau_{y_1}(\tau_b A)_+), \end{aligned} \quad (39)$$

provided the limits exist.

To prove the existence of the limit (39), consider a function $f \in C_{(0,ma)} \subset C(K_+)$ and $n > m$. Then one easily verifies, using definitions (7) and (9) of $\mathfrak{L}_a$ and $\mathfrak{L}_a^*$, that

$$\bar{\gamma}_{(0,na)}(f) = \frac{(e^{-naP}\mathfrak{L}_{na}^* \delta_\varnothing)(f)}{(e^{-naP}\mathfrak{L}_{na}^* \delta_\varnothing)(1)}, \quad (40)$$

where $\delta_\varnothing$ is defined by $\delta_\varnothing(f) = f(\varnothing)$.

From the definitions (7) of $\mathfrak{L}_a$, one verifies also that

$$\mathfrak{L}_{na}(\tau_{na} f) = f(\mathfrak{L}_{na} 1), \quad f \in C(K_+), \quad n \geq 0, \quad (41)$$

and so, using (40), (41), (25), and (26), we have

$$\begin{aligned}\lim_{\substack{y_1 \to \infty \\ y_2 \to \infty}} &\bar{\gamma}_{(0,y_1+y_2)}(\tau_{y_1}(\tau_b A)_+) \\ &= \lim_{\substack{y_1 \to \infty \\ y_2 \to \infty}} \frac{(e^{-(y_1+y_2)P}\mathfrak{L}_{y_1+y_2}^* \delta_\varnothing)(\tau_{y_1}(\tau_b A)_+)}{(e^{-(y_1+y_2)P}\mathfrak{L}_{y_1+y_2}^* \delta_\varnothing)(1)} \\ &= h(\varnothing)^{-1} \lim_{\substack{y_1 \to \infty \\ y_2 \to \infty}} (e^{-y_2 P}\mathfrak{L}_{y_2}^* \delta_\varnothing)((\tau_b A)_+(e^{-y_1 P}\mathfrak{L}_{y_1} 1)) \\ &= \nu(h(\tau_b A)_+). \end{aligned}$$

Hence we have found

$$\bar{\gamma}(A) = \nu(h(\tau_b A)_+) \quad \text{if} \quad A \in C_{(b,b')}; \quad (42)$$

this formula proves also the translational invariance of $\bar{\gamma}$.

Let now $A_1 \in C_{(b_1,b_1')}$, $A_2 \in C_{(b_2,b_2')}$. Then, using the following formula easily deduced from definition (9),

$$(\tau_x f)(X)\nu(dX) = (e^{-xP}\mathfrak{L}_x^* f\nu))(dX), \quad f \in C(K_+), \quad (43)$$

and the property that $\mathfrak{L}_x^* \nu = e^{xP}\nu$, one can deduce [taking also into account (25), (26), and (42)] that

$$\begin{aligned}\lim_{x \to \infty} &\bar{\gamma}(A_1 \tau_x A_2) \\ &= \lim_{x \to \infty} \bar{\gamma}((\tau_{b_1} A_1)_+ \tau_x (\tau_{b_2} A_2)_+) \\ &= \lim_{x \to \infty} \int h(X)(\tau_{b_1} A_1)_+(X)(\tau_x(\tau_{b_2} A_2)_+)(X)\nu(dX) \\ &= \lim_{x \to \infty} \int h(X)(\tau_{b_1} A_1)_+(X)(e^{-xP}\mathfrak{L}_x^*(\tau_{b_0} A_2)_+\nu)(dX) \\ &= \int h(X)(\tau_{b_1} A_1)_+(X)\nu(dX) \int h(X)(\tau_{b_2} A_2)_+\nu(dX) \\ &= \bar{\gamma}(A_1)\bar{\gamma}(A_2). \end{aligned}$$

Now we prove the more general cluster property (6) by induction. Suppose it is true for the product of $A_2 \in C_{(b_2,b_2')} \cdots A_n \in V_{(b_n,n')}$, and let $A_1 \in C_{(b_1,b_1')}$; using (43) and the uniformity of the convergence of $(e^{-x_1 P}\mathfrak{L}_{x_1} h(\tau_{b_1} A)_+)$ to $h\bar{\gamma}(A_1)$, and the fact that

$$\|(\tau_{b_2} A_2)_+ \cdots \tau_{x_2+\cdots+x_{n-1}}(\tau_{b_n} A_n)_+\| \leq \|A_2\| \cdots \|A_n\|,$$

we have:

$$\begin{aligned}\lim_{x_i \to \infty} &\bar{\gamma}(A_1 \tau_{x_1} A_2 \cdots \tau_{x_1+\cdots+x_{n-1}} A_n) = \lim_{x_i \to \infty} \bar{\gamma}((\tau_{b_1} A_1)\tau_{x_1}(\tau_{b_2} A_2) \cdots \tau_{x_2+\cdots+x_{n-1}}(\tau_{b_n} A_n)) \\ &= \lim_{x_i \to \infty} \int h(X)(\tau_{b_1} A_1)_+(X)\{\tau_{x_1}[(\tau_{b_2} A_2)_+ \cdots \tau_{x_2+\cdots+x_{n-1}}(\tau_{b_n} A_n)_+]\}(X)\nu(dX) \\ &= \lim_{x_i \to \infty} \int h(X)(\tau_{b_1} A_1)_+(X)\{e^{-x_1 P}\mathfrak{L}_{x_1}^*[(\tau_{b_2} A_2)_+ \cdots \tau_{x_2+\cdots+x_{n-1}}(\tau_{b_n} A_n)_+]\nu\}(dX) \\ &= \lim_{x_i \to \infty} \int [e^{-x_1 P}\mathfrak{L}_{x_1}(h(\tau_{b_1} A_1)_+)](X)[(\tau_{b_2} A_2)_+ \cdots \tau_{x_2+\cdots+x_{n-1}}(\tau_{b_n} A_n)_+](X)\nu(dX) \\ &= \lim_{x_i \to \infty\, i>1} \int \bar{\gamma}(A_1)(\tau_{b_2} A_2)_+(X) \cdots (\tau_{x_2+\cdots+x_{n-1}}(\tau_{b_n} A_n)_+)(X)\nu(dX) \\ &= \bar{\gamma}(A_1) \cdots \bar{\gamma}(A_n). \end{aligned}$$

Finally the restrictive hypothesis that $A_i \in C_{(b_i, b_{i'})}$ can be released by density arguments.

5. DESCRIPTION OF THE EQUILIBRIUM STATE $\bar{\gamma}$

We have seen in the preceding section that h and ν determine completely the equilibrium state of the system considered. In this section we describe $\bar{\gamma}$ by means of a family of density distributions[10] in the case that Φ has finite range.

We remark that if the range of the interaction is between na and $(n+1)a$, then the function h has the property that $h(Y) = h(Y \cap [0, (n+1)a))$; this is because $h = \lim_n \lambda^{-n}\mathfrak{L}_{na}1$ and $\lambda^{-n}\mathfrak{L}_{na}1$ depends only on $Y \cap [0, (n+1)a)$.

Now the state $\bar{\gamma}$ can be described by the family of density distributions $f_L(x)$ on $[0, La)$ which have the meaning of probability densities (with respect to the measure dX, $X \subset [0, La)$) for finding the configuration X inside $[0, La)$ irrespective of what happens outside.

If $L \geq n+1$, these probabilities can be defined as

$$f_L(X) = \lim_{\substack{m_1\to\infty \\ m_2\to\infty}} Z^{-1}_{(-m_1a,(L+m)a)} \int_{\substack{X_1\subset[-m_1a,0) \\ X_2\subset[L_2a,(L+m_2)a)}} dX_1\, dX_2 \times e^{-U(X_1|X)} e^{-U(X)} e^{-U(X_2|X)}. \quad (44)$$

One can define a family of operators $\tilde{\mathfrak{L}}_x$ on $C(K_-)$ which are the analogies of the $\mathfrak{L}_x$ for left-semi-infinite systems as

$$(\tilde{\mathfrak{L}}_x f)(Y) = \int_{X\subset[-x,0)} e^{-U(X|\tau_{-x}Y)} f(X \cup \tau_{-x}Y)\, dX, \qquad Y \in K_-. \quad (45)$$

The theory of these operators is exactly the same as that for $\mathfrak{L}_x$, so there exists $\tilde{h} \in C(K_-)$ such that $\tilde{\mathfrak{L}}_x \tilde{h} = e^{xP}\tilde{h}$, and $\tilde{h}(Y) = \lim_{x\to\infty} e^{-xP}\tilde{\mathfrak{L}}_x 1(Y)$ (uniformly in $Y \in K_-$). An explicit expression for $\tilde{h}$ in terms of eigenfunctions of operators of the type of $\mathfrak{L}$ can be given by considering the potential $\mathfrak{I}\Phi$ defined as the mirror image of Φ, and let $h^{\mathfrak{I}}$ be the eigenfunction of the operator $\mathfrak{L}$ corresponding to $\mathfrak{I}\Phi$; then one can prove that

$$\tilde{h}(\tau_{-La}Y) = \lim_{m\to\infty} (e^{-amP}\tilde{\mathfrak{L}}_{ma}1)(\tau_{-La}Y) = h^{\mathfrak{I}}(\mathfrak{I}Y), \quad (46)$$

where $\mathfrak{I}Y$ is the configuration obtained from $Y \subset [0, La)$ by reflecting $\tau_{-La}Y$ around the origin.

[10] D. Ruelle, J. Math. Phys. **8**, 1657, 1967.

Now (44) can be written in terms of $\mathfrak{L}$ and $\tilde{\mathfrak{L}}$ as

$$f_L(X) = \lim_{\substack{m_1\to\infty \\ m_2\to\infty}} (\mathfrak{L}^*_{(m_1+m_2+L)a}\delta_{\varnothing})(1)^{-1}(\mathfrak{L}_{m_1a}1)(Y) \times e^{-U(Y)}(\tilde{\mathfrak{L}}_{m_2a}1)(\tau_{-La}Y); \quad (47)$$

hence,

$$f_L(X) = e^{-PLa}h(\varnothing)^{-1}h(Y)h^{\mathfrak{I}}(\mathfrak{I}Y)e^{-U(Y)}. \quad (48)$$

One easily verifies, using the properties of h and $h^{\mathfrak{I}}$ as eigenvectors of $\mathfrak{L}$ and $\tilde{\mathfrak{L}}$, the normalization and compatibility conditions implicit in the meaning of f_L.

If we consider now the functional on the set $E \cap \mathfrak{L}^{\perp}$ of the translationally invariant measures on K defined as the difference between the mean entropy[11] and the mean energy, then, as a consequence of the differentiability properties of the pressure, one finds that this functional attains its maximum at one unique point of $E \cap \mathfrak{L}^{\perp}$ which coincides with $\bar{\gamma}$.[12,13] The value of this maximum is P.

This last property can easily be verified by writing

$$s(\bar{\gamma}) - U(\bar{\gamma}) = \lim_{L\to\infty} La^{-1} \int_{[0,La)} f_L(x)[-\lg f_L(X) - U(X)]\, dX,$$

and using (48).

We mention without producing the explicit calculations that one can find a sufficiently large class of elements $\epsilon \subset E \cap \mathfrak{L}^{\perp}$ and suitably parametrize its elements so that the variational equations corresponding to the extremum problem $\max_{\rho\in\epsilon} s(\rho) - U(\rho)$ give rise to the integral equation $\lambda^{-1}\mathfrak{L}_a h = h$ and to the expression (48) for the state maximizing $s(\rho) - U(\rho)$. In this context one could use the results of this paper to guarantee that the state in $\mathcal{E}$ maximizing $s(\rho) - U(\rho)$ is the true equilibrium state.[14]

ACKNOWLEDGMENTS

We are deeply indebted to D. Ruelle for suggesting the problem, for enlightening discussions, and for reading the manuscript. We also thank M. L. Motchane for his kind hospitality at IHES. One of us (S. M.-S.) is indebted to CNRS for financial support.

[11] D. Robinson and D. Ruelle, Commun. Math. Phys. **5**, 324 (1967).

[12] D. Ruelle, Commun. Math. Phys. **5**, 324 (1967).

[13] G. Gallavotti and S. Miracle-Sole, Ann. Inst. Henri Poincaré **8**, 287 (1968).

[14] S. Miracle-Sole, Thesis, University Aix-Marseille, France, 1968.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

On the Asymptotic Stability of Reactors with Arbitrary Feedback

A. Z. Akcasu and P. Akhtar
Department of Nuclear Engineering, The University of Michigan, Ann Arbor, Michigan

(Received 26 April 1969)

A theorem for the boundedness and asymptotic stability of a point reactor with an arbitrary feedback is stated and proved. The criteria obtained are shown to be essentially the same as those given by Akcasu and Dolfes. The theorem is applied to a reactor with an arbitrary linear feedback and to a xenon-controlled reactor with a flux reactivity coefficient whose feedback mechanism involves quadratic nonlinearity. It is also compared to a criterion obtained by Corduneanu in the case when delayed neutrons are ignored and the feedback mechanism is linear.

I. INTRODUCTION

Point kinetic equations for a nuclear reactor with an arbitrary feedback can be written as

$$\dot{z} = (1 + z)\delta K_f[z, t] - \sum_{i=1}^{6} a_i(z - z_i), \quad (1)$$

$$\dot{z}_i = h_i(z - z_i), \quad i = 1, 2, \cdots, 6, t \geq 0, \quad (2)$$

where t denotes time and the dot indicates the first derivative of the dependent variables $z(t)$ and $z_i(t)$ with respect to t. The functions $z(t)$ and $z_i(t)$ represent the incremental reactor power and delayed neutron precursor densities, respectively.[1] The parameters a_i and h_i are positive numbers, with the a_i satisfying the relation

$$\sum_{i=1}^{6} a_i = 1. \quad (3)$$

The symbol $\delta K_f[z, t]$ denotes a functional of the function $z(t)$ involving values of the latter in the interval $(-\infty, t)$. We shall always assume that $\delta K_f[z, t]$ can be represented as

$$\delta K_f[z, t] = \int_{-\infty}^{t} du G_1(t - u)z(u) + \int_{-\infty}^{t} du \int_{-\infty}^{t} dv G_2(t - u, t - v)z(u)z(v) + \cdots, \quad (4)$$

where $G_1(u)$, $G_2(u, v)$, etc., are linear, quadratic, etc., feedback kernels. These kernels are defined only for positive arguments, but it turns out to be convenient to define them to be identically zero for negative arguments.

Physically, $\delta K_f[z, t]$ denotes the feedback reactivity at t due to the power generation in the reactor prior to t. From the physical nature of the feedback mechanism we require $\delta K_f[z, t]$ to be a bounded function of time whenever $z(t)$ is bounded (stability of the feedback). Mathematically this condition is satisfied if we assume that the kernels $G_n(u_1, u_2, \cdots, u_n)$ are absolute integrable, i.e.,

$$\gamma_n = \int_0^{\infty} du_1 \cdots \int_0^{\infty} du_n |G_n(u_1, u_2, \cdots, u_n)|, \quad (5)$$

and that the power series $\sum_{n-1}^{\infty} \gamma_n M^n$ is convergent for all finite values of M where M is the bound of $z(t)$.

Equations (1) and (2) describe the temporal behavior of $z(t)$, $z_i(t)$ only for positive t. Since the feedback mechanism depends on the values of $z(t)$ in the interval $(-\infty, 0)$ as well as in $(0, t)$, a unique solution of this set of equations requires a specification not only of the initial value $z(0)$ and $z_i(0)$ but also of the values of $z(t)$ prior to $t = 0$. We shall refer to $z(t)$ for $t \leq 0$ as an initial curve. It is clear from the form of Eqs. (1) and (2) and the definition of $\delta K_f[z, t]$ that $z(t) = z_i(t) \equiv 0$, $t > 0$, is a solution of Eqs. (1) and (2) corresponding to the initial curve $z(t) \equiv 0$ for $t \leq 0$. We shall refer to this solution as the equilibrium state.[3] A nonzero initial curve, which is the response of the reactor to external reactivity changes and external sources for $t \leq 0$, can be regarded as a perturbation on this equilibrium state.

The question of stability involves the behavior of $z(t)$ and $z_i(t)$ for $t > 0$, and in particular when $t \to \infty$. Since these functions depend on the nature of the initial curves as well as the initial values, one

[1] The usual point reactor kinetic equations[2] can be reduced to the form of Eqs. (1) and (2) by choosing $z(t) = [P(t) - P_0]/P_0$, $z_i(t) = [C_i(t) - C_{i0}]/C_{i0}$, and by letting $\to \beta t/l$, $a_i = \beta_i/\beta$, and $h_i = l\lambda_i/\beta$. P_0 and C_{i0} denote equilibrium values of reactor power and delayed neutron precursor densities, respectively, and l, λ_i, β_i, β have their conventional meanings in the field of reactor engineering. P_0 and C_{i0} are determined by $K_0 + K_f(P_0) = 0$ and $C_{i0} = a_i P_0/\lambda_i$, where K_0 and $K_f(P_0)$ are external and equilibrium feedback reactivities. The incremental feedback reactivity functional $\delta K_f[z, t]$ appearing in Eq. (1) is then defined as $\delta K_f[z, t] \equiv K_f[P, t] - K_f(P_0)$, where $K_f[P, t]$ is the total feedback reactivity functional. We will assume that the algebraic equation relating K_0 and $K_f(P_0)$ has a unique solution. We also note that $(1 + z)$ and $(1 + z_i)$ are nonnegative, since P_0 and C_{i0} can never be negative.

[2] A. F. Henry, Nucl. Sci. Eng. 3, 52 (1958).

[3] $z(t) = z_i(t) \equiv -1$ for all t also represents an equilibrium state. Physically, this corresponds to a reactor in which there are no neutrons and delayed neutron precursors. In the course of a derivation we shall exclude this equilibrium state from discussion because it will be shown that, when conditions for boundedness as stated in the theorem in the next section are satisfied, $z(t)$ and $z_i(t)$ can never approach -1 [cf. Eqs. (14a) and (14b)] once the reactor is perturbed.

must first specify a certain class of initial curves in order to state conditions for stability. This class must be sufficiently broad to include all the possible perturbations that may arise intentionally or accidentally during the operation of the reactor. On physical grounds we choose this class of initial curves, henceforth called physically admissible initial curves, to satisfy the following properties:

(i) $z(t) \equiv 0$ for $t \leq -t_0$, $t_0 > 0$;
(ii) $z(t) > -1$, $-t_0 \leq t \leq 0$;
(iii) $z(t)$ is piecewise continuous and bounded, and its first derivative exists and is bounded at all times, from both left- and right-hand side, in the interval $(-t_0, 0)$.

Condition (i) implies that perturbations are confined to a finite time interval. Condition (ii) follows from the positivity of reactor power during reactor operation. The boundedness and piecewise continuity of $z(t)$ and the boundedness of its first derivative are indicated by the fact that an initial curve is the response of a reactor to an external disturbance either in the reactivity or in the source for $t \leq 0$. Jump discontinuities are permitted to allow for the presence of a possible pulse source in the time interval $(-t_0, 0)$.

We shall not reproduce here the various precise definitions of stability.[4] We shall be mainly concerned with asymptotic stability in the large which we define, for any physically admissible initial curve and for finite $z(0)$ and $z_i(0)$, as

$$\lim_{t \to \infty} z(t) = 0. \tag{6}$$

A criterion for the asymptotic stability of the reactors described by Eqs. (1) and (2) was derived by Akcasu and Dalfes[5] with an heuristic approach. The derivation was largely based on intuitive reasoning and the results were justified on the basis of energy considerations. In the present work a rigorous analysis is carried out and the results are stated in the form of a theorem. It is found that in addition to the criterion in Ref. 5 certain additional restrictions have to be imposed on the feedback functional $\delta K_f[z, t]$ to guarantee asymptotic stability. These restrictions, however, turn out to be not very stringent and seem to be quite compatible with physically realizable reactor systems. This will be demonstrated for reactors with linear feedback and for a xenon-controlled reactor with flux reactivity coefficient where the feedback functional is nonlinear.[6,7] In the course of the derivation of the stability criterion, it is also observed that the restrictions on $\delta K_f[z, t]$ are connected with conditions imposed on initial curves. The choice of an unrealistically broad class of initial curves for the sake of mathematical generality will yield sufficient conditions for asymptotic stability which may turn out to be too restrictive to be of any practical interest for reactor applications. This necessitates the restriction of the initial perturbations to physically admissible initial curves as defined above.

II. STABILITY THEOREMS

Theorem 1 (Boundedness): The response of a nuclear reactor, described by Eqs. (1) and (2), is always bounded for any physically admissible initial curve if:

(i) $z(0)$, $z_i(0)$ are bounded and greater than -1;
(ii) $\int_{-\infty}^{t} \delta K_f[y, t']y(t')\,dt' \leq 0$, $t > 0$, for all test functions $\{y(t)\}$ which belong to the class of physically admissible initial curves for $t \leq 0$ and are arbitrary for $t \geq 0$.[8]

Theorem 2 (Asymptotic Stability): The response $z(t)$ is asymptotically stable in the large if, in addition to Theorem 1, the feedback functional satisfies the following conditions:

(i) $\delta K_f[v, t]$ is uniformly continuous for sufficiently large t for all test functions $\{v(t)\}$ which belong to the class of physically admissible initial curves for $t \leq 0$, and are continuous, bounded, and have bounded first derivatives for $t \geq 0$.

(ii) $\lim \delta K_f[\omega, t] = 0$, as $t \to \infty$, implies $\lim \omega(t) = 0$, as $t \to \infty$, for all test functions $\{\omega(t)\}$[9] which belong to the class of physically admissible initial curves for $t \leq 0$ and, for $t \geq 0$, are continuous, bounded, and have uniformly continuous first derivatives which vanish as $t \to \infty$.

III. THE PROOF OF BOUNDEDNESS

We consider a functional of $z(t)$ defined as

$$V[z, t] = F(z) + \sum_{i=1}^{6} \frac{a_i}{h_i} F(z_i) - \int_{-\infty}^{t} \delta K_f[z, t']z(t')\,dt', \quad t \geq 0, \tag{7}$$

[4] N. N. Krasovskii, *Stability of Motion* (Stanford University Press, Stanford, California, 1963).
[5] A. Z. Akcasu and A. Dalfes, Nucl. Sci. Eng. **8**, 89 (1960).
[6] J. Chernick, G. Lellouche, and W. Wollman, Nucl. Sci. Eng. **10**, 120 (1960).
[7] A. Z. Akcasu and P. Akhtar, J. Nucl. Energy **21**, 341 (1967).
[8] It should be noted that there is no restriction on the bound of the test functions $\{y(t)\}$ for $t \geq 0$, and they may diverge as $t \to \infty$. Clearly $\{y(t)\}$ contains all possible solutions of Eqs. (1) and (2) as a subset.
[9] It may be noted that the test functions $\{\omega(t)\}$ form a subset of the functions $\{v(t)\}$, which in turn are a subset of test functions $\{y(t)\}$.

in which $F(u)$, $u = z, z_1, \cdots, z_6$, is defined as

$$F(u) \equiv u - \ln(1+u) \equiv \int_0^u \frac{x}{1+x}\,dx, \quad u > -1.$$

Variations of $F(u)$ with u are shown in Fig. 1. It is a positive definite function possessing the following properties:

(i) $F(u)$ is real and positive in $(-1, \infty)$ except for $u = 0$;
(ii) $F(0) = 0$;
(iii) $F(u)$ is continuous in $(-1, \infty)$.

The first two terms of $V[z, t]$ in Eq. (7) are thus nonnegative, since a_i, h_i are known to be positive constants. The last term is also nonnegative if condition (ii) for boundedness is imposed for all test functions $\{y(t)\}$ and thus necessarily for any possible trajectory $z(t)$ corresponding to a given physically admissible initial curve. Hence,

$$V[z, t] \geq 0. \tag{8}$$

The initial value $V(0)$ is given by

$$V(0) = F(z_0) + \sum_{i=1}^{6} \frac{a_i}{h_i} F(z_{i0}) - \int_{-\infty}^{0} \delta K_f[z, t']z(t')\,dt'. \tag{9}$$

It will be seen that $V(0)$ is finite since initial values z_0 and z_{i0} are finite (condition (i)) and the finiteness of the last term,

$$\int_{-\infty}^{0} \delta K_f[z, t']z(t')\,dt',$$

is guaranteed by the finiteness of the physically admissible initial curves and feedback stability. Moreover, by differentiating $V[z, t]$ with respect to t and using Eqs. (1) and (2), we obtain the derivative

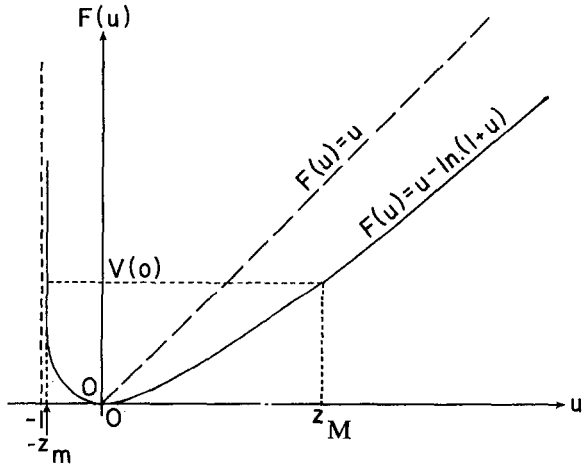


FIG. 1.

of $V[z, t]$ along a trajectory as

$$\dot{V}[z, t] = -\sum_{i=1}^{6} \frac{a_i(z - z_i)^2}{(1+z)(1+z_i)}, \quad t > 0, \tag{10}$$

which is continuous whenever $z(t) \neq -1$ and $z_i(t) \neq -1$ inasmuch as $z(t)$ and $z_i(t)$ are continuous in t by virtue of Eqs. (1) and (2). Furthermore, since $(1 + z)$ and $(1 + z_i)$ are nonnegative by definition,[1]

$$\dot{V}[z, t] \leq 0, \quad t > 0; \tag{11}$$

the equality occurs when $z = z_i$.

It is thus concluded that $V[z, t]$ is a nonnegative, monotonically decreasing function of time along a trajectory. In particular, we obtain

$$V[z, t] \leq V(0), \tag{12}$$

which indicates that $V[z, t]$ is finite for all $t \geq 0$ if $V(0)$ is finite. We observe that $V[z, t]$ is a sum of three nonnegative terms (cf. Eq. 7). Therefore, each term in Eq. (7) is smaller than $V(0)$ for all $t \geq 0$:

$$F(z) \leq V(0), \tag{13a}$$

$$F(z_i) \leq V(0)(h_i/a_i), \tag{13b}$$

$$\left| \int_{-\infty}^{t} \delta K_f[z, t']z(t')\,dt' \right| \leq V(0). \tag{13c}$$

The inequalities (13a) and (13b), with the help of Fig. 1, imply that

$$-1 < -z_m \leq z(t) \leq z_M, \tag{14a}$$

$$-1 < -z_{m_i} \leq z_i(t) \leq z_{M_i}, \tag{14b}$$

where z_m, z_M, z_{m_i}, and z_{M_i} are positive numbers such that

$$F(z_M) = F(-z_m) = V(0), \quad z_M > z_m, \tag{15a}$$

and

$$F(z_{M_i}) = F(-z_{m_i}) = V(0)h_i/a_i, \quad z_{M_i} > z_{m_i}. \tag{15b}$$

The inequalities (14a) and (14b) establish the boundedness of the reactor response to any physically admissible initial curve. Furthermore, they also prove that both $z(t)$ and $z_i(t)$ are bounded away from -1 for $t > 0$, and hence the power can never approach the shutdown equilibrium state (cf. Ref. 3).

The conditions of Theorem 1, apart from the boundedness of $z(t)$ and $z_i(t)$, also lead to the following conclusions which will be useful in proving asymptotic stability:

(i) $\dot{z}(t)$ is bounded for all $t \geq 0$; this follows from Eq. (1) and stability of the feedback; that is,

$$|\dot{z}(t)| \leq (1 + z_M)W + 2z_M, \tag{16}$$

where z_M is assumed to represent the largest value of the upperbounds $z_M, z_{M_1}, \cdots, z_{M_6}$ and W is the

upper bound of $\delta K_f[z, t]$ corresponding to the given physically admissible initial curve. Hence $z(t)$ is not only bounded but also uniformly continuous.

(ii) $z_i(t)$ is also uniformly continuous, since

$$|\dot{z}_i(t)| \leq 2h_i z_M \tag{17}$$

[cf. Eq. (2)].

(iii) $\dot{V}(t)$ is continuous for all $t \geq 0$ because $(1 + z)$ and $(1 + z_i)$ are never zero by virtue of Eqs. (14a) and (14b) [cf. Eq. (10)]. Moreover,

$$|\dot{V}(t)| \leq 4z_M^2/(1 - z_m)^2 \tag{18a}$$

and

$$|\ddot{V}(t)| \leq \frac{4z_M(1 + z_M)}{(1 - z_m)^3} \times \left((1 + z_M)W + 2z_M\left(1 + \sum_{i=1}^{6} aihi\right)\right), \tag{18b}$$

where z_m is taken to be the least value of the lower bounds z_m, $z_{m_1}, \cdots, z_{m_6}$. Hence $\dot{V}(t)$ is bounded and uniformly continuous for all $t \geq 0$.

IV. THE PROOF OF ASYMPTOTIC STABILITY

In proving the theorem for asymptotic stability we make use of a lemma by Barbalat[10] which is reproduced here for convenience.

Lemma: Let $g(t)$ be a real function of a real variable t, defined for $t > a > 0$. If

(i) $\lim g(t) = g_\infty$, as $t \to \infty$, (g_∞ is finite), and

(ii) $\dot{g}(t)$ is uniformly continuous for $t > a$, then $\lim \dot{g}(t) = 0$, as $t \to \infty$.

We start with the observation that

$$\lim_{t\to\infty} V(t) = V_\infty \geq 0 \tag{19}$$

because $V[z, t]$ is nonnegative and monotomically decreasing. Also, $\dot{V}(t)$ is uniformly continuous (cf. Eqs. 18a, 18b). The lemma is thus applicable for $g(t) = V(t)$ and we have

$$\lim_{t\to\infty} \dot{V}(t) = 0 \tag{20}$$

along a trajectory. Combining Eqs. (10) and (20) we conclude that

$$\lim_{t\to\infty} [z(t) - z_i(t)] = 0, \quad i = 1, 2, \cdots, 6, \tag{21}$$

which, by virtue of Eq. (2), leads to

$$\lim_{t\to\infty} \dot{z}_i(t) = 0. \tag{22}$$

We next note that $\dot{z}(t)$ [cf. Eq. (1)] is uniformly continuous for large t since $\delta K_f[v, t]$ is uniformly continuous for all test functions $\{v(t)\}$ [cf. condition (i) of the Theorem 2] and thus necessarily for any trajectory $z(t)$. It may be noted that limiting the requirement of uniform continuity of $\delta K_f[v, t]$ to only bounded and continuous test functions having bounded derivatives, instead of the set of test functions $\{y, (t)\}$ (which are quite arbitrary and unbounded), is made possible because we already have established the boundedness of $z(t)$ and $\dot{z}(t)$. Moreover, $\dot{z}_i(t)$ is uniformly continuous by virtue of Eq. (2) and the fact that $z(t)$ and $z_i(t)$ are uniformly continuous as shown above [cf. Eqs. (16), (17)].

The conditions of the lemma are thus satisfied for the function $g \equiv [z(t) - z_i(t)]$ with $g_\infty = 0$ and hence $\lim \dot{g} = [\dot{z} - \dot{z}_i] = 0$, as $t \to \infty$.

The latter together with Eq. (22) leads to

$$\lim_{t\to\infty} \dot{z}(t) = 0, \tag{23}$$

which, in view of Eqs. (1) and (21), and the fact that $z(t)$ cannot approach -1, leads to

$$\lim_{t\to\infty} \delta K_f[z, t] = 0. \tag{24}$$

Condition (ii) can now be applied since we have shown that $z(t)$ is continuous, bounded, and has a uniformly continuous derivative which vanishes as $t \to \infty$ and thus belongs to the subset $\{\omega(t)\}$. Hence

$$\lim_{t\to\infty} z(t) = 0$$

and the equilibrium state $z(t) \equiv 0$ is asymptotically stable.

V. APPLICATIONS

In this section we apply the stability theorem to investigate the asymptotic stability of a reactor with an arbitrary linear feedback, and of a xenon-controlled reactor with a flux reactivity coefficient whose feedback mechanism involves quadratic nonlinearity.[6,7,11]

A. Linear Feedback

The incremental feedback functional in this case reduces to

$$\delta K_f[z, t] = \int_{-\infty}^{t} G(t - u)z(u)\, du. \tag{25}$$

The stability of the feedback mechanism requires the feedback kernel to be absolutely integrable,[12] i.e.,

$$\int_0^\infty |G(u)|\, du < \infty. \tag{26}$$

[10] I. Barbalat, Rev. Math. Pures Appl. **4**, 267 (1959).

[11] G. S. Lellouche, J. Nucl. Energy **21**, 519 (1967).

[12] The condition (26) is necessary and sufficient[13] for the linear functional $\delta K_f[z, t]$ in Eq. (25) to be bound for all t and for all bounded functions $z(t)$ in $(-\infty, +\infty)$.

[13] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill Book Co., New York, 1965).

It can be further proved[6] that condition (ii) of Theorem 1, which in this case reduces to

$$\int_{-\infty}^{t} dt' \int_{-\infty}^{t'} dt'' G(t' - t'')y(t')y(t'') \leq 0, \qquad (27)$$

is satisfied for all test functions $\{y(t)\}$ if

$$R_e[\bar{G}(i\omega)] \leq 0 \qquad (28)$$

holds[7] for all ω in $(0, \infty)$; $\bar{G}(i\omega)$ is the Laplace transform of $G(t)$ with $i\omega$ as the Laplace variable. Hence the reactor response is bounded if the initial perturbation is a physically admissible initial curve, $z(0)$ and $z_i(0)$ are bounded, and relations (26) and (28) hold.

For asymptotic stability we first show the uniform continuity of $\delta K_f[v, t]$ for large t and for all the functions $\{v(t)\}$ which are continuous, bounded, and have bounded first derivatives for $t \geq 0$.

For $t \leq 0$, $v(t)$ is a physically admissible initial curve and thus has bounded left and right derivatives, though it might contain jump discontinuities. Let these discontinuities occur at $-T_1, -T_2, -T_3, \cdots, -T_N$ with $b_1, \cdots, b_N$ as respective jumps. Then, by differentiating $\delta \dot{V}_f[v, t]$ with respect to t, it can be shown that

$$|\delta \dot{K}_f[v, t]| \leq M \int_0^{\infty} |G(t)|\, dt + \sum_{i=1}^{N} |b_i|\, |G(t + T_i)|, \qquad (29)$$

where M is the maximum of $\dot{v}(t)$ for t in the entire domain $(-t_0, +\infty)$. The first term in the right-hand side of relation (29) is bounded by virtue of relation (26). The second term vanishes as $t \to \infty$ as required by relation (26). Hence, $|\delta \dot{K}_f[v, t]|$ is bounded and $\delta V_f[v, t]$ is uniformly continuous for sufficiently large times. We note that the continuity of $G(t)$ is not assumed in these discussions; it may have jump discontinuities and it may even contain delta functions in finite time intervals.

Next we have to find conditions that will ensure that

$$\lim_{t \to \infty} \int_0^{\infty} G(u)\omega(t - u) = 0 \qquad (30a)$$

implies that

$$\lim_{t \to \infty} \omega(t) = 0 \qquad (30b)$$

for all test functions $\{\omega(t)\}$ [cf. condition (ii) of Theorem 2]. These conditions immediately follow from Wiener's theorem (Pitts form),[14] which states that

$$\lim_{t \to \infty} \int_{-\infty}^{+\infty} G(t - u)\omega(u)\, du = A \int_{-\infty}^{+\infty} G(t)\, dt \qquad (31a)$$

implies that

$$\lim_{t \to \infty} \omega(t) = A \qquad (31b)$$

if:

(i) $G(t)$ is absolute integrable in $(-\infty, +\infty)$ and its Fourier transform $\bar{G}(i\omega)$ does not vanish anywhere;

(ii) $\omega(t)$ is bounded and has a derivative which remains greater than a negative constant.

The theorem is applicable since Eq. (30a) can be written in the form of Eq. (31a) [$G(u) = 0$ for $u \leq 0$] and any test function of the subset $\{\omega(t)\}$ is bounded and has bounded first derivatives. We thus obtain the following criterion from Eqs. (30) and (31). If, in addition to relations (26) and (28), $z(0)$ and $z_i(0)$ are bounded and we can demonstrate that

$$\bar{G}(i\omega) \neq 0, \quad -\infty < \omega < +\infty, \qquad (32)$$

then a reactor with linear feedback is asymptotically stable for all physically admissible initial curves. It is clear that Eqs. (28) and (32) will always be satisfied if we require that $R_e[\bar{G}(i\omega)]$ and $I_m[\bar{G}(i\omega)]$ do not vanish at the same frequency. We also note that if the equality sign in Eq. (28) is removed, then $\bar{G}(i\omega)$ cannot vanish at any frequency and Eq. (32) is always satisfied. Hence the condition

$$R_e[\bar{G}(i\omega)] < 0$$

is sufficient for asymptotic stability. This is the well-known criterion of Welton.

It may be noted that Wiener's theorem does not make explicit use of the fact that in our case $\lim \dot{\omega}(t)$, as $t \to \infty$, approaches 0. By exploiting this property, we can replace the condition (32) by a different condition which may be more easily applicable in certain specific cases of the feedback kernel $G(t)$. We first observe that, since $\omega(t)$ is bounded,

$$\left| \int_{-\infty}^{T} G(t - u)[\omega(u) - \omega(t)]\, du \right| \leq 2M \int_{t-T}^{\infty} |G(u)|\, du, \qquad (33)$$

where T is finite and M is the maximum of $\omega(t)$ in $(-t_0, T)$. The right side of relation (33) vanishes as $t \to \infty$ since $G(u)$ is absolutely integrable. Hence,

$$\lim_{t \to \infty} \int_{-\infty}^{T} G(t - u)[\omega(u) - \omega(t)]\, du = 0. \qquad (34)$$

Moreover, by expanding $\omega(t)$ as $\omega(t) = \omega(u) + (t - u)\dot{\omega}[u + \theta(u)]$, where $u \leq \theta(u) \leq t$, and choosing $T(\epsilon)$ for a given $\epsilon > 0$, no matter how small, such

[14] D. V. Widder, *The Laplace Transform* (Princeton University Press, Princeton, N.J., 1946).

that $\dot{\omega}(u) < \epsilon$ for $u > T(\epsilon)$, we have

$$\left| \int_T^t G(t-u)[\omega(u) - \omega(t)]\, du \right| \leq \epsilon \int_T^t |G(t-u)|\, |t-u|\, du \leq \epsilon \int_0^\infty |G(u)|\, u\, du. \quad (35)$$

Therefore, if we impose the condition that

$$\int_0^\infty |G(u)|\, u\, du < \infty, \quad (36)$$

then it is observed that

$$\lim_{t\to\infty} \int_T^t G(t-u)[\omega(u) - \omega(t)]\, du = 0. \quad (37)$$

Relations (34) and (37) are sufficient to show that

$$\lim_{t\to\infty} \left[\int_0^\infty G(u)\omega(t-u)\, du - \omega(t) \int_0^\infty G(u)\, du \right] = 0. \quad (38)$$

It can thus be concluded that, if the condition (36) is assumed to hold, then

$$\int_0^\infty G(u)\omega(t-u)\, du$$

behaves as $\omega(t)$ for large values of t and, therefore, by virtue of Eq. (30a),

$$\lim_{t\to\infty} \omega(t) = 0.$$

The condition (36) may be observed to be very relaxed. It is, for example, always satisfied when $G(t)$ can be expressed as a sum of exponential terms with negative exponents, such as in the case when the feedback can be described by a set of coupled linear differential equations.

B. Nonlinear Feedback

In order to illustrate the application of stability theorems to reactors with nonlinear feedback, we consider a xenon-controlled reactor with flux reactivity coefficient for which $\delta K_f[z, t]$ is quadratic and has the form[7]

$$\delta K_f[z,t] = \int_{-\infty}^t du G_1(t-u)z(u) + \int_{-\infty}^t du_1 \int_{-\infty}^t du_2 G_2(t-u_1, t-u_2)z(u_1)z(u_2), \quad (39a)$$

where

$$G_1(t) = A_1\delta(t) + K(t), \quad (39b)$$

$$K(t) = A_2 e^{-\lambda_x t} + A_3 e^{-\lambda t} + A_4 e^{-\lambda_1 t}, \quad (39c)$$

and

$$G_2(t) = A_5 \delta(t_1 - t_2) e^{-\lambda_x t}. \quad (39d)$$

$A_1, A_2, \cdots, A_5$ are constants (defined in Ref. 7) depending upon various reactor parameters, and λ_x, λ_1, λ are decay constants of ^{135}Xe, ^{135}I, and one group delayed neutron precursor concentration, respectively. $G_1(t)$ and $G_2(t_1, t_2)$ may be noted to be absolutely integrable. The reactor response $z(t)$ to any initial perturbation belonging to the class of physically admissible initial curves and for finite $z(0)$ and $z_i(0)$ is bounded if

$$\int_{-\infty}^t \delta K_f[y, t']y(t')\, dt' \leq 0 \quad (40)$$

[condition (ii) of Theorem 1], for all continuous test functions belonging to the set $\{y, (t)\}$. The condition (40), which was derived on the basis of energy considerations in Ref. 5 as a sufficient condition for asymptotic stability, was investigated for combined xenon and temperature feedback [cf. Eq. (39)] in Refs. 7 and 11, and was shown to lead to the following criterion: if the condition

$$R_e[\bar{K}(i\omega)] + A_1 - (A_5/\lambda_x)P_0 \leq 0, \quad (41)$$

where $\bar{K}(i\omega)$ is the Fourier transform of $K(t)$ [cf. Eq. (39c)] and P_0 is the equilibrium power level (cf. footnote 1), is satisfied for all real ω, then the reactor is asymptotically stable. We now demonstrate that conditions (i) and (ii) of Theorem 2 are satisfied without any additional restrictions upon feedback functional. By substituting (39b) and (39d) in (39a) and replacing $z(t)$ by the test function $v(t)$, we get

$$\delta K_f[v,t] = A_1 v(t) + \int_{-\infty}^t du v(u)(A_2 e^{-\lambda_x(t-u)} + A_3 e^{-\lambda(t-u)} A_4 e^{-\lambda_1(t-u)}) + A_5 \int_{-\infty}^t du v^2(u) e^{-\lambda_x(t-u)}. \quad (42)$$

Also, by differentiating $\delta K_f[v, t]$ with respect to t, we obtain

$$\begin{aligned} \delta \dot{K}_f[v,t] &= A_1\dot{v}(t) + v(t)[A_2 + A_3 + A_4 + A_5 v(t)] \\ &\quad - \int_{-\infty}^t du v(u)(A_2\lambda_x e^{-\lambda_x(t-u)} \\ &\quad + A_3\lambda e^{-\lambda(t-u)} + A_4\lambda_1 e^{-\lambda_1(t-u)}) \\ &\quad - A_5\lambda_x \int_{-\infty}^t du v^2(u) e^{-\lambda_x(t-u)}, \quad t > 0, \\ &\leq |A_1|\, M_1 + 2\,|A_2 + A_3 + A_4|\, M_2 \\ &\quad + 2\,|A_5|\, M_2^2, \quad t > 0, \end{aligned} \quad (43)$$

where M_1, M_2 are upper bounds of $\dot{v}(t)$ and $v(t)$. Hence, $\delta\dot{K}_f[v, t]$ is bounded, and $\delta K_f[v, t]$ is uniformly continuous for $t > 0$.

For condition (ii) of Theorem 2, we note that

$$\int_0^\infty du\omega(t-u)A_j e^{-\alpha u}, \quad j=2,3,4, \quad \alpha=\lambda_x,\lambda,\lambda_1,$$

and

$$\int_0^\infty du\omega^2(t-u)A_5 e^{-\lambda_x u}$$

behave as $(A_j/\alpha)\omega(t)$ and $(A_5/\lambda_x)\omega^2(t)$ as $t\to\infty$. This can be shown by following a similar procedure as that used in the relations (33)–(38). Hence,

$$\lim_{t\to\infty}\delta K_f[\omega,t]=0$$

implies that

$$\lim_{t\to\infty}\omega(t)\left(A_1+\frac{A_2}{\lambda_x}+\frac{A_3}{\lambda}+\frac{A_4}{\lambda_1}+\frac{A_5}{\lambda_x}\omega(t)\right)=0. \quad (44)$$

From Eq. (44) it is obvious that as $t\to\infty$, $\omega(t)$ either approaches 0 or

$$-\frac{\lambda_x}{A_5}\left(A_1+\frac{A_2}{\lambda_x}+\frac{A_3}{\lambda}+\frac{A_4}{\lambda_1}\right). \quad (45)$$

Substitution of the values of $A_1,\cdots,A_5$ and $\lambda_1,\lambda,\lambda_x$ (cf. Ref. 7) in Eq. (45) gives a constant which is always less than -1. But this value is not permissible by definition (cf. Footnote 1). Hence,

$$\lim_{t\to\infty}\delta K_f[\omega,t]=0$$

implies that

$$\lim_{t\to\infty}\omega(t)=0.$$

VI. DISCUSSION

In this analysis, sufficient criteria for the asymptotic stability of a reactor with arbitrary feedback have been obtained. The approach that is followed is similar to Liapunov's technique. However, since we are dealing with functionals instead of functions, just finding a positive definite Liapunov function with a negative first derivative is not enough, as was pointed out by Krasovskii.[4] We note that the most important condition on the feedback functional is condition (ii) of Theorem 1. This condition was obtained as a sufficient criterion for asymptotic stability in Ref. 5 on the basis of considerations of energy dissipation in passive networks. The other conditions in Theorems 1 and 2 are quite mild in nature and are expected to be satisfied in actual physical systems, as was demonstrated in applications of Theorems 1 and 2 in the case of a xenon-controlled reactor with flux reactivity coefficient.[15]

[15] Pasquantonia and Kappel have recently shown[16] that the condition in Ref. 5 is sufficient for asymptotic stability using Hale's theorem.[17]

[16] F. Di Pasquantonia and F. Kappel, Energia Nucl. (Milan) **15**, 761 (1968).

[17] J. K. Hale, J. Differential Equations **1**, 452 (1965).

It is instructive to consider the conditions obtained in Theorems 1 and 2 when delayed neutrons are ignored, i.e., a_i and h_i are identically zero. From the proof of Theorem 2 we observe that we cannot assert asymptotic stability of the reactor when $a_i=h_i\equiv 0$ because condition (ii) of Theorem 1 guarantees only the boundedness of the solutions. It is also interesting to compare Theorem 1 with a criterion obtained by Corduneanu[18] which reads as follows:

The integral equation

$$\sigma(t)=f(t)+\int_0^t l(t-z)\varphi[\sigma(z)]\,dz \quad (46)$$

has at least one solution $\sigma(t)$, defined for $t\geq 0$, which satisfies

$$\lim_{t\to\infty}\sigma(t)=0,$$

if $f(t)$, $l(t)$, and $\varphi(\sigma)$ are real functions satisfying the conditions:

(i) $f(t)$ is defined for $t\geq 0$ and $\dot{f}(t),\ddot{f}(t)\in L_1(0,\infty)$;

(ii)

$$l(t)=j(t)-\rho, \quad (47)$$

where $\rho>0$, and $j(t)$ is defined for $t\geq 0$ and

$$j(t),\frac{d}{dt}j\in L_1(0,\infty)\cap L_2(0,\infty);$$

(iii) $\varphi(\sigma)$ is continuous for all real σ and satisfies

$$\sigma\varphi(\sigma)>0 \quad (\sigma\neq 0); \quad (48)$$

(iv) There exists a $q\geq 0$ such that

$$R_e[(1+i\omega q)L(i\omega)]\leq 0 \quad (\omega\neq 0), \quad (49)$$

where

$$L(i\omega)=\int_0^\infty j(t)e^{-i\omega t}\,dt-\frac{\rho}{i\omega}.$$

Equation (46) can be reduced to the point reactor kinetic equation without delayed neutrons and with a linear feedback functional [cf. Eq. (1)]

$$\dot{z}=(1+z)\int_{-\infty}^t G(t-u)z(u)\,du \quad (50)$$

by defining

$$\sigma(t)=\ln(1+z), \quad (51a)$$

$$\varphi(\sigma)=e^\sigma-1, \quad (51b)$$

$$f(t)=\int_{-\infty}^0 du\,G(t-u)z(u), \quad (51c)$$

$$\frac{dj}{dt}=G(t), \quad (51d)$$

[18] M. C. Corduneanu, C. R. Acad. Sci. (Paris) **256**, 3564 (1963).

and

$$l(0) = j(0) - \rho = 0. \tag{51e}$$

Integrating Eq. (51d) on (t, ∞), we get

$$f(t) = -\int_t^\infty G(u)\, du. \tag{52}$$

The condition (49) thus reduces to[19]

$$R_e[(1 + i\omega q)\bar{G}(i\omega)/i\omega] \leq 0. \tag{53}$$

Hence, according to this theorem, a reactor with linear feedback is asymptotically stable in the absence of delayed neutrons if a $q \geq 0$ exists such that Eq. (53) holds and $f(t)$ and $j(t)$ satisfy conditions (i) and (ii) of Corduneanu's criterion. The conditions of Theorem 1, on the other hand, ensure only the boundedness of reactor response in the absence of delayed neutrons, although the feedback functional may be quite arbitrary and not specifically linear. If a $q \geq 0$ cannot be found such that Eq. (53) is satisfied, then Corduneanu's theorem is noncommital. This happens to be the case for a circulating fuel reactor where $\bar{G}(i\omega)$ is given by[20]

$$\bar{G}(i\omega) = (\alpha\eta/\omega^2\theta)(1 - i\omega\theta - \cos\omega\theta + i \sin\omega\theta). \tag{54}$$

Here θ is the fuel transit time in the core, α is the temperature reactivity coefficient, and η is the heat capacity of the reactor. Application of the relation (53) reduces Eq. (54) to

$$q(1 - \cos x) + \sin x/x \geq 1, \quad x \neq 0, \tag{55}$$

where we have used the fact that $\alpha < 0$ and substituted x for $\omega\theta$. Clearly Eq. (55) cannot be satisfied for all $x > 0$ for any choice of q. However, application of Theorem 1 [cf. Eq. (28)] reduces Eq. (54) to

$$1 - \cos x \geq 0. \tag{56}$$

Equation (56) ascertains the boundedness of circulating fuel reactor response to perturbations belonging to physically admissible class of initial curves.

[19] The condition (ii) of Carduneanu's theorem requires $\rho > 0$, which by virtue of (51e) implies $j(0) > 0$. The latter is the condition for the existence of a finite equilibrium power level. The condition (53) alone can not guarantee asymptotic stability.

[20] A. Z. Akcasu and L. M. Shotkin, Nucl. Sci. Eng. **28**, 72 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Bases for Irreducible Representations of the Unitary Group in the Symplectic Group Chain

V. Syamala Devi
Andhra University, Waltair, India

(Received 1 March 1968; Revised Manuscript Received 9 January 1969)

A method of obtaining the highest weight polynomials of irreducible representations (λ) of Sp_{2n} occurring in a reduction of an irreducible representation (k) of U_{2n} is described. The highest weight polynomials of equivalent representations (λ) are labeled by means of parameters which occur naturally from the Littlewood's theorem to determine the branching rules for the representations of the unitary group with respect to the symplectic subgroup, when supplemented by modification rules. The results are given explicitly for $U_4 \supset Sp_4$ and $U_6 \supset Sp_6$, the former being a canonical chain and the latter a noncanonical chain.

INTRODUCTION

The aim of this paper is to obtain a basis for any given irreducible representation (IR) of the unitary group U_{2n} in $2n$ dimensions such that, with respect to that basis, the symplectic subgroup Sp_{2n} is explicitly reduced into blocks. It is enough to find the highest weight vectors of the various IR's of Sp_{2n} occurring in the reduction of a given IR of U_{2n} since all the other basis vectors in an irreducible representation space of Sp_{2n} can be obtained by applying polynomials of the lowering generators of Sp_{2n} on the highest weight vector because linearly independent highest weight vectors give rise to linearly independent spaces. (By the reduction of an IR of U_{2n} with respect to Sp_{2n}, we mean the reduction with respect to Sp_{2n}, of the representation of Sp_{2n}, obtained by considering the restriction to Sp_{2n} of the given IR of U_{2n}.) We use a theorem of Littlewood[1] and the modification rules of

[1] D. E. Littlewood, *Theory of Group Characters* (Clarendon Press, Oxford, 1950), p. 295.

and

$$l(0) = j(0) - \rho = 0. \tag{51e}$$

Integrating Eq. (51d) on (t, ∞), we get

$$f(t) = -\int_t^\infty G(u)\, du. \tag{52}$$

The condition (49) thus reduces to[19]

$$R_e[(1 + i\omega q)\bar{G}(i\omega)/i\omega] \leq 0. \tag{53}$$

Hence, according to this theorem, a reactor with linear feedback is asymptotically stable in the absence of delayed neutrons if a $q \geq 0$ exists such that Eq. (53) holds and $f(t)$ and $j(t)$ satisfy conditions (i) and (ii) of Corduneanu's criterion. The conditions of Theorem 1, on the other hand, ensure only the boundedness of reactor response in the absence of delayed neutrons, although the feedback functional may be quite arbitrary and not specifically linear. If a $q \geq 0$ cannot be found such that Eq. (53) is satisfied, then Corduneanu's theorem is noncommital. This happens to be the case for a circulating fuel reactor where $\bar{G}(i\omega)$ is given by[20]

$$\bar{G}(i\omega) = (\alpha\eta/\omega^2\theta)(1 - i\omega\theta - \cos\omega\theta + i \sin\omega\theta). \tag{54}$$

Here θ is the fuel transit time in the core, α is the temperature reactivity coefficient, and η is the heat capacity of the reactor. Application of the relation (53) reduces Eq. (54) to

$$q(1 - \cos x) + \sin x/x \geq 1, \quad x \neq 0, \tag{55}$$

where we have used the fact that $\alpha < 0$ and substituted x for $\omega\theta$. Clearly Eq. (55) cannot be satisfied for all $x > 0$ for any choice of q. However, application of Theorem 1 [cf. Eq. (28)] reduces Eq. (54) to

$$1 - \cos x \geq 0. \tag{56}$$

Equation (56) ascertains the boundedness of circulating fuel reactor response to perturbations belonging to physically admissible class of initial curves.

[19] The condition (ii) of Carduneanu's theorem requires $\rho > 0$, which by virtue of (51e) implies $j(0) > 0$. The latter is the condition for the existence of a finite equilibrium power level. The condition (53) alone can not guarantee asymptotic stability.

[20] A. Z. Akcasu and L. M. Shotkin, Nucl. Sci. Eng. **28**, 72 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Bases for Irreducible Representations of the Unitary Group in the Symplectic Group Chain

V. Syamala Devi
Andhra University, Waltair, India

(Received 1 March 1968; Revised Manuscript Received 9 January 1969)

A method of obtaining the highest weight polynomials of irreducible representations (λ) of Sp_{2n} occurring in a reduction of an irreducible representation (k) of U_{2n} is described. The highest weight polynomials of equivalent representations (λ) are labeled by means of parameters which occur naturally from the Littlewood's theorem to determine the branching rules for the representations of the unitary group with respect to the symplectic subgroup, when supplemented by modification rules. The results are given explicitly for $U_4 \supset Sp_4$ and $U_6 \supset Sp_6$, the former being a canonical chain and the latter a noncanonical chain.

INTRODUCTION

The aim of this paper is to obtain a basis for any given irreducible representation (IR) of the unitary group U_{2n} in $2n$ dimensions such that, with respect to that basis, the symplectic subgroup Sp_{2n} is explicitly reduced into blocks. It is enough to find the highest weight vectors of the various IR's of Sp_{2n} occurring in the reduction of a given IR of U_{2n} since all the other basis vectors in an irreducible representation space of Sp_{2n} can be obtained by applying polynomials of the lowering generators of Sp_{2n} on the highest weight vector because linearly independent highest weight vectors give rise to linearly independent spaces. (By the reduction of an IR of U_{2n} with respect to Sp_{2n}, we mean the reduction with respect to Sp_{2n}, of the representation of Sp_{2n}, obtained by considering the restriction to Sp_{2n} of the given IR of U_{2n}.) We use a theorem of Littlewood[1] and the modification rules of

[1] D. E. Littlewood, *Theory of Group Characters* (Clarendon Press, Oxford, 1950), p. 295.

Newell[2] for the reduction of an IR $(k) = (k_1, k_2, \cdots)$ of U_{2n} with respect to Sp_{2n}, and we determine the highest-weight vectors of the various IR's $(\lambda) = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ of Sp_{2n} occurring in the reduction, in one-to-one association with Littlewood diagrams.

In the case of the chains[3,4] $U_4 \supset U_2 \times U_2$ and $U_3 \supset R_3$, the highest weight polynomials (h.w.p.) of the subgroups $U_2 \times U_2$ and R_3 contained in the carrier spaces of given IR's of U_4 and U_3, respectively, were obtained using Littlewood's theorems for the reduction of an IR of U_4 with respect to $U_2 \times U_2$ and an IR of U_3 with respect to R_3. We describe a similar procedure in this note for obtaining the h.w.p. of Sp_4 contained in the carrier space of a given IR of U_4 and similarly, the h.w.p. of Sp_6 belonging to the space of a given IR of U_6.

1. POLYNOMIAL BASES FOR U_{2n} AND Sp_{2n}

The infinitesimal generators of U_{2n} and Sp_{2n} can be realized in terms of boson creation operators a_{js}^{+} and annihilation operators a_s^j as follows[5]:

$$C_j^{j'} = \sum_{s=1}^{2n} a_{js}^{+} a_s^{j'}, \quad j, j' = 1, \cdots, 2n, \tag{1.1}$$

have the same commutation rules as the generators of U_{2n}. Writing

$$H_i = C_i^i - C_{i+n}^{i+n}, \quad E_{ij} = C_i^j - C_{j+n}^{i+n}, \quad F_{ii} = C_i^{i+n},$$

with $i, j = 1, \cdots, n$, and

$$G_{ij} = C_i^{j+n} + C_j^{i+n}, \quad G'_{ij} = C_{j+n}^{i} + C_{i+n}^{j}, \quad i < j = 1, \cdots, n, \tag{1.2}$$

H_i, E_{ij}, F_{ii}, G_{ij}, F'_{ii}, and G'_{ij} provide a set of generators for the subgroup Sp_{2n} of U_{2n} and are in Cartan's canonical form. The generators E_{ij} $(i<j)$, F_{ii}, G_{ij} are positive-root generators and E_{ij} $(i>j)$, F'_{ii}, and G'_{ij} the corresponding negative-root generators. The bases for IR's of (k) of U_{2n} and (λ) of Sp_{2n} can be written as polynomials in a_{js}^{+} operating on the vacuum ket.[5,6] Our aim is to find the highest weight polynomials of any IR (λ) of Sp_{2n} contained in an IR (k) of U_{2n}. A polynomial belongs to the carrier space of the IR (k) of U_{2n} if and only if it satisfies[5]

$$C_{ss}P = k_s P, \quad C_{ss'}P = 0, \quad s < s' = 1, \cdots, 2n, \tag{1.3}$$

and it is of highest weight $(\lambda_1, \lambda_2, \cdots)$ with respect to Sp_{2n} if and only if it satisfies

$$H_i P = \lambda_i P, \quad E_{\alpha_j} P = 0, \quad i, j = 1, \cdots, n, \tag{1.4}$$

where E_{α_j} are generators corresponding to a simple system of roots[7] and

$$C_{ss'} = \sum_{j=1}^{2n} a_{js}^{+} a_{s'}^{j}.$$

2. REDUCTION OF IR's OF U_{2n} WITH RESPECT TO Sp_{2n}

The reduction of the IR (k) of U_{2n} into IR's $(\lambda) = (\lambda_1, \lambda_2, \cdots)$ of the symplectic subgroup is determined by a theorem of Littlewood.[1] Consider the IR's of U_{2n} corresponding to partitions $(\lambda_1, \lambda_2, \cdots)$ and (β) where (β) is a partition whose parts occur an even number of times. If in the reduction to IR's of U_{2n} the product representation $(\lambda) \times (\beta)$ contains (k) a certain number of times which we denote by $g_{\lambda\beta k}$, then under the limitation to Sp_{2n}, the IR (k) of U_{2n} breaks up into IR's (λ) of Sp_{2n} according to the formula

$$\{k\} = \sum_{\{\beta\}} g_{\lambda\beta k} \langle\lambda\rangle, \tag{2.1}$$

the summation being over all S functions $\{\beta\}$ whose parts occur an even number of times. In using this formula, the partitions (λ) may sometimes contain more than n parts. These have to be interpreted according to modification rules. We note that

$$\begin{aligned} \langle\lambda_1, \cdots, \lambda_n, 1, 1\rangle &= -\langle\lambda_1, \cdots, \lambda_n\rangle, \\ \langle\lambda_1, \cdots, \lambda_{n-1}, 1, 1, 1, 1\rangle &= -\langle\lambda_1, \cdots, \lambda_{n-1}, 0\rangle, \end{aligned} \tag{2.2}$$

where $\langle\lambda_1, \lambda_2, \cdots\rangle$ denotes the character of the symplectic group in the IR corresponding to the partition $(\lambda_1, \lambda_2, \cdots)$. All other nonstandard symbols having not more than $2n$ parts are to be dropped.

Though the theorem as stated originally concerns itself with partitions (k) into not more than n parts, it can be extended to partitions (k) with not more than $2n$ parts by the use of modification rules (2.2). The coefficients $g_{\lambda\beta k}$ in (2.1) are determined by the Littlewood rules[8] to write the product of S functions $\{\lambda\} \times \{\beta\}$ as a sum of S functions $\{k\}$. We may restrict ourselves to IR's corresponding to partitions (k) of U_{2n} where the last part is set equal to zero, since the symplectic matrices are unimodular.

Chain $U_4 \supset Sp_4$: The Eqs. (1.3) and (1.4) in the case of $U_4 \supset Sp_4$ are

$$\begin{aligned} &C_{11}P = k_1 P, \quad C_{22}P = k_2 P, \quad C_{33}P = k_3 P, \\ &C_{12}P = C_{13}P = C_{23}P = 0, \quad (C_1^1 - C_3^3)P = \lambda_1 P, \\ &(C_2^2 - C_4^4)P = \lambda_2 P, \quad (C_1^2 - C_4^3)P = C_2^4 P = 0 \end{aligned}$$

[called (2.3)–(2.12), respectively].

[2] M. J. Newell, Proc. Roy. Irish Acad. **54**, 153 (1951).
[3] V. Syamala Devi and T. Venkatarayudu, J. Math. Phys. **9**, 1057 (1968).
[4] M. Moshinsky and V. Syamala Devi, J. Math. Phys. **10**, 455 (1969).
[5] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[6] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[7] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1962), p. 120.
[8] Ref. 1, p. 94.

3. ELEMENTARY LITTLEWOOD DIAGRAMS

Let us mark the boxes of the first row of the Young frame of the partition (λ_1, λ_2) by the symbol x and the boxes of its second row by the symbol y. Let us mark the boxes of the first and second rows of the Young frame of the partition (β) by the symbols a and b, respectively. Then, corresponding to each of the following diagrams built by the application of the symbols of (β) on the Young frame of (λ_1, λ_2) according to Littlewood's rules[8] and having rows of lengths k_1, k_2, k_3, respectively, from top to bottom, we have an IR (λ_1, λ_2) of Sp_4 which is contained in the IR (k_1, k_2, k_3) of U_4. In the diagram

$$\begin{array}{l} \overbrace{x \quad \cdots \quad x}^{\lambda_1}\ \overbrace{a \cdots a}^{x_{11}} \\ \overbrace{y \cdots y}^{\lambda_2}\ \overbrace{a \cdots a}^{x_{12}}\ \overbrace{b \cdots b}^{x_{22}} \\ \overbrace{b \cdots b}^{x_{23}}, \end{array} \tag{3.1}$$

we have

$$\lambda_1 \geq \lambda_2 + x_{12} \geq x_{23}, \quad x_{11} \geq x_{22}. \tag{3.2}$$

Since the lengths of the first, second, and third rows are, respectively, k_1, k_2, k_3, we have

$$\lambda_1 + x_{11} = k_1, \quad \lambda_2 + x_{12} + x_{22} = k_2, \quad x_{23} = k_3.$$

Also, as the total number of a's is equal to the total number of b's,

$$x_{11} + x_{12} = x_{22} + x_{23}. \tag{3.3}$$

These four equations determine all the four quantities $x_{11}, x_{12}, x_{22}, x_{23}$ uniquely in terms of k_1, k_2, k_3, λ_1, and λ_2. Hence each IR (λ_1, λ_2) of Sp_4 contained in the IR (k_1, k_2, k_3) of U_4 is contained only once.

We shift all the symbols in the second and third rows of (3.1) to the right so that (i) no symbol falls below a gap in the row above and (ii) no two identical symbols fall in a column. The rearranged diagram is

$$\begin{array}{l} \overbrace{x \quad \cdots \quad x}^{\lambda_1}\ \overbrace{a \cdots a}^{x_{11}} \\ \overbrace{y \cdots y}^{\lambda_2}\ \overbrace{a \cdots a}^{x_{12}}\ \overbrace{b \cdots b}^{x_{22}} \\ \overbrace{b \cdots b.}^{x_{23}} \end{array} \tag{3.1'}$$

We divide (3.1') columnwise into elementary permissible diagrams (e.p.d.). A diagram is said to be permissible if it contains the same number of a's and b's and if the a's and b's occur in it in lattice order [i.e., if the numbers x_{ij} and λ_1, λ_2 of the diagram satisfy the inequalities (3.2)]. The e.p.d.'s of (3.1') are

$$\begin{array}{llll} a, & xa, & x, & x, \ x \\ b & y & a & y \\ & b & b. & \end{array}$$

4. CONSTRUCTION OF THE h.w.p.

We associate with each e.p.d. a polynomial satisfying all the Eqs. (2.3)–(2.12) where the values of k_1, k_2, k_3 are those corresponding to the e.p.d. The unitary group with generators (1.1) and its symplectic subgroup with generators (1.2) form a chain for which the branching rules are furnished by the above theorem of Littlewood and the modification rules, as can be seen by applying the branching rules to the representation of lowest dimension of U_{2n} with generators (1.1) [see Ref. 9 for the definition of equivalent embeddings of a semisimple Lie algebra $\mathfrak{L}'$ in a semisimple Lie algebra $\mathfrak{L}$ and a criterion for the equivalence of two embeddings]. Hence corresponding to each e.p.d. we get a solution of equations (2.3–2.12) which is a polynomial in a^+_{js}, the numbers k_1, k_2, k_3, λ_1, λ_2 having values corresponding to the e.p.d. The polynomial corresponding to an e.p.d is found as follows. The diagram of the e.p.d is to be filled by symbols 1, 2, 3, 4 so that the numbers in the rows are nondecreasing and those in the columns are increasing. The number of 1's minus the number of 3's is the same as the number of x's in the e.p.d. The number of 2's minus the number of 4's is the same as the number of y's in the e.p.d. In general, we get several such lexical tableaux. Corresponding to each tableau we write down the polynomial $\mathbf{\Pi}\Delta^{1\cdots r}_{s_1\cdots s_r}$, where $s_1, \cdots, s_r$ are the symbols appearing in a column of the tableau, and $\mathbf{\Pi}$ is the product taken over all the columns of the tableau, and

$$\Delta^{1\cdots r}_{s_1\cdots s_r} = \det \begin{vmatrix} a^+_{s_{11}} & \cdots & a^+_{s_{1r}} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a^+_{s_r1} & \cdots & a^+_{s_rr} \end{vmatrix}.$$

Any such polynomial satisfies Eqs. (2.3)–(2.10), $k_1, k_2, k_3, \lambda_1, \lambda_2$ having the values corresponding to the e.p.d. We then find linear combinations of these polynomials so as to satisfy Eqs. (2.11) and (2.12). It is found that a unique polynomial solution of Eqs. (2.3)–(2.12) is obtained corresponding to each e.p.d. The elementary polynomial solutions corresponding to the e.p.d.'s are, respectively,

$$s = \Delta^{12}_{13} + \Delta^{12}_{24},$$

$$r = \Delta^1_1\Delta^{123}_{123} + \Delta^1_2\Delta^{123}_{124}, \Delta^{123}_{124}, \Delta^{12}_{12}, \quad \text{and} \quad \Delta^1_1.$$

[9] E. B. Dynkin, Am. Math. Soc. Transl. Ser. 2, 6, 111 (1957).

We associate with the diagram (3.1′) and hence with (3.1), the polynomial which is the product of all the polynomials associated with all the e.p.d.'s into which (3.1′) splits. Thus the polynomial corresponding to (3.1) is

$$P = (\Delta_1^1)^{\lambda_1-k_2+x_{22}}(\Delta_{12}^{12})^{\lambda_2+x_{12}-k_3}(b)^{x_{23}-x_{12}}(\Delta_{124}^{123})^{x_{12}}(s)^{x_{22}}. \tag{4.1}$$

It is easily seen that P is a solution of Eqs. (2.3) to (2.12).

The polynomials (4.1) corresponding to distinct diagrams (3.1) are linearly independent, as they are common eigenfunctions of the linear differential operators $C_1^1 - C_3^3$, $C_2^2 - C_4^4$, corresponding to distinct sets of eigenvalues (λ_1, λ_2). A full basis of U_4, with respect to which the Sp_4 subgroup is explicitly reduced into blocks, is obtained by applying on the h.w.p. the lowering operators Sp_4, since the lowering operators[10] of Sp_4 are known owing to its local isomorphism with R_5, the rotation group in five dimensions.

The author thanks Professor Moshinsky for pointing out that the h.w.p. of Sp_4 in the chain $U_4 \supset Sp_4$ could also be obtained easily by purely algebraic methods, since Sp_4 happens to be a canonical subgroup of U_4. Also, the importance of the technique developed here lies mainly in its application to noncanonical chains where the algebraic procedure becomes tedious (almost impossible). (G is said to be a canonical subgroup of U_n if, in the reduction of any IR of U_n (when restricted to G) with respect to G, each IR of G occurs only once; G is said to be noncanonical if otherwise, i.e., if G is a noncanonical subgroup of U_n, equivalent IR's of G may be contained more than once in the representation of G induced by an IR of U_n. We say that the pair U_n and G form a canonical chain or a noncanonical chain G is a canonical or noncanonical subgroup of U_n, respectively.)

Chain $U_6 \supset Sp_6$: In the following we will determine the h.w.p. of the various IR's $(\lambda_1, \lambda_2, \lambda_3)$ of Sp_6 occurring in the reduction of an IR $(k_1, \cdots, k_5)$ of U_6. A polynomial belonging to the carrier space of the IR $(k) = (k_1, \cdots, k_5)$ is of highest weight $(\lambda_1, \lambda_2, \lambda_3)$ with respect to Sp_6 if and only if it satisfies

$$C_{jj}P = k_jP, \quad C_{jj'}P = 0, \quad j < j' = 1, \cdots, 5, \tag{4.2}$$

$$(C_1^1 - C_4^4)P = \lambda_1P, \quad (C_2^2 - C_5^5)P = \lambda_2P,$$
$$(C_3^3 - C_6^6)P = \lambda_3P, \tag{4.3}$$
$$(C_1^2 - C_5^4)P = (C_2^3 - C_6^5)P = C_3^6P = 0.$$

[10] Sing Chin Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).

5. ELEMENTARY LITTLEWOOD DIAGRAMS

To determine the IR's $(\lambda_1, \lambda_2, \lambda_3)$ of Sp_6 contained in an IR (k) of U_6, we take the Young diagram of the partition of Sp_6 and mark its boxes by the symbol x. Then we apply to it, according to Littlewood's rules, the symbols of the Young diagram of the partition $(\beta) = (p, p, q, q)$, the first, second, third, and fourth row boxes of which we mark by symbols a, b, c, and d, respectively. We then obtain the following diagrams which have at most five rows of lengths $k_1, \cdots, k_5$ from top to bottom, respectively; we denote by x_{1j}, x_{2j}, x_{3j}, and x_{4j} the number of a's, the number of b's, that of c's, and that of d's, in the jth row, respectively, and by the modification rules (2.2) one has $x_{14} = x_{25} = 0$. Thus, a typical diagram is

$$\begin{array}{l} \overbrace{x \cdots x}^{\lambda_1}\ \overbrace{a \cdots a}^{x_{11}} \\ \overbrace{x \cdots x}^{\lambda_2}\ \overbrace{a \cdots a}^{x_{12}}\ \overbrace{b \cdots b}^{x_{22}} \\ \overbrace{x \cdots x}^{\lambda_3}\ \overbrace{a \cdots a}^{x_{13}}\ \overbrace{b \cdots b}^{x_{23}}\ \overbrace{c \cdots c}^{x_{33}} \\ \overbrace{b \cdots b}^{x_{24}}\ \overbrace{c \cdots c}^{x_{34}}\ \overbrace{d \cdots d}^{x_{44}} \\ \overbrace{d \cdots d}^{x_{44}}. \end{array} \tag{5.1}$$

The numbers x_{ij} satisfy

$$\begin{aligned} &x_{11} \geq x_{22} \geq x_{33} \geq x_{44}, \\ &x_{11} + x_{12} \geq x_{22} + x_{23} \geq x_{33} + x_{34}, \\ &\lambda_1 \geq \lambda_2 + x_{12} \geq \lambda_3 + x_{13} + x_{23} \geq x_{24} + x_{34} \geq x_{45}, \\ &\lambda_2 \geq \lambda_3 + x_{13} \geq x_{24}, \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} x_{11} + x_{12} + x_{13} = x_{22} + x_{23} + x_{24} = p, \\ x_{33} + x_{34} = x_{44} + x_{45} = q. \end{aligned} \tag{5.3}$$

Since the lengths of the rows are $k_1, \cdots, k_5$, we have

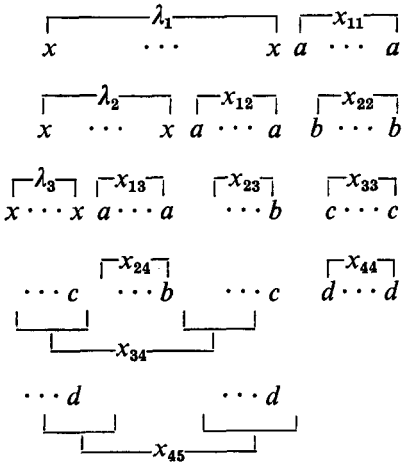
$$\begin{aligned} \lambda_1 + x_{11} = k_1, \quad \lambda_2 + x_{12} + x_{22} = k_2, \\ \lambda_3 + x_{13} + x_{23} + x_{33} = k_3, \quad x_{24} + x_{34} + x_{44} = k_4, \\ x_{45} = k_5. \end{aligned} \tag{5.4}$$

Corresponding to each diagram (5.1) we have an IR (λ) of Sp_6 contained in an IR (k) of U_6. Three independently varying x_{ij} can be chosen, say, x_{22}, x_{44}, and x_{13}, so that all the other x_{ij} can be expressed in terms of them, $k_1, \cdots, k_5$, and λ_1, λ_2, λ_3. Corresponding to distinct sets of values of the three parameters, we

have distinct diagrams (5.1) and hence, the equivalent IR's of Sp_6 contained in the IR (k) of U_6 are labeled uniquely by the three parameters x_{22}, x_{44}, x_{13}.

We rearrange the symbols in (5.1) in all the rows except the topmost one as follows. In the second row we push all the b's to the right so as not to go beyond the a's in the first row. We push all the a's to the right so as not to go beyond the x's in the first row. We then push the x's in the second row as far as the a's. In the third row, we place the a's under the x's in the second row and place the b's as far as possible under the a's in the second row; if all the places under the a's in the second row are filled with b's, then the remaining b's, if any, are placed under the x's; the c's are placed under the b's in the second row. Again, in the fourth row we place the b's as far as possible under the a's. If there is an excess of b's in the fourth row over the a's in the third row, then the excess b's are placed under the x's. Similarly, the c's are placed as far as possible under the b's, and if there is an excess of c's, they are placed under the x's; the d's in the fourth row are placed as far as possible under the c's. Finally, the d's in the fifth row are placed as far as possible under the c's in the fourth row. In the rearranged diagram, the numbers x_{ij} still satisfy the inequalities (5.2) and the Eqs. (5.3) and (5.4).

The rearranged diagram is



(5.1′)

We divide each diagram (5.1′) columnwise into elementary permissible diagrams (e.p.d.). A diagram is said to be permissible if it contains the same number of a's and b's, the same number of c's and d's, and if the symbols a, b, c, d occur in it in lattice order [in other words, if the numbers $\lambda_1, \lambda_2, \lambda_3, x_{ij}$ of the diagram satisfy the inequalities (5.2)]. A diagram is said to be elementary if it cannot be split columnwise into two permissible diagrams. The possible columns of (5.1′) are

$$\begin{array}{cccccccccccccccc} x & x & x & x & x & x & x & x & x & x & x & x & x & a & a & a & a \\ & x & x & x & x & x & x & x & x & a & x & a & a & & b & b & b \\ & & x & x & x & x & a & a & b & b & b & b & & & & c & c \\ & & & c & b & b & b & b & b & c & & & & & & & d \\ & & & & d & d & & d & & d & d & & & & & & \end{array}$$

To find all the possible e.p.d.'s let us for example find all the e.p.d.'s containing the nonpermissible column

$$\begin{array}{c} x \\ x \\ x \\ b \end{array}$$

A permissible diagram containing the column must contain at least one a in the first two rows. Hence the diagram must contain either a column $\frac{x}{a}$ or a column a in order that it is permissible. For, though a in the first two rows may occur when the diagram has any one of the columns

$$\begin{array}{ccccc} a & a & a & x & x \\ b & b & b & a & a \\ & c & c & b & b \\ & & d & & c \\ & & & & d \end{array}$$

it must again have either $\frac{x}{a}$ or the column a in order that it be permissible. Now both the diagrams

$$\begin{array}{cc} x & a \\ x & \\ x & \\ b & \end{array} \quad \text{and} \quad \begin{array}{cc} x & x \\ & x \\ & x \\ & b \end{array}$$

are permissible and elementary. Moreover, by deleting any one of these e.p.d.'s from a permissible diagram, the diagram remains permissible, since in the remaining diagram the inequalities (5.2) and Eqs. (5.3) still hold. Therefore these two are the only e.p.d.'s containing the column

$$\begin{array}{c} x \\ x \\ x \\ b \end{array}$$

Similarly, it can be seen that the following are all the possible e.p.d.'s:

$$e_1 = \begin{matrix} x \\ x \\ x \\ b \end{matrix} \qquad e_2 = \begin{matrix} x \\ x \\ x \\ d \end{matrix} \qquad e_3 = \begin{matrix} x \\ x \\ x \end{matrix} \qquad e_4 = \begin{matrix} x & x \\ x & a \\ x \end{matrix}$$

$$e_5 = \begin{matrix} x & a \\ x \\ x \\ b \end{matrix} \qquad e_6 = \begin{matrix} x & x & a \\ x & a & b \\ x & c \\ b \\ d \end{matrix} \qquad e_7 = \begin{matrix} x & a & a \\ x & b \\ x & c \\ b \\ d \end{matrix} \qquad e_8 = \begin{matrix} x & a \\ x & b \\ x \\ c \\ d \end{matrix}$$

$$e_9 = \begin{matrix} x & a \\ x & b \\ a & c \\ b \\ d \end{matrix} \qquad e_{10} = \begin{matrix} x \\ x \\ a \\ b \end{matrix} \qquad e_{11} = \begin{matrix} x & a \\ x \\ b \\ c \\ d \end{matrix} \qquad e_{12} = \begin{matrix} x \\ a \\ b \\ c \\ d \end{matrix}$$

$$e_{13} = \begin{matrix} x \\ a \\ b \end{matrix} \qquad e_{14} = \begin{matrix} x & a \\ x \\ b \end{matrix} \qquad e_{15} = \begin{matrix} a \\ b \\ c \\ d \end{matrix} \qquad e_{16} = \begin{matrix} a \\ b \end{matrix}$$

The division of (5.1′) into e.p.d.'s is not unique. However, we make it a convention to include a non-permissible column of (5.1′) in an e.p.d. by adjoining to it the nearest possible column as we proceed from left to right in (5.1′); all nonpermissible columns are exhausted one by one from left to right in (5.1′). For example,

$$\begin{matrix} x & x & x & a & a \\ x & x & a & b \\ x & x & c \\ b & b \\ d \end{matrix} \quad \text{is split as} \quad \begin{matrix} x & x, & x & a & a \\ x & a & x & b \\ x & x & c \\ b & b \\ d \end{matrix}$$

and not as

$$\begin{matrix} x & a, & x & x & a \\ x & & x & a & b \\ x & & x & & c \\ b & & b \\ & & d \end{matrix}$$

Then the diagram (5.1′) splits into q_1 e.p.d.'s e_1, q_2 e.p.d.'s e_2 and so on, where

$$q_1 = \lambda_1 - \lambda_2 - x_{12},$$
$$q_2 = \lambda_2 - \lambda_3 - x_{13} - x_{23} + m_1,$$
$$q_3 = \lambda_3 - x_{24} + x_{13} - x_{34} + m_4,$$
$$q_4 = \min (x_{24} - x_{13} - x_{45} + x_{34} + m_5, x_{12} - m_1),$$
$$q_5 = x_{24} - x_{13} - x_{45} + x_{34} + m_5 - q_4,$$
$$q_6 = \min (x_{45} - x_{34} - m_5, x_{12} - m_1 - q_4),$$
$$q_7 = x_{45} - x_{34} - m_5 - q_6,$$
$$q_8 = x_{34} - m_4, \quad q_9 = m_5, \quad q_{10} = x_{13} - m_5,$$
$$q_{11} = m_3, \quad q_{12} = m_2, \quad q_{13} = m_1 - m_2,$$
$$q_{14} = x_{23} - m_1 - m_3, \quad q_{15} = x_{44},$$
$$q_{16} = x_{22} - x_{33} - q_8, \quad \text{where} \quad m_1 = \min (x_{12}, x_{23}),$$
$$m_2 = \min (x_{12}, x_{23}, x_{34}),$$
$$m_3 = \min (x_{23} - m_1, x_{34} - m_2),$$
$$m_4 = \min (x_{23}, x_{34}), \text{ and } m_5 = \min (x_{13}, x_{45} - x_{34}).$$

6. CONSTRUCTION OF THE h.w.p.

Corresponding to each e.p.d. we find a solution of Eqs. (4.2) and (4.3) where the values for $k_1, \cdots, k_5$ and $\lambda_1, \lambda_2, \lambda_3$ are those corresponding to the e.p.d. The h.w.p. corresponding to the above e.p.d. are, respectively,

$$S_1 = (1), \quad S_2 = (12), \quad S_3 = (123),$$
$$S_4 = (12)(1235) + (13)(1236),$$
$$S_5 = (1)(1234) + (2)(1235) + (3)(1236),$$
$$S_6 = (12)(125)(12345) + (13)(136)(12346) - (16)(123)(12345) + (13)(126)(12345) + (13)(125)(12346) + (23)(125)(12356) \cdot + (23)(136)(12356),$$
$$S_7 = (1)(124)(12345) + (1)(134)(12346) + (2)(125)(12345) - (3)(124)(12356) + (3)(126)(12345) + (2)(134)(12356) + (2)(135)(12346) + (3)(136)(12346) + (2)(235)(12356) + (3)(236)(12356) + (5)(123)(12346) - (6)(123)(12345),$$
$$S_8 = (12)(12345) + (23)(12356) + (13)(12346),$$
$$S_9 = (136)(12346) + (236)(12356) + (125)(12346) - (124)(12356),$$
$$S_{10} = (1236), \quad S_{11} = (1)(12346) + (2)(12356),$$
$$S_{12} = (12356), \quad S_{13} = (125) + (136),$$
$$S_{14} = (1)(124) + (2)(125) + (3)(126) - (6)(123),$$
$$S_{15} = (1245) + (2356) + (1346),$$

and

$$S_{16} = (14) + (25) + (36),$$

where $(s_1, \cdots, s_r) = \Delta^{1 \cdots r}_{s_1 \cdots s_r}$. We associate with the diagram (5.1′), hence with the diagram (5.1), the polynomial which is the product of all the polynomials corresponding to the e.p.d.'s into which it splits. Thus

$$P = \prod_i (S_i)^{q_i} \tag{6.1}$$

is the h.w.p. of an IR (λ) of Sp_6 contained in the IR (k) of U_6. We have as many h.w.p.'s given by (6.1) as the number of times the IR (λ) of Sp_6 occurs in the reduction of the IR (k) of U_6. The h.w.p.'s (6.1) of different IR's of Sp_6 contained in the same IR (k) of U_6 are linearly independent as they are eigenfunctions of certain linear differential operators corresponding to distinct sets of eigenvalues $\lambda_1, \lambda_2, \lambda_3$. However, if an IR of Sp_6 occurs in the IR (k) of U_6 more than once, it has to be shown that the corresponding polynomials of highest weight given in (6.1) are linearly independent. This is done in the Appendix. The method can be straightforwardly generalized to $U_{2n} \supset Sp_{2n}$.

ACKNOWLEDGMENT

The author is grateful to Professor M. Moshinsky for suggesting the problem.

APPENDIX: PROOF OF LINEAR INDEPENDENCE OF THE HIGHEST WEIGHT POLYNOMIALS IN THE CASE OF $U_6 \supset Sp_6$

Let s_i be defined as $s_4 = S_4, s_5 = S_5, s_6 = S_1S_2wS_6, s_7 = S_1S_2wS_7, s_8 = S_1wS_8, s_9 = S_1S_2wS_6, s_{10} = S_{10}, s_{11} = wS_{11}, s_{12} = wS_{12}, s_{13} = S_2S_{13}, s_{14} = S_{14}, s_{15} = S_1S_2S_3S_{15}, s_{16} = S_1S_{16}$, where $w = (1234)$. We have the following relations among s_i: $s_6 = s_8s_{13}$, $s_7 = s_8s_{14} - s_3s_{11}s_{16}$, $s_9 = s_{11}s_{13} - s_2s_{12}s_{14}$, $s_{15} = s_4s_{14} - s_3s_{10}s_{16} - s_5s_{13}$, and $s_8s_{10} = s_4s_{11} - s_2s_5s_{12}$. The functions $s_1, s_2, s_3, s_4, s_5, s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{16}$ of a^+_{js}, are functionally independent. When expressed in terms of these, the polynomial P in (6.1) of the IR (λ) of Sp_6 contained in the IR (k) of U_6, has a factor $(s_1)^{\lambda_1-k_2}(s_2)^{\lambda_2-k_3}(s_3)^{\lambda_3-k_4}w^{k-5}$ which is common to all P's with the same values for (k) and (λ) and which may be omitted while considering their linear independence. Again set

$$s_{11} = t_{11}s_2s_{12}, \quad s_{13} = t_{13}s_3s_{10}, \quad t_{13}t_{11} - t_{14} = T_{14},$$
$$s_4t_{11} - s_5 = t_5, \quad s_{14} = t_{14}s_3s_{10},$$
$$P = (s_4)^{q_4}(s_4t_{11} - t_5)^{q_5}(t_{13})^{q_6+q_{13}}(T_{14})^{q_9} \times (t_5t_{13}t_{11} - t_5T_{14} - s_{16}t_{11})^{q_7}(t_{11})^{q_{11}} \times (t_{13}t_{11} - T_{14})^{q_{14}}(t_5t_{13} - s_{16} - s_4T_{14})^{q_{15}} \times (s_{16})^{q_{16}}(t_5)^{q_6+q_8}(s_2s_{12})^{k_5}(s_3s_{10})^{k_3-\lambda_3}.$$

The last two factors may be omitted. P is a polynomial in the independent functions $t_5, s_4, t_{11}, t_{13}, T_{14}, s_{16}$ and is of degree $q_5 + q_7 + q_{11} + q_{14}$ equal to $x_{11} - x_{22}$ in t_{11}. The coefficient of $(t_{11})^{x_{11}-x_{22}}$ in P is

$$Q = (s_4)^{q_4+q_5}(t_{13})^{q_6+q_{13}+q_{14}}(T_{14})^{q_9}(t_5t_{13} - s_{16})^{q_7} \times (t_5t_{13} - s_{16} - s_4T_{14})^{q_{15}}(s_{16})^{q_{16}}(t_5)^{q_6+q_8}.$$

Put $t_5t_{13} - s_{16} = t_{16}$, $t_{16} - s_4T_{14} = T$, $t_5 = T_5t_{16}$.

$$Q = (s_4)^{q_4+q_5-q_9}(t_{13})^{q_6+q_{13}+q_{14}}(t_{16} - T)^{q_9} \times (T_5t_{13} - 1)^{q_{16}}(T_5)^{q_6+q_8}(T)^{q_{15}}(t_{16})^{q_6+q_8+q_7+q_{16}}.$$

Q is of degree $q_6 + q_7 + q_8 + q_9 + q_{16}$, which is $x_{22} - x_{44}$ in t_{16}. The coefficient of the highest power of t_{16} in Q is a polynomial R in the variable s_4 of degree $k_4 - x_{44} - x_{13}$, again leaving a factor $(s_4)^{-k_5}$.

Now x_{22}, x_{44}, x_{13} can be taken as independent parameters distinguishing the equivalent IR's (λ) occurring in the IR (k). No two diagrams hence no two P's have the same set of values for these parameters. Let X_{22} be the minimum value of x_{22} in its range, and if many P's have the value X_{22} for x_{22}, then let X_{44} be the minimum value of x_{44}. When $x_{22} = X_{22}$ and $x_{44} = X_{44}$, let X_{13} be the minimum value of x_{13}. P is a polynomial of degree $x_{11} - x_{22}$ in t_{11}; the coefficient of $(t_{11})^{x_{11}-x_{22}}$ in P is a polynomial Q of degree $x_{22} - x_{44}$ in t_{16}; the coefficient of $(t_{16})^{x_{22}-x_{44}}$ in Q is a polynomial R of degree $k_4 - x_{44} - x_{13}$ in s_4. Consider an equation $\Sigma\beta(x_{22}, x_{44}, x_{13})P = 0$ where β's are numerical coefficients. The coefficient of the highest power of t_{11} in $\Sigma\beta P$ must be zero in that case. That is $\Sigma\beta(X_{22}, x_{44}, x_{13})Q = 0$. Again in $\Sigma\beta Q$, the coefficient of the highest power of t_{16} must be zero. That is, $\Sigma\beta(X_{22}, X_{44}, x_{13})R = 0$. Since the P's having the same values for x_{22} and x_{44} necessarily have distinct values for x_{13}, the polynomials R are of distinct degrees in s_4 and hence linearly independent. Therefore $\beta(X_{22}, X_{44}, x_{13}) = 0$. Considering the next least value X'_{44} of x_{44},we find by the same reasoning that the set of coefficients $\beta(X_{22}, X'_{44}, x_{13}) = 0$. Similarly all $\beta(X_{22}, x_{44}, x_{13}) = 0$ and $\beta(x_{22}, x_{44}, x_{13}) = 0$. Hence the P's corresponding to distinct diagrams (5.1) are linearly independent.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Bases for Irreducible Representations of U_{2n} in the Chain $U_{2n} \supset U_n + U_n$

V. SYAMALA DEVI AND T. VENKATARAYUDU
Andhra University, Waltair, India

(Received 5 December 1968)

The highest weight polynomials of irreducible representations of a $U_3 \times U_3$ subgroup of U_6, which occur in the reduction of a given irreducible representation of U_6, are obtained, making use of Littlewood's rules for the reduction of an irreducible representation of U_6 with respect to a $U_3 \times U_3$ subgroup. The various irreducible representations of $U_3 \times U_3$ occurring in the reduction are uniquely labeled by parameters which are obtained in a natural and obvious way from Littlewood diagrams. These polynomials could not be obtained by the algebraic methods developed so far. The method described here overcomes the division problem encountered there and makes the solution possible. It can also be directly extended to the case of $U_{2n} \supset U_n + U_n$ for any n.

INTRODUCTION

Moshinsky[1] showed that for a systematic derivation of closed formulas for the Wigner coefficients of U_n, the unitary group in n dimensions, one requires explicit construction of bases for an irreducible representation (IR) of U_{2n} consisting of subsets each of which spans a space irreducibly invariant under a particular subgroup $U_n \times U_n$. For this, it is sufficient to obtain the highest weight polynomials (h.w.p.) of the various IR's of $U_n \times U_n$ which are contained in the reduction of a given IR of U_{2n}, since the whole basis of a space can be obtained by applying lowering operators[2] of $U_n \times U_n$ on the highest weight vectors, and the irreducible spaces of $U_n \times U_n$ determined by linearly independent highest weight vectors are also linearly independent.[3] The required h.w.p. are all the linearly independent solutions of a certain set of homogeneous differential equations. Moshinsky attempted to solve them directly for polynomial solutions and obtained the h.w.p. in the case of the chain $U_4 \supset U_2 + U_2$ ($+$ is used since we are considering only the Lie algebras) after considerable labor. The algebraic methods employed there made it impossible to obtain all the polynomial solutions of the equations for the chain $U_6 \supset U_3 + U_3$. In a previous paper,[4] these polynomials were obtained much more elegantly for the chain $U_4 \supset U_2 + U_2$, and, by the same method, we will obtain them for the chain $U_6 \supset U_3 + U_3$ with the same ease. From the procedure that is described in the following sections, it will be evident that the method can be directly extended to the case of $U_{2n} \supset U_n + U_n$ for any n.

We will find that there exists a finite number of elementary polynomial solutions so that every polynomial solution of the equations is a polynomial function of the elementary solutions. The elementary solutions are easily obtained, as they are solutions of the differential equations for particular values of "h's." The particular values of "h's" to be considered will be evident from the Littlewood diagrams that one gets in the determination of the analysis of the Kronecker product of two IR's of U_n. Once the elementary solutions are obtained, it will be seen that it is quite easy to obtain all the required polynomial solutions for any general values of the "h's."

Chain $U_6 \supset U_3 + U_3$: For convenience and to save ourselves from extensive numerical work, we will restrict ourselves to partitions $(h_1, \cdots, h_4, 0, 0)$ of U_6. For any general partition $(h_1, \cdots, h_6)$, we will get only more elementary solutions. We will consider a U_6 group with generators $C^{ss'}$, $s, s' = 1, 2, 3, \bar{1}, \bar{2}, \bar{3}$, [called (i) hereafter] and two commuting U_3 subgroups with generators $C^{ss'}$, $s, s' = 1, 2, 3$, and $C^{tt'}$, $t, t' = \bar{1}, \bar{2}, \bar{3}$, respectively [called (ii) and (iii), respectively].

1. EQUATIONS DETERMINING THE h.w.p. OF $U_3 + U_3$ CONTAINED IN THE SPACE OF AN IR OF U_6

The basis vectors of an IR $(h_1, \cdots, h_4)$ of U_6 consists of a maximal linearly independent set of polynomial solutions of the equations[1]

$$C_{jj}P = h_jP, \quad C_{jj'}P = 0, \quad j < j' = 1 \cdots 4, \quad (1.1)$$

where $C^{ss'} = \sum_j a_{js}^{+} a_{js'}$, $s, s' = 1 \cdots \bar{3}$ is an operator realization of the generators of U_6 in terms of boson creation operators a_{js}^{+} and annihilation operators a_{js} and

$$C_{jj'} = \sum_{s=1,2,3,\bar{1},\bar{2},\bar{3}} a_{js}^{+} a_{j's},$$

and $P = P(a_{js}^{+})$ are polynomials in a_{js}^{+}.

Among the above polynomials, those which are of highest weight h_1', h_2', h_3' with respect to the U_3

[1] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[2] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).
[3] E. B. Dynkin, Am. Math. Soc. Transl. Ser. 2, **6**, 245 (1957).
[4] V. Syamala Devi and T. Venkatarayudu, J. Math. Phys. **9**, 1057 (1968).

group with generators (ii) are all the solutions of

$$C^{ss}P = h'_sP, \quad C^{ss'}P = 0, \quad s < s' = 1, 2, 3. \quad (1.2)$$

Similarly, those which are of highest weight $h'_{\bar{1}}, h'_{\bar{2}}, h'_{\bar{3}}$ with respect to the U_3 group with generators (iii) are all the solutions of

$$C^{tt}P = h'_tP, \quad C^{tt'}P = 0, \quad t < t' = \bar{1}, \bar{2}, \bar{3}. \quad (1.3)$$

2. REDUCTION OF AN IR $(h_1, \cdots, h_4)$ OF U_6 WITH RESPECT TO THE SUBGROUP $U_3 \times U_3$

The various IR's $(h'_1, h'_2, h'_3) \times (h'_{\bar{1}}, h'_{\bar{2}}, h'_{\bar{3}})$ of the direct product of groups $U_3 \times U_3$ contained in an IR $(h_1, \cdots, h_4)$ of U_6 are determined by the rules of Littlewood and Richardson for the decomposition of a product of S functions into S functions. We take the Young frame of the partition $(h'_1, \cdots, h'_3)$ whose first, second, and third row boxes are marked by the symbols 1, 2, and 3, respectively, and apply to it the Young frame of the partition $(h'_{\bar{1}}, \cdots, h'_{\bar{3}})$ whose first, second, and third row boxes are marked by symbols $\bar{1}$, $\bar{2}$, $\bar{3}$, respectively, according to the following rules of Littlewood and Richardson[5]:

1. After the addition of each set of identical symbols $\bar{r}$, we must have a regular Young diagram with no two identical symbols in a column;

2. If the total set of added symbols are read from right to left in the consecutive rows of the final diagram, we obtain a lattice permutation of

$$(\bar{1})^{h_{\bar{1}}'}(\bar{2})^{h_{\bar{2}}'}(\bar{3})^{h_{\bar{3}}'}.$$

Thus, corresponding to each of the following diagrams built according to the above rules,

$$\begin{array}{llll} \overbrace{1 \cdots 1}^{h'_1} & \overbrace{\bar{1} \cdots \bar{1}}^{x_{11}} & & \\ \overbrace{2 \cdots 2}^{h'_2} & \overbrace{\bar{1} \cdots \bar{1}}^{x_{12}} & \overbrace{\bar{2} \cdots \bar{2}}^{x_{22}} & \\ \overbrace{3 \cdots 3}^{h'_3} & \overbrace{\bar{1} \cdots \bar{1}}^{x_{13}} & \overbrace{\bar{2} \cdots \bar{2}}^{x_{23}} & \overbrace{\bar{3} \cdots \bar{3}}^{x_{33}} \\ \overbrace{\bar{1} \cdots \bar{1}}^{x_{14}} & \overbrace{\bar{2} \cdots \bar{2}}^{x_{24}} & \overbrace{\bar{3} \cdots \bar{3}}^{x_{34}}, & \end{array} \quad (2.1)$$

where the rows from top to bottom have lengths h_1, h_2, h_3, h_4, respectively, we have an IR $(h_1, \cdots, h_3) \times (h'_{\bar{1}}, \cdots, h'_{\bar{3}})$ of $U_3 \times U_3$ contained in an IR $(h_1, \cdots, h_4)$ of U_6. Let x_{ij} be the number of symbols $\bar{i}$ occurring in the jth row of (2.1). Then from Littlewood's rules we have

$$\begin{aligned} &h'_1 \geq h'_2 + x_{12} \geq h'_3 + x_{13} + x_{23} \geq x_{14} + x_{24} + x_{34}, \\ &h'_2 \geq h'_3 + x_{13} \geq x_{14} + x_{24}, \quad h'_3 \geq x_{14}, \\ &x_{11} \geq x_{22} \geq x_{33}. \end{aligned} \quad (2.2)$$

Since the lengths of the rows from top to bottom are $h_1, \cdots, h_4$, respectively,

$$\begin{aligned} h'_1 + x_{11} &= h_1, \\ h'_2 + x_{12} + x_{22} &= h_2, \\ h'_3 + x_{13} + x_{23} + x_{33} &= h_3, \\ x_{14} + x_{24} + x_{34} &= h_4. \end{aligned} \quad (2.3)$$

The total number of $\bar{1}$, $\bar{2}$, $\bar{3}$ are, respectively, $h'_{\bar{1}}$, $h'_{\bar{2}}$, $h'_{\bar{3}}$. Hence,

$$\begin{aligned} x_{11} + x_{12} + x_{13} + x_{14} &= h'_{\bar{1}}, \\ x_{22} + x_{23} + x_{24} &= h'_{\bar{2}}, \\ x_{33} + x_{34} &= h'_{\bar{3}}. \end{aligned} \quad (2.4)$$

Now the highest weight of the Lie algebra with generators (i) occurring in the IR of U_6 corresponding to the partition $(h_1, \cdots, h_4)$ is $(h_1, \cdots, h_4)$, and similarly, the highest weights of the two Lie algebras with generators (ii) and (iii), respectively, occurring in the IR's $(h'_1, \cdots, h'_3)$ and $(h'_{\bar{1}}, \cdots, h'_{\bar{3}})$ of the two U_3 groups are, respectively, $(h'_1, \cdots, h'_3)$ and $(h'_{\bar{1}}, \cdots, h'_{\bar{3}})$. Whether or not the reduction of IR's of a U_6 group with respect to a product of two commuting U_3 groups given in Littlewood's book applies to our present U_6 and $U_3 \times U_3$ groups with the above mentioned generators can be seen by considering the rules with reference to the basic representation[6] $(10\cdots)$ of U_6. For example, the U_6 and $U_3 \times U_3$ groups dealt with in Littlewood's book[7] are such that the basic representation $(10\cdots)$ of U_6 contains the representations $(10\cdots) \times (0\cdots0)$ and $(0\cdots0) \times (100)$ of $U_3 \times U_3$. One easily finds by considering polynomial bases that the IR of the Lie algebra with generators (i) with highest weight $(10\cdots)$ contains the IR's of $U_3 + U_3$ [the direct sum of algebras with generators (ii) and (iii)] with highest weights (100; 000) and (000; 100), where the first three components correspond to one U_3 and the last three components to the

[5] D. E. Littlewood, *Theory of Group Characters* (Clarendon Press, Oxford, 1950), p. 94.

[6] See E. B. Dynkin, Am. Math. Soc. Trans. Ser. 2, **6** (1950) for definition of equivalent embeddings of Lie algebra G' in a Lie algebra G and a criterion for the equivalence of two embeddings.

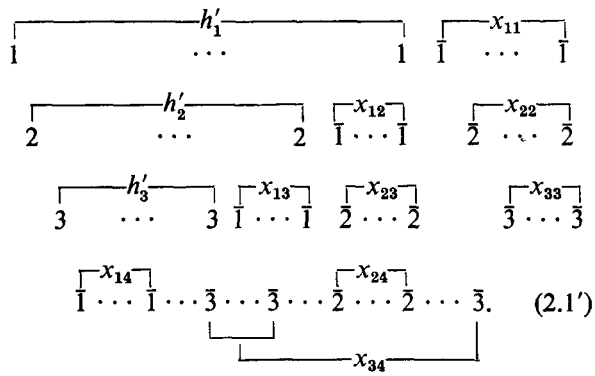
[7] Littlewood's book actually gives a way of writing the product of two S functions as a sum of S functions which directly gives an analysis of the Kronecker product of two irreducible representations of the symmetric group. To each irreducible homogeneous integral representation of degree m of the full linear group, and hence of the unitary group, corresponds an irreducible representation of the symmetric group on m symbols. Hence, the analysis of the Kronecker product of representations of the symmetric group gives a reduction procedure for an IR of U_{m+n} with respect to the direct product of two commuting subgroups $U_m \times U_n$ by considering the Frobenius theorem of reciprocality for finite groups.

other U_3. Thus, the rules given in Littlewood's book hold for the U_6 and $U_3 \times U_3$ groups in the present article.

We determine a basis for all the solutions of (1.1)–(1.3) in one-to-one association with Littlewood diagrams.

3. ELEMENTARY LITTLEWOOD DIAGRAMS

We will rearrange (2.1) as follows. In each row, the $\bar{2}$'s are placed as far as possible under the $\bar{1}$'s available in the row above; if all the places under the $\bar{1}$'s are filled, then the $\bar{2}$'s are placed under 2 or 3, whichever are available in the row above. Again, the $\bar{3}$'s are placed as far as possible under the $\bar{2}$'s available in the row above. If all the places under the $\bar{2}$'s are filled and there is still an excess of $\bar{3}$'s in the row, they are placed as far as possible under the $\bar{1}$'s. If the places under the $\bar{1}$'s are also filled and still there are $\bar{3}$'s present in the row, they are placed under the 3's. The rearranged diagram is



(2.1′)

We divide (2.1′) columnwise into elementary permissible diagrams (e.p.d.). A diagram is said to be permissible if it is a possible rearranged Littlewood diagram [that is, if the numbers x_{ij} satisfy the inequalities and Eqs. (2.2)–(2.4)]; it is said to be elementary if it cannot be split columnwise into two permissible diagrams. All the possible columns of (2.1′) are the following:

$$\begin{array}{cccccccccccccccc}
\bar{1}, & \bar{1}, & \bar{1}, & 1, & 1, & 1, & 1, & 1, & 1, & 1, & 1, & 1, & 1, & 1 \\
\bar{2} & \bar{2} & E_3 & \bar{1} & \bar{1} & \bar{1} & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\
3 & E_2 & & \bar{2} & \bar{2} & E_6 & \bar{2} & \bar{2} & \bar{1} & \bar{1} & \bar{1} & 3 & 3 & 3 \\
E_1 & & & \bar{3} & E_5 & & \bar{3} & & \bar{2} & \bar{3} & E_{11} & \bar{3} & \bar{2} & \bar{1} \\
1, & 1, & 1, & E_4 & & & & E_{10} & & & & & E_{14} & \\
2 & 2 & E_{17} & & & & & & & & & & & \\
3 & E_{16} & & & & & & & & & & & & \\
E_{15} & & & & & & & & & & & & &
\end{array}$$

Their frequencies in (2.1′) are, respectively, as follows:

$$\begin{aligned}
x_{33} = m_1, \quad x_{22} - x_{33} = m_2, \quad x_{11} - x_{22} &= m_3, \\
\min(x_{12}, x_{23}, x_{34}) &= m_4, \\
\min(x_{12}, x_{23}) - m_4 &= m_5, \\
x_{12} - m_4 - m_5 &= m_6, \\
\min(x_{23} - m_4, x_{34} - m_4) &= m_7, \\
x_{23} - m_4 - m_5 - m_7 &= m_8, \\
\min(x_{13}, x_{24}) &= m_{10}, \\
\min(x_{13} - m_{10}, x_{34} - m_4 - m_7) &= m_9, \\
x_{13} - m_9 - m_{10} &= m_{11}, \\
x_{34} - m_4 - m_7 - m_9 &= m_{12}, \\
x_{24} - m_{10} = m_{13}, \quad x_{14} &= m_{14}, \\
h_3' - x_{14} - m_{13} - m_{12} &= m_{15}, \\
h_2' - h_3' - x_{13} - m_7 - m_8 &= m_{16}, \\
h_1' - h_2' - x_{12} &= m_{17}.
\end{aligned}$$

To find all the possible e.p.d.'s, let us for example find all the e.p.d.'s containing the nonpermissible column

$$\begin{array}{c} 1 \\ 2 \\ \bar{1} \\ \bar{3} \end{array}$$

A permissible diagram containing the column must contain at least one $\bar{2}$ in the second row (and not in the third row since then the diagram would not contain the column); hence, it should contain at least one volumn $\frac{\bar{1}}{2}$. Hence, the smallest permissible diagram containing the column is

$$\begin{array}{cc} 1 & \bar{1} \\ 2 & \bar{2} \\ \bar{1} & \\ \bar{3} & \end{array}$$

Moreover, a permissible diagram containing the above e.p.d. remains permissible by leaving out the two columns of which it consists, as one finds that the inequalities (2.1) still hold in the remaining diagram. Hence,

$$\begin{array}{cc} 1 & \bar{1} \\ 2 & \bar{2} \\ \bar{1} & \\ \bar{3} & \end{array}$$

is the only e.p.d. containing the column. Similarly, one finds that the following are all the possible e.p.d.'s. Since a permissible diagram is either elementary or can be split columnwise into two permissible diagrams, it follows that any diagram is a set of the following e.p.d. put together, neglecting the order of columns in a diagram. In addition to the permissible columns

among those written already, the possible e.p.d.'s are

$1\,\bar{1}$	$1\,\bar{1}$	$1\,\bar{1}$	$1\,\bar{1}$	$1\,1$	$1\,\bar{1}$
2	2	$2\,\bar{2}$	$2\,\bar{2}$	$2\,\bar{1}$	2
$\bar{2}$	$\bar{2}$	$\bar{1}$	3	3	3
$\bar{3}$		$\bar{3}$	$\bar{3}$	$\bar{2}$	$\bar{2}$
E_7	E_8	E_9	E_{12}	E_{13}	E'_{13}

The division of (2.1′) into e.p.d.'s is not unique as is seen from the following example. The diagram

$$\begin{array}{l} 1\ 1\ \bar{1} \\ 2\ \bar{1} \\ 3 \\ \bar{2} \end{array}$$

may be thought of as consisting of either E_{13} and E_3, or E'_{13} and E_6. Then we make it a convention to include a nonpermissible column of (2.1′) in an e.p.d. by adjoining to it the possible column nearest to it as we go from left to right, and we propose to do this for every nonpermissible column, going from left to right. Thus, in the example above, we think that $(2.1') = E_{13}E_3$.

The frequencies with which the e.p.d.'s having more than one column occur in (2.1′) are

$$\begin{aligned} C_{13} &= \min(m_{13}, m_6), \\ C_{14} &= \min(m_{13} - C_{13}, m_3) = C'_{13}, \\ C_{12} &= \min(m_{12}, m_2) = m_{12}, \\ C_9 &= \min(m_9, m_2 - C_{12}) = m_9, \\ C_7 &= m_7 = \min(m_7, m_3 - C_{14}), \\ C_8 &= m_8 = \min(m_8, m_3 - C_{14} - C_7). \end{aligned}$$

4. CONSTRUCTION OF THE h.w.p.

Each e.p.d. is a diagram built by the application of the symbols of one Young diagram on another according to Littlewood's rules. We read out the values of $h'_1, \cdots, h'_{\bar{3}}$ and $h_1, \cdots, h_4$ in each e.p.d. and determine the solutions of (1.2, 1.3) by taking suitable linear combinations of the Weyl basis vectors, obtained by filling the Young diagram of $(h_1 \cdots h_4)$ by symbols 1, 2, 3, $\bar{1}$, $\bar{2}$, $\bar{3}$, so that the numbers in the rows are nondecreasing, and those in the columns are increasing (we regard that $i < j$ for all i and j). To each Weyl basis vector corresponds a polynomial[8] in a^+_{js} which is already a solution of the equations (1.1). In the present case, we get a unique solution of Eqs. (1.1)–(1.3) for each e.p.d. The solutions corresponding to the e.p.d. of (2.1′) and denoted, respectively, by $s_1, \cdots, s_{17}$ are as follows:

$$\begin{aligned} &s_1 = (\bar{1}\,\bar{2}\,\bar{3}), \quad s_2 = (\bar{1}\,\bar{2}), \quad s_3 = (\bar{1}), \quad s_4 = (1\,\bar{1}\,\bar{2}\,\bar{3}), \\ &s_5 = (1\,\bar{1}\,\bar{2}), \quad s_6 = (1\,\bar{1}), \\ &s_7 = (\bar{1})(1\,2\,\bar{2}\,\bar{3}) + (\bar{3})(1\,2\,\bar{1}\,\bar{2}) + (\bar{2})(1\,2\,\bar{3}\,\bar{1}), \\ &s_8 = (\bar{1})(1\,2\,\bar{2}) - (\bar{2})(1\,2\,\bar{1}), \\ &s_9 = (1\,2\,\bar{1}\,\bar{3})(\bar{1}\,\bar{2}) - (1\,2\,\bar{1}\,\bar{2})(\bar{1}\,\bar{3}), \\ &s_{10} = (1\,2\,\bar{1}\,\bar{2}), \quad s_{11} \doteq (1\,2\,\bar{1}), \\ &s_{12} = (1\,2\,3\,\bar{3})(\bar{1}\,\bar{2}) + (1\,2\,3\,\bar{1})(\bar{2}\,\bar{3}) + (1\,2\,3\,\bar{2})(\bar{3}\,\bar{1}), \\ &s_{13} = (1\,2\,3\,\bar{2})(1\,\bar{1}) - (1\,2\,3\,\bar{1})(1\,\bar{2}), \\ &s'_{13} = (1\,2\,3\,\bar{2})(\bar{1}) - (1\,2\,3\,\bar{1})(\bar{2}), \quad s_{15} = (1\,2\,3), \\ &s_{16} = (1\,2), \quad s_{17} = (1), \quad s_{14} = (1\,2\,3\,\bar{1}), \end{aligned}$$

where

$$(s_1 \cdots s_r) = \det \begin{vmatrix} a_1^+ s_1 & \cdots & a_1^+ s_r \\ a_r^+ s_1 & \cdots & a_r^+ s_r \end{vmatrix}.$$

We associate with (2.1′) the polynomial which is the product of all the elementary solutions corresponding to all the e.p.d.'s which make up (2.1′) that is equal to

$$\begin{aligned} P = {} & (1)^{m_{17}}(12)^{m_{16}}(123)^{m_{15}}(s_{13})^{C_{13}}(s'_{13})^{C_{14}}(s_{12})^{C_{12}}(s_9)^{C_9} \\ & \times (123\bar{1})^{m_{14}}(121)^{m_{11}}(1212)^{m_{10}}(s_7)^{C_7}(s_8)^{C_8}(11)^{m_6 - C_{13}} \\ & \times (112)^{m_5}(1123)^{m_4}(1)^{m_3 - C_{14} - C_8 - C_7} \\ & \times (12)^{m_2 - C_{12} - C_9}(123)^{m_1}. \end{aligned} \tag{4.1}$$

P is then a solution of Eqs. (1.1)–(1.3).

When the h_i, h'_j, and h'_j are fixed, all the quantities x_{ij} depend on three independently varying parameters which may be chosen as x_{34}, x_{24}, and x_{13} [as seen from Eqs. (1.2), (1.3)]. The range of their variation is fixed by the inequalities and equations (1.1)–(1.3). To distinct sets of values of x_{34}, x_{24}, and x_{13} correspond distinct diagrams (2.1). Hence, they serve to label the various equivalent IR's of $U_3 \times U_3$ occurring in the reduction of an IR of U_6.

It remains to be shown that the h.w.p. associated with diagrams (2.1) in the way described above are linearly independent. The h.w.p. corresponding to distinct sets of values for (h'_1, h'_2, h_3) and $(h'_{\bar{1}}, h'_{\bar{2}}, h'_{\bar{3}})$, while h_1, h_2, h_3, h_4 are fixed, are linearly independent, as they are eigenfunctions of linear differential operators C^{ss} and C^{tt} corresponding to distinct sets of eigenvalues. It remains to be shown that when an IR $(h'_1, \cdots, h'_{\bar{3}}) \times (h'_{\bar{1}}, \cdots, h'_{\bar{3}})$ of $U_3 \times U_3$ occurs more than once in the IR $(h_1, \cdots, h_4)$ of U_6, the polynomials P given by (4.1) are linearly independent.

5. LINEAR INDEPENDENCE OF THE h.w.p.

Let us define the functions $x_i = (\bar{i})$, $y_i = (1\,\bar{i})$, $z_i = (1\,2\,\bar{i})$, $w_i = (1\,2\,3\,\bar{i})$. The polynomials x_i, y_i,

[8] G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1459 (1963).

z_i, w_i are functionally independent. The polynomial (4.1), when expressed in terms of the new functions $x_i, \cdots, w_i$, is

$$
\begin{aligned}
&(1)^{h_1'-h_2}(12)^{h_2'-h_3}(123)^{h_3'-h_4}(w_1)^{m_{14}}(z_1w_2 - z_2w_1)^{m_{10}} \\
&\quad \times [w_1(x_2y_3 - x_3y_2) + w_2(x_3y_1 - x_1y_3) \\
&\quad + w_3(x_1y_2 - x_2y_1)]^{C_{12}} \\
&\quad \times (y_1w_2 - y_2w_1)^{C_{13}}(x_1w_2 - x_2w_1)^{C_{14}} \\
&\quad \times (x_1z_2 - x_2z_1)^{C_8}(z_1)^{m_{11}} \\
&\quad \times [(x_1y_2 - x_2y_1)(z_1w_3 - z_3w_1) \\
&\quad - (x_1y_3 - x_3y_1)(z_1w_2 - z_2w_1)]^{C_9} \\
&\quad \times [x_1(z_2w_3 - z_3w_2) + x_2(z_3w_1 - z_1w_3) \\
&\quad + x_3(z_1w_2 - z_2w_1)]^{C_7} \\
&\quad \times [y_1(z_2w_3 - z_3w_2) + y_2(z_3w_1 - z_1w_3) \\
&\quad + y_3(z_1w_2 - z_2w_1)]^{m_4} \\
&\quad \times (y_1)^{m_6-C_{13}}(y_1z_2 - y_2z_1)^{m_5}(x_1)^{m_3-C_{14}-C_7-C_8} \\
&\quad \times (x_1y_2 - x_2y_1)^{m_2-C_{12}-C_9} \\
&\quad \times [x_1(y_2z_3 - y_3z_2) + x_2(y_3z_1 - y_1z_3) \\
&\quad + x_3(y_1z_2 - y_2z_1)]^{m_1}.
\end{aligned}
$$

The first three factors may be omitted as they are common to all the polynomials having the same values for $h_1, \cdots, h_4$ and $h_1', \cdots, h_3'$. P is of degree $m_1 + m_2 + m_3 = h_1 - h_1'$ in the quantity x_1. Hence, a linear relation

$$\sum a(x_{34}, x_{24}, x_{13})P(x_{34}, x_{24}, x_{13}) = 0,$$

where $a(x_{34}, x_{24}, x_{13})$ are numerical coefficients, implies that the coefficient of $x_1^{h_1-h_1'}$ in $\sum aP$ must also be zero. This coefficient is

$$
\begin{aligned}
Q = &(Y_2W_3 - Y_3W_2)^{C_{12}}(W_2 - Y_2)^{C_{13}}(W_2)^{C_{14}}Y_2(W_3 - Z_3) \\
&- Y_3(W_2 - Z_2)^{C_9}(Z_2W_3 - Z_3W_2)^{C_7}(Z_2)^{C_8} \\
&\times (W_2 - Z_2)^{m_{10}}(Z_2 - Y_2)^{m_5}(Y_2)^{m_2-C_{12}-C_9} \\
&\times (Z_2W_3 - Z_3W_2 + Y_2Z_3 - Y_3Z_2 \\
&+ Y_3W_2 - Y_2W_3)^{m_4}(Y_2Z_3 - Y_3Z_2)^{m_1} \\
&\times (y_1)^{h_2-h_2'}(z_1)^{h_3}(w_1)^{h_4},
\end{aligned}
$$

where $y_i = y_1Y_i$, $z_i = z_1Z_i$, $w_i = w_1W_i$. $\sum aP = 0$ implies $\sum aQ = 0$, and the last three factors in Q may be omitted as they are common to all the distinct polynomials Q corresponding to all the distinct polynominals P given by (4.1) and having the same values for the h's. Q is a polynomial in W_3 of degree $C_{12} + C_9 + C_7 + m_4 = x_{34}$. The coefficient of $W_3^{x_{34}}$ in Q is

$$
\begin{aligned}
R = &(W_2 - Y_2)^{C_{13}}(W_2)^{C_{14}}(Z_2)^{C_7+C_8}(W_2 - Z_2)^{m_{10}} \\
&\times (Z_2 - Y_2)^{m_4+m_5}(Y_2)^{m_2}(Y_2Z_3 - Y_3Z_2)^{m_1}.
\end{aligned}
$$

R is again a polynomial in W_2 of degree $m_{10} + C_{13} + C_{14} = x_{24}$. The coefficient of $W_2^{x_{24}}$ in R is a polynomial T in Z_2 of degree $C_7 + C_8 + m_4 + m_5 + m_1 = h_3 - h_3' - x_{13}$.

To prove the linear independence of the P's given by (4.1), we observe that no two P's have the same set of values for the parameters x_{24}, x_{34}, and x_{13}. Let us divide all the P's into classes such that two P's belonging to a class have the same value for x_{34}, and two P's belonging to distinct classes have distinct values for x_{34}. Each class is divided into subclasses in a similar way so that two P's contained in a subclass have the same value for x_{24}, and two P's coming from two distinct subclasses have distinct values for x_{24}. We find that two P's belonging to the same subclass have distinct values for x_{13}. We know already that $\sum aP = 0$ implies $\sum aQ = 0$. Treating Q as a polynomial in W_3, the coefficient of each power of W_3 is zero. (The functions Y_i, W_i, $i = 2, 3$, are also functionally independent.) Hence, if X_{34} is the maximum possible value of x_{34} in its range, then the coefficient of $W_3^{x_{34}}$ in $\sum aQ$ must be zero. Therefore, $\sum a_jR_j = 0$, where the a_j's are a subset of all the coefficients $a(x_{34}, x_{24}, x_{13})$ and each a_j is of the form $a(X_{34}, x_{24}, x_{13})$. Again, since R_j is a polynomial in W_2, the coefficient of $W_2^{x_{24}}$ in $\sum a_jR_j$, where X_{24} is the maximum value of x_{24} when x_{34} has the value X_{34}, must be zero. That is, $\sum a_{j_k}T_k = 0$ where the a_{j_k}'s are a subset of the coefficients a_j. But T_k is a polynomial in Z_2 of degree $h_3 - h_3' - x_{13}$ and all the T_k's for the various possible k have distinct values for x_{13}. Hence, each $a_{j_k} = 0$. Considering the next highest value X_{24}' of x_{24}, when x_{34} has the value X_{34} and the coefficient of $(W_3)^{x_{34}}(W_2)^{x_{24}'}$ in $\sum aQ$ is zero, leads to the vanishing of another subset of coefficients $a_{j_p} = 0$. Similarly, fixing the value X_{34} for x_{34} and considering the possible values of x_{24} in decreasing order, we get each a_j to be zero. Again, repeating the same argument and giving values to x_{34} in decreasing order in its range, we get each coefficient $a(x_{34}, x_{24}, x_{13}) = 0$. Hence, the P's given by (4.1) corresponding to distinct diagrams, i.e., having distinct sets of values for the parameters (x_{34}, x_{24}, x_{13}), are linearly independent.

ACKNOWLEDGMENT

Thanks are due to Professor M. Moshinsky for his interest.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Energy-Dependent Boltzmann Equation in the Fast Domain*

BASIL NICOLAENKO†
Department of Nuclear Engineering, The University of Michigan, Ann Arbor, Michigan

(Received 17 March 1969)

This work presents some aspects of the static energy-dependent Boltzmann equation in plane geometry using a continuous-energy formulation. In a first part, solutions are found for a class of synthetic separable (but nondegenerate) energy-transfer kernels. Such kernels are representative, for instance, of neutron *inelastic* slowing down. In a second part, the same problem is considered with the addition of a projection kernel (typical of neutron fission); it is shown that the solutions split into space–energy separable components and nonseparable "slowing-down transients."

I. INTRODUCTION

Little progress has been made in the solution of the energy-dependent Boltzmann equation, as opposed to the status of the one-speed transport equation, where Case's method of singular normal modes[1] has yielded considerable success.

For a long time, energy-dependent investigations were specialized to finding exact solutions to the spatially dependent neutron slowing-down problem, with *elastic* scattering and without fission.[2–7]

More recent work has been applied to the neutron thermalization domain: The energy-transfer operator has been approximated by a sum of degenerate (projection) kernels, which, in turn, allows the reduction of the initial equation to a set of coupled one-speed transport equations.[8–15] Also, some work has been done on the multigroup formulation,[16–18] but the discretization of the energy variable distorts the spectrum of the Boltzmann operator: This is of prime importance in the study of the time-dependent evolution.[19]

Works trying to extend Case's method to the most general energy-dependent equation are somewhat limited in scope.[20] (The completeness theorem involved relies upon the Fredholm alternative for the inversion of operator equations, which is correct only when the energy-transfer kernel, or some iterate, is compact[21]; it fails for unbounded and noncompact kernels such as are found in neutron slowing-down theory.)

This work presents some aspects of the energy-dependent, static Boltzmann equation, in plane geometry, with a continuous energy formulation. In a first part, solutions are found for a class of synthetic, separable, but not degenerate, energy-transfer kernels: Such kernels are representative, for instance, of neutron *inelastic* slowing down. A new energy transformation is developed, which reduces the initial equation to a simple form, and an asymptotic evaluation of the Green's function is given.

In a second part, the same problem is considered with the addition of a projection kernel (typical of neutron fission); it is shown that the solutions split into (1) space–energy separable components, representative of the neutron regeneration, and asymptotically dominant; and (2) nonseparable, "slowing-down transients" solution of the initial equation without the fission-projection kernel. This generalizes the results of a previous work[22]: In order to achieve completeness for the normal modes of the Boltzmann equation with fission and slowing down, one must introduce auxiliary modes which are solutions of the ordinary slowing-down equation.

* Work supported by the National Science Foundation. Based in part on the author's Ph.D. thesis, Department of Nuclear Engineering, The University of Michigan.

† Present address: Brookhaven National Laboratory, Upton, Long Island, New York.

[1] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Reading, Mass., 1967).
[2] R. E. Marshak, Rev. Mod. Phys. **19**, 185 (1947).
[3] M. Verde and G.-C. Wick, Phys. Rev. **71**, 852 (1947).
[4] G.-C. Wick, Phys. Rev. **75**, 738 (1949).
[5] G. Hölte, Arkiv Fysik, **2**, No. 48 (1950).
[6] H. Hölte, Arkiv Fysik, **2**, No. 14 (1951).
[7] J. J. MacInerney, Nucl. Sci. Eng . **22**, 215 (1965).
[8] J. R. Mika, Nucl. Sci. Eng . **22**, 235 (1965).
[9] J. C. Stewart, I. Kuščer, and N. J. McCormick, Ann. Phys. (N.Y.) **40**, 321 (1966).
[10] M. M. R. Williams, Nucl. Sci. Eng . **18**, 260 (1964).
[11] M. M. R. Williams, Nucl. Sci. Eng . **27**, 511 (1967).
[12] J. Ferziger and A. Leonard, Ann. Phys. (N.Y.) **22**, 192 (1963).
[13] A. Leonard and J. Ferziger, Nucl. Sci. Eng . **26**, 170 (1966).
[14] A. Leonard and J. Ferziger, Nucl. Sci. Eng . **26**, 181 (1966).
[15] I. Inönü and P. F. Zweifel, *Recent Developments in Transport Theory*, Proceedings of the Ankara Symposium on Transport Theory (Academic Press Inc., New York, 1967).
[16] R. Zelazny and A. Kuszell, *Proceedings of the International Symposium on Physics of Fast and Intermediate Reactors* (IAEA, Vienna, 1962), Vol. 2, pp. 55–71.
[17] D. Metcalf, Ph.D. thesis, The University of Michigan, 1968.
[18] J. K. Shultis, Ph.D. thesis, The University of Michigan, 1968.
[19] *Proceedings of the International Symposium on Pulsed Neutron Research* (IAEA, Vienna, 1965).
[20] R. J. Bednarz and J. R. Mika, J. Math. Phys. **4**, 1285 (1963).
[21] A. E. Taylor, *Introduction to Functional Analysis* (John Wiley & Sons, New York, 1958).
[22] B. Nicolaenko and P. F. Zweifel, *Proceedings of the International Symposium on Neutron Thermalization and Reactor Spectra* (IAEA, Vienna, 1968), p. 63.

II. THE BOLTZMANN EQUATION[1] WITH A CLASS OF SYNTHETIC ENERGY-TRANSFER KERNELS

A. Introduction of a Synthetic Scattering Kernel

Many extensive solutions have been proposed to the problem of *spatial* neutron slowing down with *elastic* scattering.[2–7,23,24] Yet, little attention has been paid to the fast domain, where *inelastic* scattering is overwhelming dominant, especially for heavy nuclei: Most calculation schemes use the multigroup (discrete energy) formulation. Such a multigroup formulation distorts the spectrum of the Boltzmann operator. The necessity for a continuous energy formulation has been widely recognized in the neutron thermalization domain, where a rigorous study of the spectrum of the Boltzmann operator is fundamental for the interpretation of the time-dependent evolution of the neutron field.[19] This is also valid for the fast domain.

With the assumption of plane symmetry, we are interested in the following slowing-down transport equation:

$$\mu \frac{\partial \psi}{\partial x}(x, \mu, E) + \psi(x, \mu, E) = \frac{c_i}{2} \int_E^{+\infty} \int_{-1}^{+1} K_{in}(E' \to E)\psi(x, \mu', E')\, dE'\, d\mu' + \frac{c_e}{2} \int_{-1}^{+1} \psi(x, \mu', E)\, d\mu' + S(x, \mu, E), \quad (1)$$

where $\psi(x, \mu, E)$ is the angular neutron density, c_i is the mean number of secondaries emitted after an inelastic scattering collision times the probability of inelastic scattering, c_e is the mean number of secondaries emitted after an elastic scattering collision times the probability of elastic scattering, x is the position variable measured in optical units, E is the neutron energy, μ is the cosine of the angle between the neutron velocity vector and the x axis, and $S(x, \mu, E)$ is the source term. The kernel $K_{in}(E' \to E)$ gives the probability that a neutron of energy E' will be slowed down to a unit energy interval about the energy E by inelastic collision. Inelastic scattering is assumed isotropic in the laboratory system.

Equation (1) contains two simplifying assumptions:

(a) The cross sections are supposed to be constant above the inelastic scattering threshold (first excited level $\simeq$30 keV for heavy fissionable nuclei);

(b) Above the inelastic threshold, the energy transfer due to *elastic* scattering is considered as negligible compared to the inelastic energy degradation (which is valid for heavy nuclei).

In no way does Eq. (1) assume constant cross sections throughout the whole energy range. Below the inelastic threshold energy E_0, the inelastic scattering term becomes a *known* isotropic *source* term;

$$\frac{c_i}{2} \int_{-1}^{+1} d\mu' \int_{E_0}^{\infty} K_{in}(E' \to E)\psi(x, \mu', E)\, dE', \quad (2)$$

and we are left with the solution of a classical spatial *elastic* slowing-down problem.[2–7,23,24] Therefore, we are interested in solutions of Eq. (1) for energies greater than E_0.

The exact shape of $K_{in}(E' \to E)$ is poorly known, and,as in thermalization theory, it is advantageous to introduce a synthetic kernel. The simplest approximation is to assume

$$K_{in}(E' \to E) = f(E')g(E), \quad \text{for } E' > E,$$
$$= 0, \quad \text{for } E' < E, \quad (3)$$

where $f(E)$ and $g(E)$ are *a priori* arbitrary functions. The synthetic kernel (3) was first introduced by Okrent *et al.*,[25] in connection with Weisskopf's statistical evaporation model. Recently, it was proposed as a synthetic kernel *per se*, by Cadilhac *et al.*,[26] that is, a kernel adaptable to experimental data or more involved nuclear theory. Such a kernel has, in fact, only one arbitrary function, namely $g(E)$; this stems from the requirement of the conservation of the total inelastic cross section:

$$\int_0^{E'} K_{in}(E' \to E)\, dE = f(E') \int_0^{E'} g(E)\, dE = 1. \quad (4)$$

Defining

$$h(E) = 1/f(E), \quad (5)$$

this yields

$$g(E) = \frac{d}{dE} h(E). \quad (6)$$

In Weisskopf's statistical evaporation model, $g(E)$ assumes the shape [25]

$$g(E) = Ee^{-E/T}, \quad (7a)$$

where T is the "nuclear temperature," and

$$h(E) = \int_0^E E' \exp(-E'/T)\, dE'. \quad (7b)$$

[23] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).

[24] M. M. R. Williams, *The Slowing-Down and Thermalization of Neutrons* (Interscience Publishers, Inc., New York, 1966).

[25] S. Yiftah, D. Okrent, and P. Moldauer, *Fast Reactor Cross-Sections* (Pergamon Press, Inc., New York, 1961).

[26] M. Cadilhac and M. Pujol, J. Nucl. Energy **21**, 58 (1967).

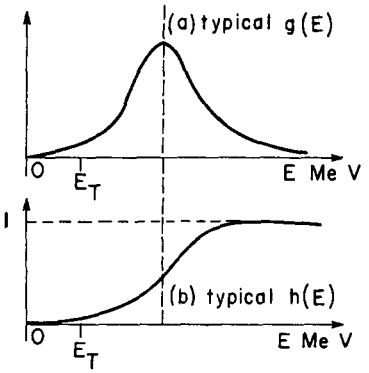


FIG. 1. Typical shapes for (a) $g(E)$ and (b) $h(E)$.

Cadilhac's approach[26] is more general; it consists in keeping $g(E)$ *a priori* arbitrary, and fitting it so that the approximate operator has the same action as the exact one on a particular reference energy spectrum. Mathematically, we need only know that

$$g(E) \to 0, \quad \text{for} \quad E \to 0+, \quad E \to +\infty,$$
$$h(E) \text{ is } \textit{bounded} \text{ for} \quad E \to +\infty.$$

The latter fact is not trivial, but holds for all suggested physical models (see Fig. 1). Similarly, one can assume that $g(E)$ is positive for $\forall E \in [0, +\infty]$.

B. Reduction of the Boltzmann Equation Through a New Energy Transformation

We consider Eq. (1) with a synthetic inelastic slowing-down kernel:

$$\mu \frac{\partial}{\partial x} \psi(x, \mu, E) + \psi(x, \mu, E) = \frac{c_i}{2} \int_{-1}^{+1} d\mu' g(E) \int_E^\infty \frac{\psi(x, \mu', E')}{h(E')} dE' + \frac{c_e}{2} \int_{-1}^{+1} \psi(x, \mu', E)\, d\mu' + S(x, \mu, E). \quad (8)$$

A classical method in *elastic* slowing-down problems consists of looking for energy eigenfunctions of the slowing-down operator (namely, exponentials of the lethargy variable) and making an expansion of the neutron density in this set of eigenfunctions (that is, a Fourier–Laplace transformation of the lethargy variable, or a Mellin transformation of the velocity variable).[23]

In the present case, the inelastic slowing-down operator has no eigenfunctions; the following equation,

$$\lambda \phi(E) = g(E) \int_E^\infty \frac{\phi(E')}{h(E')} dE', \quad (9)$$

has no solutions [Volterra integral equation with a bounded kernel; $h(\infty)$ is finite, different from zero].

However, consider the *adjoint* to the inelastic slowing-down operator; the adjoint eigenfunction equation is

$$\phi^+(E) = \frac{\lambda}{h(E)} \int_0^E g(E') \phi^+(E')\, dE'. \quad (10)$$

Equation (10) admits the following solutions:

$$\phi^+(E) = h(E)^{\lambda - 1}, \quad \forall \lambda, \quad \operatorname{Re} \lambda \geq 1 \quad (11)$$

[keeping in mind that $g(E) = dh(E)/dE$]. So, the adjoint operator (Volterra integral equation with an *unbounded* kernel) admits the set of eigenfunctions $\{h(E)^{\lambda-1}\}$. Therefore, let us make the "scalar product" of Eq. (8) by $h(E)^{\lambda-1}$; multiply both sides of the transport equation (8) by $h(E)^{\lambda-1}$ and integrate over the whole energy range; defining

$$\bar{\psi}(x, \mu, \lambda) = \int_0^\infty \psi(x, \mu, E) h(E)^{\lambda-1}\, dE, \quad (12)$$

one obtains

$$\mu \frac{\partial}{\partial x} \bar{\psi}(x, \mu, \lambda) + \bar{\psi}(x, \mu, \lambda) = \frac{c_e}{2} \int_{-1}^{+1} \bar{\psi}(x, \mu', \lambda)\, d\mu' + \bar{S}(x, \mu, \lambda) + \frac{c_i}{2} \int_{-1}^{+1} d\mu' \int_0^\infty g(E)\, h(E)^{\lambda-1} \int_E^\infty \frac{\psi(x, \mu', E')}{h(E')} dE'\, dE. \quad (13)$$

In the last (inelastic scattering) term of Eq. (13), change the order of integrations, use relations (10) and (11), and obtain

$$\mu \frac{\partial}{\partial x} \bar{\psi}(x, \mu, \lambda) + \bar{\psi}(x, \mu, \lambda) = \frac{c_e}{2} \int_{-1}^{+1} \bar{\psi}(x, \mu', \lambda)\, d\mu' + \frac{c_i}{2\lambda} \int_{-1}^{+1} \bar{\psi}(x, \mu', \lambda)\, d\mu' + \bar{S}(x, \mu, \lambda). \quad (14)$$

So, if one defines the transformation $\mathcal{M}$ by

$$\mathcal{M}\psi(x, \mu, E) = \int_0^\infty \psi(x, \mu, E) h(E)^{\lambda-1}\, dE, \quad (15)$$

this transformation reduces the initial Boltzmann equation (8) to a pseudomonokinetic equation (14), where λ is only a parameter appearing in the "multiplication coefficient" $\{c_e + (c_i/\lambda)\}$—which can take complex values, as opposed to the classical one-speed situation.

The solutions of "one-speed" equations like (14) are perfectly well known.[1] The only problem is to find an inversion formula to the transformation $\mathcal{M}$.

$\mathcal{M}$ is always defined, provided that $\mathrm{Re}\,\lambda > 1$ and that $\psi(x, \mu, E)$ is itself integrable over the whole energy range. Now define

$$\Psi(x, \mu, E) = \psi(x, \mu, E)/g(E). \tag{16}$$

Then $\mathcal{M}$ becomes

$$\bar{\psi}(x, \mu, \lambda) = \int_0^\infty \Psi(x, \mu, E) g(E) h(E)^{\lambda-1}\, dE. \tag{17}$$

Define the following change of variables:

$$V = h(E). \tag{18}$$

Normalize $h(E)$ such that $h(\infty) \equiv 1$. Then Eq. (18) defines a one-to-one mapping of

$$E \in [0, \infty] \quad \text{onto} \quad V \in [0, 1].$$

The mapping is one-to-one since the Jacobian of the transformation (18) is always different from zero:

$$\frac{dh}{dE} = g(E) \neq 0 \quad \text{for} \quad E \in]0, \infty[.$$

In terms of the new variable V, the transformation $\mathcal{M}$ can be rewritten as

$$\bar{\psi}(x, \mu, \lambda) = \int_0^1 \Psi(x, \mu, V) V^{\lambda-1}\, dV. \tag{19}$$

This is similar to a classical *Mellin transform* in terms of the new variable V with the exception that the integration range over V is restricted to [0, 1], instead of $[0, +\infty]$. The inversion formula is well known[27]:

$$\Psi(x, \mu, V) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \bar{\psi}(x, \mu, \lambda) V^{-\lambda}\, d\lambda, \tag{20a}$$

the integration path being to the right of all singularities of $\bar{\psi}(x, \mu, \lambda)$. Also,

$$\psi(x, \mu, E) = \frac{g(E)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \bar{\psi}(x, \mu, \lambda) h(E)^{-\lambda}\, d\lambda. \tag{20b}$$

We recall that, from the normalization of $h(E)$,

$$\Psi(x, \mu, V) \equiv 0, \quad \text{for} \quad V > 1.$$

This yields the following nontrivial property of the transformation $\mathcal{M}$, which is shown in Appendix A:

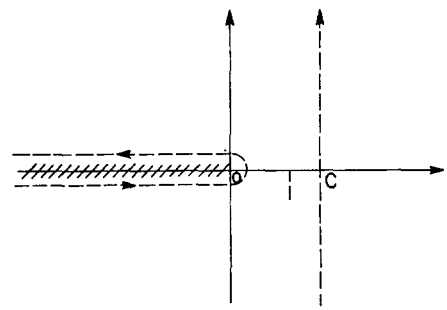


FIG. 2. The Bromwich contour in the complex λ plane.

Theorem I: $\bar{\psi}(x, \mu, \lambda)$ defined by Eq. (19) is uniformly bounded in λ, $\forall\lambda$ such that $\mathrm{Re}\,\lambda \geq 1$. As a consequence, the inversion formula (20a) yields an identically null function if $V > 1$.

If $V \leq 1$, the integration path of the inversion formula (20a) can be shifted along a Bromwich contour defined in Fig. 2, the first singularity on the real axis being $\lambda = 0$. [This can be deduced from careful inspection of Eq. (14) and its well-known solutions.]

C. Green's Function: Asymptotic Expressions

Through the use of the transformation $\mathcal{M}$ defined in Eq. (15), the slowing-down transport equation (8) has been reduced to a "one-speed" equation (14). The latter equation can be solved for a wide range of boundary conditions (full-space, half-space problems), using classical methods such as singular normal modes expansions; for calculational details, we refer to the literature.[1] Then one uses formula (20) to invert the $\mathcal{M}$ transformation and obtain the neutron distribution. As an example, we quote the exact expression for the full-space isotropic Green's function solution of Eq. (8), with the following source term:

$$S(x, \mu, E) = \tfrac{1}{2} S(E) \cdot \delta(x). \tag{21a}$$

The angle-integrated Green's function $G(|x|, E)$ is

$$G(|x|, E) = \frac{g(E)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left\{ \sum_{+\nu_j} \frac{e^{-|x|/\nu_j(\lambda)}}{N_j(\lambda)} \right\} \frac{\bar{S}(\lambda)}{2} \cdot h(E)^{-\lambda}\, d\lambda + \frac{g(E)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\bar{S}(\lambda)}{2} h(E)^{-\lambda}\, d\lambda \int_0^1 \frac{e^{-|x|/\nu}}{N(\nu, \lambda)}\, d\nu, \tag{21b}$$

where $\{\nu_j(\lambda)\}$ are the roots of

$$1 = \left(c_e + \frac{c_i}{\lambda}\right) \cdot \nu_j(\lambda) \cdot \tanh^{-1} \frac{1}{\nu_j(\lambda)} \tag{21c}$$

[27] A. Erdélyi, Ed., *Table of Integral Transforms, Vol. I* (McGraw-Hill Book Co., New York, 1954).

with

$$N_j(\lambda) = \frac{1}{2}\left(c_e + \frac{c_i}{\lambda}\right) \cdot \nu_j^3(\lambda) \cdot \left[\frac{\left(c_e + \frac{c_i}{\lambda}\right)}{\nu_j^2(\lambda) - 1} - \frac{1}{\nu_j^2(\lambda)}\right], \tag{21d}$$

$$N(\nu, \lambda) = \nu\left\{\left(1 - \left(c_e + \frac{c_i}{\lambda}\right)\nu \tanh^{-1}\nu\right)^2 + \frac{\pi^2}{4}\left(c_e + \frac{c_i}{\lambda}\right)^2\nu^2\right\}, \tag{21e}$$

$$\bar{S}(\lambda) = \mathcal{M}S(E). \tag{21f}$$

In Eq. (21b), the second term of the right-hand side corresponds to transport transients, and the first one is the (spatially) asymptotic component of the solution. The next problem is to find an asymptotic evaluation of the cumbersome contour integral in the complex λ plane. At first sight, use of the saddle-point method seems to be appropriate for the evaluation of such an integral; this is especially true if one introduces an *auxiliary* "*lethargy*" variable u associated with the inelastic slowing down and defined by

$$u = -\log[h(E)]. \tag{22a}$$

This is a one-to-one mapping of $E \in [0, +\infty]$ onto $u \in [+\infty, 0]$. Then, one gets contour integrals very similar to those encountered in spatially-dependent *elastic* slowing-down problems,[3–6,28] and classically evaluated by saddle-point methods. The position of the saddle-point λ_0 is given by the following equation[28]:

$$\left.\frac{d(1/\nu_0)(\lambda)}{d\lambda}\right|_{\lambda=\lambda_0} = \frac{u}{|x|}, \tag{22b}$$

where $\nu_0(\lambda)$ is the solution of Eq. (21c) with the largest absolute value. Unfortunately, the solution of the set of coupled implicit equations (21c) and (22b) is impracticable, unless one resorts to numerical tabulation.

Therefore, a mathematical method has been developed which yields an explicit asymptotic evaluation of formula (21b) valid for large distances, and which is briefly outlined in the next paragraph.

One can first simplify the Green's function through the following remark: Consider the full-space Green's function $G(x, \mu, E)$ solution of the ($\mathcal{M}$-transformed) equation

$$\mu\frac{\partial}{\partial x}\bar{G}(x, \mu, \lambda) + \bar{G}(x, \mu, \lambda) = \frac{1}{2}\left(c_e + \frac{c_i}{\lambda}\right)\int_{-1}^{+1}\bar{G}(x, \mu', \lambda)\,d\mu' + \tfrac{1}{2}\bar{S}(\lambda)\delta(x). \tag{23a}$$

[28] Ref. 24, p. 498ff.

Introduce the following associated equation, where the $\mathcal{M}$ transform of the source is constant:

$$\mu\frac{\partial}{\partial x}\bar{F}(x, \mu, \lambda) + \bar{F}(x, \mu, \lambda) = \frac{1}{2}\left(c_e + \frac{c_i}{\lambda}\right)\int_{-1}^{+1}\bar{F}(x, \mu', \lambda)\,d\mu' + \frac{\delta(x)}{2}. \tag{23b}$$

Then, in terms of the energy variable, one gets the following relation between $G(x, \mu, E)$, $F(x, \mu, E)$,and $S(E)$:

$$G(x, \mu, E) = g(E)\int_V^1 \{S/g\}\left(\frac{V}{w}\right)\cdot\frac{F(x, \mu, w)}{g(w)}\frac{dw}{w}, \tag{24a}$$

where

$$V = h(E), \qquad w = h(E'), \tag{24b}$$

and where $\{S/g\}(V/w)$ is the value of the function $S(E)/g(E)$ for E such that $h(E) = V/w$. Relation (24) is nothing but the "Faltung theorem" for the inverse Mellin transform,[29] similar to the convolution theorem for Laplace and Fourier transforms. Such a theorem is immediately extended to the inverse $\mathcal{M}$ transformation, which is closely related to Mellin transforms. In view of relation (24), it is sufficient to find an asymptotic evaluation of the energy Green's function $F(x, \mu, E)$ [Eq. (23b)].

The crux of the method consists in considering the inelastic scattering term in Eq. (23b) as an auxiliary source term:

$$\mu\frac{\partial}{\partial x}\bar{F}(x, \mu, \lambda) + \bar{F}(x, \mu, \lambda) = \frac{c_e}{2}\int_{-1}^{+1}\bar{F}(x, \mu', \lambda)\,d\mu' + \frac{1}{2}\left\{\delta(x) + \frac{c_i}{2\lambda}\int_{-1}^{+1}\bar{F}(x, \mu', \lambda)\,d\mu'\right\}. \tag{25}$$

Then one applies a spatial Fourier transformation to Eq. (25), and obtains a simple expression for the Fourier transform of $\bar{F}(x, \mu, \lambda)$ in terms of the well-known solution of the one-speed Boltzmann equation associated to a multiplication coefficient c_e. At this stage, in contrast to the normal-modes approach, inverse $\mathcal{M}$ transformation is immediate, and one is left with the asymptotic evaluation of $F(x, \mu, E)$ for large distances, knowing its x Fourier transform. Analytical details, being quite lengthy, are found in Appendix B, with the final asymptotic expression for the full-space, energy Green's-function solution of the inelastic slowing-down Boltzmann equation.

[29] Ref. 27, Vol. I, p. 308, relation 14.

III. ENERGY-DEPENDENT BOLTZMANN EQUATION WITH FISSION AND SLOWING-DOWN KERNELS

The addition of a fission projection kernel to the transport equation is of prime importance when it comes to studying fast multiplying systems. Physically, the classical slowing-down problem of thermal reactor theory is changed to a situation with simultaneous neutron degradation and regeneration; this may allow *self-sustaining modes*. Mathematically, this implies that the sum of a fission and slowing-down operators is likely to have a *discrete*, regular eigenfunction, which is not true for the plain slowing-down kernel.

Nevertheless, little work has been done up to now in studying simultaneous fission and slowing down in the *transport* equation. Diffusion approximations or multigroup (discrete-energy) schemes have been the rule[23]—or the problem has been reduced to a plain slowing-down situation by assimilating the fission sources to a high-energy Dirac distribution.[23]

We are interested in the following equation, where energy transfer occurs only through fission and *inelastic* scattering:

$$\mu \frac{\partial}{\partial x}\psi(x,\mu,E) + \psi(x,\mu,E) = \frac{c_e}{2}\int_{-1}^{+1}\psi(x,\mu',E)\,d\mu' + \frac{c_i}{2}\int_{-1}^{+1}d\mu' g(E)\int_E^\infty \frac{\psi(x,\mu',E')\,dE'}{h(E')} + \frac{c_F}{2}\chi(E)\int_{-1}^{+1}d\mu'\int_0^\infty \psi(x,\mu',E')\,dE' + \tfrac{1}{2}\delta(x-x_0)S(E). \quad (26)$$

The notation is the same as in Eq. (1); $\chi(E)$ is the fission spectrum, and c_F is the mean number of secondaries emitted after a fission collision times the probability of fission. In a previous work,[22] we solved an identical problem with *elastic* slowing down. In this work, quite similar results will be found for Eq. (26). Define the global energy-transfer operator $\mathcal{O}$ by

$$\mathcal{O}\phi(E) = c_F\chi(E)\int_0^\infty \phi(E')\,dE' + c_e\phi(E) + c_i g(E)\int_E^\infty \frac{\phi(E')\,dE'}{h(E')}. \quad (27a)$$

Eigenfunctions of $\mathcal{O}$ are such that

$$\mathcal{O}\phi_\nu(E) = \nu\phi_\nu(E). \quad (27b)$$

At this point, we note the conditions for a null integral.

Lemma: The necessary and sufficient condition for a function $\phi(E) \in L^1[0,\infty]$ to have a null integral,

$$\int_0^\infty \phi(E')\,dE' = 0,$$

is that

$$\bar{\phi}(1) \equiv 0,$$

where $\bar{\phi}(\lambda)$ is the $\mathcal{M}$ transform of $\phi(E)$. [This stems from the set of reciprocal formulas (15) and (20b) for the $\mathcal{M}$ transform; $\bar{\phi}(\lambda) = \int_0^\infty \phi(E)h(E)^{\lambda-1}\,dE$.]

Coming back to Eq. (27b), its $\mathcal{M}$ transform is

$$\nu\bar{\phi}_\nu(\lambda) = \left(c_e + \frac{c_i}{\lambda}\right)\bar{\phi}_\nu(\lambda) + c_F\bar{\chi}(\lambda)\bar{\phi}_\nu(1). \quad (28)$$

Solutions of Eq. (28) belong to two categories:

1. For solutions such that $\phi_\nu(1) \neq 0$, or

$$\int_0^\infty \phi_\nu(E)\,dE \neq 0,$$

there is a unique eigenvalue,

$$\nu = c_F + c_e + c_i, \quad (29a)$$

to which corresponds a single regular eigenfunction of the operator $\mathcal{O}$:

$$\overline{\mathcal{H}}(\lambda) = c_F\bar{\chi}(\lambda)\Big/\left[c_F + c_i\left(1 - \frac{1}{\lambda}\right)\right] \quad (29b)$$

[keeping in mind that $\bar{\chi}(1) = \int_0^\infty \chi(E)\,dE = 1$].

From Eq. (29), or from direct solution of Eq. (27b) (by reduction to a differential equation), one gets the following expression for $\mathcal{H}(E)$:

$$\mathcal{H}(E) = \frac{c_F}{c_F + c_i}\chi(E) + \frac{c_i}{c_F + c_i}g(E)h(E)^{-c_i/(c_F+c_i)} \times \int_E^\infty \frac{c_F}{c_F + c_i}\frac{\chi(E')}{h(E')^{c_F/(c_F+c_i)}}\,dE'. \quad (30)$$

The regular eigenfunction $\mathcal{H}(E)$ of the energy-transfer operator in Eq. (26) corresponds to the *asymptotic neutron energy spectrum.*

2. For solutions of Eq. (28) such that $\bar{\phi}_\nu(1) = 0$ or $\int_0^\infty \phi_\nu(E)\,dE = 0$, then Eq. (28) reduces to

$$\nu\bar{\phi}_\nu(\lambda) = \left(c_e + \frac{c_i}{\lambda}\right)\bar{\phi}_\nu(\lambda), \quad (31a)$$

the solutions of which are singular[30]:

$$\bar{\phi}_\nu(\lambda) = \delta(\lambda - \lambda_0), \quad \text{with} \quad \lambda_0 \neq 1, \quad (31b)$$

$$\nu = c_e + (c_i/\lambda_0). \quad (31c)$$

[30] L. Schwartz, *Théorie des distributions* (Hermann & Cie, Paris, 1966).

These distributions correspond to the ordinary slowing-down operator.

The regular eigenfunction $\mathcal{K}(E)$, Eq. (30), and the continuum of eigendistributions defined in Eq. (31) form a *complete* set in $L^1[0, \infty]$; this is expressed by the following theorem.

Theorem II: $\forall \phi(E) \in L^1[0, \infty]$ has the *unique* decomposition

$$\phi(E) = \mathfrak{S} \cdot \mathcal{K}(E) + \frac{g(E)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \bar{A}(\lambda) h(E)^{-\lambda} \, d\lambda$$

with $\bar{A}(1) \equiv 0$.

Proof: Since $\int_0^\infty \mathcal{K}(E) \, dE = 1$, one must have

$$\mathfrak{S} = \int_0^\infty \phi(E) \, dE.$$

Then, defining $\Gamma(E)$,

$$\Gamma(E) = \phi(E) - \mathfrak{S} \cdot \mathcal{K}(E),$$

one has

$$\int_0^\infty \Gamma(E) \, dE = 0;$$

and one must prove that $\Gamma(E)$ admits the representation

$$\Gamma(E) = \frac{g(E)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \bar{A}(\lambda) h(E)^{-\lambda} \, d\lambda.$$

The existence and uniqueness of such a representation is immediate, since $\Gamma(E)$ is $\mathcal{M}$ transformable:

$$\bar{A}(\lambda) = \int_0^\infty \Gamma(E) h(E)^{\lambda - 1} \, dE.$$

Finally, from the lemma,

$$\bar{A}(1) \equiv 0. \qquad \text{Q.E.D.}$$

Coming back to the transport equation (26), we see that if the source term were of the form

$$S(E) = \mathcal{K}(E),$$

then Eq. (26) would reduce to a one-speed equation, with a multiplication coefficient $c_F + c_e + c_i$, since the solution is separable into a function of space and angle times $\mathcal{K}(E)$.

For a general source term, the method consists in applying the expansion of theorem II to $S(E)$:

$$S(E) = \mathfrak{S} \cdot \mathcal{K}(E) + \Gamma(E), \tag{32a}$$

where

$$\mathfrak{S} = \int_0^\infty S(E) \, dE, \tag{32b}$$

$$\Gamma(E) = \frac{g(E)}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(\lambda) h(E)^{-\lambda} \, d\lambda, \tag{32c}$$

$$\Gamma(\lambda) = \int_0^\infty h(E)^{\lambda - 1} \Gamma(E) \, dE, \tag{32d}$$

$$\Gamma(1) = \int_0^\infty \Gamma(E) \, dE = 0. \tag{32e}$$

Since the transport equation (26) is linear, its solution can be expressed as the one-speed solution due to $\mathfrak{S} \cdot \mathcal{K}(E)$ plus the solution due to a source $\Gamma(E)$ of null integral. Call the former solution $\mathcal{K}(E) \cdot \phi_E(x, \mu)$ and the latter $\phi_{tr}(x, \mu, E)$:

$$\psi(x, \mu, E) = \mathcal{K}(E) \phi_E(x, \mu) + \phi_{tr}(x, \mu, E). \tag{33}$$

Explicitly, $\phi_E(x, \mu)$ is the solution of

$$\mu \frac{\partial}{\partial x} \phi_E(x, \mu) + \phi_E(x, \mu) = \frac{c_e + c_i + c_F}{2} \int_{-1}^{+1} \phi_E(x, \mu') \, d\mu' + \tfrac{1}{2} \delta(x - x_0) \cdot \mathfrak{S}. \tag{34}$$

As to $\phi_{tr}(x, \mu, E)$,

$$\mu \frac{\partial}{\partial x} \phi_{tr}(x, \mu, E) + \phi_{tr}(x, \mu, E) = \frac{c_e}{2} \int_{-1}^{+1} \phi_{tr}(x, \mu', E) \, d\mu' + \frac{c_i}{2} \int_{-1}^{+1} d\mu' g(E) \int_E^\infty \frac{\phi_{tr}(x, \mu', E')}{h(E')} \, dE' + \frac{\Gamma(E)}{2} \delta(x - x_0). \tag{35}$$

The crucial point is that $\phi_{tr}(x, \mu, E)$ is the solution of a *plain slowing-down* equation without any fission term; since the integral of $\Gamma(E)$ over the whole energy range is null, the same follows for $\phi_{tr}(x, \mu, E)$. To prove it, take the $\mathcal{M}$ transform of Eq. (35):

$$\mu \frac{\partial}{\partial x} \bar{\phi}_{tr}(x, \mu, \lambda) + \bar{\phi}_{tr}(x, \mu, \lambda) = \frac{1}{2} \left(c_e + \frac{c_i}{\lambda} \right) \int_{-1}^{+1} \bar{\phi}_{tr}(x, \mu', \lambda) \, d\mu' + \frac{\Gamma(\lambda)}{2} \delta(x - x_0). \tag{36}$$

Keeping in mind that $\Gamma(1) \equiv 0$ and that the $\mathcal{M}$-transformed Eq. (36) is homogeneous in λ, it follows that

$$\bar{\phi}_{tr}(x, \mu, \lambda) \equiv 0 \quad \text{for} \quad \lambda = 1.$$

The success of the decomposition of Eq. (26) into the associated Eqs. (34) and (35) had to be expected, since one has in fact made an expansion of the source $S(E)$ with the set of "eigenfunctions" of the global energy-transfer operator.

The solutions of the one-speed equation (34) are well known[1]; Eq. (35) is a plain slowing-down equation solved in Sec. II. Thus, the present problem is completely solved.

The Milne problem and criticality problems involve only Eq. (34). But the full-space and half-space Green's functions involve both solutions of Eq. (34) (space–energy separable modes) and Eq. (35) (nonseparable transients). The space–energy separable components are proportional to the characteristic energy-mode $\mathcal{K}(E)$; they are representative of self-sustaining modes in the fast-multiplying medium, and are asymptotically dominant. The nonseparable modes are "slowing down transient," solutions of an ordinary slowing-down equation; they are *not* classical one-speed "singular transport transients" (they may decay more slowly than $e^{-|x|}$, as shown in Appendix B on the asymptotic evaluation of the slowing-down Green's function), and they represent the adjustment of the neutron field from the initial source-energy distribution to the final self-sustaining asymptotic spectrum. They are likely to delay the approach to equilibrium in integral experiments on fast systems (for instance, exponential experiments). The relative importance of space–energy separable modes and "slowing-down transients" is quite sensitive to the degree of criticality of the fast system. This, in turn, limits the validity of asymptotic transport theory for the energy-dependent Boltzmann equation, since its basic assumption is space–energy separability, which leads to the omission of all "slowing-down transients."

Similar results have been found for the case of fission with *anisotropic elastic* slowing down.[22] In conclusion, in order to achieve completeness for the normal-modes solution of the Boltzmann equation with fission and slowing down, one must consider fundamental separable modes reflecting the multiplicative process, together with "slowing-down transients," solution of an ordinary slowing-down equation. In a further paper, numerical results will be presented on the relative importance of asymptotic separable modes and "slowing-down transients," in the approach to equilibrium in exponential experiments on fast-neutron-multiplying media.

APPENDIX A: PROOF OF THEOREM I

This theorem (Sec. II.B) states that $\bar{\psi}(x, \mu, \lambda)$, the $\mathcal{M}$ transform of $\psi(x, \mu, E)$, is uniformly bounded in λ, for $\forall\lambda$ such that $\operatorname{Re}\lambda \geq 1$.

Proof:

$$\bar{\psi}(x, \mu, \lambda) = \int_0^1 \Psi(x, \mu, V) V^{(\operatorname{Re}\lambda)-1} V^{i \operatorname{Im}\lambda}\, dV \quad \text{(A1)}$$

[cf. Eqs. (16)–(19)];

$$|\bar{\psi}(x, \mu, \lambda)| \leq \int_0^1 |\Psi(x, \mu, V)| \cdot |V^{(\operatorname{Re}\lambda)-1}|\, dV, \quad \text{(A2)}$$

but

$$\left.\begin{aligned} 0 \leq V \leq 1 \\ (\operatorname{Re}\lambda) - 1 \geq 0 \end{aligned}\right\} \Rightarrow |V^{(\operatorname{Re}\lambda)-1}| \leq 1. \quad \text{(A3)}$$

So

$$|\bar{\psi}(x, \mu, \lambda)| \leq \int_0^1 |\Psi(x, \mu, V)|\, dV \leq M, \quad \text{(A4)}$$

where

$$M = \int_0^\infty |\psi(x, \mu, E)|\, dE. \quad \text{(Q.E.D.)} \quad \text{(A5)}$$

APPENDIX B: ASYMPTOTIC EVALUATION OF THE FULL-SPACE GREEN'S FUNCTION

Using the notation of Sec. II.C, the full-space, energy Green's function $F(x, \mu, E)$ obeys the following $\mathcal{M}$-transformed equation:

$$\mu \frac{\partial}{\partial x} \bar{F}(x, \mu, \lambda) + \bar{F}(x, \mu, \lambda) = \frac{c_e}{2} \int_{-1}^{+1} \bar{F}(x, \mu', \lambda)\, d\mu' + \frac{1}{2}\left\{\delta(x) + \frac{c_i}{\lambda} \int_{-1}^{+1} \bar{F}(x, \mu', \lambda)\, d\mu'\right\}, \quad \text{(25)}$$

where the inelastic scattering term is considered as an extraneous source. Define

$$\bar{F}(|x|, \lambda) = \int_{-1}^{+1} \bar{F}(x, \mu', \lambda)\, d\mu'. \quad \text{(B1)}$$

Then call $G_e(|x|)$ the Green's function corresponding to the one-speed transport equation with the "multiplication coefficient" c_e (elastic scattering):

$$\mu \frac{\partial}{\partial x} G_e(x, \mu) + G_e(x, \mu) = \frac{c_e}{2} \int_{-1}^{+1} G_e(x, \mu')\, d\mu' + \frac{\delta(x)}{2}, \quad \text{(B2)}$$

where

$$G_e(|x|) = \int_{-1}^{+1} G_e(x, \mu')\, d\mu'. \quad \text{(B3)}$$

The exact expression of $G_e(|x|)$ is well known:

$$G_e(|x|) = \frac{1}{2} \frac{dK_e^2}{dc_e} \frac{e^{-K_e|x|}}{K_e} + \frac{1}{2} \int_0^1 \frac{e^{-|x|/\nu}}{\nu N(c_e, \nu)}\, d\nu, \quad \text{(B4a)}$$

where K_e is the root of

$$1 = \frac{c_e}{K_e} \tanh^{-1} K_e, \quad \text{(B4b)}$$

and

$$N(c_e, \nu^2) = \left\{(1 - c_e\nu \tanh^{-1}\nu)^2 + \frac{\pi^2\nu^2c_e^2}{4}\right\}. \quad \text{(B4c)}$$

Now, apply a *Fourier* spatial transformation to Eq. (25):

$$\bar{F}(K^2, \lambda) = \int_{-\infty}^{+\infty} F(|x|, \lambda)e^{-iKx}\,dx, \quad \text{(B5a)}$$

$$\bar{G}_e(K^2) = \int_{-\infty}^{+\infty} G_e(|x|)e^{-iKx}\,dx. \quad \text{(B5b)}$$

We get

$$(iK\mu + 1)\bar{F}(K, \mu, \lambda) = \frac{c_e}{2}\int_{-1}^{+1} \bar{F}(K, \mu', \lambda)\,d\mu' + \frac{1}{2}\left\{1 + \frac{c_i}{\lambda}\bar{F}(K^2, \lambda)\right\}. \quad \text{(B5c)}$$

Since Eq. (B5c) is *homogeneous in both variables K and* λ, this yields

$$\bar{F}(K^2, \lambda) = \bar{G}_e(K^2)\left\{1 + \frac{c_i}{\lambda}\bar{F}(K^2, \lambda)\right\} \quad \text{(B6)}$$

and

$$\bar{F}(K^2, \lambda) = \frac{\bar{G}_e(K^2)}{1 - (c_i/\lambda)\bar{G}_e(K^2)}. \quad \text{(B7)}$$

In Eq. (7), let us isolate the term $\bar{G}_e(K^2)$:

$$\bar{F}(K^2, \lambda) = \bar{G}_e(K^2) + \frac{c_i\{\bar{G}_e(K^2)\}^2}{\lambda - c_i\bar{G}_e(K^2)}. \quad \text{(B8)}$$

Recall that[23]

$$\bar{G}_e(K^2) = \frac{dK_e^2}{dc_e}\frac{1}{K_e^2 + K^2} + \int_0^1 \frac{d\nu}{(1 + K^2\nu^2)N(c_e, \nu^2)}. \quad \text{(B9)}$$

At this stage, the inverse $\mathcal{M}$ transformation of Eq. (B8) is immediate; keeping in mind that

$$\mathcal{M}\{g(E)h(E)^{-\rho}\} = \int_0^\infty g(E)h(E)^{\lambda-\rho-1}\,dE = \frac{1}{\lambda - \rho}, \quad \text{(B10)}$$

we obtain

$$\bar{F}(K^2, E) = \bar{G}_e(K^2)\delta(u) + c_i g(E)\{\bar{G}_e(K^2)\}^2 h(E)^{-c_i\bar{G}_e(K^2)}, \quad \text{(B11)}$$

where $\delta(u)$ is a Dirac distribution, and u is the "lethargy" defined in Eq.(22a),

$$u = -\log[h(E)]. \quad \text{(22a)}$$

In relation (B11), the *exact* inverse $\mathcal{M}$ transformation has been successfully performed for the energy Green's function of the infinite medium; this expression is valid for *all energies*. The first term in Eq. (B11) corresponds to the elastic scattering of the source.

The next step is to find an asymptotic expression for the spatial behavior of $F(|x|, E)$. Apply an inverse Fourier transformation to Eq. (B11):

$$F(|x|, E) = G_e(|x|)\cdot\delta(u) + \frac{c_i\cdot g(E)}{2\pi} \times \int_{-\infty}^{+\infty}\{\bar{G}_e(K^2)\}^2 e^{uc_i\bar{G}_e(K^2)}e^{+iKx}\,dK. \quad \text{(B12)}$$

In Eq. (B12), the first term $G_e(|x|)\cdot\delta(u)$ is perfectly well known [Eq. (B4)]. One is left with the evaluation of

$$R(|x|, E) = \frac{c_i g(E)}{2\pi}\int_{-\infty}^{+\infty}\{\bar{G}_e(K^2)\}^2 e^{\{uc_i\bar{G}_e(K^2)+iKx\}}\,dK. \quad \text{(B13)}$$

Make the following change of variable:

$$ik = K. \quad \text{(B14)}$$

Equation (B13) becomes

$$R(|x|, E) = \frac{c_i g(E)}{2\pi i}\int_{-i\infty}^{+i\infty}\{\bar{G}_e(k^2)\}^2 e^{\{uc_i\bar{G}_e(k^2)-kx\}}\,dk, \quad \text{(B15)}$$

where [see Eq. (B4) and (B9)]

$$\bar{G}_e(k^2) = \frac{dK_e^2}{dc_e}\frac{1}{K_e^2 - k^2} + \int_0^1 \frac{d\nu}{(1 - k^2\nu^2)N(c_e, \nu^2)} = \frac{dK_e^2}{dc_e}\frac{1}{K_e^2 - k^2} + \Omega(k^2). \quad \text{(B16)}$$

Equation (B15) involves a contour integral along the imaginary axis for a function of the complex variable k; the latter is analytic everywhere, except for

(1) the cuts $[+1, +\infty]$ and $[-\infty, -1]$ on the real axis, since these are cuts for $\Omega(k^2)$ in Eq. (B16);

(2) the *essential* singularities $k = \pm K_e$ on the real axis; this is due to the exponential blowup of the term

$$\exp uc_i\bar{G}_e(k^2),$$

which behaves as $e^{1/k\pm K_e}$ in the neighborhood of the essential singularities $\pm K_e$.

Let us now shift the integration contour in Eq. (B15) from the imaginary axis to the real one. For positive values of x, the corresponding Bromwich contour lies in the positive half-plane, since for $\mathrm{Re}\,(k) > 0$ and $x > 0$, one has $|e^{-kx}| \to 0$. Then, for $x > 0$,

$$R(x, E) = \frac{c_i g(E)}{2\pi i}\left\{\int_D + \int_C\{\bar{G}_e(k^2)\}^2 e^{\{uc_i\bar{G}_e(k^2)-kx\}}\,dk\right\} \quad \text{(B17)}$$

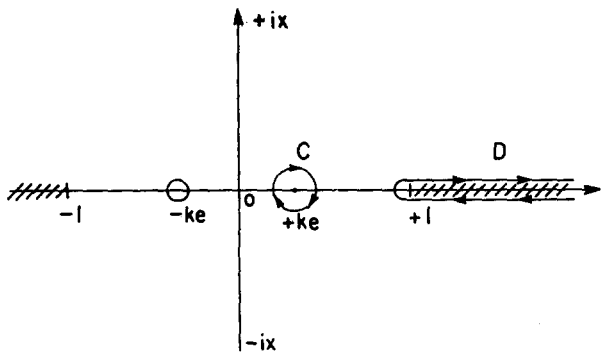


FIG. 3. The Bromwich contour for $R(x, E) - x > 0$—in the k complex plane.

(see Fig. 3). It is clear that the contribution of the contour D along the cut $[+1, +\infty]$ involves modes all decaying faster than e^{-x}; these are *transport-spatial transients*. All modes decaying more slowly than $e^{-|x|}$ are yielded by the contour integral C around the essential singularity $+K_e$; this is indeed the spatial asymptotic component one is looking for.

So, define

$$R_{as}(x, E) = \frac{c_i g(E)}{2\pi i} \oint_c \{\bar{G}_e(k^2)\}^2 e^{\{uc_i \bar{G}_e(k^2) - kx\}}\, dk. \quad \text{(B18)}$$

From Eq. (B16),

$$\{\bar{G}_e(k^2)\}^2$$

is

$$\{\bar{G}_e(k^2)\}^2 = \left(\frac{dK_e^2}{dc_e}\right)^2 \frac{1}{(K_e^2 - k^2)^2} + \frac{dK_e^2}{dc_e} \cdot \frac{2\Omega(k^2)}{K_e^2 - k^2} + \{\Omega(k^2)\}^2. \quad \text{(B19)}$$

$\Omega(K^2)$ is an analytic function on the contour C and within the domain surrounded by C. Using the three components of $\{G_e(k^2)\}^2$ in Eq. (B19), we can split $R_{as}(x, E)$ into three parts:

$$R_{as}(x, E) = R_{as}^{(1)}(x, E) + R_{as}^{(2)}(x, E) + R_{as}^{(3)}(x, E).$$

Since the procedure of next calculations is quite similar for these three parts, we outline them for $R_{as}^{(1)}(x, E)$:

$$R_{as}^{(1)}(x, E) = \frac{c_i g(E)}{2\pi i} \times \oint_c \left(\frac{dK_e^2}{dc_e}\right)^2 \frac{1}{(K_e^2 - k^2)^2} e^{\{uc_i \bar{G}_e(k^2) - kx\}}\, dk; \quad \text{(B20)}$$

this can be rewritten as

$$R_{as}^{(1)}(x, E) = \frac{c_i g(E)}{2\pi i} \oint_c \frac{L(k, u)}{(K_e - k)^2} \times \exp\left\{\frac{uc_i}{2K_e} \frac{dK_e^2}{dc_e} \frac{1}{K_e - k} - kx\right\} dk, \quad \text{(B21)}$$

where we have defined

$$L(k, u) = \left\{\frac{dK_e^2}{dc_e}\right\}^2 \frac{1}{(K_e + k)^2} \times \exp\left[(uc_i/2K_e)(dK_e^2/dc_e) \times (1/K_e + k) + uc_i\Omega(k^2)\right]. \quad \text{(B22)}$$

$L(k, u)$ is an analytic function of k on the contour C and within the domain surrounded by C. This analyticity of $L(k, u)$ enables us to replace it by $L(K_e, u)$ in the contour integral (B21), and obtain an asymptotic evaluation of $R_{as}^{(1)}x, E)$ in the same way, valid for large x:

$$L(K_e, u) = \frac{1}{4K_e^2}\left\{\frac{dK_e^2}{dc_e}\right\}^2 e^{uc_i\{(1/4Ke^2)(dKe^2/dce)+\Omega(Ke^2)\}}, \quad \text{(B23)}$$

$$R_{as}^{(1)}(x, E) \sim \frac{c_i g(E)}{2\pi i} L(K_e, u) \times \oint_c \frac{e^{\{uc_i \cdot (dKe^2/dce) \cdot 1/(Ke-k) - kx\}}}{(K_e - k)^2}\, dk. \quad \text{(B24)}$$

The idea is then to reduce the contour integral in (B24) to a *classical inverse Laplace transform in* x; for this purpose, put

$$K_e - k = p, \quad \text{(B25)}$$

$$\alpha = \frac{uc_i}{2K_e} \cdot \frac{dK_e^2}{dc_e}. \quad \text{(B26)}$$

Then Eq. (B24) reduces to

$$R_{as}^{(1)}(x, E) \sim e^{-K_e x} \cdot L(K_e^2, u) \cdot \frac{c_i g(E)}{2\pi i} \oint \frac{e^{\alpha/p}}{p^2} e^{px}\, dp \quad \text{(B27)}$$

(with counterclockwise integration, this time).

In Eq. (B27), one recognizes the following inverse Laplace transform:

$$\frac{1}{2\pi i} \oint_c \frac{e^{\alpha/p}}{p^2} e^{px}\, dp = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{\alpha/p}}{p^2} e^{px}\, dp. \quad \text{(B28)}$$

But, from Bateman's *Table of Integral Transforms*,[31]

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} p^{-\nu-1} e^{\alpha/p} e^{px}\, dp = \alpha^{-\nu/2} x^{+\nu/2} I_\nu[2(\alpha x)^{\frac{1}{2}}], \quad \text{(B29)}$$

where I_ν is the hyperbolic Bessel function of order ν.[27] One then obtains the final expression for $R_{as}^{(1)}(x, E)$, for large $x > 0$:

$$R_{as}^{(1)}(x, E) \sim c_i g(E) L(K_e^2, u) \cdot e^{-K_e x} (x/\alpha)^{\frac{1}{2}} I_1[2(\alpha x)^{\frac{1}{2}}]. \quad \text{(B30)}$$

[31] Ref. 27, Vol. I, p. 245, relation 35.

A similar procedure can be applied to $R_{as}^{(2)}(x, E)$ and $R_{as}^{(3)}(x, E)$; one uses the following set of inverse Laplace transforms:

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{1}{p}e^{\alpha/p}e^{px}\,dp = I_0[2(\alpha x)^{\frac{1}{2}}], \tag{B31a}$$

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}e^{\alpha/p}e^{px}\,dp = \delta(x) + \left(\frac{\alpha}{x}\right)^{\frac{1}{2}}I_1[2(\alpha x)^{\frac{1}{2}}]. \tag{B31b}$$

So, omitting the detailed calculations, one can write the final expression for $F_{as}(|x|, E)$, the spatially asymptotic part of the infinite-space, energy Green's function, as

$$\begin{aligned}F_{as}(|x|, E) &= \frac{1}{2}\frac{dK_e^2}{dc_e}\cdot\frac{e^{-K_e|x|}}{K_e}\cdot\delta(u) + c_i g(E)\\ &\times \exp\{uc_i[(1/4K_e^2)(dK_e^2/dc_e) + \Omega(K_e^2)]\}e^{-K_e|x|}\\ &\times\left\{\frac{1}{4}\left(\frac{1}{K_e}\frac{dK_e^2}{dc_e}\right)^2\left(\frac{|x|}{\alpha}\right)^{\frac{1}{2}}I_1[2(\alpha|x|)^{\frac{1}{2}}]\right.\\ &+\frac{1}{K_e}\cdot\frac{dK_e^2}{dc_e}\cdot\Omega(K_e^2)I_0[2(\alpha|x|)^{\frac{1}{2}}]\\ &\left.+(\Omega(K_e^2))^2\left(\frac{\alpha}{|x|}\right)^{\frac{1}{2}}I_1[2(\alpha|x|)^{\frac{1}{2}}]\right\}.\end{aligned} \tag{B32}$$

$\Omega(K^2)$ is defined in Eq. (B16), K_e in Eq. (B4), α (which is a function of u) in Eq. (B26). The expression (B32) holds for intermediate and large distances; numerical calculations have shown it to be quite accurate at distances beyond 2–3 mean free paths, in typical fast systems. It is readily seen that $F_{as}(|x|, E)$ is split into two parts: one which decays as $e^{-K_e|x|}$ and corresponds to plain elastic scattering of the source term; the second one, $R_{as}(|x|, E)$, which includes all inelastic scattering effects. Recalling that

$$I_0(0) = 1, \quad I_1(0) = 0,$$

and that hyperbolic Bessel functions are monotonically increasing, we see that $R_{as}(|x|, E)$ decays more slowly than $\exp(-K_e|x|)$. More precisely, making an asymptotic expansion of the hyperbolic Bessel functions,[27]

$$I_n[2(\alpha|x|)^{\frac{1}{2}}]_{|x|\to\infty} \sim \frac{1}{(4\pi)^{\frac{1}{2}}(\alpha|x|)^{\frac{1}{4}}}e^{2(\alpha|x|)^{\frac{1}{2}}}, \tag{B33}$$

we obtain the following expression for $R_{as}(|x|, E)$:

$$\begin{aligned}R_{as}(|x|, E)_{|x|\to\infty} &\sim c_i g(E)e^{uc_i\{(1/4K_e^2)(dK_e^2/dc_e)+\Omega(K_e^2)\}}\\ &\times 1/(4\pi)^{\frac{1}{2}}(\alpha|x|)^{\frac{1}{4}}e^{-K_e|x|+2(\alpha|x|)^{\frac{1}{2}}}\\ &\times\left\{\frac{1}{4}\left(\frac{1}{K_e}\frac{dK_e^2}{dc_e}\right)^2\cdot\left(\frac{|x|}{\alpha}\right)^{\frac{1}{2}} + \frac{1}{K_e}\frac{dK_e^2}{dc_e}\Omega(K_e^2)\right.\\ &\left.+\{\Omega(K_e^2)\}^2\left(\frac{\alpha}{|x|}\right)^{\frac{1}{2}}\right\}.\end{aligned} \tag{B34}$$

In Eq. (B34), the leading term is

$$\exp[-K_e|x| + 2(\alpha|x|)^{\frac{1}{2}}]. \tag{B35}$$

Since α is linearly increasing with the "lethargy" u [see Eq. (B26)], $F_{as}(|x|, E)$ will decay with space more slowly for low energies than for high energies. This has been verified by numerical calculations, which agrees also with the measurements of apparent relaxation lengths in the natural uranium exponential experiment[32]; in this case, one has nearly a pure slowing-down situation, and apparent relaxation lengths for high-energy neutrons ($\geq$1.0 MeV) are systematically smaller than for low-energy neutrons ($\leq$0.5 MeV).

ACKNOWLEDGMENTS

The author wishes to express his appreciation to the Michigan Memorial-Phoenix Project and to the National Science Foundation for financial support. Also, he expresses his appreciation to his thesis adviser, Professor P. F. Zweifel.

[32] J. L. Campan, P. P. Clauzon, C. P. Zaleski, "Etude du flux de neutrons à l'équilibre dans l'Uranium naturel," *Proceedings of the International Symposium on Exponential Experiments* (IAEA, Vienna, 1961), Vol. I, p. 345

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11 NUMBER 1 JANUARY 1970

Asymptotic Fields in Some Models of Quantum Field Theory. III

RAPHAEL HØEGH-KROHN*
Institute of Mathematics, University of Oslo

(Received 26 December 1967; Final Manuscript Received 29 April 1969)

We consider the model of a relativistic boson field, interacting in a local way with a nonrelativistic nucleon field. This model was studied by Nelson, who proved that after an infinite nucleon self-energy renormalization, the total energy becomes a self-adjoint operator. In this paper we prove that the asymptotic limits of the boson field exist.

1. INTRODUCTION

The model of a relativistic boson field interacting by a Yukawa coupling with a nonrelativistic nucleon field was studied by Nelson.[1] By introducing a momentum cutoff in the interaction, he obtains a self-adjoint cutoff energy operator. Using a canonical transformation due to Gross, Nelson is able to separate out a nucleon self-energy term which diverges as the cutoff is taken away. He then proves that the rest of the energy operator converges to a self-adjoint operator as the cutoff tends to infinity. The field theoretic properties of this model was studied by Cannon in his thesis.[2]

The object of this paper is to study the asymptotic limits of the quantum fields of this model. The method used is somewhat related to the method in two earlier papers[3] with the same title as this one. Here, the fact that we are dealing with a boson field leads to some complications, but these complications are already dealt with in a paper on a general class of cutoff boson interactions.[4] The basic idea of the method used to establish the existence of asymptotic fields is an adaption of Cook's method, and it was first used by Kato and Mugibayashi.[5]

2. DESCRIPTION OF THE MODEL

Our Hilbert space $\mathcal{H}$ is the tensor product of the space of nucleon wavefunctions and the Fock space of the boson field. The nucleons will be nonrelativistic particles interacting with the relativistic boson field in a local way. For the sake of simplicity we will take the nucleons to be fermi particles, although we could just as well impose boson statistics or no statistics at all on the nucleons.

An element f in $\mathcal{H}$ is represented by a sequence $\{f_n\}$ of functions $f_n(p_1, \cdots, p_n \mid x_1, \cdots, x_N)$ of $n + N$ variables p_i, x_j in R^3 such that f_n is symmetrical in $p_1, \cdots, p_n$ and antisymmetrical in $x_1, \cdots, x_N$. N is the number of nucleons, and this number will be kept fixed. The inner product in $\mathcal{H}$ is given by

$$(f, g) = \sum_{n=0}^{\infty} n! \int \cdots \int \bar{f}_n(p_1, \cdots, p_n \mid x_1, \cdots, x_N) \times g_n(p_1, \cdots, p_n \mid x_1, \cdots, x_N) \times dp_1 \cdots dp_n \, dx_1 \cdots dx_n.$$

Let M and μ be two strictly positive constants. M is the nucleon mass and μ the mass of the boson. Let Δ be the Laplacian in three dimensions regarded as a self-adjoint operator on its natural domain of definition, $D_\Delta \subset L_2(R^3)$, and let $\omega(p) = (p^2 + \mu^2)^{\frac{1}{2}}$, for $p \in R^3$.

The free-energy operator H_0 is then given by

$$(H_0 f)_n(p_1, \cdots, p_n \mid x_1, \cdots, x_N) = \left[\sum_{i=1}^{n} \omega(p_i) - \frac{1}{2M} \sum_{j=1}^{N} \Delta_j\right] f_n(p_1, \cdots, p_n \mid x_1, \cdots, x_N),$$

where Δ_j is Δ operating on x_j.

H_0 is obviously positive and self-adjoint on its natural domain of definition D_0. Introducing the boson-annihilation operator $a(p)$ by

$$[a(p)f]_n(p_1, \cdots, p_n \mid x_1, \cdots, x_N) = (n + 1) f_{n+1}(p, p_1, \cdots, p_n \mid x_1, \cdots, x_N)$$

and the boson-creation operator $a^*(p)$ as the formal adjoint of $a(p)$, we get that, for $h \in L_2(R^3)$, the operators

$$a(h) = \int h(p) a(p) \, dp \quad \text{and} \quad a^*(h) = \int h(p) a^*(p) \, dp$$

are closed operators with domains containing the domain of $H_0^{\frac{1}{2}}$, and $a(h)$ and $a^*(h)$ are adjoints of each other. We have the following well-known estimate for f in the domain of $H_0^{\frac{1}{2}}$:

$$\|a^{\#}(h) f\| \leq (1/\mu) \, \|h\|_2 \, \|(H_0^{\frac{1}{2}} + 1) f\|, \qquad (1)$$

where $a^{\#}$ denotes a^* or a.

* On leave at Palmer Physical Laboratory, Princeton University.
[1] E. Nelson, J. Math. Phys. **5**, 1190 (1964).
[2] J. T. Cannon, Ph.D. thesis, Princeton University, 1968.
[3] R. Høegh-Krohn, J. Math. Phys. **9**, 2075 (1968); **10**, 639 (1969).
[4] R. Høegh-Krohn, Commun. Math. Phys. **12**, 216 (1969).
[5] Y. Kato and N. Mugibayashi, Progr. Theoret. Phys. (Kyoto) **30**, (1963).

On the domain of H_0^2 we have the following commutation relations for g and h in $L_2(R^3)$:

$$[a(g), a(h)] = [a^*(g), a^*(h)] = 0,$$

$$[a(g), a^*(h)] = \int g(p)h(p)\,dp,$$

and, for h and $\omega(p)h(p)$ in $L_2(R^3)$,

$$[H_0, a^\#(h)] = a^\#(\pm\omega h),$$

where $+$ goes with a^* and $-$ with a.

If we define the nucleon annihilation–creation operators $\psi^\#(x)$ in a similar way, we get the symbolic expression for the free-energy operator

$$H_0 = \int \omega(p)a^*(p)a(p)\,dp - \frac{1}{2M}\int \psi^*(x)\Delta\psi(x)\,dx.$$

The boson field $\varphi(x)$ is given by

$$\varphi(x) = 2^{-\frac{1}{2}}(2\pi)^{-\frac{3}{2}}\int [e^{ipx}a(p) + e^{-ipx}a^*(p)]\omega(p)^{-\frac{1}{2}}\,dp,$$

and the momentum cutoff field is given by

$$\varphi_k(x) = 2^{-\frac{1}{2}}(2\pi)^{-\frac{3}{2}} \times \int_{|p|\le k} [e^{ipx}a(p) + e^{-ipx}a^*(p)]\omega(p)^{-\frac{1}{2}}\,dp.$$

By the remarks above, $\varphi_k(x)$ is a self-adjoint operator with domain containing the domain of $H_0^{\frac{1}{2}}$. The interaction is formally given by

$$V = g\int \varphi(x)\psi^*(x)\psi(x)\,dx,$$

where g is a real constant.

If we define the cutoff interaction by

$$V_k = g\int \varphi_k(x)\psi^*(x)\psi(x)\,dx,$$

then V_k is a symmetric operator with domain containing the domain of $H_0^{\frac{1}{2}}$, and using (1), we find for f in the domain of $H_0^{\frac{1}{2}}$ that

$$\|V_k f\| \le c_k \|(H_0^{\frac{1}{2}} + 1)f\|, \tag{2}$$

where

$$c_k = (N/\mu)2^{-\frac{1}{2}}(2\pi)^{-\frac{3}{2}}\left[\int_{|p|\le k} \omega(p)^{-1}\,dp\right]^{\frac{1}{2}}.$$

Set

$$H_k = H_0 + V_k$$

and

$$E_k = -NMg^2(2\pi)^{-3}\int_{|p|\le k} [2M\omega(p)^2 + p^2\omega(p)]^{-1}\,dp.$$

Since $H_0^{\frac{1}{2}}$ is infinitesimally small with respect to H_0, we get from (2) that, for any k and any positive ϵ, there exists a constant b such that, for all f in D_0,

$$\|V_k f\| \le \epsilon\,\|H_0 f\| + b\,\|f\|. \tag{3}$$

Hence, H_k is a self-adjoint operator with domain D_0 for all k.

Nelson proved the following theorem:

Theorem 1: There is a unique self-adjoint H on $\mathcal{H}$ such that, for all real t and all f in $\mathcal{H}$,

$$\lim_{k\to\infty} e^{-it(H_k - E_k)}f = e^{-itH}f.$$

The convergence is uniform on bounded sets in t, and H is bounded below.

For the proof of this theorem and also for more details on the model, we refer to Nelson's paper.[1]

3. THE ASYMPTOTIC BOSON FIELD

Let $D_{\frac{1}{2}}$ be the domain of $H_0^{\frac{1}{2}}$. By (3), $H_k = H_0 + V_k$ is a self-adjoint operator with the same domain, D_0, as H_0. Therefore, e^{itH_k}, as well as e^{itH_0}, leaves D_0 and $D_{\frac{1}{2}}$ invariant. For $h \in L_2(R^3)$, we define

$$a^\#_{k,t}(h) = e^{-itH_k}e^{itH_0}a^\#(h)e^{-itH_0}e^{itH_k}. \tag{4}$$

$a^\#_{k,t}(h)$ is unitarily equivalent to the closed operator $a^\#(h)$; hence it is closed. Since the domain of $a^\#(h)$ contains $D_{\frac{1}{2}}$, and $D_{\frac{1}{2}}$ is left invariant by e^{-itH_0} and e^{itH_k}, we see that the domain of $a^\#_{k,t}(h)$ contains $D_{\frac{1}{2}}$.

Let φ and ψ be in D_0 and let h be in $L_2(R^3)$.

Consider the function of one variable t given by

$$(\varphi, a^\#_{k,t}(h)\psi) = (e^{-itH_0}e^{itH_k}\varphi, a^\#(h)e^{-itH_0}e^{itH_k}\psi), \tag{5}$$

and the function of two variables s and t given by

$$(e^{-isH_0}e^{isH_k}\varphi, a^\#(h)e^{-itH_0}e^{itH_k}\psi) = (a^\#(h)^*e^{-isH_0}e^{isH_k}\varphi, e^{-itH_0}e^{itH_k}\psi). \tag{6}$$

For $s = t$, the function (6) coincides with the function (5). Since e^{itH_k} leaves D_0 invariant, we find that (6) is differentiable both with respect to s and t. The partial derivatives are given, respectively, by

$$(e^{-isH_0}iV_ke^{isH_k}\varphi, a^\#(h)e^{-itH_0}e^{itH_k}\psi)$$

and

$$(a^\#(h)^*e^{-isH_0}e^{isH_k}\varphi, e^{-itH_0}iV_ke^{itH_k}\psi).$$

V_k is a symmetric operator with domain containing D_0; hence V_k is closable, and so $e^{-isH_0}iV_ke^{isH_k}\varphi$ is strongly continuous in s. $a^\#(h)$ and $a^\#(h)^*$ are closed operators with domain containing D_0, and so $a^\#(h)e^{-itH_0}e^{itH_k}\psi$ and $a^\#(h)^*e^{-isH_0}e^{isH_k}\varphi$ are strongly continuous in t and s, respectively. Hence both the partial derivatives are continuous in s and t, yielding

that (5) is a differentiable function of t, that is,

$$\frac{d}{dt}(\varphi, a^{\#}_{k,t}(h)\psi) = (e^{-itH_0}iV_k e^{itH_k}, a^{\#}(h)e^{-itH_0}e^{itH_k}\psi) + (a^{\#}(h)^* e^{-itH_0}e^{itH_k}\varphi, e^{-itH_0}iV_k e^{itH_k}\psi).$$

It is well known that the commutation relations of H_0 and $a^{\#}(h)$ may be given in the form

$$e^{itH_0}a^{\#}(h)e^{-itH_0} = a^{\#}(h_{\pm t}), \tag{7}$$

where $+$ goes with a^* and $-$ with a, and $h_t(p) = e^{it\omega(p)}h(p)$.

By using (7), the formula for the derivative may be written as follows:

$$\frac{d}{dt}(\varphi, a^{\#}_{k,t}(h)\psi) = (iV_k e^{itH_k}\varphi, a^{\#}(h_{\pm t})e^{itH_k}\psi) + (a^{\#}(h_{\pm t})^* e^{itH_k}\varphi, iV_k e^{itH_k}\psi).$$

Let h and g be in $L_2(R^3)$ and ψ in D_0, we then have the following well-known estimate:

$$\|a^{\#}(h)a^{\#}(g)\psi\| \leq c\,\|h\|_2\,\|g\|_2\,\|(H_0+1)\psi\|, \tag{8}$$

where c depends only on μ. Consulting the definition of V_k, we get by this estimate that for h in $L_2(R^3)$, V_k maps D_0 into the domain of $a^{\#}(h)$, and $a^{\#}(h)$ maps D_0 into the domain of V_k. Hence the formula for the derivative may be written as follows:

$$\frac{d}{dt}(\varphi, a^{\#}_{k,t}(h)\psi) = -i(\varphi, e^{-itH_k}[V_k, a^{\#}(h_{\pm t})]e^{itH_k}\psi). \tag{9}$$

Using the commutation relations for $a^{\#}(h)$, we see that

$$[V_k, a^{\#}(h)] = \pm 2^{-\frac{1}{2}}(2\pi)^{-\frac{3}{2}}\int\left[\int_{|p|\leq k} e^{\pm ipx}h(p)\,dp\right]\psi^*(x)\psi(x)\,dx, \tag{10}$$

where $+$ goes with a^* and $-$ with a.

Lemma 1: Let $f \in L_\infty(R^3)$, and define

$$A(f) = \int f(x)\psi^*(x)\psi(x)\,dx.$$

Then $A(f)$ is a bounded operator on $\mathcal{H}$ such that

$$\|A(f)\| \leq N\,\|f\|_\infty.$$

Moreover, if $\{f_n\}$ is a uniformly bounded sequence in $L_\infty(R^3)$ which converges to f almost everywhere, then $A(f_n)$ converges strongly to $A(f)$.

Proof: We see that $A(f)$ is the operator of multiplication by $\sum_{i=1}^N f(x_i)$ and hence the first part of the lemma follows immediately. The second part follows from Lebesque's lemma on dominated convergence. This proves the lemma.

Since

$$\int_{|p|\leq k} e^{\pm ipx}h_{\pm t}(p)\,dp$$

is a uniformly bounded continuous function of x and t, it follows from (10) and Lemma 1 that $[V_k, a^{\#}(h_{\pm t})]$ is a bounded operator on $\mathcal{H}$ that depends in a strongly continuous way on t. Hence we may integrate (9) and get

$$(\varphi[a^{\#}_{k,t}(h) - a^{\#}(h)]\psi) = -i\left(\varphi, \int_0^t e^{-isH_k}[V_k, a^{\#}(h_{\pm s})]e^{isH_k}\psi\,ds\right),$$

where the integral is a strong integral. Since the integral on the right-hand side gives a bounded operator, we get the following formula for $h \in L_2(R^3)$:

$$a^{\#}_{k,t}(h) - a^{\#}(h) = -i\int_0^t e^{-isH_k}[V_k, a^{\#}(h_{\pm s})]e^{isH_k}\,ds, \tag{11}$$

where the integral is a strong integral.

For $h \in C_0$, we get from (10) that $[V_k, a^{\#}(h_{\pm s})]$ is independent of k for k large. From Theorem 1, we therefore get that

$$e^{-isH_k}[V_k, a^{\#}(h_{\pm s})]e^{isH_k} = e^{-is(H_k-E_k)}[V_k, a^{\#}(h_{\pm s})]e^{is(H_k-E_k)}$$

converges strongly to $e^{-isH}[V, a^{\#}(h_{\pm s})]e^{isH}$ and the convergence is uniform on bounded sets in s, where we have introduced $[V, a^{\#}(h)] = [V_k, a^{\#}(h)]$ for k large. Hence we get from (11) that, for $h \in C_0$,

$$\operatorname*{strong\,lim}_{k\to\infty}(a^{\#}_{k,t}(h) - a^{\#}(h)) = -i\int_0^t e^{-isH}[V, a^{\#}(h_{\pm s})]e^{isH}\,ds, \tag{12}$$

where the integral is a strong integral.

We now define, for $h \in L_2$,

$$a^{\#}_t(h) = e^{-itH}e^{itH_0}a^{\#}(h)e^{-itH_0}e^{itH}.$$

Lemma 2: For $h \in C_0$, we have

$$a^{\#}_t(h) - a^{\#}(h) = -i\int_0^t e^{-isH}[V, a^{\#}(h_{\pm s})]e^{isH}\,ds,$$

where the integral is a strong integral.

Proof: By (12), $a^{\#}_{k,t}(h)$ converges strongly on $D_{\frac{1}{2}}$ and, by Theorem 1, $e^{it(H_k-E_k)}$ converges strongly to

e^{itH}. Hence,

$$e^{it(H_k-E_k)}a^{\#}_{k,t}(h) = a^{\#}(h_{\pm t})e^{it(H_k-E_k)} \qquad (13)$$

converges strongly on $D_{\frac{1}{2}}$ as k tends to infinity. Since, on the other hand, $a^{\#}(h_{\pm t})$ is closed and $e^{it(H_k-E_k)}$ converges strongly to e^{itH}, we get that e^{itH} maps $D_{\frac{1}{2}}$ into the domain of $a^{\#}(h_{\pm t})$, and the strong limit on $D_{\frac{1}{2}}$ of (13) is $a^{\#}(h_{\pm t})e^{itH}$. Multiplying to the left in (13) by $e^{-it(H_k-E_k)}$, we find, on $D_{\frac{1}{2}}$, that $a^{\#}_{k,t}(h)$ converges strongly to $a^{\#}_t(h)$.

It follows from Lemma 1 and the formulas (10) and (11) that, for $h \in C_0$, $a^{\#}_{k,t}(h) - a^{\#}(h)$ is uniformly norm-bounded in k. By the observations above, $a^{\#}_{k,t}(h) - a^{\#}(h)$ tends strongly to $a^{\#}_t(h) - a^{\#}(h)$ on $D^{\frac{1}{2}}$; hence, by uniform boundedness of $a^{\#}_{k,t}(h) - a^{\#}(h)$, and Eq. (12), we see that this proves Lemma 2.

We now define L_0 to be the dense subspace of $L_2(R^3)$ consisting of those C^∞ functions with compact support which vanish in a neighborhood of the origin in R^3.

Theorem 2: For $h \in L_0$, $a^{\#}_t(h)$ tends strongly on $D_{\frac{1}{2}}$ to the limits $a^{\#}_{\pm}(h)$ as t tends to $\pm\infty$, such that $a^{\#}_t(h) - a^{\#}(h)$ converges in norm to the bounded operator $a^{\#}_{\pm}(h) - a^{\#}(h)$. Hence, the $a^{\#}_{\pm}(h)$ are closed operators with the same domain of definition as $a^{\#}(h)$.

Moreover $a^{*}_{\pm}(h)$ is the adjoint of $a_{\pm}(\bar{h})$, and $a_{-}(\bar{h})$, $a^{*}_{-}(h)$, $a(\bar{h})$, $a^{*}(h)$, $a_{+}(\bar{h})$, and $a^{*}_{+}(h)$ all have the same domain of definition.

H and $a^{\#}_{\pm}(h)$ satisfy the same commutation relation as do H_0 and $a^{\#}(h)$, in the sense that

$$e^{itH}a^{\#}_{\pm}(he)^{-itH} = a^{\#}_{\pm}(h_{\pm t}).$$

Proof: We start by proving that, if $a^{\#}_t(h) - a^{\#}(h)$ converges in norm, then the strong convergence on $D_{\frac{1}{2}}$ follows by observing that $D_{\frac{1}{2}}$ is contained in the domain of $a^{\#}(h)$. By Lemma 2, the convergence in norm of $a^{\#}_t(h) - a^{\#}(h)$ is equivalent to the norm convergence of the integral

$$\int_0^t e^{-isH}[V, a^{\#}(h_{\pm s})]e^{isH}\,ds.$$

By Lemma 1 and Eq. (10), we see that the norm of the integrand may be estimated by

$$2^{-\frac{1}{2}}(2\pi)^{-\frac{3}{2}}N \sup_x \left|\int e^{\pm i[\omega(p)t+px]}h(p)\,dp\right|,$$

and this tends to zero faster than any inverse power of t since h is in L_0. Hence the integral converges in norm to a bounded operator as t tends to $\pm\infty$. This gives us the norm convergence of $a^{\#}_t(h) - a^{\#}(h)$, and from this norm convergence we also get that $a^{*}_{\pm}(h)$ is the adjoint of $a_{\pm}(\bar{h})$.

The domain of $a^{\#}(h)$ is the closure of the subspace of $\mathcal{H}$ generated by states with a finite number of free bosons, under the norm $\|\psi\|^{\#} = \|a^{\#}(h)\psi\| + \|\psi\|$. From the fact that $a^{*}(h)$ is the adjoint of $a(\bar{h})$ and the commutation relations, we see that the norms defining the domains of $a^{*}(h)$ and $a(\bar{h})$ are equivalent. Hence the two operators have the same domain. The rest of the domain relations follow from what we have already proved, namely that $a^{\#}_{\pm}(h) - a^{\#}(h)$ is a bounded operator.

To get the commutation relations for H and $a^{\#}_{\pm}(h)$, we observe that

$$e^{itH}[a^{\#}_s(h) - a^{\#}(h)]e^{-itH} = a^{\#}_{s-t}(h_{\pm t}) - e^{itH}a^{\#}(h)e^{-itH}.$$

The left-hand side converges in norm to

$$e^{itH}a^{\#}_{\pm}(h)e^{-itH} - e^{itH}a^{\#}(h)e^{-itH},$$

as s tends to $\pm\infty$. The right-hand side converges strongly on $D_{\frac{1}{2}}$ to

$$a^{\#}_{\pm}(h_{\pm t}) - e^{itH}a^{\#}(h)e^{-itH}$$

since we already know from the proof of Lemma 2 that e^{-itH} maps $D_{\frac{1}{2}}$ into the domain of $a^{\#}(h)$. This gives us the commutation relation and hence the theorem is proved.

It would of course be interesting to know if the asymptotic limits $a^{\#}_{\pm}(h)$ satisfy the commutation relations. But here we meet with the difficulty that we do not know enough about the range of $a^{\#}_{\pm}(h)$ to form the product $a^{\#}_{\pm}(g)a^{\#}_{\pm}(h)$. As a corollary of Theorem 2 we get, however, the following weak form of the commutation relations.

Corollary: For g and h in L_0, and φ and ψ in the intersection of the domains of $a^{\#}(h)$ and $a^{\#}(g)$, we have

$$(a^{\#}_{\pm}(g)^{*}\varphi, a^{\#}_{\pm}(h)^{*}\psi) - (a^{\#}_{\pm}(h)\varphi, a^{\#}_{\pm}(g)\psi) = \pm(g, h),$$

where $+$ goes with $a^{*}_{\pm}$ and $-$ with $a_{\pm}$, and

$$(a^{\#}_{\pm}(g)^{*}\varphi, a^{\#}_{\pm}(h)\psi) - (a^{\#}_{\pm}(h)^{*}\varphi, a^{\#}_{\pm}(g)\psi) = 0.$$

Proof. Since $a^{\#}_t(h)$ and $a^{\#}(h)$ have the same domain and are unitarily equivalent, and, by Theorem 2, $a^{\#}(h)$ and $a^{\#}(h)^{*}$ have the same domain of definition; we see that $a_t(h)$ satisfies the weak form of the commutation relations as given in the corollary. By Theorem 2, $a^{\#}_t(h)$ converges strongly to $a^{\#}_{\pm}(h)$ on the domain of $a^{\#}(h)$. This proves the corollary.

ACKNOWLEDGMENT

I would here like to thank Dr. J. T. Cannon for pointing out some errors in the first version of this manuscript.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

On the Relaxation to Quantum-Statistical Equilibrium of the Wigner–Weisskopf Atom in a One-Dimensional Radiation Field. I. A Study of Spontaneous Emission

RUSSELL DAVIDSON
Center for Statistical Mechanics and Thermodynamics, The University of Texas at Austin, Austin, Texas
AND
JOHN J. KOZAK
Department of Chemistry, The University of Notre Dame, Notre Dame, Indiana

(Received 23 April 1969)

A theoretical study of the phenomenon of spontaneous emission has been carried out, using as a model the Wigner–Weisskopf atom in a one-dimensional radiation field. The calculation is performed within the framework of the Prigogine theory of nonequilibrium statistical mechanics. When the model is solved exactly to first order in the coupling parameter α and the evolution in time of the diagonal elements of the density matrix ρ is studied, it is found that the relaxation to equilibrium is characterized in part by a sequence of slowly damped oscillations. This result seems to be in agreement with the observation made by Zwanzig, namely, that exponential decay in time seems not to be universal, and may, in fact, be hidden behind some other kind of time dependence. An approximate theory is developed alongside the exact one, and corresponding terms in each treatment are compared numerically. It is found that, for small values of the coupling parameter α ($\alpha \lesssim 0.1$) and for sufficiently large values of τ, defined as $\tau = \alpha Et$ where E is frequency and t is time, the approximate theory gives a satisfactory representation of the exact solution to first order in α. The importance and relevance of the model introduced by Van Hove and coworkers, in which nonexponential behavior was also observed, will be noted but not stressed in this paper, as this relationship will be developed in considerable detail in a subsequent contribution. Finally, the possible relevance of the theory to a problem of interest in magnetic resonance is mentioned.

I. INTRODUCTION

The problems considered in this paper have been suggested in large measure by an article of Zwanzig.[1] In his 1960 Boulder lecture, Zwanzig studied the properties of an equation which has the form of what is now usually called the Prigogine–Résibois master equation.[2] Zwanzig's discussion is mostly formal, in the sense that the example considered is chosen so that the operators in the master equation can be replaced by functions, without reference to any particular physical model. Some of the conclusions reached, however, are rather disquieting in that the explicit results which he calculates seem not in accord with most of the assumptions which are usually made in developing the theory of nonequilibrium processes from the master equation. In particular, Zwanzig's analysis leads to the result that exponential decay in time seems not to be universal, and may indeed be hidden behind some other kind of time dependence. His other main suggestion—namely, that effects which are proportional to the inverse of N, the number of degrees of freedom of a large system, may become important after a certain length of time—will not be discussed in this paper. These effects, which are always ignored in usual statistical-mechanical calculations, are clearly related to the problem of recurrences. A separate paper by the authors will treat this problem explicitly for the same model which is studied below.

Nonexponential decay can appear within the context of certain mathematical models studied by means of the formalism of Prigogine and coworkers.[2,3] The model chosen here is the Wigner–Weisskopf atom: a two-level quantum system in interaction with a massless boson field.[4] This system may be considered as a very special case of the well-known Lee model.[5] It retains the main features of the Lee model for the purpose of the discussion of spontaneous emission—the decay of the two-level system from its excited state to its ground state in the presence only of the zero-point boson field. If one performs the appropriate calculations, one finds that the differences between the various models are simply ones of complication. In this paper, a one-dimensional system is treated; this is a simplication which in no way detracts from the generality of the results, inasmuch as the "collision

[1] R. W. Zwanzig, "Statistical Mechanics of Irreversibility," in *Lectures in Theoretical Physics (Boulder, 1960)*, W. E. Brittin, B. W. Downs, and J. Downs, Eds. (Interscience Publishers, Inc., New York, 1961), Vol. III.

[2] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962); P. Résibois, "A Perturbative Approach to Irreversible Statistical Mechanics," in *Physics of Many-Particle Systems* (Gordon & Breach, Science Publishers, Inc., New York, 1966), Vol. I.

[3] R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience Publishers, Inc., New York, 1963); P. Résibois, *Electrolyte Theory* (Harper and Row, New York, 1968).

[4] V. F. Weisskopf and E. P. Wigner, Z. Physik **63**, 54 (1930).

[5] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1961).

operator" developed in Sec. III needs only to be multiplied by 4π to obtain the three-dimensional result. The model does indeed contain most of the difficulties involved for the study of the interaction between one single object (the two-level system) and an infinite number of degrees of freedom, although it cannot be thought of as a model for a great number of mutually interacting objects. The concept of temperature is not relevant to the system—it is effectively at zero temperature—but nonetheless there is genuine irreversibility involved, at any rate in the limit of an infinite system.

The possibility of using the now well-developed, formal theory of Prigogine to investigate and, hopefully, to clarify the serious questions raised by Zwanzig, always within the framework of a well-defined model, justifies, we feel, the study which follows. In Sec. II, we present a complete, mathematical specification of the model under study. In Sec. III, we review those aspects of the theory of the Prigogine–Résibois master equation which are relevant to the problem at hand. In Sec. IV, the purely exponential solution will be derived and its discrepancies from the more exact solution, containing nonexponential terms, will be pointed out. The reasons for these discrepancies will be traced in Sec. V, and in Sec. VI a numerical comparison of the two solutions will be presented. Finally, in Sec. VII, we summarize the principal results of our calculation and then suggest several conclusions which might be drawn from the numerical work presented in Sec. VI. In particular, it will be suggested that the model treated below may have considerable relevance to some physical situations.

II. THE WIGNER-WEISSKOPF MODEL

The Wigner–Weisskopf model consists of a two-level fermion and a massless boson field in interaction. The term in the Hamiltonian for the bosons is chosen as for a set of harmonic oscillators in the second quantization notation:

$$H_{\text{bos}} = \sum_{\lambda} [\tfrac{1}{2}\hbar\omega_\lambda(a_\lambda^* a_\lambda + 1)]. \tag{1}$$

Here λ labels the possible modes of oscillation and ω_λ is the corresponding frequency. The creation and destruction operators a_λ^* and a_λ, respectively, are defined by their matrix elements in the occupation number representation:

$$\langle n_\lambda | a_\lambda | m_\lambda \rangle = [2(n_\lambda + 1)]^{\frac{1}{2}} \delta^{\text{Kr}}(m_\lambda - n_\lambda - 1), \tag{2a}$$

$$\langle m_\lambda | a_\lambda^* | n_\lambda \rangle = [2(n_\lambda + 1)]^{\frac{1}{2}} \delta^{\text{Kr}}(m_\lambda - n_\lambda - 1), \tag{2b}$$

where a state $|n_\lambda\rangle$ is the state with n_λ ($n = 0, 1, 2, \cdots$) photons in the λth mode. The $\delta^{\text{Kr}}(\cdots)$ is a Kronecker delta.

For the fermion, there are two quantum states which may be written $|1\rangle$ and $|2\rangle$. Then, if ϵ_1 is the energy of $|1\rangle$ and ϵ_2 that of $|2\rangle$, the fermion term in the Hamiltonian is

$$H_{\text{fer}} = \epsilon_1 \alpha\alpha^* + \epsilon_2 \alpha^*\alpha, \tag{3}$$

where the operators α and α^* are

$$\alpha = |1\rangle\langle 2|, \tag{4a}$$

$$\alpha^* = |2\rangle\langle 1|. \tag{4b}$$

It is readily checked that

$$[\alpha, \alpha^*]_+ = 1.$$

For the interaction between the fermion and the mode λ, we choose simply

$$V_\lambda = h_\lambda^* \alpha^* a_\lambda + h_\lambda \alpha a_\lambda^*, \tag{5}$$

where the coefficient h_λ^* of the first term must be the complex conjugate of the coefficient of the second term to insure hermiticity. Non-energy-conserving interactions, such as would be proportional to αa_λ or $\alpha^* a_\lambda^*$, will not be admitted. The full Hamiltonian is then

$$H = \epsilon_1 \alpha\alpha^* + \epsilon_2 \alpha^*\alpha + \sum_{\lambda} [\tfrac{1}{2}\hbar\omega_\lambda(a_\lambda^* a_\lambda + 1)] + \sum_{\lambda} (h_\lambda^* \alpha^* a_\lambda + h_\lambda \alpha a_\lambda^*). \tag{6}$$

The matrix elements of this Hamiltonian are to be taken between states of the system given by

$$|i; \{n_\lambda\}\rangle = |i\rangle \prod_{\lambda \in \Lambda} |n_\lambda\rangle,$$

with $i = 1, 2$ and with $n_\lambda = 0, 1, 2, \cdots$ for each mode λ.

We must still choose h_λ so as to specify completely the problem. This we shall do by making the model resemble as closely as possible the situation of an electron in an atom interacting with a field of electromagnetic radiation. Since there is but one dimension, the field must be scalar, and we shall describe it by a potential $\phi(x, t)$, satisfying the wave equation

$$\frac{\partial^2\phi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0.$$

This equation has a solution

$$\phi = -\sum_{\lambda} [q_\lambda(t)\phi_\lambda(x) + q_\lambda^*(t)\phi_\lambda^*(x)],$$

where

$$\frac{\partial^2 q_\lambda}{\partial t^2} + \omega_\lambda^2 q_\lambda = 0$$

and

$$\phi_\lambda(x) = ce^{ik_\lambda x}.$$

The quantity c is a normalization constant to be determined by imposing a boundary condition. In particular, by considering a system of length L, we find

$$k_\lambda = 2\pi n_\lambda / L$$

and

$$ck_\lambda = \omega_\lambda,$$

where the n_λ range over the integers.

The energy associated with the field described by $\phi(x, t)$ is

$$U = \tfrac{1}{4} \sum_\lambda [q_\lambda q_\lambda^* + q_\lambda^* q_\lambda] k_\lambda^2 c^2 L. \tag{7}$$

But Eq. (1) may be written as

$$H_{\text{bos}} = \tfrac{1}{4} \sum_\lambda \hbar\omega_\lambda [a_\lambda a_\lambda^* + a_\lambda^* a_\lambda]. \tag{8}$$

Comparison of Eqs. (7) and (8) yields as a suitable normalization

$$q_\lambda = a_\lambda,$$
$$\phi_\lambda(x) = c(\hbar / L\omega_\lambda)^{\frac{1}{2}} e^{ik_\lambda x}.$$

To complete the analogy, one sets the interaction

$$V_\lambda = -e\phi_\lambda, \tag{9}$$

where e is a kind of one-dimensional charge.

If one now compares Eq. (9) with Eq. (5), one obtains that

$$|h_\lambda|^2 = \hbar c e^2 r^2 |k_\lambda| / L \tag{10}$$

in the "dipole" approximation, with

$$r = \langle 1 | x | 2 \rangle.$$

It is readily seen that the quantity $e^2 r^2 / c\hbar$ is dimensionless and corresponds to the fine structure constant of electrodynamics. Accordingly, we set

$$\alpha = e^2 r^2 / c\hbar.$$

We then obtain

$$|h_\lambda|^2 = \alpha \hbar^2 c^2 |k_\lambda| / L. \tag{11}$$

This completes the specification.

III. THE PRIGOGINE-RÉSIBOIS MASTER EQUATION

We shall begin with the Liouville–von Neumann equation for the density matrix ρ:

$$\frac{\partial \rho}{\partial t} = \frac{1}{i\hbar} [H, \rho] = \frac{1}{i\hbar} [H\rho - \rho H], \tag{12}$$

where H is the Hamiltonian. Equation (12) has the formal solution

$$\rho(t) = e^{-Ht/\hbar} \rho(0) e^{iHt/\hbar}. \tag{13}$$

The calculations of this paper will be performed in the occupation number representation, for which we shall use the so-called (ν, N) notation[6]:

$$\langle n | A | m \rangle = A_\nu(N),$$

where

$$\nu = n - m, \quad N = \tfrac{1}{2}(n + m)$$

for any operator A. Here n, m, ν, and N are shorthand for the set of occupation numbers corresponding to each degree of freedom of the system. In this notation, the diagonal elements of ρ appear as the quantities $\rho_0(N)$.

The solution (13) is to be expressed as a perturbation series based on the splitting of the Hamiltonian:

$$H = H_0 + H_1,$$

where H_1 is the perturbation. Let us define

$$f(t) = e^{iH_0 t/\hbar} \rho(t) e^{-iH_0 t/\hbar}. \tag{14}$$

Then, the equation of evolution of $f(t)$ is

$$\frac{\partial f}{\partial t} = \frac{1}{i\hbar} [e^{iH_0 t/\hbar} H_1 e^{-iH_0 t/\hbar}, f(t)], \tag{15}$$

which yields the iterative solution

$$f(t) = f(0) + (i\hbar)^{-1} \int_0^t dt' [e^{iH_0 t/\hbar} H_1 e^{-iH_0 t/\hbar}, f(0)] + (i\hbar)^{-2} \int_0^t dt' \Big[e^{iH_0 t'/\hbar} H_1 e^{-iH_0 t'/\hbar}, \int_0^{t'} dt'' [e^{iH_0 t''/\hbar} H_1 e^{-iH_0 t''/\hbar}, f(0)] \Big] + \cdots. \tag{16}$$

This, with Eq. (14), gives the desired perturbative solution

$$\rho(t) = \sum_{n=0}^{\infty} \int_{T_n} e^{-iH_0(t-t_1)/\hbar} [H_1, e^{-iH_0(t_1-t_2)/\hbar} \times [H_1, \cdots [H_1, e^{-iH_0 t_n/\hbar} \rho(0) e^{iH_0 t_n/\hbar}] \cdots] \times e^{iH_0(t_1-t_2)/\hbar}] e^{+iH_0(t-t_1)/\hbar}, \tag{17}$$

where

$$\int_{T_n} \equiv (i\hbar)^{-n} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n.$$

From Eq. (17) we shall derive the "generalized master equation" for the system by considering the

[6] P. Résibois, Physica 27, 541 (1961).

diagonal elements of $\rho(t)$. Equation (12) gives

$$\partial_t \langle n|\, \rho(t)\, |n\rangle = (i\hbar)^{-1}\langle n|\, [H_0, \rho(t)]\, |n\rangle + (i\hbar)^{-1}\langle n|\, [H_1, \rho(t)]\, |n\rangle.$$

That is,

$$\partial_t \rho_0(N, t) = (i\hbar)^{-1}\langle N|\, [H_0, \rho(t)]\, |N\rangle + (i\hbar)^{-1}\langle N|\, [H_1, \rho(t)]\, |N\rangle.$$

Since H_0 is diagonal in this representation, the first term of the right-hand side vanishes. Using Eq. (17), we have the result

$$\partial_t \rho_0(N, t) = (i\hbar)^{-1}\sum_{m=0}^{\infty}\int_{T_m}\langle N|\, [H_1, e^{-iH_0(t-t_1)/\hbar} \times [H_1, \cdots [H_1, e^{-iH_0 t_m/\hbar}\rho(0)e^{iH_0 t_m/\hbar}]\cdots e^{iH_0(t-t_1)/\hbar}]\, |N\rangle. \quad (18)$$

Some new notation is required to simplify subsequent operations on Eq. (18). Let us note that

$$\langle n|\, [H_1, A]\, |n'\rangle = \sum_{n''}[\langle n|\, H_1\, |n''\rangle\langle n''|\, A\, |n'\rangle - \langle n|\, A\, |n''\rangle\langle n''|\, H_1\, |n'\rangle],$$

where we have used the completeness of the set of quantum states. This expression in the (ν, N) notation, with ν and N corresponding to n and n' and

$$\nu' = n'' - n',$$

is

$$\sum_{\nu'}[(H_1)_{\nu-\nu'}(N + \tfrac{1}{2}\nu')A_{\nu'}(N + \tfrac{1}{2}\nu' - \tfrac{1}{2}\nu) - (H_1)_{\nu'}(N + \tfrac{1}{2}\nu' - \tfrac{1}{2}\nu)A_{\nu-\nu'}(N + \tfrac{1}{2}\nu')].$$

By a change in the summation index of the second term, one may write this in the form

$$\sum_{\nu'}\langle \nu|\, \mathcal{H}_1(N)\, |\nu'\rangle\, A_{\nu'}(N) \quad (19)$$

with a new *operator* $\mathcal{H}_1(N)$, which acts on the N-variables of everything to its right. The new operator $\mathcal{H}_1(N)$ is defined by its matrix elements

$$\langle \nu|\, \mathcal{H}_1(N)\, |\nu'\rangle = \eta^{\nu'}(H_1)_{\nu-\nu'}(N)\eta^{-\nu} - \eta^{-\nu'}(H_1)_{\nu-\nu'}(N)\eta^{\nu},$$

where the displacement operators η act thus:

$$\eta^{\nu} f(N) = f(N + \tfrac{1}{2}\nu).$$

Further, we note that if G is a diagonal operator, then, using the notation of Eq. (19), we have

$$\begin{aligned}&\langle n|\, [H_1, e^{Gt}Ae^{-Gt}]\, |n'\rangle \\ &= \sum_{\nu'}\langle \nu|\, \mathcal{H}_1(N)\, |\nu'\rangle\langle N + \tfrac{1}{2}\nu'|\, e^{Gt}\, |N + \tfrac{1}{2}\nu'\rangle \\ &\quad \times \langle N + \tfrac{1}{2}\nu'|\, A\, |N - \tfrac{1}{2}\nu'\rangle\langle N - \tfrac{1}{2}\nu'|\, e^{-Gt}\, |N - \tfrac{1}{2}\nu'\rangle \\ &= \sum_{\nu'}\langle \nu|\, \mathcal{H}_1(N)\, |\nu'\rangle\, \mathcal{G}(\nu', N, t)\langle N + \tfrac{1}{2}\nu'|\, A\, |N - \tfrac{1}{2}\nu'\rangle,\end{aligned} \quad (20)$$

say. One can put

$$G = -iH_0/\hbar$$

and then use Eq. (20) in Eq. (18) to give

$$\begin{aligned}\partial_t \rho_0(N, t) &= (i\hbar)^{-1}\sum_{m=0}^{\infty}\sum_{N_m}\int_{T_m}\langle 0|\, \mathcal{H}_1(N)\, |\nu_1\rangle\, \mathcal{G}(\nu_1, N, t - t_1) \\ &\quad \times \langle \nu_1|\, \mathcal{H}_1(N)\, |\nu_2\rangle \cdots \langle \nu_m|\, \mathcal{H}_1(N)\, |\nu_{m+1}\rangle \\ &\quad \times \mathcal{G}(\nu_{m+1}, N, t_m)\rho_{\nu_{m+1}}(N, 0),\end{aligned} \quad (21)$$

where

$$\sum_{N_m} = \sum_{\nu_1}\cdots\sum_{\nu_{m+1}}.$$

In terms of the multiple summation over the ν-variables, we shall distinguish between those where at least one ν-variable has assumed the value zero (for *every* degree of freedom), and those where each ν has a nonzero value. Then, the right-hand side of Eq. (21) can be split into two terms as follows:

$$\begin{aligned}&\partial_t \rho_0(N, t) \\ &= (i\hbar)^{-1}\sum_{m=0}^{\infty}\sum_{f=0}^{m}\sum_{\nu_{m+1}}\int_{T_m}\langle 0|\, \mathcal{H}_1(N)G(t - t_1)\cdots G(t_{f-1} - t_f)\mathcal{H}_1(N)\, |0\rangle_{\text{irr}} \\ &\quad \times \langle 0|\, G(t_f - t_{f-1})\mathcal{H}_1(N)\cdots G(t_m)\, |\nu_{m+1}\rangle\, \rho_{\nu_{m+1}}(N, 0) \\ &\quad + (i\hbar)^{-1}\sum_{m=0}^{\infty}\sum_{\nu_{m+1}}\int_{T_m}\langle 0|\, \mathcal{H}_1(N)G(t - t_1)\cdots \mathcal{H}_1(N)G(t_m)\, |\nu_{m+1}\rangle_{\text{irr}}\, \rho_{\nu_{m+1}}(N, 0).\end{aligned} \quad (22)$$

In this expression, we have introduced the diagonal operator G, given by

$$\langle \nu''|\, G(\tau)\, |\nu''\rangle = \mathcal{G}(\nu'', N, \tau).$$

The operator products are to be expanded using the completeness notation:

$$1 = \sum_{\nu''}|\nu''\rangle\langle \nu''|.$$

The suffix "irr" is shorthand for "irreducible," and has the sense, usual in this kind of work, that the intermediate ν-variables in the expansion of an operator product may not assume the value zero.

Equation (17) can now be rewritten using the notations of Eq. (19) and (20):

$$\rho_0(N, t) = \sum_{m=0}^{\infty}\sum_{\nu_{m+1}}\int_{T_m}\langle 0|\, G(t - t_1)\mathcal{H}_1(N)\cdots \mathcal{H}_1(N)G(t_m)\, |\nu_{m+1}\rangle\, \rho_{\nu_{m+1}}(N, 0). \quad (23)$$

Equations (22) and (23) combine to yield

$$\partial_t \rho_0(N,t) = (i\hbar)^{-1} \sum_{f=0}^{\infty} \int_{T_f} \langle 0|\, \mathcal{H}_1(N) G(t-t_1) \cdots G(t_{f-1}-t_f) \mathcal{H}_1(N)\, |0\rangle_{\mathrm{irr}}\, \rho_0(N,t_f) + (i\hbar)^{-1} \sum_{m=0}^{\infty} \sum_{\nu} \int_{T_m} \langle 0|\, \mathcal{H}_1(N) G(t-t_1) \cdots \mathcal{H}_1(N) G(t_m)\, |\nu\rangle_{\mathrm{irr}}\, \rho_\nu(N,0).$$

With a few convenient changes of the t-variables, this becomes the generalized master equation

$$\partial_t \rho_0(N,t) = \int_0^t d\tau \mathcal{C}(\tau) \rho_0(N, t-\tau) + \mathfrak{D}(t, \{\rho_\nu(N,0)\}), \tag{24}$$

with

$$\mathcal{C}(\tau) = (i\hbar)^{-1} \sum_{f=0}^{\infty} \int_{T_{f-1}} \langle 0|\, \mathcal{H}_1(N) G(t-t_1) \cdots G(t_{f-1}) \mathcal{H}_1(N)\, |0\rangle_{\mathrm{irr}} \tag{25}$$

and

$$\mathfrak{D}(t, \{\rho_\nu(N,0)\}) = (i\hbar)^{-1} \sum_{m=0}^{\infty} \sum_{\nu \neq 0} \int_{T_m} \langle 0|\, \mathcal{H}_1(N) G(t-t_1) \cdots \mathcal{H}_1(N) G(t_m)\, |\nu\rangle_{\mathrm{irr}}\, \rho_\nu(N,0). \tag{26}$$

Equation (26) is the term usually called[3] the "destruction term," and it can be shown for a wide class of initial conditions $\rho_\nu(N,0)$—in particular, those which correspond to correlations built up only by the mechanical interactions between the degrees of freedom of the system—that it is a rapidly decreasing function of t. In the calculations of this paper, $\mathfrak{D}$ will be set rigorously equal to zero by the expedient of choosing an initial condition such that only diagonal elements of $\rho(0)$ may be nonzero.

The form of Eq. (24), a multiple convolution, enables one to readily take its Laplace transform

$$\psi(z) = \int_0^t dt e^{izt} \mathcal{C}(t) = (i\hbar)^{-1} \sum_{f=0}^{\infty} (i\hbar)^{-f} \langle 0|\, \mathcal{H}_1(N) [R^0(z) \mathcal{H}_1(N)]^f\, |0\rangle_{\mathrm{irr}}, \tag{27}$$

where

$$R^0(z) = \int_0^\infty dt e^{izt} G(t). \tag{28}$$

The solution to Eq. (24), for the case that $\mathfrak{D} = 0$, can be expressed in terms of this operator $\psi(z)$. Operating on Eq. (24) by

$$\int_0^\infty dt e^{izt}$$

gives

$$-\rho_0(N,0) - iz\tilde{\rho}_0(N,z) = \psi(z)\tilde{\rho}_0(N,z)$$

on using the convolution theorem. In this equation,

$$\tilde{\rho}_0(N,z) = \int_0^t dt e^{izt} \rho_0(N,t). \tag{29}$$

The relation inverse to Eq. (29) is

$$\rho_0(N,t) = (2\pi)^{-1} \int_C dz e^{-izt} \tilde{\rho}_0(N,z), \tag{30}$$

where C is a contour in the z plane parallel to the real axis and above all singularities of the integrand. With Eq. (30), the solution to Eq. (24) is

$$\rho_0(N,t) = -(2\pi)^{-1} \int_C dz e^{-izt} [\psi(z) + iz]^{-1} \rho_0(N,0). \tag{31}$$

This important result shows that, at least if $\mathfrak{D} = 0$, the dynamics of a system of many degrees of freedom is contained in the inverse operator $[\psi(z) + iz]^{-1}$. The study of this operator for the model specified in Sec. II and the case of spontaneous emission will be the object of the remainder of the paper.

With the Hamiltonian, Eq. (6), the matrix elements of the operators $R^0(z)$ and $\mathcal{H}_1(N)$ can be readily calculated. Only the results will be stated here:

$$\langle \nu|\, R^0(z)\, |\nu\rangle = i \left(E(N_p + \tfrac{1}{2}\nu_p) - E(N_p - \tfrac{1}{2}\nu_p) + \sum_\lambda \nu_\lambda \omega_\lambda - z \right)^{-1}. \tag{32}$$

This is a *diagonal* operator. The (ν, N)-variables are given suffices p for the fermion states and λ for the states of the λth mode of the bosons. Further,

$$\hbar E(n) = \epsilon_1 \delta^{\mathrm{Kr}}(n-1) + \epsilon_2 \delta^{\mathrm{Kr}}(n-2),$$

where δ^{Kr} is the Kronecker delta. Then,

$$\begin{aligned}
\langle \nu|\, \mathcal{H}_1(N)\, |\nu'\rangle &= \sum_\lambda \langle \nu|\, \mathcal{H}_1^\lambda(N)\, |\nu'\rangle \\
&= \sum_\lambda \{\delta^{\mathrm{Kr}}(N_p + \tfrac{1}{2}(\nu_p' - 3)) \cdot [\delta^{\mathrm{Kr}}(\nu_p - \nu_p' - 1) \\
&\quad \times h_\lambda^*(2N_\lambda + \nu_\lambda' + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\lambda' - \nu_\lambda - 1) \\
&\quad + \delta^{\mathrm{Kr}}(\nu_p - \nu_p' + 1) h_\lambda (2N_\lambda + \nu_\lambda' + 1)^{\frac{1}{2}} \\
&\quad \times \delta^{\mathrm{Kr}}(\nu_\lambda - \nu_\lambda' - 1)] \eta^{\nu' - \nu} \\
&\quad - \delta^{\mathrm{Kr}}(N_p - \tfrac{1}{2}(\nu_p' + 3)) \cdot [\delta^{\mathrm{Kr}}(\nu_p - \nu_p' - 1) \\
&\quad \times h_\lambda^*(2N_\lambda - \nu_\lambda' + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\lambda' - \nu_\lambda - 1) \\
&\quad + \delta^{\mathrm{Kr}}(\nu_p - \nu_p' + 1) h_\lambda (2N_\lambda - \nu_\lambda' + 1)^{\frac{1}{2}} \\
&\quad \times \delta^{\mathrm{Kr}}(\nu_\lambda - \nu_\lambda' - 1)] \eta^{\nu - \nu'}\}.
\end{aligned} \tag{33}$$

The lowest-order term in the series (27) for $\psi(z)$ is that for which $f = 1$, the $f = 0$ term vanishing identically, since

$$\langle 0|\, \mathcal{H}_1(N)\, |0\rangle = 0.$$

With Eqs. (32) and (33), this lowest-order term becomes

$$\begin{aligned}
\psi(z) &= (i\hbar)^{-2} \langle 0|\, \mathcal{H}_1(N) R^0(z) \mathcal{H}_1(N)\, |0\rangle \\
&= (i\hbar)^{-2} \sum_{\nu\neq 0} \sum_{\lambda} \sum_{\mu} \langle 0|\, \mathcal{H}_1^{\lambda}(N)\, |\nu\rangle \langle \nu|\, R^0(z)\, |\nu\rangle \langle \nu|\, \mathcal{H}_1^{\mu}(N)\, |0\rangle \\
&= (i\hbar)^{-2} \sum_{\nu\neq 0} \sum_{\lambda} \sum_{\mu} \{\delta^{\mathrm{Kr}}(N_\nu + \tfrac{1}{2}(\nu_\nu - 3))[\delta^{\mathrm{Kr}}(\nu_\nu + 1) h_\lambda^*(2N_\lambda + \nu_\lambda + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\lambda - 1) \\
&\quad + \delta^{\mathrm{Kr}}(\nu_\nu - 1) h_\lambda (2N_\lambda + \nu_\lambda + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\lambda + 1)] \eta^{\nu} \\
&\quad - \delta^{\mathrm{Kr}}(N_\nu - \tfrac{1}{2}(\nu_\nu + 3))[\delta^{\mathrm{Kr}}(\nu_\nu + 1) h_\lambda^*(2N_\lambda - \nu_\lambda + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\lambda - 1) \\
&\quad + \delta^{\mathrm{Kr}}(\nu_\nu - 1) h_\lambda (2N_\lambda - \nu_\lambda + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\lambda + 1)] \eta^{-\nu}\} \\
&\quad \times i^{-1}[E(N_\nu + \tfrac{1}{2}\nu_\nu) - E(N_\nu - \tfrac{1}{2}\nu_\nu) + \sum_{\alpha} \nu_\alpha \omega_\alpha - z]^{-1} \\
&\quad \times \{\delta^{\mathrm{Kr}}(N_\nu - \tfrac{3}{2})[\delta^{\mathrm{Kr}}(\nu_\nu - 1) h_\mu^*(2N_\mu + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\nu + 1) \\
&\quad + \delta^{\mathrm{Kr}}(\nu_\nu + 1) h_\mu (2N_\mu + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\mu - 1)] \eta^{-\nu} \\
&\quad - \delta^{\mathrm{Kr}}(N_\nu - \tfrac{3}{2})[\delta^{\mathrm{Kr}}(\nu_\nu - 1) h_\mu^*(2N_\mu + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\mu + 1) \\
&\quad + \delta^{\mathrm{Kr}}(\nu_\nu + 1) h_\mu (2N_\mu + 1)^{\frac{1}{2}} \delta^{\mathrm{Kr}}(\nu_\mu - 1)] \eta^{\nu}\}. \qquad (34)
\end{aligned}$$

When Eq. (34) is multiplied out, one finds two contributions, expressible diagrammatically as shown in Fig. 1. These contributions may be added up to yield

$$\begin{aligned}
\psi(z) = \sum_{\lambda} \Big\{ &\frac{2}{(i\hbar)^2} \cdot |h_\lambda|^2 \cdot \frac{2iz}{z^2 - (\omega_\lambda - E)^2} \\
&\times [(N_\lambda + 1)\delta^{\mathrm{Kr}}(N_\nu - 2)(1 - \eta_\lambda^{-2,2}) \\
&+ N_\lambda \delta^{\mathrm{Kr}}(N_\nu - 1)(1 - \eta_\lambda^{2,-2})] \Big\}. \qquad (35)
\end{aligned}$$

In this, $E = E(2) - E(1)$ and

$$\eta_\lambda^{i,j} f(N_\nu, N_\lambda) = f(N_\nu + \tfrac{1}{2}i, N_\lambda + \tfrac{1}{2}j).$$

We now use Eq. (35) to obtain the solution Eq. (31) for the problem of spontaneous emission. This problem uses as initial condition

$$\rho_\nu(N, 0) = \delta^{\mathrm{Kr}}(\nu)\delta^{\mathrm{Kr}}(N_\nu - 2) \prod_{\lambda} \delta^{\mathrm{Kr}}(N_\lambda). \qquad (36)$$

In Laplace transform notation, Eq. (31) may be written

$$\rho_0(N, 0) = [\psi(z) + iz]\tilde{\rho}(N, z).$$

FIG. 1. The diagrammatic representation of the two contributions to $\psi(z)$.

That is,

$$\begin{aligned}
&\rho_0(N, 0) \\
&= iz\tilde{\rho}(N, z) + [4z/(i\hbar)^2] \sum_{\lambda} \{|h_\lambda|^2 \cdot [z^2 - (\omega_\lambda - E)^2]^{-1} \\
&\quad \times [(N_\lambda + 1)\delta^{\mathrm{Kr}}(N_\nu - 2)[\tilde{\rho}(N, z) - \tilde{\rho}(N_\nu - 1, \\
&\qquad N_\lambda + 1, \{N_\mu\}, z)] \\
&\quad + N_\lambda \delta^{\mathrm{Kr}}(N_\nu - 1)[\tilde{\rho}(N, z) - \tilde{\rho}(N_\nu + 1, \\
&\qquad N_\lambda - 1, \{N_\mu\}, z)]\}. \qquad (37)
\end{aligned}$$

We shall write as $\mathcal{N}$ the following choice of the N-variables: $N_\nu = 2$, $N_\mu = 0$ for all modes μ. With this choice, Eq. (37) becomes

$$\begin{aligned}
1 = iz\tilde{\rho}(\mathcal{N}, z) &+ [4z/(i\hbar)^2] \\
&\times \sum_{\lambda} \{|h_\lambda|^2 \cdot [z^2 - (\omega_\lambda - E)^2]^{-1} \\
&\times [\tilde{\rho}(\mathcal{N}, z) - \tilde{\rho}_1(\lambda, z)]\}, \qquad (38)
\end{aligned}$$

where

$$\tilde{\rho}_1(\lambda, z) = \tilde{\rho}\,(N_\nu = 1, N_\lambda = 1, N_{\mu\neq\lambda} = 0, z).$$

Then, with the choice of N-variables in the argument of $\tilde{\rho}$ above, Eq. (37) becomes

$$\begin{aligned}
0 = iz\tilde{\rho}_1(\lambda, z) + \frac{4z}{(i\hbar)^2} \cdot |h_\lambda|^2 \cdot \frac{1}{z^2 - (\omega_\lambda - E)^2} \\
\times [\tilde{\rho}_1(\lambda, z) - \tilde{\rho}(\mathcal{N}, z)]. \qquad (39)
\end{aligned}$$

Now, in the case of an infinite system, summations over the modes λ can be replaced by integrals over the

wave number k, with the following correspondences:

$$\frac{2\pi}{L}\sum_{\lambda} \rightarrow \int_{-\infty}^{+\infty} dk,$$

$$\omega_\lambda = c\,|k|.$$

Here L is the length of the system, which will tend to infinity, and c is the phase velocity of the bosons (velocity of light). With the choice of Eq. (11) for $|h_\lambda|^2$, Eq. (39) yields, in the limit $L \rightarrow \infty$,

$$\tilde{\rho}_1(\lambda, z) = 0.$$

Hence, Eq. (38) becomes

$$1 = iz\tilde{\rho}(\mathcal{N}, z) - \frac{2iz\alpha c^2}{\pi}\int_{-\infty}^{+\infty} dk \cdot \frac{|k|}{z^2 - [c\,|k| - E]^2} \cdot \tilde{\rho}(\mathcal{N}, z). \quad (40)$$

That is,

$$\tilde{\rho}(\mathcal{N}, z) = \frac{1}{iz}\left(1 - \frac{4\alpha c^2}{\pi}\int_0^{\infty} dk \cdot \frac{k}{z^2 - (k - E)^2}\right)^{-1}. \quad (41)$$

The integral in Eq. (41) is divergent and may be written as

$$-\frac{1}{2c^2}[\ln (z^2 - (ck - E)^2)]_0^{\infty} - \frac{E}{2zc^2}\left[\ln\left(\frac{E - z}{E + z}\right) + 2\pi i\right], \quad (42)$$

where that branch of the log function is chosen such that

$$\ln\left[\frac{E - z}{E + z}\right] = 0, \quad \text{for} \quad z = 0.$$

The divergence of the integral in Eq. (41) is the well-known ultraviolet divergence[7] and can be avoided in various ways. The simplest way is to cut off the divergent part of the integral at a finite upper limit, but this procedure is most inconvenient here. We shall instead modify the coupling constant $|h_\lambda|^2$ by removing its dependence on $|k|$: We replace the factor $|k|$ by its resonant value E/c. It can be noted that this is effectively what is done in the Born approximation of time-dependent perturbation theory in quantum mechanics for such problems. A further "justification" is that when this modification is made, the first term of Eq. (42) simply drops out, inasmuch as

$$-\frac{4\alpha cE}{\pi}\int_0^{\infty} dk \cdot \frac{1}{z^2 - (ck - E)^2} = \frac{2\alpha E}{\pi z}\left[\ln\left(\frac{E - z}{E + z}\right) + 2\pi i\right]. \quad (43)$$

[7] M. L. Goldberger and K. A. Watson, *Collision Theory* (J. Wiley & Sons, New York, 1963).

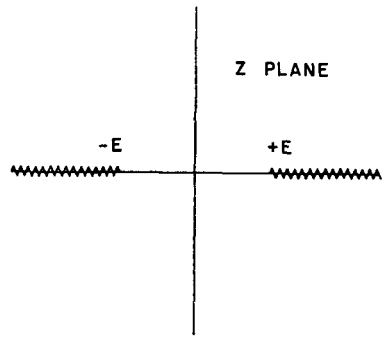


FIG. 2. The complex z plane, showing the branch points at $z = \pm E$, and the cuts along the real axis.

The second term is thus quite unaltered, and in Sec. IV it will be seen that it is this term only which contributes in the conventional theory to the solution, Eq. (41), in the limit of long times. We may hope, therefore, that our procedure preserves a good physical description of the system while ignoring the thorny problem of the ultraviolet divergence.

Accordingly, Eq. (41) becomes

$$\tilde{\rho}(\mathcal{N}, z) = \frac{1}{iz}\left[1 - \frac{2\alpha E}{\pi z}\left(\ln\left(\frac{E - z}{E + z}\right) + 2\pi i\right)\right]^{-1}.$$

That is,

$$\rho_0(\mathcal{N}, t) = -\frac{1}{2\pi}\int_C dz \times e^{-izt}\left[iz + \frac{2\alpha iE}{\pi}\ln\left(\frac{E - z}{E + z}\right) - 4\alpha E\right]^{-1}. \quad (44)$$

The integrand of Eq. (44) has branch points at $z = \pm E$, and so we shall define the integrand on the cut plane shown in Fig. 2.

There is a further singularity, the location of which may be determined by setting the real and imaginary parts of the denominator separately equal to zero. For $z = x + iy$ (x, y real), the imaginary part is

$$x + \frac{\alpha E}{\pi}\ln\left[\frac{(E - x)^2 + y^2}{(E + x)^2 + y^2}\right] = 0, \quad \text{for} \quad x = 0.$$

The real part for $x = 0$ is

$$y + \frac{2\alpha E}{\pi}\tan^{-1}\left[\frac{2yE}{y^2 - E^2}\right] + 4\alpha E.$$

The real part vanishes for only one value of y, obtainable by graphical means, or else by expanding in powers of α. To order α^2, one obtains

$$y = -4\alpha E - \frac{16\alpha^2 E}{\pi} + \cdots.$$

Thus, there is a pole with residue

$$-i(1 + 4\alpha/\pi + \cdots)$$

at

$$z = -4i\alpha E - 16i\alpha^2 E/\pi + \cdots.$$

Therefore, adding together the contributions from the cuts and this pole, Eq. (44) gives

$$\rho_0(\mathcal{N}, t) = \left(1 + \frac{4\alpha}{\pi}\right) \exp\left[-\left(4\alpha E + \frac{16\alpha^2 E}{\pi}\right)t\right] - \frac{1}{2\pi}\int_E^\infty dx \left\{ \frac{12\alpha E \cos xt + [2x + (4\alpha E/\pi) \ln |(E - x)/(E + x)|] \sin xt}{36\alpha^2 E^2 + [x + (2\alpha E/\pi) \ln |(E - x)/(E + x)|]^2} - \frac{4\alpha E \cos xt + [2x + (4\alpha E/\pi) \ln |(E - x)/(E + x)|] \sin xt}{4\alpha^2 E^2 + [x + (2\alpha E/\pi) \ln |(E - x)/(E + x)|]^2} \right\}. \quad (45)$$

IV. THE EXPONENTIAL SOLUTION

In this section, the customary, purely exponential, solution to our problem will be derived by making an "asymptotic" approximation. This approximation is valid for long times, by which we shall mean times of the order of $(\alpha E)^{-1}$. It is shown in many places (see, for example, Ref. 2) that, for long terms, the function $\mathcal{C}(t)$ of Eq. (25) will be very small, and so the master equation, Eq. (24), may be simplified to

$$\partial_t \rho_0(N, t) = \int_0^\infty d\tau \mathcal{C}(\tau)\rho_0(N, t) = \psi(+i0)\rho_0(N, t),$$

where $\psi(+i0)$ means that $\psi(z)$ is evaluated in the limit that $z \to 0$ from above, down the imaginary axis. Using Eq. (35) for $\psi(z)$, we obtain from this simplified master equation the result

$$\partial_t \rho_0(\mathcal{N}, t) = -\frac{4i\alpha c^2}{\pi}\int_0^\infty dk \cdot \frac{zk}{z^2 - (ck - E)^2}\bigg|_{z=+i0} \rho_0(\mathcal{N}, t) \quad (46)$$

and

$$\partial_t \rho_1(\lambda, t) = 0,$$

both in the infinite-system limit.

If in the integrand of Eq. (46) one sets $z = +i\epsilon$ $(\epsilon > 0)$ and

$$ck - E = \mathcal{E},$$

then one may note that

$$\lim_{\epsilon \to +0} \frac{\epsilon}{\pi(\epsilon^2 + \mathcal{E}^2)} = \delta(\mathcal{E}),$$

where δ is the Dirac delta function. Then, Eq. (46) becomes

$$\partial_t \rho_0(\mathcal{N}, t) = -4\alpha c^2 \int_0^\infty dk \cdot k \cdot \delta(ck - E)\rho_0(\mathcal{N}, t) = -4\alpha E \rho_0(\mathcal{N}, t). \quad (47)$$

Hence,

$$\rho_0(\mathcal{N}, t) = e^{-4\alpha E t} \quad (48)$$

with the given initial condition. Equation (48) is the usual exponential solution.

If the integral appearing in Eq. (46) is evaluated before proceeding to the limit $z = +i0$, then

$$-\frac{4i\alpha c^2}{\pi}\int_0^\infty dk \cdot \frac{kz}{z^2 - (ck - E)^2} = iz\left\{\frac{2\alpha}{\pi} \ln\left[\frac{z^2 - (c\mu - E)^2}{z^2 - E^2}\right] + \frac{2\alpha E}{\pi z}\left[\ln\left(\frac{E - z}{E + z}\right) + 2\pi i\right]\right\},$$

where μ is an upper cut-off for the divergent part of the integral. It is seen that for $z = +i0$ the first term on the right-hand side vanishes, as was remarked in Sec. III for the justification of the use of the modified $|h_\lambda|^2$. The second term yields $-4\alpha E$, in accord with Eq. (47).

If one compares the approximate solution, Eq. (48), with the solution obtained by an exact treatment of the master equation, Eq. (45), there are certain notable features. First, the exponent in the first term of Eq. (45) agrees with that in Eq. (48) only to first order. Inasmuch as the theory developed so far considers the dynamical operator $\psi(z)$ to its lowest order only, this is not, perhaps, very worrying. More serious is the presence of the second term in Eq. (45), which leads to the kind of nonexponential decay discussed in the previously quoted paper of Zwanzig,[1] in which he showed that this decay was slower than an exponential one. This clearly implies that, for sufficiently long times, this contribution will dominate, that is, it, not the exponential term, should be the long-time asymptotic solution of the master equation. An even more distressing feature is that the second term in Eq. (45) is not definite in sign, and so presents the possibility of obtaining a negative value for a quantity which by its physical nature must be nonnegative, being a probability for the occupation of a quantum state.

An examination of the assumptions of the asymptotic theory casts some light on these difficulties. In the approach elaborated by Prigogine and George,[8] the inverse operator $[\psi(z) + iz]^{-1}$ is expanded as

$$[\psi(z) + iz]^{-1} = \frac{1}{iz} \sum_{n=0}^{\infty} \left[\frac{-\psi(z)}{iz}\right]^n.$$

In the regime of long times, one then assumes that all the singularities of $\psi(z)$ lie in the lower half-plane of z. Then, since the contributions of these singularities to the inverse Laplace transform integral will all be damped by decaying exponentials, only the multiple pole at the origin need be considered. But, in Sec. III, we have seen that the singularities of $\psi(z)$ are not confined to the lower half-plane. The contributions from the cuts are evaluated along the real axis, and it is they which cause the breakdown of the asymptotic solution. Besides, the first term of Eq. (45), arising from the pole of $[\psi(z) + iz]^{-1}$, is indeed reproduced to first order by the asymptotic theory as given in this section, and will no doubt be reproduced to higher order by the more refined asymptotic theory of Prigogine and George (this will be seen to the next order in the following section). The absence of the second term of Eq. (45) from the solution, Eq. (48), is thus explained.

The possibility of negative values of $\rho_0(\mathcal{N}, t)$ remains, however. It is customary[7] to ignore this problem by claiming that the contributions from the cuts are in any event very small and probably nonphysical, since they appear to be proportional to a higher power of α than the exponential term. This should be a reason for neglecting these contributions in a treatment where only the lowest order of $\psi(z)$ is considered. Alternatively, it has been claimed[9] that, by considering higher orders in $\psi(z)$, one obtains further contributions to the solution, which, on being added to what one has already, yield a positive definite result. But the approach in the paper of Henin is so different that it is difficult to see whether or not exactly the same problem is being discussed. In any event, the second term in Eq. (45) has a very complicated dependence on α, so that it is not clear what part of it, if any, should be retained in our present approximation scheme. In the next section, this point will be clarified, and the reason for the confusion in the approximation will be pointed out.

[8] I. Prigogine, F. Henin, C. George, and F. Mayné, Physica **82**, 1828 (1966); F. Henin, I. Prigogine, and C. George, Physica **32**, 1873 (1966); I. Prigogine and F. Henin, "Kinetic Equation, Quasiparticles and Entropy," in *Statistical Mechanics*, T. A. Bak, Ed. (W. A. Benjamin, Inc., New York, 1967); C. George, Bull. Acad. Sci. Belg. **53**, 623 (1967); Physica **37**, 182 (1967); **39**, 251 (1968).

[9] F. Henin, Bull. Acad. Sci. Belg. **54**, 585 (1968); F. Henin and M. De Haan, Physica **40**, 399 (1968).

V. THE DISCREPANCY BETWEEN THE EXACT AND THE STRICTLY EXPONENTIAL SOLUTION

So far the independent parameters of the solution that we have obtained to the master equation have been taken as α, a dimensionless coupling constant, and t, the time, usually scaled by the frequency E and thus occurring in the combination Et. However, a little thought shows that this is a poor choice. Since the lowest-order term of $\mathcal{C}(t)$, Eq. (25), is proportional to α, we may write

$$\mathcal{C}(t) = \alpha\hat{\mathcal{C}}(t),$$

where $\hat{\mathcal{C}}(t)$ has a term independent of α. Then the master equation can be written

$$\frac{\partial}{\partial(\alpha t)} \rho_0(N, \alpha t) = \int_0^{\alpha t} d\tau \hat{\mathcal{C}}(\tau)\rho_0(N, \alpha t - \tau),$$

where now t always occurs in the combination αt. Of course, t may be associated with higher powers of α. This means that in the master equation t is *scaled* by α. Thus, if one tries to keep α and t as independent parameters and then to consider only the lowest-order terms in α, the higher-order terms in t are thrown away with those in α. But, since it is the aim of our calculations to examine the full time-dependence, this is unsatisfactory. It is clear that the proper procedure is to consider as independent, dimensionless parameters α and, say, $\alpha Et = \tau$. In this way, one may consistently retain all dependence on τ, and disregard all but the lowest-order dependence in α.

Let us return to Eq. (31) and carry out this scheme. We shall suppress, for convenience, the N-dependence in the following manipulations. Conjugate to $\tau = \alpha Et$, we define the variable $\xi = z/\alpha E$ and then make the further definitions:

$$\rho_0(t) = \rho_0(\tau/\alpha E) = \rho(\tau),$$

$$\hat{\rho}(\xi) = \int_0^{\infty} d\tau e^{i\xi\tau}\rho(\tau) = dE \int_0^{\infty} dt e^{i(\alpha E\xi)t}\rho_0(t)$$

$$= \alpha E\tilde{\rho}(\alpha E\xi). \tag{49}$$

Now, from Eq. (31),

$$[\psi(z) + iz]\tilde{\rho}(z) = -\rho_0(0). \tag{50}$$

Hence,

$$[\psi(\alpha E\xi) + i\alpha E\xi]\tilde{\rho}(\alpha E\xi) = -\rho_0(0).$$

Further, if we put

$$\psi(\alpha E\xi) = \alpha E\hat{\psi}(\xi), \tag{51}$$

then

$$[\hat{\psi}(\xi) + i\xi]\hat{\rho}(\xi) = -\rho_0(0). \tag{52}$$

For the moment, we shall assume that all the quantities in Eq. (52) can be expanded as power series in α, this with a view to retaining only the lowest-order terms. It can be seen at once that the lowest term of $\psi(\xi)$ is indeed independent of α. We write

$$\hat{\psi}(\xi) = \sum_{n=0}^{\infty} \alpha^n \psi^{(n)}(\xi), \tag{53a}$$

$$\hat{\rho}(\xi) = \sum_{n=0}^{\infty} \alpha^n \hat{\rho}_n(\xi) \tag{53b}$$

with the coefficients of α in these series *independent* of α. The equating of coefficients of powers of α when the relations, Eqs. (53), are substituted in Eq. (52) yields

$$[\psi^{(0)}(\xi) + i\xi]\hat{\rho}_0(\xi) = -\rho_0(0), \tag{54a}$$

$$\psi^{(1)}(\xi)\hat{\rho}_0(\xi) + \psi^{(0)}(\xi)\hat{\rho}_1(\xi) + i\xi\hat{\rho}_1(\xi) = 0. \tag{54b}$$

For the case of spontaneous emission, Eq. (44) may be written

$$\left[\frac{2i\alpha E}{\pi} \ln\left(\frac{E-z}{E+z}\right) - 4\alpha E + iz\right]\tilde{\rho}(z) = -1. \tag{55}$$

This has the *form* of Eq. (50), although the full operator $\psi(z)$ has not been written down. We may thus use the results of Eq. (54), with

$$\hat{\rho}(\xi) = \alpha E \tilde{\rho}(\alpha E \xi)$$

and

$$\hat{\psi}(\xi) = \frac{2i}{\pi} \ln\left(\frac{1-\alpha\xi}{1+\alpha\xi}\right) - 4. \tag{56}$$

Now, $\hat{\psi}(\xi)$ can be expanded as a series in α:

$$\hat{\psi}(\xi) = -4 - (4i/\pi)[\alpha\xi + \tfrac{1}{3}\alpha^3\xi^3 + \cdots] \tag{57}$$

for $|\xi| < \alpha^{-1}$. So, then $\psi^{(0)}(\xi) = -4$ and

$\psi^{(1)}(\xi) = -4i\xi/\pi +$ a term from the next term in the series for $\psi(z)$.

We shall, for the moment, ignore the term for the α^2 term in $\psi(z)$, but merely note that consistency demands that it be included here.

With the values found for $\psi^{(0)}(\xi)$ and $\psi^{(1)}(\xi)$, Eq. (54) beomes

$$(-4 + i\xi)\hat{\rho}_0(\xi) = -1 \tag{58a}$$

and

$$-(4i\xi/\pi)\hat{\rho}_0(\xi) + (-4 + i\xi)\hat{\rho}_1(\xi) = 0. \tag{58b}$$

That is,

$$\hat{\rho}_0(\xi) = (4 - i\xi)^{-1}$$

and

$$\hat{\rho}_1(\xi) = (4 - i\xi)^{-2} \cdot (-4i\xi/\pi).$$

Inverse Laplace transformation yields

$$\rho^0(\tau) = e^{-4\tau}, \tag{59a}$$

$$\rho^1(\tau) = (4/\pi) \cdot e^{-4\tau}(1 - 4\tau), \tag{59b}$$

where the expansion

$$\rho(\tau) = \sum_{n=0}^{\infty} \alpha^n \rho^n(\tau)$$

corresponds to Eq. (53b). In terms of α and t, Eq. (59a) becomes, to the order calculated,

$$\rho_0(t) = \rho(\alpha E t) = e^{-4\alpha E t}[1 + (4\alpha/\pi)(1 - 4\alpha E t)]. \tag{60}$$

To first order in α, this is equal to

$$\rho_0(\tau) = [1 + 4\alpha/\pi] \exp(-4\alpha E t - 16\alpha^2 E t/\pi). \tag{61}$$

We have so far ignored the restriction of Eq. (57); that is, for the expansion to be valid, one must have $|\xi| < 1/\alpha$. Now, this restriction is clearly *not* satisfied for all values of ξ on the contour C of the inverse Laplace transform integral, and so the results, Eqs. (59), cannot be considered trustworthy. However, for small α, the parts of the contour where $|\xi| > 1/\alpha$ will be distant from the origin, and so (in particular, for τ not too small) will yield contributions to the integral which will be incoherent and may be unimportant. It can be seen, too, that the expansion of $\hat{\psi}(\xi)$ in powers of α is effectively the same expansion as that performed in the asymptotic theory of Prigogine and George[8] and does indeed yield agreement to the next order in α, with the exponent obtained in the exact solution [Eqs. (45) and (61)], as remarked in the preceding section.

Whether or not Eq. (60) is a good approximation to the exact solution, Eq. (45), cannot, of course, be settled by the kind of arguments posed in the preceding paragraph. We are seeking a consistent approximation in terms of the parameters α and τ, and must therefore express Eq. (45) itself in terms of these, bearing in mind, of course, that $\rho^1(\tau)$ in Eq. (59) may be consistently included only if something of the α^2 term in $\psi(z)$ is accounted for, and this is certainly not so in Eq. (45). This latter equation then becomes

$$\rho(\tau) = \left(1 + \frac{4\alpha}{\pi}\right)e^{-4\tau - 16\alpha\tau/\pi} - \frac{1}{2\pi}\int_0^{\alpha} d\xi \left(\frac{12\cos(\tau/\xi) + 2[\xi^{-1} + (2/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]\sin(\tau/\xi)}{36\xi^2 + [1 + (2\xi/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]^2} - \frac{4\cos(\tau/\xi) + 2[\xi^{-1} + (2/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]\sin(\tau/\xi)}{4\xi^2 + [1 + (2\xi/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]^2}\right). \tag{62}$$

The reason for the confusion in our approximation scheme and for the breakdown of the series expansion for $\psi(\xi)$ is now apparent from the integral term in Eq. (62). The reason is that this term is *nonanalytic* in α at $\alpha = 0$. Were the integrand a well-behaved function, the integral would clearly be proportional to α in the lowest order, and the solution, Eqs. (59), would hold. But, as it is, we must ask whether the integral in Eq. (62) is for small α comparable to the expression $(4\alpha/\pi)e^{-4\tau}$, which is, to appropriate order in α, the correction to the purely exponential term $e^{-4\tau-16\alpha\tau/\pi}$ in Eq. (60). The arguments advanced earlier would suggest that this will be so if τ is not too small. The answer to this question is considered in the next section.

VI. NUMERICAL COMPARISON

In this section, we describe briefly the results of several numerical experiments which have been performed which relate to the various questions raised in the preceding paragraphs. To first order in α, the difference between the approximate solution, Eq. (60), and the exact solution, Eq. (62), may be computed directly. For convenience of notation, we denote the approximate correction to the strictly exponential term as $\rho_D(\tau)_A$, where

$$\rho_D(\tau)_A = (4\alpha/\pi)\cdot e^{-4\alpha Et} = (4\alpha/\pi)\cdot e^{-4\tau}. \quad (63)$$

Similarly, we denote the exact correction to the exponential terms as $\rho_D(\tau)_E$, where

$$\rho_D(\tau)_E = -\frac{1}{2\pi}\int_0^\alpha d\xi \left(\frac{12\cos(\tau/\xi) + 2[\xi^{-1} + (2/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]\sin(\tau/\xi)}{36\xi^2 + [1 + (2\xi/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]^2} - \frac{4\cos(\tau/\xi) + 2[\xi^{-1} + (2/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]\sin(\tau/\xi)}{4\xi^2 + [1 + (2\xi/\pi)\ln|(\xi-\alpha)/(\xi+\alpha)|]^2} \right). \quad (64)$$

In Figs. 3–5 we present the results obtained upon computing $\rho_D(\tau)_E$ and $\rho_D(\tau)_A$ for three representative values of the coupling parameter α, corresponding to weak coupling ($\alpha = 0.20$), intermediate coupling ($\alpha = 0.40$), and strong coupling ($\alpha = 0.80$). It is evident from the study of these figures that the profile of the exact correction term is characterized by oscillations which gradually damp out with increasing τ. Since all three figures were constructed to span a range of τ values from 0.0 to 5.0, it is seen immediately that the smaller the coupling parameter, the more rapid are the oscillations. Phrased differently, the oscillatory character of $\rho_D(\tau)_E$ persists longer for strongly coupled systems than for weakly coupled systems.

On the other hand, one notices that the profile of the approximate correction term $\rho_D(\tau)_A$ decreases monotonically with increasing τ. Since the structures of the curves representing $\rho_D(\tau)_E$ and $\rho_D(\tau)_A$ are so different, we employ the following device to aid in their comparison: We look for that value of τ beyond which the absolute value of the particular correction term under

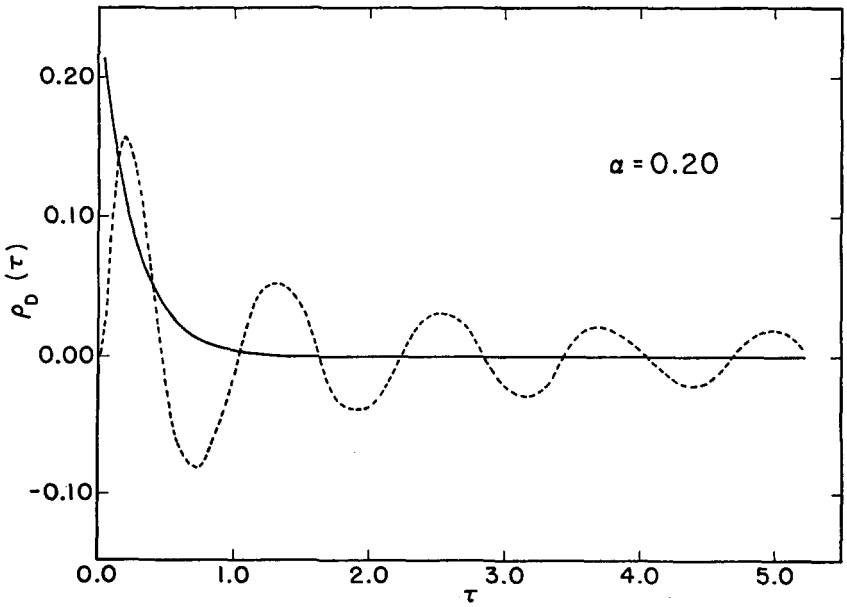


FIG. 3. Comparison of the exact (indicated by a dashed line) vs approximate (indicated by a solid line) corrections to the strictly exponential solution for $\alpha = 0.20$.

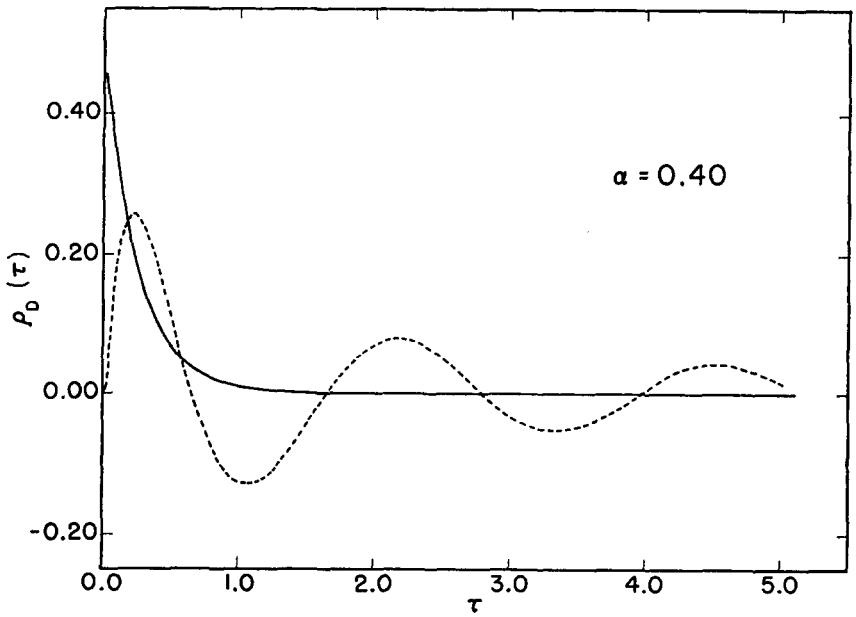


FIG. 4. Comparison of the exact (dashed line) vs approximate (solid line) corrections to the strictly exponential solution for $\alpha = 0.40$.

consideration, say $|\rho_D(\tau)|$, finally falls below a certain predetermined value. We then assert that any differences between the two curves, which occur in regions characterized by a value of $|\rho_D(\tau)|$ less than this preassigned value, can be neglected. The use of such a device corresponds to a kind of "high-resolution" versus "low-resolution" probe for determining the success or failure of the suggested approximation scheme, as will become clear in the next paragraph.

Two values of $|\rho_D(\tau)|$ were chosen as representative, 0.25 and 0.025. The criterion suggested above was applied to the curves obtained for the following six values of α: 0.01, 0.10, 0.20, 0.40, 0.60, and 0.80. In the previous section, it was suggested that, for small α and for τ sufficiently large, the calculation of $\rho_D(\tau)_E$ and $\rho_D(\tau)_A$ should yield results of comparable orders of magnitude. Examination of Table I reveals that this is approximately true, that is, the apparent differences between the approximate and exact correction terms are least important when the coupling parameter is small. This information (on the range of values of α and τ for which the approximate theory provides a

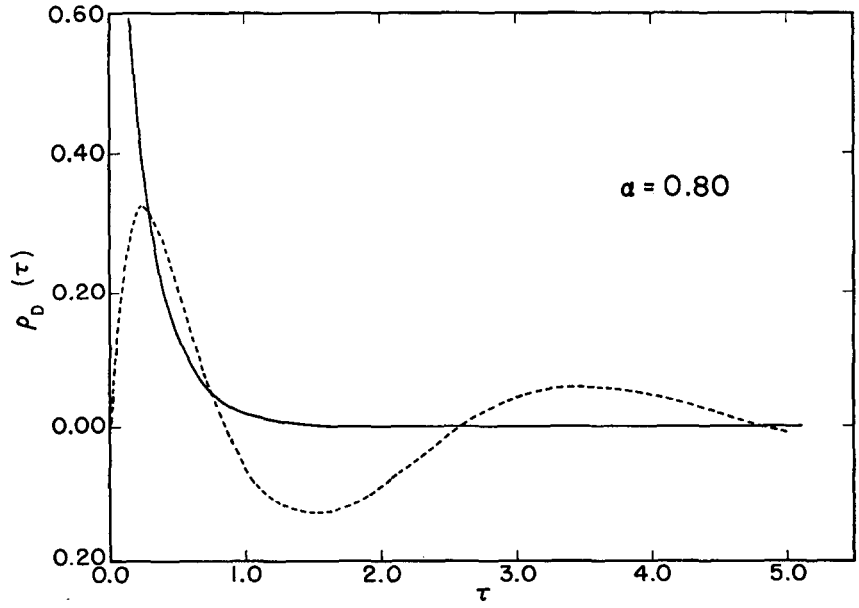


FIG. 5. Comparison of the exact (dashed line) vs approximate (solid line) corrections to the strictly exponential solution for $\alpha = 0.80$.

TABLE I. Comparison of exact vs approximate correction terms.

	$\lvert\rho_D(\tau)\rvert < 0.25$		$\lvert\rho_D(\tau)\rvert < 0.025$	
α	Approximate	Exact	Approximate	Exact
0.01	$\tau > 0$	$\tau > 0$	$\tau > 0$	$\tau > 0$
0.10	$\tau > 0$	$\tau > 0$	$\tau > 0.40$	$\tau > 0.50$
0.20	$\tau > 0.05$	$\tau > 0$	$\tau > 0.55$	$\tau > 3.25$
0.40	$\tau > 0.18$	$\tau > 0.28$	$\tau > 0.75$	$\tau > 7$
0.60	$\tau > 0.28$	$\tau > 0.38$	$\tau > 0.85$	$\tau > 10$
0.80	$\tau > 0.35$	$\tau > 0.45$	$\tau > 0.93$	$\tau > 20$

reasonable representation of the exact solution) will be utilized in a detailed way in a subsequent paper when the present calculation is extended to deal with more general situations (see Sec. VII).

VII. DISCUSSION

In this last section, we summarize and discuss the principal results obtained in this paper and then mention several areas of study, related to the present problem, which are now under investigation by the authors.

Perhaps the most interesting result of this work is the apparent confirmation of the Zwanzig suggestion, that exponential decay in time may be hidden behind some other, more complicated kind of time dependence. Indeed, for the model under study here, which has been solved exactly to first order in the coupling parameter α, a rather complicated nonexponential behavior is observed when one studies the decay in τ of the diagonal elements of the density matrix ρ. Again, this behavior is illustrated in Figs. 3–5 for three representative values of α.

The approximate theory, developed in Sec. IV, compares favorably with the exact theory for small values of α and for τ sufficiently long. Even though the actual structure of the two correction terms $\rho_D(\tau)_E$ and $\rho_D(\tau)_A$ is quite different, provided one is looking at the behavior of these two terms relative to a certain preassigned "decay level," the data given in Table I indicate that $\rho_D(\tau)_E$ and $\rho_D(\tau)_A$ can be considered of comparable magnitude for $\alpha < 0.20$.

It can be said that the approximate correction term is, in some sense, more "physical" than the exact correction term. This is because $\rho_D(\tau)_E$ oscillates about zero and, hence, there exist values of τ for which the correction term is negative. As indicated earlier, there are really two comments one can make concerning these negative values of $\rho_D(\tau)_E$.

The first suggestion relates to the possibility that the contributions from the cuts are numerically small compared to other, purely exponential contributions to the full solution, Eq. (62). An examination of the full solution, Eq. (62), for a coupling of $\alpha = 0.10$ reveals that, relative to the term $e^{-4\tau}(1 - 16\alpha\tau/\pi)$, the correction $\rho_D(\tau)_E$ is indeed negligible if one neglects all contributions for which $|\rho_D(\tau)_E| < 0.025$. Incidentally, from Table I we know that this corresponds to a value of $\tau \geq 0.5$. If one did not insist on this requirement, then even the full solution would eventually exhibit negative values, since, as may be inferred from Figs. 3–5, oscillatory behavior persists long after the exponential-type solutions have died away. In applications of the theory to problems of physical interest, the appropriate cut-off will be determined by the sensitivity of the instrument and the measurement being performed.

The second possible suggestion to account for negative contributions to $\rho_D(\tau)_E$ is the observation that, in fact, the above calculation has been performed only to lowest order in the coupling parameter α. As a consequence, there may well be contributions from other higher-order terms which add in such a way as to cancel out the over-all effect of these anomalous, unphysical contributions to $\rho_D(\tau)_E$. This, as nearly as we understand it, is the suggestion of Henin.[9] The importance of this suggestion is that it relaxes the requirement that the correction term $\rho_D(\tau)_E$ must necessarily be small as compared to other terms in the full solution. The straightforward way to check this suggestion, within the framework of the present model, is to extend the calculation to the next higher order in the coupling parameter and see if there is any evidence for this kind of cancellation. This work is now in progress.

There is yet a third possible explanation for the apparently unphysical behavior of the exact correction term. This explanation relates to the recent work of Lebowitz and Percus.[10] These authors investigated the time evolution of the distribution function of a labeled particle in a one-dimensional system of hard rods of diameter a. They discovered that if one expanded the inverse of the Laplace transform of the velocity autocorrelation function as a power series in $n = \rho/(1 - a\rho)$, then the resulting series, of the form $\sum_l B_l n_l / z^{l-1}$, diverged for $l \geq 2$ and when $z \to 0$. This result was in contrast to the one obtained when the problem was solved exactly. In the latter case, an explicit closed expression for the time-displaced, self-distribution function ρ_s was obtained. With regard to

[10] J. L. Lebowitz and J. K. Percus, Phys. Rev. **155**, 122 (1967).

[11] L. Van Hove, "Master Equation and Approach to Equilibrium for Quantum Systems," in *Fundamental Problems in Statistical Mechanics*, E. D. G. Cohen, Ed. (North-Holland Publ. Co., Amsterdam, 1962),Vol. I; L. Van Hove and E. Verboven, Physica **27**, 418 (1961); A. Janner, L. Van Hove, and E. Verboven, *ibid.* **28**, 1341 (1962); E. Verboven and L. Buyst, *ibid.* **29**, 653 (1963).

the present investigation, perhaps the Lebowitz–Percus result is an indication of the fact that, even though our calculation is exact to first order in α, it has still been carried out within the framework of a perturbation-type expansion, and, as such, would necessarily suffer from the kind of difficulties mentioned in the preceding paragraphs.

Nonexponential behavior has also been observed in the model introduced by Van Hove and coworkers[11]—namely, the evolution to quantum-statistical equilibrium of an electron interacting with a collection of randomly distributed, static, elastic scattering centers. As it will be the chief purpose of one of the papers in this series to investigate in detail the relation between our model and the one introduced by Van Hove, we shall not enter into a detailed discussion here. Suffice it to say that the model of Van Hove was also investigated in the limits of weak, intermediate, and strong coupling, and both exponential and nonexponential behavior was exhibited by the diagonal part of the transition probability density, depending on the coupling.

Calculations similar in spirit to that of this paper have been performed by Rubin.[12] He has considered in detail the mechanics of a single heavy particle in a lattice, a problem with many similarities, both physical and formal, to that treated here. Error bounds have been calculated for the expression of the momentum autocorrelation function as a pure exponential in several cases, including that of a finite system. Further, Ullersma,[13] in the context of an exactly solvable model for Brownian motion, has examined the circumstances under which a bound electron in an electromagnetic field may be described by a Langevin equation (implying essentially exponential behavior) and has given exhaustive sets of conditions which are needed for such a description.

Finally, we mention that the theory presented in this paper might possibly be applied to clarify certain aspects of the Ernst method for enhancement of sensitivity in magnetic resonance.[14] Indeed, a preliminary study has revealed that the above theory, properly generalized, may well provide a basis for a more fundamental understanding of pulse experiments in magnetic resonance.

ACKNOWLEDGMENT

The authors are indebted to Professor I. Prigogine for many interesting and stimulating discussions.

[12] R. J. Rubin, J. Math. Phys. **1**, 309 (1960); R. J. Rubin and P. Ullersma, *ibid.* **7**, 1877 (1966); R. J. Rubin, J. Am. Chem. Soc. **90**, 3061 (1968).

[13] P. Ullersma, Physica **32**, 27, 56, 74, 90 (1966).

[14] R. R. Ernst and W. A. Anderson, Rev. Sci. Instr. **37**, 93 (1966); R. R. Ernst, "Sensitivity Enhancement in Magnetic Resonance," in *Advances in Magnetic Resonance*, J. S. Waugh, Ed. (Academic Press Inc., New York, 1966), Vol. II.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Frequency Spectrum of a One-Dimensional Lattice with an Isotopic Impurity

EARL T. KINZER
Auburn University, Auburn, Alabama

(Received 7 May 1969)

The frequency spectrum of a one-dimensional lattice with a substitutional isotopic impurity is calculated directly and compared with previous calculations for the correction to the spectrum of a perfect lattice. The calculation makes use of certain properties of the delta function to facilitate the derivation of the frequency spectrum. Minor corrections to previous results are presented.

I. INTRODUCTION

In the past ten years several calculations have been made of the correction $\Delta g(\omega)$ to the frequency spectrum $g(\omega)$ for a one-dimensional lattice in the harmonic approximation when one of the lattice particles possesses a mass which is different from all the rest (isotopic substitutional impurity). The principal calculations have been presented by Mahanty, Maradudin, and Weiss,[1] Litzman,[2] and Maradudin.[3] All of these have been special cases of a more general calculation of additive functions of the normal-mode frequencies. It is the purpose of this paper to present a more direct and intuitively obvious calculation of the spectrum for the defect lattice and to compare this calculation with previous results.

The equations of motion for the lattice of N particles with an isotope impurity are

$$m\frac{d^2\phi(n,t)}{dt^2} = K[\phi(n+1,t) + \phi(n-1,t) - 2\phi(n,t)], \quad n \neq p,$$

$$M\frac{d^2\phi(p,t)}{dt^2} = K[\phi(p+1,t) + \phi(p-1,t) - 2\phi(p,t)]. \tag{1}$$

The displacement from equilibrium of the nth particle is given by $\phi(n,t)$, where K is the spring constant, M is the mass of the impurity particle, and m is the mass of each of the remaining lattice particles. In addition, we require that the periodic boundary condition be satisfied, i.e.,

$$\phi(n,t) = \phi(n+N,t). \tag{2}$$

For the special case in which $M = m$, the equations may be solved by assuming solutions of the form $e^{i(n\theta-\omega t)}$. They lead to a dispersion relation

$$\omega = \Omega\,|\sin(\tfrac{1}{2}\theta)|, \quad \Omega = 2(K/m)^{\frac{1}{2}}, \tag{3}$$

and an equation for the determination of the allowed values of θ,

$$[\sin(\tfrac{1}{2}N\theta)]^2 = 0. \tag{4}$$

From this we conclude that $\theta(k) = 2\pi k/N$, where k is an integer which is eventually identified as the wavenumber. Equation (4) can also be used to confirm that the frequencies are doubly degenerate for this model and k can be restricted to the first Brillouin zone so that

$$\omega(k) = \Omega \sin(\tfrac{1}{2}\theta(k)), \quad 0 \leq \theta(k) \leq \pi. \tag{5}$$

For the impurity problem, the usual methods of solution lead, with slightly more difficulty, to a similar dispersion relation

$$\omega = \Omega\,|\sin(\tfrac{1}{2}\theta)|, \quad \Omega = 2(K/m)^{\frac{1}{2}}, \tag{6}$$

and, with considerable calculation,[4,5] to an equation for the determination of the allowed values of θ,

$$\sin(N\theta/2)[\sin(\tfrac{1}{2}N\theta) - \epsilon\tan(\tfrac{1}{2}\theta)\cos(N\theta/2)] = 0, \tag{7}$$

where $\epsilon = 1 - (M/m)$. It is clear that the antisymmetric modes (those having a node at the impurity) are unaffected, while the symmetric modes are altered by the presence of the impurity. For the antisymmetric modes, $\theta(k) = 2\pi k/N$ as before, while, for the symmetric modes, $\theta(k)$ is determined by

$$\tan(N\theta/2) = \epsilon\tan(\tfrac{1}{2}\theta), \tag{8}$$

which can be solved by numerical methods. A linear relationship between $\theta(k)$ and k no longer holds; but Eq. (8) can be used to find the frequency spectrum through a differential relationship between $\theta(k)$ and k.

[1] J. Mahanty, A. Maradudin, and G. Weiss, Progr.Theoret. Phys. (Kyoto) **20**, 369 (1958).
[2] O. Litzman, Czech. J. Phys. **9**, 692 (1959).
[3] A. Maradudin, *Phonons and Phonon Interactions*, T. Bak, Ed. (W. A. Benjamin, Inc., New York, 1964), p. 424.
[4] E. W. Montroll and R. B. Potts, Phys. Rev. **100**, 525 (1955).
[5] J. Hori and T. Asahi, Progr. Theoret. Phys. (Kyoto) **17**, 523 (1957).

In Sec. II this calculation is performed and discussed and in Sec. III it is compared with previous results.

II. THE FREQUENCY SPECTRUM

Bowers and Rosenstock[6] have shown that the frequency spectrum may be represented as a sum of δ functions

$$g(\omega) = \frac{1}{N}\sum_k \delta(\omega - \omega(k)). \tag{9}$$

Using well-known properties of the δ function, we can write Eq. (9) in terms of θ as

$$g(\omega) = \frac{1}{N}\sum_k \delta(\theta - \theta(k)) \Big/ \left|\frac{d\omega}{d\theta}\right|_{\theta=\theta(k)}. \tag{10}$$

If we define $\xi(k) = 2\pi k/N$, Eq. (10) can be written

$$g(\omega) = \frac{1}{N}\sum_k \delta(\xi - \xi(k)) \Big/ \left(\left|\frac{d\omega}{d\theta}\right|\left|\frac{d\theta}{d\xi}\right|\right)_{\xi=\xi(k)}. \tag{11}$$

Making use of another property of the δ function, we obtain

$$g(\omega) = \left(N\left|\frac{d\omega}{d\theta}\right|\left|\frac{d\theta}{d\xi}\right|\right)^{-1}\sum_k \delta(\xi - \xi(k)). \tag{12}$$

We can also write

$$\frac{1}{N}\sum_k \delta(\xi - \xi(k)) = \frac{1}{2\pi}\sum_k \delta(\xi - \xi(k))\Delta\xi(k), \tag{13}$$

where $\Delta\xi(k) = 2\pi\Delta k/N = 2\pi/N$. As N becomes large, the sum in Eq. (13) tends to an integral, i.e.,

$$\frac{1}{2\pi}\sum_k \delta(\xi - \xi(k))\Delta\xi(k) \to \frac{1}{2\pi}\int_0^\pi \delta(\xi - x)\,dx. \tag{14}$$

Thus we have, for the in-band frequencies,

$$\begin{aligned} g(\omega) &= \left(2\pi\left|\frac{d\omega}{d\theta}\right|\left|\frac{d\theta}{d\xi}\right|\right)^{-1}, && \omega < \Omega', \\ &= 0, && \omega > \Omega', \end{aligned} \tag{15}$$

where Ω' is the largest real solution of Eq. (8). If, however, one of the frequencies is modified to such an extent that $\omega(d) > \Omega'$ (a localized mode), the corresponding term in Eq. (12) must be handled separately. The frequency spectrum is then given by

$$\begin{aligned} g(\omega) &= \left(2\pi\left|\frac{d\omega}{d\theta}\right|\left|\frac{d\theta}{d\xi}\right|\right)^{-1}, && \omega < \Omega', \\ &= \frac{1}{N}\delta(\omega - \omega(d)), && \omega > \Omega'. \end{aligned} \tag{16}$$

In this case, the maximum frequency has been shifted out of the band of allowed frequencies for the perfect lattice and appears as a localized mode frequency associated with the defect.

As examples, we treat first the perfect lattice, obtaining the well-established result, and then the same lattice with an isotopic substitutional impurity.

For the perfect lattice, we have

$$\omega = \Omega\,|\sin(\tfrac{1}{2}\theta)|, \quad 0 < \theta < \pi, \tag{17}$$

and, in this range of θ, we obtain

$$\left|\frac{d\omega}{d\theta}\right| = \tfrac{1}{2}\Omega\cos\tfrac{1}{2}\theta = \tfrac{1}{2}(\Omega^2 - \omega^2)^{\frac{1}{2}}. \tag{18}$$

To obtain $d\theta/d\xi$, we refer to Eq. (4) which implies that $\theta(k) = 2\pi k/N$. From the definition of $\xi(k)$, we see that $\theta(k) = \xi(k)$, and the values of ξ become dense in the range $0 < \xi < \pi$ so that $|d\theta/d\xi| = 1$ in this range. Since the calculation is identical for the symmetric and the antisymmetric modes, the spectrum is given by

$$\begin{aligned} g(\omega) &= (2/\pi)(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega, \\ &= 0, && \omega > \Omega, \end{aligned} \tag{19}$$

where the factor of two in the numerator comes from counting each frequency twice.

For the lattice with an isotopic defect, we again have

$$\omega = \Omega\,|\sin(\tfrac{1}{2}\theta)|, \quad 0 \leq \theta \leq \pi, \tag{20}$$

and, as before,

$$\left|\frac{d\omega}{d\theta}\right| = \tfrac{1}{2}(\Omega^2 - \omega^2)^{\frac{1}{2}}. \tag{21}$$

To obtain $d\theta/d\xi$, we recall that the antisymmetric modes are unaffected and, for these modes, $d\theta/d\xi = 1$. For the symmetric modes we refer to Eq. (8), which can be written as

$$\tan(\tfrac{1}{2}N\theta) - \epsilon\tan(\tfrac{1}{2}\theta) = 0.$$

Although the equation cannot be solved explicitly for θ, an implicit differentiation can be performed providing θ is made single-valued. This is done by considering the solutions, for integer values of k, of the equation

$$\tan(\tfrac{1}{2}N\theta - \pi k) - \epsilon\tan(\tfrac{1}{2}\theta) = 0. \tag{22}$$

In effect, we have restricted the tangent functions to their principal ranges [$\tan(\frac{1}{2}\theta)$ was already so restricted by the condition $0 \leq \theta \leq \pi$]. In terms of $\xi = 2\pi k/N$, we have

$$\tan(\tfrac{1}{2}N)(\theta - \xi) - \tan(\tfrac{1}{2}\theta) = 0. \tag{23}$$

On differentiating Eq. (23) with respect to ξ and solving for $d\theta/d\xi$, we obtain

$$\left|\frac{d\theta}{d\xi}\right| = \left(1 - \frac{\epsilon}{N}\frac{1}{\cos^2(\frac{1}{2}\theta) + \epsilon^2\sin^2(\frac{1}{2}\theta)}\right)^{-1}. \tag{24}$$

[6] W. Bowers and H. Rosenstock, J. Chem. Phys. 18, 1056 (1950).

In terms of ω, Eq. (24) becomes

$$\left|\frac{d\theta}{d\xi}\right| = \left(1 - \frac{\epsilon}{N}\frac{1}{\Omega^2 + (\epsilon^2 - 1)\omega^2}\right)^{-1}. \quad (25)$$

When $-\infty < \epsilon \leq 0$, Eq. (8) has $\frac{1}{2}N$ distinct real solutions for integer values of k. When $0 \leq \epsilon \leq 1$, Eq. (8) has $\frac{1}{2}N - 1$ distinct real solutions for integer values of k and one complex solution which corresponds to a frequency greater than the maximum frequency of the perfect lattice. As Montroll[4] has noted, the defect frequency corresponds to $\theta = \pi + iz$ and, for large N, is given by

$$\omega(d) = \Omega(1 - \epsilon^2)^{-\frac{1}{2}}. \quad (26)$$

Thus, for $-\infty \leq \epsilon \leq 0$,

$$\begin{aligned} g(\omega) &= \pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega, \\ &= +\pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \\ &\quad \times \{1 - (\epsilon\Omega^2/N)[\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1}\}, && \omega \leq \Omega', \quad (27) \\ &= 0, && \omega > \Omega, \end{aligned}$$

and for $0 \leq \epsilon \leq 1$,

$$\begin{aligned} g(\omega) &= \pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega, \\ &= +\pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \\ &\quad \times \{1 - (\epsilon\Omega^2/N)[\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1}\}, && \omega \leq \Omega', \quad (28) \\ &= +N^{-1}\delta(\omega - \Omega(1 - \epsilon^2)^{-\frac{1}{2}}), && \omega \geq \Omega, \end{aligned}$$

where Ω' denotes the largest real solution of Eq. (8).

From the derivation, it is clear that $g(\omega)$ for the perfect lattice can be written variously in the equivalent forms

$$\begin{aligned} g(\omega) &= 2\pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega, \\ &= 0, && \omega > \Omega, \quad (19) \end{aligned}$$

$$\begin{aligned} g(\omega) &= 2\pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega', \\ &= 2N^{-1}\delta(\omega - \Omega), && \omega > \Omega', \quad (29) \end{aligned}$$

and

$$\begin{aligned} g(\omega) &= \pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega, \\ &= +\pi^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}}, && \omega \leq \Omega', \\ &= +N^{-1}\delta(\omega - \Omega), && \omega > \Omega', \quad (30) \end{aligned}$$

where Ω' is the second largest allowed frequency. It is also clear that, in each case, $\int_0^\infty g(\omega)\, d\omega = 1$. Making use of these results, we are in a position to analyze the frequency spectrum for the lattice with an isotope defect.

For the case $-\infty \leq \epsilon \leq 0$ we see that, as N tends to infinity, Eq. (27) tends to Eq. (19); the correction tends to zero as N^{-1}. As $\epsilon \to 0$ (perfect lattice), Eq. (27) also tends to Eq. (19); the correction tends to zero as ϵ. For the case $0 \leq \epsilon \leq 1$, as N tends to infinity, Eq. (28) tends to Eq. (19); the corrections go to zero as N^{-1}. As $\epsilon \to 0$, Eq. (28) tends to Eq. (30); the "spike" corresponding to the defect frequency rejoins the passing band of frequencies in its role as maximum allowed frequency and the other correction term tends to zero as ϵ.

Thus, the spectrum $g(\omega)$ is a continuously varying function of the parameter ϵ in the sense that Eq. (27) changes in a smooth fashion, as ϵ is varied, into Eq. (28). As the value of $\epsilon = 0$ is approached through the negative range of ϵ, a "spike" appears because the defect frequency separates from the rest of the frequencies and appears as a localized mode frequency above the passing band of the perfect lattice. As ϵ is further increased, the "spike" separates from the passing band more and more and, as $\epsilon \to 1$, the "spike" moves to infinity, indicating the loss of a mode of motion by the system.

III. PREVIOUS RESULTS

In order to compare the results of this calculation with those of other authors, the correction $\Delta g(\omega)$ to the spectrum must be calculated. We can ignore the contribution to the spectrum of the antisymmetric frequencies since they are unaffected by the defect and thus make no contribution to $\Delta g(\omega)$. For the symmetric frequencies in the perfect lattice, Eq. (12) can be written either as

$$g_0(\omega) = 2N^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \sum_k \delta(\xi - \xi(k)) \quad (31)$$

or

$$g_0(\omega) = 2N^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \times \sum_{k'} \delta(\xi - \xi(k)) + N^{-1}\delta(\omega - \Omega), \quad (32)$$

where the sum now contains one less term. For the lattice with a defect, we have, for $-\infty \leq \epsilon \leq 0$,

$$\begin{aligned} g_1(\omega) &= 2N^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \\ &\quad \times \{1 - \epsilon\Omega^2 N^{-1}[\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1}\} \\ &\quad \times \sum_k \delta(\xi - \xi(k)) \quad (33) \end{aligned}$$

and, for $0 \leq \epsilon \leq 1$,

$$\begin{aligned} g_1(\omega) &= 2N^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \\ &\quad \times \{1 - \epsilon\Omega^2 N^{-1}[\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1}\} \\ &\quad \times \sum_{k'} \delta(\xi - \xi(k)) + N^{-1}(\omega - \Omega(1 - \epsilon^2)^{-\frac{1}{2}}). \\ & \quad (34) \end{aligned}$$

In calculating $\Delta g(\omega) = g_1(\omega) - g_0(\omega)$, we must be careful to pair the "spikes" which correspond to each

frequency in the perfect lattice and its counterpart in the defect lattice. Subtracting Eq. (31) from Eq. (33), we obtain, for $-\infty \leq \epsilon \leq 0$,

$$\Delta g(\omega) = -\epsilon\Omega^2 N^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \times [\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1} 2N^{-1} \sum_k \delta(\xi - \xi(k)) \tag{35}$$

and, for $0 \leq \epsilon \leq 1$, subtracting Eq. (32) from Eq. (34), we have

$$\Delta g(\omega) = -\epsilon\Omega^2 N^{-1}(\Omega^2 - \omega^2)^{-\frac{1}{2}} \times [\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1} 2N^{-1} \sum_{k'} \delta(\xi - \xi(k)) + N^{-1}\delta(\omega - \Omega(1 - \epsilon^2)^{-\frac{1}{2}}) - N^{-1}\delta(\omega - \Omega). \tag{36}$$

As we let N become large, the sums become integrals as before and we find, for $-\infty \leq \epsilon \leq 0$,

$$\Delta g(\omega) = -(\epsilon\Omega^2/N\pi)(\Omega^2 - \omega^2)^{-\frac{1}{2}}[\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1}, \quad \omega \leq \Omega',$$
$$= 0, \quad \omega > \Omega', \tag{37}$$

and, for $0 \leq \epsilon \leq 1$,

$$\Delta g(\omega) = -(\epsilon\Omega^2/N\pi)(\Omega^2 - \omega^2)^{-\frac{1}{2}}[\Omega^2 + (\epsilon^2 - 1)\omega^2]^{-1}, \quad \omega \leq \Omega',$$
$$= +N^{-1}\delta(\omega - \Omega(1 - \epsilon^2)^{-\frac{1}{2}}) - N^{-1}\delta(\omega - \Omega), \quad \omega > \Omega', \tag{38}$$

where Ω' denotes the largest real solution of Eq. (8).

It is a natural temptation to subtract Eq. (19) from Eq. (27) and Eq. (30) from Eq. (28), and thus easily obtain an expression for $\Delta g(\omega)$. Such a procedure would lead to incorrect results, because the actual subtraction must be done with the spectra expressed in their true representation as sums of δ functions.

Our results are substantially in agreement with those of other authors, although with important differences in that previous results contain an additional term $-(2N)^{-1}\delta(\omega - \Omega)$ in Eq. (37) and show the last term in Eq. (38) as $-(2N)^{-1}\delta(\omega - \Omega)$. The initial result, by Mahanty, Maradudin, and Weiss,[1] is in error due to an improperly performed contour integration around a branch point leading to their Eq. (A11). We have not endeavored to determine where Litzman[2] went astray. It should be noted, however, that his result was derived after discarding significant terms in a sum which led him to the large-N limit of our Eq. (8). He was further restricted to values of ω for which $\omega > \Omega$ and hence the subsequent derivation is only pertinent to terms involving $\omega(d)$. Since he was therefore restricted to $N^{-1} = 0$ and $\omega > \Omega$, it is somewhat startling that he obtained any result at all.

The third derivation, by Maradudin,[3] is actually correct except for the last step in which a discontinuous function is (improperly) differentiated on an interval containing the point of discontinuity. In fact, in a general discussion earlier in his paper, Maradudin has essentially stated the correct result. Another expression, presented by Takeno,[7] does not contain the term $-N^{-1}\delta(\omega - \Omega)$ for the case $0 \leq \epsilon \leq 1$ and is in error as to the valid range of ω for the term proportional to ϵ. Ludwig[8] has criticized the result of Takeno on the basis of an argument concerning the normalization of the various terms in the frequency spectrum; however, the argument is clearly incorrect. To see this, one need only recall that the term proportional to ϵ is bounded and consider the behavior of the spectrum as $\epsilon \to 0$.

[7] S. Takeno, Progr. Theoret. Phys. (Kyoto) Suppl. **23**, 94 (1962).
[8] W. Ludwig, *Theory of Crystal Defects*, B. Grubner, Ed. (Academic Press, New York, 1966), p. 57.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Exact Solution of the Faddeev Equations for a One-Dimensional System

L. R. DODD
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia

(Received 9 May 1969)

The Faddeev equations for the one-dimensional system consisting of three identical particles interacting through δ-function potentials are shown to be exactly solvable. The scattering solutions in the form of half-off-shell rearrangement and elastic amplitudes are constructed explicitly and the ground state of the system is derived.

I. INTRODUCTION

Faddeev[1] has provided a rigorous, general formulation of the three-body problem in nonrelativistic quantum mechanics. In the Faddeev approach three-particle scattering amplitudes are determined from the off-shell two-particle transition operators by coupled integral equations. In recent years many numerical calculations[2] based on various simplifications of the Faddeev equations have been performed for systems of physical interest. The results of these calculations, which have given the broad features of experiment, have been encouraging. On the other hand, an interesting question, which has not received much attention,[3] is whether the Faddeev equations admit exact solutions for particular, special systems. As in the case of two-particle systems and the Schrödinger equation, such exact solutions should elucidate the structure of the Faddeev equations as well as providing useful tests of approximation procedures.

The problem of a one-dimensional system of three identical (but distinguishable) particles interacting through δ-function potentials is exactly solvable. The bound and scattering states for this system have been found by McGuire[4] and by Yang[5] by special methods. McGuire's method of solution is based on the reduction of the three-body problem to an equivalent problem in geometrical optics. Yang's algebraic method rests on the assumption that an exact solution may be represented as a linear combination of a finite number of planewaves (Bethe's hypothesis).

In this paper we show that the Faddeev equations are exactly soluble for the system considered by McGuire and Yang. An outline of our approach is as follows. In Sec. II McGuire's solution for the scattering wavefunction is recalled. Using the separable property of the δ-function interaction and following Lovelace,[6] we reduce the Faddeev equations to a pair of coupled one-dimensional integral equations in Sec. III. In Sec. IV off-shell, three-particle elastic and rearrangement amplitudes are constructed from McGuire's solution of Sec. II. These amplitudes are then shown to satisfy the Faddeev equations of Sec. III. The proof depends on the cancellation of the two-body branch cut in the kernel of the integral equations by a compensating factor in the exact amplitudes. This property of the exact amplitudes (which is probably peculiar to this problem) is apparently the counterpart of Bethe's hypothesis used by Yang and McGuire in their derivations. Some remarks on the high-energy limit of the amplitudes and the validity of the Born approximation are also included. The bound state of the three-body system is found in Sec. V by considering the homogeneous form of the Faddeev equations.

For completeness the probabilities for rearrangement and elastic scattering are derived from the exact amplitudes in Appendix A and shown to agree with the results of Yang and McGuire. In Appendix B, the bound state found in Sec. V is identified as the ground state of the three-body system.

II. McGUIRE'S SOLUTION

The three-body system of interest consists of three identical (but distinguishable) particles moving on a line and interacting through equal-strength, attractive

[1] L. D. Faddeev, Zh. Eksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys.—JETP **12**, 1014 (1961)]; *Mathematical Aspects of the Three-Body Problem* (Daniel Davey and Co., Inc., New York, 1965).

[2] For a review of calculations in nuclear and particle physics using separable potentials, see K. M. Watson and J. Nuttall, *Topics in Several Particle Dynamics* (Holden-Day, Inc., San Francisco, 1967); for an application to atomic physics using Sturmian expansions, see J. Ball, T. C. Y. Chen, and D. Y. Wong, Phys. Rev. **173**, 202 (1968).

[3] The Faddeev equations in three dimensions for the case of potentials of vanishing range have been considered by several authors. In the zero-range limit the integral equations differ only slightly from those of the one-dimensional problem discussed here. However, there is the additional complication that the kernels are singular. See, for example, *Application of Mathematics to Problems in Theoretical Physics*, F. Lurçat, Ed. (Gordon and Breach, Science Publishers, Inc., New York, 1967).

[4] J. B. McGuire, J. Math. Phys. **5**, 622 (1964).

[5] C. N. Yang, Phys. Rev. Letters **19**, 1312 (1967); Phys. Rev. **168**, 1920 (1968).

[6] C. Lovelace, Phys. Rev. **135**, B1225 (1964).

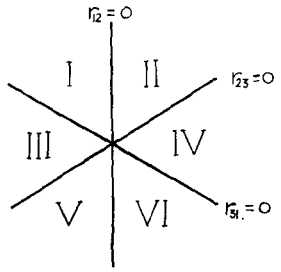


FIG. 1. Regions of the three-particle configuration space.

δ-function potentials. We shall use the usual Jacobi coordinates, which are defined in terms of the momenta k_i and the positions x_i of the three particles, by relations

$$p_k = \tfrac{1}{3}(k_j + k_i - 2k_k), \quad k_{ij} = \tfrac{1}{2}(k_i - k_j),$$
$$q_k = \tfrac{1}{2}(x_j + x_i - 2x_k), \quad r_{ij} = x_i - x_j$$

(i, j, k being any cyclic permutation of 1, 2, 3), and we take the center of mass at rest:

$$k_1 + k_2 + k_3 = 0.$$

The Hamiltonian is

$$H = H_0 + V_1 + V_2 + V_3,$$

where the kinetic energy is

$$H_0 = \tfrac{3}{4}p_i^2 + k_{jk}^2,$$

and the two-body interaction V_i is

$$V_i = -g\delta(r_{jk})$$

with g the strength of the Dirac δ function. For convenience we have chosen units such that the masses of the particles $m_j = \frac{1}{2}$, and Planck's constant $\hbar = 1$.

In an interesting paper,[4] McGuire has found the solutions of Schrödinger's equation for this system by elementary methods. The solutions are most easily exhibited in the configuration space of the three particles (Fig. 1). If the center-of-mass coordinate is ignored, the configuration space is a plane which is divided symmetrically into six segments by the lines $r_{12} = 0$, $r_{23} = 0$, and $r_{31} = 0$, which constitute all those points where the particles interact.

From ray-tracing arguments McGuire shows that a solution Ψ of $H\Psi = E\Psi$ is a linear combination of six possible planewave types, corresponding to the six possible permutations of the initial momenta of the three particles

$$\Psi = \sum_{\text{perm}\,(1,2,3)} C_{ijk} \exp\,(ik_1x_i + ik_2x_j + ik_3x_k). \quad (1)$$

Classically speaking, Eq. (1) assumes that the three particles may only exchange momenta during the scattering process; no new momenta are generated by collisions. The coefficients C_{ijk} are chosen in each of the six segments of configuration space so that the total wavefunction Ψ is continuous everywhere and so that its gradient has the correct discontinuity across the lines of interaction $r_{12} = 0$, $r_{23} = 0$, and $r_{31} = 0$. The complete specification of a particular solution therefore reduces to the enumeration of the thirty-six coefficients which determine the amplitudes of the six different wave types in each of the six regions of configuration space.

To construct the amplitudes which appear in the Faddeev equations of Sec. III, we need the scattering solution of Schrödinger's equation which describes one of the particles incident on a bound state of the other two. For definiteness we label the free particle in the initial state as particle 3. For the attractive δ interaction there is just one bound state of the two-particle system with binding energy $-\frac{1}{4}g^2$ and wavefunction $\phi_3(r_{12}) = (\frac{1}{2}g)^{\frac{1}{2}}e^{-\frac{1}{2}g|r_{12}|}$. Then, if particle 3 has momentum p_3 initially with respect to the center of mass of particles 1 and 2, the wavefunction describing the three-body system before collision is

$$\psi_{p_3}(r_{12}, q_3) = (g/4\pi)^{\frac{1}{2}}e^{iq_3p_3}e^{-\frac{1}{2}g|r_{12}|}. \quad (2)$$

The scattering solution $\Psi_{p_3}(r_{12}, q_3)$ of Schrödinger's equation (see Table I) is

$$H\Psi_{p_3} = E\Psi_{p_3},$$

with energy

$$E = \tfrac{3}{4}p_3^2 - \tfrac{1}{4}g^2,$$

which satisfies the boundary conditions

$$\Psi_{p_3} \sim \psi_{p_3} + \text{outgoing waves},$$

for large distances from the origin in configuration space as given by McGuire. In the first column the exponents of the six possible wave types are listed. Each row then gives the coefficients of a particular wave type in the various regions of configuration space.

TABLE I. Wavefunction for rearrangement scattering.

	I	II	III	IV	V	VI
$-\frac{1}{2}gr_{12} + iq_3p_3$	—	1	—	T_2	—	T_1T_2
$-\frac{1}{2}gr_{31} + iq_2p_2$	R_2	R_1T_2	—	—	—	—
$-\frac{1}{2}gr_{23} + iq_1p_1$	—	—	T_2R_1	—	—	—
$\frac{1}{2}gr_{12} + iq_3p_3$	1	—	—	—	T_1T_2	—
$\frac{1}{2}gr_{31} + iq_2p_2$	—	—	—	R_1T_2	—	—
$\frac{1}{2}gr_{23} + iq_1p_1$	T_2R_1	R_2	—	—	—	—

The reflection coefficients R_i and the transmission coefficients T_i of the table are defined by

$$T_i = S_i/(1 + S_i), \quad R_i = T_i - 1 = -1/(1 + S_i), \tag{3}$$

where the parameters S_i are given in terms of the relative momenta in the intial state by

$$S_1 = 2ik_{23}/g \quad \text{and} \quad S_2 = 2ik_{13}/g.$$

Since in the initial state $k_{23} - k_{13} = -k_{12} = -\frac{1}{2}ig$ and $k_{23} + k_{13} = 3p_3$, we have the relations

$$S_1 - S_2 = 1$$

and

$$S_1 + S_2 = 6ip_3/g, \tag{4}$$

so that in fact all the amplitudes of the table may be expressed in terms of a single parameter S_1, which is fixed by the incident momentum p_3 and the strength g of the two-particle interaction.

The probabilities for elastic and rearrangement scattering are easily deduced from the table; the reader is referred to McGuire's paper and Appendix A of this paper for further details.

III. FADDEEV-LOVELACE EQUATIONS

With the usual definition of the three-particle transition operators for elastic and rearrangement scattering,[6]

$$U_{ij}^{+}(E) = v_i + v_i G^{+}(E) v_j, \tag{5}$$

where

$$v_i = V_j + V_k \quad \text{and} \quad G^{+}(E) = (E + i\epsilon - H)^{-1}$$

is the full resolvent operator, the Faddeev equations may be written in the form[7]

$$\begin{aligned} U_{13}^{+} &= V_3 + t_2 G_0^{+} U_{23}^{+} + t_3 G_0^{+} U_{33}^{+}, \\ U_{23}^{+} &= V_3 + t_1 G_0^{+} U_{13}^{+} + t_3 G_0^{+} U_{33}^{+}, \\ U_{33}^{+} &= t_1 G_0^{+} U_{13}^{+} + t_2 G_0^{+} U_{23}^{+}. \end{aligned} \tag{6}$$

The free Green's operator for this problem is

$$G_0^{+}(E) = (E + i\epsilon - \tfrac{3}{4}p_i^2 - k_{jk}^2)^{-1}$$

and the two-particle transition operators (acting in the three-particle space) are defined by

$$t_i(E) = V_i + V_i G_0^{+}(E) t_i(E).$$

Now, since

$$\begin{aligned} \langle r_{ij}|\, V_k \,|r'_{ij}\rangle &= -g\delta(r_{ij})\delta(r'_{ij} - r_{ij}) \\ &= -g\delta(r_{ij})\delta(r'_{ij}), \end{aligned}$$

the two-particle δ-function interaction is separable and may be expressed as

$$V_i = -(2/g^2)\,|\chi_i\rangle\langle\chi_i|, \tag{7}$$

with the form factor

$$|\chi_i\rangle = (g^3/4\pi)^{\frac{1}{2}} \int dk_{jk}\, |k_{jk}\rangle. \tag{8}$$

The two-particle transition operator (acting in the three-particle space) then take the familiar form[6]

$$t_i(E) = \int_{-\infty}^{\infty} |\chi_i p_i\rangle \tau(E + i\epsilon - \tfrac{3}{4}p_i^2) \langle p_i \chi_i|\, dp_i. \tag{9}$$

In this case the propagator

$$\tau(E) = -\frac{2}{g^2} \frac{2\sqrt{E}}{2\sqrt{E} - ig} \tag{10}$$

has a pole at $\sqrt{E} = \frac{1}{2}ig$ associated with the single bound state of the attractive δ potential, which has binding energy $E_B = -\frac{1}{4}g^2$. Furthermore, with the appropriate normalization of the form factor, the bound state ϕ_i of particles j and k is given in general by the relation

$$G_0^{+}(E)\,|\chi_i p_i\rangle = |\phi_i p_i\rangle, \tag{11}$$

with

$$E = -\tfrac{1}{4}g^2 + \tfrac{3}{4}p_i^2,$$

which is easily verified in this particular problem, since

$$|\phi_i\rangle = \int_{-\infty}^{\infty} (-\tfrac{1}{4}g^2 - k_{jk}^2)^{-1} (g^3/4\pi)^{\frac{1}{2}}\, |k_{jk}\rangle\, dk_{jk}$$

or

$$\langle r_{jk} |\, \phi_i\rangle = -(\tfrac{1}{2}g)^{\frac{1}{2}} e^{-\frac{1}{2}g|r_{12}|}.$$

If we define the half-on-shell amplitudes

$$X_{ij}(p_i, p_j) = \langle \chi_i p_i|\, G_0^{+}(E) U_{ij}^{+}(E) G_0^{+}(E)\, |\chi_j p_j\rangle, \tag{12}$$

with

$$E = -\tfrac{1}{4}g^2 + \tfrac{3}{4}p_j^2,$$

and the "potentials"

$$\langle \chi_i p_i|\, G_0^{+}(E)\, |\chi_i p_j\rangle = Z(p_i, p_j; E), \tag{13}$$

$$\langle \chi_i p_i|\, G_0^{+}(E) V_3 G_0^{+}(E)\, |\chi_i p_j\rangle = Y(p_i p_j), \tag{14}$$

with

$$E = -\tfrac{1}{4}g^2 + \tfrac{3}{4}p_j^2,$$

substitute the separable form Eq. (9) for t_i in Eqs. (6), and use the identity of the particles, the Faddeev equations (6) are reduced to the following form:

$$X_{33}(p'_3, p_3) = 2\int_{-\infty}^{\infty} Z(p'_3, p; E)\tau(E - \tfrac{3}{4}p^2) X_{23}(p, p_3)\, dp, \tag{15a}$$

$$X_{23}(p'_2, p_3) = Y_{23}(p'_2, p_3) + \int_{-\infty}^{\infty} dp Z(p'_2, p; E)\tau(E - \tfrac{3}{4}p^2) \times [X_{23}(p, p_3) + X_{33}(p, p_3)]. \tag{15b}$$

[7] The inhomogeneous terms of our Eqs. (6) are different from Lovelace's Eqs. (3.9). But Eqs. (6) of this paper are easily obtained from Eqs. (3.9) of Lovelace's by expressing $U_{\alpha\beta}^{-}$ in terms of $U_{\alpha\beta}^{+}$ in the second of Lovelace's Eqs. (3.9).

The kernel and the inhomogeneous term are obtained explicitly from Eq. (10) for the propagator and the definitions (13) and (14):

$$Z(p_i, p_j; E) = \frac{g^3}{4\pi} \frac{1}{E + i\epsilon - p_i^2 - p_j^2 - p_i p_j} \quad (16)$$

and

$$Y_{23}(p_2', p_3) = \frac{g^3}{\pi} \frac{ig + 2t}{t[(ig + 2t)^2 + (p_2' - p_3)^2]}, \quad (17)$$

with

$$t = (E - \tfrac{3}{4}p_2'^2 + i\epsilon)^{\frac{1}{2}}.$$

From the relation (11) and the definition (12), we note that the amplitudes X_{23} and X_{33}, when taken fully on shell, are equal to the usual on-shell rearrangement and elastic scattering amplitudes

$$X_{23}(p_2', p_3) = \langle \phi_2 p_2' | U_{23}^+ | \phi_3 p_3 \rangle, \quad \text{if} \quad |p_2'| = |p_3|,$$
and (18)
$$X_{33}(p_3', p_3) = \langle \phi_3 p_3' | U_{33}^+ | \phi_3 p_3 \rangle, \quad \text{if} \quad |p_3'| = |p_3|.$$

IV. EXACT SOLUTION OF THE FADDEEV-LOVELACE EQUATIONS

In this section the unknown amplitudes X_{23} and X_{33} are constructed from the wavefunction of Table I and are shown to satisfy the Faddeev–Lovelace equations (15).

First, we see that the half-off-shell amplitudes X_{23} and X_{33} may be expressed in terms of the scattering state $|p_3 +\rangle$, which is the scattering solution of the complete Schrödinger equation describing particle 3 incident on the bound state of particles 1 and 2. Since the integral form of Schrödinger's equation is

$$|p_3 +\rangle = (1 + G^+(E)v_3)\,|\phi_3 p_3\rangle,$$

provided that $E = -\frac{1}{4}g^2 + \frac{3}{4}p_3^2$, we have, using Eqs. (5), (11), and (12),

$$X_{33}(p_3', p_3) = \langle \chi_3 p_3' | \, G_0^+(V_1 + V_3)\,|p_3 +\rangle,$$
$$X_{23}(p_2', p_3) = \langle \chi_2 p_2' | \, G_0^+(V_1 + V_2)\,|p_3 +\rangle. \quad (19)$$

Now $\langle r_{12} q_3 \,|\, p_3 +\rangle = \Psi_{p_3}(r_{12}, q_3)$ of Table I and it is straightforward to form X_{23} and X_{33} from Eqs. (19). For example,

$$\begin{aligned} X_{33}(p, p_3) &= (g^3/4\pi)^{\frac{1}{2}} \int_{-\infty}^{\infty} \frac{dk_{12}}{E - k_{12}^2 - \frac{3}{4}p_3^2 + i\epsilon} \\ &\quad \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dr_{12}\, dq_3 e^{-ik_{12}r_{12} - ip_3 q_3} \\ &\quad \times [-g\delta(q_3 - \tfrac{1}{2}r_{12}) - g\delta(r_{12})]\Psi_{p_3}(r_{12}, q_3) \\ &= \frac{ig^2}{\pi}\left(\frac{ig}{2t}\right)\left(\frac{T_2}{2t + p + ig - p_3}\right. \\ &\quad \left. + \frac{T_2 T_1}{2t - p + ig + p_3} + \frac{R_1 T_2}{2t + p + 2p_3}\right), \end{aligned} \quad (20a)$$

with

$$t = (\tfrac{3}{4}(p_3^2 - p^2) - \tfrac{1}{4}g^2)^{\frac{1}{2}}.$$

Similarly,

$$\begin{aligned} X_{23}(p, p_3) &= \frac{ig^2}{2\pi}\left(\frac{ig}{2t}\right)\left(\frac{T_2 R_1 + T_1 T_2}{2t + p + 2p_3} + \frac{T_2 T_1 + T_2 R_1 + R_2}{2t + ig - p + p_3}\right. \\ &\quad \left. + \frac{1}{2t - p - 2p_3} + \frac{T_2}{2t + ig - p_3 + p}\right). \end{aligned} \quad (20b)$$

We remark parenthetically that these amplitudes when taken on shell according to Eqs. (18) yield cross sections which agree with the results of McGuire and Yang. This is demonstrated in Appendix A with the aid of the asymptotic form of the Lippmann–Schwinger equation.

In order to show that the amplitudes (20) provide the exact solution of the Faddeev equations, we are faced with the difficulty of performing the integrations in Eqs. (15); the branch cut of the propagator τ is apparently troublesome. The essential step here is to realize that the exact amplitudes (20) contain an implicit factor which exactly cancels the propagator in the kernel of the integral equations. With some algebra the amplitudes (20) may be rewritten as

$$X_{33}(p, p_3) = \frac{g}{2\pi} \frac{\tau^{-1}(t^2)F(p)}{(p - p_1)(p - p_2)(p - p_3)} \quad (21a)$$

and

$$X_{23}(p, p_3) = \frac{g}{4\pi} \frac{\tau^{-1}(t^2)G(p)}{(p - p_1)(p - p_2)(p - p_3)}, \quad (21b)$$

where

$$F(p) = (p - p_3)R_1 T_2 + (p - p_1)T_2 T_1 + (p - p_2)T_2,$$
$$G(p) = (p - p_3)(R_1 T_2 + T_1 T_2 + 1) + (p - p_2)T_2 + (p - p_1)(T_2 T_1 + T_2 R_1 + R_2),$$

and

$$p_2 = -\tfrac{1}{2}(p_3 - ig) \quad \text{and} \quad p_1 = -\tfrac{1}{2}(p_3 + ig).$$

It is now straightforward to verify that the amplitudes (21) satisfy the integral equations. We outline the proof for the first of the coupled equations, Eq. (15a). After substitution of Eq. (21b), the right-hand side of Eq. (15a) becomes

$$(g^4/32\pi^2)\int_{-\infty}^{\infty} G(p)[(p - p_0)(p - \bar{p}_0)(p - p_1) \times (p - p_2)(p - p_3)]^{-1}\, dp.$$

The poles at $p_0 = -\frac{1}{2}p_3 + t$ and $\bar{p}_0 = -\frac{1}{2}p_3 - t$ arise from the kernel (16). The integration is performed by completing the contour in the upper half of the complex p plane. There are three contributions to the integral from the three simple poles at p_0, p_3,

and p_2. After some lengthy algebra, using the relations (3) and (4) to express all terms containing p_3 in terms of S_1, and rearranging terms, the result of the integration may be shown to equal X_{33} as required.

The interesting question of the validity of the Born approximation for rearrangement collisions[8] may be considered in the context of our exactly soluble model. From either Eq. (20b) or Eq. (21b), we have

$$X_{23}(p_2p_3)|_{p_2=p_3} = -\frac{g}{4\pi}\frac{(2S_1-1)(S_1-1)}{S_1(S_1+1)} \sim -\frac{g}{2\pi}$$

at high energies (p_3 large). The Born approximation for rearrangement is obtained from the first term of Eq. (5):

$$\langle\phi_2 p_2|\,(V_3+V_1)\,|\phi_3 p_3\rangle|_{p_2=p_3} = -\frac{g}{2\pi}+\frac{g}{4\pi(S_1-1)S_1} \sim -\frac{g}{2\pi}$$

at high energies. So for this simple system the Born approximation yields the correct high-energy behavior of the rearrangement amplitude. Note, however, that the lowest-order approximation from the Faddeev equations (6) is

$$\langle\phi_2 p_2|\,V_3\,|\phi_3 p_3\rangle = \frac{g}{4\pi(S_1-1)S_1} \sim 0$$

at high energies, which is incorrect. Nevertheless, when the Faddeev equations are iterated once, the resulting approximate solution is

$$\langle\phi_2 p_2|\,(V_3+t_1G_0V_3)\,|\phi_3 p_3\rangle,$$

which does give the correct high-energy behavior.

V. THREE-BODY BOUND STATE

Each of the exact three-particle amplitudes of Eqs. (20) and (21) has a pole at $p_3 = ig$ associated with a three-body bound state of binding energy $E_B = -\frac{1}{4}g^2 - \frac{3}{4}p_3^2 = -g^2$.

The existence of this bound state may be confirmed by considering the homogeneous form of the Faddeev equations (6). Equations (15) are then replaced by a single homogeneous equation:

$$X(p) = 2\int_{-\infty}^{\infty} dp' Z(p,p',E)\tau(E-\tfrac{3}{4}p'^2)X(p'), \quad (22)$$

with

$$E = \tfrac{3}{4}p_3^2 - \tfrac{1}{4}g^2.$$

The amplitude $X(p)$ is related to the three-body bound state Φ by [c.f. Eqs. (19)]

$$X(p) = \langle p\chi_3|\,G_0(V_1+V_2)\,|\Phi\rangle. \quad (23)$$

Following our discussion of the scattering equations, we assume that X contains a factor which exactly cancels the propagator:

$$X(p) = \tau^{-1}(E-\tfrac{3}{4}p^2)W(p). \quad (24)$$

Equation (22) then becomes

$$W(p) = -\frac{g}{\pi}\frac{2(E-\frac{3}{4}p^2)^{\frac{1}{2}}}{2(E-\frac{3}{4}p^2)^{\frac{1}{2}}-ig} \times\int_{-\infty}^{\infty}\frac{W(p')\,dp'}{E-pp'-p^2-p'^2}. \quad (25)$$

This equation has been previously formulated by Eyges,[9] whose approach to the three-body bound state problem is similar to Faddeev's.

The exact solution $W_0(p) = (p^2+g^2)^{-1}$ for energy $E = -g^2$ has been found by Jasperse,[10] and may easily be checked by contour integration. The consistency of this solution with the bound-state wavefunction of McGuire[4] and Yang,[5]

$$\langle r_{12}, r_{23}, r_{31}\,|\,\Phi\rangle \alpha \exp\left[-\tfrac{1}{2}g(|r_{12}|+|r_{23}|+|r_{31}|)\right],$$

is also easily checked with the help of Eqs. (23) and (24).

Equation (25) has a polar kernel and therefore may be transformed to an equivalent equation with a symmetric kernel. In Appendix B, it is shown that the transformed kernel is real and square-integrable for three-body energies less than the two-particle binding energy, $E < -\frac{1}{4}g^2$. The general theory of Weinberg[11] is also applied to show that the above solution W_0 for binding energy $E = -g^2$ corresponds to the ground state of the system.

ACKNOWLEDGMENT

The author wishes to thank H. O. Girotti for help with calculations in the early stages of this work.

APPENDIX A

In this appendix the probabilities for elastic and rearrangement scattering are found from the formal theory. The Lippmann–Schwinger equation for the scattering state is

$$|p_3+\rangle = |\phi_3 p_3\rangle + (E-H_0-V_3+i\epsilon)^{-1}V_3\,|p_3+\rangle.$$

The elastically scattered component of the wavefunction is then

$$\langle\phi_3 q_3\,|\,p_3+\rangle = \langle q_3\,|\,p_3\rangle + \int_{-\infty}^{\infty} dq_3'\langle q_3|\,(E-\tfrac{3}{4}p_3^2+i\epsilon)^{-1}\,|q_3'\rangle \times\langle q_3'\phi_3|\,v_3\,|p_3+\rangle.$$

[8] R. Aaron, R. D. Amado, and B. W. Lee, Phys. Rev. **121**, 319 (1961); K. R. Greider and L. R. Dodd, *ibid.*, **146**, 671 (1966); K. Dettmann and G. Leibfried, *ibid.*, **148**, 1271 (1966).

[9] L. Eyges, J. Math. Phys. **6**, 1320 (1965).

[10] J. R. Jasperse, Phys. Rev. **159**, 69 (1967).

[11] S. Weinberg, Phys. Rev. **131**, 440 (1963).

But,

$$\langle q_3|\,(E - \tfrac{3}{4}p_3^2 + i\tau)^{-1}\,|q_3'\rangle = -(2i/3p_3)e^{ip_3|q_3-q_3'|} = -(4\pi i/3p_3)\langle q_3\,|\,p_3\rangle\langle p_3\,|\,q_3'\rangle,$$

for $q_3 > q_3'$. Thus, for q_3 large and positive,

$$\langle \phi_3 q_3\,|\,p_3\,+\rangle = \langle q_3\,|\,p_3\rangle[1 - (4\pi i/3p_3)\langle p_3\phi_3|\,v_3\,|p_3\,+\rangle].$$

From Eq. (19) and Eq. (11) the matrix element in the square brackets is equal to $X_{33}(p_3, p_3)$, which is found explicitly by putting $p = p_3$ in Eq. (20a) or Eq. (21a). Thus the probability for transmission of particle 3 through the bound state of 1 and 2 is

$$\left|1 - \frac{4\pi i}{3p_3}X_{33}(p_3, p_3)\right|^2 = \left|\frac{S_1 - 1}{S_1 + 1}\right|^2 = |T_2T_1|^2.$$

By similar manipulations of the Lippmann–Schwinger equation the following results are obtained. The probability for reflection of particle 3 from the bound state is

$$|1 - (4\pi i/3p_3)X_{33}(-p_3, p_3)|^2 = 0.$$

The probability for rearrangement where particle 3 replaces particle 2 in the bound state, particle 2 emerging with momentum equal to the incident momentum of particle 3, is

$$\left|-\frac{4\pi i}{3p_3}X_{23}(p_3, p_3)\right|^2 = \left|-\frac{S_1 - 1}{S_1(S_1 + 1)}\right|^2 = |T_2R_1|^2.$$

The probability for rearrangement where particle 2 emerges with momentum equal in magnitude but in the opposite sense to the incident momentum of particle 3, is

$$|-(4\pi i/3p_3)X_{23}(-p_3, p_3)|^2 = 0.$$

The rearrangement probabilities where particle 1 is free in the final state are, of course, equal to those given above. The above probabilities, obtained from the formal theory, may be checked directly using the wavefunction of Table I. For example, consider the common border of regions II and IV at large distances from the origin, which corresponds to a configuration where particles 1 and 3 are together and particle 2 is distant from them. The wavefunction is $T_2e^{-\frac{1}{2}ar_{12}+iq_3p_3} + R_1T_2e^{iq_2p_2}$.

The first term is very small along the line $r_{13} = 0$ at large distances from the origin so that the probability for rearrangement is $|R_1T_2|^2$ as required.

APPENDIX B

If the kernel of Eq. (25) is symmetrized[12] and the substitutions $x = p(-E)^{-\frac{1}{2}}$, $y = p'(-E)^{-\frac{1}{2}}$, and $\mu = 2(-E)^{\frac{1}{2}}/g$ made, Eq. (22) takes the form

$$\tilde{W}(x) = \int_{-\infty}^{\infty} K(x, y; \mu)\tilde{W}(y)\,dy, \tag{B1}$$

where

$$K(x, y; \mu) = \eta^{\frac{1}{2}}(x; \mu)k(x, y)\eta^{\frac{1}{2}}(y; \mu)$$

and

$$\tilde{W}(x) = \eta^{-\frac{1}{2}}(x; \mu)W(x),$$

with

$$\eta(x; \mu) = (1 + \tfrac{3}{4}x^2)^{\frac{1}{2}}/[\mu(1 + \tfrac{3}{4}x^2)^{\frac{1}{2}} - 1]$$

and

$$k(x, y) = (2/\pi)(x^2 + y^2 + xy + 1)^{-1}.$$

We seek solutions of the equation for values of $\mu > 1$ which correspond to three-body energies $E < -\frac{1}{4}g^2$, which is the two-particle binding energy.

The dimensionless form of Eq. (B1) shows that the number of three-body bound states is independent of the strength g of the two-particle potentials. Also, with the aid of the inequality

$$\eta(x; \mu) \leq (\mu - 1)^{-1} \tag{B2}$$

valid for all x and $\mu > 1$, it is easy to show that

$$\mathrm{Tr}\,[KK^\dagger] \leqslant (\mu - 1)^{-2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[k(x, y)]^2\,dx\,dy \leqslant 1.49 \times (\mu - 1)^{-2}, \tag{B3}$$

showing that the kernel is compact.

Following Weinberg,[11,12] we consider next the eigenvalue problem for the kernel, for fixed μ:

$$\lambda_n(\mu)K(\mu)\phi_n(\mu) = \phi_n(\mu).$$

The solutions of Eq. (B1) occur for those values of μ for which $\lambda_n(\mu) = 1$. The spectrum of the kernel is discrete, real, and bounded from above.

From the properties of the spectrum and Eq. (B3), we have the sum rule

$$\sum_{n=1}^{\infty}\lambda_n^2(\mu) = \mathrm{Tr}\,[K^2] = \mathrm{Tr}\,[KK^\dagger] \leqslant 1.49 \times (\mu - 1)^{-2}. \tag{B4}$$

Also, if λ_1 is chosen as the largest of the eigenvalues,

$$\lambda_1(\mu) \geqslant (\varphi, K\varphi)/(\varphi, \varphi),$$

for any φ. Choose $\varphi = \eta^{-\frac{1}{2}}W_0$ where $W_0(x) = (1 + x^2)^{-1}$ is the solution of Eq. (25) of Sec. V.

[12] R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill Book Co., Inc., New York, 1966), Chap. 9.

Then, since

$$K\varphi = \eta^{\frac{1}{2}}(\mu)\eta^{-1}(2)W_0, \tag{B5}$$

$$\begin{aligned}\lambda_1(\mu) &\geqslant (W_0, \eta^{-1}(2)W_0)/(W_0, \eta^{-1}(\mu)W_0)\\ &= 1 + (W_0, [\eta^{-1}(2) - \eta^{-1}(\mu)]W_0)/(W_0, \eta^{-1}(\mu)W_0)\\ &= 1 + (2 - \mu)(W_0, W_0)(W_0, \eta^{-1}(\mu)W_0)^{-1}. \end{aligned}\tag{B6}$$

But, from Eq. (B2),

$$(W_0, \eta^{-1}(\mu)W_0) \geqslant (\mu - 1)(W_0, W_0).$$

Thus, for $\mu \geqslant 2$,

$$\lambda_1(\mu) \geqslant (\mu - 1)^{-1}. \tag{B7}$$

Combining the inequalities (B7) and (B4), we have the result that

$$\sum_{n=2}^{\infty} \lambda_n^2(\mu) \geqslant 0.49 \times (\mu - 1)^{-2}, \quad \text{for} \quad \mu \geqslant 2,$$

showing that λ_1 is the only eigenvalue which may be larger than unity for $\mu \geqslant 2$ and, therefore, there is at most one bound state with energy $E \leqslant -g^2$. But from Eq. (B5)

$$K\varphi(2) = \varphi(2) \quad \text{when} \quad \mu = 2 \text{ (i.e., when } E = -g^2),$$

so that $\eta^{-\frac{1}{2}}(x; 2)(1 + x^2)^{-1}$ is the solution of Eq. (B1), which corresponds to the ground state of the system.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Renormalized Perturbation Theory for the Weakly Nonlinear Oscillator

S. Ø. Aks* and R. A. Carhart
Department of Physics, University of Illinois at Chicago Circle, Chicago, Illinois

(Received 2 December 1968)

The problem of finding solutions for the weakly nonlinear quantum oscillator is investigated in the Heisenberg representation in one dimension. The perturbation method developed makes allowance for the nonisochronous nature of nonlinear oscillations and avoids at any level of approximation the secularity—terms increasing without bound as $t \to \infty$—intrinsic to the usual type of iteration scheme. The treatment of the general quantum equation $(d^2/dt^2 + \omega^2)x(t) = \epsilon f(x)$, for the Heisenberg position operator $x(t)$, is first motivated by the classical analog. The iteration equations for the quantum case are derived, and the case $f(x) = x^3$ is studied fully to order ϵ, and partially to order ϵ^2.

1. INTRODUCTION

Considerable interest and effort has been directed toward solving the quantum field theory, whose field operators obeys

$$(\Box + m^2)\phi = \epsilon\phi^3. \tag{1.1}$$

This is the self-coupled scalar-field problem, usually called the "ϕ^3" theory.[1] It is one of the simplest nontrivial field theories (nontrivial means it allows for scattering and production) and as such would seem to offer a logical first theory to try to solve fully.

As part of a program to shed light on the nature of a solution to this theory, this paper presents a further study of the nonlinear harmonic oscillator in one-dimensional quantum mechanics.[2] In particular, a method is presented which generates successive approximations to a solution for the position operator $x(t)$, which must satisfy the equation

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x(t) = \epsilon x^3(t). \tag{1.2}$$

This method differs from the previous method presented by one of the authors in avoiding from the start any secular behavior in solutions to the hierarchy of approximation equations and in giving quite directly the "renormalized" frequency and the modified amplitude of the fundamental harmonic—which are q-numbers as previously argued—as power series in ϵ. It also allows systematic calculation to arbitrary order in ϵ, whereas with the previous method it is not clear how to proceed to higher order.

In Sec. 2 the method, as developed by Bogoliubov and Krylov[3] for classical systems, is applied to the classical equations of the type

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x(t) = \epsilon f(x(t)) \tag{1.3}$$

to demonstrate its salient features in a familiar case.

In Sec. 3 we present a systematic formulation of the approximation scheme leading to quantum solutions of Eq. (1.3). Finally, in Sec. 4, this scheme is applied to the case $f(x) = x^3$ and full results to first order in ϵ are given, along with partial results to order ϵ^2.

2. THE METHOD OF BOGOLIUBOV AND KRYLOV FOR CLASSICAL NONLINEAR OSCILLATORS

Bogoliubov and Krylov[3] have developed a very useful method for developing approximate solutions of (c-number) nonlinear differential equations of the form

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x = \epsilon F\left(x, \frac{dx}{dt}\right),$$

where ϵ is a small parameter. In this section we give a short survey of their work in which those aspects of the method that carry over to the case of nonlinear quantum equations of motion are stressed.

In the case of quantum mechanics, interest is almost exclusively limited to nonlinear terms which depend on x alone; thus we restrict attention (in this section) to the c-number equations

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x = \epsilon f(x). \tag{2.1}$$

The method of Bogoliubov and Krylov reduces to finding an approximation of the form

$$x = \rho\cos\psi + \sum_{n\geq 1}\epsilon^n u_n(\rho, \psi). \tag{2.2}$$

* Research supported in part by the National Science Foundation under grants GP8734 and GP7328.

[1] The ϕ^3 current model was introduced by P. T. Matthews, Phil. Mag. **41**, 185 (1950). A similar problem, the ϕ^2 current model, was studied by C. A. Hurst, Proc. Cambridge Phil. Soc. **48**, 625 (1952); W. Thirring, Helv. Phys. Acta **26**, 33 (1953).

[2] S. Ø. Aks, Fortsch. Physik **15**, 661 (1967).

[3] N. Krylov and N. Bogoliubov, *Introduction to Non-linear Mechanics* (Princeton University Press, Princeton, N.J., 1947), p. 41.

The quantities ψ and ρ are taken to satisfy the equations

$$\frac{d\psi}{dt} = \omega + \sum_{n\geq 1} \epsilon^n b_n(\rho), \tag{2.3}$$

$$\frac{d\rho}{dt} = 0. \tag{2.4}$$

The latter equation takes this simple form as a result of the assumption that the nonlinear driving term is independent of dx/dt. The solutions of these equations are each determined up to an additive constant which is found using the initial conditions. If we give the differential equation the usual interpretation, then the equation describes the motion of an anharmonic oscillator. The additive constants are then related to the initial phase and amplitude of the motion of the fundamental frequency term $\rho \cos \psi$.

The following results are obtained from Eq. (2.2):

$$\frac{dx}{dt} = -\rho \frac{d\psi}{dt} \sin \psi + \sum_{n\geq 1} \epsilon^n \frac{d\psi}{dt} \frac{\partial u_n}{\partial \psi},$$

$$\frac{d^2x}{dt^2} = -\rho \left(\frac{d\psi}{dt}\right)^2 \cos \psi + \sum_{n\geq 1} \epsilon^n \left(\frac{d\psi}{dt}\right)^2 \frac{\partial^2 u_n}{\partial \psi^2},$$

$$\begin{aligned} \frac{d^2x}{dt^2} + \omega^2 x = {} & \rho\left[\omega^2 - \left(\frac{d\psi}{dt}\right)^2\right] \cos \psi \\ & + \omega^2 \sum_{n\geq 1} \epsilon^n \left(\frac{d^2u_n}{\partial \psi^2} + u_n\right) \\ & - \sum_{n\geq 1} \epsilon^n \left[\omega^2 - \left(\frac{d\psi}{dt}\right)^2\right] \frac{\partial^2 u_n}{\partial \psi^2}. \end{aligned} \tag{2.5}$$

The complete dependence of Eq. (2.5) on ϵ is obtained by substituting Eq. (2.3) into Eq. (2.5).

The right-hand side of Eq. (2.1) can also be expressed in powers of ϵ as

$$\begin{aligned} \epsilon f(x) &= \epsilon f\left(\rho \cos \psi + \sum_{n\geq 1} \epsilon^n u_n(\rho, \psi)\right) \\ &= \epsilon f(\rho \cos \psi) + \sum_{n\geq 1} \epsilon^{n+1} u_n(\rho, \psi) \frac{d^n f}{dx^n}(\rho \cos \psi). \end{aligned} \tag{2.6}$$

In accordance with Eq. (2.1), the equations to be satisfied by $u_1, u_2, \cdots$ are found by equating like powers of ϵ in Eq. (2.5) and (2.6). The equations take the form

$$\omega^2 \left(\frac{\partial^2 u_n}{\partial \psi^2} + u_n\right) = f_{n-1}(\rho, \psi) + 2\omega b_n(\rho) \rho \cos \psi, \tag{2.7}$$

where

$$f_0(\rho, \psi) = f(\rho \cos \psi),$$

$$f_1(\rho, \psi) = u_1 \frac{df}{dx}(\rho \cos \psi) + b_1(\rho)^2 \rho \cos \psi - 2\omega b_1(\rho) \frac{\partial^2 u_1}{\partial \psi^2},$$

and similarly for the others.

In solving these equations, two additional criteria are imposed by Bogoliubov and Krylov. The first condition requires that (in nth order) the coefficient $b_n(\rho)$ is to be chosen so that the right-hand side of Eq. (2.7) is free of terms proportional to $\cos \psi$. The presence of such terms would lead to secular behavior in u_n (i.e., u_n would then contain trigonometric functions with coefficients which are polynomials in t). (For further details on secular behavior see the monograph by Bogoliubov and Mitripolsky[4]; for a shorter discussion of the problem, particularly as it relates to quantum mechanical problems, see the paper by S. Ø. Aks.[2]) From Eq. (2.3) it follows that $b_n(\rho)$ is the nth-order correction to the fundamental frequency of the oscillation and, as expected, it depends on ρ.

The second condition on the solutions of Eq. (2.7) is that each u_n be orthogonal to $\cos \psi$ in the sense that

$$\int_0^{2\pi} u_n(\rho, \psi) \cos \psi \, d\psi = 0, \quad n \geq 1.$$

This means that ρ is the full amplitude of the fundamental-frequency term.

With these conditions, the method of Bogoliubov and Krylov becomes a systematic procedure for finding nonsecular approximate solutions of Eq. (2.1). The approximations thus obtained depend on two parameters, ρ and the integration constant associated with the amplitude (2.3). They are, in principle, found using the initial conditions which are imposed on the solution. It should be noted that the actual determination of ρ (for an nth-order approximation) may involve solving a very complicated algebraic equation. To find ρ in nth-order, it may be useful, in view of the assumption that ϵ is small, to expand ρ as

$$\rho = \sum_{k=0}^{n} \epsilon^k \rho_k.$$

The initial conditions are handled in this way in the case of quantum mechanics.

The monograph by Bogoliubov and Mitropolsky[4] contains a number of applications of the approximation procedure described above and the interested reader is directed there for further information. The presentation here is offered primarily as a pattern for the approximation scheme which is introduced in the next section for use in connection with problems in quantum mechanics.

3. THE APPROXIMATION THEORY FOR QUANTUM-MECHANICAL NONLINEAR OSCILLATORS

In this section we develop a method analogous to the method of Bogoliubov and Krylov for application to q-number equations of the type

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x = \epsilon f(x), \tag{3.1}$$

where the q-number x is taken to satisfy the initial condition

$$\left[x(0), \frac{dx(0)}{dt}\right] = i\hbar. \tag{3.2}$$

In analogy with the work of Bogoliubov and Krylov, we seek approximate solutions of the form

$$x(t) = \rho e^{-i\Omega t}a + a^\dagger e^{i\Omega t}\rho + \sum_{n\geq 1}\epsilon^n[S_n(t; N, a) + \tilde{S}_n(t; N, a^\dagger)]. \tag{3.3}$$

The following definitions and conditions are placed on the parameters appearing in Eq. (3.3):

(a) a, $a^\dagger$: a and $a^\dagger$ are the usual annihilation and creation operators, respectively; thus they satisfy the commutation relation

$$[a, a^\dagger] \equiv aa^\dagger - a^\dagger a = I.$$

(b) N: $N = a^\dagger a$; N is the standard number operator.

(c) ρ: $\rho = \rho^\dagger$, $\rho = \rho(N)$, $d\rho/dt \equiv 0$; ρ is a symmetric operator which depends only on N and is constant in the time t.

(d) Ω: $\Omega = \Omega^\dagger$,

$$\Omega = \omega + \sum_{n\geq 1}\epsilon^n B_n(N);$$

Ω is a symmetric operator, depending on N, which equals ω when $\epsilon = 0$ is satisfied.

(e) $S_n(t; N, a)$: S_n is function of the variable t, N, and a in special order; i.e., S_n is written in the form

$$S_n(t; N, a) = \sum_{k\geq 0} S_{n,k}(t; N)a^k.$$

(f) $\tilde{S}_n(t; N; a^\dagger)$: $\tilde{S}_n = S_n^\dagger$; thus $\tilde{S}_n$ is also in special order and can be written in the form

$$\tilde{S}_n(t; N, a^\dagger) \equiv \sum_{k\geq 0} a^{\dagger k}\tilde{S}_{n,k}(t; N) = \sum_{k\geq 0} a^{\dagger k}S^*_{n,k}(t, N),$$

where the asterisk (*) denotes complex conjugation.

Condition (d) is of particular importance in what follows and is directly analogous to the condition expressed in Eq. (2.3) for the c-number case. The requirement that Ω is dependent on N was established in a previous paper[2] for the case where $f(x) = x^3$. We shall show that this is, in general, a consistent assumption. As in the method of Bogliubov and Krylov, the role of the quantities $B_n(N)$, $n = 1, 2, 3, \cdots$, is to remove secular behavior. We note that secularity arises in the approximation procedure through auxiliary equations of the form

$$\left(\frac{d^2}{dt^2} + \alpha^2\right)f = p(ae^{-i\alpha t}, e^{i\alpha t}a^\dagger) = q^{(-)}(a, a^\dagger)e^{-i\alpha t} + e^{i\alpha t}q^{(+)}(a, a^\dagger) + \text{terms with other dependence on } t.$$

The secularity comes from the inhomogeneous terms which are solutions of the homogeneous equation and these terms occur when the number of factors of the form $ae^{-i\alpha t}$ is greater or less, by unity, than the number of factors $e^{i\alpha t}a^\dagger$. In such terms the number of creation operators is greater or less by unity than the number of annihilation operators; hence $q^{(-)}(a, a^\dagger)$ can be written as $q^{(-)}(N)a$; $q^{(+)}(a, a^\dagger)$ can be written as $a^\dagger q^{(+)}(N)$. In renormalizing the frequency to avoid secular behavior, such inhomogeneous terms are brought to the left-hand side of the equation and are then lumped with the frequency terms as contributions to the quantities B_n, $n = 1, 2, \cdots$. This is the basis of the requirement that Ω depends only on N.

The dependence of ρ on N was also established in the paper mentioned above.[2] The special ordering conditions on S_n and $\tilde{S}_n$ are introduced for technical reasons which become evident in the following. We remark that any function of the creation and annihilation operators can be rewritten in this special order; a proof of this is given in the Appendix.

The choice of a concrete representation of the commutation relation $[a, a^\dagger] = I$ is of some interest. We note that in the c-number case, when Eq. (2.3) is integrated, there occurs an integration constant which plays the role of an initial phase, whereas the analog of the integration constant is omitted in Eq. (3.3). Suppose that an initial phase ϕ, dependent on N and symmetric, is included in Eq. (3.3) which then

[4] N. Bogoliubov and Yu. A. Mitropolsky, *Asymptotic Methods in the Theory of Non-linear Oscillations* (Hindustan Publishing Corp., Delhi, 1961).

becomes

$$x(t) = \rho e^{-i(\Omega t+\phi)}a + a^\dagger e^{i(\Omega t+\phi)}\rho + \sum_{n\geq 1}\epsilon^n(S_n + \tilde{S}_n)$$
$$= \rho e^{-i\Omega t}(e^{-i\phi}a) + (e^{-i\phi}a)^\dagger e^{i\Omega t} + \sum_{n\geq 1}\epsilon^n(S_n + \tilde{S}_n). \tag{3.4}$$

The commutator of $e^{-i\phi}a$ and its adjoint is given by

$$[e^{-i\phi}a, (e^{-i\phi}a)^\dagger] = e^{-i\phi}aa^\dagger e^{+i\phi} - a^\dagger e^{+i\phi}e^{-i\phi}a$$
$$= e^{-i\phi}(N+I)e^{+i\phi} - N$$
$$= N + I - N$$
$$= I.$$

We have made use of the fact that ϕ is a function of N to obtain this result. It follows that $e^{-i\phi}a$ and $(e^{+i\phi}a)^\dagger$ are new creation and annihilation operators, respectively; thus the inclusion of a phase dependent on N induces an equivalent (but trivially related) representation of the commutation relations of a and $a^\dagger$. Consequently, the choice of a concrete representation for a and $a^\dagger$ amounts to a choice of the phase constant.

Starting with Eq. (3.3), we evaluate the following quantities:

$$\frac{dx}{dt} = -i\Omega\rho e^{-i\Omega t}a + a^\dagger e^{i\Omega t}\rho i\Omega + \sum_{n\geq 1}\epsilon^n\left(\frac{dS_n}{dt} + \frac{d\tilde{S}_n}{dt}\right), \tag{3.5}$$

$$\frac{d^2x}{dt^2} = -\Omega^2\rho e^{-i\Omega t}a - a^\dagger e^{i\Omega t}\rho\Omega^2 + \sum_{n\geq 1}\epsilon^n\left(\frac{d^2S_n}{dt^2} + \frac{d^2\tilde{S}_n}{dt^2}\right), \tag{3.6}$$

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x = (\omega^2 - \Omega^2)\rho e^{-i\Omega t}a + a^\dagger e^{i\Omega t}\rho(\omega^2 - \Omega^2) + \sum_{n\geq 1}\epsilon^n\left[\left(\frac{d^2S_n}{dt^2} + \Omega^2 S_n\right) + \left(\frac{d^2\tilde{S}_n}{dt^2} + \tilde{S}_n\Omega^2\right)\right] + \sum_{n\geq 1}\epsilon^n[(\omega^2 - \Omega^2)S_n + \tilde{S}_n(\omega^2 - \Omega^2)]. \tag{3.7}$$

The completed expansion of Eq. (3.7) in powers of ϵ is obtained by using Eq. (3.4d):

$$\omega^2 - \Omega^2 = -2\epsilon\omega B_1(N) - \epsilon^2(2\omega B_2(N) + B_1(N)^2) - \cdots \tag{3.8}$$

and

$$\left(\frac{d^2}{dt^2} + \omega^2\right)x = \epsilon\left[\left(\frac{d^2S_1}{dt^2} + \Omega^2 S_1\right) + \left(\frac{d^2\tilde{S}_1}{dt^2} + \tilde{S}_1\Omega^2\right) - 2\omega B_1\rho e^{-i\Omega t}a - a^\dagger e^{i\Omega t}\rho 2\omega B_1\right]$$
$$+ \epsilon^2\left[\left(\frac{d^2S_2}{dt^2} + \Omega^2 S_2\right) + \left(\frac{d^2\tilde{S}_2}{dt^2} + \tilde{S}_2\Omega^2\right) - (2\omega B_2 + B_1^2)\rho e^{-i\Omega t} - a^\dagger e^{i\Omega t}\rho(2\omega B_2 + B_1^2) - 2\omega B_1 S_1 - \tilde{S}_1 2\omega B_1\right]$$
$$+ \cdots$$
$$+ \epsilon^n\left[\left(\frac{d^2S_n}{dt^2} + \Omega^2 S_n\right) + \left(\frac{d^2\tilde{S}_n}{dt^2} + \tilde{S}_n\Omega^2\right) + \cdots\right]$$
$$+ \cdots. \tag{3.9}$$

It should be noted that each term of Eq. (3.9) is in special order.

The inhomogeneous term $\epsilon f(x)$ in Eq. (3.1) is also expanded in powers of ϵ. We denote

$$y_0(t) = \rho e^{-i\Omega t}a + a^\dagger e^{i\Omega t}\rho,$$
$$y_n(t) = S_n + \tilde{S}_n, \quad n = 1, 2, \cdots,$$

and expand $\epsilon f(x)$ as follows:

$$\epsilon f\left(\sum_{n\geq 0}\epsilon^n y_n\right) = \epsilon f(y_0) + \epsilon^2\left[y_1 \circ \frac{df}{dx}\right](y_0) + \cdots.$$

The notation $y_1 \circ df/dx$ is employed to denote the modified chain rule for use in the case of noncommuting quantities. As an example of the use of the notation, we consider the case $f(x) = x^k$. Then,

$$y_1 \circ \frac{dx^k}{dx} = y_1x^{k-1} + xy_1x^{k-2} + \cdots + x^{k-1}y_1.$$

The result obtained is not, in general, in special order. In order to reexpress such terms in special order, we make use of the equations

$$ag(N) = g(N+1)a \tag{3.10a}$$

and

$$h(N)a^\dagger = a^\dagger h(N+1), \tag{3.10b}$$

where it is assumed that g and h are power series in N.

Let us suppose that ϵf has been expanded in powers of ϵ and has been placed in special order. We denote the resulting expansion by

$$\epsilon f(x) = \epsilon[F_0(y_0) + \tilde{F}_0(y_0)] + \epsilon^2[F_1(y_0) + \tilde{F}_1(y_0)] + \cdots, \tag{3.11}$$

where the terms $F_i(y_0)$ and $\tilde{F}_i(y_0)$, $i = 1, 2, \cdots$, are in special order, as defined in conditions (e) and (f), respectively.

The equations to be satisfied by the quantities S_n and $\tilde{S}_n$, $n = 1, 2, \cdots$, can now be found by comparing coefficients of similar powers in ϵ in Eq. (3.1)

by using Eqs. (3.9) and (3.11). The results are

$$\frac{d^2S_1}{dt^2} + \Omega^2 S_1 = F_0(y_0) + 2\omega B_1 \rho e^{-i\Omega t}, \tag{3.12a}$$

$$\frac{d^2S_2}{dt^2} + \Omega^2 S_2 = F_1(y_0) + 2\omega B_2 \rho e^{-i\Omega t} a + B_1^2 \rho e^{-i\Omega t} a + 2\omega B_1 S_1, \tag{3.12b}$$

$$\frac{d^2S_n}{dt^2} + \Omega^2 S_n = F_{n-1}(y_0) + 2\omega B_n \rho e^{-i\Omega t} a + \cdots. \tag{3.12c}$$

A similar set of equations holds for $\tilde{S}_n$, $n = 1, 2, \cdots$, but the equations are not written out since the solutions are obtained from the solutions of Eqs. (3.12) by taking adjoints, in accordance with condition (f).

As in the method of Bogoliubov and Krylov for the c-number case, two conditions are placed on Eqs. (3.12). The quantity B_n appears first in Eq. (3.12c) for S_n and is determined by the requirement that the right-hand side of the equation is to be free of terms proportional to $e^{-i\Omega t}a$. By choosing B_n so as to remove such terms from the equation, we arrive at an equation whose solution does not show any secular behavior. It is useful to note that B_n may be readily found once B_1 through B_{n-1} and S_1 through S_{n-1} are known; thus, the renormalized frequency is essentially known to one order in ϵ better than the approximation for the solution of the equation of motion.

The other condition requires that the solution of the equations for S_n, $n = 1, 2, \cdots$, be without terms which are solutions of the homogeneous equation. Therefore, all terms of the approximate solution which are proportional to $e^{\pm i\Omega t}$ are contained within the zeroth-order approximation $\rho e^{-i\Omega t}a + a^\dagger e^{i\Omega t}\rho$. As in the c-number case, this condition means that ρ is the full amplitude of the lowest-frequency contribution to the approximate solution. Consequently, the first condition determines the renormalized frequency and the second condition determines what is equivalent to, in the one-dimensional systems being studied here, wavefunction renormalization. We have made the provision in the approximation procedure for the renormalized frequency and amplitude to depend on the number operator N as is suggested by the results of a previous paper noted above.[2]

We stress that the ultimate dependence of the renormalized frequency and amplitude on N is a consequence of the inhomogeneous term of Eq. (3.1) and is forced by the requirement that the approximations we look for are free of secular behavior. It is, of course, always possible [contrary to the previously considered case where $f(x) = x^3$] that the renormalized frequency and amplitude are trivial functions of N raised to the zero power, or, in other words, c-numbers. In the next section, we apply the approximation method described here to concrete examples, and we find that this supposition is fortuitous.

We close this section with a remark on the evaluation of the quantity ρ. To this point, no use has been made of the initial condition embodied in Eq. (3.2). We are, therefore, free to use it to determine ρ. To this end, we require that the Kth-order approximation given by Eq. (3.3) is to satisfy Eq. (3.2) with an error of at most order ϵ^{K+1}. It should be noted that the amplitude ρ is also found in the method of Bogoliubov and Krylov, in the c-number case, by employing the initial conditions and at the same state of the calculation. We add that the actual determination of ρ in either the c-number or q-number cases involves practical difficulties. For example, in the c-number equation with $f(x) = x^3$, the evaluation of ρ involves, for a first-order approximation, the solution of a cubic equation. For a second-order approximation in ϵ, ρ is a root of a fifth-degree polynomial. Consequently, although the approximation procedure laid out above is systematic in that it may, in principle, be carried to any order in ϵ, it, along with most other approximation methods, is limited to the calculation of only the first few orders in ϵ.

Applications of the approximation scheme developed above are carried out in the next section.

4. THE EXAMPLE $f(x) = x^3$

We now specialize the discussion to the case $f(x) = x^3$. This is the quantum version of the first-order correction to a physical pendulum, based on the approximation $\sin\theta = \theta - \frac{1}{6}\theta^3$ rather than $\sin\theta = \theta$. It is also a case which generalizes directly to the self-interacting field theory with a $\lambda\phi^4$ interaction Lagrangian. The method developed in Sec. 3 is employed to find the first- and second-order corrections in ϵ to the frequency Ω (i.e., the operators B_1 and B_2) and the first-order correction to the amplitude ρ and the "position operator." To this end we first determine the function F_0 of Eq. (3.11) in the expansion of f. To determine F_1 will require a knowledge of S_1. Thus,

$$F_0(y_0) + \tilde{F}_0(y_0) = y_0^3. \tag{4.1}$$

We must carry out the product y_0^3 and arrange the result in normal order. The terms containing a higher power of a than of $a^\dagger$ are then represented by the function F_0. The hermitian property of y_0^3 guarantees that the remaining terms give $\tilde{F}_0$. Since the operators ρ and Ω are assumed to be functions of the operator N,

let us use the notations

$$\rho = \rho(N) = \rho_N$$

and

$$\Omega = \Omega(N) = \Omega_N.$$

The special-ordering theorem, Eqs. (3.10), implies the following kind of result:

$$a\rho_N = \rho_{N+1}a, \tag{4.2a}$$

$$e^{i\Omega_N t}a^\dagger = a^\dagger e^{i\Omega_{N+1}t}. \tag{4.2b}$$

Using this theorem, we bring the term y_0^3 into normal order, obtaining

$$\begin{aligned} y_0^3 = \{&\rho_N\rho_{N+1}\rho_{N+2}e^{-i(\Omega_N+\Omega_{N+1}+\Omega_{N+2})t}a^3 \\ &+ [\rho_N\rho_{N-1}^2N + \rho_N^3(N+1) \\ &+ \rho_N\rho_{N+1}^2(N+2)]e^{-i\Omega_N t}a + \text{h.c.}\}. \end{aligned} \tag{4.3}$$

We can immediately identify F_0 as

$$\begin{aligned} F_0(y_0) = &\rho_N\rho_{N+1}\rho_{N+2}e^{-i\Phi_N t}a^3 \\ &+ [N\rho_{N-1}^2 + (N+1)\rho_N^2 \\ &+ (N+2)\rho_{N+1}^2]\rho_N e^{-i\Omega_N t}a, \end{aligned} \tag{4.4}$$

where

$$\Phi_N = \Omega_N + \Omega_{N+1} + \Omega_{N+2}.$$

In order to find S_1 and B_1, we must consider Eq. (3.12a). Thus,

$$\left(\frac{d^2}{dt^2} + \Omega_N^2\right)S_1 = F_0(y_0) + 2\omega B_1\rho_N e^{-i\Omega_N t}a. \tag{4.5}$$

The value of B_1 follows from Eqs. (4.4) and (4.5) due to the condition that the right-hand side of Eq. (4.5) contains no terms proportional to the fundamental. Thus,

$$B_1 = -(2\omega)^{-1}[N\rho_{N-1}^2 + (N+1)\rho_N^2 + (N+2)\rho_{N+2}^2]. \tag{4.6}$$

Then the equation for S_1 simplifies to

$$\left(\frac{d^2}{dt^2} + \Omega_N^2\right)S_1 = \rho_N\rho_{N+1}\rho_{N+2}e^{-i\Phi_N t}a^3. \tag{4.7}$$

The expansion of S_1 in powers of a from Condition (a) (Sec. 3), when substituted here, gives

$$\sum_{k=0}^{\infty}\left(\frac{d^2}{dt^2} + \Omega_N^2\right)S_{1,k}a^k = G_N e^{-i\Phi_N t}a^3, \tag{4.8}$$

where $G_N = \rho_N\rho_{N+1}\rho_{N+2}$. The linear independence of the monomials a^k (in the Hilbert space) leads to a reduced set of equations:

$$\left(\frac{d^2}{dt^2} + \Omega_N^2\right)S_{1,k} = 0, \quad k \geq 0,\ k \neq 3, \tag{4.9a}$$

$$\left(\frac{d^2}{dt^2} + \Omega_N^2\right)S_{1,3} = G_N e^{-i\Phi_N t}. \tag{4.9b}$$

In light of the condition that ρ_N be the full amplitude of the fundamental frequency contribution, the solutions of Eqs. (4.7) are

$$S_{1,k} = 0, \quad k \geq 0,\ k \neq 3, \tag{4.10a}$$

$$S_{1,3} = D_N G_N e^{-i\Phi_N t}, \tag{4.10b}$$

where

$$D_N = (\Omega_N^2 - \Phi_N^2)^{-1}.$$

We require that $\Omega_N^2 - \Phi_N^2$ is a nonsingular operator which is reasonable for sufficiently small values of ϵ.

Thus, we have obtained the first-order approximation to the operator x and the frequency Ω in terms of the amplitude ρ. An explicit representation of S_1 and B_1 in terms of N requires a knowledge of the explicit form of the amplitude ρ_N. An adequate approximation for ρ, to order ϵ, will be found by satisfying the commutation relation to order ϵ^2. But to maintain continuity, we proceed first to a determination of B_2.

Referring to Eq. (3.11) and the preceding remarks, we have

$$\begin{aligned} F_1(y_0) + \tilde{F}_1(y_0) &= \left[y_1 \circ \frac{df}{dx}\right](y_0) \\ &= y_1y_0^2 + y_0y_1y_0 + y_0^2y_1, \end{aligned} \tag{4.11}$$

where

$$y_1 = S_1 + \tilde{S}_1.$$

We have seen that S_1 is of the form

$$S_1 = S_1(t; N, a) = D_N G_N \exp(-i\Phi_N t)a^3,$$

and in conjunction with the form of y_0 and the special-ordering theorem, we can collect, from the right-hand side of Eq. (4.11), the terms which contain powers of a in special order as opposed to those which contain $a^\dagger$. The terms in a give F_1; those in $a^\dagger$ give $\tilde{F}_1$. The result is as follows:

$$\begin{aligned} F_1(y_0) = &(D_N + D_{N+1} + D_{N+2})G_N\rho_{N+3}\rho_{N+4}e^{-i\Psi_N t}a^5 \\ &+ \{[N\rho_{N-1}^2 + (N+1)\rho_N^2]G_N \\ &+ N\rho_{N-1}\rho_{N+2}G_{N-1} + (N+4)\rho_N\rho_{N+3}G_{N+1} \\ &+ [(N+4)\rho_{N+3}^2 + (N+3)\rho_{N+2}^2]G_N\}e^{-i\Phi_N t}a^3 \\ &+ \{N(N-1)\rho_{N-2}^2\rho_{N-1}^2D_{N-2} \\ &+ N(N+2)\rho_{N-1}^2\rho_{N+1}^2D_{N-1} \\ &+ (N+2)(N+3)\rho_{N+1}^2\rho_{N+2}^2D_N\}\rho_N e^{-i\Omega_N t}a, \end{aligned} \tag{4.12}$$

where

$$\Psi_N = \Phi_N + \Omega_{N+3} + \Omega_{N+4}.$$

The coefficient B_2 is determined from Eq. (3.12b):

$$\begin{aligned} \left(\frac{d^2}{dt^2} + \Omega_N^2\right)S_2 = &F_1(y_0) + 2\omega B_1 S_1 \\ &+ (B_1^2 + 2\omega B_2)\rho_N e^{-i\Omega_N t}a. \end{aligned} \tag{4.13}$$

This coefficient must be chosen so that the right-hand side of Eq. (4.13) contains no term proportional to $e^{-i\Omega_N t}a$ as in the case of B_1. Since the $2B_1S_1$ term contains only $e^{-i\Phi_N t}a^3$, this requirement becomes

$$[N(N-1)\rho_{N-2}^2\rho_{N-1}^2D_{N-2} + N(N+2)\rho_{N-1}^2\rho_{N+1}^2D_{N-1} + (N+2)(N+3)\rho_{N+1}^2\rho_{N+2}^2D_N]\rho_N e^{-i\Omega_N t}a + (2\omega B_2 + B_1^2)\rho_N e^{-i\Omega_N t}a = 0. \quad (4.14)$$

From Eqs. (4.6) and (4.10b), we can substitute expressions for B_1 and $S_{1,3}^N$ into Eq. (4.14) and solve for B_2. We obtain the following result:

$$B_2 = -\frac{1}{2\omega}\Big\{N(N-1)\rho_{N-2}^2\rho_{N-1}^2D_{N-2} + N(N+2)\rho_{N-1}^2\rho_{N+1}^2D_{N-1} + (N+2)(N+3)\rho_{N+1}^2\rho_{N+2}^2D_N + \frac{1}{4\omega^2}[N\rho_{N-1}^2 + (N+1)\rho_N^2 + (N+2)\rho_{N+1}^2]\Big\}. \quad (4.15)$$

If one desires the order ϵ^2 correction to x, i.e., the function (operator) S_2, then one must solve Eq. (4.13) using only the a^3 and a^5 terms from $F_1(y_0)$ and the term $2\omega B_1S_1$ on the right-hand side.

Now we proceed to obtain an explicit approximate expression for ρ_N in terms of N by satisfying the commutation relation (3.2), thus giving explicit approximations to ρ_N to order ϵ, and Ω_N to order ϵ^2. We first carry out the procedure to order unity and then to order ϵ to indicate how one would handle the general case.

However, there is a fundamental ambiguity in any such approximation in powers of ϵ. The zeroth-order values of ρ_N and Ω_N are just those for the ordinary linear harmonic oscillator, $(\hbar/2\omega)^{\frac{1}{2}}$ and ω, respectively, so that $\rho_N^2\Omega_N = \frac{1}{2}\hbar$. Any operator differing from these values by terms of order ϵ will satisfy Eq. (3.2) to order ϵ, i.e., $\rho_N^2\Omega_N = \frac{1}{2}\hbar + O(\epsilon)$. There is, in fact, no unique way to make the approximation order by order. We require only that the method selected give accurate results to the order under consideration, the same results given by every consistent scheme.

The functions ρ_N and Ω_N are themselves power series in ϵ, an explicit fact for Ω_N, but also true implicitly for ρ_N. The value for ρ_N must be chosen in order n to satisfy the commutator to order $(n+1)$. This fixes the form of the operator ρ_N to order ϵ^n. An equivalent process is necessary in actually evaluating the approximate solution in the classical case. This point is not made explicit by Bogoliubov and Krylov, although they do handle the solution this way in carrying out the first-order approximation. The procedure we have followed is next presented in detail.

In finding Ω_N to order ϵ^n, the coefficients B_n, $B_{n-1}, \cdots, B_1$ must be known to order $\epsilon^0, \epsilon^1, \cdots, \epsilon^n$, respectively. From the *exact* expression for B_n in terms of Ω_N and ρ_N, we can thus use their order ϵ^{n-1} values to find B_n to order ϵ^n; their order ϵ^{n-2} values to find B_{n-1} to order ϵ^n; etc.

In computing x to order ϵ^n, we use Ω_N to order ϵ^n *wherever* it occurs in an exponential, but we substitute ρ_N (and Ω_N wherever it is not in an exponential) to order ϵ^n in the fundamental, ϵ^{n-1} in S_1, and so on until in S_n we use Ω_N to order $\epsilon^0 = 1$. Thus, we generate the solution x as a series of harmonics of the *renormalized* frequency to order n, but we do not retain unnecessary amplitude corrections. Since no exponentials appear in the commutation relation, in evaluating the commutator, ρ_N and Ω_N are handled on an equivalent basis. This procedure differs from other consistent ones only by terms of order ϵ^{n+1} and higher, when calculating in order ϵ^n. In what follows, we shall apply Ω_N to a complete calculation to order ϵ^2 of the operators in terms of N explicitly.

To order ϵ^0 we have

$$\Omega_N^{(0)} = \omega, \quad (4.16a)$$

which implies

$$x_0 = \rho_N^{(0)}\exp(-i\omega t)a + a^\dagger\exp(i\omega t)\rho_N^{(0)}. \quad (4.16b)$$

Temporarily dropping the superscript zero, the commutator of x and $\dot{x}$ at $t = 0$ is as follows:

$$[x(0), \dot{x}(0)] = 2i\omega[N\rho_{N-1}^2 - (N+1)\rho_N^2]. \quad (4.17)$$

This commutator is required to be $i\hbar + O(\epsilon)$, leading to the condition

$$N\rho_{N-1}^2 - (N+1)\rho_N^2 = (\hbar/2\omega) + O(\epsilon). \quad (4.18)$$

The final question is whether or not ρ_N can have terms of order ϵ^0 which depend on N. If ρ_N is taken to be a c-number, it follows at once that $\rho_N = (\hbar/2\omega)^{\frac{1}{2}}$. We express ρ_N^2 as a power series in the operator N by

$$\rho_N^2 = \sum_{j=0}^{\infty} T_j N^j, \quad (4.19a)$$

where

$$T_j = \sum_{k=0}^{\infty} T_{jk}\epsilon^k. \quad (4.19b)$$

Then we obtain the following relationship:

$$(N+1)\rho_N^2 - N\rho_{N-1}^2 = T_{00} + \sum_{j=1}^{\infty} N[N^{j-1} + N^j - (N-1)^j]T_{j0} + O(\epsilon). \quad (4.20)$$

The terms in the summation are of such a nature that no choice of the $\{T_{j0}\}$ can cause the sum to be a c-number. Thus,

$$T_{00} = \hbar/2\omega, \tag{4.21a}$$

$$T_{j0} = 0, \quad j \geq 1, \tag{4.21b}$$

and, finally, we recover the well-known result for the ordinary harmonic oscillator:

$$\rho_N^{(0)} = (\hbar/2\omega)^{\frac{1}{2}}. \tag{4.22}$$

The complete zeroth-order approximation is contained in Eqs. (4.16) and (4.22). But it should be emphasized that this approximation was obtained by the method outlined above, and not by setting $\epsilon = 0$ and solving Eq. (2.1). Thus,

$$\Omega_N^{(0)} = \omega, \tag{4.23a}$$

$$\rho_N^{(0)} = (\hbar/2\omega)^{\frac{1}{2}}, \tag{4.23b}$$

$$x_0 = (\hbar/2\omega)^{\frac{1}{2}}[e^{-i\omega t}a + a^\dagger e^{i\omega t}]. \tag{4.23c}$$

In carrying out the next higher order, accurate to order ϵ, from Eq. (4.6) we have

$$B_1 = -(2\omega)^{-1}[N(\rho_{N-1}^{(0)})^2 + (N+1)(\rho_N^{(0)})^2 + (N+2)(\rho_{N+1}^{(0)})^2] \tag{4.24a}$$

$$= -(3\hbar/4\omega^2)(N+1), \tag{4.24b}$$

and the renormalized frequency to order ϵ is

$$\Omega_N^{(1)} = \omega[1 - \epsilon(3\hbar/4\omega^3)(N+1)]. \tag{4.25}$$

According to the procedure established above, the first-order correction to x takes on the following form:

$$S_1 = -(8\omega^2)^{-1}(\hbar/2\omega)^{\frac{3}{2}}e^{-i3\Omega_N{}^{(1)}t}a^3. \tag{4.26}$$

The entire first-order x-operator is, then,

$$x_1(t) = \rho_N^{(1)}e^{-i\Omega_N{}^{(1)}t}a - \epsilon(\hbar/16\omega^3)\rho_N^{(0)}e^{-i3\Omega_N{}^{(1)}t}a^3 + \text{h.c.} \tag{4.27}$$

Again, dropping temporarily the superscript *one* on ρ_N, the commutator becomes

$$\begin{aligned}[x_1(0), \dot{x}_1(0)] &= i\omega(2(N+1)\rho_N^2 - N\rho_{N-1}^2 - N\rho_{N+1}^2 - \epsilon(3\hbar/4\omega^3) \\ &\quad \times [2(N+1)^2\rho_N^2 - N^2\rho_{N-1}^2 - N(N+2)\rho_{N+1}^2]) \\ &\quad + \epsilon(i\omega)(\{(3\hbar/4\omega^3)\rho_N\rho_{N+1} + (\hbar/4\omega^3)(\hbar/2\omega)^{\frac{1}{2}} \\ &\quad \times [N\rho_{N-1} - (N+3)\rho_{N+2}]\}a^2 + \text{H.c.}) \\ &\quad + \epsilon(i\omega)[(\hbar/8\omega^3)(\hbar/2\omega)^{\frac{1}{2}}(\rho_N - \rho_{N+3})a^4 + \text{h.c.}].\end{aligned} \tag{4.28}$$

The polynomials in a and $a^\dagger$ are linearly independent in Hilbert space. Thus, three equations must be satisfied by ρ_N to various orders in ϵ:

$$\begin{aligned}&2(N+1)\rho_N^2 - N\rho_{N-1}^2 - N\rho_{N+1}^2 - \epsilon(3\hbar/4\omega^3) \\ &\quad \times [2(N+1)^2\rho_N^2 - N^2\rho_{N-1}^2 - N(N+2)\rho_{N+1}^2] \\ &\qquad = (\hbar/\omega) + O(\epsilon^2),\end{aligned} \tag{4.29a}$$

$$(3\hbar/4\omega^3)\rho_N\rho_{N+1} + (\hbar/4\omega^3)(\hbar/2\omega)^{\frac{1}{2}} \times [N\rho_N - (N+3)\rho_{N+2}] = 0 + O(\epsilon), \tag{4.29b}$$

$$\rho_N - \rho_{N+3} = 0 + O(\epsilon). \tag{4.29c}$$

The third, Eq. (4.29c), is immediate since

$$\rho_N = (\hbar/2\omega)^{\frac{1}{2}} + O(\epsilon).$$

The second equation also is satisfied for the same reason. The first equation, Eq. (4.29a), will give the first-order correction to ρ_N. Without presenting the details, when one refers to the expansion of Eq. (4.19), the coefficients T_{j1} are determined (T_{jk} for $k \geq 2$ are arbitrary, but we choose them as zero) to be

$$T_{01} = T_{11} = (3\hbar^2/4\omega^3), \tag{4.30a}$$

$$T_{j1} = 0, \quad j \geq 2. \tag{4.30b}$$

Thus we can list the first-order results for Ω_N, ρ_N, and x:

$$\Omega_N^{(1)} = \omega[1 - \epsilon(3\hbar/4\omega^3)(N+1)], \tag{4.31a}$$

$$\rho_N^{(1)} = (\hbar/2\omega)^{\frac{1}{2}}[1 + \epsilon(3\hbar/8\omega^3)(N+1)], \tag{4.31b}$$

$$\begin{aligned}x_1(t) &= \rho_N^{(1)}e^{-i\Omega_N{}^{(1)}t}a + a^\dagger e^{i\Omega_N{}^{(1)}t}\rho_N^{(1)} \\ &\quad - \epsilon(\hbar/16\omega^3)(\hbar/2\omega)^{\frac{1}{2}}[e^{-i3\Omega_N{}^{(1)}t}a^3 \\ &\qquad + a^{\dagger 3}e^{+i3\Omega_N{}^{(1)}t}].\end{aligned} \tag{4.31c}$$

Finally, we give the result from Eq. (4.15) for B_2, obtained as above by making the approximations $\rho_N \cong \rho_N^{(0)} = (\hbar/2\omega)^{\frac{1}{2}}$ and $\Omega_N \cong \Omega_N^{(0)} = \omega$. We also list $\Omega_N^{(2)}$ for completeness:

$$B_2 = -(3\hbar^2/64\omega^5)(5N^2 + 10N + 4), \tag{4.32a}$$

$$\begin{aligned}\Omega_N^{(2)} = \omega[1 &- \epsilon(3\hbar/4\omega^3)(N+1) \\ &- \epsilon^2(3\hbar^2/64\omega^5)(17N^2 + 34N + 21)].\end{aligned} \tag{4.32b}$$

5 CONCLUSIONS

The equation of motion for x studied in Sec. 4 can be derived from the Hamiltonian

$$H = \tfrac{1}{2}(p^2 + \omega^2x^2) - \tfrac{1}{4}\epsilon x^4.$$

If one now takes the first-order expressions obtained for ρ_N and Ω_N and calculates the first-order Hamiltonian, one obtains

$$H^{(1)} = (N + \tfrac{1}{2})\hbar\omega - (3\epsilon\hbar/8\omega^3) \times (N^2 + N + \tfrac{1}{2})\hbar\omega, \tag{5.1}$$

where $N = a^\dagger a$. To this order, the Hamiltonian has the unperturbed eigenstates $|n\rangle$ as eigenstates, where

$$H^{(0)}|n\rangle = (n + \tfrac{1}{2})\hbar\omega\,|0\rangle$$

and (5.2)

$$|n\rangle = (n!)^{-\frac{1}{2}}(a^\dagger)^n\,|0\rangle.$$

The first-order corrected energy eigenvalues are

$$E_n^{(1)} = (n + \tfrac{1}{2})\hbar\omega - (3\epsilon\hbar/8\omega^3)(n^2 + n + \tfrac{1}{2})\hbar\omega. \quad (5.3)$$

It is interesting to compare this result with that of the time-independent perturbation theory in the Schrödinger representation. There, one finds that

$$E_n^{(1)} - (n + \tfrac{1}{2})\hbar\omega = \int_{-\infty}^{\infty} dx\phi_n^*(x)\left(-\frac{\epsilon}{4}x^4\right)\phi_n(x)$$
$$= -(3\epsilon\hbar/8\omega^3)(n^2 + n + \tfrac{1}{2})\hbar\omega, \quad (5.4)$$

in agreement with Eq. (5.3). This agreement provides a partial check on the assumptions made in our method, such as the existence of a power series for Ω_n and ρ_n in the operator $N = a^\dagger a$.

Thus, there are other methods,[2] many in fact, by which one can obtain first-order results for the one-dimensional problems considered in this paper. Also, one could work in Schrödinger representation to obtain the perturbed energies E_n and the perturbed eigenstates ϕ_n to arbitrary order in ϵ, assuming convergence. However, the neutral scalar field problem with interaction Hamiltonian density $\frac{1}{4}\epsilon\phi^4(x)$ is known to lead to an infinite set of coupled, nonlinear oscillators when quantized in a box. Since there is no equivalent Schrödinger theory for such a problem, a method which generalizes to it must involve the Heisenberg operator. One of the coupling terms for each of the denumerable infinity of such oscillators is the one we have chosen to investigate in Sec. 4. Since we have shown that a reasonable long-time approximation requires a q-number renormalized frequency and a q-number renormalization of the coordinate amplitude, it is appropriate to expect a q-number renormalized mass and a q-number wave function renormalization in the self-coupled field.

Such q-number renormalizations are new to the perturbation solutions of field theory. Previously, they have been excluded in part due to the handling of the singularities in Green's function integral solutions by a method called "adiabatic switching."

The reader is referred elsewhere for a detailed demonstration of the existence of a q-number mass renormalization in the $\epsilon\phi^4$ field theory.[5]

Finally, it should be emphasized that the present method gives a consistent procedure for calculating, to arbitrary order in ϵ, a condition one requires for the generalization of the method to field theory in order to include production processes and radiative corrections.

[5] S. Ø. Aks and R. A. Carhart, "Asymptotic Behaviour and Operator Mass Renormalization in the ϕ^4 Model," Nuovo Cimento (to be published).

APPENDIX

It is shown here that every polynomial over the variables a and $a^\dagger$ can be written as a polynomial in special order. We start with the well-known result that such polynomials can be re-expressed as polynomials in normal order, as linear combinations of terms of the form $(a^\dagger)^k a^\ell$. We prove below that these terms can be rewritten as polynomials in special order. Then, using the fact that composite polynomial functions of polynomials are again polynomials, we have the desired result.

It remains for us to prove that $(a^\dagger)^k a^\ell$ can be written as a polynomial in special order. The proof is given in two parts. The case $k \geq \ell$ is considered first and is carried out using mathematical induction on the integers $M = k + \ell$.

For $M = 1$, remembering that $k \geq \ell$, the only term to consider is $a^\dagger$ which is already in special order.

Using the induction hypothesis, i.e., all terms with $n = k + \ell$ can be written as polynomials in special order, we show that all terms with $k + \ell = M + 1$ can be written as polynomials in special order.

The only term occurring in the case where $k = M + 1, \ell = 0$ [i.e., $(a^\dagger)^{M+1}$] is already in special order. For $k + \ell = M + 1$, with $\ell \geq 1$, we write

$$(a^\dagger)^k a^\ell[(a^\dagger)^k a^{\ell-1}]a.$$

The term in the bracket is of the type $K + (\ell - 1) = n$ and can, in accordance with the induction hypothesis, be expressed as a polynomial in special order. Thus $(a^\dagger)^k a^\ell$, $k + \ell = n + 1$, and $\ell \geq 1$ is a linear combination of terms of the form

$$[(a^\dagger)^r N^s]a = (a^\dagger)^r a(N + 1)^s = (a^\dagger)^{r-1}N(N + 1)^s,$$

where r and s are positive integers (in particular, $r \geq 1$).

Summarizing to this point, we have shown by mathematical induction that all normal-ordered terms of the form $(a^\dagger)^k a^\ell$, with $k \geq \ell$, can be expressed as a polynomial in special order.

Terms of the form $(a^\dagger)^\ell a^k$, with $k \geq \ell$, remain to be considered. We note that

$$(a^\dagger)^\ell a^k = [(a^\dagger)^k a^\ell]^\dagger,$$

and the term in the bracket is of the form considered above. Thus it can be put into the form of a polynomial in special order. Consequently, its adjoint can also be written a polynomial in special order. This completes the argument.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Spherically Symmetric Boundary-Value Problems in One-Speed Transport Theory*

K. M. CASE,† R. ZELAZNY‡
Department of Physics, The University of Michigan, Ann Arbor, Michigan
AND
M. KANAL
Space Physics Laboratory, The University of Michigan, Ann Arbor, Michigan

(Received 25 April 1969)

A Green's function technique is applied to one-velocity neutron problems with spherical symmetry. The main advantage of the approach is that it bypasses the need to construct explicitly the appropriate spherical eigenfunctions. Indeed, these can be directly deduced from this new formulation, if one so desires.

1. INTRODUCTION

The eigenfunction expansion technique has achieved considerable success in dealing with boundary-value problems in one-speed linear transport theory. The method developed by one of the authors,[1] though applied extensively to boundary-value problems with planar boundary conditions,[1-4] is readily adapted to more general geometrical configurations. The essential feature of this technique lies in constructing a complete set of eigenfunctions (normal modes) of the appropriate transport equation, expanding the neutron angular density in terms of the complete set, and finding the expansion coefficients from the boundary conditions. There are some drawbacks in this kind of treatment. They are, among others:

(i) The set of eigenfunctions (for example, the energy-dependent transport equation) may not form a complete set, which means one must construct appropriate additional functions to make the set complete.

(ii) In most cases it is not always easy to prove completeness.

In a recent paper by Case,[5] a fresh approach has been introduced. It draws on analogy with the Green's function technique in dealing with classical boundary-value problems. The advantages of this approach, among others, are:

(i) It incorporates the normal mode expansion technique in the scheme.

(ii) The eigenfunctions arise in a rather natural way, and thus the necessity of proving their completeness (if they form a complete set) is eliminated.

This paper utilizes the new approach to deal with various spherically symmetric boundary-value problems in one-speed transport theory. In particular, we treat albedo, critical, and Milne problems for the interior of a sphere, and the Milne problem for the exterior of a black sphere. In formulating the boundary-value problems for specific cases, we encounter an apparent difficulty in managing the regular integral equations by analytic methods. To circumvent this difficulty, we introduce a reduction operator which permits us to transform these regular integral equations into integral equations with singular kernels, but with the original coefficients. In other words, the reduction operator essentially reduces the spherical eigenfunctions in the integral equations to the planar ones. The resulting singular integral equations are then solved for the coefficients by the conventional method developed for planar problems.

2. CONSTRUCTION OF THE GREEN'S FUNCTION

The time-independent transport equation in the one-speed approximation is

$$(1 + \mathbf{\Omega} \cdot \nabla)\psi(\mathbf{r}, \mathbf{\Omega}) = \frac{c}{4\pi} \int d\Omega' \psi(\mathbf{r}, \mathbf{\Omega}') + Q(\mathbf{r}, Q), \tag{1}$$

where $\mathbf{\Omega} = \mathbf{v}/v$ is the unit velocity vector, $\mathbf{r}$ is the position vector, $Q(\mathbf{r}, \mathbf{\Omega})$ is a given neutron-source function, and c is the average number of secondary neutrons per collision produced by a neutron of velocity $\mathbf{v}$. In this treatment we will assume that c is a known constant.

* The work reported here was supported in part by the National Science Foundation and by NASA contracts NAS5-9113 and NAS8-21086.

† Present address: The Rockefeller University, New York, N.Y. 10021.

‡ Present address: Institute of Nuclear Research of the Polish Academy of Science, Institute of Theoretical Physics, Warsaw University, Warsaw, Poland.

[1] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley Pub. Co., Inc., Reading, Mass., 1967).

[2] J. R. Mika, Nucl. Sci. Eng . **11**, 415 (1961).

[3] G. J. Mitsis, Nucl. Sci. Eng . **17**, 55 (1963).

[4] R. Zelazny, J. Math. Phys. **2**, 4 538 (1961).

[5] K. M. Case, *On the Boundary Value Problems of Linear Transport Theory* (The University of Michigan Press, Ann Arbor, 1967).

In the treatment of the boundary-value problems, the standard way of incorporating the boundary conditions is to convert the differential equation into an integral equation. With this end in mind, consider the Green's function for the above transport equation:

$$(1 + \boldsymbol{\Omega} \cdot \nabla)G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_0, \boldsymbol{\Omega}_0) = \frac{c}{4\pi}\int d\Omega' G(\mathbf{r}, \boldsymbol{\Omega}', \mathbf{r}_0, \boldsymbol{\Omega}_0) + \delta(\mathbf{r} - \mathbf{r}_0)\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0), \tag{2}$$

where $\boldsymbol{\Omega}_0$ is the direction of the monodirectional point source located at $\mathbf{r}_0$ and $\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0)$ is the surface δ function defined in the usual manner; i.e.,

$$\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0) = 0, \quad \boldsymbol{\Omega} \neq \boldsymbol{\Omega}_0, \tag{3a}$$

$$\int d\boldsymbol{\Omega} f(\boldsymbol{\Omega})\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0) = f(\boldsymbol{\Omega}_0), \tag{3b}$$

if somewhere in the domain of integration $\boldsymbol{\Omega}_0 = \boldsymbol{\Omega}$. By construction, the solution[5] of Eq. (1) is

$$\psi(\mathbf{r}, \boldsymbol{\Omega}) = \int d\Omega'\, d^3r' G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}', \boldsymbol{\Omega}')Q(\mathbf{r}', \boldsymbol{\Omega}') + \int dS\, d\Omega_s G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_s, \boldsymbol{\Omega}_s)\hat{\mathbf{n}}(r_s) \cdot \boldsymbol{\Omega}_s \psi(\mathbf{r}_s, \boldsymbol{\Omega}_s), \tag{3c}$$

where $\mathbf{r}_s$ and $\boldsymbol{\Omega}_s$ are position and velocity vectors of the neutron at the boundary surface, respectively, and $\hat{\mathbf{n}}(\mathbf{r}_s)$ is the corresponding normal pointing toward the region where the solution of the transport equation is being sought.

Let us now construct the Green's function by taking the Fourier transform of Eq. (2) with respect to $\mathbf{r}$; i.e., let

$$G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_0, \boldsymbol{\Omega}_0) = \frac{1}{(2\pi)^3}\int d^3k \exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_0)]G_k(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0). \tag{4}$$

Equation (2) then becomes

$$G_k(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0) = \frac{1}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}}\frac{c}{4\pi}\int d\Omega' G_k(\boldsymbol{\Omega}', \boldsymbol{\Omega}_0) + \frac{\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0)}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}}. \tag{5}$$

By integrating both sides of Eq. (5) with respect to $\boldsymbol{\Omega}$, we obtain

$$\int d\Omega G_k(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0) = \frac{1}{\Lambda(k) \cdot (1 + i\mathbf{k} \cdot \boldsymbol{\Omega})}, \tag{6}$$

where

$$\Lambda(k) = 1 - \frac{c}{4\pi}\int \frac{d\Omega}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}} \tag{7}$$

is the familiar dispersion function.

Substituting the integral in Eq. (5) by the expression (6), we get

$$G_k(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0) = [\Lambda(1 + i\mathbf{k} \cdot \boldsymbol{\Omega})(1 + i\mathbf{k} \cdot \boldsymbol{\Omega}_0)]^{-1} + \delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0)[1 + i\mathbf{k} \cdot \boldsymbol{\Omega}]^{-1}.$$

The Green's function is then simply given by

$$G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}_0, \boldsymbol{\Omega}_0) = \frac{1}{(2\pi)^3}\int d^3k \exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_0)] \times \left[\frac{c}{4\pi\Lambda}[(1 + i\mathbf{k} \cdot \boldsymbol{\Omega})(1 + i\mathbf{k} \cdot \boldsymbol{\Omega}_0)]^{-1} + \frac{\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_0)}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}}\right]. \tag{8}$$

This is the fundamental Green's function which will serve to determine the solution of the one-speed transport equation for any given source and any incident distribution. We, therefore, turn to Eq. (3), which represents such a solution, and cast it in a more useable form.

Let us introduce the explicit expression for G [as given by Eq. (8)] in Eq. (3c) and rearrange the terms to obtain

$$\psi(\mathbf{r}, \boldsymbol{\Omega}) = \psi_q(\mathbf{r}, \boldsymbol{\Omega}) + \psi_0(\mathbf{r}, \boldsymbol{\Omega}) + \psi_c(\mathbf{r}, \boldsymbol{\Omega}), \tag{9}$$

where

$$\psi_q(\mathbf{r}, \boldsymbol{\Omega}) = \int d\Omega'\, d^3r' G(\mathbf{r}, \boldsymbol{\Omega}; \mathbf{r}', \boldsymbol{\Omega}')Q(\mathbf{r}', \boldsymbol{\Omega}') \tag{10}$$

is a known function,

$$\psi_0(\mathbf{r}, \boldsymbol{\Omega}) = \frac{1}{4\pi^2}\int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}} H, \tag{11}$$

$$H = \frac{1}{2\pi}\int dS\, d\Omega_s \hat{\mathbf{n}}(\mathbf{r}_s) \cdot \boldsymbol{\Omega}_s \psi(\mathbf{r}_s, \boldsymbol{\Omega}_s)\delta(\boldsymbol{\Omega} \cdot \boldsymbol{\Omega}_s)e^{-i\mathbf{k}\cdot\mathbf{r}_s} = \frac{1}{2\pi}\int dS \hat{\mathbf{n}}(\mathbf{r}_s) \cdot \boldsymbol{\Omega}_s \psi(\mathbf{r}_s, \boldsymbol{\Omega})e^{-i\mathbf{k}\cdot\mathbf{r}_s}, \tag{12}$$

$$\psi_c(\mathbf{r}, \boldsymbol{\Omega}) = \frac{c}{8\pi^2}\int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}}\frac{T}{\Lambda}, \tag{13}$$

and

$$T = \frac{1}{4\pi^2}\int dS\, d\Omega_s \hat{\mathbf{n}}(\mathbf{r}_s) \cdot \boldsymbol{\Omega}_s \psi(\mathbf{r}_s, \boldsymbol{\Omega}_s)\frac{e^{-i\mathbf{k}\cdot\mathbf{r}_s}}{1 + i\mathbf{k} \cdot \boldsymbol{\Omega}_s}. \tag{14}$$

From Eq. (9) we see that the solution of any boundary-value problem is known, provided we can find the surface distribution $\psi(\mathbf{r}_s, \boldsymbol{\Omega})$ or, equivalently, the coefficients T and H. Of course, if $\psi(\mathbf{r}_s, \boldsymbol{\Omega})$ is known *a priori*, then we are done. However, this is not always possible, for, in most instances, we only know

either the incident or the outgoing distribution, but not both. It is, therefore, necessary to supplement Eq. (9) with another equation which determines $\psi(\mathbf{r}_s, \mathbf{\Omega})$. This is easily done by passing to the limit [in Eq. (9)] as $\mathbf{r} \to \mathbf{r}_s$ from within the region of interest; i.e.,

$$\psi(\mathbf{r}_s, \mathbf{\Omega}) = \psi_q(\mathbf{r}_s, \mathbf{\Omega}) + \psi_0(\mathbf{r}_s, \mathbf{\Omega}) + \psi_c(\mathbf{r}_s, \mathbf{\Omega}). \quad (15)$$

For further discussion of the equation that determines the surface distribution, see Ref. 5.

The rest of this paper is devoted to the study of the integral equation (9) in conjunction with Eq. (15), subject to various boundary conditions in spherically symmetric problems. Specifically, we shall consider two categories of problems, the interior and the exterior of a sphere.

3. SPHERICALLY SYMMETRIC PROBLEMS

By this we mean that the angular density ψ will be a function of r and $\mu = \hat{\mathbf{r}} \cdot \mathbf{\Omega}$, with $\mu > 0$ corresponding to the outgoing neutrons and $\mu < 0$ to the incoming ones. Under the spherical symmetry, the integral equation (9) is considerably simplified by carrying out the appropriate angular integrations. Thus, noting that T, given by Eq. (14), is a function of the magnitude of k, we can write

$$\psi_c(r, \mu) = \frac{c}{8\pi^2} \int_0^\infty dk k^2 \frac{T(k)}{\Lambda(k)} \int_0^{2\pi} d\varphi \int_{-1}^1 \frac{dt}{1 + ikt} \times \exp\left(ik\{t\mu + [(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}} \cos(\varphi - \varphi_k)\}\right),$$

or

$$\psi_c(r, \mu) = \frac{c}{8\pi} \int_{-\infty}^\infty dk k^2 I(k, r, \mu) \frac{T(k)}{\Lambda(k)}, \quad (16)$$

where

$$I(k, r, \mu) = \int_{-1}^1 \frac{dt}{1 + ikt} e^{iktr\mu} J_0\{kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\}. \quad (17)$$

Similarly,

$$\psi_0(r, \mu) = \frac{1}{4\pi} \int_{-\infty}^\infty dk k^2 \int_{-1}^1 \frac{dt}{1 + ikt} e^{iktr\mu} \times J_0\{kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\} H(k, t), \quad (18)$$

where

$$H(k, t) = \int_{-1}^1 d\mu' \mu' \psi(r_s, \mu') e^{-iktr_s\mu'} \times J_0\{kr_s[(1 - \mu'^2)(1 - t^2)]^{\frac{1}{2}}\} \quad (19)$$

and r_s denotes the radius of the sphere. The equation that determines the angular density is then

$$\psi(r, \mu) = \psi_q(r, \mu) + \psi_0(r, \mu) + \psi_c(r, \mu). \quad (20)$$

Before we express $\psi(r, \mu)$ (for interior and exterior problems) in terms of eigenfunctions of the transport equation, let us examine the analytical properties of $\Lambda(k)$, $I(k, r, u)$, and $T(k)$ in the complex k plane. From Eq. (7) it is clear that Λ is sectionally holomorphic in the k plane with the branch cuts extending from $-i\infty$ to $-i$ and i to $i\infty$. This property is shared by the functions $I(k, r, \mu)$ and $T(k)$. (We shall use this fact in constructing the eigenfunctions of the transport equation.) The zeros of $\Lambda(k)$ are either purely real or purely imaginary, depending on whether $c < 1$ or $c > 1$. It may seem peculiar, at first, that the angular density [or more appropriately, the Green's function given by Eq. (8)] is not uniquely determined when the zeros of Λ are real. However, we will show later (when we deal with the critical problem) that it is *not necessary* to prescribe any one particular recipe for treating the real zeros of Λ; that is, all prescriptions lead to a unique determination of the angular density. Finally, we note that, for complex values of k, the functions $I(k, r, \mu)$, $T(k)$, and $H(k, t)$ diverge at infinity. However, we show in the following sections that these functions can always be written as a sum of two, one of which converges in the upper half k plane and the other in the lower half.

A. Interior Problems

In Eq. (20), let us first consider $\psi_c(r, \mu)$ given by Eq. (16). We wish to express $\psi_c(r, \mu)$ in terms of eigenfunctions. To do that, we need to change the path of integration from the real axis to the contour surrounding the cut, as shown in Fig. 1. Since $r < r_s$, it is necessary to decompose $T(k)$ only. The decomposition is readily obtained by expanding the exponential in Eq. (14) in terms of spherical harmonics. Thus,

$$e^{-i\mathbf{k}\cdot\mathbf{r}_s} = 4\pi \sum_{n,m} i^n j_n(kr_s) Y^*_{nm}(\hat{\mathbf{k}} \cdot \mathbf{\Omega}_s) Y_{nm}(-\hat{\mathbf{r}}_s \cdot \mathbf{\Omega}_s). \quad (21)$$

The expression for $T(k)$ now becomes

$$T(k) = \frac{1}{\pi} \int dS\, d\Omega_s \hat{\mathbf{n}}(\mathbf{r}_s) \cdot \mathbf{\Omega}_s \frac{\psi(\mathbf{r}_s, \mu_s)}{1 + i\mathbf{k} \cdot \mathbf{\Omega}_s} \times \sum_{n,m} i^n j_n(kr_s) Y^*_{nm}(\hat{\mathbf{k}} \cdot \mathbf{\Omega}_s) Y_{nm}(-\hat{\mathbf{r}}_s \cdot \mathbf{\Omega}_s).$$

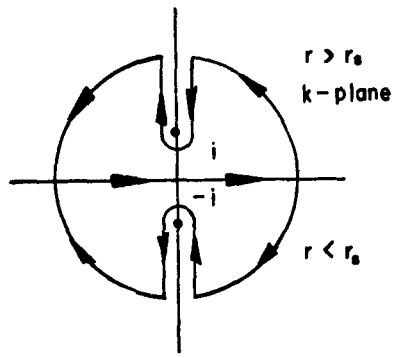


FIG. 1. Contours for $r > r_s$ and $r < r_s$.

Because of the azimuthal symmetry of the surface angular density $\psi(r_s, \mu_s)$, the only term in the sum over m which is nonzero is that for which $m = 0$. Hence,

$$T(k) = \int_{-1}^{1} d\mu_s \mu_s \psi(r_s, \mu_s) \times \sum_{n=0}^{\infty} i^n (2n + 1) j_n(kr_s) P_n(-\mu_s) Q_n(k), \quad (22)$$

where $Q_n(k)$ is the Legendre function of the second kind defined by

$$Q_n(k) = \int_{-1}^{1} dt \frac{P_n(t)}{1 + ikt}. \quad (23)$$

Now the spherical Besssel function $j_n(z)$ can be expressed as a sum of Hankel functions of the first and the second kind. In Sommerfeld's notation,[6]

$$2j_n(z) = \zeta_n^{(1)}(z) + \zeta_n^{(2)}(z). \quad (24)$$

Putting (24) into (23), we get the following decomposition for $T(k)$:

$$T(k) = \tfrac{1}{2}[T_1(k) + T_2(k)], \quad (25)$$

where

$$T_{1,2}(k) = \int_{-1}^{1} d\mu_s \mu_s \psi(r_s, \mu_s) \zeta^{(1),(2)}(k, r_s, -\mu_s) \quad (26)$$

and

$$\zeta^{(1),(2)}(k, r_s, -\mu_s) = \sum_{n=0}^{\infty} i^n (2n + 1) \zeta_n^{(1),(2)}(kr_s) P_n(-\mu_s) Q_n(k). \quad (27)$$

Let us note that

$$T_2(-k) = T_1(k). \quad (28)$$

The expression (16) for $\psi_c(r, \mu)$ may now be rewritten as a sum of two integrals by inserting the decomposition of $T(k)$ given by Eq. (25). One may, then, readily show that one of the integrals converges in the upper half k plane and the other in the lower half. To simplify the subsequent calculations, we use the relation (28) and write

$$\psi_c(r, \mu) = \frac{c}{8\pi} \int_{-\infty}^{\infty} dk k^2 \frac{T_1(k)}{\Lambda} I(k, r, \mu). \quad (29)$$

In much the same way, the expression for $\psi_0(r, \mu)$ [see Eq. (18)] may be cast in the following form:

$$\psi_0(r, \mu) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dk k^2 \int_{-1}^{1} \frac{dt}{1 + ikt} e^{iktr\mu} \times J_0\{kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\} H_1(k, t), \quad (30)$$

where

$$H_1(k, t) = \int_{-1}^{1} d\mu' \mu' \psi(r_s, \mu') \times \sum_{n=0}^{\infty} i^n (2n + 1) \zeta_n^{(1)}(kr_s) P_n(-\mu') P_n(t). \quad (31)$$

Now we can change the path of integration from the real axis to a contour surrounding the cut in the upper half k plane. For $\psi_c(r, \mu)$, we have

$$\psi_c(r, \mu) = \frac{c}{8\pi} \int_i^{i\infty} dk k^2 \left(\frac{I^- T_1^-}{\Lambda^-} - \frac{I^+ T_1^+}{\Lambda^+} \right) + \text{d. c.} \quad (32)$$

where $-$ $(+)$ denotes the boundary value as we approach the cut from the right(left), and $D \cdot C$ is the discrete contribution arising from the zero of Λ. For the present we have assumed that $c < 1$; i.e., the zeros of $\Lambda(k)$ are purely imaginary. Let us simplify the integrand by constructing the eigenfunctions. This we do as follows. First we note that, from Plemelj's formula, we have

$$I^{\pm} = \mathcal{I} I \pm i\pi I_\delta, \quad (33)$$

where

$$\mathcal{I} I = \mathcal{I} \int_{-1}^{1} dt \frac{e^{iktr\mu}}{1 + ikt} J_0\{kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\}, \quad (34a)$$

$$I_\delta = \int_{-1}^{1} dt e^{iktr\mu} J_0\{kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\} \delta(1 + ikt) = \frac{i}{k} e^{-r\mu} J_0\{r[(1 - \mu^2)(k^2 + 1)]^{\frac{1}{2}}\}, \quad (34b)$$

and k is purely imaginary. Thus,

$$A = (I^- T_1^-/\Lambda^-) - (I^+ T_1^+/\Lambda^+) = (T_1^-/\Lambda^-)(\mathcal{I} I + i\pi I_\delta) - (T_1^+/\Lambda^+)(\mathcal{I} I - i\pi I_\delta).$$

To construct the eigenfunctions, we introduce $\lambda(k)$ in A as follows:

$$A = (T_1^-/\Lambda^-)[\mathcal{I} I + \lambda I_\delta - (\lambda - i\pi) I_\delta] - (T_1^+/\Lambda^+)[\mathcal{I} I + \lambda I_\delta - (\lambda + i\pi) I_\delta].$$

Now choose λ such that

$$\phi(k, r, \mu) = \mathcal{I} I + \lambda I_\delta \quad (35)$$

are spherical eigenfunctions of the transport equation with continuous spectrum. By straightforward calculations, one may readily show that the appropriate λ is

$$\lambda = i\pi(\Lambda^+ + \Lambda^-)/(\Lambda^+ - \Lambda^-). \quad (36)$$

If we choose λ in this way, then A becomes

$$A = [(T_1^-/\Lambda^-) - (T_1^+/\Lambda^+)]\phi(k, r, \mu) + 2\pi i(T_1^- - T_1^+/\Lambda^- - \Lambda^+) I_\delta.$$

[6] A. Sommerfeld, *Lectures on Theoretical Physics, Vol. 6: Partial Differential Equations in Physics* (Academic Press, New York, 1949).

Equation (32) for $\psi_c(r, \mu)$ now becomes

$$\psi_c(r, \mu) = \int_i^{i\infty} \frac{dk}{k} \phi(k, r, \mu)\Gamma_<(k) + \frac{ic}{4}\int_i^{i\infty} dk k^2 \frac{T_1^- - T_1^+}{\Lambda^- - \Lambda^+} I_\delta + \phi(ik_0, r, \mu)\Gamma^0_<(k_0), \quad (37)$$

where

$$\Gamma_<(k) = \frac{k^3 c}{8\pi}\left(\frac{T_1^-}{\Lambda^-} - \frac{T_1^+}{\Lambda^+}\right), \quad (38)$$

and we have written down the explicit expression for the discrete contribution in which

$$\Gamma^0_<(k_0) = -\frac{ic}{4} k_0^2 \frac{T_1(ik_0)}{\Lambda'(ik_0)}, \quad (39)$$

and

$$\phi(ik_0, r, \mu) = \int_{-1}^{1} \frac{dt}{1 - k_0 t} e^{-k_0 r\mu} I_0\{k_0 r[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\} \quad (40)$$

is the discrete eigenfunction of the transport equation with the eigenvalue ik_0 the zero of $\Lambda(k)$; i.e.,

$$\Lambda(ik_0) = 0, \quad 0 < k_0 < 1.$$

For $\psi_0(r, \mu)$ [see Eq. (30)], we get

$$\psi_0(r, \mu) = \frac{2\pi i}{4\pi}\int_i^{i\infty} dk k^2 \int_{-1}^{1} dt \times e^{ikt r\mu} J_0\{kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}\} \times \delta(1 + ikt) H(k, t),$$

or

$$\psi_0(r, \mu) = -\tfrac{1}{2} e^{-r\mu}\int_i^{i\infty} dk k J_0\{r[(1 - \mu^2)(k^2 + 1)]^{\frac{1}{2}}\} \times H\left(k, -\frac{1}{ik}\right). \quad (41)$$

Now one may easily show that $\psi_0(r, \mu)$, as given by Eq. (41), is equal to minus the second term on the right-hand side of Eq. (37). With this in mind, Eq. (20), for the angular density $\psi(r, \mu)$, becomes

$$\psi(r, \mu) = \psi_q(r, \mu) + \int_0^1 d\nu \frac{1}{\nu}\Gamma_<\left(\frac{i}{\nu}\right)\phi\left(\frac{i}{\nu}, r, \mu\right) + \phi\left(\frac{i}{\nu_0}, r, \mu\right)\Gamma^0_<\left(\frac{1}{\nu_0}\right), \quad (42)$$

where we have set $k = i/\nu$ and $k_0 = i/\nu_0$.

An equation that determines the coefficients $\Gamma_<$ and $\Gamma^0_<$ is obtained by letting $r \to r_s$ in Eq. (42). Thus,

$$\psi(r_s, \mu) = \psi_q(r_s, \mu) + \int_0^1 d\nu \frac{1}{\nu}\Gamma_<\left(\frac{i}{\nu}\right)\phi\left(\frac{i}{\nu}, r_s, \mu\right) + \phi\left(\frac{i}{\nu_0}, r_s, \mu\right)\Gamma^0_<\left(\frac{1}{\nu_0}\right). \quad (43)$$

This is a regular integral equation. Its solution is difficult to discuss. Therefore, we shall seek the help of an operator (the reduction operator) which, when applied to Eq. (20), produces an auxiliary equation with the same coefficients as that in Eq. (43), but with a singular kernel. Before we present such an operator, let us first consider the class of exterior problems.

B. Exterior Problems

Now since $r > r_s$, in order to express $\psi_c(r, \mu)$ [see Eq. (16)] in terms of eigenfunctions by the change of path of integration to a contour surrounding the cut in the upper half plane, we need to decompose $I(k, r, \mu)$ [see Eq. (17)] and then follow the same procedure as that in the interior problems. Thus, to decompose $I(k, r, \mu)$, we write

$$e^{ikt r\mu} J_0(kr[(1 - \mu^2)(1 - t^2)]^{\frac{1}{2}}) = \tfrac{1}{2}\sum_{n=0}^{\infty} i^n(2n + 1)[\zeta_n^{(1)}(kr) + \zeta_n^{(2)}(kr)]P_n(\mu)P_n(t), \quad (44)$$

where we have used the decomposition (24) of the spherical Bessel function. Then,

$$I(k, r, \mu) = \tfrac{1}{2}[\zeta^{(1)}(k, r, \mu) + \zeta^{(2)}(k, r, \mu)], \quad (45)$$

where $\zeta^{(1),(2)}(k, r, \mu)$ is given by (27) with r_s replaced by r and $-\mu_s$ by μ. Equation (16) for $\psi_c(r, \mu)$ now becomes

$$\psi_c(r, \mu) = \frac{c}{8\pi}\int_{-\infty}^{\infty} dk k^2 \zeta^{(1)}(k, r, \mu)\frac{T(k)}{\Lambda(k)}. \quad (46)$$

The resulting integral equation for $\psi(r, \mu)$ is given by

$$\psi(r, \mu) = \psi_q(r, \mu) + \int_0^1 d\nu(1/\nu)\Gamma_>(i/\nu)Z(i/\nu, r, \mu) + Z(i/\nu_0, r, \mu)\Gamma^0_>(1/\nu_0). \quad (47)$$

An equation that determines the coefficients $\Gamma_>(i/\nu)$ and $\Gamma^0_>(1/\nu_0)$ is

$$\psi(r_s, \mu_s) = \psi_q(r_s, \mu) + \int_0^1 d\nu(1/\nu)\Gamma_>(i/\nu)Z(i/\nu, r_s, \mu) + Z(i/\nu_0, r_s, \mu)\Gamma^0_>(1/\nu_0), \quad (48)$$

where

$$\Gamma_>(i/\nu) = \frac{c}{8\pi i\nu^3}\left[\frac{T^-(i/\nu)}{\Lambda^-(i/\nu)} - \frac{T^+(i/\nu)}{\Lambda^+(i/\nu)}\right], \quad (49)$$

$$Z(i/\nu, r, \mu) = \sum_{n=0}^{\infty} i^n(2n + 1)\zeta_n^{(1)}(ir/\nu)P_n(\mu)\left[\mathcal{P}\int_{-1}^{1}\frac{dt P_n(t)\nu}{\nu - t} + \nu P_n(\nu) i\pi\frac{\Lambda^+ + \Lambda^-}{\Lambda^+ - \Lambda^-}\right], \quad (50)$$

and

$$\Gamma^0_>(1/\nu_0) = \frac{-ic}{4\nu_0^2}\frac{T(i/\nu_0)}{\Lambda'(i/\nu_0)}. \quad (51)$$

In Eq. (47), the regular eigenfunctions occur implicitly in the coefficients $\Gamma_>(i/\nu)$ and $\Gamma^0_>(1/\nu_0)$. Thus, from Eq. (22) for $T(k)$, we have

$$T(k) = \int_{-1}^{1} d\mu_s \mu_s \psi(r_s, \mu_s) I(k, r_s, -\mu_s).$$

Therefore,

$$\frac{T^-(k)}{\Lambda^-} - \frac{T^+(k)}{\Lambda^+}$$
$$= \int_{-1}^{1} d\mu_s \mu_s \psi(r_s, \mu_s) \left[\frac{I^-(k, r_s, -\mu_s)}{\Lambda^-} - \frac{I^+(k, r_s, -\mu_s)}{\Lambda^+} \right],$$

where k is a purely imaginary number (i.e., $k = i/\nu$). Now, by definition,

$$[I^-(k, r_s, -\mu_s)/\Lambda^-] - [I^+(k, r_s, -\mu_s)/\Lambda^+]$$
$$= (1/\Lambda^-)[\mathcal{P}I(k, r_s, -\mu_s) + i\pi I_\delta(k, r_s, -\mu_s)]$$
$$- (1/\Lambda^+)[\mathcal{P}I(k, r_s, -\mu_s) - i\pi I_\delta(k, r_s, -\mu_s)]$$
$$= (1/\Lambda^+\Lambda^-)\{\mathcal{P}I(k, r_s, -\mu_s)$$
$$+ i\pi[(\Lambda^+ + \Lambda^-)/(\Lambda^+ - \Lambda^-)] I_\delta(k, r_s, -\mu_s)\}$$
$$= (1/\Lambda^+\Lambda^-)\phi(k, r_s, -\mu_s).$$

We have already mentioned that $\phi(k, r, \mu)$, as given by Eq. (35), are the regular eigenfunctions of the transport equation with continuous spectrum. Here,

$$\phi(k, r_s, -\mu_s) = \mathcal{P}I(k, r_s, -\mu_s)$$
$$+ i\pi[(\Lambda^+ + \Lambda^-)/(\Lambda^+ - \Lambda^-)] I_\delta(k, r_s, -\mu_s) \quad (52)$$

are the regular eigenfunctions of the adjoint equation with the same continuous spectrum. Thus, explicitly,

$$\Gamma_>\left(\frac{i}{\nu}\right) = \frac{c}{8\pi i \nu^3} \int_{-1}^{1} d\mu_s \mu_s \psi(r_s, \mu_s) \frac{1}{\Lambda^+\Lambda^-} \phi\left(\frac{i}{\nu}, r_s, -\mu_s\right). \quad (53)$$

Similarly,

$$\Gamma^0_>\left(\frac{1}{\nu_0}\right) = \frac{-ic}{4\nu_0^2} \frac{1}{\Lambda'(i/\nu_0)}$$
$$\times \int_{-1}^{1} d\mu_s \mu_s \psi(r_s, \mu_s) \phi\left(\frac{i}{\nu_0}, r_s, -\mu_s\right), \quad (54)$$

where

$$\phi\left(\frac{i}{\nu_0}, r_s, \mu_s\right)$$
$$= \nu_0 \int_{-1}^{1} \frac{dt}{\nu_0 - t} e^{r_s\mu_s/\nu_0} I_0\left\{\frac{r_s}{\nu_0}[(1 - \mu_s^2)(1 - t^2)]^{\frac{1}{2}}\right\} \quad (55)$$

is the regular discrete eigenfunction of the adjoint equation with the point spectrum ν_0 ($ik_0 = i/\nu_0$ being the zero of Λ).

To solve the integral equation (48) for the coefficients, we shall seek the help of the reduction operator. In the next section, we present such an operator.

4. THE REDUCTION OPERATOR

Let us consider an operator θ, given by

$$\theta \equiv \lim_{r \to r_s} \int_{-\infty}^{\infty} dr' r' K(r - r', \mu)$$
$$\times \left(1 + \mu \frac{\partial}{\partial r'} + \frac{1 - \mu^2}{r'} \frac{\partial}{\partial \mu}\right), \quad (56)$$

where the kernel $K(r - r', \mu)$ is

$$K(r - r', \mu) = (1/\mu) e^{-(r-r')/\mu} \Theta(r - r') \Theta(\mu)$$
$$- (1/\mu) e^{-(r-r')/\mu} \Theta(r' - r) \Theta(-\mu). \quad (57)$$

Let us write the operator θ, formally, as

$$\theta \equiv \lim_{r \to r_s} KS, \quad (58)$$

where

$$S \equiv 1 + \mu \frac{\partial}{\partial r} + \frac{1 - \mu^2}{r} \frac{\partial}{\partial \mu}$$

is the streaming part of the transport operator.

A. Application of the Operator θ to Interior Problems

Owing to the linearity of the operator θ, its application to Eq. (20) gives us

$$\theta\psi(r, \mu) = \theta\psi_q(r, \mu) + \theta\psi_0(r, \mu) + \theta\psi_c(r, \mu). \quad (59)$$

First, consider the left-hand side of Eq. (59). Since θ is a product of two operators [Eq. (58)], K and S, let us apply S first. Thus,

$$S\psi(r, \mu)\Theta(r_s - r)$$
$$= \psi(r, \mu)\Theta(r_s - r) + \mu\Theta(r_s - r)\frac{\partial \psi}{\partial r}$$
$$- \mu\psi\delta(r_s - r) + \frac{1 - \mu^2}{r} \Theta(r_s - r) \frac{\partial \psi}{\partial \mu}.$$

For $\mu < 0$,

$$\theta[\psi(r, \mu)\Theta(r_s - r)] = r_s\psi(r_s, \mu). \quad (60)$$

Next consider $\theta\psi_c$. We have, from Eq. (16),

$$\theta\psi_c(r, \mu) = \frac{c}{8\pi} \int_{-\infty}^{\infty} dk k^2 \frac{T(k)}{\Lambda(k)} [\theta I(k, r, \mu)]. \quad (61)$$

Now,

$$SI(k, r, \mu) = (2 \sin kr)/kr.$$

Therefore, for $\mu < 0$,

$$KSI(k, r, \mu) = -\frac{2}{r} \int_{-\infty}^{\infty} dr' r' e^{-(r-r')/\mu} \Theta(r' - r) \frac{\sin kr'}{kr'}$$

or

$$KSI(k, r, \mu) = \frac{i}{k}\left[-\frac{e^{ikr}}{1 + ik\mu} + \frac{e^{-ikr}}{1 - ik\mu}\right]. \quad (62)$$

If we put Eq. (62) into Eq. (60), the result is

$$\theta\psi_c(r,\mu) = \frac{ic}{8\pi}\lim_{r\to r_s}\int_{-\infty}^{\infty} dkk\,\frac{T(k)}{\Lambda(k)} \times\left[-\frac{e^{ikr}}{1+ik\mu}+\frac{e^{-ikr}}{1-ik\mu}\right]. \quad (63)$$

Now, decomposing $T(k)$ into the sum of $T_1(k)$ and $T_2(k)$ as before [see Eq. (25)] and following the same procedure of integration around the cut in the upper half k plane, we obtain

$$\begin{aligned}\theta\psi_c(r,\mu) = {}& \frac{ic}{8\pi}\int_i^{i\infty} dkk[\phi^0(-k,r_s,\mu)-\phi(k,r_s,\mu)] \\ & \times\left(\frac{T_1^-}{\Lambda^-}-\frac{T_1^+}{\Lambda^+}\right) \\ & -\frac{c}{4}\int_i^{i\infty} dkk[e^{-ikr_s}\delta(1-ik\mu) \\ & -e^{ikr_s}\delta(1+ik\mu)]\left(\frac{T_1^- - T_1^+}{\Lambda^- - \Lambda^+}\right)+\text{|d.c.}\end{aligned} \quad (64)$$

where

$$\begin{aligned}\phi^0(k,r_s,\mu) = {}& e^{ikr_s}\{\mathcal{P}(1+ik\mu)^{-1} \\ & + i\pi[(\Lambda^+ + \Lambda^-)/(\Lambda^+ - \Lambda^-)]\delta(1+ik\mu)\}\end{aligned} \quad (65)$$

are the planar eigenfunctions corresponding to the continuous k spectrum. Also, for $\mu < 0$, these eigenfunctions are regular. However, if the sign of k is reversed, then these planar eigenfunctions become irregular.

In Eq. (64), let us introduce the coefficients $\Gamma_<$ and $\Gamma_<^0$ as given by Eqs. (38) and (39), respectively. Thus,

$$\begin{aligned}\theta\psi_c(r,\mu) = {}& i\int_i^{i\infty}\frac{dk}{k^2} \\ & \times[\phi^0(-k,r_s,\mu)-\phi^0(k,r_s,\mu)]\Gamma_<(k) \\ & -\frac{c}{4}\int_i^{i\infty} dkk\,\frac{T_1^- - T_1^+}{\Lambda^- - \Lambda^+} \\ & \times[e^{-ikr_s}\delta(1-ik\mu)-e^{ikr_s}\delta(1+ik\mu)] \\ & +\left(\frac{1}{k_0}\right)[\phi^0(-ik_0,r_s,\mu) \\ & -\phi^0(ik_0,r_s,\mu)]\Gamma_<(k_0),\end{aligned} \quad (66)$$

where

$$\phi^0(ik_0,r_s,\mu) = e^{-ik_0r_s}/(1-k_0\mu), \quad 0<k_0<1, \quad (67)$$

is the planar discrete eigenfunction corresponding to the point spectrum k_0.

One may readily show again that the second term on the right-hand side of Eq. (66) is equal to $-\theta\psi_0(r,\mu)$. Consequently, the reduced angular density given by Eq. (59) may now be written as follows:

$$\begin{aligned}r_s\psi(r_s,\mu) = {}& \theta\psi_0(r,\mu)\int_0^1 d\nu\left[\phi^0\left(-\frac{i}{\nu},r_s,\mu\right)\right. \\ & \left.-\phi^0\left(\frac{i}{\nu},r_s,\mu\right)\right]\Gamma_<\left(\frac{i}{\nu}\right) \\ & +\nu_0\left[\phi^0\left(-\frac{i}{\nu_0},r_s,\mu\right)-\phi^0\left(\frac{i}{\nu_0},r_s,\mu\right)\right] \\ & \times\Gamma_<^0\left(\frac{i}{\nu_0}\right), \quad \mu<0.\end{aligned} \quad (68)$$

Putting the explicit forms of the planar eigenfunctions in Eq. (67), we re-express this equation in the standard notation.[1] Thus,

$$\begin{aligned}&\frac{1}{2}\left[\Lambda^+\left(-\frac{i}{\mu}\right)+\Lambda^-\left(-\frac{i}{\mu}\right)\right]\tilde{\Gamma}_<\left(-\frac{i}{\mu}\right) \\ &\quad+\frac{1}{2\pi i}\left[\Lambda^+\left(-\frac{i}{\mu}\right)-\Lambda^-\left(-\frac{i}{\mu}\right)\right]\mathcal{P}\int_0^1\frac{d\nu}{\nu+\mu}\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) \\ &\quad=\left[\Lambda^+\left(-\frac{i}{\nu}\right)-\Lambda^-\left(-\frac{i}{\nu}\right)\right]f(\mu), \quad \mu<0,\end{aligned} \quad (69)$$

where

$$\begin{aligned}f(\mu) = {}& \frac{1}{2\pi i}\left[\nu_0^2\Gamma_<^0\left(\frac{1}{\nu_0}\right)\left(\frac{e^{-r_s/\nu_0}}{\nu_0-\mu}-\frac{e^{r_s/\nu_0}}{\nu_0+\mu}\right)\right. \\ & +\int_0^1\frac{d\nu}{\nu-\mu}e^{-2r_s/\nu}\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) \\ & \left.+r_s\psi(r_s,\mu)-\theta\psi_0(r,\mu)\right]\end{aligned} \quad (70)$$

and

$$\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) = \nu e^{r_s/\nu}\Gamma_<\left(\frac{i}{\nu}\right). \quad (71)$$

Equation (69) is the auxiliary equation which can be solved for the coefficients $\tilde{\Gamma}_<$ and $\Gamma_<^0$ by the standard technique developed for planar problems.[1] We note that the existence of such a solution guarantees the completeness of the regular spherical eigenfunctions given by Eqs. (35) and (40) in the sense stated in the following theorem.

Theorem 1: "Any" function $\psi(r,\mu)$ in the domain $-\mu<1$ and regular at $r=0$ may be expanded in terms of the regular spherical eigenfunctions $\phi(i/\nu, r, \mu)$ and $\phi^0(i/\nu_0, r, \mu)$, corresponding to the continuous spectrum $0<\nu<1$ and the point spectrum ν_0.

Let us mention that this is the half-range completeness of the eigenfunctions. The proof of the theorem is demonstrated by constructing the coefficients $\Gamma_<$ and $\Gamma_<^0$ from Eq. (69). For further details on this we refer the reader to Ref. 1. Here we shall merely use the

results of such a solution in various specific boundary-value problems. Before we do that, however, let us complete our presentation by giving the analog of Eq. (69) for the class of exterior problems.

B. Application of the Operator θ to Exterior Problems

Consider the left-hand side of Eq. (59). For $r > r_s$, we have

$$S\{\psi(r,\mu)\Theta(r-r_s)\} = \psi(r,\mu)\Theta(r-r_s) + \mu\Theta(r-r_s)\frac{\partial\psi}{\partial r} + \mu\psi\delta(r-r_s) + \frac{1-\mu^2}{r}\Theta(r-r_s)\frac{\partial\psi}{\partial\mu}.$$

For $\mu > 0$, we get

$$\theta\{\psi(r,\mu)\Theta(r-r_s)\} = r_s\psi(r_s,\mu). \tag{72}$$

In the right-hand side of Eq. (59) consider $\theta\psi_c$:

$$\theta\psi_c(r,\mu) = \frac{c}{8\pi}\int_{-\infty}^{\infty}\frac{dk}{k^2}\frac{T(k)}{\Lambda(k)}[\theta I(k,r,\mu)].$$

For $\mu > 0$, we have

$$KSI(k,r,\mu) = K\frac{2\sin kr}{kr} = \frac{2}{k\mu}\int_{-\infty}^{r} dr' e^{-(r-r')/\mu}\sin kr' = \frac{1}{ki}\left[\frac{e^{ikr}}{1+ik\mu} - \frac{e^{-ikr}}{1-ik\mu}\right].$$

Therefore,

$$\theta\psi_c(r,\mu) = \frac{c}{8\pi i}\lim_{r\to r_s}\int_{-\infty}^{\infty} dkk\frac{T(k)}{\Lambda}\times\left[\frac{e^{ikr}}{1+ik\mu} - \frac{e^{-ikr}}{1-ik\mu}\right],$$

or

$$\theta\psi_c(r,\mu) = \frac{c}{4\pi i}\lim_{r\to r_s}\int_{-\infty}^{\infty} dkk\frac{T(k)}{\Lambda}\frac{e^{ikr}}{1+ik\mu}. \tag{73}$$

Again, by changing the path of integration to a contour surrounding the cut in the upper half k plane, we cast Eq. (73) into the form given below:

$$\theta\psi_c(r,\mu) = \frac{c}{4\pi i}\int_{i}^{i\infty} dkk\phi^0(k,r_s,\mu)\left(\frac{T^-}{\Lambda^-} - \frac{T^+}{\Lambda^+}\right) + \frac{c}{2}\int_{i}^{i\infty} dkke^{ikr_s}\delta(1+ik\mu)\times\frac{T^- - T^+}{\Lambda^- - \Lambda^+} + \text{d. c.} \tag{74}$$

where $\phi^0(k,r_s,\mu)$ is the planar eigenfunction represented by Eq. (65). Now let us introduce the coefficients $\Gamma_>$ and $\Gamma^0_>$ defined by Eqs. (50) and (51), respectively, and put the form of $\theta\psi_c$ given by Eq. (73) into Eq. (59). We obtain

$$r_s\psi(r_s,\mu) = \theta\psi_q(r,\mu) - 2\int_0^1 d\nu\Gamma_>\left(\frac{i}{\nu}\right)\phi^0\left(\frac{i}{\nu},r_s,\mu\right) - 2\nu_0\phi^0\left(\frac{i}{\nu_0},r_s,\mu\right)\Gamma^0_>\left(\frac{i}{\nu_0}\right), \tag{75}$$

where $\phi^0(i/\nu_0,r_s,\mu)$ is the discrete planar eigenfunction given by Eq. (67).

Let us cast Eq. (74) into a more usable form by introducing the explicit expressions for the eigenfunctions. Thus,

$$\frac{1}{2}\left[\Lambda^+\left(\frac{i}{\mu}\right) + \Lambda^-\left(\frac{i}{\mu}\right)\right]\tilde{\Gamma}_>\left(\frac{i}{\mu}\right) + \frac{1}{2\pi i}\left[\Lambda^+\left(\frac{i}{\mu}\right) - \Lambda^-\left(\frac{i}{\mu}\right)\right]\mathcal{S}\int_0^1\frac{d\nu}{\nu-\mu}\tilde{\Gamma}_>\left(\frac{i}{\nu}\right) = (\Lambda^+ - \Lambda^-)g(\mu), \quad \mu > 0, \tag{76}$$

where

$$g(\mu) = \frac{1}{4\pi i}\left[-\nu_0^2\Gamma^0_>\left(\frac{i}{\nu_0}\right)e^{-r_s/\nu_0}/(\nu_0-\mu) + \theta\psi_q(r,\mu) - r_s\psi(r_s,\mu)\right] \tag{77}$$

and

$$\tilde{\Gamma}_>(i/\mu) = \mu e^{-r_s/\mu}\Gamma_>(i/\mu). \tag{78}$$

In the next section, we consider some specific interior and exterior problems.

5. APPLICATIONS

In the integral equation (69), we notice that, except for the terms involving the incident distribution and the distribution due to source(s), the rest of the features are common to all interior problems. This is also true for Eq. (76) for the class of exterior problems. For this reason it is convenient to write down the most general solutions for the corresponding coefficients and treat just the distinguishing part separately for each problem. Thus, for the class of interior problems, the solution of the singular integral equation (69) is[1]

$$\tilde{\Gamma}_>\left(\frac{i}{\mu}\right) = -\frac{1}{2\pi i}\left[\frac{1}{X^+(-\mu)} - \frac{1}{X^-(-\mu)}\right]\left\{\nu_0^2\Gamma^0_>\left(\frac{1}{\nu_0}\right)\times\left[e^{r_s/\nu_0}\frac{X(\nu_0)}{\mu+\nu_0} - e^{-r_s/\nu_0}\frac{X(-\nu_0)}{\mu-\nu_0}\right] + \int_0^1 d\nu e^{-2r_s/\nu}\frac{\tilde{\Gamma}_<(i/\nu)}{\nu-\mu}X(-\nu) + \int_0^1 d\mu'\frac{X^+(\mu') - X^-(\mu')}{\mu'+\mu}\times[r_s\psi(r_s,-\mu') - \theta\psi_q(r,-\mu')]\right\}, \quad \mu < 0. \tag{79}$$

The equation that determines $\Gamma^0_<$ is

$$\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right)[X(\nu_0)e^{-r_s/\nu_0} - X(-\nu_0)e^{-r_s/\nu_0}] - \int_0^1 d\nu X(-\nu)\exp(-2r_s/\nu)\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) - \int_0^1 d\mu'[X^+(\mu') - X^-(\mu')] \times [r_s\psi(r_s, -\mu') - \theta\psi_q(r, -\mu')] = 0, \quad (80)$$

where

$$X(z) = \frac{1}{1-z}\exp\left[\frac{1}{\pi}\int_0^1 d\mu \frac{\arg\Lambda^+(\mu)}{\mu - z}\right]. \quad (81)$$

Similarly, for the exterior problems, Eq. (76) has the solution

$$\tilde{\Gamma}_>\left(\frac{i}{\mu}\right) = -\frac{1}{4\pi i}\left(\frac{1}{X^+(\mu)} - \frac{1}{X^-(\mu)}\right) \times \left\{-\nu_0^2\Gamma^0_>\left(\frac{1}{\nu_0}\right)e^{-r_s/\nu_0}\frac{X(+\nu_0)}{\mu - \nu_0} + \int_0^1 d\mu' \frac{X^+(\mu') - X^-(\mu')}{\mu' - \mu} \times [\theta\psi(r, \mu') - r_s\psi(r_s, \mu')]\right\}, \quad \mu > 0. \quad (82)$$

The equation that determines $\Gamma^0_>$ is

$$\nu_0^2\Gamma^0_>\left(\frac{1}{\nu_0}\right)X(+\nu_0)e^{-r_s/\nu_0} + \int_0^1 d\mu'[X^+(\mu') - X^-(\mu')] \times [\theta\psi_q(r, \mu') - r_s\psi(r_s, \mu')] = 0. \quad (83)$$

Let us now consider some specific problems and determine the coefficients $\tilde{\Gamma}_{\lessgtr}$ and $\Gamma^0_{\lessgtr}$ explicitly. We treat the interior problems first.

A. The Albedo Problem

The albedo problem involves the determination of neutron angular density everywhere inside the source-free sphere $[\psi_q(r, \mu) \equiv 0]$ with an incident distribution given by

$$\psi(r_s, \mu) = (r_s^2)^{-1}\delta(\mu - \mu_0), \quad \mu < 0, \quad \mu_0 < 0. \quad (84)$$

Under this boundary condition, Eq. (79) for $\tilde{\Gamma}_<$ becomes

$$\tilde{\Gamma}\left(\frac{i}{\mu}\right) = -\frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)} - \frac{1}{X^-(-\mu)}\right)\left\{\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right) \times \left[\exp\left(\frac{r_s}{\nu_0}\right) - \exp\left(\frac{-r_s}{\nu_0}\right)\frac{X(-\nu_0)}{\mu - \nu_0}\right] + \int_0^1 d\nu \exp(-2r_s/\nu)\frac{\tilde{\Gamma}_<(i/\nu)}{\nu - \mu}X(-\nu) + \frac{1}{r_s}\frac{X^+(-\mu_0) - X^-(-\mu_0)}{\mu - \mu_0}\right\}. \quad (85)$$

For $\Gamma^0_<$, we have

$$\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right)\left[X(\nu_0)\exp\left(\frac{r_s}{\nu_0}\right) - X(-\nu_0)\exp\left(-\frac{r_s}{\nu_0}\right)\right] - \int_0^1 d\nu X(-V)\exp\left(\frac{-2r_s}{\nu}\right)\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) - \frac{1}{r_s}[X^+(-\mu_0) - X^-(-\mu_0)] = 0. \quad (86)$$

Equations (85) and (86) are well suited for asymptotic expansions of the angular density. Thus, for a large sphere, one may neglect the integral term in Eq. (85) involving $\exp(-2r_s/\nu)$, and solve the equation by iteration. In particular, in the zeroth approximation, we have

$$\tilde{\Gamma}_<\left(\frac{i}{\mu}\right) \cong -\frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)} - \frac{1}{X^-(-\mu)}\right) \times \left\{\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right)\left[\exp\left(\frac{r_s}{\nu_0}\right)\frac{X(\nu_0)}{\mu + \nu_0} - \exp\left(\frac{-r_s}{\nu_0}\right)\frac{X(-\nu_0)}{\mu - \nu_0}\right] + \frac{1}{r_s}\frac{X^+(-\mu_0) - X^-(-\mu_0)}{\mu - \mu_0}\right\} = 0 \quad (87)$$

and

$$\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right) \doteq \frac{[X^+(-\mu_0) - X^-(-\mu_0)]}{r_s[X(\nu_0)\exp(r_s/\nu_0) - X(-\nu_0)\exp(-r_s/\nu_0)]}. \quad (88)$$

Now, by eliminating $\nu_0^2\Gamma^0_<(1/\nu_0)$ from Eq. (87), we obtain the explicit expression for $\tilde{\Gamma}_<$. The angular density is then readily obtained from Eq. (42).

B. Milne Problem for the Interior of a Sphere

The Milne problem involves the determination of neutron angular density everywhere inside the sphere with a source at the center and zero incident distribution. Thus, the boundary condition is

$$\psi(r_s, \mu) = 0, \quad \mu < 0. \quad (89)$$

We assume an isotropic source; i.e.,

$$Q(\mathbf{r}, \boldsymbol{\Omega}) = \delta(\mathbf{r}). \quad (90)$$

Putting this expression for Q into Eq. (10), we obtain

$$\psi_q(r,\mu) = \frac{1}{8\pi^2}\int_{-\infty}^{\infty}\frac{dkk^2}{\Lambda} I(k,r,\mu), \tag{91}$$

where $I(k, r, \mu)$ is given by Eq. (17).

As before, we may express ψ_q in terms of boundary values about the branch cut in the upper half k plane by first decomposing $I(k, r, \mu)$ as shown in Eq. (45). To avoid repetition of calculations, we merely state the answer. Thus,

$$\psi_q(r,\mu) = -\frac{c}{8\pi}\int_i^{i\infty}\frac{dkk}{\Lambda^+\Lambda^-} Z(k,r,\mu) - \frac{ik_0^2}{4\pi\Lambda'(ik_0)}\zeta^{(1)}(ik_0,r,\mu), \tag{92}$$

where $Z(k, r, \mu)$ and $S^{(1)}(ik_0, r, \mu)$ are given by Eqs. (50) and (27), respectively. Put $k = i/\nu$ and $k_0 = 1/\nu_0$ to get

$$\psi_q(r,\mu) = \frac{c}{8\pi}\int_0^1\frac{d\nu}{\nu^3\Lambda^+\Lambda^-} Z\left(\frac{i}{\nu}, r, \mu\right) + \frac{1}{4\pi^2 i\nu_0\Lambda'}\zeta^{(1)}\left(\frac{i}{\nu_0}, r, \mu\right). \tag{93}$$

After the application of the reduction operator to Eq. (91) for $\mu < 0$, we get

$$\theta\psi_q(r,\mu) = \frac{c}{4\pi}\int_0^1\frac{d\nu}{\nu\Lambda^+\Lambda^-}\frac{e^{-r_s/\nu}}{\nu-\mu} + \frac{1}{2\pi i\Lambda'}\frac{e^{-r_s/\nu_0}}{\nu_0-\mu}, \quad \mu<0. \tag{94}$$

Now let us subject Eq. (79) to the boundary condition (92) and insert the expression (94) for $\theta\psi_q$. The result is

$$\tilde{\Gamma}_<\left(\frac{i}{\mu}\right) = -\frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)}-\frac{1}{X^-(-\mu)}\right) \times\left\{\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right)\left[\exp\left(\frac{r_s}{\nu_0}\right)\frac{X(\nu_0)}{\mu+\nu_0} - \exp\left(\frac{-r_s}{\nu_0}\right)\frac{X(-\nu_0)}{\mu-\nu_0}\right] - \frac{1}{2\pi i\Lambda'}\exp\left(\frac{-r_s}{\nu_0}\right)\frac{X(-\nu_0)}{\mu-\nu_0} + \int_0^1 d\nu\exp\left(\frac{-2r_s}{\nu}\right)\tilde{\Gamma}_<\left(\frac{i}{\nu}\right)\frac{X(-\nu)}{\nu-\mu} + \frac{c}{4\pi}\int_0^1\frac{d\nu}{\nu\Lambda^+\Lambda^-}\exp\left(\frac{-r_s}{\nu}\right)\frac{X(-\nu)}{\nu-\mu}\right\}, \quad \mu<0. \tag{95}$$

From Eq. (79), we get

$$\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right)\left[\exp\left(\frac{r_s}{\nu_0}\right)X(\nu_0) - \exp\left(\frac{-r_s}{\nu_0}\right)X(-\nu_0)\right] - \frac{1}{2\pi i\Lambda'}\exp\left(\frac{-r_s}{\nu_0}\right)X(-\nu_0) - \int_0^1 d\nu\exp\left(\frac{-2r_s}{\nu}\right)X(-\nu)\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) - \frac{c}{4\pi}\int_0^1\frac{d\nu}{\nu\Lambda^+\Lambda^-}\exp\left(\frac{-r_s}{\nu}\right)X(-\nu) = 0. \tag{96}$$

For a large sphere, we may neglect the integrals in Eqs. (95) and (96) involving $\exp(-2r_s/\nu)$. Thus, in the zeroth approximation, Eq. (96) becomes

$$\nu_0^2\Gamma^0_<\left(\frac{1}{\nu_0}\right) \simeq \frac{1}{2\pi i\Lambda'}\exp\left(\frac{-2r_s}{\nu_0}\right)\frac{X(-\nu_0)}{X(\nu_0)} + \frac{c}{4\pi}\exp\left(\frac{-r_s}{\nu_0}\right)\int_0^1\frac{d\nu}{\nu\Lambda^+\Lambda^-} \times\exp\left(\frac{-r_s}{\nu}\right)\frac{X(-\nu)}{X(\nu_0)}, \tag{97}$$

while Eq. (95) becomes

$$\tilde{\Gamma}\left(\frac{i}{\mu}\right) \simeq -\frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)}-\frac{1}{X^-(-\mu)}\right) \times\left[\frac{1}{\pi i\Lambda'}\exp\left(\frac{-r_s}{\nu_0}\right)\frac{\nu_0 X(-\nu_0)}{(\mu+\nu_0)(\mu-\nu_0)} + \frac{c}{4\pi(\mu+\nu_0)}\int_0^1\frac{d\nu'}{\nu'\Lambda^+\Lambda^-} \times\exp\left(\frac{-r_s}{\nu'}\right)\frac{X(-\nu')}{\nu'-\mu}(\nu'+\nu_0)\right]. \tag{98}$$

These are precisely the coefficients which occur in the half-space Milne problem,[1] as expected. The first order approximation of $\Gamma^0_<$ and $\tilde{\Gamma}_<$ may now be readily obtained by computing the integrals previously neglected in Eqs. (96) and (95) from the first

iterations (97) and (98). Thus,

$$\nu_0^2\tilde{\Gamma}_<\left(\frac{1}{\nu_0}\right) \cong \frac{1}{2\pi i\Lambda'}\exp\left(\frac{-2r_s}{\nu_0}\right)\frac{X(-\nu_0)}{X(\nu_0)} + \frac{c}{4\pi}\exp\left(\frac{-r_s}{\nu_0}\right)\int_0^1\frac{d\nu}{\nu\Lambda^+\Lambda^-}\exp\left(\frac{-r_s}{\nu}\right)\frac{X(-\nu)}{X(\nu_0)} - \frac{1}{2\pi i}\int_0^1 d\nu\exp\left(\frac{-r_s}{\nu}\right)X(-\nu)\left(\frac{1}{X^+(\nu)}-\frac{1}{X^-(\nu)}\right) \times\left[\frac{1}{\pi i\Lambda'}\exp\left(\frac{-r_s}{\nu_0}\right)\frac{\nu_0X(-\nu_0)}{(\nu_0-\nu)(\nu_0+\nu)} - \frac{c}{4\pi(\nu_0-\nu)}\int_0^1\frac{d\nu'}{\nu'\Lambda^+\Lambda^-} \times\exp\left(\frac{-r_s}{\nu'}\right)\frac{X(-\nu')}{\nu'+\nu}(\nu'+\nu_0)\right] \tag{99}$$

and

$$\tilde{\Gamma}_<\left(\frac{i}{\mu}\right) \cong \frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)}-\frac{1}{X^-(-\mu)}\right) \times\left\{\frac{1}{\pi i\Lambda'}\exp\left(\frac{-r_s}{\nu_0}\right)\frac{\nu_0X(-\nu_0)}{(\nu_0-\nu)(\nu_0+\nu)} - \frac{c}{4\pi(\nu_0-\nu)}\int_0^1\frac{d\nu'}{\nu'\Lambda^+\Lambda^-} \times\exp\left(\frac{-r_s}{\nu'}\right)\frac{X(-\nu')}{\nu'+\nu}(\nu'+\nu_0) - \frac{1}{2\pi i}\int_0^1\frac{d\nu}{\nu-\mu} \times\exp\left(\frac{-2r_s}{\nu}\right)X(-\nu)\left(\frac{1}{X^+(\nu)}-\frac{1}{X^-(\nu)}\right) \times\left[\frac{1}{\pi i\Lambda'}\exp\left(\frac{r_s}{\nu_0}\right)\frac{\nu_0X(-\nu_0)}{(\nu_0-\nu)(\nu_0+\nu)} - \frac{c}{4\pi(\nu_0-\nu)}\int_0^1\frac{d\nu'}{\nu'\Lambda^+\Lambda^-} \times\exp\left(\frac{-r_s}{\nu'}\right)\frac{X(-\nu')}{\nu'+\nu}(\nu'+\nu_0)\right]\right\}. \tag{100}$$

C. The Critical Problem

We mentioned earlier that, when the zeros of the dispersion function are real ($c > 1$), the Green's function [Eq. (5)] is not uniquely determined. The angular density, however, is still uniquely determined regardless of the manner we choose to treat the singularities. To illustrate this point, let us consider the critical problem. Assuming no volume sources, the integral equation (20) becomes

$$\psi(r,\mu) = \psi_0(r,\mu) + \psi_c(r,\mu). \tag{101}$$

The dispersion function occurs only in ψ_c, which is given by Eq. (29). We rewrite this equation as follows:

$$\psi_c(r,\mu) = \frac{c}{8\pi}\int_{-\infty}^{\infty}dkk^2T_1(k)I(k,r,\mu) \times\left[\mathcal{P}\frac{1}{\Lambda}+\lambda_1\delta(k-k_1)+\lambda_2\delta(k-k_2)\right], \tag{102}$$

where $\mathcal{P}$ implies the Cauchy principal value, λ_1 and λ_2 are some arbitrary functions of k, and k_1, k_2 are the real zeros of $\Lambda(k)$. Since $\Lambda(-k) = \Lambda(k)$, $-k_2 = k_1 = k_0 > 0$. Also from Eq. (28), we have $T_1(-k) = T_2(k)$. With this in mind, let us re-express Eq. (101) in the following form:

$$\psi_c(r,\mu) = \frac{c}{8\pi}\int_{-\infty}^{\infty}dkk^2\frac{T_1(k)}{\Lambda}I(k,r,\mu) + \phi(k_0,r,\mu)\Gamma^c(k_0), \tag{103}$$

where we have omitted writing the principal value symbol and

$$\Gamma^c(k_0) = \frac{c}{8\pi}k_0^2[\lambda_1T_1(k_0)+\lambda_2T_2(k_0)] \tag{104}$$

and

$$\phi(k_0,r,\mu) = \int_{-1}^{1}\frac{dt}{1+ik_0t}e^{ik_0tr\mu}J_0(k_0r[(1-\mu^2)(1-t^2)]^{\frac{1}{2}}) \tag{105}$$

is the discrete regular eigenfunction which is of oscillatory type in contrast to the eigenfunctions constructed previously.

The calculation of the integral in Eq. (103) may now be carried out in exact analogy with the interior problems for $c < 1$. Here we merely state the final result. Thus,

$$\psi(r,\mu) = \int_0^1\frac{d\nu}{\nu}\Gamma_<\left(\frac{i}{\nu}\right)\phi\left(\frac{i}{\nu},r,\mu\right) + \phi(k_0,r,\mu)\Gamma^c(k_0), \tag{106}$$

where the coefficient $\Gamma_<$ is given by Eq. (38), and, as before, $\phi(i/\nu, r, \mu)$ [see Eq. (35)] are the regular eigenfunctions corresponding to the continuous spectrum.

The implication of our previous statement as to the uniqueness of the angular density should now be obvious. In particular, we see from Eq. (106) that the coefficients $\Gamma_<$ and Γ^c are determined *uniquely by the*

boundary conditions, and the angular density, therefore, *does not* depend on how we set up these coefficients. Equation (104) illustrates such an arbitrariness in Γ^c.

The steps involved in obtaining the auxiliary equation for the coefficients are exactly those involved in the interior problems for $c < 1$. Thus, Eq. (69) represents that equation with $\theta\psi_q \equiv 0$, Γ^c replacing $\Gamma^0_<$, and the discrete eigenfunctions by their oscillatory counterpart. In fact, from Eq. (79), which represents the most general solution of Eq. (69), we have

$$\begin{aligned}\tilde{\Gamma}_<\left(\frac{i}{\mu}\right) = &-\frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)} - \frac{1}{X^-(-\mu)}\right)\\ &\times\left\{\nu_0^2\Gamma^c\left(\frac{1}{\nu_0}\right)\left[\exp\left(\frac{ir_s}{\nu_0}\right)\frac{X(-i\nu_0)}{\mu - i\nu_0}\right.\right.\\ &\left.-\exp\left(\frac{-ir_s}{\nu_0}\right)\frac{X(i\nu_0)}{\mu + i\nu_0}\right]\\ &+\int_0^1 d\nu \exp\left(\frac{-2r_s}{\nu}\right)\frac{\tilde{\Gamma}_<(i/\nu)}{\nu - \mu}X(-\nu)\\ &\left.+\int_0^1 d\mu'\frac{X^+(\mu') - X^-(\mu')}{\mu' + \mu}r_s\psi(r_s, \mu)\right\}.\end{aligned} \tag{107}$$

The equation that determines Γ^c is

$$\begin{aligned}&\nu_0^2\Gamma^c\left(\frac{1}{\nu_0}\right)\left[X(-i\nu_0)\exp\left(\frac{-ir_s}{\nu_0}\right) - X(i\nu_0)\exp\left(\frac{ir_s}{\nu_0}\right)\right]\\ &-\int_0^1 d\nu X(-\nu)\exp\left(\frac{-2r_s}{\nu}\right)\tilde{\Gamma}_<\left(\frac{i}{\nu}\right)\\ &-\int_0^1 d\mu'[X^+(\mu') - X^-(\mu')]r_s\psi(r_s, -\mu') = 0.\end{aligned} \tag{108}$$

The boundary condition for the critical problem is

$$\psi(r_s, \mu) = 0, \quad \mu < 0. \tag{109}$$

Inserting this boundary condition into Eq. (107), we get

$$\begin{aligned}\tilde{\Gamma}_<\left(\frac{i}{\mu}\right) = &-\frac{1}{2\pi i}\left(\frac{1}{X^+(-\mu)} - \frac{1}{X^-(-\mu)}\right)\\ &\times\left\{\nu_0^2\Gamma^c\left(\frac{1}{\nu_0}\right)\left[\exp\left(\frac{ir_s}{\nu_0}\right)\frac{X(-i\nu_0)}{\mu - i\nu_0}\right.\right.\\ &\left.-\exp\left(\frac{-ir_s}{\nu_0}\right)\frac{X(i\nu_0)}{\mu + i\nu_0}\right]\\ &\left.+\int_0^1 d\nu \exp\left(\frac{-2r_s}{\nu}\right)\frac{\tilde{\Gamma}_<(i/\nu)}{\nu - \mu}X(-\nu)\right\},\end{aligned} \tag{110}$$

and, from Eq. (107), we get

$$\begin{aligned}&\nu_0^2\Gamma^c\left(\frac{1}{\nu_0}\right)\left[X(-i\nu_0)\exp\left(\frac{ir_s}{\nu_0}\right) - X(i\nu_0)\exp\left(\frac{-ir_s}{\nu_0}\right)\right]\\ &-\int_0^1 d\nu X(-\nu)\exp\left(\frac{-2r_s}{\nu}\right)\tilde{\Gamma}_<\left(\frac{i}{\nu}\right) = 0.\end{aligned} \tag{111}$$

For a large sphere, if we neglect the integral in Eq. (111) involving $\exp(-2r_s/\nu_0)$, we get

$$\exp(-ir_s/\nu)X(+i\nu_0) - X(-i\nu_0)\exp(ir_s/\nu_0) = 0, \tag{112}$$

which merely states that the asymptotic density is to vanish at the extrapolated end point. This problem has been extensively treated for planar geometry by the normal mode expansion technique.[1,4]

As a final application of the Green's function technique, let us consider the Milne problem for the exterior of a black sphere.

D. Milne Problem for the Exterior of a Black Sphere

The problem under consideration involves the determination of the neutron angular density outside a purely absorbing sphere (black sphere). Far away from the sphere, there is a source which supplies the neutrons. Since the black sphere implies zero emergent distribution, the appropriate boundary condition is

$$\psi(r_s, \mu) = 0, \quad \mu > 0. \tag{113}$$

In calculating the angular density $\psi_q(r, \mu)$ in Eq. (47), let us assume that a spherically symmetric source is located at some distance R outside the black sphere. Thus, let

$$Q(r, \Omega) = (q_0/R^2)\delta(r - R), \quad R > r_s. \tag{114}$$

Putting this source function into Eq. (10), we get

$$\psi_q(r, \mu) = \frac{q_0}{4\pi Ri}\int_{-\infty}^{\infty}\frac{dkk}{\Lambda}e^{ikR}I(k, r, \mu), \tag{115}$$

where $I(k, r, \mu)$ is given by Eq. (17).

Let us split Eq. (115) into two parts as follows:

$$\begin{aligned}\psi_q(r, \mu) = \frac{q_0}{4\pi iR}\left[\int_{-\infty}^{\infty}\frac{dkk}{\Lambda}e^{ikR}I(k, r, \mu)\Theta(R - r)\right.\\ \left.+\int_{-\infty}^{\infty}\frac{dkk}{\Lambda}e^{ikR}I(k, r, \mu)\Theta(r - R)\right].\end{aligned} \tag{116}$$

Now, if we push the source to infinity (i.e., let $R \to \infty$), we see that the second integral in Eq. (116) will make no contribution. Furthermore, in the same limit, the modes with continuous spectrum must also disappear. Thus, in Eq. (110), if we choose

$$q_0 = R\,e^{k_0R}i[k_0\Lambda'(k_0)]^{-1}, \tag{117}$$

where $\Lambda(ik_0) = 0$, and let $R \to \infty$, we get

$$\psi_q(r, \mu) = -\phi(ik_0, r, \mu). \tag{118}$$

The application of the reduction operator for $\mu > 0$ to Eq. (116) gives us, in the limit $R \to \infty$ and the

same choice of q_0 as given in Eq. (117),

$$\theta\psi_c(r,\mu) = \frac{1}{2k_0}\left(\frac{e^{-k_0 r_s}}{1-k_0\mu} - \frac{e^{k_0 r_s}}{1+k_0\mu}\right), \quad \mu > 0. \tag{119}$$

The coefficients $\tilde{\Gamma}_>$ and $\Gamma^0_>$ given by Eqs. (82) and (83), respectively, which solve the integral equation (47), may now be readily obtained. Thus, inserting the boundary condition (113) and Eq. (119) into Eq. (82), we get

$$\tilde{\Gamma}_>\left(\frac{i}{\mu}\right) = -\frac{1}{2\pi i}\left(\frac{1}{X^+(\mu)} - \frac{1}{X^-(\mu)}\right) \times \left\{\frac{\nu_0^2 X(\nu_0)}{\mu-\nu_0} e^{-r_s/\nu_0}\left[1-\Gamma_>\left(\frac{1}{\nu_0}\right)\right] + \frac{\nu_0^2 X(-\nu_0)}{\mu+\nu_0} e^{r_s/\nu_0}\right\}. \tag{120}$$

The equation that determines $\Gamma^0_>$ is similarly obtained from Eq. (82). Thus,

$$\Gamma^0_>(1/\nu_0) = 1 + [X(-\nu_0)/X(\nu_0)] \exp(2r_s/\nu_0). \tag{121}$$

One usually writes

$$X(-\nu_0)/X(\nu_0) = -\exp(-2r_0/\nu_0), \tag{122}$$

where r_0 is the so called extrapolation distance which determines the distance where the asymptotic neutron density vanishes.

APPENDIX

In the main body of this paper, we have dealt with the media with regeneration property. In the course of the treatment we encountered certain complicated looking functions such as $\zeta^{(1),(2)}$, defined by the infinite series (27), which are hard to relate to any classically known functions. In this section we consider the Green's function for media without regeneration property (i.e., $c = 0$). (For a geometrical interpretation of the Green's function for purely absorbing media, see Ref. 1.) Here also we encounter similar type of functions which do not seem to have classical analogs, but their certain integrals are related to Dirac's delta function. Consequently, they give rise to some interesting mathematical identities and completeness relations (half and full range). For the planar geometry, the completeness relations are rather trivial. However, for the sake of comparing the hierarchy of complexity, we also present these trivial completeness relations.

Derivation of Identities

Let us begin with the Green's function (for $c = 0$) in the form of a Fourier integral

$$G(\mathbf{r},\boldsymbol{\Omega};\mathbf{r}_0,\boldsymbol{\Omega}_0) = \frac{1}{(2\pi)^3}\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)\int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{R}}}{1+i\mathbf{k}\cdot\boldsymbol{\Omega}}, \tag{A1}$$

where $\mathbf{R} = \mathbf{r} - \mathbf{r}_0$. The integration may be carried out in a straightforward manner by first resolving $\mathbf{k}$ and $\mathbf{R}$ as follows:

$$\mathbf{k} = \boldsymbol{\Omega}\mathbf{k}\cdot\boldsymbol{\Omega} + \mathbf{k}_\perp, \quad \text{such that} \quad \mathbf{k}_\perp\cdot\boldsymbol{\Omega} = 0,$$

and

$$\mathbf{R} = \boldsymbol{\Omega}\mathbf{R}\cdot\boldsymbol{\Omega} + \mathbf{R}_\perp, \quad \text{such that} \quad \mathbf{R}_\perp\cdot\boldsymbol{\Omega} = 0.$$

Then we can write

$$\mathbf{k}\cdot\mathbf{R} = \mathbf{k}\cdot\boldsymbol{\Omega}\mathbf{R}\cdot\boldsymbol{\Omega} + \mathbf{k}_\perp\cdot\mathbf{R}_\perp.$$

Equation (1) then becomes

$$G = \delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)\frac{1}{2\pi} \times \int_{-\infty}^{\infty}\frac{d(\mathbf{k}\cdot\boldsymbol{\Omega})\exp(i\mathbf{k}\cdot\boldsymbol{\Omega}\mathbf{R}\cdot\boldsymbol{\Omega})}{1+i\mathbf{k}\cdot\boldsymbol{\Omega}}\cdot\frac{1}{(2\pi)^2} \times \int d^2k_\perp \exp(i\mathbf{k}_\perp\cdot\mathbf{R}_\perp).$$

Separate parts of the integrals are

$$\frac{1}{(2\pi)^2}\int d^2k_\perp \exp(i\mathbf{k}_\perp\cdot\mathbf{R}_\perp) = \delta(\mathbf{R}_\perp)$$

and

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d(\mathbf{k}\cdot\boldsymbol{\Omega})}{1+i\mathbf{k}\cdot\boldsymbol{\Omega}} e^{i\mathbf{k}\cdot\boldsymbol{\Omega}\mathbf{R}\cdot\boldsymbol{\Omega}} = e^{-\mathbf{R}\cdot\boldsymbol{\Omega}}\Theta(\mathbf{R}\cdot\boldsymbol{\Omega}),$$

where

$$\Theta(\mathbf{R}\cdot\boldsymbol{\Omega}) = 1, \quad \mathbf{R}\cdot\boldsymbol{\Omega} > 0, \qquad = 0, \quad \mathbf{R}\cdot\boldsymbol{\Omega} < 0. \tag{A2}$$

The expression for the Green's function now may be written as

$$G = \exp(-\mathbf{R}\cdot\boldsymbol{\Omega})\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)\delta(\mathbf{R}_\perp)\Theta(\mathbf{R}\cdot\boldsymbol{\Omega}). \tag{A3}$$

This equation holds for any arbitrary values of r and r_0.

Let us consider Eq. (1) again and carry out the integration in a manner parallel to the treatment of interior and exterior problems. Define I as

$$I = \frac{1}{(2\pi)^3}\int\frac{d^3k\exp[i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)]}{1+i\mathbf{k}\cdot\boldsymbol{\Omega}}, \tag{A4}$$

and let $I_<$ denote this integral when $r < r_0$, and $I_>$ when $r > r_0$. First consider $r < r_0$. Expanding

$\exp(-i\mathbf{k}\cdot\mathbf{r}_0)$ in terms of spherical harmonics and using the cosine formula to express $\hat{\mathbf{k}}\cdot\mathbf{r} = t$ and $\hat{\mathbf{r}}\cdot\boldsymbol{\Omega} = \mu$, we get

$$I_< = \frac{1}{2\pi}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n e^{\frac{1}{2}im\pi} Y_{nm}(-\hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega}) \times \int_{-\infty}^{\infty} dk k^2 \zeta_n^{(2)}(kr_0)\int_{-1}^{1}\frac{dt}{1+ikt} Y^*_{nm}(t,\varphi) \cdot e^{iktr\mu} J_m\{kr[(1-\mu^2)(1-t^2)]^{\frac{1}{2}}\}. \quad \text{(A5)}$$

(Here we have used the decomposition of the spherical Bessel function in terms of spherical Hankel functions $\zeta_n^{(1)}$ and $\zeta_n^{(2)}$.) Now the distortion of the path of integration (with respect to k) to the path surrounding the branch cut in the lower half k plane yields

$$I_< = -e^{-r\mu}\int_{-1}^{0}\frac{d\nu}{\nu^3}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n e^{-\frac{1}{2}im\pi}\zeta_n^{(2)}\left(i\frac{r_0}{\nu}\right) \times Y_{nm}(-\hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega})Y^*_{nm}(\nu,\varphi) \times I_m\{r[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\}. \quad \text{(A6)}$$

Since $G = I\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)$, the comparison of Eq. (6) with Eq. (3) gives us the first identity.

Identity 1:

$$-\exp(-r\mu)\int_{-1}^{0}\frac{d\nu}{\nu^3}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n \exp\left(-im\frac{\pi}{2}\right)\zeta_n^{(2)}\left(i\frac{r_0}{\nu}\right) \times Y_{nm}(-\hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega})Y^*_{nm}(\nu,\varphi) \times I_m\{r[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\} = \exp(-\mathbf{R}\cdot\boldsymbol{\Omega})\delta(\mathbf{R}_\perp)\Theta(\mathbf{R}\cdot\boldsymbol{\Omega}), \quad r < r_0, \quad \mu \equiv \hat{\mathbf{r}}\cdot\boldsymbol{\Omega}.$$

Similarly, for $r > r_0$, we have the second.

Identity 2:

$$\exp(r_0\mu_0)\int_0^1\frac{d\nu}{\nu^3}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n e^{\frac{1}{2}im\pi}\zeta_n^{(1)}\left(i\frac{r}{\nu}\right) \times Y_{nm}(\hat{\mathbf{r}}\cdot\boldsymbol{\Omega})Y^*_{nm}(\nu,\varphi_0) \times I_m(r_0[(1-\mu_s^2)(1/\nu^2-1)]^{\frac{1}{2}}) = \exp(-\mathbf{R}\cdot\boldsymbol{\Omega})\delta(\mathbf{R}_\perp)\Theta(\mathbf{R}\cdot\boldsymbol{\Omega}), \quad r > r_0, \quad \mu_s \equiv \hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega}.$$

We remark here that Identities 1 and 2 so far are not restricted to any geometry. Also note that the Green's functions $G_<$ ($r < r_0$) and $G_>$ ($r > r_0$) are related to $I_<$ and $I_>$, respectively, by

$$G_{\lessgtr} = \delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0)I_{\lessgtr}. \quad \text{(A7)}$$

Now if we express the delta function in Eq. (A7) in the form

$$\delta(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}_0) = \delta(\mu_s-\mu_0)\delta(\varphi_\Omega - \varphi_{\Omega_0}), \quad \text{(A8)}$$

where $\mu_s = \hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega}$, $\mu_0 = \hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega}_0$, and in Eq. (A3) write $\delta(\mathbf{R}_\perp)$ in cylindrical coordinate system, i.e.,

$$\delta(\mathbf{R}_\perp) = r^{-1}(1-\mu^2)^{-\frac{1}{2}} \times \delta[r(1-\mu^2)^{\frac{1}{2}} - r_0(1-\mu_s^2)^{\frac{1}{2}}]\delta(\varphi-\varphi_s), \quad \text{(A9)}$$

and integrate Eq. (A7) over all angles except $\mu = \hat{\mathbf{r}}\cdot\boldsymbol{\Omega}$ and $\mu_0 = \hat{\mathbf{r}}_0\cdot\boldsymbol{\Omega}_0$, we get Identities 3 and 4 corresponding to the spherical geometry.

Identity 3:

$$-\tfrac{1}{2}\exp(-r\mu)\int_{-1}^{0}\frac{d\nu}{\nu^3}S^{(2)}\left(\frac{ir_0}{\nu},-\mu_0,\nu\right) \times I_0\{r[(1-\mu^2)(1/\nu^2-1)]^{\frac{1}{2}}\} = \exp[-(r\mu - r_0\mu_0)] \times \frac{\delta[r(1-\mu^2)^{\frac{1}{2}} - r_0(1-\mu_0^2)^{\frac{1}{2}}]}{r(1-\mu^2)^{\frac{1}{2}}}\Theta(r\mu - r_0\mu_0), \quad r < r_0;$$

Identity 4:

$$\tfrac{1}{2}\exp(r_0\mu_0)\int_0^1\frac{d\nu}{\nu^3}S^{(1)}\left(\frac{ir}{\nu},\mu,\nu\right) \times I_0\{r_0((1-\mu_0^2)[(1/\nu^2)-1])^{\frac{1}{2}}\} = \exp[-(r\mu - r_0\mu_0)] \times \frac{\delta[r(1-\mu^2)^{\frac{1}{2}} - r_0(1-\mu_0^2)^{\frac{1}{2}}]}{r(1-\mu^2)^{\frac{1}{2}}}\Theta(r\mu - r_0\mu_0), \quad r > r_0,$$

where

$$S^{(1),(2)}\left(\frac{ir}{\nu},\mu,\nu\right) = \sum_{n=0}^{\infty} i^n(2n+1)\zeta_n^{(1),(2)}\left(\frac{ir}{\nu}\right)P_n(\mu)P_n(\nu). \quad \text{(A10)}$$

A more convenient form of these identities is obtained if we use the formula

$$\delta(f(x)) = \sum_n \frac{\delta(x-x_n)}{|f'(x_n)|}, \quad f(x_n) = 0,$$

to re-express the right-hand sides. This gives us:

Identity 3':

$$-\tfrac{1}{2}\exp(-r\mu)\int_{-1}^{0}\frac{d\nu}{\nu^3}S^{(2)}\left(\frac{ir_0}{\nu},-\mu_0,\nu\right) \times I_0\{r[(1-\mu^2)(1/\nu^2-1)]^{\frac{1}{2}}\} = \exp[-(r\mu - r_0\mu_0)]\frac{1}{r^2|\mu|} \times [\delta(\mu-\mu_1)+\delta(\mu+\mu_1)]\Theta(r\mu - r_0\mu_0), \quad r < r_0;$$

Identity 4′:

$$\tfrac{1}{2}\exp(r_0\mu_0)\int_0^1 \frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir}{\nu},\mu,\nu\right) \times I_0\{r_0[(1-\mu_0^2)(1/\nu^2-1)]^{\frac{1}{2}}\} = \exp[-(r\mu-r_0\mu_0)](r^2|\mu|)^{-1} \times [\delta(\mu-\mu_1)+\delta(\mu+\mu_1)]\Theta(r\mu-r_0\mu_0), \quad r>r_0,$$

where

$$\mu_1 = [1-(r_0^2/r^2)(1-\mu_0^2)]^{\frac{1}{2}}.$$

Completeness Theorems

Let us note that the left-hand and right-hand sides of identities (3′) and (4′) are two representations of $G_<$ and $G_>$, respectively. The half-range completeness theorems follow from the limits of $G_<$ and $G_>$ as r approaches r_0. Specializing to various particular values of μ and μ_0, let us first obtain a number of useful results. Consider $G_<$ first.

The right-hand side representation of $G_<$ is

$$G_< = \exp[-(r\mu-r_0\mu_0)]\frac{1}{r^2\mu} \times [\delta(\mu+\mu_1)+\delta(\mu-\mu_1)]\Theta(r\mu-r_0\mu_0). \tag{A11}$$

1. $\mu_0>0, \quad \mu>0$

The argument of the Θ function is positive if $\mu > (r_0/r)\mu_0$. Now the first δ function makes no contribution, since its argument cannot vanish. The second δ function can contribute if

$$\mu=\mu_1=\{(r_0^2/r^2)\mu_0^2-[(r_0^2/r^2)-1]\}^{\frac{1}{2}} < (r_0/r)\mu_0, \quad (r_0/r)>1.$$

But then $\Theta(r\mu-r_0\mu_0)=0$. Hence,

$$G_<=0, \quad \mu_0>0, \quad \text{and} \quad \mu>0. \tag{A12a}$$

2. $\mu_0>0, \quad \mu<0$

For these values of μ and μ_0, $\Theta(r\mu-r_0\mu_0)=0$. Hence, again

$$G_<=0, \quad \mu_0>0, \quad \mu<0. \tag{A12b}$$

3. $\mu_0<0, \quad \mu>0$

For this $\Theta=1$, $\delta(\mu+\mu_1)$ makes no contribution. Therefore, only $\delta(\mu-\mu_1)$ may contribute. Hence,

$$G_< = e^{-(r\mu-r_0\mu_0)}(r^2|\mu|)^{-1}\delta(\mu-\mu_1), \quad \mu_0<0, \quad \mu>0. \tag{A12c}$$

4. $\mu_0<0, \quad \mu<0$

Then $r\mu-r_0\mu_0 = r_0|\mu_0|-r|\mu|>0$ if

$$|\mu|<(r_0/r)|\mu_0|.$$

Clearly $\delta(\mu-\mu_1)$ makes no contribution. Hence, the possible contribution may come from $\delta(\mu+\mu_1)$. This is easily seen from putting $\mu+\mu_1$ equal to zero:

$$\mu+\mu_1 = 0 = -|\mu|+\mu_1$$

or

$$|\mu| = \{(r_0^2/r^2)\mu_0^2-[(r_0^2/r^2)-1]\}^{\frac{1}{2}} < (r_0/r)|\mu_0|.$$

The last inequality shows that $\Theta(r\mu-r_0\mu_0)=1$ is satisfied. Hence,

$$G_< = e^{-(r\mu-r_0\mu_0)}(r^2|\mu|)^{-1}\delta(\mu+\mu_1), \quad \mu_0<0, \quad \mu<0. \tag{A12d}$$

By exactly the same argument, one may show that $G_>$, in the right-hand side representation (see identity 4′), can be written as

$$G_> = e^{-(r\mu-r_0\mu_0)}(r^2|\mu|)^{-1}\delta(\mu-\mu_1), \quad \mu_0>0, \quad \mu>0, \tag{A13a}$$

$$G_> = 0, \quad \mu_0>0, \quad \mu<0, \tag{A13b}$$

$$G_> = [e^{-(r\mu-r_0\mu_0)}/r^2|\mu|]\delta(\mu-\mu_1), \quad \mu_0<0, \quad \mu>0, \tag{A13c}$$

$$G_> = 0, \quad \mu_0<0, \quad \mu<0. \tag{A13d}$$

Now in Eqs. (A12a)–(A12d) and (A13a)–(A13d) let $r\to r_0$. Denoting this limit of $G_{\lessgtr}$ by $G_{\mp}$, we obtain the following set of results:

$$G_- = 0, \quad \mu_0>0, \quad \mu>0, \tag{A14a}$$

$$G_- = 0, \quad \mu_0>0, \quad \mu<0, \tag{A14b}$$

$$G_- = (e^{-2r_0\mu}/r_0^2\mu)\delta(\mu+\mu_0), \quad \mu_0<0, \quad \mu>0, \tag{A14c}$$

$$G_- = -(r_0^2\mu)^{-1}\delta(\mu-\mu_0), \quad \mu_0<0, \quad \mu<0, \tag{A14d}$$

$$G_+ = (r_0^2\mu)^{-1}\delta(\mu-\mu_0), \quad \mu_0>0, \quad \mu>0, \tag{A15a}$$

$$G_+ = 0, \quad \mu_0>0, \quad \mu<0, \tag{A15b}$$

$$G_+ = (e^{-2r_0\mu}/r_0^2\mu)\delta(\mu+\mu_0), \quad \mu_0<0, \quad \mu>0, \tag{A15c}$$

$$G_+ = 0, \quad \mu_0<0, \quad \mu<0. \tag{A15d}$$

Let us remark here that the purpose of tabulating $G_{\pm}$ for various sets of values of μ, μ_0 (instead of using Heavyside theta function) is that only those representations of $G_{\pm}$ give rise to completeness relations which correspond to the same sign of μ, μ_0 –for instance, Eqs. (A14a), (A14d), (A15a), and (A15d). The rest give rise to mere identities.

In the left-hand side representations of $G_{\mp}$ we have from identities (A3′) and (A4′),

$$G_- = -\tfrac{1}{2}e^{-r_0\mu}\int_{-1}^{0}\frac{d\nu}{\nu^3}\,S^{(2)}\left(\frac{ir_0}{\nu}, -\mu_0, \nu\right) \times I_0\{r_0[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\}, \quad \text{(A16)}$$

$$G_+ = \tfrac{1}{2}e^{r_0\mu_0}\int_{0}^{1}\frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir_0}{\nu}, \mu, \nu\right) \times I_0\{r_0[(1-\mu_0^2)(1/\nu^2-1)]^{\frac{1}{2}}\}. \quad \text{(A17)}$$

From Eqs. (A14d) and (A15a) we may now conclude our half-range completeness theorems.

Theorem 1 (Half-Range Completeness): For $\mu_0 < 0$ and $\mu < 0$,

$$\frac{r_0^2}{2}\,e^{-r_0\mu}\int_{-1}^{0}\frac{d\nu}{\nu^3}\,S^{(2)}\left(\frac{ir_0}{\nu}, -\mu_0, \nu\right) \times I_0\{r_0[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\} = \delta(\mu-\mu_0).$$

Theorem 2 (Half-Range Completeness): For $\mu_0 > 0$ and $\mu > 0$,

$$\tfrac{1}{2}r_0^2\mu e^{r_0\mu_0}\int_{0}^{1}\frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir_0}{\nu}, \mu, \nu\right) \times I_0\{r_0[(1-\mu_0^2)(\nu^{-2}-1)]^{\frac{1}{2}}\} = \delta(\mu-\mu_0).$$

The full-range completeness theorems may now be readily obtained by taking the appropriate differences of G_+ and G_-. Thus, subtracting (A16) from (A17), we have (in the left-hand side representation)

$$G_+ - G_- = \tfrac{1}{2}e^{r_0\mu_0}\int_{0}^{1}\frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir_0}{\nu}, \mu, \nu\right) \times I_0\{r_0[(1-\mu_0^2)(\nu^{-2}-1)]^{\frac{1}{2}}\} + \tfrac{1}{2}e^{-r_0\mu}\int_{-1}^{0}\frac{d\nu}{\nu^3}\,S^{(2)}\left(\frac{ir_0}{\nu}, -\mu_0, \nu\right) \times I_0\{r_0[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\}. \quad \text{(A18)}$$

This equation can be cast into a more symmetric form by means of the following relations:

$$e^{-r\mu}I_0\{r_0[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\} = \frac{1}{2}\left[S^{(1)}\left(\frac{ir_0}{\nu}, \mu, \nu\right) + S^{(2)}\left(\frac{ir_0}{\nu}, \mu, \nu\right)\right], \quad \text{(A19)}$$

$$e^{r\mu}I_0\{r_0[(1-\mu^2)(\nu^{-2}-1)]^{\frac{1}{2}}\} = \frac{1}{2}\left[S^{(1)}\left(\frac{ir_0}{\nu}, -\mu, \nu\right) + S^{(2)}\left(\frac{ir_0}{\nu}, -\mu, \nu\right)\right] \quad \text{(A20)}$$

$$S^{(1)}\left(-\frac{ir_0}{\nu}, \mu, -\nu\right) = S^{(2)}\left(\frac{ir_0}{\nu}, \mu, \nu\right). \quad \text{(A21)}$$

We may now rewrite Eq. (A8) in the form

$$G_+ - G_- = \frac{1}{4}\int_{-1}^{1}\frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir_0}{\nu}, \mu, \nu\right)S^{(2)}\left(\frac{ir_0}{\nu}, -\mu_0, \nu\right). \quad \text{(A22)}$$

In the right-hand side representations of G_+ and G_- we have, from (A15a), (A14a) and (A15d), and (A14d),

$$G_+ - G_- = (r_0^2\mu)^{-1}\delta(\mu-\mu_0), \quad \mu_0 > 0, \quad \mu > 0, \quad \mu_0 < 0, \quad \mu < 0. \quad \text{(A23)}$$

The full-range completeness theorems may now be readily concluded from Eqs. (A22) and (A23).

Theorem 3 (Full-Range Completeness): For any μ, μ_0,

$$\frac{r_0^2}{4}\int_{-1}^{1}\frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir_0}{\nu}, \mu, \nu\right)S^{(2)}\left(\frac{ir_0}{\nu}, -\mu_0, \nu\right) = \frac{1}{\mu}\,\delta(\mu-\mu_0).$$

Theorem 4 (Full-Range Completeness): For any μ, μ_0,

$$-\tfrac{1}{4}r_0^2\int_{-1}^{1}\frac{d\nu}{\nu^3}\,S^{(1)}\left(\frac{ir_0}{\nu}, -\mu, \nu\right)S^{(2)}\left(\frac{ir_0}{\nu}, \mu_0, \nu\right) = \frac{1}{\mu}\,\delta(\mu-\mu_0).$$

We note that Theorem 1 is adjoint to Theorem 2 in the sense that they imply each other under the reflection of μ and μ_0. In particular, this equivalence also exists under the interchange of μ and μ_0. We may, therefore, combine the two theorems into a symmetric form.

Theorem 5 (Full-Range Completeness—Symmetric form): For any μ, μ_0,

$$\tfrac{1}{8}r_0^2\int_{-1}^{1}\frac{d\nu}{\nu^3}\Big[S^{(1)}\Big(\frac{ir_0}{\nu},\mu,\nu\Big)S^{(2)}\Big(\frac{ir_0}{\nu},-\mu_0,\nu\Big) - S^{(1)}\Big(\frac{ir_0}{\nu},-\mu,\nu\Big)S^{(2)}\Big(\frac{ir_0}{\nu},\mu_0,\nu\Big)\Big] = \frac{1}{\mu}\delta(\mu-\mu_0).$$

For the sake of comparison we present the corresponding trivial theorems for the planar geometry:

$$-\mu\int_{-1}^{0}\frac{d\nu}{\nu}\,\delta(\nu-\mu_0)\delta(\nu-\mu) = \delta(\mu-\mu_0), \qquad \mu_0<0,\quad \mu<0.$$

$$\mu\int_{0}^{1}\frac{d\nu}{\nu}\,\delta(\nu-\mu_0)\delta(\nu-\mu) = \delta(\mu-\mu_0), \qquad \mu_0>0,\quad \mu>0.$$

$$\mu\int_{-1}^{1}\frac{d\nu}{\nu}\,\delta(\nu-\mu_0)\delta(\nu-\mu_0) = \delta(\mu-\mu_0), \qquad \text{for any } \mu,\mu_0.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

On the Theory of Quantum Corrections to the Equations of State and to the Particle-Distribution Functions

G. NIENHUIS
Institute for Theoretical Physics, University of Utrecht, The Netherlands

(Received 20 September 1968)

The first-order quantum corrections to the equations of state of an almost-classical N-particle system are calculated to all orders in the particle density by expanding the normalized Wigner distribution function in powers of $\hbar^2$. In this way one avoids the expansion of the partition function, which has the unsatisfactory property that the correction terms diverge in the thermodynamic limit. Similarly, the first-order quantum correction to the pair distribution function is derived.

1. INTRODUCTION

A well-known concept in the quantum statistical treatment of an N-particle system is the Wigner distribution function (WDF[1]) $f(\mathbf{r},\mathbf{p})$, where $\mathbf{r}$ and $\mathbf{p}$ are the $3N$-dimensional position and momentum vectors. It is defined in such a way that the quantum statistical ensemble average of an arbitrary operator A is given by

$$\langle A\rangle_{\mathrm{qu}} = \mathrm{Tr}\,\rho A/\mathrm{Tr}\,\rho = \int a(\mathbf{r},\mathbf{p})f(\mathbf{r},\mathbf{p})\,d\mathbf{r}\,d\mathbf{p}\Big/\int f(\mathbf{r},\mathbf{p})\,d\mathbf{r}\,d\mathbf{p}. \quad (1)$$

Here ρ is the density operator and $a(\mathbf{r},\mathbf{p})$ represents the classical quantity corresponding to the quantum mechanical operator A. If, in particular, the correspondence between a and A is established according to Weyl's rule (cf., e.g., Ref. 2), one finds for $f(\mathbf{r},\mathbf{p})$ the expression

$$f(\mathbf{r},\mathbf{p}) = (\pi\hbar)^{-3N}\int\rho(\mathbf{r}-\mathbf{y},\mathbf{r}+\mathbf{y})\exp(2i\hbar^{-1}\mathbf{p}\cdot\mathbf{y})\,d\mathbf{y}, \quad (2)$$

where $\rho(\mathbf{r},\mathbf{r}')$ is the density operator in coordinate representation.[3] It will be obvious that ρ and $f(\mathbf{r},\mathbf{p})$ may be multiplied by an arbitrary temperature-dependent factor.

In the case of statistical equilibrium described by a canonical ensemble, the (unnormalized) density operator is given by

$$\rho = \exp(-\beta H), \quad \beta = (kT)^{-1}. \quad (3)$$

This operator satisfies the so-called Bloch equation

$$\frac{\partial\rho}{\partial\beta} = -H\rho. \quad (4)$$

[1] E. Wigner, Phys. Rev. **40**, 749 (1932).
[2] K. Schram and B. R. A. Nijboer, Physica **25**, 733 (1959).
[3] In fact, the correspondence (2) and Weyl's rule are equivalent, as several authors showed independently; see, e.g., Ref. 2, Ref. 5, and Boris Leaf, J. Math. Phys. **9**, 65 (1968).

Theorem 5 (Full-Range Completeness—Symmetric form): For any μ, μ_0,

$$\tfrac{1}{8}r_0^2\int_{-1}^{1}\frac{d\nu}{\nu^3}\Big[S^{(1)}\Big(\frac{ir_0}{\nu},\mu,\nu\Big)S^{(2)}\Big(\frac{ir_0}{\nu},-\mu_0,\nu\Big) - S^{(1)}\Big(\frac{ir_0}{\nu},-\mu,\nu\Big)S^{(2)}\Big(\frac{ir_0}{\nu},\mu_0,\nu\Big)\Big] = \frac{1}{\mu}\delta(\mu-\mu_0).$$

For the sake of comparison we present the corresponding trivial theorems for the planar geometry:

$$-\mu\int_{-1}^{0}\frac{d\nu}{\nu}\,\delta(\nu-\mu_0)\delta(\nu-\mu) = \delta(\mu-\mu_0), \qquad \mu_0<0,\quad \mu<0.$$

$$\mu\int_{0}^{1}\frac{d\nu}{\nu}\,\delta(\nu-\mu_0)\delta(\nu-\mu) = \delta(\mu-\mu_0), \qquad \mu_0>0,\quad \mu>0.$$

$$\mu\int_{-1}^{1}\frac{d\nu}{\nu}\,\delta(\nu-\mu_0)\delta(\nu-\mu_0) = \delta(\mu-\mu_0), \qquad \text{for any } \mu,\mu_0.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

On the Theory of Quantum Corrections to the Equations of State and to the Particle-Distribution Functions

G. NIENHUIS
Institute for Theoretical Physics, University of Utrecht, The Netherlands

(Received 20 September 1968)

The first-order quantum corrections to the equations of state of an almost-classical N-particle system are calculated to all orders in the particle density by expanding the normalized Wigner distribution function in powers of $\hbar^2$. In this way one avoids the expansion of the partition function, which has the unsatisfactory property that the correction terms diverge in the thermodynamic limit. Similarly, the first-order quantum correction to the pair distribution function is derived.

1. INTRODUCTION

A well-known concept in the quantum statistical treatment of an N-particle system is the Wigner distribution function (WDF[1]) $f(\mathbf{r},\mathbf{p})$, where $\mathbf{r}$ and $\mathbf{p}$ are the $3N$-dimensional position and momentum vectors. It is defined in such a way that the quantum statistical ensemble average of an arbitrary operator A is given by

$$\langle A\rangle_{\mathrm{qu}} = \mathrm{Tr}\,\rho A/\mathrm{Tr}\,\rho = \int a(\mathbf{r},\mathbf{p})f(\mathbf{r},\mathbf{p})\,d\mathbf{r}\,d\mathbf{p}\Big/\int f(\mathbf{r},\mathbf{p})\,d\mathbf{r}\,d\mathbf{p}. \quad (1)$$

Here ρ is the density operator and $a(\mathbf{r},\mathbf{p})$ represents the classical quantity corresponding to the quantum mechanical operator A. If, in particular, the correspondence between a and A is established according to Weyl's rule (cf., e.g., Ref. 2), one finds for $f(\mathbf{r},\mathbf{p})$ the expression

$$f(\mathbf{r},\mathbf{p}) = (\pi\hbar)^{-3N}\int\rho(\mathbf{r}-\mathbf{y},\mathbf{r}+\mathbf{y})\exp(2i\hbar^{-1}\mathbf{p}\cdot\mathbf{y})\,d\mathbf{y}, \quad (2)$$

where $\rho(\mathbf{r},\mathbf{r}')$ is the density operator in coordinate representation.[3] It will be obvious that ρ and $f(\mathbf{r},\mathbf{p})$ may be multiplied by an arbitrary temperature-dependent factor.

In the case of statistical equilibrium described by a canonical ensemble, the (unnormalized) density operator is given by

$$\rho = \exp(-\beta H), \quad \beta = (kT)^{-1}. \quad (3)$$

This operator satisfies the so-called Bloch equation

$$\frac{\partial\rho}{\partial\beta} = -H\rho. \quad (4)$$

[1] E. Wigner, Phys. Rev. **40**, 749 (1932).
[2] K. Schram and B. R. A. Nijboer, Physica **25**, 733 (1959).
[3] In fact, the correspondence (2) and Weyl's rule are equivalent, as several authors showed independently; see, e.g., Ref. 2, Ref. 5, and Boris Leaf, J. Math. Phys. **9**, 65 (1968).

The (unnormalized) WDF, which in the sense of (2) corresponds to (3), satisfies the equation

$$\frac{\partial f}{\partial \beta} = -\cos\left\{\frac{\hbar}{2}\left(\frac{\partial}{\partial \mathbf{p}_H}\cdot\frac{\partial}{\partial \mathbf{r}_f} - \frac{\partial}{\partial \mathbf{r}_H}\cdot\frac{\partial}{\partial \mathbf{p}_f}\right)\right\} H_{\mathrm{cl}}(\mathbf{r},\mathbf{p}) f(\mathbf{r},\mathbf{p}), \tag{5}$$

where $H_{\mathrm{cl}}(\mathbf{r},\mathbf{p})$ is the classical Hamiltonian function, and $\partial/\partial \mathbf{r}_f$ means $\partial/\partial \mathbf{r}$ operating on f only, etc. This equation was first derived by Schram and Nijboer[2] and independently by Imre, Özizmir, Rosenbaum, and Zweifel[4] (IORZ).

2. EXPANSION OF THE UNNORMALIZED WDF

In the latter paper, IORZ used Eq. (5) as a starting point to study the quantum corrections to the (thermal) equation of state for gases like H_2, He, and Ne at low (but not too low) temperatures. The authors introduce a function χ by

$$f(\mathbf{r},\mathbf{p}) = \chi(\mathbf{r},\mathbf{p}) \exp(-\beta H_{\mathrm{cl}}) \tag{6}$$

and expand χ in even powers of $\hbar$ according to

$$\chi = 1 + \hbar^2\chi_1 + \hbar^4\chi_2 + \cdots. \tag{7}$$

From (5), one obtains in a straightforward way a differential equation for χ, from which the χ_i can be calculated. If it is assumed that $H_{\mathrm{lc}} = (2m)^{-1}\mathbf{p}^2 + \Phi(\mathbf{r})$, χ_1 is found to be given by[5]

$$\chi_1 = \frac{1}{24m}\left[\frac{\beta^3}{m}\left(\mathbf{p}\cdot\frac{\partial}{\partial \mathbf{r}}\right)^2\Phi - 3\beta^2\frac{\partial^2\Phi}{\partial \mathbf{r}^2} + \beta^3\left(\frac{\partial\Phi}{\partial \mathbf{r}}\right)^2\right]. \tag{8}$$

If now one tries to expand the partition function

$$Z = \mathrm{Tr}\, e^{-\beta H} = \int f(\mathbf{r},\mathbf{p})\, d\mathbf{r}\, d\mathbf{p} \tag{9}$$

again as a power series in $\hbar^2$,

$$Z = Z_{\mathrm{cl}}(1 + \hbar^2 C_1 + \hbar^4 C_2 + \cdots), \tag{10}$$

where

$$Z_{\mathrm{cl}} = \int e^{-\beta H_{\mathrm{cl}}}\, d\mathbf{r}\, d\mathbf{p}, \tag{11}$$

one finds

$$C_1 = \langle\chi_1\rangle_{\mathrm{cl}}\left(= -\frac{\beta^2}{24m}\left\langle\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}}\right), \quad C_2 = \langle\chi_2\rangle_{\mathrm{cl}}, \text{ etc.}, \tag{12}$$

where

$$\langle a(\mathbf{r},\mathbf{p})\rangle_{\mathrm{cl}} = Z_{\mathrm{cl}}^{-1}\int a(\mathbf{r},\mathbf{p}) e^{-\beta H_{\mathrm{cl}}}\, d\mathbf{r}\, d\mathbf{p}. \tag{13}$$

It now turns out, as was noticed by IORZ,[4] that C_1 is of order N, C_2 is of order N^2, etc., so that these correction factors to the partition function blow up for a macroscopic system. The reason is, of course, that the free energy $F = \beta^{-1}\log Z$ should be proportional to N in the thermodynamic limit, so that if we had expanded the quantity $N^{-1}\log Z$ in a power series in $\hbar^2$, the quantum corrections should be expected to be small quantities for an almost-classical system.

Nevertheless, it will be obvious from the theory of power series that, if the expansion of $N^{-1}\log Z$ converges, the expansion (10) will also converge; hence, one may conclude that the coefficients of the expansion of $N^{-1}\log Z$ may be found from those of (10) by formal application of the expansion

$$\log(1 + x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 \cdots, \tag{14}$$

although for large N, we find that $\hbar^2 C_1$, $\hbar^4 C_2$, etc., may be much bigger than 1.

One can now derive the quantum corrections to the equations of state. Expanding the pressure according to

$$p = p_{\mathrm{cl}} + \hbar^2 A_1 + \hbar^4 A_2 + \cdots, \tag{15}$$

one finds from (12)

$$A_1 = \frac{1}{\beta}\frac{\partial}{\partial V}\langle\chi_1\rangle_{\mathrm{cl}}. \tag{16}$$

In order to evaluate this expression from (8) and (13), one should notice that $\langle\chi_1\rangle$ depends on the volume V only through the limits of integration. One then proceeds in the well-known way: Supposing that the volume is a cube with edge L, one makes a change in variables

$$\mathbf{r} = L\mathbf{r}', \quad \frac{\partial}{\partial V} = \frac{1}{3L^2}\frac{\partial}{\partial L}, \tag{17}$$

in such a way that the limits of integration for $\mathbf{r}'$ no longer depend on V. One then finds

$$A_1 = \frac{-\beta}{72mV}\left\langle \mathbf{r}\cdot\frac{\partial}{\partial \mathbf{r}}\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} + \frac{\beta^2}{72mV}\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial \mathbf{r}}\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} - \frac{\beta^2}{72mV}\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial \mathbf{r}}\right\rangle_{\mathrm{cl}}\left\langle\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}}. \tag{18}$$

In the derivation we used the equality

$$\left\langle\frac{\partial^2\Phi}{\mathbf{r}\partial^2}\right\rangle_{\mathrm{cl}} = \beta\left\langle\left(\frac{\partial\Phi}{\partial \mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}}, \tag{19}$$

which follows from partial integration.

For the caloric equation of state, one expands the internal energy

$$U = U_{\mathrm{cl}} + \hbar^2 E_1 + \hbar^4 E_2 + \cdots, \tag{20}$$

[4] K. Imre, E. Özizmir, M. Rosenbaum, and P. F. Zweifel, J. Math. Phys. **8**, 1097 (1967).

[5] The last term of Eq. (8) was omitted by IORZ [cf. their Eq. (A8)], possibly due to a printing error. We may add here that their Eqs. (A24) for the free energy and (A25) for the pressure are also incorrect.

where

$$U = -\frac{\partial \log Z}{\partial \beta}.$$

Applying the equality

$$\frac{\partial}{\partial \beta}\langle h\rangle_{\mathrm{cl}} = \left\langle \frac{\partial h}{\partial \beta}\right\rangle_{\mathrm{cl}} + \langle h\rangle_{\mathrm{cl}}\langle H\rangle_{\mathrm{cl}} - \langle hH\rangle_{\mathrm{cl}}, \quad (21)$$

where $h(\mathbf{r}, \mathbf{p})$ is any function in phase space, one finds from (8) for the first-order correction

$$E_1 = \frac{\beta}{12m}\left\langle \frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} + \frac{\beta^2}{24m}\langle\Phi\rangle_{\mathrm{cl}}\left\langle \frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} - \frac{\beta^2}{24m}\left\langle \Phi\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}}, \quad (22)$$

in perfect agreement with the result of Wigner.[1] As far as we know, the result (18) has not been published before.

Higher-order corrections to the equations of state may be evaluated in the same way, but the calculations become much more complicated.

3. EXPANSION OF THE NORMALIZED WDF

In this section, we wish to point out that an alternative method, namely expansion of the normalized WDF instead of the unnormalized one, avoids the unsatisfactory features of the above derivation and leads to the same results.

The normalized WDF is

$$g(\mathbf{r}, \mathbf{p}) = f(\mathbf{r}, \mathbf{p})\left\{\int f(\mathbf{r}, \mathbf{p})\, d\mathbf{r}\, d\mathbf{p}\right\}^{-1} = Z^{-1}f(\mathbf{r}, \mathbf{p}). \quad (23)$$

It corresponds to $Z^{-1}e^{-\beta H}$ in the sense of Eq. (2) and it satisfies an equation similar to (5):

$$\frac{\partial g}{\partial \beta} = -\cos\left\{\frac{\hbar}{2}\left(\frac{\partial}{\partial \mathbf{p}_H}\cdot\frac{\partial}{\partial \mathbf{r}_g} - \frac{\partial}{\partial \mathbf{r}_H}\cdot\frac{\partial}{\partial \mathbf{p}_g}\right)\right\}H_{\mathrm{cl}}(\mathbf{r}, \mathbf{p})g(\mathbf{r}, \mathbf{p}) + \langle H\rangle_{\mathrm{qu}}g(\mathbf{r}, \mathbf{p}), \quad (24)$$

where $\langle\ \rangle_{\mathrm{qu}}$ is the quantum statistical ensemble average.

Introducing now in a similar way the classical normalized distribution function

$$g_{\mathrm{cl}}(\mathbf{r}, \mathbf{p}) = Z_{\mathrm{cl}}^{-1}\exp(-\beta H_{\mathrm{cl}}),$$

we can define the ratio

$$\psi(\mathbf{r}, \mathbf{p}) = g(\mathbf{r}, \mathbf{p})/g_{\mathrm{cl}}(\mathbf{r}, \mathbf{p}), \quad (25)$$

and, as before, we expand

$$\psi(\mathbf{r}, \mathbf{p}) = 1 + \hbar^2\psi_1 + \hbar^4\psi_2 + \cdots. \quad (26)$$

Now (24) leads immediately to a differential equation for $\psi(\mathbf{r}, \mathbf{p})$, and, considering in particular the term in $\hbar^2$, we have

$$\frac{\partial \psi_1}{\partial \beta} = \frac{1}{8m}\left[\frac{\beta^2}{m}\left(\mathbf{p}\cdot\frac{\partial}{\partial \mathbf{r}}\right)^2\Phi - 2\beta\frac{\partial^2\Phi}{\partial \mathbf{r}^2} + \beta^2\left(\frac{\partial\Phi}{\partial \mathbf{r}}\right)^2\right] + \langle\psi_1 H\rangle_{\mathrm{cl}}. \quad (27)$$

Equation (27) differs from the corresponding equation for χ_1 only in the last term, which does not depend on $(\mathbf{r}, \mathbf{p})$. Therefore

$$\psi_1 = \chi_1 + a(\beta). \quad (28)$$

Equation (27) can be solved directly, but it is easier to use the normalization of g and g_{cl}:

$$1 = \int d\mathbf{r}\, d\mathbf{p} g(\mathbf{r}, \mathbf{p}) = \int d\mathbf{r}\, d\mathbf{p} g_{\mathrm{cl}}(\mathbf{r}, \mathbf{p})\{1 + \hbar^2\psi_1 + \cdots\} = 1 + \hbar^2\langle\psi_1\rangle_{\mathrm{cl}} + \cdots, \quad (29)$$

from which it follows that

$$\langle\psi_1\rangle_{\mathrm{cl}} = 0, \quad \langle\psi_2\rangle_{\mathrm{cl}} = 0, \cdots; \quad (30)$$

then (28) and (30) lead to

$$\psi_1 = \chi_1 - \langle\chi_1\rangle_{\mathrm{cl}}. \quad (31)$$

Higher-order corrections can be calculated in a similar way.[6]

In order to calculate the quantum corrections to the equation of state from the expansion of the normalized distribution function, we make use of the virial theorem, which states that

$$pV = \frac{2}{3}\frac{\langle \mathbf{p}^2\rangle}{2m} - \frac{1}{3}\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial \mathbf{r}}\right\rangle \quad (32)$$

in a system in thermal equilibrium. Equation (32) is valid in quantum statistics as well as in classical statistics. Here we assume that Φ is only the potential of the intermolecular forces. One finds from (32) and (31) the first correction to the equation of state:

$$A_1 = \frac{\beta}{36mV}\left\langle \frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} + \frac{\beta^2}{36mV}\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial \mathbf{r}}\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} - \frac{\beta^2}{72mV}\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial \mathbf{r}}\right\rangle_{\mathrm{cl}}\left\langle \frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} - \frac{\beta^3}{72mV}\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial \mathbf{r}}\left(\frac{\partial\Phi}{\partial \mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}}. \quad (33)$$

One may prove that (33) is equivalent to (18) if one uses the equalities

$$\left\langle \mathbf{r}\cdot\frac{\partial}{\partial \mathbf{r}}\frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} = \frac{\beta}{2}\left\langle \mathbf{r}\cdot\frac{\partial}{\partial \mathbf{r}}\left(\frac{\partial\Phi}{\partial \mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}} - \left\langle \frac{\partial^2\Phi}{\partial \mathbf{r}^2}\right\rangle_{\mathrm{cl}} \quad (34)$$

[6] From the definitions of ψ and χ it follows immediately that $\chi = \psi\langle\chi\rangle_{\mathrm{cl}}$; hence, one concludes that ψ_i can be expressed in χ_j and $\langle\chi_j\rangle_{\mathrm{cl}}$ with $j \leq i$.

and

$$\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial\mathbf{r}}\frac{\partial^2\Phi}{\partial\mathbf{r}^2}\right\rangle_{\mathrm{cl}} = \beta\left\langle \mathbf{r}\cdot\frac{\partial\Phi}{\partial\mathbf{r}}\left(\frac{\partial\Phi}{\partial\mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}} - \left\langle\left(\frac{\partial\Phi}{\partial\mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}} - \frac{1}{2}\left\langle \mathbf{r}\cdot\frac{\partial}{\partial\mathbf{r}}\left(\frac{\partial\Phi}{\partial\mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}}, \quad (35)$$

the validity of which can be shown by partial integration. From (31) one calculates at once the correction of order $\hbar^2$ to the caloric equation of state; one finds

$$E_1 = \frac{\beta}{24m}\left\langle\frac{\partial^2\Phi}{\partial\mathbf{r}^2}\right\rangle_{\mathrm{cl}} - \frac{\beta^2}{12m}\left\langle\Phi\frac{\partial^2\Phi}{\partial\mathbf{r}^2}\right\rangle_{\mathrm{cl}} + \frac{\beta^2}{24m}\langle\Phi\rangle_{\mathrm{cl}}\left\langle\frac{\partial^2\Phi}{\partial\mathbf{r}^2}\right\rangle_{\mathrm{cl}} + \frac{\beta^3}{24m}\left\langle\Phi\left(\frac{\partial\Phi}{\partial\mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}}. \quad (36)$$

Equation (36) is equivalent to (22), as it should be; this is seen by using the equality

$$\left\langle\Phi\left(\frac{\partial\Phi}{\partial\mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}} = \frac{1}{\beta}\left\langle\left(\frac{\partial\Phi}{\partial\mathbf{r}}\right)^2\right\rangle_{\mathrm{cl}} + \frac{1}{\beta}\left\langle\Phi\frac{\partial^2\Phi}{\partial\mathbf{r}^2}\right\rangle_{\mathrm{cl}}. \quad (37)$$

We wish to emphasize that (18) [or (33)] and (22) [or (36)] give the corrections to the equations of state to all orders in the particle density. Therefore, they are valid also for dense systems.[7]

If $\Phi(\mathbf{r})$ can be written as a sum of two-particle interactions $v(ij) = v(\mathbf{r}_i - \mathbf{r}_j)$, A_1 and E_1 can be expressed in terms of the reduced classical particle distribution functions

$$n_s(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_s) = \frac{N!}{(N-s)!}\int g_{\mathrm{cl}}(\mathbf{r}, \mathbf{p})\, d\mathbf{r}_{s+1}\, d\mathbf{r}_{s+2}\cdots d\mathbf{r}_N\, d\mathbf{p}. \quad (38)$$

One finds from (18), if we write $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$,

$$\begin{aligned} A_1 = &-\frac{\beta}{72mV}\int n_2(\mathbf{r}_1, \mathbf{r}_2)\mathbf{r}_{12}\cdot\frac{\partial}{\partial\mathbf{r}_{12}}\frac{\partial^2 v(12)}{\partial\mathbf{r}_{12}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2 \\ &+\frac{\beta^2}{72mV}\left(\int n_2(\mathbf{r}_1, \mathbf{r}_2)\mathbf{r}_{12}\cdot\frac{\partial v(12)}{\partial\mathbf{r}_{12}}\frac{\partial^2 v(12)}{\partial\mathbf{r}_{12}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2\right. \\ &+2\int n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)\mathbf{r}_{12}\cdot\frac{\partial v(12)}{\partial\mathbf{r}_{12}}\frac{\partial^2 v(13)}{\partial\mathbf{r}_{13}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 \\ &+\tfrac{1}{2}\int n_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)\mathbf{r}_{12} \\ &\cdot\frac{\partial v(12)}{\partial\mathbf{r}_{12}}\frac{\partial^2 v(34)}{\partial\mathbf{r}_{34}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4 \\ &-\tfrac{1}{2}\int n_2(\mathbf{r}_1, \mathbf{r}_2)\mathbf{r}_{12}\cdot\frac{\partial v(12)}{\partial\mathbf{r}_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2 \\ &\left.\times\int n_2(\mathbf{r}_3, \mathbf{r}_4)\frac{\partial^2 v(34)}{\partial\mathbf{r}_{34}^2}\, d\mathbf{r}_3\, d\mathbf{r}_4\right). \end{aligned} \quad (39)$$

[7] One should realize, however, that for dense systems the effects of the symmetry conditions on the wavefunctions may be important. These effects are neglected here.

In the same way one finds from (22)

$$\begin{aligned} E_1 = &\frac{\beta}{12m}\int n_2(\mathbf{r}_1, \mathbf{r}_2)\frac{\partial^2 v(12)}{\partial\mathbf{r}_{12}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2 \\ &-\frac{\beta^2}{24m}\left(\int n_2(\mathbf{r}_1, \mathbf{r}_2)v(12)\frac{\partial^2 v(12)}{\partial\mathbf{r}_{12}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2\right. \\ &+2\int n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)v(12)\frac{\partial^2 v(13)}{\partial\mathbf{r}_{13}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 \\ &+\tfrac{1}{2}\int n_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)v(12)\frac{\partial^2 v(34)}{\partial\mathbf{r}_{34}^2}\, d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4 \\ &-\tfrac{1}{2}\int n_2(\mathbf{r}_1, \mathbf{r}_2)v(12)\, d\mathbf{r}_1\, d\mathbf{r}_2 \\ &\left.\times\int n_2(\mathbf{r}_3, \mathbf{r}_4)\frac{\partial^2 v(34)}{\partial\mathbf{r}_{34}^2}\, d\mathbf{r}_3\, d\mathbf{r}_4\right). \end{aligned} \quad (40)$$

The last two terms in (39) are proportional to V in the thermodynamic limit, but taken together they have a finite limit. This is easily seen if one introduces a set of functions $k_s(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_s)$ in the following way:

$$\begin{aligned} n_1(\mathbf{r}_1) &= \rho k_1(\mathbf{r}_1), \\ n_2(\mathbf{r}_1, \mathbf{r}_2) &= \rho^2[k_2(\mathbf{r}_1, \mathbf{r}_2) + k_1(\mathbf{r}_1)k_1(\mathbf{r}_2)], \\ n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) &= \rho^3[k_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) + k_1(\mathbf{r}_1)k_2(\mathbf{r}_2, \mathbf{r}_3) \\ &\quad + k_1(\mathbf{r}_2)k_2(\mathbf{r}_1, \mathbf{r}_3) + k_1(\mathbf{r}_3)k_2(\mathbf{r}_1, \mathbf{r}_2) \\ &\quad + k_1(\mathbf{r}_1)k_1(\mathbf{r}_2)k_1(\mathbf{r}_3)], \cdots. \end{aligned} \quad (41)$$

Then the functions k_s have the cluster property: If the positions $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_s$ can be divided in at least two groups such that positions in different groups are further apart than the range of the interaction between the particles, then $k_s = 0$. One may say that $k_s(\mathbf{r}_1, \cdots, \mathbf{r}_s)$ is zero unless all $\mathbf{r}_i$, $i = 1, 2, \cdots, s$, are close to each other.

If one substitutes (41) into (39), one easily verifies that the terms which are of order V in the thermodynamic limit cancel each other. In the same way one sees that the sum of the last two terms of (40), which are of order V^2 in the thermodynamic limit, gives a contribution to E_1 which is of order V.

It is well known that the first term of the expansion of $n_s(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_s)$ in powers of the density is

$$\rho^s \exp\left[-\beta\sum_{i\le 1<j\le s} v(ij)\right]. \quad (42)$$

Therefore, from (39), one finds for dilute systems that

$$\begin{aligned} A_1^{(a)} &= -\frac{\beta\rho^2}{72m}\int\exp[-\beta v(\mathbf{r})]\mathbf{r}\cdot\frac{\partial}{\partial\mathbf{r}}\frac{\partial^2 v(\mathbf{r})}{\partial\mathbf{r}^2}\, d_3\mathbf{r} \\ &\quad +\frac{\beta^2\rho^2}{72m}\int\exp[-\beta v(\mathbf{r})]\mathbf{r}\cdot\frac{\partial v(\mathbf{r})}{\partial\mathbf{r}}\frac{\partial^2 v(\mathbf{r})}{\partial\mathbf{r}^2}\, d_3\mathbf{r} \\ &= \frac{\beta\rho^2}{24m}\int\exp[-\beta v(\mathbf{r})]\frac{\partial^2 v}{\partial\mathbf{r}^2}\, d_3\mathbf{r}. \end{aligned} \quad (43)$$

$A_1^{(d)}$ is the first-order quantum correction to the equation of state to second order in the density, and is therefore the first quantum correction to the second virial coefficient. Equation (43) is in agreement with the results of Uhlenbeck and Beth[8] and of Green.[9]

In the same way one obtains for dilute systems

$$E_1^{(d)} = \frac{\beta V}{12m}\int \exp\left[-\beta v(\mathbf{r})\right]\frac{\partial^2 v}{\partial \mathbf{r}^2}\,d_3\mathbf{r} - \frac{\beta^2 V}{24m}\int \exp\left[-\beta v(\mathbf{r})\right]v(\mathbf{r})\frac{\partial^2 v}{\partial \mathbf{r}^2}\,d_3\mathbf{r}, \quad (44)$$

in agreement with the result of Green.[9]

The expansion of the normalized WDF can be used to evaluate quantum corrections to the ensemble average of any quantity which is a function of $(\mathbf{r}, \mathbf{p})$, and also to the reduced particle distribution functions.

As an example, we derive the first quantum correction to the pair-distribution function. Expanding

$$N_2(\mathbf{r}_1, \mathbf{r}_2) = n_2(\mathbf{r}_1, \mathbf{r}_2) + \hbar^2 b_1(\mathbf{r}_1, \mathbf{r}_2) + \cdots, \quad (45)$$

where[10]

$$N_2(\mathbf{r}_1, \mathbf{r}_2) = N(N-1)\int g(\mathbf{r}, \mathbf{p})\,d\mathbf{r}_3\,d\mathbf{r}_4 \cdots d\mathbf{r}_N\,d\mathbf{p}, \quad (46)$$

one finds (to all orders in the particle density)

$$b_1(\mathbf{r}_1, \mathbf{r}_2) = \left\{\frac{\beta^2}{24m}\left[\beta\left(\frac{\partial v_{12}}{\partial \mathbf{r}_1}\right)^2 - 2\frac{\partial^2 v_{12}}{\partial \mathbf{r}_1^2} + \tfrac{1}{2}\int n_2(\mathbf{r}_3, \mathbf{r}_4)\frac{\partial^2 v_{34}}{\partial \mathbf{r}_3^2}\,d\mathbf{r}_3\,d\mathbf{r}_4\right] n_2(\mathbf{r}_1, \mathbf{r}_2)\right.$$

$$+ \frac{\beta^2}{12m}\int n_3(r_1, r_2, r_3)\left[\beta\frac{\partial v_{12}}{\partial \mathbf{r}_1}\frac{\partial v_{13}}{\partial \mathbf{r}_1} + \beta\left(\frac{\partial v_{13}}{\partial \mathbf{r}_1}\right)^2 + \tfrac{1}{2}\beta\frac{\partial v_{31}}{\partial \mathbf{r}_3}\frac{\partial v_{32}}{\partial \mathbf{r}_3} - 2\frac{\partial^2 v_{13}}{\partial \mathbf{r}_1^2}\right] d\mathbf{r}_3$$

$$+ \frac{\beta^2}{24m}\int n_4(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)\left[\beta\frac{\partial v_{13}}{\partial \mathbf{r}_1}\frac{\partial v_{14}}{\partial \mathbf{r}_1} + \tfrac{1}{2}\beta\left(\frac{\partial v_{34}}{\partial \mathbf{r}_3}\right)^2 + 2\beta\frac{\partial v_{31}}{\partial \mathbf{r}_3}\frac{\partial v_{34}}{\partial \mathbf{r}_3} - \frac{\partial^2 v_{34}}{\partial \mathbf{r}_3^2}\right] d\mathbf{r}_3\,d\mathbf{r}_4$$

$$\left. + \frac{\beta^3}{48m}\int n_5(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4, \mathbf{r}_5)\frac{\partial v_{34}}{\partial \mathbf{r}_3}\frac{\partial v_{35}}{\partial \mathbf{r}_3}\,d\mathbf{r}_3\,d\mathbf{r}_4\,d\mathbf{r}_5\right\} + \{\text{I.T.}\}, \quad (47)$$

where {I.T.} is equal to the first term between the braces with $\mathbf{r}_1$ and $\mathbf{r}_2$ interchanged. Applying (41) again, one shows that the terms, which are of order V in the thermodynamic limit, compensate each other.

ACKNOWLEDGMENTS

As a conclusion I would like to thank Dr. A. Rahman (Argonne National Laboratory, Argonne, Illinois) for having drawn my attention to the difficulties connected with the expansion of the partition function and Professor B. R. A. Nijboer (Institute for Theoretical Physics, Utrecht, The Netherlands) for his guidance and advice. This investigation is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie" and was made possible by financial support from the "Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek."

[8] G. E. Uhlenbeck and E. Beth, Physica **3**, 729 (1936); **4**, 915 (1937).

[9] H. S. Green, J. Chem. Phys. **19**, 955 (1951).

[10] One should notice that (46) has the same physical meaning as the classical two-particle distribution function; it represents the Slater sum integrated over $N-2$ coordinates.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Functional Formulation of the Coherence Theory

ANDRZEJ ZARDECKI
Institute of Physics, The Warsaw Technical University, Warsaw, Poland*

(Received 5 August 1968; Revised Manuscript Received 2 May 1969)

A functional formalism for coherence theory is developed in terms of the (quasi) probability functional of the radiation field. The formalism is closely connected to the P representation of the density operator. For the fields confined within a spatial volume of finite size, this connection is expressed as a multidimensional Fourier integral. A certain type of correspondence, illustrated by explicit examples, between the classical and the quantum mechanical description of the field statistics is established. The recent results of Titulaer and Glauber concerning the (1, 1)th-order coherent fields are analyzed from the point of view of the functional formulation.

1. INTRODUCTION

The optical coherence theory is concerned with the statistical description of fluctuations of the electromagnetic field. Optical coherence phenomena may be said to be manifestations of correlations between these fluctuations.[1] In order to study correlations in optical fields in a systematic way, a general statistical description of the field is needed. This has been provided both in classical terms[2] and in terms of the quantized field.[3–5] A classical approach to specifying the randomness of radiation consists in defining the multifold probability distribution w_N, from which all moments at N space–time points $\mathbf{r}_i$, t_i can be calculated. When we describe the electromagnetic field in quantum-mechanical terms, we must think of the field vectors as operators which satisfy the Maxwell equations. Since the field, in general, is in a mixed state, the averaging must be performed by means of the density operator ρ, which describes the field completely. It has been found especially convenient to make use of the set of coherent states in describing the quantum state of the field. A particular representation of the density operator, in terms of products of the coherent-state vectors, is called the P representation (or the diagonal representation in Sudarshan's terminology[6]). The use of the P representation in describing fields brings out a correspondence between the results of quantum electrodynamics and classical theory. The analogy between both types of predictions can be made even closer by introducing configuration-space probability densities[7] which correspond to any given form of the P representation. In this connection an important question arises whether this correspondence, persisting throughout the theory of coherence, may be regarded as demonstrating the complete equivalence of the classical and quantum-mechanical approaches. Although explicit calculations relating to black-body radiation[8] seem to support this statement, we should bear in mind, however, that the field described by the density operator in the P representation does not necessarily have a classical analog.[9] Another controversy is related to the question of universal validity of the P representation. Mehta and Sudarshan[10] express the density operator as the limit of an infinity sequence of P representations. Statistical averages may then be evaluated for any density operator by carrying out an appropriate limiting procedure. It is, however, stressed by Mollow and Glauber[11] that the usefulness of this approach is not clear. It is not our intention to resolve this problem here. All the results presented in this paper will be qualified by the assumption that the P representation exists.

The probability distribution w_N, whether defined classically or derived quantum-mechanically from the density operator in the P representation, contains, as $N \to \infty$, all the information we possess about the random electromagnetic field. In the limit $N \to \infty$, w_N becomes a probability functional $W[\mathcal{E}(\mathbf{r}, t)] = W[\mathcal{E}]$. The problem of determining the equations governing the probability functional W of a random field or the characteristic functional Φ (a functional Fourier transform of W) has been extensively developed in the theory of turbulence.[12–14] An attempt to formulate the coherence theory in terms of the functional

* Warsaw, Koszykowa 75

[1] L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

[2] E. Wolf, *Proceedings of the Symposium on Optical Masers* (John Wiley & Sons, Inc., New York, 1963).

[3] R. J. Glauber, Phys. Rev. Letters **10**, 84 (1963).

[4] R. J. Glauber, Phys. Rev. **130**, 2529 (1963).

[5] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).

[6] E. C. G. Sudarshan, Phys. Rev. Letters **10**, 277 (1963).

[7] See Ref. 1, p. 247.

[8] C. L. Mehta and E. Wolf, Phys. Rev. **134A**, 1143, 1149 (1964).

[9] This takes place when the singularities of the function $P(\{\alpha_k\})$ specifying the P representation are of types stronger than those of δ functions; cf. Ref. 5, Footnote 11.

[10] C. J. Mehta and E. C. G. Sudarshan, Phys. Rev. **138**, B274 (1965).

[11] B. R. Mollow and R. J. Glauber, Phys. Rev. **160**, 1076 (1967).

[12] E. Hopf, J. Ratl. Mech. **1**, 87 (1952).

[13] R. M. Lewis and R. M. Kraichnan, Commun. Pure. Appl. Math. **15**, 397 (1962).

[14] A. S. Monin and A. M. Yaglom, *Statistical Hydrodynamics* (Moscow Publishing House, Moscow, 1967) (in Russian).

equations is indicated in the book by Beran and Parrent.[15] Characteristic functionals have been used for different purposes by Glauber,[16] Keller,[17] and Mehta and Sudarshan.[10] In this study a unified functional formalism for coherence is proposed.

Since the probability functional represents the probability of observation of the field as a whole, it certainly gives the over-all statistical characteristics of the field. Once it is known, all correlation functions can be evaluated. Vice versa, if the set of correlation functions is specified, we face the moment problem of finding the probability functional on the basis of its moments. It is worth mentioning that, since the positive- and negative-frequency parts of the field are to be regarded as independent of each other, the (N, M)-type correlation functions, which contain unequal numbers of creation and annihilation operators, must be taken into account. Only the totality of these functions forms the complete set of moments of the probability functional.

The purpose of this paper is to show that the knowledge of the density operator in the P representation enables one to construct the probability functional of the electric field. The most convenient way to do this is to deal with a discrete set of variables rather with a continuum of them. This is attained by assuming that the field is confined within a spatial volume of finite size. All functional integrals can then be reduced to multiple integrals of denumerably infinite type. In this case the characteristic functional Φ appears to be a multiple Fourier transform of the function $P(\{\alpha_k\})$, specifying the P representation. When the characteristic functional is found, the P representation may be rederived. On the other hand, if the characteristic functional is given in classical terms, then, using the Fourier theorem, the P representation can in principle be derived. In this sense, it may be said that apart from the possible negativeness of the probability functional the correspondence between the classical and quantum description is one-to-one. It should be stressed that the inverse problem of determining the density matrix for a field when its characteristic functional is known has already been solved by Keller[17] from a different standpoint.

The general methods of the outlined procedure are illustrated in Sec. 3 by some specific examples where the P representations corresponding to Gaussian fields, an ideal laser field, and fully coherent fields are considered. In Sec. 4 the probability functional for the (1, 1)th-order[18] coherent field is derived.

2. PROBABILITY FUNCTIONAL OF THE FIELD

The classical theory of coherence may be described in terms of a probability distribution W (probability functional) in the space of positive-frequency part $\mathcal{E}_\mu(x)$ and negative-frequency part $\mathcal{E}_\mu^*(x)$ of electric vector fields. Here x stands for the space–time point $\mathbf{r}$, t, and $\mu = 1, 2, 3$ labels the Cartesian components of the field. We employ the summation convention over repeated indices. If $\zeta_\mu(x)$ is an arbitrary complex vector field whose Fourier expansion contains only positive frequency amplitudes and $\zeta_\mu^*(x)$ is its complex conjugate, we define real scalar products as

$$\begin{aligned}(\zeta, \mathcal{E}^*) &= \lim_{T\to\infty} \frac{1}{2T} \int_V \int_{-T}^{T} \zeta_\mu(x)\mathcal{E}_\mu^*(x)\, d\mathbf{r}\, dt \\ &\equiv \int \zeta_\mu(x)\mathcal{E}_\mu^*(x)\, dx, \\ (\zeta^*, \mathcal{E}) &= \lim_{T\to\infty} \frac{1}{2T} \int_V \int_{-T}^{T} \zeta_\mu^*(x)\mathcal{E}_\mu(x)\, d\mathbf{r}\, dt \\ &\equiv \int \zeta_\mu^*(x)\mathcal{E}_\mu(x)\, dx, \end{aligned} \tag{2.1}$$

where the domain V of $\mathbf{r}$ is a certain bounded volume of space. The characteristic functional of W is defined by

$$\begin{aligned}\Phi[\zeta, \zeta^*] &= \langle e^{i(\zeta,\mathcal{E}^*)+i(\zeta^*,\mathcal{E})}\rangle \\ &= \int W[\mathcal{E}^*, \mathcal{E}] e^{i(\zeta,\mathcal{E}^*)+i(\zeta^*,\mathcal{E})}\, d\mathcal{E}^*\, d\mathcal{E}, \end{aligned} \tag{2.2}$$

where the integration is over the entire space of electric vector fields lying inside the considered volume.[19]

To gain an insight into the physical meaning of the probability functional, imagine space–time to be divided into discrete cells centered at $x_1, \cdots, x_N$. If we suppress tensor indices, the values of the functions $\mathcal{E}^*(x)$ and $\mathcal{E}(x)$ at those particular points

$$\begin{aligned}&\mathcal{E}^*(x_1), \quad \mathcal{E}(x_1), \cdots, \mathcal{E}^*(x_N), \\ &\mathcal{E}(x_N) = \mathcal{E}_1^*, \quad \mathcal{E}_1, \cdots, \mathcal{E}_N^*, \mathcal{E}_N\end{aligned}$$

are analogous to the variables of a multivariable distribution function. In the limit, as $N \to \infty$ and each cell contains only one point, the probability of obtaining a particular set of values $\mathcal{E}_1^*\mathcal{E}_1, \cdots, \mathcal{E}_N^*\mathcal{E}_N$ in the range $d\mathcal{E}_1^*\, d\mathcal{E}_1, \cdots, d\mathcal{E}_N^*\, d\mathcal{E}_N$ goes over into

[15] M. J. Beran and G. B. Parrent, *Theory of Partial Coherence* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964).

[16] R. J. Glauber, in *Quantum Optics and Electronics*, C. deWitt *et al.*, Eds. (Gordon and Breach Science Publishers, Inc., New York, 1965).

[17] E. F. Keller, Phys. Rev. **139**, B202 (1965).

[18] The first order in Glauber's terminology.

[19] The integral in (2.2) is to be understood as a functional integral. The technique of functional integration is supposed to be known. For the details see, e.g., K. O. Friedrichs, *Mathematical Aspects of the Quantum Field Theory* (Interscience Publishers, Inc., New York, 1953); K. Symanzik, Z. Naturforsch. **9a**, 809 (1954); R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Co., New York, 1965).

probability of observing particular functions $\mathcal{E}^*\mathcal{E}$ in the range $d\mathcal{E}^*\, d\mathcal{E}$. In a similar manner, the functional integral on the right-hand side of (2.2) can be thought of as a symbolic generalization of

$$\Phi(\zeta_1\zeta_1^*, \cdots, \zeta_N\zeta_N^*) = \int W(\mathcal{E}_1^*\mathcal{E}_1, \cdots, \mathcal{E}_N^*\mathcal{E}_N) \times \exp\left[i\sum_{k=1}^{N}(\zeta_k\mathcal{E}_k^* + \zeta_k^*\mathcal{E}_k)\right] d\mathcal{E}_1^*\, d\mathcal{E}_1 \cdots d\mathcal{E}_N^*\, d\mathcal{E}_N$$

to infinitely many variables. By functional differentiation of (2.2) we obtain

$$\frac{\delta^{N+M}\Phi}{\delta\zeta_{\mu_1}(x_1)\cdots\delta\zeta_{\mu_N}(x_N)\delta\zeta^*_{\mu_{N+1}}(x_{N+1})\cdots\delta\zeta^*_{\mu_{N+M}}(x_{N+M})} = \langle i^{N+M}\mathcal{E}^*_{\mu_1}(x_1)\cdots\mathcal{E}^*_{\mu_N}(x_N)\mathcal{E}_{\mu_{N+1}} \times (x_{N+1})\cdots\mathcal{E}_{\mu_{N+M}}(x_{N+M})e^{i(\zeta\mathcal{E}^*)+i(\zeta^*\mathcal{E})}\rangle. \quad (2.3)$$

Hence the (N, M)th-order correlation function is

$$G^{(N,M)}_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}) = i^{-(N+M)}\frac{\delta^{N+M}\Phi}{\delta\zeta_{\mu_1}(x_1)\cdots\delta\zeta_{\mu_N}(x_N)\delta\zeta^*_{\mu_{N+1}}(x_{N+1})\cdots\delta\zeta^*_{\mu_{N+M}}(x_{N+M})}\bigg|_{\zeta=\zeta^*=0}. \quad (2.4)$$

A formal representation for $\Phi[\zeta\zeta^*]$, in terms of the correlation functions, is the Volterra expansion[20]

$$\Phi[\zeta, \zeta^*] = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty}\frac{i^{N+M}}{N!\,M!}\tilde{G}^{(N,M)}[\zeta, \zeta^*], \quad (2.5)$$

where

$$\tilde{G}^{(N,M)}[\zeta, \zeta^*] = \int G^{(N,M)}_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}) \times \zeta_{\mu_1}(x_1)\cdots\zeta_{\mu_N}(x_N) \times \zeta^*_{\mu_{N+1}}(x_{N+1})\cdots\zeta^*_{\mu_{N+M}}(x_{N+M})\, dx_1\cdots dx_{N+M} \quad (2.6)$$

is the Wightman-type functional[21] of the (N, M)th order. By the Fourier inversion formula, it follows from (2.2) that

$$W[\mathcal{E}^*, \mathcal{E}] = \int\Phi[\zeta, \zeta^*]e^{-i[(\zeta,\mathcal{E}^*)+(\zeta^*,\mathcal{E})]}\, d\left(\frac{\zeta}{2\pi}\right) d\left(\frac{\zeta^*}{2\pi}\right). \quad (2.7)$$

Hence, by virtue of (2.5), we get

$$W[\mathcal{E}^*, \mathcal{E}] = \sum_{N=0}^{\infty}\sum_{M=0}^{\infty}\frac{i^{N+M}}{N!\,M!}\int\tilde{G}^{(N,M)}[\zeta, \zeta^*] \times e^{-i[(\zeta,\mathcal{E}^*)+(\zeta^*,\mathcal{E})]}\, d\left(\frac{\zeta}{2\pi}\right) d\left(\frac{\zeta^*}{2\pi}\right). \quad (2.8)$$

Therefore, W can be considered as a functional not only of $\mathcal{E}^*$ and $\mathcal{E}$ but also of all correlation functions $G^{(N,M)}$, i.e.,

$$W = W\{\mathcal{E}^*(x), \mathcal{E}(x); [G^{(N,M)}_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}} \times (x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M})]\}. \quad (2.9)$$

In the quantum description, the electric field operator $E_\mu(x)$ is analogously separated into its positive-frequency part $E^{(+)}_\mu(x)$ and its negative-frequency part $E^{(-)}_\mu(x)$, which is a Hermitian adjoint of $E^{(+)}_\mu$. The operator $E^{(+)}_\mu$ has right eigenstates of the form

$$E^{(+)}_\mu(x)\,|\{\alpha_k\}\rangle = \mathcal{E}_\mu(x, \{\alpha_k\})\,|\{\alpha_k\}\rangle \quad (2.10)$$

and the corresponding eigenvalue function $\mathcal{E}_\mu$ is a linear form in the variables $\{\alpha_k\}$,

$$\mathcal{E}_\mu(x, \{\alpha_k\}) = \sum_k e_\mu(x, k)\alpha_k, \quad (2.11)$$

where

$$e_\mu(x, k) = i(\tfrac{1}{2}\hbar\omega_k)^{\frac{1}{2}}u_{k\mu}(\mathbf{r})e^{-i\omega_k t}. \quad (2.12)$$

The state $|\{\alpha_k\}\rangle \equiv \prod_k |\alpha_k\rangle_k$ is a direct product of the individual modes, eigenstates of the annihilation operators a_k,

$$a_k\,|\alpha_k\rangle_k = \alpha_k\,|\alpha_k\rangle_k. \quad (2.13)$$

Given now the P representation of the density operator

$$\rho = \int P(\{\alpha_k\})\,|\{\alpha_k\}\rangle\langle\{\alpha_k\}|\, d^2\{\alpha_k\}, \quad (2.14)$$

the (N, M)th-order correlation function, defined by

$$G^{(N,M)}_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}) = \mathrm{Tr}\,\{\rho E^{(-)}_{\mu_1}(x_1)\cdots E^{(-)}_{\mu_N}(x_N) \times E^{(+)}_{\mu_{N+1}}(x_{N+1})\cdots E^{(+)}_{\mu_{N+M}}(x_{N+M})\}, \quad (2.15)$$

can be written as

$$G^{(N,M)}_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}) = \int P(\{\alpha_k\})\mathcal{E}^*_{\mu_1}(x_1, \{\alpha_k\})\cdots\mathcal{E}^*_{\mu_N}(x_N, \{\alpha_k\}) \times \mathcal{E}_{\mu_{N+1}}(x_{N+1}, \{\alpha_k\})\cdots\mathcal{E}_{\mu_{N+M}}(x_{N+M}, \{\alpha_k\})\, d^2\{\alpha_k\}. \quad (2.16)$$

[20] V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations* (Dover Publications, Inc., New York, 1959).
[21] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

If we introduce a configuration-space probability density

$$w^{(N,M)}\{\mathcal{E}^*_{\mu_1}(x_1), \cdots, \mathcal{E}^*_{\mu_N}(x_N); \mathcal{E}_{\mu_{N+1}}(x_{N+1}), \cdots, \mathcal{E}_{\mu_{N+M}}(x_{N+M})\}$$
$$= \int P(\{\alpha_k\}) \prod_{j=1}^{N} \delta\left\{\mathcal{E}^*_{\mu_j}(x_j) - \sum_k e^*_{\mu_j}(x_j, k)\alpha^*_k\right\} \times \prod_{j=N+1}^{N+M} \delta\left\{\mathcal{E}_{\mu_j}(x_j) - \sum_k e_{\mu_j}(x_j, k)\alpha_k\right\} d^2\{\alpha_k\}, \quad (2.17)$$

which is the distribution function corresponding to $P(\{\alpha_k\})$, the correlation function can be expressed in the following form:

$$G^{(N,M)}_{\mu_1 \cdots \mu_N \mu_{N+1} \cdots \mu_{N+M}}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M})$$
$$= \int w^{(N,M)}\{\mathcal{E}^*_{\mu_1}(x_1), \cdots, \mathcal{E}^*_{\mu_N}(x_N); \mathcal{E}_{\mu_{N+1}}(x_{N+1}), \cdots, \mathcal{E}_{\mu_{N+M}}(x_{N+M})\}$$
$$\times \mathcal{E}^*_{\mu_1}(x_1) \cdots \mathcal{E}^*_{\mu_N}(x_N)\mathcal{E}_{\mu_{N+1}}(x_{N+1}) \cdots \mathcal{E}_{\mu_{N+M}}(x_{N+M})$$
$$\times d\mathcal{E}^*_{\mu_1}(x_1) \cdots d\mathcal{E}^*_{\mu_N}(x_N)\, d\mathcal{E}_{\mu_{N+1}}$$
$$\times (x_{N+1}) \cdots d\mathcal{E}_{\mu_{N+M}}(x_{N+M}). \quad (2.18)$$

It should be noted that the distribution function (2.17) gives the joint probability of the functions

$$\mathcal{E}^*_{\mu_1}(x_1, \{\alpha_k\}), \cdots, \mathcal{E}^*_{\mu_N}(x_N, \{\alpha_k\}); \quad \mathcal{E}_{\mu_{N+1}}(x_{N+1}, \{\alpha_k\}), \cdots, \mathcal{E}_{\mu_{N+M}}(x_{N+M}, \{\alpha_k\})$$

taking on the values

$$\mathcal{E}^*_{\mu_1}(x_1), \cdots, \mathcal{E}^*_{\mu_N}(x_N); \quad \mathcal{E}_{\mu_{N+1}}(x_{N+1}), \cdots, \mathcal{E}_{\mu_{N+M}}(x_{N+M})$$

at $N + M$ fixed space–time points $x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}$. Since we consider $\mathcal{E}^*$ and $\mathcal{E}$ as independent functions, introduced in place of real and imaginary parts Re ($\mathcal{E}$) and Im ($\mathcal{E}$) of the field, one-dimensional δ functions of independent arguments α_k and α^*_k introduced in place of Re (α_k) and Im (α_k) are used, but otherwise these distribution functions are identical with those derived by Glauber for the field amplitudes at one and two space–time points.[22]

The correlation function can be also written in terms of the probability functional W as a functional integral,

$$G^{(N,M)}_{\mu_1 \cdots \mu_N \mu_{N+1} \cdots \mu_{N+M}}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M})$$
$$= \int W[\mathcal{E}^*, \mathcal{E}]\mathcal{E}^*_{\mu_1}(x_1) \cdots \mathcal{E}^*_{\mu_N}(x_N) \times \mathcal{E}_{\mu_{N+1}}(x_{N+1}) \cdots \mathcal{E}_{\mu_{N+M}}(x_{N+M})\, d\mathcal{E}^*\, d\mathcal{E}, \quad (2.19)$$

where the probability functional is given by

$$W[\mathcal{E}^*, \mathcal{E}] = \int P(\{\alpha_k\})\delta\left[\mathcal{E}^*(x) - \sum_k e^*(x, k)\alpha^*_k\right] \times \delta\left[\mathcal{E}(x) - \sum_k e(x, k)\alpha_k\right] d^2\{\alpha_k\}. \quad (2.20)$$

Here $\delta[\mathcal{E}^*]$ and $\delta[\mathcal{E}]$ are the δ functionals which can be regarded as generalizations of $\delta(\mathcal{E}^*_1, \cdots, \mathcal{E}^*_N)$ and $\delta(\mathcal{E}_1, \cdots, \mathcal{E}_N)$ to infinitely many variables. They can be expressed as the functional Fourier transforms of unity,

$$\delta[\mathcal{E}] = \int e^{i(\zeta^*, \mathcal{E})}\, d\left(\frac{\zeta^*}{2\pi}\right),$$
$$\delta[\mathcal{E}^*] = \int e^{i(\zeta, \mathcal{E}^*)}\, d\left(\frac{\zeta}{2\pi}\right).$$

The meaning of functional integration in (2.19) is the same as in (2.2). Inserting (2.20) in (2.2) we obtain

$$\Phi[\zeta, \zeta^*] = \int P(\{\alpha_k\}) \exp\left[i \sum_k (\alpha^*_k\zeta_k + \alpha_k\zeta^*_k)\right] d^2\{\alpha_k\}, \quad (2.21)$$

where

$$\zeta_k = (\zeta, e^*(x, k)), \quad \zeta^*_k = (\zeta^*, e(x, k)). \quad (2.22)$$

Hence the characteristic functional, which in this case can be written as $\Phi[\zeta, \zeta^*] = \Phi(\{\zeta_k\})$, is simply a multidimensional Fourier transform of the function $P(\{\alpha_k\})$. On applying the Fourier inversion formula to Eq. (2.21), we obtain the following expression for $P(\{\alpha_k\})$:

$$P(\{\alpha_k\}) = \int \Phi(\{\zeta_k\}) \exp\left\{-i \sum_k (\alpha^*_k\zeta_k + \alpha_k\zeta^*_k)\right\} d^2\left\{\frac{\zeta_k}{2\pi}\right\}. \quad (2.23)$$

Equations (2.21) and (2.23) establish the above mentioned correspondence between the classical and quantum mechanical approaches to the coherence theory. The characteristic functional may be considered as a classical quantity which can be determined in classical terms, whereas the function $P(\{\alpha_k\})$ is inherently of quantum mechanical origin. In contradistinction to a functional formalism for turbulence, neither the configuration space probability density $w^{(N,M)}$ nor the probability functional W are necessarily nonnegative. Both of them inherit behavior of the function $P(\{\alpha_k\})$ from which they are derived and are subject to the same limitations that apply to $P(\{\alpha_k\})$. It is known that in some cases $P(\{\alpha_k\})$ can become quite singular—more singular, for example, than any tempered distribution[23]—and we can expect the same will hold for the probability functional.

[22] See Ref. 16, Eqs. (14.44) and (14.61).

[23] K. Cahill, Phys. Rev. **138**, B1566 (1965).

3. SOME SPECIFIC EXAMPLES

We illustrate the former results with the examples of characteristic functionals and probability functionals for certain types of fields of special interest.

A. Gaussian Fields

The density operator is specified by

$$P(\{\alpha_k\}) = \prod_k \frac{1}{\pi\langle n_k\rangle} e^{-|\alpha_k|^2/\langle n_k\rangle}, \tag{3.1}$$

where $\langle n_k\rangle$ is the mean number of photons in the kth mode. From (2.21) we find

$$\Phi = \prod_k \Phi_k, \tag{3.2}$$

where

$$\Phi_k = e^{-|\zeta_k|^2\langle n_k\rangle}. \tag{3.3}$$

Making use of (2.22) and the relation[24]

$$G_{\mu\nu}^{(1,1)}(x_1, x_2) = \sum_k e_\mu^*(x_1, k)\langle n_k\rangle e_\nu(x_2, k), \tag{3.4}$$

we obtain

$$\Phi[\zeta, \zeta^*] = \exp\{-\tilde{G}^{(1,1)}[\zeta, \zeta^*]\}. \tag{3.5}$$

With the aid of (2.7) it can be shown that

$$W[\mathcal{E}^*, \mathcal{E}] = \prod_k \frac{1}{2\pi\lambda_k} \exp\left\{-\int G_{\mu\nu}^{-1}(x_1, x_2)\mathcal{E}_\mu(x_1)\mathcal{E}_\nu^*(x_2)\, dx_1\, dx_2\right\}, \tag{3.6}$$

where λ_k are the eigenvalues of the kernel $G_{\mu\nu}(x_1, x_2)$, and $G_{\mu\nu}^{-1}(x_1, x_2)$ is the reciprocal kernel, i.e.,

$$\int G_{\mu\sigma}(x_1, x)G_{\sigma\nu}^{-1}(x_1, x_2)\, dx = \delta_{\mu\nu}^T(x_1 - x_2) \tag{3.7}$$

$\delta_{\mu\nu}^T$ being the transverse δ function.[25]

B. Fully Coherent Fields

In this case the density operator in the P representation is

$$P(\{\alpha_k\}) = \prod_k \delta(\alpha_k^* - \beta_k^*)\delta(\alpha_k - \beta_k), \tag{3.8}$$

which after substitution into (2.21) leads again to the factorized form (3.2) of Φ. Here we have

$$\Phi_k = \exp\left[i(\zeta_k\beta_k^* + \zeta_k^*\beta_k)\right] \tag{3.9}$$

and

$$\Phi = \exp\left[i\int \zeta_\mu(x)\mathcal{E}_\mu^*(x, \{\beta_k\})\, dx + i\int \zeta_\mu^*(x)\mathcal{E}_\mu(x, \{\beta_k\})\, dx\right]. \tag{3.10}$$

The probability functional, as also can be seen directly from (2.20), is of the form

$$W[\mathcal{E}^*, \mathcal{E}] = \delta[\mathcal{E}^*(x) - \mathcal{E}^*(x, \{\beta_k\})]\delta[\mathcal{E}(x) - \mathcal{E}(x, \{\beta_k\})]. \tag{3.11}$$

C. An Ideal Laser Field

In this case we have

$$P(\{\alpha_k\}) = (2\pi|\beta_l|)^{-1}\delta(|\alpha_l| - |\beta_l|)\prod_{k\neq l}(2\pi|\alpha_k|)^{-1}\delta(|\alpha_k|). \tag{3.12}$$

From (2.21), after a straightforward calculation, we obtain

$$\Phi = I_0\left\{2i\left[\int G_{\mu\nu}^{(\mathrm{I,L.})}(x_1, x_2)\zeta_\mu(x_1)\zeta_\nu^*(x_2)\, dx_1\, dx_2\right]^{\frac{1}{2}}\right\}, \tag{3.13}$$

where I_0 is the Bessel function of imaginary argument and zeroth order, and the correlation function of the (1, 1)th order for the ideal laser field is given by

$$G_{\mu\nu}^{(\mathrm{I,L.})}(x_1, x_2) = e_\mu^*(x_1, l)\,|\beta_l|^2\, e_\nu(x_2, l). \tag{3.14}$$

For the probability functional we obtain a rather complicated formula involving functional derivatives of the δ functionals,

$$W[\mathcal{E}^*, \mathcal{E}] = I_0\left\{2\left[\int G_{\mu\nu}^{(\mathrm{I,L.})}(x_1, x_2)\frac{\delta}{\delta\mathcal{E}_\mu^*(x_1)}\frac{\delta}{\delta\mathcal{E}_\nu(x_2)}\, dx_1\, dx_2\right]^{\frac{1}{2}}\right\} \times \delta[\mathcal{E}^*]\delta[\mathcal{E}], \tag{3.15}$$

which can also be written as

$$W[\mathcal{E}^*, \mathcal{E}] = \frac{1}{2\pi}\int_0^{2\pi} \delta[\mathcal{E}^*(x) - e^*(x, l)\,|\beta_l|\, e^{-i\theta_l}] \times \delta[\mathcal{E}(x) - e(x, l)\,|\beta_l|\, e^{i\theta_l}]\, d\theta_l, \tag{3.16}$$

if we use (3.12) and (2.20) rather than (2.7).

4. PROBABILITY FUNCTIONAL FOR COHERENT FIELDS

In Sec. 3 we were concerned with determining the probability functional on the basis of an explicit form of the P representation. Here we wish to describe the application of the functional formalism to the case where the only information about the field is the knowledge of the distribution moments. An example is provided by the (1, 1)th-order coherent fields, which have been studied extensively by Titulaer and Glauber.[26,27]

The (1, 1)th-order coherence condition

$$|G_{\mu\nu}^{(1,1)}(x_1, x_2)|^2 = G_{\mu\mu}^{(1,1)}(x_1, x_1)G_{\nu\nu}^{(1,1)}(x_2, x_2), \tag{4.1}$$

[24] See Ref. 16, p. 150, Eq. (14.32).
[25] See, for example, W. H. Louisell, *Radiation and Noise in Quantum Electronics* (McGraw-Hill Book Co., New York, 1964).
[26] U. M. Titulaer and R. J. Glauber, Phys. Rev. **140**, B676 (1965).
[27] U. M. Titulaer and R. J. Glauber, Phys. Rev. **145**, 1041 (1966).

provided it holds for all space–time points, implies the pair of identities obeyed by the density operator

$$E_{\mu}^{(+)}(x)\rho = \frac{G_{\mu_0\mu}^{(1,1)}(x_0, x)}{G_{\mu_0\mu_0}^{(1,1)}(x_0, x_0)} E_{\mu_0}^{(+)}(x_0)\rho, \tag{4.2}$$

$$\rho E^{(-)}(x) = \frac{G_{\mu\mu_0}^{(1,1)}(x, x_0)}{G_{\mu_0\mu_0}^{(1,1)}(x_0, x_0)} \rho E_{\mu_0}^{(-)}(x_0). \tag{4.3}$$

These identities may be used to shift the arguments of the (N, M)th-order correlation function to a common reference point x_0. By a direct extension of Titulaer and Glauber results, we obtain

$$G_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}}^{(N,M)}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}) = \frac{G_{\mu_0\cdots\mu_0}^{(N,M)}(x_0, \cdots, x_0)}{[G_{\mu_0\mu_0}^{(1,1)}(x_0, x_0)]^{\frac{1}{2}(N+M)}} \times \frac{\prod_{j=1}^{N} G_{\mu_j\mu_0}^{(1,1)}(x_j, x_0) \prod_{j=N+1}^{N+M} G_{\mu_0\mu_j}^{(1,1)}(x_0, x_j)}{[G_{\mu_0\mu_0}^{(1,1)}(x_0, x_0)]^{\frac{1}{2}(N+M)}}. \tag{4.4}$$

If we introduce the vector function

$$\mathcal{F}_{\mu}(x, \{\beta_k\}) = \frac{G_{\mu_0\mu}^{(1,1)}(x_0, x)}{[G_{\mu_0\mu_0}^{(1,1)}(x_0, x_0)]^{\frac{1}{2}}}, \tag{4.5}$$

where

$$\beta_k = \int \frac{u_{k\mu}^*(\mathbf{r})e^{i\omega_k t}}{i(\frac{1}{2}\hbar\omega_k)^{\frac{1}{2}}} \frac{G_{\mu_0\mu}^{(1,1)}(\mathbf{r}_0, t_0; \mathbf{r}, t)}{[G_{\mu_0\mu_0}^{(1,1)}(\mathbf{r}_0, t_0; \mathbf{r}_0, t_0)]^{\frac{1}{2}}} d\mathbf{r}, \tag{4.6}$$

we may write Eq. (4.4) in the form

$$G_{\mu_1\cdots\mu_N\mu_{N+1}\cdots\mu_{N+M}}^{(N,M)}(x_1, \cdots, x_N; x_{N+1}, \cdots, x_{N+M}) = g^{(N,M)} \prod_{j=1}^{N} \mathcal{F}_{\mu_j}^*(x_j, \{\beta_k\}) \prod_{j=N+1}^{N+M} \mathcal{F}_{\mu_j}(x_j, \{\beta_k\}). \tag{4.7}$$

It is worth noting that the constants

$$g^{(N,M)} = \frac{G_{\mu_0\cdots\mu_0}^{(N,M)}(x_0, \cdots, x_0)}{[G_{\mu_0\mu_0}^{(1,1)}(x_0, x_0)]^{\frac{1}{2}(N+M)}} \tag{4.8}$$

are not necessarily real numbers.

When the expression (4.7) is substituted in the Volterra expansion (2.5) of the characteristic functional, we find

$$\Phi[\zeta, \zeta^*] = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{i^{N+M} g^{(N,M)}}{N!\,M!} \times \prod_{j=1}^{N} \int \zeta_{\mu_j}(x_j)\mathcal{F}_{\mu_j}^*(x_j, \{\beta_k\})\, dx_j \times \prod_{j=N+1}^{N+M} \int \zeta_{\mu_j}^*(x_j)\mathcal{F}_{\mu_j}(x_j, \{\beta_k\})\, dx_j. \tag{4.9}$$

According to (2.1), this may be written as

$$\Phi[\zeta, \zeta^*] = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{i^{N+M} g^{(N,M)}}{N!\,M!} (\zeta, \mathcal{F}^*)^N (\zeta^*, \mathcal{F})^M, \tag{4.10}$$

and the Fourier inversion formula (2.7) yields

$$W[\mathcal{E}^*, \mathcal{E}] = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{i^{N+M} g^{(N,M)}}{N!\,M!} \times \int (\zeta, \mathcal{F}^*)^N e^{-i(\zeta, \mathcal{E}^*)}\, d\left(\frac{\zeta}{2\pi}\right) \times \int (\zeta^*, \mathcal{F})^M e^{-i(\zeta^*, \mathcal{E})}\, d\left(\frac{\zeta^*}{2\pi}\right). \tag{4.11}$$

This is the desired general expression for the probability functional representing the (1, 1)th-order coherent fields. The functional integration in (4.11) can be performed to give

$$W[\mathcal{E}^*, \mathcal{E}] = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{g^{(N,M)}}{N!\,M!} \times \left[-\int \mathcal{F}_{\mu}^*(x) \frac{\delta}{\delta\mathcal{E}_{\mu}^*(x)}\, dx\right]^N \delta[\mathcal{E}^*] \times \left[-\int \mathcal{F}_{\mu}(x) \frac{\delta}{\delta\mathcal{E}_{\mu}^*(x)}\right]^M \delta[\mathcal{E}], \tag{4.12}$$

or in a more extended form,

$$W[\mathcal{E}^*, \mathcal{E}] = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{g^{(N,M)}}{N!\,M!} \times \int (-1)^N \mathcal{F}_{\mu_1}^*(x_1) \cdots \mathcal{F}_{\mu_N}^*(x_N) \times \frac{\delta^N \delta[\mathcal{E}^*]}{\delta\mathcal{E}_{\mu_1}^*(x_1) \cdots \delta\mathcal{E}_{\mu_N}^*(x_N)}\, dx_1 \cdots dx_N \times \int (-1)^M \mathcal{F}_{\mu_1}(x_1) \cdots \mathcal{F}_{\mu_M}(x_M) \times \frac{\delta^M \delta[\mathcal{E}]}{\delta\mathcal{E}_{\mu_1}(x_1) \cdots \delta\mathcal{E}_{\mu_M}(x_M)}\, dx_1 \cdots dx_M. \tag{4.13}$$

With the aid of (2.19), it may be readily verified that the expression (4.13) leads to the correlation functions of the form (4.7), which is characteristic for the (1, 1)th-order coherent fields. In order to demonstrate this, use must be made of the relations

$$\int dx_1' \cdots dx_N' \mathcal{F}_{\nu_1}^*(x_1') \cdots \mathcal{F}_{\nu_N}^*(x_N') \times \int \frac{(-1)^N \delta^N \delta[\mathcal{E}^*]}{\delta\mathcal{E}_{\nu_1}^*(x_1') \cdots \delta\mathcal{E}_{\nu_N}^*(x_N')} \mathcal{E}_{\mu_1}^*(x_1) \cdots \mathcal{E}_{\mu_N}^*(x_N)\, d\mathcal{E}^* = N! \int dx_1' \cdots dx_N' \mathcal{F}_{\nu_1}^*(x_1') \cdots \mathcal{F}_{\nu_N}^*(x_N') \times \delta_{\mu_1\nu_1}\delta(x_1 - x_1') \cdots \delta_{\mu_N\nu_N}\delta(x_N - x_N') = N!\, \mathcal{F}_{\mu_1}^*(x_1) \cdots \mathcal{F}_{\mu_N}^*(x_N), \tag{4.14}$$

and the similar ones, obtained by replacing $\mathcal{F}^*$, $\mathcal{E}^*$, and N by $\mathcal{F}$, $\mathcal{E}$, and M, respectively.

The case of full coherence corresponds to $g^{(N,M)} = 1$, for all N and M. It follows then from (4.12) that

$$W[\mathcal{E}^*, \mathcal{E}] = \delta[\mathcal{E}^*(x) - \mathcal{F}^*(x, \{\beta_k\})] \times \delta[\mathcal{E}(x) - \mathcal{F}(x, \{\beta_k\})], \quad (4.15)$$

which agrees with (3.11).

Finally, if we write the characteristic functional in the form

$$\Phi(\{\zeta_k\}) = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{i^{N+M} g^{(N,M)}}{N!\,M!} \left(\sum_k \zeta_k \beta_k^*\right)^N \left(\sum_k \zeta_k^* \beta_k\right)^M, \quad (4.16)$$

then, by virtue of (2.23), we find

$$P(\{\alpha_k\}) = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{g^{(N,M)}}{N!\,M!} \left(-\sum_k \beta_k^* \frac{\partial}{\partial \alpha_k^*}\right)^N \delta(\{\alpha_k^*\}) \times \left(-\sum_k \beta_k \frac{\partial}{\partial \alpha_k}\right)^M \delta(\{\alpha_k\}). \quad (4.17)$$

This formula may be regarded as a counterpart of the expression for the density operator representing the (1, 1)th-order coherent fields in terms of the n-photon states.[28]

ACKNOWLEDGMENT

The author wishes to thank Dr. B. Karczewski for his interest and advice.

[28] See Ref. 27, Eq. (2.15).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Relativistic Analogs of Scalar-Tensor Cosmologies

C. B. G. McIntosh
Department of Mathematics, University of New England, Armidale, N.S.W., Australia

(Received 23 December 1968)

Relativistic analogs of Brans–Dicke and Hoyle–Narlikar scalar–tensor cosmologies are given in terms of two- or three-fluid models. Each fluid has an equation of state of the form pressure = const × density. The models discussed have the matter and scalar fluids interacting and are expressed by different forms of E_i, where this is the total new rate of transfer of energy per unit volume to the ith fluid from all other fluids.

Vajk[1] has commented recently on relativistic cosmological models with one or more relativistic fluids, each with an equation of state of the form

$$P = (\nu - 1)\rho c^2, \quad 1 \leq \nu \leq 2, \quad (1)$$

where ν is a constant. The models he considers are homogeneous and isotropic with conservation equation

$$\frac{d}{dt}(\rho c^2 R^3) + P \frac{d}{dt}(R^3) = 0, \quad (2)$$

where R is the scale factor of the Robertson–Walker line element

$$ds^2 = c^2\,dt^2 - R^2(t)\left\{\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta\,d\phi^2)\right\}. \quad (3)$$

Here r, θ, and ϕ are dimensionless, comoving coordinates and k is 0, ± 1. In a one-fluid model, Eqs. (1) and (2) give

$$\rho = BR^{-3\nu}, \quad (4)$$

where B is a constant. When this is inserted in the Friedmann differential equation with zero cosmological constant

$$\kappa\rho c^4 = 3(\dot{R}^2 + kc^2)R^{-2}, \quad (5)$$

it can be seen that, for $k \neq 0$, the metric can be expressed in terms of elementary functions and elliptic functions only if

$$\nu = \tfrac{1}{3}n, \quad n = 3, 4, 5, 6. \quad (6)$$

If a model contains n noninteracting fluids, then each one satisfies Eq. (4) and the total density and pressure are

$$\rho = \sum_{i=1}^{n} B_i R^{-3\nu_i}, \quad P = \sum_{i=1}^{n} (\nu_i - 1)c^2 B_i R^{-3\nu_i}. \quad (7)$$

Numerous papers have been published recently on models containing the two noninteracting fluids of radiation ($\nu_1 = \frac{4}{3}$, $P = \frac{1}{3}\rho c^2$) and dust ($\nu_2 = 1$, $P = 0$) in order to describe a universe which was dominated by radiation at early epochs but is now dominated by

[1] J. P. Vajk, J. Math. Phys. **10**, 1145 (1969).

The case of full coherence corresponds to $g^{(N,M)} = 1$, for all N and M. It follows then from (4.12) that

$$W[\mathcal{E}^*, \mathcal{E}] = \delta[\mathcal{E}^*(x) - \mathcal{F}^*(x, \{\beta_k\})] \times \delta[\mathcal{E}(x) - \mathcal{F}(x, \{\beta_k\})], \quad (4.15)$$

which agrees with (3.11).

Finally, if we write the characteristic functional in the form

$$\Phi(\{\zeta_k\}) = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{i^{N+M} g^{(N,M)}}{N!\,M!} \left(\sum_k \zeta_k \beta_k^*\right)^N \left(\sum_k \zeta_k^* \beta_k\right)^M, \quad (4.16)$$

then, by virtue of (2.23), we find

$$P(\{\alpha_k\}) = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} \frac{g^{(N,M)}}{N!\,M!} \left(-\sum_k \beta_k^* \frac{\partial}{\partial \alpha_k^*}\right)^N \delta(\{\alpha_k^*\}) \times \left(-\sum_k \beta_k \frac{\partial}{\partial \alpha_k}\right)^M \delta(\{\alpha_k\}). \quad (4.17)$$

This formula may be regarded as a counterpart of the expression for the density operator representing the (1, 1)th-order coherent fields in terms of the n-photon states.[28]

ACKNOWLEDGMENT

The author wishes to thank Dr. B. Karczewski for his interest and advice.

[28] See Ref. 27, Eq. (2.15).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Relativistic Analogs of Scalar-Tensor Cosmologies

C. B. G. McIntosh
Department of Mathematics, University of New England, Armidale, N.S.W., Australia

(Received 23 December 1968)

Relativistic analogs of Brans–Dicke and Hoyle–Narlikar scalar–tensor cosmologies are given in terms of two- or three-fluid models. Each fluid has an equation of state of the form pressure = const × density. The models discussed have the matter and scalar fluids interacting and are expressed by different forms of E_i, where this is the total new rate of transfer of energy per unit volume to the ith fluid from all other fluids.

Vajk[1] has commented recently on relativistic cosmological models with one or more relativistic fluids, each with an equation of state of the form

$$P = (\nu - 1)\rho c^2, \quad 1 \leq \nu \leq 2, \quad (1)$$

where ν is a constant. The models he considers are homogeneous and isotropic with conservation equation

$$\frac{d}{dt}(\rho c^2 R^3) + P \frac{d}{dt}(R^3) = 0, \quad (2)$$

where R is the scale factor of the Robertson–Walker line element

$$ds^2 = c^2\,dt^2 - R^2(t)\left\{\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta\,d\phi^2)\right\}. \quad (3)$$

Here r, θ, and ϕ are dimensionless, comoving coordinates and k is 0, ± 1. In a one-fluid model, Eqs. (1) and (2) give

$$\rho = BR^{-3\nu}, \quad (4)$$

where B is a constant. When this is inserted in the Friedmann differential equation with zero cosmological constant

$$\kappa\rho c^4 = 3(\dot{R}^2 + kc^2)R^{-2}, \quad (5)$$

it can be seen that, for $k \neq 0$, the metric can be expressed in terms of elementary functions and elliptic functions only if

$$\nu = \tfrac{1}{3}n, \quad n = 3, 4, 5, 6. \quad (6)$$

If a model contains n noninteracting fluids, then each one satisfies Eq. (4) and the total density and pressure are

$$\rho = \sum_{i=1}^{n} B_i R^{-3\nu_i}, \quad P = \sum_{i=1}^{n} (\nu_i - 1)c^2 B_i R^{-3\nu_i}. \quad (7)$$

Numerous papers have been published recently on models containing the two noninteracting fluids of radiation ($\nu_1 = \frac{4}{3}$, $P = \frac{1}{3}\rho c^2$) and dust ($\nu_2 = 1$, $P = 0$) in order to describe a universe which was dominated by radiation at early epochs but is now dominated by

[1] J. P. Vajk, J. Math. Phys. **10**, 1145 (1969).

matter (see, for example, Refs. 2–7). Other papers[8–10] have been published on models containing these two fluids but in the case where there is conversion from one to the other. Where there are n interacting fluids, Eq. (2) can be written as

$$\sum_{i=1}^{n} E_i = 0, \tag{8}$$

where

$$E_i = R^{-3}\left[\frac{d}{dt}(\rho_i c^2 R^3) + P_i \frac{d}{dt}(R^3)\right] = (\dot{\rho}_i + 3\nu_i H \rho_i)c^2. \tag{9}$$

$H = \dot{R}/R$ is the total net rate of transfer of energy per unit volume from the $n-1$ fluids $(1, 2, \cdots, i-1, i+1, \cdots, n)$ to the ith fluid. Equation (7) thus holds when all the E_i are zero.

It is shown in this paper how two- or three-fluid relativistic cosmological models can be used to describe both Dicke's (D)[11,12] and Hoyle and Narlikar's (HN)[13] scalar–tensor cosmologies. The relativistic models used are ones which are solutions of Einstein's field equations

$$R_{ij} - \tfrac{1}{2}g_{ij}R = -\kappa T_{ij}, \quad \kappa = 8\pi G/c^4, \tag{10}$$

which are to be understood in the normal way. This is done by writing the density as

$$\rho = \rho_m + \rho_r + \rho_c, \tag{11}$$

where the subscript m denotes matter (or dust), r denotes radiation, and c denotes a third fluid with $\nu = 2$, $P_c = \rho_c c^2$. The matter–energy tensor is now

$$\begin{aligned} T_{ij} &= -Pg_{ij} + (P + \rho c^2)u_i u_j \\ &= -(\rho_c + \tfrac{1}{3}\rho_r)c^2 g_{ij} \\ &\quad + (\rho_m + \tfrac{4}{3}\rho_r + 2\rho_c)c^2 u_i u_j, \end{aligned} \tag{12}$$

$u^i = dx^i/ds$. The field equations and the metric (3) now give

$$3(\dot{R}^2 + kc^2)/R^2 = \kappa c^4(\rho_m + \rho_r + \rho_c) \tag{13}$$

and

$$2\ddot{R}/R + (\dot{R}^2 + kc^2)/R^2 = -\kappa c^4(\rho_c + \tfrac{1}{3}\rho_r). \tag{14}$$

In the following, k will be taken as zero.

The two scalar–tensor cosmologies satisfy Eqs. (13) and (14) together with the conditions

$$\text{HN: } E_r = 0 \text{ and } \begin{cases} E_m = E_c = 0, & \text{(15a)} \\ E_m = -E_c = 3c^{-2}fH, & \text{(15b)} \end{cases}$$

$$\text{D: } E_r = 0, \quad E_m = -E_c = \pm\tfrac{1}{2}\rho_m \rho_c^{\frac{1}{2}} Mc^4(-2/f)^{\frac{1}{2}}, \tag{16}$$

where M and f are constants described below. In the original versions of both theories, P_c is expressed in terms of a scalar field C by splitting the matter–energy tensor into a dust and radiation part $T_{m\,ij}$ and a scalar part $T_{c\,ij}$ such that

$$T_{c\,ij} = -f(C_i C_j - \tfrac{1}{2}g_{ij}C_k C^k), \quad C_k = C_{,k}. \tag{17}$$

This leads to

$$\rho_c c^2 = -\tfrac{1}{2}fC_k C^k \tag{18}$$

so that, when C is a function only of t,

$$\rho_c c^4 = -\tfrac{1}{2}f\dot{C}^2. \tag{19}$$

Here, f and C are the same here as in the HN theory, though Hoyle and Narlikar deal with the case $\rho_r = 0$. In their model, the conservation equation $T^{ij}{}_{;j} = 0$ means that

$$(\dot{\rho}_m + 3H\rho_m)c^4 = f\dot{C}(\ddot{C} + 3H\dot{C}) \tag{20}$$

and the extra condition required by their theory is

$$f \,\square\, C = c^2 j^i{}_{;i}, \quad j^i = \rho c u^i, \tag{21}$$

giving

$$(\dot{\rho}_m + 3H\rho_m)c^4 = f(\ddot{C} + 3H\dot{C}). \tag{22}$$

Equations (20) and (22) agree only when $E_m = E_c = 0$ or $\dot{C} = 1$. The first case is Eq. (15a). When $\rho_r = 0$, this is Nariai's (B) type solution and the metric has been written out by various authors including Vajk.[1] The second case leads to

$$\rho_c c^4 = -\tfrac{1}{2}f, \quad \rho_m c^4 = f + \text{const}/R^3, \tag{23}$$

and is the case (15b). When $\rho_r = 0$, this is Nariai's (A_1) type solution[14] when the constant in Eq. (23) is zero and his (A_2) type solution otherwise. The (A_1) model is Hoyle and Narlikar's steady-state model. In both these (A) models, there is continual creation of matter but only because energy from $\rho_c c^2$ is converted to energy in $\rho_m c^2$.

In the D theory,

$$MC = \ln \lambda, \quad f = -(2\omega + 3)M^2 c^4/16\pi G, \tag{24}$$

[2] P. J. E. Peebles, Astrophys. J. **142**, 1317 (1965).
[3] R. A. Alpher, G. Gamow, and R. Herman, Proc. Natl. Acad. Sci. (U.S.) **58**, 2179 (1967).
[4] A. D. Chernin, Astron. Zh. **42**, 1124 (1965) [Sov. Astron.—A.J. **9**, 871 (1966)].
[5] K. C. Jacobs, Nature **215**, 1156 (1967).
[6] C. B. G. McIntosh, Monthly Notices Roy. Astron. Soc. **138**, 423 (1968).
[7] Ya. B. Zel'dovich, Usp. Fiz. Nauk **89**, 647 (1966) [Sov. Phys.—Usp. **9**, 602 (1967)].
[8] W. Davidson and J. V. Narlikar, Rept. Progr. Phys. **29**, 539 (1966).
[9] C. B. G. McIntosh, Nature **215**, 36 (1967).
[10] C. B. G. McIntosh, Monthly Notices Roy. Astron. Soc. **140**, 461 (1968).
[11] R. H. Dicke, Phys. Rev. **125**, 2163 (1962).
[12] R. H. Dicke, Astrophys. J. **152**, 1 (1968).
[13] F. Hoyle and J. V. Narlikar, Proc. Roy. Soc. (London) **A273**, 1 (1963).
[14] H. Nariai, Progr. Theoret. Phys. (Kyoto) **32**, 450 (1964).

where M is a constant with the dimension of (time)$^{-1}$. Equation (16) can be written as

$$E_m = -\tfrac{1}{2}\rho_m M c^2 \dot{C} = -\tfrac{1}{2}\rho_m c^2 \dot{\lambda}/\lambda \qquad (25)$$

and leads to

$$\lambda^{\frac{1}{2}}\rho_m R^3 = \text{const} = (\lambda^2\rho_m)(\lambda^{-\frac{1}{2}}R)^3. \qquad (26)$$

Dicke[12] interprets this as $\bar{\rho}_m \bar{R}^3 = \text{const}$, where $\bar{\rho}_m$ and $\bar{R}$ are measured in ordinary (cgs) units and the ρ_m and R in gravitational units. The general form of the condition (16) is

$$\Box\, C = -\rho_m c^2 M/(2f). \qquad (27)$$

In the D theory, ρ_c is positive for $\omega > -\frac{3}{2}$, but in the HN theory, ρ_c is negative for f positive. It is interesting to note here that Noerdlinger[15] has stated that there is a "theoretical necessity of a positive value" for ω. This point of view would eliminate the HN model.

Dicke[12] has shown that, in his model, the ratio of helium to hydrogen produced by the cosmic fireball can be reduced to almost zero. It would be interesting to study the effects, in general, on the production of helium of the inclusion of a fluid with $v = 2$ into cosmological models containing matter and radiation. The author does not know of any work where HN models containing radiation have been studied.

In neither theory is it necessary to introduce the scalar field C as in Eq. (17). Hoyle and Narlikar[16] have noted, "It is also of interest that if creation of matter is postulated it is possible to obtain the steady-state cosmology *without* introducing the C field used in previous papers. It turns out that the mass fields can play an almost identical role to the C field."

When looked at in this way, the two theories discussed could be accused of having ad hoc choices for the conditions (15) and (16). It must be admitted, however, that the conditions (21) and (27) can be derived from variational principles. The two theories do appear far more similar than would generally be believed and it would be worthwhile to compare them to see how results applying to one apply to the other and also to general 3-fluid models. Both Brans and Dicke[17] (in the original D model) and Hoyle and Narlikar developed their theories partly to understand the relationship between Mach's principle and general relativity. The question now arises as to how a general three-fluid model helps us study such a relationship.

Also, it must be noted that all authors saw their theories as satisfying Einstein's equations but only when the original Lagrangian from which the equations were derived was altered. What is clear is that there is no need to change this Lagrangian in order to obtain the two different kinds of cosmologies. It could be worthwhile to examine various other ad hoc conditions such as (15) and (16) to see what types of models can be formed and whether or not corresponding Lagrangians can be found.

The Dicke theory was formed from the Brans–Dicke one by means of a conformal mapping in the metric tensor. A similar mapping was made by Hoyle and Narlikar[18] which yields field equations similar to the Brans–Dicke ones, though these field equations are smooth-fluid approximations of their particle equations and are not valid in the neighborhood of a particle.

ACKNOWLEDGMENT

I thank Dr. N. W. Taylor for useful discussions on this work.

[15] P. D. Noerdlinger, Phys. Rev. **170**, 1175 (1968).
[16] F. Hoyle and J. V. Narlikar, Proc. Roy. Soc. (London) **A299**, 188 (1967).
[17] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[18] F. Hoyle and J. V. Narlikar, Proc. Roy. Soc. (London) **A294**, 138 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

On Crossing Relations for Helicity Amplitudes

YASUO HARA
Department of Physics, Tokyo University of Education, Tokyo, Japan

(Received 30 April 1968)

Crossing relations between s-channel helicity amplitudes and u-channel helicity amplitudes are derived, following the derivation of the crossing relations between s-channel helicity amplitudes and t-channel helicity amplitudes by Trueman and Wick [T. L. Trueman and G.-C. Wick, Ann. Phys. (N.Y.) **26**, 322 (1964)]. The over-all phase factor of crossing relations, which was not given by Trueman and Wick, is also derived by specifying the path of analytic continuation and by applying these relations in the static approximation.

1. INTRODUCTION

In 1963, crossing relations for helicity amplitudes (crossing relations between s-channel helicity amplitudes and t-channel helicity amplitudes) were derived by Trueman and Wick.[1] These relations have been found to be powerful tools in particle physics. For example, they have been used to study the analytic properties of helicity amplitudes[2,3] and to derive the so-called conspiracy relations.[4]

According to Trueman and Wick, helicity amplitudes $G_{\delta\beta,\gamma\alpha}$ of the s-channel reaction $A + C \to B + D$ are expressed as linear combinations of helicity amplitudes $F_{\delta\gamma,\beta\alpha}$ of the t-channel reaction $A + \bar{B} \to \bar{C} + D$ (see Sec. 2).

One of the purposes of this paper is to derive crossing relations between s-channel helicity amplitudes and u-channel helicity amplitudes, i.e., to express helicity amplitudes $G_{\delta\beta,\gamma\alpha}$ of the s-channel reaction $A + C \to B + D$ as linear combinations of helicity amplitudes $H_{\gamma\beta,\delta\alpha}$ of the u-channel reaction $A + \bar{D} \to B + \bar{C}$ (see Sec. 3). These crossing relations between G and H are easily obtained if we follow the derivation of the Trueman–Wick crossing relations between G and F. Since the derivation of the Trueman–Wick crossing relations is complicated, we believe that it is worth deriving the crossing relations between G and H.

Another purpose of this paper is to find the over-all phase factor of the relations. The factor is dependent on the path of the analytic continuation. It is dependent on the position of the real point of the trajectory in the s–t plane. In order to avoid the ambiguity, we assume that the real point lies in the domain of the s–t plane which reduces to the triangle defined by the inequalities

$$4m^2 > s, t \quad \text{and} \quad u > 0 \tag{1}$$

when all external and internal particles have a common mass m. We determine the over-all phase factor by applying our crossing relations to cases in which the crossing relations are explicitly known, i.e., to the static model (Sec. 4) and to elastic $p\bar{p}$ scattering (one pion exchange process) (Sec. 5). Discussions are given in Sec. 6.

2. CROSSING RELATIONS BETWEEN s-CHANNEL HELICITY AMPLITUDES AND t-CHANNEL HELICITY AMPLITUDES

In this section we write the crossing relations between helicity amplitudes $G_{\delta\beta,\gamma\alpha}$ of the s-channel reaction $A + C \to B + D$ (Fig. 1) and helicity amplitudes $F_{\delta\gamma,\beta\alpha}$ of the t-channel reaction $A + \bar{B} \to \bar{C} + D$ (Fig. 2)[1] for convenience as

$$G_{\delta\beta,\gamma\alpha}(s = (p_a + p_c)^2, t = (p_a - p_b)^2) = \epsilon(-1)^{\eta} \sum_{\alpha'\beta'\gamma'\delta'} (-1)^{\eta'} \times d^D_{\delta'\delta}(\psi_d) d^C_{\gamma'\gamma}(\psi_c) d^B_{\beta'\beta}(\psi_b) d^A_{\alpha'\alpha}(\psi_a) F_{\delta'\gamma',\beta'\alpha'}(t = (p_a + P_b)^2, s = (p_a - P_c)^2), \tag{2}$$

where

$$\begin{aligned}
\cos\psi_a &= \frac{(s + m_a^2 - m_c^2)(t + m_a^2 - m_b^2) - 2m_a^2(m_a^2 - m_b^2 - m_c^2 + m_d^2)}{4p_s p_t (st)^{\frac{1}{2}}}, \\
\cos\psi_b &= \frac{-(s + m_b^2 - m_d^2)(t + m_b^2 - m_a^2) - 2m_b^2(m_a^2 - m_b^2 - m_c^2 + m_d^2)}{4p'_s p_t (st)^{\frac{1}{2}}}, \\
\cos\psi_c &= \frac{-(s + m_c^2 - m_a^2)(t + m_c^2 - m_d^2) - 2m_c^2(m_a^2 - m_b^2 - m_c^2 + m_d^2)}{4p_s p'_t (st)^{\frac{1}{2}}}, \\
\cos\psi_d &= \frac{(s + m_d^2 - m_b^2)(t + m_d^2 - m_c^2) - 2m_d^2(m_a^2 - m_b^2 - m_c^2 + m_d^2)}{4p'_s p'_t (st)^{\frac{1}{2}}},
\end{aligned} \tag{3}$$

[1] T. L. Trueman and G.-C. Wick, Ann. Phys. (N.Y.) **26**, 322 (1964). [2] Y. Hara, Phys. Rev. **136B**, 507 (1964).
[3] L. L. Wang, Phys. Rev. **142**, 1187 (1966). [4] See, for example, Y. Hara, Progr. Theoret. Phys. (Kyoto) **37**, 941 (1967).

$$\sin\psi_a = m_a p_t \sin\theta_t / p_s(s)^{\frac{1}{2}},$$
$$\sin\psi_b = m_b p_t \sin\theta_t / p_s'(s)^{\frac{1}{2}},$$
$$\sin\psi_c = m_c p_t' \sin\theta_t / p_s(s)^{\frac{1}{2}}, \quad (4)$$
$$\sin\psi_d = m_d p_t' \sin\theta_t / p_s'(s)^{\frac{1}{2}},$$

$$p_s = [s-(m_a+m_c)^2]^{\frac{1}{2}}[s-(m_a-m_c)^2]^{\frac{1}{2}}/2(s)^{\frac{1}{2}},$$
$$p_s' = [s-(m_b+m_d)^2]^{\frac{1}{2}}[s-(m_b-m_d)^2]^{\frac{1}{2}}/2(s)^{\frac{1}{2}},$$
$$p_t = [t-(m_a+m_b)^2]^{\frac{1}{2}}[t-(m_a-m_b)^2]^{\frac{1}{2}}/2(t)^{\frac{1}{2}}, \quad (5)$$
$$p_t' = [t-(m_c+m_d)^2]^{\frac{1}{2}}[t-(m_c-m_d)^2]^{\frac{1}{2}}/2(t)^{\frac{1}{2}},$$

and

$$\cos\theta_t = \left[2st + t^2 - t\sum_i m_i^2 + (m_a^2-m_b^2)(m_c^2-m_d^2)\right] \times (4tp_tp_t')^{-1}. \quad (6)$$

In Eq. (2), p_i is the momentum of the particle i in the s channel, and p_a, P_b $(=-p_b)$, P_c $(=-p_c)$, and p_d are momenta of particles in the t channel. The α, β, γ, and $\delta(\alpha', \beta', \gamma',$ and $\delta')$ are helicities of particles a, b, c, and d in the s (t) channel. The superscript A of the function $d^A_{\alpha'\alpha}$ in Eq. (2) stands for the spin of the particle A. The sign of $\sin\theta_t$ is determined by the relation

$$s^{\frac{1}{2}}p_sp_s'\sin\theta_s = t^{\frac{1}{2}}p_tp_t'\sin\theta_t. \quad (7)$$

The left-hand side of Eq. (7) must be continued analytically from the physical region of the s-channel reaction to the physical region of the t-channel reaction on the condition that $\sin\theta_s > 0$.

The quantities η and η' depend on which particles are taken to be "particle 2" in G and F. The values of

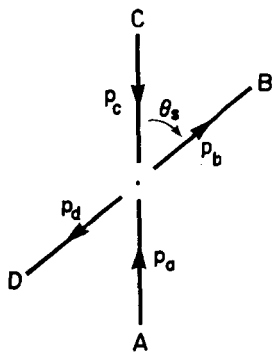


FIG. 1. The s-channel reaction: $A + C \to B + D$. p_a:$(\omega_s; 0, 0, p_s)$, p_b:$(\omega_s'; p_s'\sin\theta_s, 0, p_s'\cos\theta_s)$, p_c:$(E_s, 0, 0, -p_s)$, p_d:$(E_s'; -p_s'\sin\theta_s, 0, -p_s'\cos\theta_s)$.

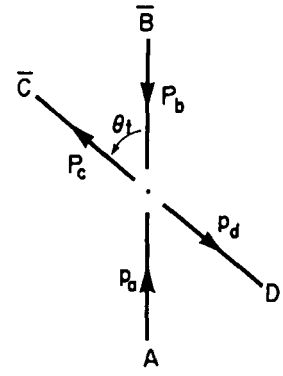


FIG. 2. The t-channel reaction: $A + \bar{B} \to \bar{C} + D$. p_a:$(\omega_t; 0, 0, p_t)$, P_b:$(\omega_t'; 0, 0, -p_t)$, P_c:$(E_t; -p'\sin\theta_t, 0, p_t'\cos\theta_t)$, p_d:$(E_t'; p_t'\sin\theta_t, 0, -p_t'\cos\theta_t)$.

η and η' for the various possible choices in the definition of G and F are tabulated below.

"particle 2" in G	η
A, B	0
A, D	$-\beta+\delta$
C, B	$-\alpha+\gamma$
C, D	$-\alpha-\beta+\gamma+\delta$

"particle 2" in F	η'
$A, \bar{C}$	0
A, D	$\gamma'-\delta'$
$\bar{B}, \bar{C}$	$\alpha'-\beta'$
$\bar{B}, D$	$\alpha'-\beta'+\gamma'-\delta'$

The quantity $\epsilon = (-1)^\Lambda$ is the over-all factor and is expressed as

$$\Lambda = aA + bB + cC + dD. \quad (8)$$

The factor ϵ can be determined only when the s-channel reaction $A + C \to B + D$ is identical to the t-channel reaction $A + \bar{B} \to \bar{C} + D$, i.e., $B = \bar{C}$. In Secs. 4 and 5 it will be found that

$$\Lambda = 2A = 2D \quad (9)$$

when particles B and C ($\bar{B}$ and $\bar{C}$) are taken to be "particle 2" in both G and F.

3. CROSSING RELATIONS BETWEEN s-CHANNEL HELICITY AMPLITUDES AND u-CHANNEL HELICITY AMPLITUDES

In this section we derive crossing relations between helicity amplitudes $G_{\delta\beta,\gamma\alpha}$ of the s-channel reaction $A + C \to B + D$ (Fig. 1) and helicity amplitudes

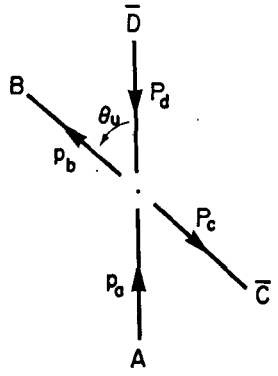


Fig. 3. The u-channel reaction: $A + \bar{D} \to B + \bar{C}$. p_a:$(\omega_u; 0, 0, p_u)$, p_b:$(\omega'_u; -p'_u \sin\theta_u, 0, p'_u \cos\theta_u)$, P_c:$(E'_u; p'_u \sin\theta_u, 0, -p'_u \cos\theta_u)$, P_d:$(E_u; 0, 0, -p_u)$.

$H_{\gamma\beta,\delta\alpha}$ of the u-channel reaction $A + \bar{D} \to B + \bar{C}$ (Fig. 3). If we compare Figs. 1 and 2 with Figs. 1 and 3, we find that the crossing relations between G and H are obtained from the crossing relations between G and F of the previous section through the following substitutions:

$$\begin{aligned}
&\theta_s \to \theta_s + \pi,\\
&\theta_t \to \theta_u + \pi,\\
&s \to s,\\
&t \to u,\\
&A \to A, \quad \alpha \to \alpha, \quad \alpha' \to \alpha',\\
&B \to D, \quad \beta \to \delta, \quad \beta' \to \delta',\\
&C \to C, \quad \gamma \to \gamma, \quad \gamma' \to \gamma',\\
&D \to B, \quad \delta \to \beta, \quad \delta' \to \beta'.
\end{aligned} \tag{10}$$

Thus we obtain the following crossing relations[5]:

$$\begin{aligned}
&G_{\delta\beta,\gamma\alpha}(s = (p_a + p_c)^2, u = (p_a - p_d)^2)\\
&= \epsilon'(-1)^{\xi} \sum_{\alpha'\beta'\gamma'\delta'} (-1)^{\xi'}\\
&\times d^D_{\delta'\delta}(\chi_d) d^C_{\gamma'\gamma}(\chi_c) d^B_{\beta'\beta}(\chi_b) d^A_{\alpha'\alpha}(\chi_a)\\
&\times H_{\gamma'\beta',\delta'\alpha'}(u = (p_a + P_d)^2, s = (p_a - P_c)^2),
\end{aligned} \tag{11}$$

where

$$\begin{aligned}
\cos\chi_a &= \frac{(s + m_a^2 - m_c^2)(u + m_a^2 - m_d^2) - 2m_a^2(m_a^2 + m_b^2 - m_c^2 - m_d^2)}{4p_s p_u (su)^{\frac{1}{2}}},\\
\cos\chi_b &= \frac{(s + m_b^2 - m_d^2)(u + m_b^2 - m_c^2) - 2m_b^2(m_a^2 + m_b^2 - m_c^2 - m_d^2)}{4p'_s p'_u (su)^{\frac{1}{2}}},\\
\cos\chi_c &= \frac{-(s + m_c^2 - m_a^2)(u + m_c^2 - m_b^2) - 2m_c^2(m_a^2 + m_b^2 - m_c^2 - m_d^2)}{4p_s p'_u (su)^{\frac{1}{2}}},\\
\cos\chi_d &= \frac{-(s + m_d^2 - m_b^2)(u + m_d^2 - m_a^2) - 2m_d^2(m_a^2 + m_b^2 - m_c^2 - m_d^2)}{4p'_s p_u (su)^{\frac{1}{2}}},
\end{aligned} \tag{12}$$

$$\begin{aligned}
\sin\chi_a &= m_a p_u \sin\theta_u / p_s (s)^{\frac{1}{2}},\\
\sin\chi_b &= m_b p'_u \sin\theta_u / p'_s (s)^{\frac{1}{2}},\\
\sin\chi_c &= m_c p'_u \sin\theta_u / p_s (s)^{\frac{1}{2}},\\
\sin\chi_d &= m_d p_u \sin\theta_u / p'_s (s)^{\frac{1}{2}},
\end{aligned} \tag{13}$$

$$\begin{aligned}
p_u &= [u - (m_a + m_d)^2]^{\frac{1}{2}}[u - (m_a - m_d)^2]^{\frac{1}{2}}/2(u)^{\frac{1}{2}},\\
p'_u &= [u - (m_b + m_c)^2]^{\frac{1}{2}}[u - (m_b - m_c)^2]^{\frac{1}{2}}/2(u)^{\frac{1}{2}},
\end{aligned} \tag{14}$$

$$p_c = -P_c, \quad \text{and} \quad p_d = -P_d. \tag{15}$$

The sign of $\sin\theta_u$ is determined by the relation

$$s^{\frac{1}{2}} p_s p'_s \sin\theta_s = u^{\frac{1}{2}} p_u p'_u \sin\theta_u. \tag{16}$$

The left-hand side of the relation must be continued analytically from the physical region of the s-channel reaction to the physical region of the u-channel reaction on the condition that $\sin\theta_s > 0$. The values of ξ and ξ' are tabulated below.[5]

"particle 2" in G	ξ
A, D	$\alpha - \beta + \gamma - \delta$
A, B	$\alpha + \gamma - 2\delta$
C, D	$-\beta + 2\gamma - \delta$
C, B	$2\gamma - 2\delta$

"particle 2" in H	ξ'
$A, \bar{C}$	$-\alpha' + \beta' - \gamma' + \delta'$
A, B	$-\alpha' + \delta'$
$\bar{D}, \bar{C}$	$\beta' - \gamma'$
$\bar{D}, B$	0

The quantity $\epsilon' = (-1)^{\Lambda'}$ is the over-all phase factor and is expressed as

$$\Lambda' = aA + dB + cC + bD. \tag{17}$$

In Secs. 4 and 5 it will be found that

$$\Lambda' = 2A = 2B \tag{18}$$

when particles C and D ($\bar{C}$ and $\bar{D}$) are taken to be "particle 2" in both G and H.

[5] In deriving (11) and (13) from (2) and (4), we have used the relation $d_{\lambda\mu}(-\theta) = (-1)^{\lambda-\mu} d_{\lambda\mu}(\theta)$.

4. CROSSING MATRIX IN THE STATIC MODEL AND THE DETERMINATION OF ϵ

Let us consider the following reaction involving π^0:

$$A + \pi^0 \to B + \pi^0. \tag{19}$$

Since the reaction (19) is identical to its u-channel reaction, we can determine ϵ' in this case. If particles A and B are much heavier than pion ($m_a \approx m_b \gg m_\pi$) and if $|s - m_a^2|$, $|s - m_b^2|$, and $|t|$ are far smaller than m_a^2, then the static model is a good approximation. For convenience, let us assume that $m_a = m_b = m$ in the following. In the static model

$$\begin{aligned} s &= (\omega_s + m)^2 - p_s^2 = m^2 + 2m\omega_s + m_\pi^2, \\ t &\approx -2p_s^2(1 - z_s), \\ u &= 2m^2 + 2m_\pi^2 - s - t \\ &\approx m^2 - 2m\omega_s + m_\pi^2 + 2p_s^2(1 - z_s) \end{aligned} \tag{20}$$

in the s channel, and

$$\begin{aligned} u &= m^2 + 2m\omega_u + m_\pi^2, \\ t &\approx -2p_u^2(1 - z_u), \\ s &\approx m^2 - 2m\omega_u + m_\pi^2 + 2p_u^2(1 - z_u) \end{aligned} \tag{21}$$

in the u channel. The ω_i and p_i are the energy and momentum of the pion in the i reaction, and z_i is the cosine of the scattering angle θ_i in the i channel. We find that

$$z_s \approx z_u, \tag{22}$$

since $p_s^2 = \omega_s^2 - m_\pi^2 \approx (-\omega_u)^2 - m_\pi^2 = p_u^2$ in the static model. We also find that

$$\sin\theta_s \approx \sin\theta_u \tag{23}$$

from (16) if we assume that the real point of the path of analytic continuation lies in the domain of the s–t plane bounded by $t = 0$ and $su = (m^2 - m_\pi^2)^2$. Hence, we find

$$\begin{aligned} \cos\chi_a &= \cos\chi_b = -z_s = -z_u, \\ \cos\chi_c &= \cos\chi_d = 1, \\ \sin\chi_a &= \sin\chi_b = \sin\theta_s = \sin\theta_u, \\ \sin\chi_c &= \sin\chi_d = 0, \end{aligned} \tag{24}$$

and

$$\chi_a = \chi_b = \pi - \theta_s = \pi - \theta_u \equiv \pi - \theta \tag{25}$$

in the static model.

Therefore, the crossing relations (11) become

$$\begin{aligned} G_{\beta,\alpha}(s, u) = \epsilon'(-1)^{-\beta} \sum_{\alpha'\beta'} (-1)^{\beta'} \\ \times d^B_{\beta'\beta}(\pi - \theta) d^A_{\alpha'\alpha}(\pi - \theta) H_{\beta',\alpha'}(u, s), \end{aligned} \tag{26}$$

if we take particles C and D to be "particle 2" in G and H. Since

$$G_{\beta,\alpha}(s, u) = (2\pi m/p_s) \sum_J (J + \tfrac{1}{2}) T_{\beta,\alpha}(\omega, J) d^J_{\alpha,\beta}(\theta) \tag{27}$$

and

$$\begin{aligned} H_{\beta',\alpha'}(u, s) \approx (2\pi m/p_s) \\ \times \sum_J (J + \tfrac{1}{2}) T_{\beta',\alpha'}(-\omega, J) d^J_{\alpha',\beta'}(-\theta), \end{aligned} \tag{28}$$

we find the following crossing relations among partial-wave helicity amplitudes:

$$\begin{aligned} T_{\beta,\alpha}(\omega, J) = \epsilon'(-1)^{A-B} \sum_{\alpha'\beta'J'J''} (-1)^{\alpha'+\beta'} \left(\frac{2J' + 1}{2J + 1}\right) \\ \times C(J'AJ''; \beta', \alpha) C(J'AJ''; \alpha', -\alpha') \\ \times C(J''BJ; \alpha + \beta', -\beta') \\ \times C(J''BJ; 0, \beta) T_{\beta',\alpha'}(-\omega, J'). \end{aligned} \tag{29}$$

Helicity states $|JM; \lambda_1\lambda_2\rangle$ and the states used in the conventional static model $|JM; LS\rangle$ are connected through transformation matrix

$$\begin{aligned} \langle JM; LS \,|\, JM; \lambda_1\lambda_2\rangle \\ = \left(\frac{2L + 1}{2J + 1}\right)^{\frac{1}{2}} C(LSJ; 0, \lambda) C(s_1 s_2 s; \lambda_1, -\lambda_2). \end{aligned} \tag{30}$$

Thus, we obtain the following crossing matrix:

$$\begin{aligned} &T(\omega, J, L = 1, B, A) \\ &= \epsilon'(-1)^{A-B} 9(2J + 1)^{-2} C(1BJ; 0, \beta) C(1AJ; 0, \alpha) \\ &\quad \times \sum_{\alpha\beta\alpha'\beta'J'J''} (-1)^{\alpha'+\beta'} C(J'AJ''; \beta', \alpha) C(J'AJ''; \alpha', -\alpha') \\ &\quad \times C(J''BJ; \beta' + \alpha, -\beta') C(J''BJ; 0, \beta) C(1BJ'; 0, \beta') \\ &\quad \times C(1AJ'; 0, \alpha') T(-\omega, J', L = 1, B, A) \\ &= \epsilon' \sum_{J'} (2J' + 1) \begin{Bmatrix} A, 1, J' \\ B, 1, J \end{Bmatrix} T(-\omega, J', L = 1, B, A), \end{aligned} \tag{31}$$

where we have used relations

$$\sum_\beta C(1BJ; 0, \beta) C(J''BJ; 0, \beta) = \tfrac{1}{3}(2J + 1)\delta_{1J''}, \tag{32}$$

$$\begin{aligned} \sum_{\alpha'} (-1)^{\alpha'} C(1AJ'; 0, \alpha') C(J'A1; \alpha', -\alpha') \\ = (-1)^{J'-1}(2J' + \tfrac{1}{3})^{\frac{1}{2}} \end{aligned} \tag{33}$$

and the definition of the 6-j symbol [6]

$$\begin{aligned} \begin{Bmatrix} j_1, j_2, j_{12} \\ j_3, J, j_{23} \end{Bmatrix} = [(2j_{12} + 1)(2j_{23} + 1)]^{-\frac{1}{2}} (-1)^{j_1+j_2+j_3+J} \\ \times \sum_{m_1 m_2} C(j_1 j_2 j_{12}; m_1, m_2) \\ \times C(j_{12} j_3 J; m_1 + m_2, M - m_1 - m_2) \\ \times C(j_2 j_3 j_{23}; m_2, M - m_1 - m_2) \\ \times C(j_1 j_{23} J; m_1, M - m_1). \end{aligned} \tag{34}$$

If we compare the crossing matrix (31) with the $SU(2)$ crossing matrix,[7] we find

$$\epsilon' = (-1)^{2A} = (-1)^{2B}. \tag{35}$$

[6] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J. 1957).

[7] See, for example, V. Singh and B. M. Udgaonkar, Phys. Rev. **149**, 1164 (1966).

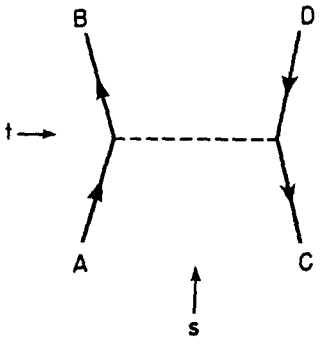


FIG. 4. One-pion-exchange diagram in $p\bar{p}$ scattering.

5. DETERMINATION OF ϵ

In the previous section we have found that

$$\Lambda = 2A + (b + c)C = 2D + (b + c)C, \qquad B = \bar{C}, \quad (36)$$

and

$$\Lambda' = 2A + (b + c)C = 2B + (b + c)C, \qquad C = \bar{D}, \quad (37)$$

when crossed particles are taken to be "particle 2" in both G and F (or H). In order to find $b + c$ we have only to consider the one-pion-exchange diagram of $p\bar{p}$ elastic scattering (Fig. 4). Then its contribution to G and F is given by

$$G_{+-,-+} = -G_{++,--} = F_{++,++} = -F_{++,--} = t/[4(t - m_\pi^2)], \quad (38)$$

where $+(-)$ stands for $+\frac{1}{2}(-\frac{1}{2})$. From Eqs. (2), (36), and (38) we obtain $b + c = 0$ and (9) and (18).

6. DISCUSSION

In order to determine ϵ, we do not have to apply crossing relations to cases of which the crossing relations are known. All that we have to do is to prove

$$G_{\delta\beta,\gamma\alpha}(p_d, -P_b; -P_c, p_a) = (-1)^{2A-\beta+\gamma} F_{\delta\gamma,\beta\alpha}(p_d, P_c; P_b, p_a) \quad (39)$$

when particles B and C ($= \bar{B}$) are taken to be "particle 2" in both G and F. This problem will be discussed elsewhere.

For other possible choices of the position of the real point of the trajectory in the s–t plane, we can determine ϵ by making use of our knowledge on the analytic properties of helicity amplitudes. The analytic properties of helicity amplitudes were studied by assuming the crossing relations.[2] However, what we have used in Ref. 2 is the magnitude of $\cos\psi_i$ and the fact that the helicity does not change sign in the crossing process. That is, actually we have not assumed the crossing relations to study the analytic properties of helicity amplitudes, but we have assumed the Lorentz invariance of the theory and the assumption that scalar–scalar scattering amplitudes satisfy Mandelstam representations.[1]

Note added in Proof: The knowledge of the sine and cosine of ψ_i is not sufficient to determine ψ_i. In this paper we have assumed that $\psi_a = \psi_d$ and $\psi_b = \psi_c$ in the limit $m_a = m_d$ and $m_b = m_c$.

By defining phases of initial and final helicity states and by proving (39), we have determined the over-all phase factor of the crossing relations for any processes in International Centre for Theoretical Physics, Trieste, Preprint IC/69/38.

ACKNOWLEDGMENT

The author thanks A. Sakai for discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Sine-Gordon Equation

JULIO RUBINSTEIN*
Physics Department, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

(Received 4 March 1968; Revised Manuscript Received 2 June 1969)

The equation $\phi_{tt} - \phi_{xx} + m^2 \sin \phi = 0$ is presented as a model field theory and studied in detail. It exhibits discrete conserved quantities and extended particle states, with the proper behavior regarding covariance, stability, etc. Those particles do not interact with small oscillations, and we show that this (plus a few reasonable requirements) defines the model uniquely. A connection with the Korteweg–de Vries equation, which has similar properties, is established.

1. INTRODUCTION

We study a one-dimensional nonlinear field equation, in which the range of the field is a 1-sphere. Skyrme[1] obtained essentially the same equation by working on a nonlinear theory of strong interactions, in which the components ϕ_ρ of a field ϕ are subject to the constraint $\sum_1^N \phi_\rho^2 = \text{const}$, so ϕ describes a 1-sphere if $N = 2$. In a subsequent paper[2] the model is quantized by introducing discontinuous functions as localized particle states. The classical equation is analyzed in great detail by Perring and Skyrme,[3] who give a number of exact solutions with obvious physical interpretations. Part of their results is obtained independently by Enz.[4]

Finkelstein and Misner[5] propose the above-mentioned model in the course of a study of topological conservation laws; in their work, as in ours, only continuous fields appear, i.e., continuous mappings of space–time into the field space Φ. For each $t = \text{const}$ those mappings determine *a fortiori* continuous fields $\phi(x)$, to which conserved discrete quantities can be assigned.[5] Restricting Φ to be a manifold, a point $\phi \in \Phi$ can be represented by n real numbers ϕ_α, the "components" of the field, with n constant over Φ. If a manifold Φ is a vector space, we say that it is trivial because it can be shown[5] that it does not exhibit discrete conserved quantities in the above sense. In particular, the *only* nontrivial one-dimensional ($n = 1$) manifold is the 1-sphere, and much of this paper will deal with a one-(space)-dimensional equation in which the field ϕ ranges over the one-sphere, i.e., $\phi = 2\pi$ and $\phi = 0$ are considered identical expressions.

In Sec. 2 we give the motivation of the problem, and certain particlelike states of the system are shown in Sec. 3; we call those particle states *kinks*, and the interaction between kinks and between a kink and a perturbation are analyzed in Secs. 4 and 5. In Sec. 6 the system is quantized, and it is shown that the classical behavior found before is essentially unchanged. In Sec. 7 we prove that a set of plausible conditions lead uniquely to the problem developed above. Finally, we show in Sec. 8 that the Korteweg–de Vries equation, which appears in several independent branches of physics, is intimately connected to our problem.

Some of the results obtained here could be concisely expressed using topological terms, as done elsewhere.[6] Here we shall, however, renounce compactness in favor of a more intuitive presentation.

2. THE ONE-DIMENSIONAL MODEL

The simplest topologically nontrivial manifold is the 1-sphere, i.e., the space of the real numbers modulo 2π with the usual metric, or the subspace of the real Euclidean plane (x_1, x_2) given by $x_1^2 + x_2^2 = 1$. In what follows we assume that the range of the field $\phi(x, t)$ is such a manifold, if not otherwise stated.

The Klein–Gordon Lagrangian density in one space dimension,

$$\mathfrak{L} = -\tfrac{1}{2}(\phi_x^2 - \phi_t^2 + m^2\phi^2), \quad \hbar = c = 1, \quad (2.1)$$

is not admissible in this case, since we have $\mathfrak{L}(\phi) \neq \mathfrak{L}(\phi + 2\pi)$. One way of fixing this is to modify (2.1) as follows:

$$\begin{aligned}\mathfrak{L} &= -\tfrac{1}{2}(\phi_x^2 - \phi_t^2 + 2m^2(1 - \cos\phi)) \\ &= -\tfrac{1}{2}(\phi_x^2 - \phi_t^2 + 4m^2 \sin^2 \phi/2); \qquad (2.2)\end{aligned}$$

that is, we replace the mass term of (2.1) by a periodic function of ϕ, with period 2π. Of all such functions, $2m^2(1 - \cos\phi)$ is the simplest one which becomes $m^2\phi^2$ for small ϕ, giving the Klein–Gordon equation in the low-amplitude limit.

* Supported by the National Research Council of Canada. Present address: Centre for Research in Experimental Space Science, York University, Toronto.

[1] T. H. R. Skyrme, Proc. Roy. Soc. (London) **A247**, 260 (1958).
[2] T. H. R. Skyrme, Proc. Roy. Soc. (London) **A262**, 237 (1961).
[3] J. K. Perring and T. H. R. Skyrme, Nucl. Phys. **31**, 550 (1962).
[4] U. Enz, Phys. Rev. **131**, 1392 (1963).
[5] D. Finkelstein and C. W. Misner, Ann. Phys. (N.Y.) **6**, 230, (1959).
[6] D. Finkelstein and J. Rubinstein, J. Math. Phys. **9**, 1762 (1968).

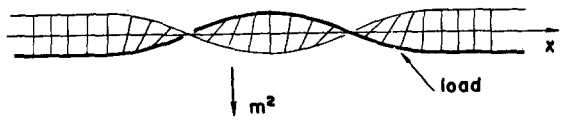


FIG. 1. A classical system described by the sine-Gordon equation: An infinite elastic ribbon has a load on one of the edges and is in a gravitational field perpendicular to its axis; the only allowed motion is torsion around this axis. The angle between the load and the gravitational field is the field variable ϕ. If ϕ is defined by looking from $x = \infty$, the figure shows a 1-kink state, namely $(2\pi)^{-1}[\phi(\infty) - \phi(-\infty)] \equiv N = 1$. Given $\phi(x)$, any continuous deformation $\phi_s(x)$ [e.g., the time evolution $\phi(x, t)$] which preserves $\phi(\pm\infty)$ will conserve N, and its possible values are the integers if $\phi(\pm\infty) = 0 \pmod{2\pi}$.

Notice that ϕ in (2.2) is considered dimensionless and that

$$\mathcal{H} = \tfrac{1}{2}(\phi_t^2 + \phi_x^2 + 4m^2 \sin^2 \tfrac{1}{2}\phi)$$

has the correct sign and dimension.

If we impose the boundary conditions

$$\phi(-\infty, t) = \phi(\infty, t) = 0 \pmod{2\pi}, \qquad (2.3)$$

the solutions will exhibit a conserved quantity, namely,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\partial\phi(x, t)}{\partial x}\, dx = \frac{1}{2\pi}[\phi(\infty, t) - \phi(-\infty, t)] = N, \quad \text{an integer.} \qquad (2.4)$$

N is additive in the following sense (all fields are taken at the same time, which we omit): Let $a < b < c$ be real numbers; any field $\phi_1(x)$ satisfying (2.3) can be deformed continuously, keeping $\phi_1(\pm\infty)$ constant in the process, into a field $\phi_1'(x)$ such that $\phi_1'(x) = 0 \pmod{2\pi}$ for $x \notin (a, b)$; such a field has the same N, say N_1, as $\phi_1(x)$. Analogously, a field $\phi_2(x)$ can be deformed into $\phi_2'(x)$, with $\phi_2'(x) = 0$ for $x \notin (b, c)$. Then the juxtaposition

$$\begin{aligned} \phi(x) &= \phi_1'(x), && x \in (a, b), \\ &= \phi_2'(x), && x \in (b, c), \\ &= 0 \pmod{2\pi}, && x \notin (a, c), \end{aligned}$$

has

$$N = N_1 + N_2; \qquad (2.5)$$

upon interchange of the supports of ϕ_1' and ϕ_2' we still have (2.5). Also, if $N_1 = 1$ and $N_2 = -1$, any N can be obtained juxtaposing N ϕ_1' fields or $(-N)$ ϕ_2' fields. The system described by (2.2) can also be thought of as a strip[5] of semiwidth 1, torsion modulus 1, mass density 1 along one edge, placed in a uniform gravitational field of strength m^2; $\phi(x, t)$ is the angular displacement at (x, t) (Fig. 1). If (2.3) is satisfied, $2\pi N$ is the total torsion of the strip, from $x = -\infty$ to $x = \infty$. In terms of this visualization, the process developed in the preceding paragraph can be described as "ironing $\phi_1(x)$ into the interval (a, b)," etc.

3. BOUND STATES

The equation of motion corresponding to (2.2) is

$$\phi_{tt} - \phi_{xx} + m^2 \sin\phi = 0. \qquad (3.1)$$

We shall call (3.1), for obvious reasons, the sine-Gordon equation[7]; in the static case it becomes

$$\phi^0_{xx} = m^2 \sin\phi^0. \qquad (3.2)$$

It is easy to see[1,3–5] that (3.2) admits only two bound [i.e., such that $\phi \to 0 \pmod{2\pi}$ as $|x| \to \infty$] solutions, except for a displacement, and they correspond to $N = \pm 1$:

$$\phi^0 = 4 \tan^{-1} e^{\pm m(x-q)}, \qquad (3.3)$$

with integration constants $e^{\mp mq}$. We see that ϕ^0 satisfies our boundary condition and that $\phi^0 =$ "max $(\text{mod } 2\pi)$" at $x = q$, so that q can be interpreted as the position of our "ϕ-packet" or "kink," as we shall call it in the following.[6] The $\pm$ signs correspond evidently to the two possible helicities of the kink.

Putting $\phi(x, t) = \phi^v(x - vt)$ in the sine-Gordon equation, we get

$$(v^2 - 1)\phi^{v\prime\prime} + m^2 \sin\phi^v = 0,$$

and, upon comparison with (3.2) and (3.3),

$$\phi^v = 4 \tan^{-1} \exp\left[\frac{m}{\pm(1 - v^2)^{\frac{1}{2}}}(x - q(t))\right], \qquad (3.4)$$

where $q(t) = q + vt = q(0) + vt$ is the position of the kink at time t. The width of this "running kink" is $(1 - v^2)^{\frac{1}{2}}(1/m)$, i.e., $(1 - v^2)^{\frac{1}{2}}$ times the static kink width. This Lorentz contraction is, of course, a consequence of the relativistic invariance of the theory.

The energy of the running kink is

$$\begin{aligned} E &= \int_{-\infty}^{\infty} dx \mathcal{H}(\phi^v) \\ &= \frac{1}{2}\int_{-\infty}^{\infty} dx\left(\phi_t^{v2} + \phi_x^{v2} + 4m^2 \sin^2 \tfrac{1}{2}\phi^v\right) \\ &= \frac{4m}{(1 - v^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty} \cosh^{-2} y\, dy \\ &= \frac{8m}{(1 - v^2)^{\frac{1}{2}}}. \end{aligned}$$

In particular, the rest mass of the kink is $M = 8m$. Note, again, that E exhibits the right relativistic dependence on the speed v of the kink. The factor 8 is due to the choice of an over-all factor $\frac{1}{2}$ in (2.2), and to that extent somewhat arbitrary.

Thus, the classical (c-number) kink is a relativistically invariant model of an extended body in one dimension.[4]

[7] The name is due to Professor Martin Kruskal.

4. INTERACTION BETWEEN KINKS

For short periods of time the deformation of several kinks in interaction will be small and we may look for solutions of the form

$$\phi(x, t) = \psi(x, t) + \sum_{i=1}^{n} \phi_i(x, t) \tag{4.1}$$

with

$$\phi_i(x, t) = 4 \tan^{-1} \exp N_i \frac{m}{(1 - v_i^2)^{\frac{1}{2}}} (x - v_i t - q_i(0)), \tag{4.2}$$

$$N_i = \pm 1, \tag{4.3}$$

and $\psi \ll 1$. In particular, with

$$v_i = \partial_t \psi(x, 0) = \psi(x, 0) = 0, \tag{4.4}$$

such a solution will tell us the direction and strength of the force acting on each kink at rest at $t = 0$.

For simplicity we restrict ourselves to $n = 2$, and we put $-q_1 = q_2 = q$; inserting (4.1) into the sine-Gordon equation, we obtain, at $t = 0$,

$$\begin{aligned} &\frac{1}{m^2} \partial_t^2 \psi(x, 0) \\ &= \sum \sin \phi_i - \sin \sum \phi_i \\ &= \sin \phi_1 (1 - \cos \phi_2) + \sin \phi_2 (1 - \cos \phi_1) \\ &= -4 \frac{N_1 \sinh m(x + q) + N_2 \sinh m(x - q)}{\cosh^2 m(x + q) \cosh^2 m(x - q)}. \end{aligned} \tag{4.5}$$

In Figs. 2 and 3 we plot $\phi(x, 0) = \phi_1(x, 0) + \phi_2(x, 0)$ in full line and $\phi(x, t) \simeq \phi(x, 0) + \frac{1}{2} \partial_t^2 \psi(x, 0) t^2$ (t small) in dotted line; this shows that the kinks repel if $N_1 = N_2$ and attract if $N_1 = -N_2$ (the cases $N_1 = N_2 = -1$ and $-N_1 = N_2 = +1$ are represented by Figs. 2 and 3 with $\phi \to -\phi$).

If $q \gg 1/m$ and for $x \simeq q$, we get from (4.5) the change in ϕ_2:

$$\begin{aligned} \frac{1}{m^2} \partial_t^2 \psi(x, 0) &\simeq \frac{-4N_1 \sinh m(x + q)}{\cosh^2 m(x + q) \cosh^2 m(x - q)} \\ &\simeq \frac{-4N_1 \tanh 2mq}{\cosh 2mq \cosh^2 m(x - q)} \\ &\simeq \frac{-8N_1 e^{-2mq}}{\cosh^2 m(x - q)}. \end{aligned} \tag{4.6}$$

From (4.2), the change in $\phi_2(x, 0)$ generated by a small change in q, i.e., a rigid displacement of the kink, is

$$\delta\phi_2 = \frac{\partial \phi_2(x, 0)}{\partial q} \cdot \delta q = \frac{-2N_2 m}{\cosh m(x - q)} \delta q, \tag{4.7}$$

and, if δq changes in time,

$$\partial_t^2 (\delta\phi_2) = \frac{-2N_2 m}{\cosh m(x - q)} \delta\ddot{q}. \tag{4.8}$$

Comparing (4.8) and (4.6) at $x \simeq q$, we see that under their interaction the kinks tend to contract and that the force acting on ϕ_2 is

$$(8m)(\delta\ddot{q}) \simeq 32 m^2 N_1 N_2 e^{-m(2q)}. \tag{4.9}$$

For completeness we give here two exact solutions of (3.1) found by Perring and Skyrme[3]: Putting $(1 - v^2)^{-\frac{1}{2}} = \gamma$, $m = 1$, and $\alpha + 2\pi = \phi$, they are

$$\begin{aligned} \phi_a &= 4 \tan^{-1} - \frac{\cosh \gamma v t}{v \sinh \gamma x} \\ &\xrightarrow[x \to -\infty]{} 4 \tan^{-1} \frac{2}{v} e^{\gamma x} \cosh \gamma v t \\ &\xrightarrow[t \to -\infty]{} 4 \tan^{-1} \frac{1}{v} e^{\gamma(x - vt)} \\ &\xrightarrow[t \to \infty]{} 4 \tan^{-1} \frac{1}{v} e^{\gamma(x + vt)}, \end{aligned} \tag{4.10}$$

$$\begin{aligned} \phi_a &= 4 \tan^{-1} \frac{v \sinh \gamma x}{\cosh \gamma v t} + 2\pi \\ &\xrightarrow[x \to \infty]{} 4 \tan^{-1} \frac{v e^{\gamma x}}{2 \cosh \gamma v t} + 2\pi \\ &\xrightarrow[t \to -\infty]{} 4 \tan^{-1} v e^{\gamma(x + vt)} + 2\pi \\ &\xrightarrow[t \to \infty]{} 4 \tan^{-1} v e^{\gamma(x - vt)} + 2\pi, \end{aligned}$$

$$\begin{aligned} \phi_b &= 4 \tan^{-1} \frac{v \cosh \gamma x}{\sinh \gamma v t} \\ &\xrightarrow[t \to -\infty]{} 4 \tan^{-1} 2v \cosh \gamma x e^{\gamma v t} \\ &\xrightarrow[x \to -\infty]{} 4 \tan^{-1} v e^{-\gamma(x - vt)} \\ &\xrightarrow[x \to \infty]{} 4 \tan^{-1} v e^{\gamma(x + vt)}, \end{aligned} \tag{4.11}$$

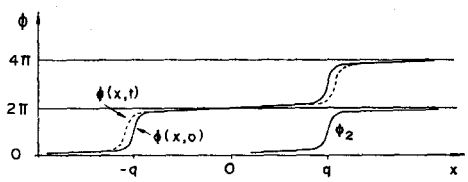


FIG. 2. Interaction between kinks: Two $N = 1$ kinks are initially at rest at $x = -q$ and $x = q$; the dotted line is $\phi(x, t)$ for $t > 0$ but small, showing that the kinks repel each other.

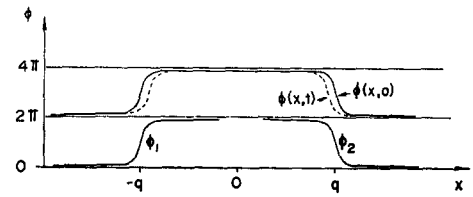


FIG. 3. As in Fig. 2, except that now we have an antikink at $x = q$ ($N_2 = -1$); here $\phi(x, t)$ shows that a kink and an antikink attract each other.

$$\phi_b = 4\tan^{-1}\frac{\sinh \gamma vt}{v\cosh \gamma x} + 2\pi$$
$$\xrightarrow[t\to\infty]{} 4\tan^{-1}\frac{e^{\gamma vt}}{2v\cosh \gamma x} + 2\pi$$
$$\xrightarrow[x\to-\infty]{} 4\tan^{-1}\frac{1}{v}e^{\gamma(x+vt)} + 2\pi$$
$$\xrightarrow[x\to\infty]{} 4\tan^{-1}\frac{1}{v}e^{-\gamma(x-vt)} + 2\pi.$$

Except for the unimportant factors $1/v$ and v (which can be absorbed in the phase), (4.10) shows that ϕ_a represents two $N = 1$ kinks with speeds $\pm v$ approaching the origin for $t \to -\infty$ and moving away from it for $t \to +\infty$. If v is small enough ($\lesssim 0.05$), ϕ_a departs little from a linear combination of two kinks, with the distance of closest approach larger than twice the kink size, so that ϕ_a can be recognized as the bouncing of two kinks. For larger v's, ϕ_a is strongly distorted and no such interpretation is possible.

Analogously, ϕ_b represents, at $t = -\infty$, an $N = -1$ kink approaching the origin with speed v from $x = -\infty$ and an $N = 1$ kink coming from $x = \infty$ with speed $-v$. Those pass through each other and, at $t = \infty$, ϕ_{-1} is at $x = \infty$ and ϕ_{+1} at $x = -\infty$; $\phi_b(x, +\infty)$ has the form of $\phi(x, 0)$ of Fig. 3.

As said in Sec. 1, we are interested in continuous fields only, so what follows is presented as just a curiosity. A discontinuous type of solution of (3.1) is[1]

$$\phi = n\pi \sum_i \pm\, \theta(x \pm t - x_i^0), \qquad (4.12)$$

with $\theta(x) = 0$ for $x \leq 0$, and $= 1$ for $x > 0$, and the x_i^0 real constants. In particular, $\phi = 2\pi\theta(x - t)$ can be interpreted as the limit as $v \to 1$ of ϕ^v:

$$\phi^v = 4\tan^{-1}\exp\frac{x - vt}{(1 - v^2)^{\frac{1}{2}}} \xrightarrow[v\to 1]{} 2\pi\theta(x - t). \qquad (4.13)$$

Since elements like $2\pi\theta(x \pm t - x_n^0)$ do not interact with each other nor with ϕ, where ϕ satisfies the sine-Gordon equation, we can have an indeterminate number of "nothings" (for them $\phi = 0 \bmod 2\pi$), with an arbitrary total particle number and an energy

$$\sum \lim_{v_i\to 1} \frac{1}{(1 - v_i^2)^{\frac{1}{2}}} = \infty.$$

Although this strongly suggests a nonquantum approach to the divergence problem in field theory, we have not pursued this line further.

5. SMALL OSCILLATIONS

We next analyze small departures from the single kink state, i.e., we consider $\phi = \phi^0 + \psi$, $\psi \ll 1$. The more general case of taking ϕ^v instead of ϕ^0 can be reduced to this one by a Lorentz transformation. Also, the substitution $x - q \to x$ allows us to take $q = 0$ in ϕ^0.

Expanding $\sin\phi = \sin\phi^0 + \psi\cos\phi^0$, sine-Gordon gives

$$\psi_{tt} - \psi_{xx} + (m^2\cos\phi^0)\psi$$
$$= \psi_{tt} - \psi_{xx} + m^2(1 - 2\cosh^{-2} mx)\psi = 0, \quad (5.1)$$

in which we put $\psi(x, t) = \psi(x)e^{-i\omega t}$, obtaining

$$\psi'' + (\omega^2 - m^2 + 2m^2\cosh^{-2} mx)\psi = 0. \quad (5.2)$$

Equation (5.2) is solved by Landau and Lifschitz[8] (p. 76); there is only one possible bound level, with

$$\omega^2 = 0. \qquad (5.3)$$

The corresponding solution is a combination of hypergeometric functions which in our case reduces to

$$\psi_b(x) \sim 1/\cosh mx. \qquad (5.4)$$

It is easier to check directly that (5.4) is a solution of (5.2) if (5.3) holds. This is the only bound solution because a one-dimensional bound problem is nondegenerate.

From (4.7) we see that (5.4) is, except for a constant, the change in $\phi^0|_{q=0}$ generated by a rigid displacement of the kink.

The above comments show that ϕ^0 is stable against almost every bound variation, the only exception being a rigid translation; against it ϕ^0 is in neutral equilibrium, because $\omega^2 = 0$. Note, again, the analogy with particles.

The remaining (unbounded, scattering) solutions of (5.2) are obtained by Morse and Feshbach,[9] also as a combination of hypergeometric functions; after some manipulations they can be reduced to

$$\psi_k(x) = (2\pi)^{\frac{1}{2}}\omega e^{ikx}(k + im\tanh mx), \qquad (5.5)$$

where $k^2 = \omega^2 - m^2$ and $-\infty < k < \infty$.

The fact that (5.2) has this type of solutions means that the potential $-(m^2/2)\cosh^{-2} mx$, for which (5.2) is the Schrödinger equation (particle of mass = 1), is completely transparent at any energy of the incoming particle. This is a remarkable property indeed, and we shall elaborate on it later on.

It is easy to check that (5.5) is a solution of (5.2) and

[8] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Non-Relativistic Theory* (Pergamon Press, London, 1958).

[9] P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., New York, 1953).

also that

$$\int_{-\infty}^{\infty} \psi_k^*(x)\psi_{k'}(x)\,dx = \delta(k-k'),$$
$$\int_{-\infty}^{\infty} \psi_k(x)\psi_b(x)\,dx = 0, \qquad (5.6)$$
$$\int_{-\infty}^{\infty} \psi_b(x)\psi_b(x)\,dx = 8m.$$

For completeness we give here $\mathcal{H}$ in terms of ψ and $\dot{\psi} = \pi$:

$$\mathcal{H} = \tfrac{1}{2}(\phi_x^{02} + 4m^2 \sin^2 \tfrac{1}{2}\phi^0) + \tfrac{1}{2}(\dot{\psi}^2 + \psi_x^2 + m^2(1 - 2\cosh^{-2} mx)\psi^2) \qquad (5.7)$$

and the closure relation

$$\int_{-\infty}^{\infty} dk\,\psi_k^*(x)\psi_k(x') + \frac{1}{8m}\psi_b(x)\psi_b(x') = \delta(x-x').$$

Notice that (5.5) does not satisfy boundary conditions (i.e., $\phi_0 + \psi_k$ does not), but it can be understood in the usual way as a contribution to a packet $\psi(x) = \int a(k)\psi_k(x)\,dk$ so that $\psi(x) \to 0$, $|x| \to \infty$.

6. QUANTIZATION

In what follows we consider ϕ^0 a c number and subject ψ to the usual boson quantization procedure. We expand

$$\psi(x,t) = \int_{-\infty}^{\infty} \frac{dk}{(2\omega)^{\frac{1}{2}}}\,[a_k e^{-i\omega t}\psi_k(x) + a_k^+ e^{i\omega t}\psi_k^*(x)] + q\psi_b(x), \qquad (6.1)$$

with a_k time-dependent and $q = q^+$ satisfying $\ddot{q} = 0$. Then (6.1) is the general Hermitian solution of (5.1). Using (3.3), (5.6), (5.7), and (6.1), we get, after some computations,

$$H = \int_{-\infty}^{\infty} dx\mathcal{H} = 8m + \frac{8m}{2}\dot{q}^2 + \frac{1}{2}\int_{-\infty}^{\infty} dk\omega(a_k a_k^+ + a_k^+ a_k). \qquad (6.2)$$

The last term in (6.2) has the usual form for bosons, whereas the former ones are the nonrelativistic approximation for the Hamiltonian of a free particle of mass $8m$ and position operator q. There is no interaction between the bosons and the kink, as should have been expected in virtue of the remarks made after (5.5).

Imposing on

$$i[\psi(x,t), \psi(x',t')] = D(x,t,x',t'), \qquad (6.3)$$

say, the standard causal ($D(x,t,x',t) = 0$), dynamical (5.1), and canonical ($i[\pi(x,t), \psi(x',t)] = \delta(x-x')$) conditions, we obtain

$$D = i\int_{-\infty}^{\infty} \frac{dk}{2\omega}\,[e^{-i\omega(t-t')} - e^{i\omega(t-t')}]\psi_k(x)\psi_k^*(x') + \frac{t-t'}{8m}\psi_b(x)\psi_b(x'). \qquad (6.4)$$

Comparing (6.3) and (6.4), we get

$$8m[q,\dot{q}] = -i, \quad [a_k, a_k^+] = \delta(k-k'),$$
$$[a_k, a_{k'}] = [a_k, q] = [a_k, \dot{q}] = 0. \qquad (6.5)$$

The relations (6.5) imply that (6.2) contains an infinite zero-point energy (as in the $\phi^0 = 0$ case) and that $8m\dot{q}$ can be interpreted, in this approximation, as the *bona fide* momentum of the kink.

One point of caution should be made about the quantization process: Although ψ is assumed small in (5.1), it is easy to see that (6.1) gives in the vacuum (for instance)

$$\langle\psi\rangle_0 = 0, \quad \text{but} \quad \langle\psi^2\rangle_0 = \infty.$$

Nevertheless, we may redefine $\langle\psi^2\rangle_0 = \langle\dot{\psi}^2\rangle_0 = \langle\psi_x^2\rangle_0 = 0$ and keep the expression (5.7) for $\mathcal{H}$, so that $\langle H\rangle_0 = 8m$. Although arbitrary, this is consistent with the usual zero-point energy subtraction in (6.2), corresponding to the case in which there is no kink.

Our method of quantization is different from that of Skyrme,[2] who introduces singular functions in order to represent a localized kink, and from Finkelstein and Misner,[5] who quantize in a finite box of size Ω—as a consequence of which the contribution of the kink to the total energy disappears when $\Omega \to \infty$.

7. TRANSPARENCY OF KINKS

The remarks made just after (5.5) lead us to the following question: Is the absence of reflection an accidental feature arising from the shape of the mass term in the sine-Gordon equation, as stated by Perring and Skyrme[3]? Although we do not fully answer the question, we will show that, with some qualifications, the sine-Gordon mass term is the only one that gives zero reflection and is compatible with our ϕ manifold; this suggests that the sine-Gordon equation could be a unique result of some set of phenomenological requirements, one of which would be the strength of the coupling of the ψ field to the ϕ^0 particle.

In this section we do *not* assume from the start that ϕ ranges over a 1-sphere. Let the one-dimensional field equation

$$\phi_{tt} - \phi_{xx} + f(\phi) = 0 \qquad (7.1)$$

admit the static solution (or solutions) $\phi_0(x)$:

$$\phi_0'' = (\phi_0); \qquad (7.2)$$

expanding

$$\phi = \phi_0 + \psi, \quad f(\phi) = f(\phi_0) + f'(\phi_0)\psi,$$
$$\psi(x, t) = e^{-i\omega t}\psi(x),$$

we obtain

$$-\psi'' + [f'(\phi_0) - \omega^2]\psi = 0. \tag{7.3}$$

We shall assume that:

(a) $f(0) = 0$;
(b) $f'(0) = m^2$ (say);
(c) $\phi_0 \to 0$ as $x \to -\infty$;
(d) ϕ_0 is stable against small perturbations and
(e) in neutral equilibrium against a rigid translation; (7.4)
(f) (7.3) admits solutions behaving as $e^{i(\pm\alpha+kx)}$, $x \to \pm\infty$, $-\infty < k < \infty$; i.e., that (7.3) is a Schrödinger equation with zero reflection coefficient *at all k*.

By (a) and (b), (7.1) becomes, for small ϕ, a Klein–Gordon equation; (c) is an arbitrary choice of reference value, and (d) and (e) assure proper particlelike behavior of ϕ_0, while (f) amounts to saying that the field ψ does not interact, in this approximation, with the particle represented by ϕ_0.

In addition to those "running" solutions (f), (7.3) will also admit, in general, a discrete set of bound solutions:

$$-\psi_i'' + [f'(\phi_0) - \omega_i^2]\psi_i = 0. \tag{7.5}$$

The mathematical problem posed by (7.4f) has been completely solved by Kay and Moses,[10] who showed that the ψ_i *must* be one of the n solutions of

$$-\psi'' + [V_n(x) + k_i^2]\psi_i = 0, \tag{7.6}$$

with

$$V_n(x) = -2d^2/dx^2 \log\det\left(\delta_{ij} + \frac{(A_iA_j)^{\frac{1}{2}}}{k_i + k_j} e^{(k_i+k_j)x}\right) \leq 0, \tag{7.7}$$

and A_i and k_i are $2n$ positive constants, otherwise arbitrary.

Comparing (7.5) and (7.6), we obtain (after reordering if necessary)

$$f'(\phi_0(x)) = V_n(x) + k_i^2 + \omega_i^2. \tag{7.8}$$

Since $V_n(x) \to 0$ as $|x| \to \infty$, (7.4b) and (7.4c) give

$$\omega_i^2 + k_i^2 = m^2.$$

Numbering the ω_i so that $\omega_i^2 \leq \omega_{i+1}^2$, from (7.4d) and (7.4e) we obtain $\omega_1 = 0$; then $k_1 = m$ and

$$\omega_i^2 = m^2 - k_i^2, \quad i = 1, \cdots, n, \quad k_i \geq k_{i+1}. \tag{7.9}$$

We can take the k_i to be all different from each other without loss of generality, because, as is shown in the Appendix, if there are groups of r, s, $\cdots$ k_i equal to each other, $V_n(x)$ reduces to $V_{n-r+1-s+1\cdots}(x)$. Then (7.9) says that there are n ω_i for each V_n, and $0 \leq \omega_i < m$. In a quantized version this would have an obvious physical meaning: The fields $\phi_0(x) + \psi_i(x, t)$ describe bound states of a ϕ_0 particle with ψ particles, with the binding energy

$$E_i = m - \omega_i, \quad \hbar = c = 1.$$

From (7.2) and (7.8),

$$-\phi_0''' + f'(\phi_0)\phi_0' = -(\phi_0')'' + [V_n(x) + k_1^2](\phi_0') = 0, \tag{7.10}$$

i.e., by (7.6),

$$\phi_0'(x) \sim \psi_1(x), \tag{7.11}$$

as should have been expected since $\omega_1 = 0$ corresponds to a rigid translation of ϕ_0 and

$$\phi_0'(x) = -\frac{d}{dx_0}\phi_0(x - x_0)|_{x=x_0}.$$

Then, by (7.4c),

$$\phi_0(\infty) = T_n, \text{ say,}$$
$$= \int_{-\infty}^{\infty} \phi_0'(x)\,dx \sim \int_{-\infty}^{\infty} \psi_1(x)\,dx \neq 0, \tag{7.12}$$

by the well-known oscillation theorem, since ψ_1 is the "ground-state" wavefunction of (7.6). This, incidentally, also implies that ϕ_0 is monotonic. T_n is finite because $V(x) \to 0$ as $|x| \to \infty$, so (7.6) shows that $\psi_1(x) \to e^{\pm k_1 x}$, as $x \to \mp\infty$, and the $\int$ in (7.12) converges. This shows that our conditions are sufficient for the system to have bound states, in the sense that $\lim \phi_0$, as $x \to -\infty$, and $\lim \phi_0$, as $x \to \infty$, exist and are finite.

Case $n = 1$: It can be shown[10] that $\bar{\psi}_i = \psi_i(A_i)^{\frac{1}{2}}$ satisfies, for any n,

$$\left(\delta_{ij} + \frac{(A_jA_j)^{\frac{1}{2}}}{k_i + k_j} e^{(k_i+k_j)x}\right)\bar{\psi}_j = -A_i^{\frac{1}{2}} e^{k_i x}; \tag{7.13}$$

then, for $n = 1$ we have

$$\psi_1 = \frac{-A_1 e^{k_1 x}}{1 + (2k_1)^{-1}A_1 e^{2k_1 x}} = \frac{\alpha}{\cosh k_1(x - x_0)},$$
$$x_0 = \frac{1}{2k_1}\log\frac{2k_1}{A_1}. \tag{7.14}$$

[10] I. Kay and H. E. Moses, J. Appl. Phys. 27, 1503 (1956). I am indebted to Professor J. B. Keller for this reference.

Then

$$\phi_0(x) = \alpha' \int_{-\infty}^{x} \psi_1(x)\, dx = \frac{2T_1}{\pi} \tan^{-1} e^{k_1(x-x_0)},$$

and (7.2) becomes

$$f\left(\frac{2T_1}{\pi} \tan^{-1} e^{k_1(x-x_0)}\right) = -\frac{T_1 k_1^2}{\pi} \frac{\tanh k_1(x-x_0)}{\cosh k_1(x-x_0)} = +\frac{T_1 k_1^2}{2\pi} \sin 4 \tan^{-1} e^{k_1(x-x_0)} = \frac{T_1}{2\pi} k_1^2 \sin \frac{2\pi}{T_1}\left(\frac{2T_1}{\pi} \tan^{-1} e^{k_1(x-x_0)}\right),$$

or

$$f(\phi) = \frac{T_1}{2\pi} k_1^2 \sin \frac{2\pi}{T_1} \phi, \tag{7.15}$$

by which (7.1) becomes the sine-Gordon equation; the arbitrariness in T_1 corresponds to the possibility of a scale transformation of the 1-sphere, namely $(0, 2\pi) \to (0, T_1)$.

Indeed, we have proven (7.15) only for ϕ_0—i.e., for $0 \le \phi \le T_1$—but we may, for instance, extend it to all real ϕ by analytic continuation.

Case $n > 1$: Normalizing $\psi_1(x)$ so as to make (7.11) an equality, we have

$$f^{(m)}(\phi_0) = \frac{1}{\psi_1}\frac{d}{dx} f^{(m-1)}(\phi_0) = \cdots = \left(\frac{1}{\psi_1}\frac{d}{dx}\right)^{m-1} f'(\phi_0) = \left(\frac{1}{\psi_1}\frac{d}{dx}\right)^{m-1} V_n(x), \quad m > 1. \tag{7.16}$$

The expression (7.7) can be brought into the form (3.7) of[10]

$$V_n(x) = -2\frac{\left\{\sum_{\nu=1}^{M} \beta_\nu^2 \alpha_\nu e^{\beta_\nu x} + \frac{1}{2}\sum_{\nu,\mu} \alpha_\nu \alpha_\mu (\beta_\nu - \beta_\mu)^2 e^{(\beta_\nu+\beta_\mu)x}\right\}}{1 + 2\sum_{\nu} \alpha_\nu e^{\beta_\nu x} + \sum_{\nu,\mu} \alpha_\nu \alpha_\mu e^{(\beta_\nu+\beta_\mu)x}}, \tag{7.17}$$

where the α's are positive constants related to the A's and k's, and the β's are all the possible sums without repetition of k's, times 2; ordering them so that $\beta_\nu > \beta_{\nu+1}$, we have

$$\beta_1 = 2\sum_{1}^{n} k_i, \quad \beta_2 = 2\sum_{1}^{n-1} k_i, \quad \beta_3 = 2\sum_{1}^{n-2} k_i + 2k_n, \quad \text{etc.}, \tag{7.18}$$

so that, in a first approximation,

$$V_n(x) \to -2\frac{\alpha_1\alpha_2(\beta_1 - \beta_2)^2 e^{(\beta_1+\beta_2)x}}{\alpha_1^2 e^{2\beta_1 x}} \sim e^{(\beta_2-\beta_1)x} = e^{-2k_n x}, \quad x \to \infty, \quad \to -2\beta_M \alpha_M e^{\beta_M x} \sim e^{\beta_M x} = e^{2k_n x}, \quad x \to -\infty. \tag{7.19}$$

Also, as we saw before,

$$\psi_1(x) \to e^{\mp k_1 x}, \quad x \to \pm\infty. \tag{7.20}$$

Then

$$f'''(T_n, 0) = \lim_{x\to\pm\infty}\left(\frac{1}{\psi_1}\frac{d}{dx}\right)^2 V_n(x) \sim \lim_{x\to\pm\infty}\left(e^{\pm k_1 x}\frac{d}{dx}\right)^2 e^{\mp 2k_n x} \sim \lim_{x\to\pm\infty} e^{\pm k_1 x}\frac{d}{dx} e^{\pm(k_1-2k_n)x} \sim \lim_{x\to\pm\infty} e^{\pm 2(k_1-k_n)x} = \infty, \tag{7.21}$$

unless

$$2k_n = k_1, \tag{7.22}$$

because then the last step in (7.21) is incorrect; (7.22) is then a necessary condition for $f(\phi)$ to be three times differentiable in the interval $[0, T_n]$.

Proceeding as in (7.19), we obtain, in a second approximation,

$$V_n(x) \to a_\pm e^{\pm 2k_n x} + b_\pm e^{\mp 2k_{n-1} x}, \quad x \to \pm\infty.$$

Then, even if (7.22) holds, we again obtain $f'''(T_n, 0) = \infty$, because $k_{n-1} \neq k_n$ (Appendix), unless $k_{n-1} = k_1$ or

$$n = 2. \tag{7.23}$$

To sum up, for $f(\phi)$ to be three times differentiable in the interval $[0, T_n]$, we must have $n = 1$, which gives the sine-Gordon equation, or $n = 2$ and $2k_2 = k_1$.

Case $n = 2, 2k_2 = k_1$: Putting $e^{k_1 x} = y$, $9k_1/A_2 = p$, and $18k_1/A_1 = q$, we obtain, after some computation,

$$\psi_1 \sim \frac{1}{\Delta} y(y + \tfrac{1}{3}p), \tag{7.24}$$

with

$$\Delta = y^3 + py^2 + qy + \tfrac{1}{9}pq.$$

We could now differentiate ψ_1 (to get ϕ_0'') and integrate it (to get ϕ_0), and substituting in (7.2) obtain an equation for $f(\phi)$, but here we are out of luck, because the equation is in general much more involved than in the case $n = 1$ (where we knew the solution to start with!) and we are not able to solve it.

However, we can compute $f''(\phi_0)$:

$$f''(\phi_0) \sim \frac{1}{\psi_1}\frac{d}{dx} V_2(x) \sim \frac{1}{\psi_1}\frac{d^3}{dx^3}\log\Delta.$$

Δ has three different roots if $p^2 \neq 3q$; writing

$$\Delta = \prod_1^3 (y - y_i),$$

we get

$$f''(\phi_0) \sim \frac{\Delta}{(y + \frac{1}{3}p)} \sum_1^3 \frac{y_i(y + y_i)}{(y - y_i)^3}$$

$$\xrightarrow[y\to\infty]{} y^2 \sum \frac{y_i}{y^2} = -p,$$

$$\xrightarrow[y\to 0]{} \frac{\frac{1}{9}pq}{\frac{1}{3}p} \sum \frac{y_i^2}{-y_i^3} = \frac{3q}{p}, \quad (7.25)$$

and since p and q are positive we have

$$f''(0) \neq f''(T_2). \quad (7.26)$$

If we now assume the 1-sphere topology for the space spanned by ϕ and want ϕ_0 to be bounded (i.e., going to "0" at $x = \pm\infty$), by (7.4c) and (7.12) we have

$$\phi_0(-\infty) = \phi_0(\infty) \pmod{T_2},$$

which means that $f(\phi)$, and *a fortiori* its derivatives, ought to be periodic with period T_2; this is inconsistent with (7.26) if $f'''(\phi)$ is continuous in $[0, T_2]$.

We have proven, then, that the sine-Gordon equation is the only one compatible with assumptions (7.4), if the ϕ manifold is a 1-sphere and $\phi_0 \to 0 \pmod{T}$ for $x \to \pm\infty$.

For certain values of the parameters, $f(\phi)$ can be easily found: Let $p^2 = 3q$; then $\Delta = (y + \frac{1}{3}p)^3$, and putting $\frac{1}{3}p = c$, we get $y_i = -c$ and

$$\phi_0 \sim \int_{-\infty}^{x} \psi_1\, dx \sim \int_c^{y+c} \frac{dz}{z^2} \sim \frac{y}{y + c},$$

or

$$\phi_0 = T\frac{y}{y + c} = T\frac{e^{mx}}{e^{mx} + c},$$

and from (7.25),

$$f''(\phi_0) \sim (y + c)^2 3\frac{-c(y - c)}{(y + c)^3} \sim \frac{y - c}{y + c} = \frac{2}{T}\phi_0 - 1,$$

$$f'(\phi_0) \sim \frac{1}{T}\phi_0^2 - \phi_0 + \alpha,$$

$$f(\phi_0) \sim \frac{1}{3T}\phi_0^3 - \tfrac{1}{2}\phi_0^2 + \alpha\phi_0,$$

or, upon imposing the boundary conditions $f(T) = 0$, $f'(0) = m^2$,[11]

$$f(\phi) = m^2\left(\frac{2}{T^2}\phi^3 - \frac{3}{T}\phi^2 + \phi\right), \quad (7.27)$$

with which (7.1) becomes what is known in the literature as the "Thirring model." This is an interesting aspect of our problem, but to explore it further would take us too far afield.

8. RELATION TO THE KORTEWEG–DE VRIES EQUATION

The Korteweg–de Vries equation,

$$u_t + uu_x + u_{xxx} = 0, \quad (8.1)$$

of ubiquitous presence in physics,[12] has properties strikingly similar to the ones we have found for the sine-Gordon equation: the existence of bound traveling states, their emerging unaltered from (nonlinear) interaction with each other, the decay of an initial pulselike disturbance into a well-defined number of bound states, depending on the width and amplitude of the initial disturbance,[13] etc. Some of these results can be obtained using certain conserved densities that (8.1) exhibits in addition to mass, momentum, and energy,[13,14] and recent results[15] indicate that the sine-Gordon equation also has such "unphysical" conserved densities.

This makes us suspect that (8.1) and sine-Gordon may be closely related equations, but serious problems arise if we try to prove it: for instance, (8.1) is Galileo-invariant but sine-Gordon is Lorentz-invariant; (8.1) is of first order in time and sine-Gordon of second, etc.

We can prove, however, that the respective *reduced* equations, namely, the ones we get by putting $\phi(x, t) = \phi(x - vt)$ and $u(x, t) = u(x - wt)$, can be transformed into each other; more generally, if the equation for ϕ is

$$\phi_{tt} - \phi_{xx} + f(\phi) = 0, \quad (7.1)$$

the reduced equations, after adequate scaling of variables and functions, are

$$\phi'' = f(\phi) \quad (8.2)$$

and

$$\pm u' + uu' + u''' = 0; \quad (8.3)$$

[11] Notice that $f(T) = 0$ is not an additional assumption:
$f(T) = \int_0^T f'(\psi)\, d\psi = \int_\infty^\infty \phi_0''(x)\, dx \sim \int_{-\infty}^{\infty} \psi_1''(x)\, dx = 0.$

[12] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Letters **19**, 1095 (1967), and references therein.

[13] Yu. A. Berezin and V. I. Karpman, Sov. Phys.—JETP **24**, 1049 (1967).

[14] R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).

[15] M. Kruskal, private communication.

we now set

$$u = af'(\phi) + b = a\frac{df}{d\phi} + b \tag{8.4}$$

and easily find that a necessary and sufficient condition for u to satisfy (8.3) (except for scalings), if ϕ satisfies (8.2), is that

$$2f^{IV}f^P + 3f'''f = \alpha f''f' + \beta f'', \tag{8.5}$$

where $f^P = \int f(\phi)\, d\phi$, and α and β are real numbers, with $\alpha + 1 \neq 0$.

It is easy to see that $f(\phi) = \sin\phi$ is one solution of (8.5) and also that the other explicit solution to the transparency problem obtained in Sec. 8, namely

$$f(\phi) = 2\phi^3 - 3\phi^2 + \phi \tag{7.27'}$$

(m and T have been rescaled away), satisfies (8.5) too.

Although we do not have an explicit expression for the general ($p^2 \neq 3q$) solution $f(\phi)$ of the transparency problem, all the ingredients in (8.5) can be written in terms of x, $\wp_1(x)$ [see (7.24)] and its derivatives, so (8.5) could still be checked as an expression in x. The calculation is trivial but lengthy and has not been carried out; we suspect, however, that the answer is affirmative.

Notice that (8.2) and (8.4) give also two lower-order equations in u, which can be considered as constraints on (8.3) and limit our choice of constants of integration. For instance, $f = e^{\phi}$ satisfies (8.5) but does not allow solitary waves in u, for which $u(\infty) = u'(\infty) = u''(\infty) = 0$, because $\phi'' = e^{\phi}$ has only periodic solutions. However, the two f's that we examined above do give solitary waves in u, if we choose b/a adequately in (8.4).

We have, then, made less surprising the fact that the Korteweg–de Vries, sine-Gordon, and Thirring equations do all have bound traveling solutions, but the other similarities are still not fully understood. Work in that direction is currently being done.[15]

ACKNOWLEDGMENT

The author is indebted to Professor D. Finkelstein for motivation of this work and constructive discussion during its performance.

APPENDIX

Kay and Moses's expression [see Ref. 10, (2.11) and (3.4)] for V_n can be written in a slightly more explicit form, namely,

$$\begin{aligned} V_n(x) &= -2\frac{d^2}{dx^2}\log\left(1 + \sum_1^n a_\nu\right) \\ &= -2\frac{d^2}{dx^2}\log\left(1 + \sum_{\nu=1}^{n}\sum_{\binom{n}{\nu}}\left[\left(\prod^{(\nu)} A_i e^{2k_i x}\right)\times\frac{\prod^{(\nu)}_{i>j}(k_i - k_j)^2}{\prod^{(\nu)}_{i,j}(k_i + k_j)}\right]\right), \end{aligned} \tag{A1}$$

where $\binom{n}{\nu}$ indicates that the sum is over all the possible combinations of ν elements taken from the given n, and (ν) indicates that the corresponding product is between the elements of the given combination.

If we now build up a V_{n+1} adding to the A's and k's (of V_n) A_{n+1} and $k_{n+1} = k_n$, we get

$$V_{n+1}(x) = -2\frac{d^2}{dx^2}\log\left(1 + \sum_1^{n+1} a'\right), \tag{A2}$$

and it is easy to see from (A1) that, if we rename $(A_n + A_{n+1}) \to A_n$, we obtain

$$a'_\nu \to a_\nu, \quad \nu = 1, \cdots, n,$$
$$a_{n+1} = 0.$$

Then, since the A's can take any positive value, (A2) becomes (A1). Iteration of this reasoning gives the result stated in the text, namely, that we can take the k_i all different from each other without loss of generality.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Anisotropy of Scattering in Neutron Transport Theory by the Two-Sided Laplace Integral Method

V. C. BOFFI
Comitato Nazionale per l'Energia Nucleare, Centro di Calcolo di Bologna, Bologna, Italy

(Received 7 August 1967)

In this paper the effects of arbitrarily anisotropic scattering in establishing the eigenvalue spectrum of the operator associated with the stationary monoenergetic neutron transport equation in a medium of infinite extent are investigated through an extensive application of the two-sided Laplace integral theory. How the contribution of the discrete eigenvalues imbedded in the continuous part of the spectrum can explicitly be evaluated when inverting the bilateral Laplace transform of the sought distribution is shown by resorting to the Plemelj formulas, which are in order in the theory of the Cauchy integrals.

INTRODUCTION

Space-angle dependent problems in neutron transport theory have been the object of several investigations during the last few years. Functional analysis and eigenfunction expansion methods have mostly been used and extended to various physical cases (see, for instance, the review article by Zelazny[1]). In comparison, integral transform methods have received much less attention, even if they have successfully been applied in the past to solve neutron transport problems with isotropic scattering[2] and have recently been resumed to study the effects of anisotropic scattering once the transformed flux is expanded into a series of Legendre polynomials.[3] Integral transform methods are undoubtedly an interesting alternative to the two methods mentioned above when the Green's function for an infinite medium is dealt with. Actually, solutions of finite geometry problems have also been satisfactorily worked out by means of special integral transform techniques.[4]

With the aim of providing a deeper insight into the mathematical structure of the spatially-dependent neutron transport theory, a new method is here proposed in order to construct the Green's function for an infinite medium in the case of the stationary monoenergetic space-angle dependent form of the Boltzmann equation with arbitrarily anisotropic scattering. Such a method largely refers to the two-sided Laplace integral theory and does not resort to any supplementary expansion of the transformed flux distribution. The point is that, proceeding along the lines of the two-sided Laplace integral theory, we can first reduce the original transport equation to a Fredholm linear integral equation of the second kind for the bilateral transform of the distribution. Because the kernel of this integral equation is degenerate, the solution is readily obtained by standard technique and then transformed back through an extensive application of the integration rule to yield the sought space-angle flux distribution. The spectrum of the singularities of the bilateral transform which is to be inverted is seen to consist of poles and branch points. It will be verified that, because of the role played by the streaming term of the neutron balance equation, one of the poles always lies in those regions which, in order to insure the analyticity of the considered transform, are cut off the complex plane when the equivalent contour of integration is drawn to circumnavigate the branch points. Such a pole will henceforth be referred to as principal pseudopole.[5] Furthermore, as a consequence of special values that can be assigned to the Legendre moments of the scattering kernel and to the scattering probability of the medium, other poles, which will be referred to as additional or secondary pseudopoles, may analogously turn out to be imbedded in the deformed integration path connected with the branch points. In terms of functional analysis, separation of poles, pseudopoles, and branch points implies that the spectrum of the eigenvalues corresponding to the unsymmetric operator associated with the present form of the source-free Boltzmann equation separates into a discrete and a continuous part and some of the discrete eigenvalues are imbedded in the continuous part of the spectrum.

The existence of principal and secondary pseudopoles has first been recognized by Mika[6] through an extension of the eigenfunction expansion method and successively by Lathrop,[3] who has given conditions for the presence or absence of the secondary pseudopoles when inverting the Fourier transform of the

[1] R. Zelazny, Nukleonika 11, 79 (1966).
[2] A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chain Reactors* (University of Chicago Press, Chicago, 1958).
[3] K. D. Lathrop, Los Alamos Scientific Lab. Report LA-3051, 13 May 1964.
[4] J. M. Norwood and R. N. Little, J. Nucl. Energy 17, 245 (1963).
[5] V. C. Boffi, C.N.E.N. report, Doc. CEC (67)11, May 1967.
[6] J. R. Mika, Nucl. Sci. Engr. 11, 415 (1961).

fundamental harmonic of the flux.[7] It is worth noting that the explicit evaluation of the contribution due to the general pseudopole does strictly depend on the adopted method of solution. Thus, for instance, by Mika's procedure it is possible to evaluate explicitly the contribution of the principal pseudopole, whereas only an implicit expression has been given for the contribution of the secondary pseudopoles. In Lathrop's work the situation is practically reversed. The disappearance of the principal pseudopole contribution in the latter case is, of course, merely fictitious, as it is in some way contained in the final result for the solution. The situation is quite similar to the one occurring in the isotropic scattering case when the principal pseudopole is treated in the framework of the so-called B_0 approximation.[8] The gap between the Mika and Lathrop results is here filled up as we succeed in showing by the present integral contour approach that the contribution of any pseudopole can explicitly be resolved by means of the Plemelj formulas, which are of basic significance in the theory of Cauchy integrals.

1. THE SOLUTION OF THE TRANSFORMED EQUATION

For the physical situation already sketched in the Introduction, the form of the Boltzmann equation to be considered is[6]

$$\left(1 + \mu \frac{d}{dx}\right) f(x, \mu) = \frac{1}{4\pi}\,\delta(x) + \tfrac{1}{2}c \sum_{n=0}^{N} b_n P_n(\mu) \int_{-1}^{1} P_n(\mu') f(x, \mu')\, d\mu', \quad (1)$$

where $f(x, \mu)$ is the angular flux; x is the optical distance measured from the source plane at $x = 0$; μ is the cosine of the angle between the neutron velocity v and the positive direction of the x axis; c is the scattering probability of the medium ($0 < c < 1$); $S(x, \mu) \equiv \delta(x)/4\pi$ is the monochromatic, isotropic, infinite plane δ-like source; $P_n(\mu)$ is the Legendre polynomial of nth order[9]; and $b_n = (2n + 1)g_n$, g_n being the nth Legendre moment of the scattering kernel ($g_0 = 1$).

The problem of solving Eq. (1) for $f(x, \mu)$ can be stated as follows.

By the symmetry of the problem we may confine ourselves to positive values of x and seek for a solution of Eq. (1) of the form

$$f(x, \mu) = \frac{1}{4\pi} \cdot \left\{ \sum_{m=0}^{M} a_m F(k_m, \mu) e^{-k_m x} + \int_1^{\infty} A(\eta) G(\eta, \mu) e^{-x\eta}\, d\eta \right\}. \quad (2)$$

In terms of functional analysis, the solution [Eq. (2)] is recognized to be expressed as a linear combination of regular and singular eigenfunctions, of coefficients a_m and $A(\eta)$, respectively. In particular, k_m is the general eigenvalue belonging to the discrete part of the spectrum and $F(k_m, \mu)$ is the corresponding regular eigenfunction. The integral term in the bracket represents, on the contrary, the contribution of the continuous part of the spectrum, and $G(\eta, \mu)$ is the corresponding singular eigenfunction which may behave like a distribution. In the eigenfunction expansion method, the coefficients a_m and $A(\eta)$ are determined, by using the orthogonality theorems for the eigenfunctions, from the uniqueness of the solution of a singular integral equation, which is derived by separately imposing the boundary condition at the source plane. Some difficulty in calculating explicitly the coefficients a_m and $A(\eta)$ arises when discrete eigenvalues are present in the continuous part of the spectrum, as the required conditions for doing it are only implicitly known.[6]

A solution of the form as in Eq. (2) can, however, be obtained by a straightforward application of the two-sided Laplace integral theory, as will be discussed here in some details. Another goal of the present discussion will be a more explicit knowledge of the integral term of Eq. (2) in the case when secondary pseudopoles are present.

We begin thus with taking a bilateral Laplace transform of both sides of Eq. (1) with respect to the space variable; assume the convergence of the integral[10]

$$\mathfrak{L}_{II}\{f(x, \mu)\} \equiv f(p, \mu) = \int_{-\infty}^{\infty} e^{-px} f(x, \mu)\, dx, \quad (3)$$

where $p = \alpha + i\omega$ is a complex parameter.

As the medium is of infinite extent, we get from Eq. (1)

$$(1 + p\mu) \cdot f(p, \mu) = \frac{1}{4\pi} + \tfrac{1}{2}c \sum_{n=0}^{N} b_n P_n(\mu) \int_{-1}^{1} P_n(\mu') f(p, \mu')\, d\mu', \quad (4)$$

which is to be solved for the transform $f(p, \mu)$.

[7] Actually the concept of pseudopole is not used in Refs. 3 and 6. This concept is well suited to and better understood in the context of the present treatment.

[8] V. C. Boffi, Nukleonik **11**, 95 (1968).

[9] The $P_n(\mu)$'s could actually represent any set of polynomials orthogonal with respect to the interval $(-1, 1)$. For the sake of simplicity, however, they are referred to as Legendre polynomials.

[10] B. Van Der Pol and H. Bremmer, *Operational Calculus Based on the Two-sided Laplace Integral* (University Press, Cambridge, England, 1964).

It is clear that the way by which $f(p, \mu)$ is extracted from Eq. (4) must have a strong influence in establishing the spectrum of the singularities associated with the analytic behavior of $f(p, \mu)$ in the complex plane, $p = \alpha + i\omega$. In fact, if we resorted to the usual method of solving Eq. (4) by means of an expansion of $f(p, \mu)$ into a series of Legendre polynomials, we would cause the factor $1 + p\mu$ on the left-hand side of Eq. (4) to disappear from the successive calculations.[3] Now, $1 + p\mu$ is one of the two factors which constitute the transform of the rimotion term appearing in the operator associated with the homogeneous form of Eq. (1). It follows that the singularity, if any, corresponding to such a factor will be missing in the spectrum of the singularities of $f(p, \mu)$, even if its contribution is in some manner retained in the final result for $f(x, \mu)$.[11]

There actually exists a direct way of solving Eq. (4) for $f(p, \mu)$. Dividing both sides of Eq. (4) by $1 + p\mu$ yields

$$f(p, \mu) = \frac{1}{4\pi} \cdot \frac{1}{1 + p\mu} + \tfrac{1}{2}c \int_{-1}^{1} \sum_{n=0}^{N} P_n^*(\mu) P_n(\mu') f(p, \mu')\, d\mu', \quad (5)$$

where we have set

$$P_n^*(\mu) = b_n \frac{P_n(\mu)}{1 + p\mu}. \quad (6)$$

Equation (5) is nothing else but a Fredholm linear integral equation of the second kind with a degenerate kernel, in which the functions $P_n^*(\mu)$, as the functions $P_n(\mu')$, are linearly independent. As is well known from the theory of the degenerate equations, their solutions reduce to algebraic equations of the first degree.[12]

Following Ref. 12 we put

$$f_n(p) = 2\pi \int_{-1}^{1} P_n(\mu) f(p, \mu)\, d\mu, \quad (7)$$

which is the usual Legendre moment of nth order of $f(p, \mu)$, so that Eq. (5) becomes

$$f(p, \mu) = \frac{1}{4\pi} \cdot \left\{ \frac{1}{1 + p\mu} + c \sum_{n=0}^{N} P_n^*(\mu) f_n(p) \right\}. \quad (8)$$

Once the $f_n(p)$'s are known, Eq. (8) represents the correct and complete form of the solution of Eq. (4). Inserting Eq. (8) into Eq. (5) and equating coefficients for the linearly independent functions $P_n^*(\mu)$, we find that the moments $f_n(p)$ must satisfy, for any fixed p, the system of $N + 1$ algebraic equations:

$$f_m(p) - c \sum_{n=0}^{N} b_n I_{mn}(p) f_n(p) = I_{m0}(p), \qquad m = 0, 1, \cdots, N, \quad (9)$$

where

$$I_{m0}(p) \equiv I_{0m}(p) = \frac{1}{2} \int_{-1}^{1} \frac{P_m(\mu)}{1 + p\mu}\, d\mu \quad (10a)$$

and

$$I_{mn}(p) \equiv I_{nm}(p) = \frac{1}{2} \int_{-1}^{1} \frac{P_m(\mu) P_n(\mu)}{1 + p\mu}\, d\mu \quad (10b)$$

are certain functions to be specified according to the values the parameter p takes in the complex plane.

For later reference, let us restrict our attention to the strip

$$-1 < \mathrm{rep} < 1. \quad (11)$$

We know then by successive iteration that[8]

$$I_{m0}(p) \equiv (-1)^m \frac{1}{p} \mathcal{Q}_m\left(\frac{1}{p}\right) = (-1)^m \cdot \left[I_{00}(p) \cdot P_m\left(\frac{1}{p}\right) - W_{m-1}\left(\frac{1}{p}\right) \frac{1}{p} \right], \quad (12)$$

where

$$I_{00}(p) = \frac{\coth^{-1}(1/p)}{p} \quad (13a)$$

and

$$\mathcal{Q}_m\left(\frac{1}{p}\right) = P_m\left(\frac{1}{p}\right) \mathcal{Q}_0\left(\frac{1}{p}\right) - W_{m-1}\left(\frac{1}{p}\right) \quad (13b)$$

is the nth Legendre function of the second kind when Eq. (11) holds.[13] In Eq. (13b), $P_m(z)$ is the nth Legendre polynomial of complex argument and

$$W_{m-1}(z) = \sum_{l=1}^{m} \frac{1}{l} P_{l-1}(z) P_{m-l}(z), \quad (13c)$$

as listed in Ref. 13.

As far as the $I_{mn}(p)$'s [Eq. (10b)] are concerned, two recursion formulas can be established, as is usually done in the framework of the B_N approximation.[14] We in fact distinguish the relation relative to the second index, namely,

$$\frac{2n + 1}{p} I_{mn}(p) + (n + 1) I_{m,n+1}(p) + n I_{m,n-1}(p) = \frac{\delta_{mn}}{p}, \quad (14a)$$

[11] We shall see later on that the singular point corresponding to $1 + p\mu$ is just the principal pseudopole of $f(p, \mu)$.

[12] V. I. Smirnov, *A Course of Higher Mathematics* (Pergamon Press, Ltd., Oxford, England, 1964), Vol. 4.

[13] E. Jahnke, F. Emde, and F. Lösch, *Tables of Higher Functions* (B. G. Taubner, Stuttgart, Germany, 1960).

[14] J. H. Ferziger and P. F. Zweifel, *The Theory of Neutron Slowing Down in Nuclear Reactors* (Pergamon Press, Ltd., Oxford, England, 1966).

from the one relative to the first index,

$$\frac{2m+1}{p} I_{mn}(p) + (m+1) I_{m+1,n}(p) + m I_{m-1,n}(p) = \frac{\delta_{mn}}{p}, \quad (14b)$$

which follows inasmuch as the $I_{mn}(p)$'s are symmetric in m and n.

For the purposes of the present discussion, we select from Eqs. (14) the expressions

$$I_{mn}(p) = (-1)^n I_{0m}(p) P_n(1/p), \quad \text{for } n \leq m-2,$$

$$I_{m,m-1}(p) = (-1)^m \left[I_{0,m-1}(p) P_m(1/p) - \frac{(-1)^{m-1}}{mp} \right], \quad \text{for } n = m-1,$$

$$I_{mn}(p) = (-1)^m I_{0n}(p) P_m(1/p), \quad \text{for } n \geq m. \quad (15)$$

At this point, in order to render explicit the knowledge of the bilateral Laplace transform $f(p, \mu)$ [Eq. (8)], there remains only to determine the $f_n(p)$'s. How the parameter p enters the inversion of the system, Eq. (9), must also be considered.

2. THE SOLUTION $f_n(p)$ OF THE SYSTEM, EQ. (9)

By Cramer's rule, the general solution $f_n(p)$ of the system (9) is, with obvious notation,

$$f_n(p) = \frac{\Delta^{(n)}(p, c, g_0, g_1, \cdots, g_{n-1}, g_{n+1}, \cdots, g_N)}{\Delta(p, c, g_0, g_1, g_2, \cdots, g_N)}, \quad n = 0, 1, \cdots, N, \quad (16)$$

where $\Delta^{(n)}$ and Δ are determinants of $(N+1)$th order. If the determinant of the coefficients Δ were different from zero, Eq. (9) would give with any $I_{m0}(p)$ definite values for the $f_n(p)$'s. As a function of the complex parameter p, the determinant $\Delta(p, c, g_0, g_1, g_2, \cdots, g_N)$ could indeed vanish for some particular values of p. But these values of p will act at most as poles or pseudopoles when transforming back both sides of Eq. (8). This important correlation between the degeneracy of the kernel of Eq. (5) and the spectrum of the singularities of the transform $f(p, \mu)$ [Eq. (8)] will properly be taken into account later on so that we may for now solve formally the system, Eq. (9), by supposing $\Delta \neq 0$.

A convenient form for the general $f_n(p)$ [Eq. (16)] is

$$f_n(p) = R_n(1/p) + S_n(1/p) f_0(p), \quad (17)$$

as can be verified *a posteriori* (see Appendix). Moreover, since

$$\Delta(p, c, g_0, g_1, g_2, \cdots, g_N) = 1 - c \sum_{n=0}^{N} b_n I_{0n}(p) S_n(1/p), \quad (18)$$

one recognizes also that

$$f_0(p) \equiv \frac{\Delta^{(0)}}{\Delta} = \frac{I_{00}(p) + c \sum_{n=1}^{N} b_n I_{0n}(p) R_n(1/p)}{1 - c \sum_{n=0}^{N} b_n I_{0n}(p) S_n(1/p)}. \quad (19)$$

In Eqs. (17) and (19), the functions

$$R_n\left(\frac{1}{p}\right) \equiv R_n\left(\frac{1}{p}, c, g_1, g_2, \cdots, g_{n-1}\right) \quad (20a)$$

and

$$S_n\left(\frac{1}{p}\right) \equiv S_n\left(\frac{1}{p}, c, g_0, g_1, g_2, \cdots, g_{n-1}\right) \quad (20b)$$

are polynomials of nth degree in p^{-1} and can be constructed by recursion from the formula

$$(n+1)pH_{n+1}(1/p) + (2n+1)(1 - cg_n)H_n(1/p) + npH_{n-1}(1/p) = 0, \quad (21)$$

where $H_n(1/p)$ stands for either $R_n(1/p)$, $n = 1, 2, \cdots, N$, or $S_n(1/p)$, $n = 0, 1, 2, \cdots, N$. Equation (21) is the same as obtained by Lathrop[3] in defining the fundamental harmonic $f_0(p)$ [Eq. (19)] through an expansion of the Fourier transform of his distribution into a series of Legendre polynomials. In the present case, for the form in Eq. (17) to be really useful in the sequel, we can expand as far as we need for the R_n and S_n to be given a compact expression which is independent of Eq. (21). Since

$$R_0(1/p) = 0, \quad S_0(1/p) = 1, \quad (22)$$

we find that

$$R_n(1/p) = (-1)^{[n/2]} \sum_{m=1}^{[(n+1)/2]} (-1)^{m+1} \frac{T^{(1)}_{n,m}(c, g_m)}{p^{2m-\delta_{n0}}}, \quad n = 2, 3, \cdots, N, \quad (23a)$$

$$S_n(1/p) = (-1)^{[(n+1)/2]} \sum_{m=0}^{[n/2]} (-1)^m \frac{T^{(0)}_{n,m}(c, g_m)}{p^{2m+\delta_{n0}}}, \quad n = 1, 2, \cdots, N, \quad (23b)$$

where the symbol $[M/2]$ indicates the maximum integer contained in $M/2$ and δ_{n0} is the Kronecker index for the oddness of n; that is,

$$\delta_{n0} = 0, \text{ for } n \text{ even}, \qquad \delta_{n0} = 1, \text{ for } n \text{ odd}. \quad (24)$$

As far as the dimensionless coefficients $T_{n,m}$ are concerned, we have

$$T^{(1)}_{n,m}(c, g_m) = \sum_{I_{[(n+1)/2]-m}=1}^{2m-\delta_{n0}} \sum_{I_{[(n+1)/2]-m-1}=I_{[(n+1)/2]-m}+2}^{2m+2-\delta_{n0}} \cdots \sum_{I_1=I_2+2}^{2[(n-2)/2]-\delta_{n0}} \frac{(I_{[(n+1)/2]-m}+1)\cdots(I_1+1)}{(I_{[(n+1)/2]-m}+2)\cdots(I_1+2)} \times \prod_{I_0=1}^{n-1} \frac{2I_0+1}{I_0+1}(1 - cg_{I_0}), \quad (25a)$$

provided the index I_0 running from 1 to $n-1$ in the product is taken to satisfy the inequalities

$$\begin{aligned} &I_0 \neq I_{[(n+1)/2]-m},\ I_{[(n+1)/2]-m}+1,\\ &\cdots\\ &I_0 \neq I_2,\ I_2+1,\\ &I_0 \neq I_1,\ I_1+1. \end{aligned} \tag{25b}$$

When $m = [(n+1)/2]$, Eq. (25a) reduces to

$$T^{(1)}_{n,[(n+1)/2]}(c, g_{[(n+1)/2]}) = \prod_{I_0=1}^{n-1} \frac{2I_0+1}{I_0+1}\cdot(1-cg_{I_0}). \tag{25c}$$

For $T^{(0)}_{n,m}(c, g_m)$ we have the slightly different expression

$$T^{(0)}_{n,m}(c, g_m) = \sum_{I_{[n/2]-m}=0}^{2m+\delta_{n0}} \sum_{I_{[n/2]-m-1}=I_{[n/2]-m}+2}^{2m+2+\delta_{n0}} \cdots \sum_{I_1=I_2+2}^{2[(n-2)/2]+\delta_{n0}} \frac{(I_{[n/2]-m}+1)\cdots(I_1+1)}{(I_{[n/2]-m}+2)\cdots(I_1+2)} \cdot \prod_{I_0=0}^{n-1}\frac{2I_0+1}{I_0+1}(1-cg_{I_0}), \tag{26a}$$

where I_0 still obeys Eqs. (25b) with $[n/2]$ in place of $[\frac{n+1}{2}]$. When $m = [n/2]$, we get

$$T^{(0)}_{n,[n/2]}(c, g_{[n/2]}) = \prod_{I_0=0}^{n-1}\frac{2I_0+1}{I_0+1}(1-cg_{I_0}). \tag{26b}$$

The first few of R_n and S_n are listed below:

$$\begin{aligned} R_1\left(\frac{1}{p}\right) &= \frac{1}{p};\\ R_2\left(\frac{1}{p}\right) &= -\frac{3}{2}\frac{1-cg_1}{p^2};\\ R_3\left(\frac{1}{p}\right) &= \frac{5}{2}\frac{(1-cg_1)(1-cg_2)}{p^3} - \frac{2}{3p}; \end{aligned} \tag{27a}$$

and

$$\begin{aligned} S_1\left(\frac{1}{p}\right) &= -\frac{1-cg_0}{p};\\ S_2\left(\frac{1}{p}\right) &= \frac{3}{2}\frac{(1-cg_0)(1-cg_1)}{p^2} - \frac{1}{2};\\ S_3\left(\frac{1}{p}\right) &= -\left\{\frac{5}{2}\frac{(1-cg_0)(1-cg_1)(1-cg_2)}{p^2} - [\tfrac{5}{6}(1-cg_2) + \tfrac{2}{3}(1-cg_0)]\frac{1}{p}\right\}. \end{aligned} \tag{27b}$$

In the isotropic case, that is, when $b_n = g_n = 0$ for $n \geq 1$, Eqs. (23) sum up to known functions. There results

$$\begin{aligned} R^{(i)}_n\left(\frac{1}{p}\right) &= (-1)^{n+1}\frac{1}{p}W_{n-1}\left(\frac{1}{p}\right),\\ S^{(i)}_n\left(\frac{1}{p}\right) &= (-1)^n\left[P_n\left(\frac{1}{p}\right) - \frac{c}{p}W_{n-1}\left(\frac{1}{p}\right)\right], \end{aligned} \tag{28}$$

where P_n and W_{n-1} are the same as in Eq. (13b).

3. THE BILATERAL LAPLACE TRANSFORM $f(p, \mu)$, EQ. (8)

Inserting Eq. (17) for $f_n(p)$ into the right-hand side of Eq. (8) and then using Eq. (19) for $f_0(p)$ and Eq. (12) for $I_{m0}(p)$ yield for the bilateral Laplace transform of the sought distribution

$$f(p,\mu) = \frac{1}{4\pi}\left\{\frac{\phi_0(p)}{1+\mu p}\cdot\sum_{n=0}^{N} b_n P_n(\mu) S_n\left(\frac{1}{p}\right) - \frac{1}{1+\mu p}\sum_{n=1}^{N} b_n P_n(\mu) S^*_n\left(\frac{1}{p}\right)\right\}, \tag{29}$$

where

$$\phi_0(p) \equiv 1 + cf_0(p) = \frac{1 - \frac{c}{p}\sum_{n=1}^{N}(-1)^n b_n S^*_n\left(\frac{1}{p}\right)\mathfrak{Q}_n\left(\frac{1}{p}\right)}{1 - \frac{c}{p}\sum_{n=0}^{N}(-1)^n b_n S_n\left(\frac{1}{p}\right)\mathfrak{Q}_n\left(\frac{1}{p}\right)} \tag{30}$$

and

$$\begin{aligned} S^*_n(1/p) &\equiv S_n(1/p) - cR_n(1/p)\\ &= (-1)^{[(n+1)/2]}\sum_{m=0}^{[n/2]}(-1)^m \cdot \frac{T^{*(0)}_{n,m}(c, g_m)}{p^{2m+\delta_{n0}}} \end{aligned} \tag{31}$$

is the same polynomial as the $S_n(1/p)$ [Eq. (23b)] provided the factor $(1-cg_0)$ appearing in each coefficient of the latter is replaced by unity. Therefore the coefficients $T^{*(0)}_{n,m}(c, g_m)$ of Eq. (31) are still given by Eq. (26a), in which the index I_0 of the product now runs from 1 to $n-1$.

By taking into account Eq. (23b) for $S_n(1/p)$ and Eq. (31) for $S^*_n(1/p)$, the two sums in the brackets of Eq. (29) become

$$\sum_{n=0}^{N} b_n P_n(\mu) S_n(1/p) = \sum_{n=0}^{N}(-1)^n \frac{U_n(\mu, c)}{p^n}, \tag{32a}$$

$$\sum_{n=1}^{N} b_n P_n(\mu) S^*_n(1/p) = -1 + \sum_{n=0}^{N}(-1)^n \frac{U^*_n(\mu, c)}{p^n}, \tag{32b}$$

where $U_n(\mu, c)$ and $U^*_n(\mu, c)$ are certain dimensionless

coefficients defined as

$$U_n(\mu, c) \equiv U_n(\mu, c, g_0, \cdots, g_N) = \sum_{m=0}^{[(N-n)/2]} (-1)^m b_{2m+n} \cdot T^{(0)}_{2m+n,[n/2]} \cdot P_{2m+n}(\mu), \tag{33a}$$

$$U^*_n(\mu, c) \equiv U^*_n(\mu, c, g_1, \cdots, g_N) = \sum_{m=0}^{[(N-n)/2]} (-1)^m b_{2m+n} \cdot T^{*(0)}_{2m+n,[n/2]} \cdot P_{2m+n}(\mu). \tag{33b}$$

Setting

$$F_0(p, \mu) = \frac{\phi_0(p)}{1 + \mu p}, \tag{34a}$$

$$G_0(p, \mu) = \frac{1}{p + \mu^{-1}} \tag{34b}$$

and using Eqs. (32), Eq. (29) is then rewritten as

$$f(p, \mu) \equiv f_1(p, \mu) + f_2(p, \mu) = \frac{1}{4\pi} \cdot \sum_{n=0}^{N} (-1)^n U_n(\mu, c) \cdot \frac{F_0(p, \mu)}{p^n} + \frac{1}{4\pi} \cdot \frac{1}{\mu}\left[G_0(p, \mu) - \sum_{n=0}^{N} (-1)^n U^*_n(\mu, c) \frac{G_0(p, \mu)}{p^n} \right], \tag{35}$$

which is a very convenient form for $f(p, \mu)$ to be transformed back. By the way, each of the two components $f_1(p, \mu)$ and $f_2(p, \mu)$ of Eq. (35) will be separately considered and inverted.

4. INVERSION OF $f_1(p, \mu)$, EQ. (35)

If $F_0(x, \mu)$ is the original of $F_0(p, \mu)$, by invoking the integration rule for the bilateral Laplace transforms, we know that[10]

$$\mathcal{L}^{-1}_{II}\left\{ \frac{F_0(p, \mu)}{p^n} \right\} = \frac{1}{(n-1)!} \int_{\pm\infty}^{x} F_0(u, \mu)(x - u)^{n-1}\, du, \tag{36}$$

where the positive sign holds for rep < 0, the negative sign for rep > 0. Then, we obtain

$$\mathcal{L}^{-1}_{II}\{f_1(p, \mu)\} \equiv f_1(x, \mu) = \frac{1}{4\pi}\left\{ U_0(\mu, c) F_0(x, \mu) + \sum_{n=1}^{N} (-1)^n \frac{U_n(\mu, c)}{(n-1)!} \times \int_{\pm\infty}^{x} F_0(u, \mu)(x - u)^{n-1}\, du \right\}, \tag{37}$$

as follows from the linearity of the considered transform. In order to determine $F_0(x, \mu)$, we need to identify not only the spectrum of the singularities of $F_0(p, \mu)$ in the complex plane $p = \alpha + i\omega$, but also its strip of convergence as required by the integration rule, Eq. (36). In fact, the constant of integration, involved by the two opposite signs in Eq. (36), must properly be specified. We have

$$F_0(p, \mu) = \frac{1}{1 + \mu p} \cdot \frac{1 - \frac{c}{p} \sum_{n=1}^{N} (-1)^n b_n S^*_n(1/p) \mathfrak{Q}_n(1/p)}{1 - \frac{c}{p} \sum_{n=0}^{N} (-1)^n b_n S_n(1/p) \mathfrak{Q}_n(1/p)}. \tag{38}$$

According to the definition of $S^*_n(1/p)$, $S_n(1/p)$, and $\mathfrak{Q}_n(1/p)$, it is easily seen that

$$F_0(p, \mu) = F_0(-p, -\mu), \qquad \lim_{|p| \to \infty} F_0(p, \mu) = 0. \tag{39}$$

The spectrum of the singularities of $F_0(p, \mu)$ consists then of the following points:

(i) The poles for those values of p at which the even function of p,

$$\Delta(p, c, g_0, g_1, \cdots, g_N) \equiv 1 - \frac{c}{p} \sum_{n=0}^{N} (-1)^n b_n S_n\left(\frac{1}{p}\right) \mathfrak{Q}_n\left(\frac{1}{p}\right) = 0. \tag{40}$$

This is the equivalent form which expresses the vanishing of the determinant of the coefficients of the system, Eq. (9), once Eq. (12) is inserted into Eq. (18). It has been shown[6] that the roots of Eq. (40), when $0 < c < 1$, are real and their number is $2(M + 1)$ with $M \leq N$, N being the order of the considered anisotropy. We assume that all these roots are simple, lie in the interval $(-1, 1)$, and are ordered as

$$0 < k_0 < k_1, \cdots, < k_M < 1. \tag{41}$$

(ii) The pole at $p = -1/\mu$.

(iii) The essential singularity at $p = \mp 1$, which is a branch point for the multivalued function $\mathfrak{Q}_n(1/p)$.

We can thus conclude that, since $F_0(p, \mu)$ is even with respect to p and μ and has no imaginary poles and since the strip of its convergence must contain the imaginary axis, its inverse $F_0(x, \mu)$ is even, so that we can restrict our calculations to the positive values of x.

Let us go further on: The imaginary axis is from now on taken to separate the integration contour which is in order, by the Jordan lemma, in the half-plane rep < 0 when positive values of x are considered from the integration contour in the half-plane rep > 0 when $x < 0$. (See Fig. 1.) In particular, referring to the

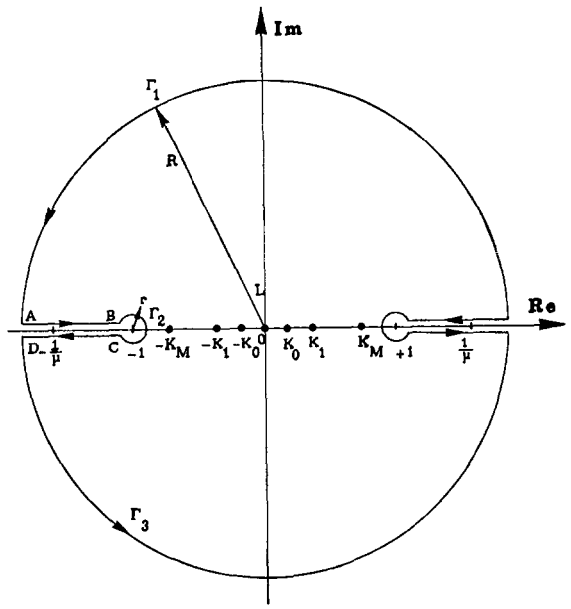


FIG. 1. The equivalent contour for the complex integral, Eq. (44).

values rep < 0, that is, $x > 0$, we realize that

$$-1 < \text{rep} < 0 \tag{42}$$

is, except, at most, for the poles at $p = -k_0$, $-k_1, \cdots, -k_M$, one of the two strips of convergence into which the domain of convergence of $F_0(p, \mu)$ is divided because of Eq. (36). Actually, the lines rep $= -k_m$, $m = 0, 1, \cdots, M$, separate the strip, Eq. (52), into the subdomains

$$\begin{aligned} -k_0 &< \text{rep} < 0, \\ -k_1 &< \text{rep} < -k_0, \\ &\cdots \\ -k_M &< \text{rep} < -k_{M-1}, \end{aligned} \tag{43a}$$

and

$$-1 < \text{rep} < -k_M. \tag{43b}$$

As far as the behavior of $F_0(p, \mu)$ on the boundaries of these subdomains is concerned, we note that $F_0(p, \mu)$ has a first-order pole on the finite part of the left boundary of the general strip $-k_m < \text{rep} < -k_{m-1}$, $m = 0, 1, \cdots, M$, and an essential singularity on the finite part of the left boundary of the last one, Eq. (43b). The behavior of $F_0(p, \mu)$ at infinity [see the second of Eqs. (39)] allows in the present case the crossing of the boundaries; we therefore conclude that the Mellin–Fourier integral

$$\mathcal{L}_{II}^{-1}\{F_0(p, \mu)\} \equiv F_0(x, \mu) = \frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} e^{xp} F_0(p, \mu)\, dp, \tag{44}$$

with $-k_0 < \alpha \leq 0$, can be split up into the sum

$$F_0(x, \mu) = \sum_{m=0}^{M} F_{0,m}^{(p)}(x, \mu) + F_0^{(e.s.)}(x, \mu), \tag{45}$$

where $F_{0,m}^{(p)}(x, \mu)$ and $F_0^{(e.s.)}(x, \mu)$ represent the contribution due to the pole at $p = -k_m$, $m = 0, 1, \cdots, M$, and to the essential singularity at $p = -1$, respectively.

Another consideration is now necessary and concerns the peculiar behavior of the pole at $p = -1/\mu$ when $\mu > 0$. In examining the branch cut in Fig. 1 which makes the function $e^{xp} F_0(p, \mu)$ continuous, we notice that the pole at $p = -1/\mu$ is left out from the integration path when circumnavigating the essential singularity at $p = -1$. The question then arises as to how the contribution of the pole at $p = -1/\mu$ to the complex integral, Eq. (44), can be drawn out. We shall soon see that this contribution does not appear in Eq. (45) as a term of the kind as $F_{0,m}^{(p)}(x, \mu)$, but as a term which is extracted from $F_0^{(e.s.)}(x, \mu)$ when the Cauchy-type integrals, occurring in evaluating $F_0^{(e.s.)}(x, \mu)$ and having a singularity just at $p = -1/\mu$, are resolved by means of the Plemelj formulas. As it is, the pole at $p = -1/\mu$ will be named principal pseudopole. Owing to the method adopted for solving Eq. (4), this principal pseudopole is always present in the singularity spectrum of $f(p, \mu)$, whereas it would disappear if, for instance, we had expanded $f(p, \mu)$ into a series of Legendre polynomials.

We can finally pass on to evaluate the two terms on the right-hand side of Eq. (45). They will be inserted into Eq. (37) to give the desired result for $f_1(x, \mu)$.

As far as the contribution due to the general pole at $p = -k_m$ is concerned, by still confining ourselves to $x > 0$, a straightforward application of the residue theorem yields

$$F_{0,m}^{(p)}(x, \mu) = \psi(k_m, c) \frac{e^{-k_m x}}{1 - \mu k_m}, \quad m = 0, 1, \cdots, \tag{46}$$

where we set

$$\begin{aligned} &\psi(k_m, c) \\ &\equiv \psi(k_m, c, g_0, g_1, \cdots, g_N) \\ &= \left[1 - \frac{c}{k_m} \sum_{n=1}^{N} (-1)^n b_n S_n^*\left(\frac{1}{k_m}\right) \Omega_n\left(\frac{1}{k_m}\right)\right] \\ &\quad \times \left\{\frac{d}{dp}\left[1 - \frac{c}{p} \sum_{n=0}^{N} (-1)^n b_n S_n\left(\frac{1}{p}\right) \Omega_n\left(\frac{1}{p}\right)\right]\right\}_{p=-k_m}^{-1}. \end{aligned} \tag{46'}$$

In dealing with the contribution of the essential singularity at $p = -1$, we first rewrite the inversion

integral, Eq. (44), with $\alpha = 0$, as

$$F_0^{(e.s.)}(x,\mu) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{e^{xp}}{1+\mu p}\cdot\left[X^*(1/p) - \frac{c}{p}\mathcal{Q}_0(1/p)Y^{**}(1/p)\Big/ X(1/p) - \frac{c}{p}\mathcal{Q}_0(1/p)Y(1/p)\right]dp, \quad (47)$$

where, by Eqs. (13b) and (38),

$$X\left(\frac{1}{p}\right) = 1 + \frac{c}{p}\sum_{n=1}^{N}(-1)^n b_n W_{n-1}\left(\frac{1}{p}\right)S_n\left(\frac{1}{p}\right),$$
$$Y\left(\frac{1}{p}\right) = 1 + \sum_{n=1}^{N}(-1)^n b_n P_n(1/p)S_n(1/p). \quad (48a)$$

$X^*(1/p)$ and $Y^*(1/p)$ are analogously defined in terms of $S_n^*(1/p)$ and

$$Y^{**}(1/p) = Y^*(1/p) - 1. \quad (48b)$$

Referring to the equivalent contour $C = L + (\Gamma_1 + \Gamma_3) + AB + \Gamma_2 + CD$ of Fig. 1, we remember that[13]

$$\mathcal{Q}_0\left(\frac{1}{p}\right) = Q_0\left(\frac{1}{p}\right) + i\tfrac{1}{2}\pi \quad (49a)$$

along AB (upper edge of the cut) and

$$\mathcal{Q}_0(1/p) = Q_0(1/p) - i\tfrac{1}{2}\pi \quad (49b)$$

along CD (lower edge of the cut) with

$$Q_0(1/p) = \tanh^{-1}(1/p). \quad (49c)$$

From the application of the Cauchy integral theorem to the contour C we deduce that

$$\frac{1}{2\pi i}\int_{-iR}^{iR} e^{xp}F_0(p,\mu)\,dp + \frac{1}{2\pi i}\int_{\Gamma_1+\Gamma_3} e^{xp}F_0(p,\mu)\,dp + \frac{1}{2\pi i}\int_{\Gamma_2} e^{xp}\cdot F_0(p,\mu)\,dp = -\frac{1}{2\pi i}\int_{-R+ir}^{-1+ir}\frac{\varphi^+(x,p)}{1+\mu p}\,dp - \frac{1}{2\pi i}\int_{-1-ir}^{-R-ir}\frac{\varphi^-(x,p}{1+\mu p}\,dp. \quad (50)$$

In Eq. (50), $F_0(p,\mu)$—which is given by Eq. (38)—is a function vanishing at infinity as $1/|p|$, when $|p|$ tends to ∞ on $\Gamma_1 + \Gamma_3$, and bounded in the neighborhoods of $p = -1$, whereas $\varphi^+(x,p)$ and $\varphi^-(x,p)$ are continuous and regular functions associated, respectively, with the upper and lower edge of the cut and given by

$$\varphi^{\pm}(x; c, g_0, g_1, \cdots, g_N; p) \equiv \varphi^{\pm}(x,p) = e^{xp}[v(p) \pm i(\pi c/2p)w(p)], \quad (51)$$

with

$$v(c, g_0, g_1, \cdots, g_N; p) \equiv v(p) = \frac{Z^{**}(1/p)Z(1/p) + (\pi c/2p)^2 Y^{**}(1/p)Y(1/p)}{Z^2(1/p) + (\pi c/2p)^2 Y^2(1/p)},$$

$$w(c, g_0, g_1, \cdots, g_N; p) \equiv w(p) = \left[Z^2\left(\frac{1}{p}\right) + \left(\frac{\pi c}{2p}\right)^2 Y^2\left(\frac{1}{p}\right)\right]^{-1}, \quad (52a)$$

and

$$Z^{**}(1/p) = X^*(1/p) - (c/p)Q_0(1/p)Y^{**}(1/p),$$
$$Z(1/p) = X(1/p) - (c/p)Q_0(1/p)Y(1/p). \quad (52b)$$

When $\mu < 0$, by passing to the limit for $R \to \infty$ and $r \to 0$ in both sides of Eq. (50), we get by straightforward algebra

$$F_0^{(e.s.)}(x,\mu) = \frac{1}{2}c\int_{-1}^{-\infty}\frac{e^{xp}}{p}\cdot\frac{1}{1+\mu p}\cdot\frac{dp}{Z^2(1/p) + (\pi c/2p)^2 Y^2(1/p)}, \quad (53)$$

which is a real integral to be performed numerically for any 0, any μ between -1 and 0, and any assigned set of the numbers $c, g_1, g_2, \cdots, g_N$.

Of course, the result of Eq. (52) is valid only if

$$Z^2(1/p) + (\pi c/2p)^2 Y^2(1/p) \neq 0 \quad (54)$$

for $-\infty \leq \text{rep} \leq -1$. We shall later on resume the case when the condition, Eq. (54), is violated. For now we stress that Eq. (54) is just the condition for the transcendental equation $\Delta(p; c, g_0, g_1, \cdots, g_N) = 0$, Eq. (40), to have all its own roots lying in the interval $(-1, 1)$.[6]

When $\mu > 0$, as for $r \to 0$ the integrals along the upper and lower edges of the cut become singular in the Cauchy sense, the evaluation of $F_0^{(e.s.)}(x,\mu)$ requires a more sophisticated treatment. Following the theory of Cauchy integrals as developed in the Muskhelishvili monograph,[15] we observe that

$$\varphi^+(x,p)[\varphi^-(x,p)]$$

satisfies the Hölder condition on the infinite arc $(-\infty, -1)[(-1, -\infty)]$ and then forms the sectionally holomorphic functions

$$\phi^+(x,z) = \frac{1}{2\pi i}\int_{-\infty}^{-1}\frac{\varphi^+(x,p)}{p-z}\,dp,$$
$$\phi^-(x,z) = \frac{1}{2\pi i}\int_{-1}^{-\infty}\frac{\varphi^-(x,p)}{p-z}\,dp, \quad (55)$$

[15] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Ltd., Groningen, The Netherlands, 1953).

when z is not on the corresponding arc. Furthermore, $\varphi^{\pm}(x, -\infty) = 0$, so that both $\phi^{+}(x, z)$ and $\phi^{-}(x, z)$ vanish at infinity.

It is easily inferred now, if r is sufficiently small, that

$$\lim_{r\to 0}\frac{1}{2\pi i}\int_{-\infty+ir}^{-1+ir}\frac{\varphi^{+}(x,p)}{p+(1/\mu)}\,dp = \phi^{+}\left(x, -\frac{1}{\mu} - i0\right),$$
$$\lim_{r\to 0}\frac{1}{2\pi i}\int_{-1-ir}^{-\infty-ir}\frac{\varphi^{-}(x,p)}{p+(1/\mu)}\,dp = \phi^{-}\left(x, -\frac{1}{\mu} + i0\right). \tag{56}$$

The problem is thus reduced to the one of determining the limiting values of the functions $\phi^{+}(x, z)$ and $\phi^{-}(x, z)$ from the right of their own arc of discontinuity. If $0 < \mu < 1$, the Plemelj formulas give[15]

$$\phi^{\pm}(x, -(1/\mu) \mp i0)$$
$$= -\tfrac{1}{2}\varphi^{\pm}[x, -(1/\mu)] \mp \frac{1}{2\pi i}\,\mathrm{P}\int_{-1}^{-\infty}\frac{\varphi^{\pm}(x,p)}{p+(1/\mu)}\,dp, \tag{57}$$

where the capital P indicates that the principal value of the considered integral is to be taken.

By passing to the limit for $R \to \infty$ and $r \to 0$ in both sides of Eq. (50) and taking into account Eqs. (52), (55), and (57), for the contribution of the essential singularity, when x and μ are both positive, we obtain

$$F_0^{(e.s.)}(x,\mu) = v\left(\frac{1}{\mu}\right)\frac{e^{-x/\mu}}{\mu} + \tfrac{1}{2}c\,\mathrm{P}\int_{-1}^{-\infty}\frac{e^{xp}w(p)}{1+\mu p}\,dp, \tag{58}$$

in which we have used the evenness of $v(p)$, Eq. (52a). With the change to the real variable $\eta = -p$, Eq. (58) becomes also

$$F_0^{(e.s.)}(x,\mu) = v(\mu)\frac{e^{-x/\mu}}{\mu} + \tfrac{1}{2}c\,\mathrm{P}\int_{1}^{\infty}e^{-x\eta}\chi(\eta,\mu,c)\,d\eta, \tag{59}$$

where

$$\chi(\eta,\mu,c) \equiv \chi(\eta,\mu;c,g_0,g_1,\cdots,g_N) = \left\{\eta(1-\mu\eta)\left[Z^2\left(\frac{1}{\eta}\right) + \left(\frac{\pi c}{2\eta}\right)^2 Y^2\left(\frac{1}{\eta}\right)\right]\right\}^{-1}. \tag{60}$$

5. EVALUATION OF $f_1(x, \mu)$, EQ. (37)

Entering Eq. (45) into Eq. (37), we find that for $x > 0$

$$f_1(x,\mu) = \frac{1}{4\pi}\Bigg\{U_0(\mu,c)\cdot\left[\sum_{m=0}^{M}F_{0,m}^{(p)}(x,\mu) + F_0^{(e.s.)}(x,\mu)\right] + \sum_{n=1}^{N}(-1)^n\frac{U_n(\mu,c)}{(n-1)!}\cdot\int_{\infty}^{x}\left[\sum_{m=0}^{M}F_{0,m}^{(p)}(u,\mu) + F_0^{(e.s.)}(u,\mu)\right]\cdot(x-u)^{n-1}\,du, \tag{61}$$

as follows from Eq. (36). The integrals involved in the right-hand side of Eq. (61) are readily evaluated. By reference to Eqs. (46) and (59) they are, respectively,

$$\int_{\infty}^{x}e^{-k_m u}\cdot(x-u)^{n-1}\,du = (-1)^n\frac{(n-1)!}{k_m^n}\,e^{-k_m x}, \tag{62a}$$

$$\int_{\infty}^{x}e^{-u/\mu}\cdot(x-u)^{n-1}\,du = (-1)^n\cdot(n-1)!\,\mu^n\cdot e^{-x/\mu}, \tag{62b}$$

$$\tfrac{1}{2}c\,\mathrm{P}\int_{1}^{\infty}\chi(\eta,\mu,c)\,d\eta\int_{\infty}^{x}e^{-\eta u}(x-u)^{n-1}\,du = (-1)^n\cdot(n-1)!\,\tfrac{1}{2}c\,\mathrm{P}\int_{1}^{\infty}\frac{\chi(\eta,\mu,c)}{\eta^n}\,e^{-x\eta}\,d\eta, \tag{62c}$$

for $n = 1, 2, \cdots, N$. Equation (61) then becomes

$$f_1(x,\mu) = \frac{1}{4\pi}\Bigg\{\sum_{m=0}^{M}\frac{\psi(k_m,c)}{1-\mu k_m}\,e^{-k_m x}\cdot\sum_{n=0}^{N}\frac{U_n(\mu,c)}{k_m^n} + \frac{e^{-x/\mu}}{\mu}\cdot\sum_{n=0}^{N}U_n(\mu,c)\mu^n + \tfrac{1}{2}c\,\mathrm{P}\sum_{n=0}^{N}U_n(\mu,c)\int_{1}^{\infty}\chi(\eta,\mu,c)\frac{e^{-x\eta}}{\eta^n}\,d\eta\Bigg\}, \tag{63}$$

which determines $f_1(x, \mu)$ when $\mu > 0$. When $\mu < 0$, the second term in the brackets vanishes and the integral term is not to be taken any longer as principal value.

6. INVERSION OF $f_2(p, \mu)$, EQ. (35)

Let us now turn our attention to the second component, $f_2(p, \mu)$, of $f(p, \mu)$, Eq. (35). We know that, when $\mu > 0$, the strip of convergence of the image $G_0(p, \mu)$, Eq. (34b), is

$$-1/\mu < \mathrm{rep} < \infty, \tag{64}$$

so that it includes the strip $-k_0 < \mathrm{rep} \leq 0$ with respect to which the inversion of $F_0(p, \mu)$, Eq. (44), has been performed to yield Eq. (63) for $f_1(x, \mu)$. Therefore the integration rule, Eq. (36), can still be applied termwise to obtain the original of $f_2(p, \mu)$ for $x > 0$. As a matter of fact, the strip $-1 < \mathrm{rep} < 0$ with the specifications of Eqs. (43) is the strip of convergence for the whole $f(p, \mu)$, Eq. (35). With this strip of convergence the original $f(x, \mu)$ of $f(p, \mu)$ is uniquely determined by

$$f(x,\mu) = f_1(x,\mu) + f_2(x,\mu). \tag{65}$$

For[10]

$$\mathfrak{L}_{\Pi}^{-1}\{G_0(p,\mu)\} = e^{-x/\mu}\cdot 1(x), \tag{66}$$

where $1(x)$ is the unit function, we get for x and $\mu > 0$

$$f_2(x,\mu) = \frac{1}{4\pi}\cdot\frac{e^{-x/\mu}}{\mu}\cdot\left\{1 - \sum_{n=0}^{N}U_n^{*}(\mu,c)\cdot\mu^n\right\}, \tag{67}$$

as follows from Eqs. (36) and (62b).

It is interesting to remark that the point $p = -1/\mu$, which acts as a pseudopole when transforming back $f_1(p, \mu)$, is a true pole when dealing with the inversion of $f_2(p, \mu)$. Of course, if $\mu < 0$, $f_2(x, \mu)$ is identically zero.

7. THE SPACE-ANGLE FLUX DISTRIBUTION $f(x, \mu)$

According to Eq. (65), we are now able to write down the solution of Eq. (1). Referring to positive values of x and to the validity of Eq. (54), we have

$$f(x,\mu) = \frac{1}{4\pi} \cdot \left\{ \sum_{m=0}^{M} \frac{\psi(k_m, c)}{1 - \mu k_m} e^{-k_m x} \cdot \sum_{n=0}^{N} \frac{U_n(\mu, c)}{k_m^n} + \frac{e^{-x/\mu}}{\mu} \cdot \left[1 + \sum_{n=0}^{N} (v(\mu) U_n(\mu, c) - U_n^*(\mu, c)) \mu^n \right] + \tfrac{1}{2} c \, \mathrm{P} \sum_{n=0}^{N} U_n(\mu, c) \cdot \int_1^\infty \frac{e^{-x\eta}}{\eta^n} \chi(\eta, \mu, c) \, d\eta \right\} \tag{68}$$

when $\mu > 0$, and

$$f(x,\mu) = \frac{1}{4\pi} \cdot \left\{ \sum_{m=0}^{M} \frac{\psi(k_m, c)}{1 - \mu k_m} \cdot e^{-k_m x} \sum_{n=0}^{N} \frac{U_n(\mu, c)}{k_m^n} + \tfrac{1}{2} c \sum_{n=0}^{N} U_n(\mu, c) \int_1^\infty \frac{e^{-x\eta}}{\eta^n} \chi(\eta, \mu, c) \, d\eta \right\} \tag{69}$$

when $\mu < 0$.

Dual expressions are in order when negative values of x are considered. We observe that both the solutions, Eqs. (68) and (69), are of the form as proposed in Eq. (2) and the coefficients $U_n(\mu, c)$ and $U_n^*(\mu, c)$ follow readily from Eqs. (33). As far as the coefficients k_m of the exponential decays are concerned, we remember that in most problems of neutron physics $M = 0$, so that the only root $\pm k_0$ of Eq. (40) can be extracted, as done by Holte[16] when $|1 - c| \ll 1$, and as resumed successively by other authors.[17–19]

Concentrating on Eq. (68), we conclude with the following considerations:

(i) The sum with respect to $e^{-k_m x}$ is the contribution of the poles lying on the real axis between -1 and 0. The term in $e^{-k_0 x}$ survives asymptotically as x increases to infinity and the reciprocal of k_0 is the usual diffusion length of the considered medium in terms of the mean free path.

(ii) The term in $e^{-x/\mu}$ is the direct consequence of having solved Eq. (4) in the form of Eq. (8). For the solution, Eq. (68), in its whole, we do not need to know whether the contribution in $e^{-x/\mu}$ comes from the principal pseudopole at $p = -1/\mu$ of $f_1(p, \mu)$ or from the true pole, still at $p = -1/\mu$, of $f_2(p, \mu)$. It is just the presence of the term in $e^{-x/\mu}$ which might imply the existence of a physically anomalous diffusion length. In other words, the term in $e^{-x/\mu}$ can be regarded as a pseudopole effect.

(iii) The integral term, which is the transient component of the distribution, represents at last the contribution of the essential singularity.

Let us check now how Eq. (68) reduces in the limit of isotropic scattering, that is, when $N = 0$. It is found that

$$f^{(i)}(x,\mu) = \frac{1}{4\pi} \cdot \left\{ \frac{k_0(1 - k_0^2)}{k_0^2 - (1 - c)} \cdot \frac{e^{-k_0 x}}{1 - \mu k_0} + \frac{1 - c\mu Q_0(\mu)}{[1 - c\mu Q_0(\mu)]^2 + (\pi c \mu/2)^2} \cdot \frac{e^{-x/\mu}}{\mu} + \tfrac{1}{2} c \, \mathrm{P} \int_1^\infty \frac{e^{-x\eta}}{1 - \mu\eta} \cdot \frac{\eta \, d\eta}{[\eta - c Q_0(1/\eta)]^2 + (\pi c/2)^2} \right\}, \tag{70}$$

which coincides with the Case result, as obtained by means of the eigenfunction expansion method.[20]

8. VIOLATION OF THE CONDITION, EQ. (54)

It may happen that, for special values of the scattering probability c and of the Legendre moments $g_1, g_2, \cdots, g_N$ of the scattering kernel, the condition expressed by Eq. (54) is not fulfilled any longer. This means, as follows from Eqs. (40) and (48), that the function

$$\Delta(p, c, g_0, g_1, \cdots, g_N) = X(1/p) - (c/p) Q_0(1/p) Y(1/p) \tag{71}$$

may vanish at the edges of the cut of Fig. 1; that is, it may possess some roots outisde the interval $(-1, 1)$. This is the case when $X(1/p)$ and $Y(1/p)$, which are polynomials of $2N$th order, have common zeros in the real intervals $(-\infty, -1)$ and $(1, \infty)$. It can be shown[6] that if $2(M + 1)$ is the number of the real roots in $(-1, 1)$ and $2M^*$ is the number of the real roots outside $(-1, 1)$ of Eq. (71), then $M + M^* \leq N$. Conditions for the absence or presence of the $2M^*$ additional roots are given in Ref. 3. We do not enter here the details of such a problem, but prefer on the contrary to underline the following points.

As $\Delta(p, c, g_0, g_1, \cdots, g_N)$, Eq. (71), is still the determinant, Eq. (18), related to the degeneracy of

[16] G. Holte, Arkiv Mat. Astron. Fys. **35A**, No. 36 (1948).
[17] I. Kŭscer, J. Math. Phys. **34**, 256 (1955).
[18] E. Inonü and A. I. Usseli, Nucl. Sci. Eng . **23**, 251 (1965).
[19] V. C. Boffi and T. Trombetti, Nuovo Cimento **47B**, 210 (1967).
[20] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).

the kernel of Eq. (5), we can thus argue on the properties of the roots of Eq. (71):

(i) The real roots, $\mp k_0, \mp k_1, \cdots, \mp k_M$, lying in the interval $(-1, 1)$,act as true poles when transforming back the solution $f(p, \mu)$, Eq. (8), of Eq. (5). Their contributions are asymptotic in the sense of the stability of the solution.

(ii) The real roots outside the interval $(-1, 1)$ behave like pseudopoles when dealing with the evaluation of the essential singularity, Eq. (50), of the function $F_0(p, \mu)$ [this is the only part of the solutions, Eqs. (68) and (69), to be modified].

The location of all these $2(M + M^* + 1)$ roots depends on the set of numbers $c, g_0, g_1, \cdots, g_N$ and their number depends on the degree N of the considered anisotropy. We realize that the principal pseudopole at $p = -1/\mu$ is not included among the zeros of $\Delta(p, c, g_0, g_1, \cdots, g_N)$ and is therefore independent of the characteristics of scattering. It depends only on the purely angular variable associated with the streaming term of the transformed balance equation. Comparing Eq. (68) with Eq. (71) makes us aware that both the anisotropic and isotropic scattering solutions consist, in principle, of the sum of three different contributions due to the pole, the pseudopole and the essential singularity, respectively. It is only a special combination of the values of c and $g_0, g_1, \cdots, g_N$, which, in the case of anisotropic scattering, may cause $X(1/p)$ and $Y(1/p)$ to have common real zeros outside the interval $(-1, 1)$ and, hence, the condition, Eq. (54), to be violated. It follows that additional or secondary pseudopoles will appear in the spectrum of singularities of $f(p, \mu)$, Eq. (8).

Let us pass on now to resolve the contribution of secondary pseudopoles. In order to simplify the notations, we operate the conformal transformation $u = -1/p$ in Eq. (47) so that Eq. (50) can be written as

$$F_0^{(e,s.)}(x, \mu) = \lim_{r \to 0} \left\{ -\frac{1}{2\pi i} \int_{0+ir}^{1+ir} \frac{e^{-x/u}}{u} \cdot \frac{1}{u - \mu} \cdot \frac{Z^{**}(u) + i\frac{1}{2}\pi c u Y^{**}(u)}{Z(u) + i\frac{1}{2}\pi c u Y(u)} du - \frac{1}{2\pi i} \int_{1-ir}^{0-ir} \frac{e^{-x/u}}{u} \cdot \frac{1}{u - \mu} \cdot \frac{Z^{**}(u) - i\frac{1}{2}\pi c u Y^{**}(u)}{Z(u) - i\frac{1}{2}\pi c u Y(u)} du \right\}. \quad (72)$$

The interval $(-\infty, -1)$ $[(-1, -\infty)]$ thus changes into the finite interval $(0, 1)$ $[(1, 0)]$.

We first examine the case when $X(u)$ and $Y(u)$ have a single common real root k_1^* in $(0, 1)$. Let $\nu_1^* = \nu_1 + 1$ be the order of such a root. We may then write

$$X(u) = (u - k_1^*)^{\nu_1^*} \cdot \sum_{n=0}^{2N-\nu_1^*} \frac{X^{(n+\nu_1^*)}(k_1^*)}{(n + \nu_1^*)!} \cdot (u - k_1^*)^n, \quad (73a)$$

where $X^{(n+\nu_1^*)}$ denotes the $(n + \nu_1^*)$th derivative of $X(u)$ at $u = k_1^*$ and

$$X(k_1^*) = X^{(1)}(k_1^*) = \cdots = X^{(\nu_1)}(k_1^*) = 0. \quad (73b)$$

Analogous expressions hold for $Y(u)$.

Setting

$$\phi^{\pm}(x; c, g_0, g_1, \cdots, g_N) \equiv \phi^{\pm}(x, u) = \frac{e^{-x/u}}{u} \cdot [Z^{**}(u) \pm i\tfrac{1}{2}cuY^{**}(u)] \cdot \left\{ \sum_{n=0}^{2N-\nu_1^*} [X^{(n+\nu_1^*)}(k_1^*) - cuQ_0(u)Y^{(n+\nu_1^*)}(k_1^*) \pm i\tfrac{1}{2}\pi c u Y^{(n+\nu_1^*)}(k_1^*)] \frac{(u - k_1^*)^n}{(n + \nu_1^*)!} \right\}^{-1}, \quad (74)$$

Eq. (72) becomes

$$F_0^{(e,s.)}(x, \mu) = \lim_{r \to 0} \left\{ -\frac{1}{2\pi i} \int_{0+ir}^{1+ir} \frac{\phi^+(x, u)}{(u - \mu)(u - k_1^*)^{\nu_1^*}} du - \frac{1}{2\pi i} \int_{1-ir}^{0-ir} \frac{\phi^-(x, u)}{(u - \mu)(u - k_1^*)^{\nu_1^*}} du \right\}. \quad (75)$$

When $\mu < 0$, we further set

$$\phi^{\pm}(x, \mu; u) = \frac{\phi^{\pm}(x, u)}{u - \mu} \quad (76)$$

and then obtain

$$F_0^{(e,s.)}(x, \mu) = \lim_{r \to 0} \left\{ -\frac{1}{2\pi i} \int_{0+ir}^{1+ir} \frac{\phi^+(x, \mu; u)}{(u - k_1^*)^{\nu_1+1}} du - \frac{1}{2\pi i} \int_{1-ir}^{0-ir} \frac{\phi^-(x, \mu; u)}{(u - k_1^*)^{\nu_1+1}} du \right\}. \quad (77)$$

Because $\phi^+(x, \mu; u)[\phi^-(x, \mu; u)]$ is continuous and regular with all its derivatives of any order on the edges of the cut and it satisfies the Hölder condition on the arc $(0, 1)$ $[(1, 0)]$, we may then form the functions

$$F_{\nu_1^*}^+(x, \mu; s) = \frac{1}{2\pi i} \int_0^1 \frac{\phi^+(x, \mu; u)}{(u - s)^{\nu_1+1}} du,$$

$$F_{\nu_1^*}^-(x, \mu; s) = \frac{1}{2\pi i} \int_1^0 \frac{\phi^-(x, \mu; u)}{(u - s)^{\nu_1+1}} du, \quad (78)$$

where s is any point of the complex plane outside the arcs $(0, 1)$ and $(1, 0)$, respectively.

Arguing as in the case of the principal pseudopole, we recognize that

$$\lim_{r\to 0}\frac{1}{2\pi i}\int_{0+ir}^{1+ir}\frac{\phi^+(x,\mu;u)}{(u-k_1^*)^{\nu_1+1}}\,du = F_{\nu_1}^{+}{}^*(x,\mu;k_1^*-i0),$$
$$\lim_{r\to 0}\frac{1}{2\pi i}\int_{1-ir}^{0-ir}\frac{\phi^-(x,\mu;u)}{(u-k_1^*)^{\nu_1+1}}\,du = F_{\nu_1}^{-}{}^*(x,\mu;k_1^*+i0). \tag{79}$$

On the other hand, since

$$F_{\nu_1}^{+}{}^*(x,\mu;s) = \frac{1}{\nu_1!}F_1^{+(\nu_1)}(x,\mu;s) \equiv \frac{1}{\nu_1!}\frac{d^{\nu_1}}{ds^{\nu_1}}\left\{\frac{1}{2\pi i}\int_0^1\frac{\phi^+(x,\mu;u)}{u-s}\,du\right\},$$
$$F_{\nu_1}^{-}{}^*(x,\mu;s) = \frac{1}{\nu_1!}F_1^{-(\nu_1)}(x,\mu;s) \equiv \frac{1}{\nu_1!}\frac{d^{\nu_1}}{ds^{\nu_1}}\left\{\frac{1}{2\pi i}\int_1^0\frac{\phi^-(x,\mu;u)}{u-s}\,du\right\}, \tag{80}$$

the problem expressed by Eqs. (79) reduces to investigating the behavior of the ν_1th derivative of Cauchy integrals

$$F_1^+(x,\mu;s) = \frac{1}{2\pi i}\int_0^1\frac{\phi^+(x,\mu;u)}{u-s}\,du,$$
$$F_1^-(x,\mu;s) = \frac{1}{2\pi i}\int_1^0\frac{\phi^-(x,\mu;u)}{u-s}\,du, \tag{81}$$

when s approaches their arc of discontinuity from the right-hand side.

That the limiting value of the ν_1th derivative of the above Cauchy integrals exists when s tends to k_1^* is insured[21] if the general derivative, from the first to the ν_1th order, of the functions $\phi^+(x,\mu;u)$ and $\phi^-(x,\mu;u)$ satisfies the Hölder condition on the arcs (0, 1) and (1, 0), respectively, as occurs in the present context. Thus Eqs. (78) represent two sectionally holomorphic functions.

Let us illustrate how the limits on the right-hand side of Eqs. (79) can be evaluated in the case $\nu_1 = 1$. Then

$$F_2^+(x,\mu;s) \equiv F_1^{+(1)}(x,\mu;s) = \frac{1}{2\pi i}\int_0^1\frac{\phi^+(x,\mu;u)}{(u-s)^2}\,du,$$
$$F_2^-(x,\mu;s) \equiv F_1^{-(1)}(x,\mu;s) = \frac{1}{2\pi i}\int_1^0\frac{\phi^-(x,\mu;u)}{(u-s)^2}\,du, \tag{82}$$

and it is at once verified that[21]

$$\frac{1}{2\pi i}\int_0^1\frac{\phi^+(x,\mu;u)}{(u-s)^2}\,du = -\left[\frac{\phi^+(x,\mu;u)}{u-s}\right]_{u=0}^{u=1} + \frac{1}{2\pi i}\int_0^1\frac{\phi^{+(1)}(x,\mu;u)}{u-s}\,du,$$
$$\frac{1}{2\pi i}\int_1^0\frac{\phi^-(x,\mu;u)}{(u-s)^2}\,du = -\left[\frac{\phi^-(x,\mu;u)}{u-s}\right]_{u=1}^{u=0} + \frac{1}{2\pi i}\int_1^0\frac{\phi^{-(1)}(x,\mu;u)}{u-s}\,du. \tag{83}$$

If k_1^* does not coincide with the ends of the given arc of discontinuity and if, in the limit of s tending to k_1^*, we apply the Plemelj formulas to the right-hand side of Eq. (83), we find

$$F_1^{\pm}(x,\mu;k_1^*\mp i0) = \mp\left[\frac{\phi^{\pm}(x,\mu;1)}{1-k_1^*}+\frac{\phi^{\pm}(x,\mu;0)}{k_1^*}\right] - \tfrac{1}{2}\phi^{\pm(1)}(x,\mu;k_1^*) \pm \frac{1}{2\pi i}\,\mathrm{P}\int_0^1\frac{\phi^{\pm(1)}(x,\mu;u)}{u-k_1^*}\,du. \tag{84}$$

Using Eqs. (79) and (84) in Eq. (77), we obtain for the contribution of the essential singularity for $x > 0$ and $\mu < 0$, when a secondary pseudopole of multiplicity 2 is present,[22] the explicit expression

$$F_0^{(\mathrm{e.s.})}(x,\mu) = \frac{\phi^+(x,\mu;0)-\phi^-(x,\mu;0)}{k_1^*} + \frac{\phi^+(x,\mu;1)-\phi^-(x,\mu;1)}{1-k_1^*} + \tfrac{1}{2}[\phi^{+(1)}(x,\mu;k_1^*)-\phi^{-(1)}(x,\mu;k_1^*)] - \frac{1}{2\pi i}\,\mathrm{P}\int_0^1\frac{\phi^{+(1)}(x,\mu;u)-\phi^{-(1)}(x,\mu;u)}{u-k_1^*}\,du, \tag{85}$$

which holds when k_1^* is in the interior of the interval (0, 1) and if $\mu \neq k_1^*$. When $\mu \equiv k_1^*$, it is sufficient to increase the order of multiplicity of the secondary pseudopole at $s = k_1^*$ by unity.

The case $\nu_1 > 1$ can be treated analogously by proceeding progressively to the higher-order derivatives.

When $\mu > 0$, one has to go back to Eq. (75). If we denote $\mu \neq k_1^*$ by k_0^*, a partial fraction analysis allows us to write the expansion

$$\frac{1}{(u-k_0^*)(u-k_1^*)^{\nu_1^*}} = \sum_{m=0}^{1}\sum_{n=1}^{\nu_m^*}\frac{A_{mn}}{(u-k_m^*)^{\nu_m^*-n+1}}, \tag{86}$$

[21] N. I Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity* (P. Noordhoff, Ltd., Groningen, The Netherlands, 1963).

[22] The case $\mu < 0$ with a first-order secondary pseudopole is analogous to the case when only the principal pseudopole is present.

where $\nu_0^* = 1$ and A_{mn} are certain coefficients given by

$$A_{mn} = \frac{1}{(n-1)!}\left[\frac{d^{n-1}}{du^{n-1}}\frac{(u-k_m^*)^{\nu_m^*}}{(u-k_0^*)(u-k_1^*)^{\nu_1^*}}\right]_{u=k_m^*}. \tag{87}$$

By using Eq. (86), each term on the right-hand side of Eq. (75) reduces to a linear combination of the derivatives of the functions $F_1^+(x,\mu;s)$ and $F_1^-(x,\mu;s)$, Eqs. (81), respectively. No difficulty arises then in extending the procedure illustrated above for calculating the limiting value of such linear combinations when s tends to k_0^* and k_1^*, separately.

The general case, when $X(u)$ and $Y(u)$ have more than one single common root in (0, 1), is also amenable to the simpler one before considered by still resorting to a partial fraction expansion.

ACKNOWLEDGMENT

The author wishes to thank Dr. T. Trombetti for useful discussions on the argument of this work.

APPENDIX

That $f_n(p)$, as given by Eq. (17), is a solution of the system, Eq. (9), can be verified as follows. By introducing Eq. (17) into Eq. (9), we find the condition that

$$f_0(p) = \frac{I_{m0} - R_m(1/p) + c\sum_{n=1}^{N} b_n I_{mn}(p) R_n(1/p)}{S_m(1/p) - c\sum_{n=0}^{N} b_n I_{mn}(p) S_n(1/p)} \tag{A1}$$

for any m between 0 and N. By using Eqs. (15) we recognize that the numerator and the denominator of Eq. (A1) take, respectively, the form

$$(-1)^m P_m(1/p)\cdot\left[I_{00}(p) + c\sum_{n=1}^{N} b_n I_{0n}(p) R_n(1/p)\right] \tag{A2a}$$

and

$$(-1)^m P_m(1/p)\cdot\left[1 - c\sum_{n=0}^{N} b_n I_{0n}(p) S_n(1/p)\right], \tag{A2b}$$

provided

$$(-1)^{m+1} W_{m-1}(1/p) - R_m(1/p) + c\sum_{n=1}^{m} b_n I_{mn}(p) R_n(1/p) = (-1)^m P_m(1/p) c\cdot\sum_{n=1}^{m} b_n I_{0n}(p) R_n(1/p) \tag{A3a}$$

in the case of Eq. (A2a) and

$$S_m(1/p) - c\sum_{n=0}^{m-1} b_n I_{mn}(p) S_n(1/p) = (-1)^m P_m(1/p)\cdot\left[1 - c\sum_{n=1}^{m-1} b_n I_{0n}(p) S_n(1/p)\right] \tag{A3b}$$

in the case of Eq. (A2b).

After appropriate manipulations and using Eqs. (12), (13), and (15), Eqs. (A3) can be rewritten as

$$H_m\left(\frac{1}{p}\right) - H_m^{(i)}\left(\frac{1}{p}\right) = (-1)^m \frac{c}{p}\sum_{n=1}^{m-1}(-1)^n\cdot b_n\cdot\left[P_m\left(\frac{1}{p}\right)W_{n-1}\left(\frac{1}{p}\right) - W_{m-1}\left(\frac{1}{p}\right)P_n\left(\frac{1}{p}\right)\right]\cdot H_n\left(\frac{1}{p}\right), \tag{A4}$$

where H_m (and $H_m^{(i)}$) stands for either R_m or S_m (and for $R_m^{(i)}$ or $S_m^{(i)}$).

Equation (A4), which permits expressing the polynomials H_m of the anisotropic problem as a linear combination of the polynomials $H_m^{(i)}$ of the isotropic case, reduces to the identity when using in it each of the pairs R_m, $R_m^{(i)}$ and S_m, $S_m^{(i)}$ given by Eqs. (23) and (28). The condition, Eq. (A1), for $f_n(p)$ [Eq. (17)] to be a solution of the system, Eq. (9), is thus satisfied.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Generalized Young Tableaux and the General Linear Group

R. C. KING
Mathematics Department, The University, Southampton, England

(Received 30 October 1968)

A generalization of the Young tableau is defined, and use is made of this in the study of some of the properties of the irreducible representations (IR's) of each of the linear groups in n dimensions induced in a space defined by mixed tensors without recourse to lowering or raising of indices. A formula for the dimensions of any IR of L_n is given. Procedures are derived for the reduction of the outer product of such IR's and for the decomposition of these IR's into IR's of some subgroups of interest in the theory of elementary particles.

1. INTRODUCTION

The relationship between the symmetric group on a symbols Σ_a and the general linear group in n dimensions, $GL(n)$, was first recognized by Schur. The theory of this relationship was greatly developed by Weyl[1] who made use of the work of Young concerning the representations of Σ_a. It was proved by Weyl that the imposition of maximal symmetry conditions on the a indices of the covariant or contravariant tensors, defining a space in which is induced a representation of $GL(n)$, decomposes that space into invariant subspaces in which are induced irreducible representations (IR's) of $GL(n)$. The maximal symmetry conditions are just those defined by the Young symmetrizers corresponding to a Young tableau which specifies an irreducible representation (IR) of Σ_a.

By making use of the theory of characters, Murnaghan[2] and Littlewood[3] were able to reduce the inner and outer products of IR's of the symmetric groups into IR's of the symmetric groups. The importance of Young's work was particularly stressed by Robinson[4] who derived procedural rules for carrying out these reductions. These rules were expressed solely in terms of Young tableaux and thus obviated the necessity of calculating characters. The results obtained by these authors are applicable to those IR's of $GL(n)$ whose bases are either covariant or contravariant tensors.

More generally, mixed tensors may be used to form the bases of IR's of $GL(n)$, but it has been customary to raise or lower indices appropriately in order to make use of the duality of Σ_a and $GL(n)$ as expressed in the Young tableaux. In Sec. 2 of this paper, a generalization of the Young tableau is defined and use is made of this in Secs. 3–6 to study some of the properties of the IR's of $GL(n)$ induced in a space defined by mixed tensors, without recourse to raising or lowering of indices.

The motivation for this work was provided by the application of group theory to the study of elementary particles. In these applications, the elementary particles are considered to be the basis states of the IR's of the appropriate symmetry group. In order to account for the multiplicity of particles and their strong interactions within such a symmetry scheme, it is necessary to know the dimensions of these IR's and the reductions of their outer products. These aspects of the theory of the IR's of $GL(n)$ and of some of its subgroups relevant to the theory of elementary particles are discussed in Secs. 3 and 4.

It is known that the symmetry schemes are broken so that it is also necessary to discuss some of the subgroups of the symmetry group. In particular, the decomposition formulas which give the IR's of the subgroups $GL(m) \circ GL(n)$ and $GL(m) \otimes GL(n)$ contained in the IR's of $GL(m+n)$ and $GL(mn)$, respectively, are required. The analysis of these decompositions is given by the reduction of the outer and inner products of the IR's of the symmetric group for those IR's of the general linear group induced in a space defined by either covariant or contravariant tensors. In Secs. 5 and 6, decomposition formulas are derived which are appropriate to the generalization to mixed tensors.

The results are discussed in Sec. 7 with particular reference to applications to the theory of elementary particles.

2. INEQUIVALENT IRREDUCIBLE REPRESENTATIONS OF $GL(n)$

The set of mixed tensors

$$T^{\beta_1\beta_2\cdots\beta_b}_{\alpha_1\alpha_2\cdots\alpha_a}$$

forms the basis of a representation of $GL(n)$ if all the indices may take the values $1, 2, \cdots, n$ and if the

[1] H. Weyl, *The Classical Groups, Their Invariants and Representations* (Princeton University Press, Princeton, N.J., 1939).
[2] F. D. Murnaghan, *The Theory of Group Representations* (The Johns Hopkins Press, Baltimore, 1938).
[3] D. E. Littlewood, *The Theory of Group Characters* (Oxford University Press, London, 1940).
[4] G. De B. Robinson, *Representation Theory of the Symmetric Group* (The University Press, Edinburgh, 1961).

linear transformation

$$T^{\beta_1\beta_2\cdots\beta_b}_{\alpha_1\alpha_2\cdots\alpha_a} \to T'^{\beta_1\beta_2\cdots\beta_b}_{\alpha_1\alpha_2\cdots\alpha_a} = A^{\alpha_1'}_{\alpha_1} A^{\alpha_2'}_{\alpha_2} \cdots A^{\alpha_a'}_{\alpha_a} A^{-1\beta_1}_{\beta_1'} \cdots A^{-1\beta_b}_{\beta_b'} T^{\beta_1'\beta_2'\cdots\beta_b'}_{\alpha_1'\alpha_2'\cdots\alpha_a'} \tag{2.1}$$

is associated with every nonsingular $n \times n$ matrix A. The representation is irreducible if the tensors are traceless and if the sets of covariant indices $\alpha_1, \alpha_2, \cdots, \alpha_a$ and contravariant indices $\beta_1, \beta_2, \cdots, \beta_b$ satisfy the permutational symmetry conditions defined by the Young tableaux $[\mu]_a$ and $[\nu]_b$, corresponding to IR's of the symmetric groups Σ_a and Σ_b, respectively.

The set of tensors (2.1) forms the basis of IR's of the unitary group $U(n)$ and the pseudounitary group $U(n-m, m)$ if the appropriate restrictions are placed on A. Similarly, if the additional restriction $\det A = 1$ is imposed, this set forms the basis of IR's of the unimodular groups $SL(n)$, $SU(n)$, and $SU(n-m, m)$.[5] This restriction leads to equivalences between IR's. For example, the totally antisymmetric pseudotensor $\epsilon_{\gamma_1\gamma_2\cdots\gamma_n}$ may be used to lower indices so that the basis of every finite-dimensional inequivalent IR of SL_n may be chosen to be a set of covariant tensors $T_{\alpha_1\alpha_2\cdots\alpha_a}$, whose index symmetry is specified by a single regular Young tableau[6]

$$[\mu]_a = [\mu_1, \mu_2, \cdots, \mu_p]_a = (\mu_1', \mu_2', \cdots, \mu_q')_a, \tag{2.2}$$

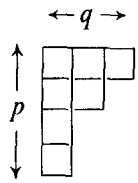
where

$$\mu_1 + \mu_2 + \cdots + \mu_p = \mu_1' + \mu_2' + \cdots + \mu_q' = a$$

and

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_p > 0,$$
$$\mu_1' \geq \mu_2' \geq \cdots \geq \mu_q' > 0,$$

with $p = \mu_1'$ and $q = \mu_1$. Diagrammatically, the tableau $[\mu]_a$ takes the typical form



with μ_i and μ_j' "boxes" in the ith row and jth column, respectively. Such a tableau $[\mu]_a$ specifies an IR of Σ_a and also an IR of each of the groups L_n, with $n \geq p$. However, for the groups L_n with $n < p$, no such IR's exist since the corresponding tensors all vanish identically. Distinct diagrams $[\mu]_a$ with p rows specify inequivalent IR's of Σ_a and of each L_n with $n > p$. However, although all the distinct tableau

$$[\mu + \epsilon]_{a+n\epsilon} = [\mu_1 + \epsilon, \mu_2 + \epsilon, \cdots, \mu_p + \epsilon, \epsilon, \epsilon, \cdots, \epsilon]_{a+n\epsilon}, \tag{2.3}$$

with ϵ any integer such that $\epsilon \geq -\mu_p$, specify inequivalent IR's of GL_n, they also specify equivalent IR's of SL_n. Thus, all inequivalent finite-dimensional IR's of SL_n are specified by the distinct tableaux $[\mu]_a$, with $p < n$.

The tableau conjugate to $[\mu]_a$ is defined by

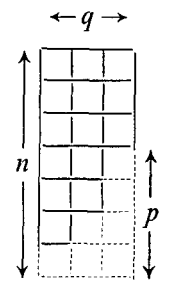
$$[\mu]_a' = [\mu']_a = [\mu_1', \mu_2', \cdots, \mu_q']_a = (\mu_1, \mu_2, \cdots, \mu_p)_a \tag{2.4}$$

and is obtained from $[\mu]_a$ by interchanging rows and columns. Corresponding to $[\mu']_a$, there exists an IR of each of the groups L_n with $n \geq q$.

The IR of L_n, complementary to that specified by $[\mu]_a$, is defined by the tableau

$$[\mu]_a^* = [\mu^*]_{nq-a} = [\mu_1^*, \mu_2^*, \cdots, \mu^*_{n-\mu'_q}]_{nq-a} = (\mu_{1'}^*, \mu_{2'}^*, \cdots, \mu_q^{*\prime})_{nq-a} \tag{2.5}$$

with $\mu_j^{*\prime} = n - \mu'_{q-j+1}$ for $j = 1, 2, \cdots, q$. Diagrammatically, $[\mu^*]_{nq-a}$ takes the typical form



In the same way, $\epsilon^{\gamma_1\gamma_2\cdots\gamma_n}$ may be used to raise indices so that an alternative basis of every inequivalent finite-dimensional IR of SL_n may be chosen to be a set of contravariant tensors

$$T^{\beta_1\beta_2\cdots\beta_b}$$

whose index symmetry is specified by a single regular Young tableau[7]

$$[\nu]^b = [\nu_1, \nu_2, \cdots, \nu_r]^b = (\nu_1', \nu_2', \cdots, \nu_s')^b, \tag{2.6}$$

where

$$\nu_1 + \nu_2 + \cdots + \nu_r = \nu_1' + \nu_2' + \cdots + \nu_s' = b$$

[5] Hereafter, the linear groups $GL(n)$, $U(n)$, and $U(n-m, m)$ are denoted collectively by GL_n, and the unimodular linear groups $SL(n)$, $SU(n)$, and $SU(n-m, m)$ are denoted by SL_n. In addition, GL_n and SL_n are denoted collectively by L_n wherever this is appropriate.

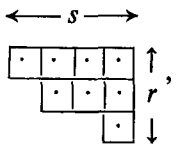
[6] In the notation of Ref. 1, the tableau $[\mu]_a$ has signature $(\mu_1, \mu_2, \cdots, \mu_p, 0, 0, \cdots, 0)$, and is given by $T(\mu_1, \mu_2, \cdots, \mu_p, 0, 0, \cdots, 0)$.

[7] In the notation of Ref. 1, the tableau $[\nu]^b$ has signature $(0, 0, \cdots, 0, -\nu_r, \cdots, -\nu_2, -\nu_1)$, and is given by $T(\nu_1, \nu_2, \cdots, \nu_r, 0, 0, \cdots, 0)$.

and

$$\nu_1 \geq \nu_2 \geq \cdots \geq \nu_r > 0,$$
$$\nu'_1 \geq \nu'_2 \geq \cdots \geq \nu'_s > 0,$$

with $\nu_1 = s$ and $\nu'_1 = r$. It is convenient to distinguish between the covariant and contravariant bases of the IR's of L_n, not only by the use of lower and upper indices a and b in $[\mu]_a$ and $[\nu]^b$, but also by adopting the convention whereby the tableau $[\nu]^b$ takes the typical form



with ν_k and ν'_l "dotted boxes" in the kth row and lth column, respectively. Such a tableau $[\nu]^b$ specifies an IR of Σ_b and also an IR of each of the groups L_n, with $n \geq r$.

The tableau adjoint to $[\mu]^b$ is defined to be the "dotted" tableau

$$\overline{[\mu]}_a = [\mu]^a = [\mu_1, \mu_2, \cdots, \mu_p]^a = (\mu'_1, \mu'_2, \cdots, \mu'_q)^a. \tag{2.7}$$

This tableau $[\mu]^a$ specifies the same IR of Σ_a as does the tableau $[\mu]_a$. However, it also specifies an IR of each L_n with $n \geq p$, which is adjoint or contragredient to that specified by $[\mu]_a$.

The tableau conjugate to $[\nu]^b$ is defined by

$$[\nu]^{b'} = [\nu']^b = [\nu'_1, \nu'_2, \cdots, \nu'_s]^b = (\nu_1, \nu_2, \cdots, \nu_r)^b, \tag{2.8}$$

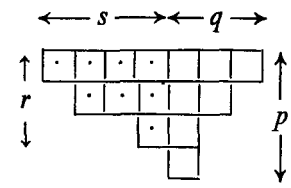
and the IR of L_n, complementary to that specified by $[\nu]^b$, is defined by the tableau

$$[\nu]^{b*} = [\nu^*]^{ns-b} = [\nu^*_1, \nu^*_2, \cdots, \nu^*_{n-\nu'_s}]^{ns-b}$$
$$= (\nu_1^{*\prime}, \nu_2^{*\prime}, \cdots, \nu_s^{*\prime})^{ns-b}, \tag{2.9}$$

with $\nu_l^{*\prime} = n - \nu'_{s-l+1}$ for $l = 1, 2, \cdots, s$. Clearly, with these definitions the adjoint and complement of an IR of SL_n are equivalent, whereas for GL_n they are inequivalent. Indeed, to complete the classification of all the inequivalent finite-dimensional IR's of GL_n, the diagrammatic notation may be extended so that the general mixed tensors (2.1) are defined by a regular composite tableau[8]

$$[\nu; \mu]^b_a = [\nu_1, \nu_2, \cdots, \nu_r; \mu_1, \mu_2, \cdots, \mu_p]^b_a$$
$$= (\nu'_1, \nu'_2, \cdots, \nu'_s; \mu'_1, \mu'_2, \cdots, \mu'_q)^b_a, \tag{2.10}$$

which takes the typical form



so that it is composed of the two regular tableaux $[\mu]_a$ and $[\nu]^b$ joined back-to-back.[9] The corresponding set of tensors (2.1) is such that $[\mu]_a$ describes the symmetry of the covariant indices $\alpha_1, \alpha_2, \cdots, \alpha_a$, $[\nu]^b$ describes the symmetry of the contravariant indices $\beta_1, \beta_2, \cdots, \beta_b$, and the back-to-back notation is taken to indicate that the tensors (2.1) are traceless. The tracelessness condition implies that the only admissible, regular, composite tableaux corresponding to IR's of L_n are those for which $p + r \leq n$.

Distinct, admissible, regular, composite tableaux specify inequivalent IR's of GL_n, but for SL_n the IR corresponding to $[\nu; \mu]^b_a$ is equivalent to the IR specified by

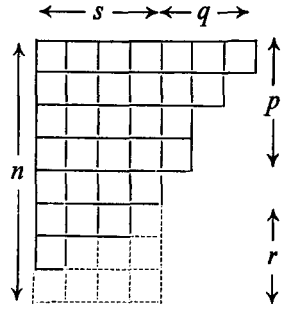
$$[\lambda]_c = [\lambda_1, \lambda_2, \cdots]_c = (\lambda'_1, \lambda'_2, \cdots, \lambda'_{s+q})_c, \tag{2.11}$$

with

$$\lambda'_m = n - \nu'_{s-m+1}, \quad m = 1, 2, \cdots, s,$$
$$= \mu'_{m-s}, \quad m = s+1, s+2, \cdots, s+q,$$

so that $c = ns - b + a$ and the tableau $[\lambda]_c$ is formed by joining the diagram $[\mu]_a$ to the last column of the tableau $[\nu^*]_{sn-b}$. The resulting tableau $[\lambda]_c$ is always regular since $p + r \leq n$.

Diagrammatically, $[\lambda]_c$ takes the typical form



The IR of L_n which is adjoint to that specified by $[\nu; \mu]^b_a$ is specified by $[\mu; \nu]^a_b$.

3. THE DIMENSIONS OF THE IRREDUCIBLE REPRESENTATIONS OF L_n

As pointed out in Sec. 2, the set of tensors (2.1) forms the basis of IR's of each of the groups L_n if the appropriate restrictions are placed on A. Since these

[8] The notation adopted here contrasts both with Ref. 15, in which the tableau $[\nu; \mu]^b_a$ is specified by the Schur function $\{\mu; \nu\}$, and with Ref. 1, in which the tableau $[\nu; \mu]^b_a$ has signature $(\mu_1, \mu_2, \cdots, \mu_p, 0, 0, \cdots, 0, -\nu_r, \cdots, -\nu_2, -\nu_1)$, and is given by $T(\lambda_1, \lambda_2, \cdots, 0)$ with $[\lambda]$ defined by (2.11).

[9] The very useful back-to-back diagrammatic notation was suggested to me by J. Abramsky.

restrictions do not affect the number of independent components of the tensors, all these IR's are of the same dimension. Probably the simplest formula for the dimension $D_n(\mu)$ of the IR of L_n specified by $[\mu]_a$ is given, in the notation of (2.2), by[4,10]

$$D_n(\mu) = \prod_{i=1}^{p} \prod_{j=1}^{q} \frac{n-i+j}{1+\mu_i+\mu'_j-i-j}. \quad (3.1)$$

Thus, for example,

$$D_n(4,2,1) = \frac{n}{6}\frac{n+1}{4}\frac{n+2}{2}\frac{n+3}{1} \times \frac{n-1}{3}\frac{n}{1} \times \frac{n-2}{1}.$$

Since mutually contragredient IR's of L_n are of the same dimension, it follows that the dimension of the IR of L_n specified by $[\nu]^b$ is $D_n(\nu)$, which is given in the notation of (2.6) by

$$D_n(\nu) = \prod_{k=1}^{r} \prod_{l=1}^{s} \frac{n-k+l}{1+\nu_k+\nu'_l-k-l}. \quad (3.2)$$

Furthermore, the adjoint and complement of an IR of an SL_n are equivalent and are therefore of the same dimension. It then follows that any IR of L_n and its complement are of the same dimension. Thus, in the notation of (2.5),

$$D_n(\mu^*) = D_n(\mu). \quad (3.3)$$

Similarly, since the IR's of SL_n, specified by $[\nu;\mu]_a^b$ as defined in (2.10) and by $[\lambda]_c$ as defined in (2.11), are equivalent, it follows that the IR's of L_n specified by these two distinct tableaux are of the same dimension, i.e.,

$$D_n(\nu;\mu) = D_n(\lambda). \quad (3.4)$$

The application of (3.1) to (3.4), together with the use of (3.3), then gives the formulas

$$D_n(\nu;\mu) = D_n(\mu)D_{n-p}(\nu) \times \prod_{i=1}^{p}\prod_{l=1}^{s} \frac{n+\mu_i-i+l}{n+\mu_i-\nu'_l-i+l},$$

$$D_n(\nu,\mu) = D_{n-r}(\mu)D_n(\nu) \times \prod_{j=1}^{q}\prod_{k=1}^{r} \frac{n+\nu_k-k+j}{n+\nu_k-\mu'_j-k+j}. \quad (3.5)$$

Using (3.5), it has been proved by Jahn and El Samra[11] that

$$D_n(\nu;\mu) = \prod_{i=1}^{p}\prod_{j=1}^{q}\prod_{k=1}^{r}\prod_{l=1}^{s} \frac{(n+1+\mu_i+\nu_j-i-j)(n-1-\mu'_k-\nu'_l+k+l)}{(1+\mu_i+\mu'_j-i-j)(1+\nu_k+\nu'_l-k-l)}, \quad (3.6)$$

with $\nu_j = 0$ for $j > r$ and $\mu'_k = 0$ for $k > q$. Thus, for example,

$$D_n(4,3,1;2^2,1) = \frac{n}{1}\frac{n-2}{3}\frac{n-3}{4}\frac{n-5}{6}\frac{n+5}{4}\frac{n+3}{2} \times \frac{n}{1}\frac{n-1}{2}\frac{n-3}{4}\frac{n+4}{3}\frac{n+2}{1} \times \frac{n}{1}\frac{n+2}{1}$$

so that $D_8(4,3,1;2^2,1) = 2\,0, 0\,2\,0, 0\,0\,0$.

4. REDUCTION OF THE OUTER PRODUCT OF IRREDUCIBLE REPRESENTATIONS OF L_n

The reduction of the outer product of two IR's of L_n, corresponding to $[\mu]_a$ and $[\nu]_b$, into a sum of IR's of L_n, corresponding to $[\lambda]_{a+b}$, is determined by the reduction theorem for the decomposition of the outer products of the corresponding representations of the symmetric group, namely

$$[\nu]_b \cdot [\mu]_a = \sum_{\lambda} m_{\nu\mu,\lambda}[\lambda]_{a+b}. \quad (4.1)$$

The procedure P for determining all $[\lambda]_{a+b}$ contained in this decomposition and their multiplicities $m_{\nu\mu,\lambda}$ given by Littlewood and Richardson[12] and proved by Robinson[13] is well known. This procedure is given by P: The tableau $[\nu]_b$ is written down and μ_1 letters a, μ_2 letters b, μ_3 letters c, etc., are added to it one by one, alphabetically, in all possible ways such that, at each stage in this procedure,

(i) the resulting figure corresponds to a regular Young tableau if each letter is interpreted as a box,

(ii) the resulting figure contains no two identical letters in the same column, and

(iii) the series of letters, obtained by reading from right to left along each row taken, in turn, from top to bottom of the resulting figure, is a lattice permutation of the added letters, i.e., if the series is terminated at any point and the truncated series contains ρ_1, ρ_2, ρ_3, $\cdots$ letters $a, b, c, \cdots$, then $\rho_1 \geq \rho_2 \geq \rho_3 \geq \cdots$ for all possible points of termination.

The multiplicity $m_{\nu\mu,\lambda}$ is just the number of distinctly labeled figures corresponding to the tableau $[\lambda]_{a+b}$ obtained by this procedure.

[10] J. S. Frame, G. De B. Robinson, and R. M. Thrall, C. J. Math. 6, 316 (1954).

[11] H. A. Jahn and N. El Samra (unpublished); N. El Samra, thesis, The University, Southampton, England, 1969.

[12] D. E. Littlewood and A. R. Richardson, Phil. Trans. Roy. Soc. London A233, 99 (1934).

[13] G. De B. Robinson, Am. J. Math. 60, 745 (1938).

The corresponding reduction theorem for L_n is obtained from (4.1) by retaining only those terms in the sum for which the number of rows λ_1' of $[\lambda]_{a+b}$ is such that

$$\lambda_1' \leq n. \tag{4.2}$$

Similarly,

$$[\nu]^b \cdot [\mu]^a = \sum_\lambda m^{\nu\mu,\lambda}[\lambda]^{a+b}, \tag{4.3}$$

with $m^{\nu\mu,\lambda} = m_{\nu\mu,\lambda}$.

The procedure P may be used to carry out the reduction

$$[\nu^*]_{sn-b} \cdot [\mu]_a = \sum_\sigma m_{\nu^*\mu,\sigma}[\sigma]_{sn-b+a}, \tag{4.4}$$

and then using the equivalence of the IR's of SL_n specified by (2.5) and (2.7) and the equivalence of the IR's of SL_n specified by (2.10) and (2.11), it is straightforward to carry out the related reduction

$$[\nu]^b \cdot [\mu]_a = \sum_{\lambda\rho} m_{\mu,\lambda}^{\nu,\rho}[\rho;\lambda]_{a-t}^{b-t}. \tag{4.5}$$

For each value of t between zero and the minimum of a and b, the procedure Q for determining all $[\rho;\lambda]_{a-t}^{b-t}$ contained in this decomposition and their multiplicities $m_{\mu,\lambda}^{\nu,\rho}$ is given by Q: The tableaux $[\nu]^b$ and $[\mu]_a$ are written down back-to-back with the dotted boxes of $[\nu]^b$ replaced by dots and the undotted boxes of $[\mu]_a$ replaced by $\mu_1, \mu_2, \mu_3, \cdots$ letters $a, b, c, \cdots$ in the first, second, third, $\cdots$ rows, respectively. An overbar is set above all possible sets of t letters of this figure, and all the t-bar letters are then superposed one by one, alphabetically, on the dots of the figure in all possible ways such that, at each stage in this procedure, we have the following:

(i) the resulting figure corresponds to a regular composite tableau if each unbarred letter, each dot, each barred letter, and each superposed letter is interpreted as an undotted box, a dotted box, an undotted box which has been removed, and a dotted box which has been removed, respectively;

(ii) the resulting figure contains no two identical superposed letters in the same column;

(iii) the series of letters, obtained from the resulting figure by reading the unbarred letters from right to left along each row taken in turn from top to bottom and then reading the superposed letters from right to left along each row taken, in turn, from *bottom to top*, is a lattice permutation of those letters.

The multiplicity $m_{\mu,\lambda}^{\nu,\rho}$ is just the number of distinctly labeled figures corresponding to the tableau $[\rho;\lambda]_{a-t}^{b-t}$ obtained by this procedure.

The corresponding reduction theorem for L_n is obtained from (4.5) by retaining only those terms in the sum for which the number of rows, λ_1', of $[\lambda]_{a-t}$, the number of rows, ρ_1', of $[\rho]^{b-t}$, and the number of the row, i_1', of $[\mu]_a$ from which came the uppermost letter superposed on a dot in the first column of $[\nu]^b$, counting from the right, are such that

$$\lambda_1' + \rho_1' \leq n \tag{4.6}$$

and

$$i_1' + \rho_1' \leq n. \tag{4.7}$$

These restrictions arise in the derivation of Q from P based on the equivalence of (4.4) and (4.5) as shown in the Appendix. An example of the application of this procedure Q and the restrictions (4.6) and (4.7) is provided by the reduction of $[4, 3, 1]^8 \cdot [2^2, 1]_5$. The resulting figures, obtained by using Q for all possible values of t, together with the corresponding minimum value n_0 of n obtained by using (4.6) and (4.7), are given by

$t = 0$, $n_0 = 6$,

. . . . a a
. . . b b ,
. c

$t = 1$, $n_0 = 6$,

. . . . a a b . . . a a
b . . b $\bar{b}$ . . . b $\bar{b}$,
. c . c

$n_0 = 5$,

. . . . a a . . . . a a
. . . b $\bar{b}$ . . . b b ,
b c c $\bar{c}$

. . . . a a c . . . a a
c . . b b . . . b b ,
. $\bar{c}$. $\bar{c}$

$t = 2$, $n_0 = 6$,

b . . a $\bar{a}$
a . . b $\bar{b}$,
. c

$n_0 = 5$,

. . . . a $\bar{a}$ b . . . a $\bar{a}$ . . . . a a
b . . b $\bar{b}$ . . . b $\bar{b}$ b . . b $\bar{b}$,
a c a c c $\bar{c}$

. . . . a a c . . . a a b . . . a a
b c . b $\bar{b}$ b . . b $\bar{b}$ c . . b $\bar{b}$,
. $\bar{c}$. $\bar{c}$. $\bar{c}$

b . . . a a
. . . b $\bar{b}$,
c $\bar{c}$

$n_0 = 4$,

. . . . a a c . . . a a
c . . b $\bar{b}$ . . . b $\bar{b}$,
b $\bar{c}$ b $\bar{c}$

$t = 3$, $n_0 = 5$,

b . . . a $\bar{a}$ b . . . a $\bar{a}$ b c . . a $\bar{a}$
a . . b $\bar{b}$ a c . b $\bar{b}$ a . . b $\bar{b}$,
c $\bar{c}$. $\bar{c}$. $\bar{c}$

$n_0 = 4$, $\begin{array}{lll} \ldots a\bar{a} & c\ldots a\bar{a} & b\ldots a\bar{a} \\ bc.b\bar{b} & b..b\bar{b} & c..b\bar{b}, \\ .\ a\bar{c} & a\bar{c} & a\bar{c} \end{array}$

$\begin{array}{lll} \ldots aa & c\ldots aa & b\ldots aa \\ bc.\bar{b}\bar{b} & b..\bar{b}\bar{b} & c..\bar{b}\bar{b}, \\ b\bar{c} & b\bar{c} & b\bar{c} \end{array}$

$\begin{array}{ll} c\ldots aa & bc..aa \\ bb.\bar{b}\bar{b} & b..\bar{b}\bar{b}, \\ .\bar{c} & .\bar{c} \end{array}$

$t = 4$, $n_0 = 4$, $\begin{array}{lll} b\ldots a\bar{a} & bc..a\bar{a} & bc..a\bar{a} \\ ac.\bar{b}\bar{b} & a..\bar{b}\bar{b} & ab.\bar{b}\bar{b}, \\ b\bar{c} & b\bar{c} & .\bar{c} \end{array}$

$n_0 = 3$, $\begin{array}{ll} c\ldots a\bar{a} & bc..a\bar{a} \\ bb.\bar{b}\bar{b} & b..\bar{b}\bar{b}, \\ a\bar{c} & a\bar{c} \end{array}$

$t = 5$, $n_0 = 3$, $\begin{array}{l} bc..\bar{a}\bar{a} \\ ab.\bar{b}\bar{b}. \\ a\bar{c} \end{array}$

Hence, the corresponding reduction formula is

$$\begin{aligned}
&[4,3,1]^8 \cdot [2^2,1]_5 \\
&= 1_6[4,3,1;2^2,1]^8_5 + 1_5[4,3;2^2]^7_4 + 1_5[4,3;2,1^2]^7_4 \\
&\quad + 1_5[4,2,1;2^2]^7_4 + 1_6[4,2,1;2,1^2]^7_4 \\
&\quad + 1_5[3^2,1;2^2]^7_4 + 1_6[3^2,1;2,1^2]^7_4 \\
&\quad + 2_{5+4}[4,2;2,1]^6_3 + 1_5[4,2;1^3]^6_3 \\
&\quad + 1_5[4,1^2;2,1]^6_3 + 2_{5+4}[3^2;2,1]^6_3 + 1_5[3^2;1^3]^6_3 \\
&\quad + 2_{5+5}[3,2,1;2,1]^6_3 + 1_6[3,2,1;1^3]^6_3 + 1_4[4,1;2]^5_2 \\
&\quad + 1_4[4,1;1^2]^5_2 + 2_{4+4}[3,2;2]^5_2 + 3_{5+4+4}[3,2;1^2]^5_2 \\
&\quad + 1_4[3,1^2;2]^5_2 + 1_5[3,1^2;1^2]^5_2 + 1_4[2^2,1;2]^5_2 \\
&\quad + 1_5[2^2,1;1^2]^5_2 + 2_{4+3}[3,1;1]^4_1 + 2_{4+3}[2^2;1]^4_1 \\
&\quad + 1_4[2,1^2;1]^4_1 + 1_3[2,1;0]^3_0, \qquad (4.8)
\end{aligned}$$

where the subscripts added to the multiplicities $m^{\nu,\rho}_{\mu,\lambda}$ indicate the minimum values n_0 of n, for which each of the $m^{\nu,\rho}_{\mu,\lambda}$ tableau specifies an IR of L_n which should be included in the reduction formula for the outer product of two IR's of L_n. Thus, the reduction formula for L_n with $n \geq 6$ is given by (4.8) with the subscripts removed from the multiplicities. However, the corresponding reduction formula for L_n with, e.g., $n = 4$ is given by

$$\begin{aligned}
&L_4\colon [4,3,1]^8 \cdot [2^2,1]_5 \\
&\quad = [4,2;2,1]^6_3 + [3^2;2,1]^6_3 + [4,1;2]^5_2 \\
&\qquad + [4,1;1^2]^5_2 + 2[3,2;2]^5_2 + 2[3,2;1^2]^5_2 \\
&\qquad + [3,1^2;2]^5_2 + [2^2,1;2]^5_2 + 2[3,1;1]^4_1 \\
&\qquad + 2[2^2;1]^4_1 + [2,1^2;1]^4_1 + [2,1;0]^3_0. \qquad (4.9)
\end{aligned}$$

The reduction formula (4.5) specifies the reduction in terms of an irreducible set of mixed tensors of the outer product of irreducible sets of covariant and contravariant tensors. Thus the tensors associated with $[\lambda;\rho]^{b-t}_{a-t}$ are traceless tensors constructed from the tensors associated with $[\mu]_a$ and $[\nu]^b$ by taking a trace over a set of t indices. It is therefore instructive to consider the direct product subgroups $\Sigma_{a-t} \odot \Sigma_t$ and $\Sigma_{b-t} \odot \Sigma_t$ of Σ_a and Σ_b, respectively. The formula for the decomposition of the IR of Σ_a, specified by the tableau $[\mu]_a$ into a sum of IR's of $\Sigma_{a-t} \odot \Sigma_t$, is given by

$$[\mu]_a \to \sum_\lambda ([\lambda]_{a-t}, [\mu/\lambda]_t), \qquad (4.10)$$

where $[\mu/\lambda]_t$ denotes[14] the tableau which is obtained from the regular tableau $[\mu]_a$ by the removal of the regular tableau $[\lambda]_{a-t}$. Such a tableau $[\mu/\lambda]_t$ is, in general, a skew tableau corresponding to a reducible representation of Σ_t. The reduction into IR's of Σ_t is specified by

$$[\mu/\lambda]_t = \sum_\sigma m_{\lambda\sigma,\mu}[\sigma]_t. \qquad (4.11)$$

Hence,

$$[\mu]_a \to \sum_\sigma m_{\lambda\sigma,\mu}([\lambda]_{a-t}, [\sigma]_t). \qquad (4.12)$$

Similarly, the formula for the decomposition of the IR of Σ_b, specified by the tableau $[\nu]^b$ into a sum of IR's of $\Sigma_{b-t} \odot \Sigma_t$, is given by

$$[\nu]^b \to \sum_{\rho\tau} m^{\rho\tau,\nu}([\rho]^{b-t}, [\tau]^t). \qquad (4.13)$$

If sets of t indices of the tensors associated with $[\mu]_a$ and $[\nu]^b$ have their symmetry specified by $[\sigma]_t$ and $[\tau]^t$, respectively, it is only possible to take a trace over these sets if $[\sigma]_t$ and $[\tau]^t$ specify the same IR of Σ_t. This corresponds to the obvious statement in the notation of (4.5) that

$$m^{\tau,0}_{\sigma 0,} = \delta^\tau_\sigma, \qquad (4.14)$$

where 0 refers to an empty tableau corresponding to a scalar which may be denoted by $[0]_0$, $[0]^0$, or even $[0;0]^0_0$. Such a tableau is conventionally represented diagrammatically by 1. It then follows from (4.12) and (4.13) that

$$[\nu]^b \cdot [\mu]_a = \sum_{\lambda\rho\sigma} m_{\lambda\sigma,\mu} m^{\rho\sigma,\nu}[\rho;\lambda]^{b-t}_{a-t}, \qquad (4.15)$$

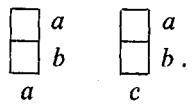
and, therefore, in (4.5)

$$m^{\nu,\rho}_{\mu,\lambda} = \sum_\sigma m_{\lambda\sigma,\mu} m^{\rho\sigma,\nu}. \qquad (4.16)$$

The evaluation of $m^{\nu,\rho}_{\mu,\lambda}$ may thus be carried out by using the relationship (4.16) together with the procedure P, rather than by using the procedure Q. For example, the coefficient of $[3,2;1^2]^5_2$ in (4.8) may be evaluated by reducing the skew tableaux $[2^2, 1/1^2]_3$ and $[4,3,1/3,2]^3$. The procedure P may be used to generate the tableau $[2^2,1]_5$ from the tableau $[1^2]_2$ in

[14] The notation used here is that of Ref. 3 which contrasts with that of Ref. 4, in which $[\mu/\lambda]$ is denoted by $[\mu] - [\lambda]$.

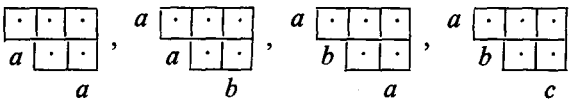
just two ways. The corresponding resulting figures are



Thus,

$$[2^2, 1/1^2]_3 = [2, 1]_3 + [1^3]_3. \qquad (4.17)$$

Similarly, the tableau $[4, 3, 1]^8$ may be generated from the tableau $[3, 2]^5$ in four ways using the procedure P. The corresponding resulting figures are



Thus,

$$[4, 3, 1/3, 2]^3 = [3]^3 + 2[2, 1]^3 + [1^3]^3. \qquad (4.18)$$

A comparison of (4.17) and (4.18) indicates that

$$[4, 3, 1]^8 \cdot [2^2, 1]_5 = \cdots + 3[3, 2; 1^2]^5_2 + \cdots. \qquad (4.19)$$

The multipicity of three occurring in (4.19), derived from (4.15) by the use of the procedure P, is of course the same as that occurring in association with the tableau $[3, 2; 1^2]^5_2$ in (4.8), derived from (4.5) by the use of the procedure Q. However, it is assumed in the derivation of (4.15) that the sets of tensors involved are the bases of IR's of L_n with n sufficiently large for all the tensors to exist. As a result, it follows that the information relevant to applications to IR's of L_n with n small, implicit in the subscripts associated with the multiplicities in reductions like (4.8), is not contained in reductions like (4.19). In fact, it can be seen from (4.6) and (4.7) that (4.5) and thus also (4.15) are applicable without restriction only to IR's of L_n with $n \geq \mu'_1 + \nu'_1$, where μ'_1 and ν'_1 denote the number of rows of $[\mu]_a$ and $[\nu]^b$, respectively.

In the reduction formula (4.5), it follows from the procedure Q that $m^{\nu,\nu}_{\mu,\mu} = 1$ and, therefore,

$$[\nu; \mu]^b_a = [\nu]^b \cdot [\mu]_a - \sum_{\substack{\lambda \neq \mu \\ \rho \neq \nu}} m^{\nu,\rho}_{\mu,\lambda} [\rho; \lambda]^{b-t}_{a-t}, \qquad (4.20)$$

where the summation excludes any term corresponding to $t = 0$. Similarly,

$$[\rho; \lambda]^{b-t}_{a-t} = [\rho]^{b-t} \cdot [\lambda]_{a-t} - \sum_{\substack{\sigma \neq \lambda \\ \tau \neq \rho}} m^{\rho,\tau}_{\lambda,\sigma} [\tau; \sigma]^{b-t-u}_{a-t-u}, \qquad (4.21)$$

where the summation excludes any term corresponding to $u = 0$. It follows that any regular composite tableau may be expressed as a sum of outer products of regular Young tableau, i.e.,

$$[\nu; \mu]^b_a = \sum_{\lambda\rho} n^{\nu,\rho}_{\mu,\lambda} [\rho]^{b-t} \cdot [\lambda]_{a-t}. \qquad (4.22)$$

It is possible to determine all the outer products contained in this expansion and their multiplicities $n^{\nu,\rho}_{\mu,\lambda}$ by using the procedure Q to evaluate all the coefficients in (4.20), (4.21), and all the relevant formulas of the same type. However, it has been shown by Littlewood[15] that

$$[\nu; \mu]^b_a = \sum_{\lambda\rho\sigma} (-1)^t m_{\lambda\sigma,\mu} m^{\rho\sigma',\nu} [\rho]^{b-t} \cdot [\lambda]_{a-t}, \qquad (4.23)$$

where $[\sigma]_t$ and $[\sigma']_t$ are mutually conjugate tableaux. Moreover, $m_{\mu\nu,\lambda} = m_{\mu'\nu',\lambda'}$, so that from (4.23) and (4.5)

$$[\nu; \mu]^b_a = \sum_{\lambda\rho} (-1)^t m^{\nu',\rho'}_{\mu,\lambda} [\rho]^{b-t} \cdot [\lambda]_{a-t}. \qquad (4.24)$$

This expansion formula may thus be evaluated by making use of the procedure Q to carry out the reduction

$$[\nu']^b \cdot [\mu]_a = \sum_{\lambda\rho'} m^{\nu',\rho'}_{\mu,\lambda} [\rho'; \lambda]^{b-t}_{a-t}, \qquad (4.25)$$

from which the coefficients $m^{\nu',\rho'}_{\mu,\lambda}$ may be obtained. If the initial tableau $[\nu; \mu]^b_a$ is admissible and thus specifies an IR of L_n, then the expansion formula (4.24) specifies the expansion of this IR without any restriction like those of (4.6) and (4.7). As an example, it follows from (4.8) that

$$\begin{aligned}
&[3, 2^2, 1; 2^2, 1]^8_5 \\
&\quad = [3, 2^2, 1]^8 \cdot [2^2, 1]_5 - [3, 2^2]^7 \cdot [2^2]_4 \\
&\qquad - [3, 2^2]^7 \cdot [2, 1^2]_4 - [3, 2, 1^2]^7 \cdot [2^2]_4 \\
&\qquad - [3, 2, 1^2]^7 \cdot [2, 1^2]_4 - [2^3, 1]^7 \cdot [2^2]_4 \\
&\qquad - [2^3, 1]^7 \cdot [2, 1^2]_4 + 2[3, 2, 1]^6 \cdot [2, 1]_3 \\
&\qquad + [3, 2, 1]^6 \cdot [1^3]_3 + [3, 1^3]^6 \cdot [2, 1]_3 \\
&\qquad + 2[2^3]^6 \cdot [2, 1]_3 + [2^3]^6 \cdot [1^3]_3 \\
&\qquad + 2[2^2, 1^2]^6 \cdot [2, 1]_3 + [2^2, 1^2]^6 \cdot [1^3]_3 \\
&\qquad - [3, 2]^5 \cdot [2]_2 - [3, 2]^5 \cdot [1^2]_2 \\
&\qquad - [3, 1^2]^5 \cdot [2]_2 - [3, 1^2]^5 \cdot [1^2]_2 \\
&\qquad - 2[2^2, 1]^5 \cdot [2]_2 - 3[2^2, 1]^5 \cdot [1^2]_2 \\
&\qquad - [2, 1^3]^5 \cdot [2]_2 - [2, 1^3]^5 \cdot [1^2]_2 \\
&\qquad + [3, 1]^4 \cdot [1]_1 + 2[2^2]^4 \cdot [1]_1 \\
&\qquad + 2[2, 1^2]^4 \cdot [1]_1 - [2, 1]^3 \cdot [0]_0. \qquad (4.26)
\end{aligned}$$

Clearly, $[3, 2^2, 1; 2^2, 1]^8_5$ is only an admissible tableau corresponding to an IR of L_n if $n \geq 7$, and for these values of n the expansion (4.26) is valid.

The reduction of an outer product of two IR's of L_n specified by admissible, regular, composite tableaux may be carried out, firstly, by expanding these tableaux as two outer products of dotted and undotted tableaux using (4.24); secondly, by reducing the outer

[15] D. E. Littlewood, Phil. Trans. Roy. Soc. London **A235**, 387 (1943).

products of the pairs of dotted tableaux and the pairs of undotted tableaux using (4.3) and (4.1); and finally, by reducing the outer products of the resulting tableaux using (4.5). Thus,

$$[\nu;\mu]^b_a \cdot [\rho;\lambda]^d_c = \sum_{\sigma,\tau} m^{\nu\rho,\tau}_{\mu\lambda,\sigma}[\tau;\sigma]^f_e, \qquad (4.27)$$

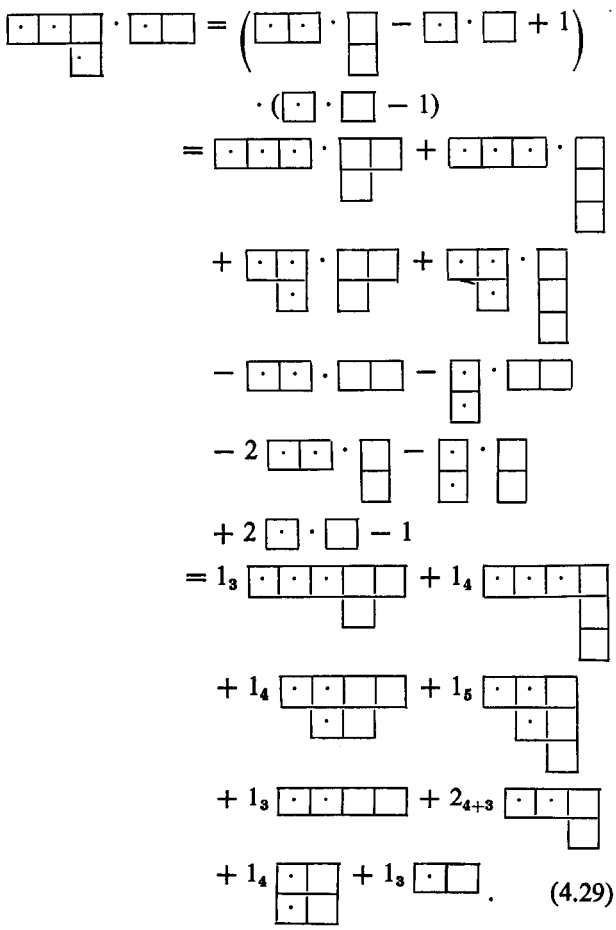
with

$$m^{\nu\rho,\tau}_{\mu\lambda,\sigma} = \sum (-1)^t m^{\nu',\beta'}_{\mu,\alpha}(-1)^u m^{\rho',\delta'}_{\lambda,\gamma} m_{\alpha\gamma,\theta} m^{\beta\delta,\phi} m^{\phi,\tau}_{\theta,\sigma}, \qquad (4.28)$$

where the summation is over all the tableaux $[\alpha]_{a-t}$, $[\beta]^{b-t}$, $[\gamma]_{c-u}$, $[\delta]^{d-u}$, $[\theta]_{a+c-t-u}$, and $[\phi]^{b+d-t-u}$, so that $e = a + c - t - u - v$ and $f = b + d - t - u - v$.

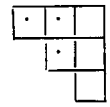
It should be noted that, in the evaluation of the coefficients $m^{\nu\rho,\tau}_{\mu\lambda,\sigma}$ appropriate to the decomposition of an outer product of IR's of L_n specified by admissible composite tableaux, the only restrictions for small values of n occur in the last stage of the process involving the evaluation of $m^{\phi,\tau}_{\theta,\sigma}$.

For example,

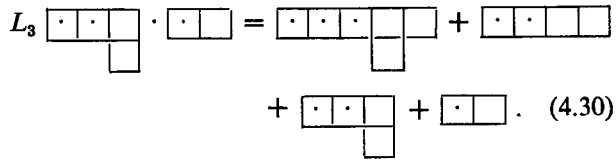


(4.29)

From (4.6) and (4.7), it follows that this expansion formula is applicable to the IR's of all L_n with $n \geq 5$, but that, for $n = 4$, the tableau



is excluded while, for $n = 3$, the expansion is



(4.30)

Of course, for $n = 2$, no such expansion exists.

5. DECOMPOSITION OF THE IRREDUCIBLE REPRESENTATIONS OF L_{m+n} WITH RESPECT TO $L_m \odot L_n$

An IR of L_{m+n} may be decomposed into IR's of the subgroup $L_m \odot L_n$ and the associated sets of tensors with indices α_i, such that $\alpha_i = 1, 2, \cdots, m+n$ are correspondingly decomposed into sets of tensors of the same rank but with each index α_i such that $\alpha_i = 1, 2, \cdots, m$ or $\alpha_i = m+1, m+2, \cdots, m+n$. For the IR of L_{m+n} specified by $[\lambda]_c$, it follows that this decomposition is determined by the decomposition of the corresponding IR of Σ_c into IR's of all the subgroups $\Sigma_a \odot \Sigma_b$ with $a + b = c$. Thus, from (4.12), the relevant decomposition formula is[16,17]

$$[\lambda]_c \to \sum_{\mu\nu} m_{\mu\nu,\lambda}([\mu]_a, [\nu]_b). \qquad (5.1)$$

Similarly, of course,

$$[\lambda]^c \to \sum_{\mu\nu} m^{\mu\nu,\lambda}([\mu]^a, [\nu]^b). \qquad (5.2)$$

The corresponding decomposition formula for the composite tableau $[\tau;\sigma]^f_e$ is established by expanding, using (4.24), then decomposing, using (5.1) and (5.2), and finally reducing, using (4.5). Hence,

$$[\tau;\sigma]^f_e \to \sum_{\mu\nu\lambda\rho} p^{\tau,\nu\rho}_{\sigma,\mu\lambda}([\nu;\mu]^{b-t}_{a-t}, [\rho;\lambda]^{d-u}_{c-u}), \qquad (5.3)$$

with

$$p^{\tau,\nu\rho}_{\sigma,\mu\lambda} = \sum (-1)^v m^{\tau',\phi'}_{\sigma,\theta} m_{\alpha\gamma,\theta} m^{\beta\delta,\phi} m^{\beta,\nu}_{\alpha,\mu} m^{\delta,\rho}_{\gamma,\lambda}, \qquad (5.4)$$

where the summation is over all the tableau $[\alpha]_a$, $[\beta]^b$, $[\gamma]_c$, $[\delta]^d$, $[\theta]_{e-v}$, and $[\phi]^{f-v}$, so that $e - v = a + c$ and $f - v = b + d$.

Formulas (5.1), (5.2), and (5.3) specify the decompositions of IR's of L_{m+n} into IR's of the subgroup $L_m \odot L_n$, provided that the restrictions (4.2), (4.6),

[16] M. L. Whippman, J. Math. Phys. **6**, 1534 (1965).
[17] C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).

and (4.7) are used to exclude certain terms from the summations involved.

The outer-product multiplication rule (4.1) is both commutative and associative so that

$$m_{\mu\nu,\lambda} = m_{\nu\mu,\lambda} \tag{5.5}$$

and

$$\sum_{\sigma} m_{\mu\nu,\sigma} m_{\lambda\sigma,\rho} = \sum_{\sigma} m_{\mu\sigma,\rho} m_{\nu\lambda,\sigma}. \tag{5.6}$$

Furthermore, from (4.5) and (4.24) it follows that

$$\sum_{\lambda\rho} (-1)^t m^{\nu',\rho'}_{\mu,\lambda} m^{\rho,\tau}_{\lambda,\sigma} = \delta_{\mu\sigma}\delta^{\nu\tau}. \tag{5.7}$$

Using (5.5), (5.6), and (5.7), together with (4.16), we straightforwardly derive from (5.4) the very much simpler result conjectured by Abramsky[18]:

$$[\tau;\sigma]^f_c \to \sum m^{\tau,\phi}_{\sigma,\theta} m_{\mu\lambda,\theta} m^{\nu\rho,\phi} ([\nu;\mu]^b_a, [\rho,\lambda]^d_c), \tag{5.8}$$

where the summation is over all the tableaux $[\theta]_{e-v}$ and $[\phi]^{f-v}$, so that $e - v = a + c$ and $f - v = b + d$. It should be stressed that the formula (5.8) may not be applicable to the decomposition of the IR's of L_{m+n} if m or n is small, since the derivation of (5.8) depends upon (4.15). However, the reduction theorem corresponding to (4.15) for IR's of L_n is obtained from (4.15) by retaining in the sum all the terms for which $n \geq \rho'_1 + \lambda'_1$, provided $n \geq \mu'_1 + \nu'_1 - 1$. It follows that (5.8) is applicable to the decomposition of IR's of L_{m+n} if the terms included in the sum are only those for which

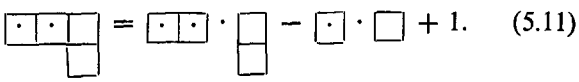
$$m \geq \mu'_1 + \nu'_1 \quad \text{and} \quad n \geq \lambda'_1 + \rho'_1, \tag{5.9}$$

provided that

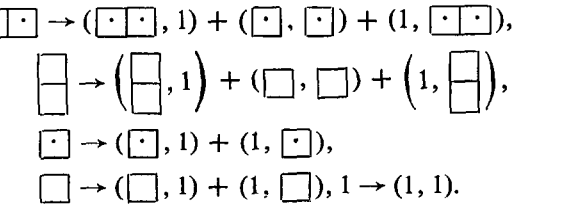
$$m \geq \sigma'_1 + \tau'_1 - 1 \quad \text{and} \quad n \geq \sigma'_1 + \tau'_1 - 1, \tag{5.10}$$

where μ'_1, ν'_1, λ'_1, ρ'_1, σ'_1,and τ'_1 denote the number of rows of the tableaux $[\mu]_a$, $[\nu]^b$, $[\lambda]_e$, $[\rho]^f$, $[\sigma]_e$, and $[\tau]^f$.

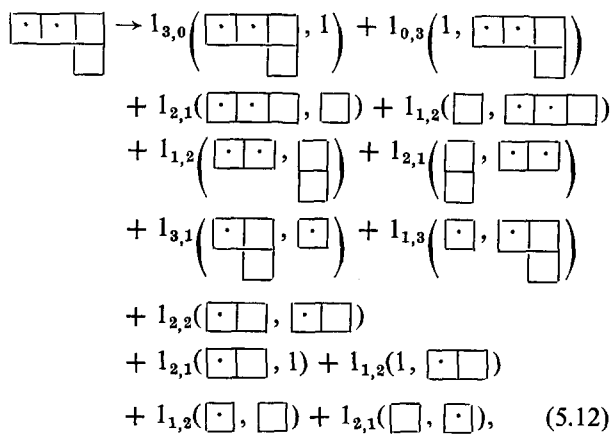
An example of the application of (5.3) is provided by the decomposition of

 (5.11)

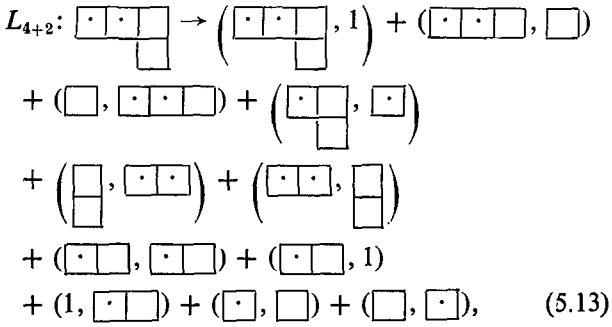
From (5.1),

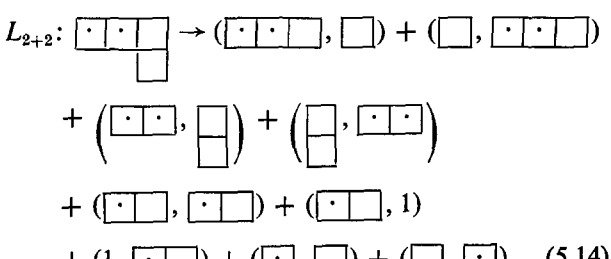


Therefore, using (4.5), we have

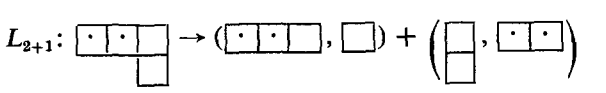
 (5.12)

where the subscripts on the values of the coefficients $p^{\tau,\nu\rho}_{\sigma,\mu\lambda}$ indicate the minimum values of m and n for which the corresponding pair of tableaux specify an IR of $L_m \odot L_n$, which should be included in the decomposition of the IR of L_{m+n}. Thus, the corresponding decomposition formulas for various values of m and n are, e.g., given by the following:

L_{4+2}:  (5.13)

L_{2+2}:  (5.14)

and

L_{2+1}:  (5.15)

Results (5.13) and (5.14) may be obtained more quickly by the use of (5.8), since in these cases conditions (5.10) are satisfied. In fact, (5.8) yields (5.12), without, of course, giving the values of the subscripts, and the application of restrictions (5.9)–(5.12) gives

[18] J. Abramsky, Ph.D. thesis, The University, Southampton, England, 1969.

(5.13) and (5.14) for the appropriate values of m and n. However, (5.15) cannot be obtained by the use of (5.8) since conditions (5.10) are violated.

All these results may be readily checked by the use of the formulas for the dimensions of the IR's of L_n given in Sec. 3.

6. DECOMPOSITION OF THE IRREDUCIBLE REPRESENTATIONS OF L_{mn} WITH RESPECT TO $L_m \otimes L_n$

An IR of L_{mn} may be decomposed into IR's of the subgroup $L_m \otimes L_n$. The associated tensors, with indices α_i such that $\alpha_i = 1, 2, \cdots, mn$, are correspondingly decomposed into tensors of twice the rank obtained by replacing α_i by a pair of indices $\gamma_i\delta_i$ such that $\gamma_i = 1, 2, \cdots, m$ and $\delta_i = 1, 2, \cdots, n$. For the IR of L_{mn} specified by $[\lambda]_a$, it follows that this decomposition is determined by the reduction formula for the inner products of IR's of Σ_a. This reduction formula is specified by[16,17]

$$[\mu]_a \times [\nu]_a = \sum_{\lambda} k_{\mu\nu,\lambda}[\lambda]_a, \tag{6.1}$$

where the procedure K for obtaining all $[\lambda]_a$ contained in this reduction and their multiplicities $k_{\mu\nu,\lambda}$ is well known.[19] The associated decomposition formula for the IR's of L_{mn} is specified by

$$[\lambda]_a \Rightarrow \sum_{\mu\nu} k_{\mu\nu,\lambda}([\mu]_a; [\nu]_a). \tag{6.2}$$

Similarly,

$$[\lambda]^a \Rightarrow \sum_{\mu\nu} k^{\mu\nu,\lambda}([\mu]^a; [\nu]^a), \tag{6.3}$$

with

$$k^{\mu\nu,\lambda} = k_{\mu\nu,\lambda}.$$

The corresponding decomposition formula for the composite tableau $[\tau; \sigma]^f_e$ is established by expanding, using (4.24), then decomposing, using (6.2) and (6.3), and finally reducing, using (4.5). Hence,

$$[\tau; \sigma]^f_e \Rightarrow \sum_{\mu\nu\lambda\rho} k^{\tau,\nu\rho}_{\sigma,\mu\lambda}([\nu; \mu]^{b-t}_{a-t}; [\rho; \lambda]^{b-u}_{a-u}), \tag{6.4}$$

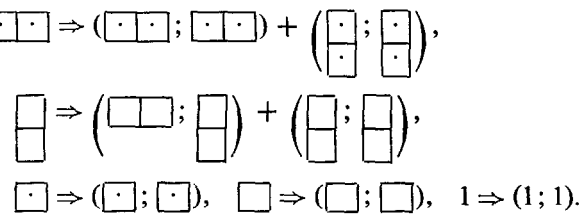
with

$$k^{\tau,\nu\rho}_{\sigma,\mu\lambda} = \sum (-1)^v m^{\tau',\phi'}_{\sigma,\theta} k_{\alpha\gamma,\theta} k^{\beta\delta,\phi}_{\alpha,\mu} m^{\beta,\nu}_{\gamma,\lambda} m^{\delta,\rho}_{\gamma,\lambda}, \tag{6.5}$$

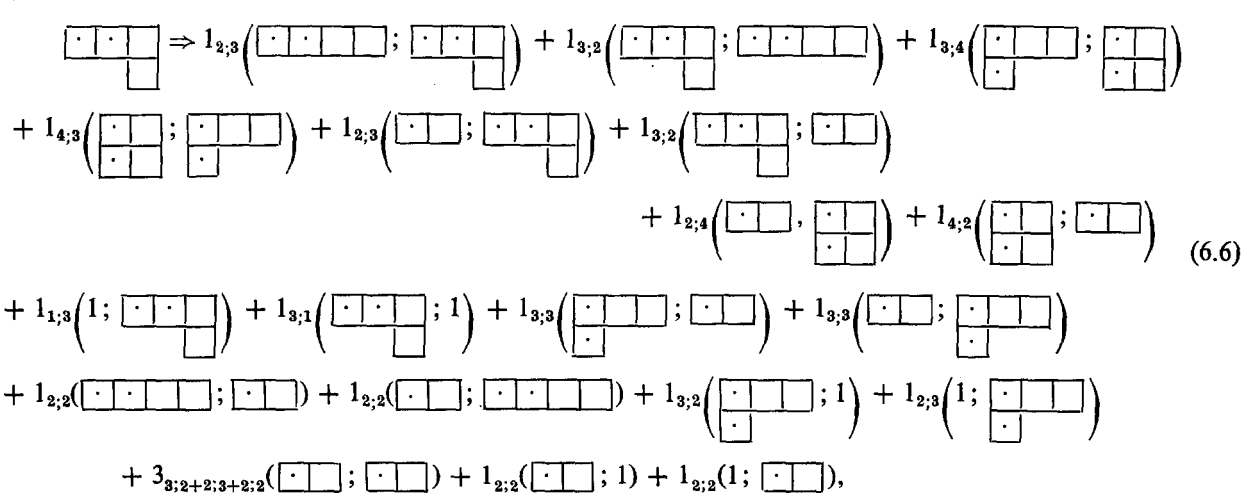
where the summation is over all the tableaux $[\alpha]_a$, $[\beta]^b$, $[\gamma]_a$, $[\delta]^b$, $[\theta]_{e-v}$, and $[\phi]^{f-v}$, so that $a = e - v$ and $b = f - v$.

Formulas (6.2), (6.3), and (6.4) specify the decompositions of IR's of L_{mn} into IR's of $L_m \otimes L_n$ provided that the restrictions (4.2), (4.6), and (4.7) are used to exclude certain terms from the summations involved in (6.4) and (6.5).

The application of (6.4) to the decomposition of (5.11) involves the use of the following decompositions which may be derived from (6.2):



Then, using (4.5), we have

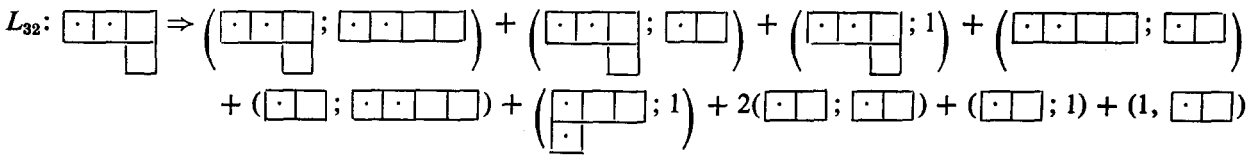


(6.6)

where the subscripts on the values of the coefficients $k^{\tau,\nu\rho}_{\sigma,\mu\lambda}$ indicate the minimum values of m and n for which the corresponding pair of tableaux specify an IR of $L_m \otimes L_n$, which should be included in the decomposition of the IR of L_{mn}. Thus, the corresponding decomposition formulas for various values of m and n are, e.g.,

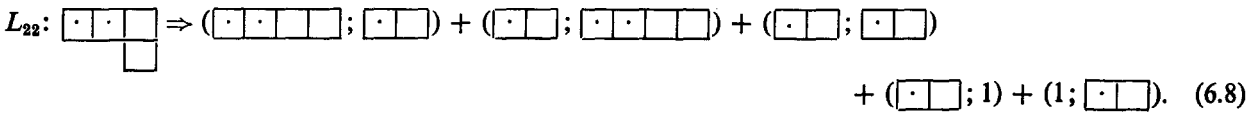
[19] G. De B. Robinson and O. E. Taulbee, Proc. Natl. Acad. Sci. U.S. 40, 723 (1954).

given by



(6.7)

and



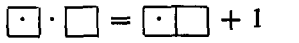
(6.8)

Once again all these results may be checked by making use of the formulas of Sec. 3.

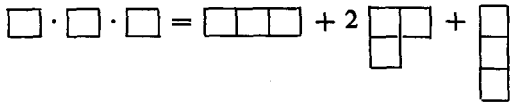
7. DISCUSSION

In the preceding sections of this paper an attempt has been made to establish rules appropriate to the evaluation of some of the properties of the IR's of L_n induced in a space defined by mixed tensors. The key result is the statement of the procedure Q together with the restrictions (4.6) and (4.7). The derivation of this result depends upon the equivalence of certain IR's of SL_n, but the result itself is applicable to the IR's of each of the groups L_n for all values of n. The same is true of other results such as (4.28), (5.4), and (6.5) which depend upon the reduction (4.5) and the expansion (4.24).

This is of importance in the applications of group theory to the study of elementary particles. For example, in quark models based on $SU(3)$, the mesons and baryons of lowest mass are thought to be composed of a quark–antiquark pair and of three quarks, respectively. It is, therefore, natural to specify the corresponding IR's of the symmetry group, whether it be $SU(3)$ itself or a larger group such as $SU(6)$, $U(6)$, or $SU(6, 6)$, by the tableaux occurring in the reductions



and



for mesons and baryons, respectively. It then follows that any composite tableau such as $[\nu; \mu]_a^b$ specifies an IR of the symmetry group corresponding to particles composed of a quarks and b antiquarks having baryon number $B = \frac{1}{3}(a - b)$. Some applications of the composite tableau notation to the consideration of particles composed of one antiquark and four quarks, within the context of a model based on the symmetry group $SU(6, 6)$, have been published elsewhere.[20] It was found that the evaluation of the reductions, expansions, and decompositions involved in this work just cited could be carried out very simply using the results presented in this paper. In fact, the alternative to the use of composite tableaux such as $[\nu; \mu]_a^b$ with $a = 4$ and $b = 1$ is, in this model, the use of conventional tableau such as $[\lambda]_c$ with $c = 15$. Unfortunately, the multiplicities $m_{\mu\nu,\lambda}$ and $k_{\mu\nu,\lambda}$ have only been tabulated for all tableaux such that, in the notation of (4.1), $a + b \leq 10$,[17,21,22] and in the notation of (6.1), $a \leq 8$.[17,23]

Another example of the use of composite tableaux is provided by the decomposition of the IR of $SU(6, 6)$ specified by $[3, 1^9]_{12}$ into IR's of $SU(4) \otimes SU(3)$. This decomposition lies outside the range of the tables referred to, but may readily be evaluated by carrying out the decomposition of the equivalent IR of $SU(6, 6)$ specified by $[2; 1^2]_2^2$. In fact, this decomposition is given by (6, 6) which, by means of equivalence relationships, may be written in the form

$$\begin{aligned}[3, 1^9]_{12} \Rightarrow\ & ([3^2, 2]_8 + [2, 1^2]_4; [4, 2]_6) \\ & + ([4, 2^2]_8 + [2, 1^2]_4 + [0]_0; [3^2]_6) \\ & + ([2^2]_4 + [2, 1^2]_4 + [0]_0; [3]_3) \\ & + ([3^2, 2]_8 + [3, 1]_4 + [2, 1^2]_4; [0]_0) \\ & + ([4, 2^2]_8 + [3^2, 2]_8 + [3, 1]_4 + [2^2]_4 \\ & + 3[2, 1^2]_4 + [0]_0; [2, 1]_3). \end{aligned} \tag{7.1}$$

The decompositions of a large number of the IR's of $SU(6)$ into IR's of $SU(4) \odot SU(2)$ and into IR's of $SU(3) \otimes SU(2)$ have been tabulated by Hagen and Macfarlane.[24,25] The decompositions (5.13) and (6.7)

[20] J. Abramsky and R. C. King, Phys. Rev. Letters 20, 1408 (1968).
[21] F. D. Murnaghan, Am. J. Math. 59, 437 (1947).
[22] F. D. Murnaghan, Am. J. Math. 60, 44 (1938).
[23] F. D. Murnaghan, Am. J. Math. 60, 761 (1938).
[24] C. R. Hagen and A. J. Macfarlane, J. Math. Phys. 6, 1366 (1965).
[25] C. R. Hagen and A. J. Macfarlane, J. Math. Phys. 6, 1355 (1965).

correspond to the decompositions of the IR of $SU(6)$ specified by $[3, 1^3]_6$ of dimension **280**. However, the slightly more general results (5.12) and (6.6) are expressed in a form which is eminently suitable for application to quark models in which the symmetry group is extended, e.g., from $SU(6)$ to $SU(6, 6)$. This is made clear by the derivation of (7.1). Moreover, the composite tableaux notation immediately yields the specification of the hypercharge Y associated with each IR of $SU(4) \odot SU(2)$ occurring in (5.13). In fact, in the notation of (5.3),

$$Y = \tfrac{1}{3}(a - b) - \tfrac{2}{3}(c - d). \tag{7.2}$$

This result (7.2) also applies to the decomposition of the IR's of $SU(3)$ into IR's of $SU(2) \odot U(1)$ when the IR's of $SU(2)$ correspond to isotopic spin multiplets.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor H. A. Jahn for his continual encouragement and for the communication of his dimensionality formula, to Dr. I. J. Ketley for a number of helpful discussions, and to J. Abramsky whose applications of the author's preliminary ideas on composite tableaux to the study of the symmetries of elementary particles proved invaluable.

APPENDIX

In order to derive the procedure Q from the procedure P, it is useful to define a third procedure R which may be used to carry out the reduction

$$[\nu]_b \cdot [\mu]_a = \sum_{\rho\lambda} m_{\nu\mu,\rho\lambda}[\rho]_{b+t}[\lambda]_{a-t}, \tag{A1}$$

where $[\rho]_{b+t}[\lambda]_{a-t}$ denotes a tableau, which in general is irregular, formed by the adjunction of two regular tableaux $[\rho]_{b+t}$ and $[\lambda]_{a-t}$. For each value of t between zero and a, the procedure R for determining all $[\rho]_{b+t}[\lambda]_{a-t}$, contained in this reduction, and their multiplicities $m_{\nu\mu,\rho\lambda}$, is given by R: The tableau $[\nu]_b$ is written down and $\mu_1, \mu_2, \mu_3, \cdots$ letters $a, b, c, \cdots$ are added to the first, second, third, $\cdots$ rows, respectively, in such a way that these letters form the regular tableau $[\mu]_a$ with the first letter of each row, reading from the left, appearing in the column immediately to the right of the last column of $[\nu]_b$, again counting from the left. An overbar is put above all possible sets of t letters of this figure, and all the t-barred letters are then added to the columns of the tableau $[\nu]_b$, one by one, alphabetically in all possible ways such that, at each stage in this procedure,

(i) the resulting figure corresponds to the adjunction of two regular tableaux if each unbarred and each added letter is interpreted as a box and if each barred letter is disregarded,

(ii) the resulting figure contains no two identical added letters in the same column, and

(iii) the series of letters, obtained from the resulting figure by reading the unbarred letters from right to left along each row taken, in turn, from top to bottom and then reading the added letters also from right to left along each row taken, in turn, from top to bottom, is a lattice permutation of those letters.

The multiplicity $m_{\nu\mu,\rho\lambda}$ is just the number of distinctly labeled figures corresponding to the adjunction of $[\rho]_{b+t}$ and $[\lambda]_{a-t}$ obtained by this procedure.

The equivalence of the application of P to (4.1) and the application of R to (A1), together with any restrictions which may be necessary, depends upon the interpretation of the adjunction of $[\rho]_{b+t}$ and $[\lambda]_{a-t}$ as a regular tableau $[\sigma]_{a+b}$. In order for such an interpretation to be possible, it is necessary to match up the symbols in the last column of $[\rho]_{b+t}$ and those in the first column of $[\lambda]_{a-t}$, both counted from the left. In the notation of Sec. 2, μ_1', ν_1', λ_1', and ρ_1' denote the lengths of the first columns of $[\mu]_a$, $[\nu]_b$, $[\lambda]_{a-t}$, and $[\rho]_{b+t}$. In addition, ν_s' and ρ_s' denote the lengths of the last columns of $[\nu]_b$ and $[\rho]_{b+t}$. It is also convenient to use i_1' to denote the number of the row of $[\mu]_a$ from which came the last letter, *counting from the top*, added to $[\nu]_b$ in the last column, counting from the left.

With this notation, if

$$\nu_s' \geq \mu_1', \tag{A2}$$

the matching is trivial since the adjunction $[\rho]_{b+t}[\lambda]_{a-t}$ always corresponds to a regular tableau, and the figures are all such that no added letters appear in any row containing either barred or unbarred letters. Thus, if the barred letters are disregarded, there is a one-to-one correspondence between all the figures obtained by the application of R to (A1) and those obtained by the application of P to (4.1), provided $[\mu]_a$ and $[\nu]_b$ are such that (A2) is satisfied.

For example, the use of R may give rise to the following typical figures:

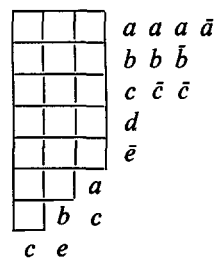


FIG. 1.

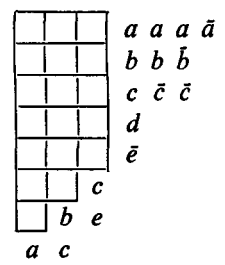


FIG. 2.

In these figures,

	μ_1'	ν_1'	λ_1'	ρ_1'	ν_s'	ρ_s'	i_1'
FIG. 1	5	7	4	8	5	7	3
FIG. 2	5	7	4	8	5	7	5

Clearly, both these figures may be arrived at by the use of P.

Similarly, if

$$\nu_s' \geq \lambda_1', \tag{A3}$$

the adjunction $[\rho]_{b+t}[\lambda]_{a-t}$ always corresponds to a regular tableau, and the figures obtained using R are such that no added letters appear in any row containing unbarred letters. Thus, there is a one-to-one correspondence between the figures satisfying (A3) obtained by the application of R to (A1) and by the application of P to (4.1).

The following are typical figures which satisfy (A3) but not (A2):

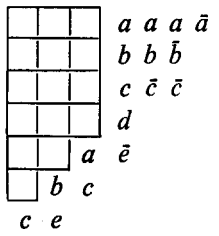


FIG. 3.

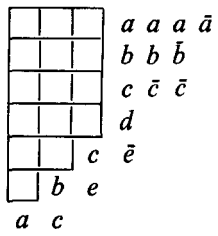


FIG. 4.

In these figures,

	μ_1'	ν_1'	λ_1'	ρ_1'	ν_s'	ρ_s'	i_1'
FIG. 3	5	6	4	7	4	6	3
FIG. 4	5	6	4	7	4	6	5

Once again, both these figures may be obtained by the use of P.

It should be noted that in all the figures satisfying (A3), the only letters added to the mth row, counting from the top, are barred letters from the topmost m rows of $[\mu]_a$. In particular, it follows that

$$i_1' \leq \rho_s'. \tag{A4}$$

This inequality is clearly satisfied by Figs. 1–4.

If the condition (A3) is not satisfied, but rather

$$\nu_s' < \lambda_1' \leq \rho_s', \tag{A5}$$

the adjunction $[\rho]_{b+t}[\lambda]_{a-t}$ always corresponds to a regular tableau, but the figures obtained using R are such that both added and unbarred letters appear in at least one row. The necessary and sufficient condition that in each such row all these letters are in alphabetical order reading from left to right is just the condition (A4). In general, the procedure R gives rise to figures satisfying (A5) but not (A4). However, there is a one-to-one correspondence between the figures satisfying both (A5) and (A4) obtained by the application of R to (A1) and by the application of P to (4.1).

This one-to-one correspondence is valid despite the fact that for these figures the series of letters relevant to (iii) of R and P are, in general, different. In fact, if these series are each terminated at some position in the mth row of $[\rho]_{b+t}$, then the numbers of each of the first m letters of the alphabet in each series are the same, but the series relevant to (iii) of R contains in addition the unbarred letters from $[\lambda]_{a-t}$ which are later in the alphabet than the mth letter. However, since $[\lambda]_{a-t}$ is a regular tableau, it follows that the series relevant to (iii) of R and P are such that either both or neither are lattice permutations.

The following typical figures satisfy (A5):

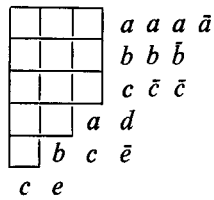


FIG. 5.

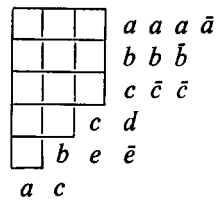


FIG. 6.

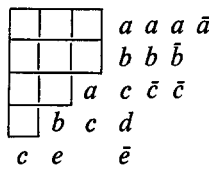


FIG. 7.

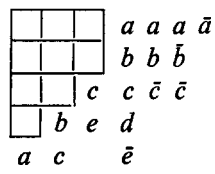


FIG. 8.

In these figures,

	μ_1'	ν_1'	λ_1'	ρ_1'	ν_s'	ρ_s'	i_1'
FIG. 5	5	5	4	6	3	5	3
FIG. 6	5	5	4	6	3	5	5
FIG. 7	5	4	4	5	2	4	3
FIG. 8	5	4	4	5	2	4	5

Figures 5, 6, and 7 may all be arrived at by the use of P even though the series of letters obtained from Fig. 7, relevant to (iii) of P, is not the same as that relevant to (iii) of R. Figure 8, on the other hand, may not be arrived at by the use of P, and it is the only one of these figures which does not satisfy (A4).

Finally, if the condition

$$\lambda_1 \leq \rho_s' \tag{A6}$$

is not satisfied, the adjunction $[\rho]_{b+t}[\lambda]_{a-t}$ never corresponds to a regular tableau. The following are typical

figures which do not satisfy (A6):

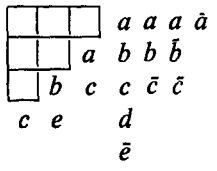


FIG. 9.

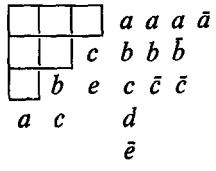


FIG. 10.

In these figures,

	μ_1'	ν_1'	λ_1'	ρ_1'	ν_s'	ρ_s'	i_1'
FIG. 9	5	3	4	4	1	3	3
FIG. 10	5	3	4	4	1	3	5

Neither of these figures may be arrived at by the use of P even though Fig. 9 satisfies (A4).

From these results it follows that if, in the application of R to (A1), only those terms are retained which satisfy both (A4) and (A6), then the final set of figures corresponds exactly to that obtained by the application of P to (4.1).

It is known that the corresponding reduction theorem for IR's of SL_n is obtained from the application of P to (4.1) by retaining only those terms in the sum for which (4.2) is satisfied. This reduction theorem can therefore be obtained from the application of R to (A1) by retaining only those terms in the sum which satisfy (A4), (A6), and the condition

$$\rho_1' \leq n. \tag{A7}$$

Moreover, the definitions of R and Q are such that the application of R to (A1), together with the restriction (A7), is exactly equivalent to the application of Q to

$$[\nu^*]^{sn-b} \cdot [\mu]_a = \sum_{\rho,\lambda} m_{\mu,\lambda}^{\nu^*;\rho^*} [\rho^*; \lambda]_{a-t}^{sn-b-t}, \tag{A8}$$

where $[\nu^*]^{sn-b}$ and $[\rho^*]^{sn-b-t}$ specify IR's of SL_n which are equivalent to those specified by $[\nu]_b$ and $[\rho]_{b+t}$. In addition, if (A6) is satisfied, the composite tableau $[\rho^*; \lambda]_{a-t}^{sn-b-t}$ specifies an IR of SL_n which is equivalent to that specified by the adjunction $[\rho]_{b+t}[\lambda]_{a-t}$ which is, itself, a regular tableau as a consequence of (A6). The corresponding reduction theorem for IR's of SL_n is therefore obtained from the application of Q to (A8) by retaining only those terms satisfying (A4) and (A6).

The following typical figures may be obtained using Q:

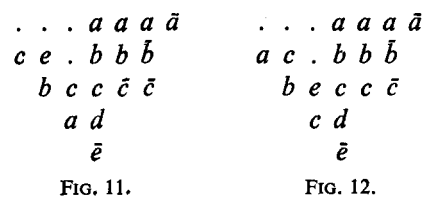


FIG. 11.

FIG. 12.

In these figures,

	μ_1'	ν_1'	λ_1'	ρ_1'	ν_s'	ρ_s'	i_1'
FIG. 11	5	$n-2$	4	$n-1$	$n-4$	$n-2$	3
FIG. 12	5	$n-2$	4	$n-1$	$n-4$	$n-2$	5

Figures 11 and 12 are equivalent for the appropriate values of n between nine and five, to all the odd-numbered and all the even-numbered figures given previously. From (A4) and (A6), it follows that Figs. 11 and 12 only specify IR's of SL_n which should be included in the reduction (A8) if n is greater than or equal to six and seven, respectively.

It should perhaps be stressed that the geometric transformation connecting $[\rho]_{b+t}[\lambda]_{a-t}$ and $[\rho^*; \lambda]_{a-t}^{sn-b-t}$ is such that reading [in (iii) of R] the added letters from right to left along each row taken, in turn, *from top to bottom* gives the same result as reading [in (iii) of Q] the added letters from right to left along each row taken, in turn, *from bottom to top*. Furthermore, although the columns of $[\rho]_{b+t}$ are counted in the usual way *from left to right*, the dotted diagram notation is such that the columns of $[\rho^*]^{sn-b-t}$ are counted *from right to left*. Thus, if $[\nu^*]^{sn-b}$ and $[\rho^*]^{sn-b-t}$ are replaced by $[\nu]^b$ and $[\rho]^{b-t}$, respectively, (A8) is transformed into (4.5) and ρ_s' is transformed into $(n - \rho_1')$. Therefore, (A4) and (A6) are transformed into (4.7) and (4.6), respectively, and the corresponding reduction theorem for IR's of SL_n is obtained from the application of Q to (4.5) by retaining only those terms satisfying (4.6) and (4.7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Considerations on the Lorentz–Dirac Equation

JACK COHN
Dept. of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73069

(Received 28 March 1969)

Charged particles whose motions are described by the Lorentz–Dirac equation are considered. It is proved that, for an arbitrary collection of such classical charges moving solely under mutual electromagnetic interaction, the position of each charge and the force acting on each are infinitely differentiable functions of the proper time throughout any time interval for which the particle separations are all nonzero. A corollary of this result is then shown to be that discontinuous-, step-, or δ-function-type external boundary forces which might act on nonzero-size charged particles are not permissible.

I. INTRODUCTION

The Lorentz–Dirac equation which governs the motion of classical (nonquantum) radiating charges has by now been derived from several quite different points of view.[1–4] This, together with recent considerations of Plass[5] which show that meaningful solutions do exist for very general types of force fields actually encountered in physical situations, means that these equations must be heeded seriously. Therefore, for instance, the provocative prediction of the "preacceleration" effect that arises from this equation must be considered carefully. However, it seems fair to say that the entire notion of causality as related to this effect seems poorly understood at present.

The motivation for the present work arose from a desire to understand such effects in the customary context of classical physics. As will be seen, even though the problem is not resolved here, several results will be discussed which may have an important bearing on the matter.

In the following it will be proved that, for an arbitrary collection of classical charged particles (each governed by the Lorentz–Dirac equation) moving solely under mutual electromagnetic interaction, the position of each charge and the force acting on each are infinitely differentiable functions of the proper time throughout any time interval for which the particle separations are all nonzero. As a corollary it will then be shown that, therefore, discontinuous-, step-, or δ-function-type external boundary forces which might act on nonzero-size charged particles are not permissible.

II. DIFFERENTIABILITY THEOREM

Consider a system of *classical* charged particles (with motions governed by the Lorentz–Dirac equation) moving *solely* under mutual electromagnetic interaction. The particles may be radiating and may be moving relativistically. It is, however, stipulated that they are moving in such a manner that mutual particle separations are nonzero throughout some specified time interval. (The unfortunate consequent restriction that, for certain situations, the specified time interval cannot be infinite is unavoidable, if the present method of proof is to be valid. For situations where the particle separations are always nonzero, the time interval may be taken as infinite. In other cases, like the "simple" case of one charge approaching another along a straight line, the time interval cannot be infinite.) Then it will be shown that the position of each charge as well as the force acting on each are infinitely differentiable functions of the proper time throughout the specified time interval.

Throughout the present work all Greek indices take on values from 1 to 4; Latin indices take on values from 1 to 3; the invariant arc length $d\tau$ is defined by

$$-c^2\,d\tau^2 = \sum_{i=1}^{3} dx^i\,dx^i - dx^4\,dx^4, \tag{1}$$

where $x^4 = ct$, and the metric $g_{\mu\nu}$ is taken as the diagonal matrix $(1, 1, 1, -1)$. For the most part, the notation to be used will be that found in Rohrlich's recent book.[6]

We single out one of the particles and refer to its coordinates as x^μ, $\mu = 1, \cdots, 4$. The equation of motion of this charge is then

$$m\frac{d^2X^\mu}{d\tau^2} = \frac{2}{3}\frac{e^2}{c^3}\frac{d^3X^\mu}{d\tau^3} - \frac{2}{3}\frac{e^2}{c^5}\frac{d^2X^\sigma}{d\tau^2}\frac{d^2X_\sigma}{d\tau^2}\frac{dX^\mu}{d\tau} + \frac{e}{c}F^{\mu\nu}\frac{dX_\nu}{d\tau}, \tag{2}$$

where $F^{\mu\nu}$ is the retarded-field tensor produced by all the other charges at the location of the "preferred"

[1] P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1938).
[2] J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **21**, 425 (1949); **17**, 157 (1945).
[3] F. Rohrlich, Ann. Phys. (N.Y.) **13**, 93 (1961).
[4] J. Cohn, Am. J. Phys. **35**, 949 (1967).
[5] G. N. Plass, Rev. Mod. Phys. **33**, 37 (1961).
[6] F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965); see esp. Chaps. 4 and 6.

charge and where e without subscript refers to the preferred charge.

We can rewrite this equation as follows:

$$\dot{V}^\mu - b^{-1}\ddot{V}^\mu + b^{-1}c^{-2}\dot{V}^\sigma V_\sigma V^\mu = ec^{-1}F^{\mu\nu}V^\nu, \quad (3)$$

where $V^\mu \equiv dX^\mu/d\tau$, the dot refers to differentiation with respect to τ, and $b^{-1} = 2e^2/3mc^3$.

In general, $F^{\mu\nu}$ is given by the expression

$$F^{\mu\nu} = \sum_k \{(e^k/\rho_k^2 c)(V_k^\mu U_k^\nu - V_k^\nu V_k^\mu) + (e^k/\rho_k c^2) \times [(A_k^\mu V_k^\nu - A_k^\nu V_k^\mu)c^{-1} - U_k^\mu(V_k^\nu A_{u_k}/c + A_k^\nu) + U_k^\nu(V_k^\mu A_{u_k}/c + A_k^\mu)]\}_{\text{ret}}, \quad (4)$$

where the sum is over all charges e^k other than the preferred one, U^μ is a spacelike 4-vector orthogonal to V^μ and A^μ is the acceleration 4-vector, ρ_k is the distance from the retarded location of particle k to the field point (in that charge's rest system), and $A_{u_k} \equiv A_k^\sigma(U_\sigma)_k$. Now,

$$\frac{dF^{\mu\nu}}{d\tau} = \frac{\partial F^{\mu\nu}}{\partial X^\sigma}\frac{dX^\sigma}{d\tau} \quad (5)$$

and we can evaluate $\partial F^{\mu\nu}/\partial X^\sigma$ by considering the following relations (where we momentarily drop the superscript k):

$$\frac{\partial V^\mu}{\partial X^\sigma} = \left(-\frac{A^\mu}{\rho c}R_\sigma\right)_{\text{ret}}, \quad \frac{\partial A^\mu}{\partial X^\sigma} = \left(-\frac{\dot{A}^\mu}{\rho c}R_\sigma\right)_{\text{ret}},$$

etc., and

$$\frac{\partial \rho}{\partial X^\sigma} = \left(U_\sigma + \frac{A_u R_\sigma}{c^2}\right)_{\text{ret}}, \quad (6)$$

$$\frac{\partial U^\mu}{\partial X^\sigma} = \left[-\frac{1}{\rho^2}\frac{\partial \rho}{\partial X^\sigma}R^\mu + \frac{1}{\rho}\left(\delta_\sigma^\mu + V^\mu\frac{R_\sigma}{\rho c}\right) + \frac{1}{c^2}A^\mu\frac{R_\sigma}{\rho}\right]_{\text{ret}},$$

where R^μ is the null vector going from the respective charge to the field point.

From the above relations we see that the quantity

$$\frac{d}{d\tau}(F^{\mu\nu}V_\nu) = F^{\mu\nu}A_\nu + V_\nu\frac{dF^{\mu\nu}}{d\tau} \quad (7)$$

depends only on the following quantities: V^μ, A^μ, $A_k^\mu(\text{ret})$, $\dot{A}_k^\mu(\text{ret})$, $U_k^\mu(\text{ret})$, $V_k^\mu(\text{ret})$, $\rho_k \equiv \rho_k(\text{ret})$.

We now construct the following argument: In Eq. (3) we know that $F^{\mu\nu}V_\nu$ exists for all τ in the specified interval, since the particles do not have zero separation. Therefore, $\dot{V}^\mu$, $\ddot{V}^\mu$, V^μ exist for all such τ, where quantities without the subscript k refer to the preferred charge. Similarly then, we can assert that the $\dot{V}_k^\mu$, $\ddot{V}_k^\mu$, V_k^μ exist (for all k) for all such τ. Therefore, the quantity

$$\frac{d}{d\tau}(F^{\mu\nu}V_\nu)$$

exists for all such τ. Therefore, $\dddot{V}^\mu$ exists for all such τ. Similarly then, we can again assert that $\dddot{V}_k^\mu$ (for all k) exists for all such τ. Therefore, the quantity

$$\frac{d^2}{d\tau^2}(F^{\mu\nu}V_\nu)$$

exists for all such τ, $\cdots$ etc. Therefore, derivatives of $F^{\mu\nu}V_\nu$ and X^μ exist to all orders in the specified time interval.

It is interesting to note that this derivation would not work if the Schott term (involving $\ddot{V}$) were absent from the Lorentz–Dirac equation.

It is again to be noted that it was necessary to require that the particle separations be nonzero, since otherwise the field tensor may become singular and the method of proof would not work.

III. COROLLARY

The above theorem together with a particular stipulation leads to an interesting corollary.

This ansatz we take to be the following statement: Since the Lorentz–Dirac equation is a formulation only involving classical electromagnetic fields, we can allow only such fields even as boundary or initial conditions. That is, even the boundary forces can only arise from mutual classical electromagnetic interactions.

This requirement, then, precludes boundary or initial forces which are either nonelectromagnetic, or short-range forces which necessarily arise from quantized systems. Thus, electron emission from an atom or metal or collisions of electrons with atoms or gravitational interactions comprise inadmissible boundary conditions.

It follows as a corollary to the above theorem that the boundary forces themselves must be infinitely differentiable functions of the time, for nonzero size particles (that is, for particles whose separations are always nonzero).

Therefore, the well-known δ-function and step-function forces that have been used as boundary conditions on the Lorentz–Dirac equation (see Ref. 5) are not admissible for such particles.

However, it is by the use of such boundary conditions (see Refs. 1 and 5) that one customarily demonstrates the existence of the preacceleration effect with the theory. Unfortunately, the problem is not thereby disposed of, since the preacceleration effect supposedly still exists even for boundary forces which are infinitely differentiable (See Ref. 6, p. 151).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Asymptotic Theory of Transition Radiation*

ROBERT M. LEWIS AND JACK K. COHEN
Courant Institute of Mathematical Sciences, New York University, New York, N.Y.

(Received 26 April 1966; Revised Manuscript Received 7 April 1969)

We treat the electromagnetic fields arising from the passage of a point source through an interface separating two dissimilar dispersive media. Since the media are dispersive, they are partially characterized by a characteristic frequency λ. Our expansions are valid for $\lambda a/c \gg 1$, where a is a characteristic length and c is the speed of light.

1. INTRODUCTION

Transition radiation is an electromagnetic phenomenon associated with the passage of a point source through an interface between two media with different electromagnetic properties. It is well known that, if such a source travels with sufficiently high speed in a dispersive medium, then it will emit Cerenkov radiation. Compared to this radiation, the transition radiation is a second-order effect.

For a dispersive medium, the time-dependent equations governing the electromagnetic field are an integro-differential system. The system, as treated here, involves a large parameter λ which is a characteristic frequency of the medium and which will be our expansion parameter. Since the notion of a "characteristic frequency" may be an unfamiliar one to the reader, we cite the example of an isotropic plasma in which case the dielectric permeability is given by $1 - \phi^2/(\omega^2 + i\nu\omega)$. In this case the "plasma frequency" ϕ might serve as the characteristic frequency λ. (For a more general example, see Lewis,[1] Appendix D.) An equivalent dimensionless parameter is $\lambda_0 = \lambda a/c$, where "a" is the distance between the source and observation points and "c" is the speed of light. Our aim is to obtain an asymptotic expression for the transition radiation fields for large λ (or large λ_0).

The phenomenon of transition radiation has been previously discussed by several Russian physicists.[2,3] These authors have examined the radiation fields which arise from sources such as monopoles or dipoles moving on a straight line with constant velocity through a plane interface. The analysis given here pertains to a general class of point sources which includes as special cases all multipoles. Moreover, our source moves on an arbitrarily prescribed trajectory which intersects a plane interface. An advantage of treating the case of an arbitrary trajectory is that the source may be supposed to start and stop smoothly. Thus, we may avoid assumptions involving infinite path lengths or infinite accelerations. There are two basic underlying mathematical differences between our procedure for obtaining the asymptotic expansion and the procedures used previously. First of all, we exploit the method of multiple stationary phase,[4] while the earlier authors considered only problems which contained enough symmetry so that the integral representation of the solution could be reduced to a form in which stationary-phase (or saddle-point) evaluation was needed only for single integrals. A second difference lies in the distinction between the "explicit" and "parametric" viewpoints (see Sec. 3), both of which may be employed in our problem. When the method of stationary phase is applied to the integral representation of the solution, it leads naturally to a parametric representation of the asymptotic solution involving certain straight lines called "rays." This parametric representation shows that we may think of transition radiation as being a burst of radiation emitted from the "transition point" (i.e., the space–time point at which the source intersects the interface) and which is carried on rays from that point. In Fig. 1 we have depicted this family of rays and also the family of rays associated with the Cerenkov radiation. When appropriately specialized, the results given here agree with those appearing in the literature.

In Sec. 2 we formulate the interface problem and give an integral representation for its solution for an arbitrary source. We also carry out some preliminary steps in the asymptotic expansion of the integral representation which are valid for any source localized in space. Finally, we consider the special case of a moving point source whose trajectory intersects the interface. In Sec. 3 we complete the asymptotic expansion to obtain an expression for the fields associated

* This research was supported by the Office of Naval Research under Contract No. NONR 285(48). Reproduction in whole or in part is permitted for any purpose of the United States Government.

[1] R. M. Lewis, Arch. Ratl. Mech. Anal. **20**, 191 (1965).

[2] G. M. Garibian, Zh. Eksp. Teor. Fiz. **33**, 1403 (1957) [Sov. Phys.—JETP **6**, 1079 (1958)].

[3] V. E. Pafomov Zh. Eksp. Teor. Fiz. **36**, 1853 (1959) [Sov. Phys.—JETP **9**, 1321 (1959)].

[4] R. M. Lewis, in *Asymptotic Solutions of Differential Equations and Their Applications* (John Wiley & Sons, New York, 1964).

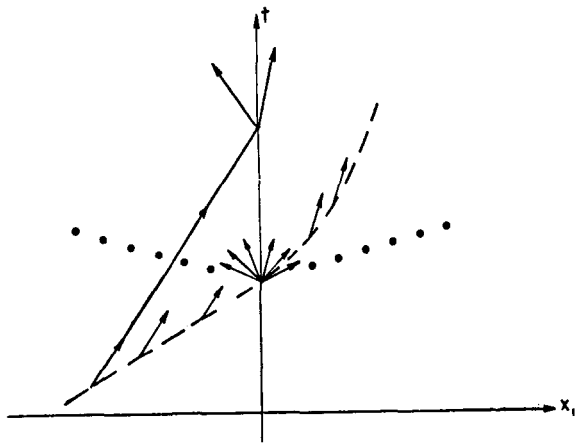


FIG. 1. Projection of ray configuration in $t-x_1$ space. The particle trajectory is shown as dashed. In general, the Cerenkov rays (shown here for a single emission frequency) are emitted from all points along the trajectory. The transition rays are all emitted from the "transition point" (see text) with all speeds between $\pm c$. The limiting rays are shown as dotted.

with transition radiation. In the Appendix we list and discuss all the assumptions made about the dielectric permeability function.

This paper was motivated by the research of Lewis[1] into the applicability of "ray methods" to problems involving integro-differential systems such as the one which appears in this paper. Using the expansion derived in Ref. 1 and the results given here and by Handelsman,[5] one could obtain the radiation fields arising when the source considered here moves in a medium consisting of two smoothly varying inhomogeneous media separated by a smooth interface. However, we do not carry out this procedure here.

2. INTERFACE PROBLEMS FOR MAXWELL'S EQUATIONS IN DISPERSIVE MEDIA

In a homogeneous isotropic medium, Maxwell's equations in Gaussian units take the form

$$\frac{\partial}{\partial t}\mathfrak{D} - c\nabla \times \mathfrak{H} = -4\pi \mathbf{J}, \tag{2.1}$$

$$\frac{\partial}{\partial t}\mathfrak{B} + c\nabla \times \mathfrak{E} = 0, \tag{2.2}$$

$$\nabla \cdot \mathfrak{D} = 4\pi\rho, \tag{2.3}$$

$$\nabla \cdot \mathfrak{B} = 0. \tag{2.4}$$

The source terms $\rho(t, \mathbf{X})$ and $\mathbf{J}(t, \mathbf{X})$ must satisfy the continuity equation

$$\frac{\partial}{\partial t}\rho + \nabla \cdot \mathbf{J} = 0. \tag{2.5}$$

By taking the divergence of (2.1) and (2.2), we see that (2.3) and (2.4) are automatically satisfied if they are satisfied at any given time t_0. In a homogeneous, isotropic, electrically dispersive medium the constitutive relations are[6]

$$\mathfrak{D}(t, \mathbf{X}) = \int f(t)\mathfrak{E}(t-\tau, \mathbf{X})\, d\tau, \tag{2.6}$$

$$\mathfrak{B}(t, \mathbf{X}) = \mu\mathfrak{H}(t, \mathbf{X}). \tag{2.7}$$

Here μ is a real, positive constant and $f(t)$ is a real-valued "memory" function,[7] for which we assume the causality condition

$$f(t) \equiv 0, \quad \text{for} \quad t < 0. \tag{2.8}$$

The dielectric permeability $\hat{\epsilon}$ is then given by

$$\hat{\epsilon}(\hat{\omega}) = \int e^{i\hat{\omega}t} f(t)\, dt. \tag{2.9}$$

We now set

$$\omega = \hat{\omega}/\lambda, \quad \hat{\epsilon}(\hat{\omega}) = \tilde{\epsilon}(\omega). \tag{2.10}$$

Here λ is to be a characteristic frequency of the problem, while ω is to be dimensionless and independent of λ. Thus, Eq. (2.9) may be written in the form

$$f(t) = \frac{\lambda}{2\pi}\int e^{-i\lambda\omega t}\tilde{\epsilon}(\omega)\, d\omega. \tag{2.11}$$

We shall seek expansions for the transition radiation fields valid in the limit $\lambda \to \infty$. In Ref. 1 it is shown that the true physical interpretation of such an expansion is that it is valid for large values of the dimensionless parameter

$$\lambda_0 = \lambda a/c.$$

Here "c" is the speed of light in vacuo and "a" is a characteristic length for the given problem. Following Lewis,[1] we will always impose the "initial conditions"

$$\mathfrak{E}(t, \mathbf{X}) \equiv 0, \quad \mathfrak{H}(t, \mathbf{X}) \equiv 0, \quad \text{for} \quad t \leq t_0. \tag{2.12}$$

In this case it follows from (2.6), (2.7), and (2.8) that

$$\mathfrak{D}(t, \mathbf{X}) \equiv 0, \quad \mathfrak{B}(t, \mathbf{X}) \equiv 0, \quad \text{for} \quad t < t_0. \tag{2.13}$$

Our problem involves a plane interface at $x_1 = 0$. Thus, if $\mathbf{N}$ is a unit vector along the positive x_1 axis, we must satisfy the interface conditions

$$[\mathbf{N} \cdot \mathfrak{D}] = [\mathbf{N} \cdot \mathfrak{B}] = [\mathbf{N} \times \mathfrak{E}] = [\mathbf{N} \times \mathfrak{H}] = 0. \tag{2.14}$$

Here we have employed the notation

$$[h(t, \mathbf{X})] = \lim_{x_1 \downarrow 0} h(t, \mathbf{X}) - \lim_{x_1 \uparrow 0} h(t, \mathbf{X}) \tag{2.15}$$

[5] R. Handelsman, J. Math. Phys. 7, 1982 (1966).

[6] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, New York, 1960).

[7] More precisely, $f(t) = \delta(t) + f_1(t)$, where $f_1(t)$ is real-valued and integrable.

and have assumed that the source term $\mathbf{J}$ is continuous at $x_1 = 0$.

We will assume that our source function

$$\mathbf{J}(t, \mathbf{X}; \lambda) = (j_1, j_2, j_3) \tag{2.16}$$

satisfies[8]

$$\mathbf{J}(t, \mathbf{X}; \lambda) \equiv 0, \quad \text{for} \quad t \leq t_-, \quad t \geq t_+, \tag{2.17}$$

where

$$t_0 < t_- < t_+. \tag{2.18}$$

If we now define ρ consistently with (2.5) and (2.17), it follows that ρ vanishes for $t \leq t_-$. Thus, by (2.18) and (2.13), it follows that (2.3) and (2.4) are satisfied.

Thus, for each given $\mathbf{J}(t, \mathbf{X}; \lambda)$, we need only satisfy (2.1), (2.2), (2.12), and (2.14). In order to be able to express the solution to this problem in a convenient form, we make the following definitions:

$$\boldsymbol{\varkappa} = (0, k_2, k_3), \quad \text{for} \quad k_2, k_3 \text{ real}, \tag{2.19}$$

$$' = \partial/\partial x, \tag{2.20}$$

$$\tilde{m}^2 = (\omega^2/c^2)\mu\tilde{\epsilon}, \tag{2.21}$$

$$\tilde{n}^2 = \tilde{m}^2 - \kappa^2, \tag{2.22}$$

$$\mathbf{G}(t, \mathbf{X}, k_2, k_3; \lambda) = (g_1, g_2, g_3) = \left(j_1, \frac{k_3}{\kappa^2} j_2 - \frac{k_2}{\kappa^2} j_3, \frac{k_2}{\kappa^2} j_2 + \frac{k_3}{\kappa^2} j_3\right), \tag{2.23}$$

and

$$\begin{aligned} \tilde{\epsilon} &= \tilde{\epsilon}_+, \quad x_1 > 0, \quad \mu = \mu_+, \quad x_1 > 0, \\ &= \tilde{\epsilon}_-, \quad x_1 < 0, \quad = \mu_-, \quad x_1 < 0, \\ \tilde{n}^2 &= \tilde{n}_+^2, \quad x_1 > 0, \quad \tilde{m}^2 = \tilde{m}_+^2, \quad x_1 > 0, \\ &= \tilde{n}_-^2, \quad x_1 < 0, \quad = \tilde{m}_-^2, \quad x_1 < 0. \end{aligned} \tag{2.24}$$

Also we define $\tilde{n}$ as the square root of $\tilde{n}^2$ which has a positive-imaginary part at $\omega = 0$. By the use of Fourier-transform techniques and by the introduction of the Hertz polarization potential, we find that the solution to our problem can be expressed by

$$\begin{aligned} \mathcal{E}(t, \mathbf{X}) = \left(\frac{\lambda}{2\pi}\right)^3 \int_c d\omega \int dk_2 \int dk_3 \\ \times \exp\{i\lambda(k_2 x_2 + k_3 x_3 - \omega t)\} \\ \times \frac{1}{\tilde{\epsilon}_\pm} \{(p_1'' + \lambda^2 \tilde{m}_\pm^2 p_1 + i\lambda\kappa^2 p_3')\mathbf{N} \\ + \lambda^2 \tilde{m}_\pm^2 p_2(\boldsymbol{\varkappa} \times \mathbf{N}) + i\lambda(p_1' - i\lambda\tilde{n}_\pm^2 p_3)\boldsymbol{\varkappa}\}, \end{aligned} \tag{2.25}$$

$$\begin{aligned} \mathcal{H}(t, X) = \left(\frac{\lambda}{2\pi}\right)^3 \int_c d\omega \int dk_2 \int dk_3 \\ \times \exp\{i\lambda(k_2 x_2 + k_3 x_3 - \omega t)\} \\ \times \frac{i\lambda\omega}{c} \{i\lambda\kappa^2 p_2 \mathbf{N} - (i\lambda p_1 - p_3') \\ \times (\boldsymbol{\varkappa} \times \mathbf{N}) - p_2' \boldsymbol{\varkappa}\}, \end{aligned} \tag{2.26}$$

[8] t_+ could be $+\infty$.

where for $j = 1, 2, 3$,

$$\begin{aligned} p_j(\omega, x_1, k_2, k_3) \\ = \pm \frac{4\pi}{i\lambda\omega} \int_{t_-}^{t_+} d\tau \int d\xi_2 \int d\xi_3 \\ \times \exp\{-i\lambda(k_2\xi_2 + k_3\xi_3 - \omega\tau)\} \\ \times \left\{ \tilde{a}_\pm^j \int_0^{x_1} d\xi_1 \exp\{\pm i\lambda\tilde{n}_\pm(x_1 - \xi_1)\} \right. \\ \times g_j(\tau, \xi_1, \xi_2, \xi_3, k_2, k_3; \lambda) \\ + \tilde{a}_\pm^j \int_{x_1}^{\pm\infty} d\xi_1 \exp\{\mp i\lambda\tilde{n}_\pm(x_1 - \xi_1)\} \\ \times g_j(\tau, \xi_1, \xi_2, \xi_3, k_2, k_3; \lambda) \\ + \tilde{b}_\pm^j \int_0^{\pm\infty} d\xi_1 \exp\{\pm i\lambda\tilde{n}_\pm(x_1 + \xi_1)\} \\ \times g_j(\tau, \xi_1, \xi_2, \xi_3, k_2, k_3; \lambda) \\ \times \frac{1}{\tilde{w}_\pm^j} \int_{\mp\infty}^{0} d\xi_1 \exp\{\pm i\lambda(\tilde{n}_\pm x_1 - \tilde{n}_\mp \xi_1)\} \\ \left. \times g_j(\tau, \xi_1, \xi_2, \xi_3, k_2, k_3; \lambda) \right\} \end{aligned} \tag{2.27}$$

and

$$(\tilde{a}_\pm^1, \tilde{a}_\pm^2, \tilde{a}_\pm^3) = \frac{1}{2i\lambda\tilde{n}_\pm}(1, 1, 1), \tag{2.28}$$

$$\begin{aligned} (\tilde{b}_\pm^1, \tilde{b}_\pm^2, \tilde{b}_\pm^3) \\ = \frac{1}{2i\lambda\tilde{n}_\pm}\left(\frac{\tilde{n}_\pm\tilde{\epsilon}_\mp - \tilde{n}_\mp\tilde{\epsilon}_\pm}{\tilde{n}_\pm\tilde{\epsilon}_\mp + \tilde{n}_\mp\tilde{\epsilon}_\pm}, \frac{\tilde{n}_\pm\mu_\mp - \tilde{n}_\mp\mu_\pm}{\tilde{n}_\pm\mu_\mp + \tilde{n}_\mp\mu_\pm}, \frac{\tilde{n}_\mp\tilde{\epsilon}_\pm - \tilde{n}_\pm\tilde{\epsilon}_\mp}{\tilde{n}_\pm\tilde{\epsilon}_\mp + \tilde{n}_\mp\tilde{\epsilon}_\pm}\right), \end{aligned} \tag{2.29}$$

$$\begin{aligned} \left(\frac{1}{\tilde{w}_\pm}, \frac{1}{\tilde{w}_\pm^2}, \frac{1}{\tilde{w}_\pm^3}\right) \\ = \frac{1}{i\lambda}\left(\frac{\tilde{\epsilon}_\pm}{\tilde{n}_\pm\tilde{\epsilon}_\mp + \tilde{n}_\mp\tilde{\epsilon}_\pm}, \frac{\mu_\pm}{\tilde{n}_\pm\mu_\mp + \tilde{n}_\mp\mu_\pm}, \frac{\tilde{n}_\mp}{\tilde{n}_\pm}\frac{\tilde{\epsilon}_\pm}{\tilde{n}_\mp\tilde{\epsilon}_\pm + \tilde{n}_\pm\tilde{\epsilon}_\mp}\right). \end{aligned} \tag{2.30}$$

In Eqs. (2.25)–(2.30) the upper signs in the subscript $\pm$ or $\mp$ refer to $x_1 > 0$, while the lower signs refer to $x_1 < 0$. The contour c in (2.25) and (2.26) is parallel to and just above the real ω axis. We may note that the exact solution [(2.25)–(2.30)] simplifies considerably in the special case where $j_2 = j_3 = 0$—for then $g_2 = g_3 = 0$ and, hence, $p_2 = p_3 = 0$.

From (2.27) we see that the p_j consist of the sum of four integrals. Since by (2.25) and (2.26) the solution to our problem consists of a multiple integral of a linear combination of the p_j, we conclude that the fields $\mathcal{E}$ and $\mathcal{H}$ also consist of a linear combination of four corresponding integrals. These four integrals can be given a useful physical interpretation. In the order given in (2.27), the first two integrals correspond[9] to a "primary wave" emitted by the source, the third to a "reflected wave" produced by the incidence of the primary wave on the interface, and the fourth to a "transmitted wave" coming from the other side of the

[9] From (2.27) the ξ_1 integrals of the first two terms can be combined in the form

$$\tilde{a}_\pm^j \int_0^{\pm\infty} d\xi_1 \exp\{\pm i\lambda\tilde{n}_\pm |x_1 - \xi_1|\} g_j.$$

interface. Moreover, of the two integrals representing the primary wave, the first corresponds to a wave which is moving *away* from the interface, while the second corresponds to a wave moving towards the interface and which is thus *incident* on the interface. This identification will be borne out in the expansions we will develop.

We will henceforth assume that the media which occur in our problem are "weakly dissipative," that is,

$$\tilde{\epsilon}(\omega) \sim \epsilon(\omega) + (i/\lambda)\delta(\omega), \quad \lambda \to \infty, \tag{2.31}$$

where $\epsilon(\omega)$ and $\delta(\omega)$ are real functions for real ω. It then follows from (2.22) that

$$\tilde{n} \sim n + (1/\lambda)n', \tag{2.32}$$

where

$$n = [(\omega^2\mu\epsilon/c^2) - \kappa^2]^{\frac{1}{2}}, \quad n' = \omega^2\mu\delta/2c^2n. \tag{2.33}$$

It follows easily that we obtain a preliminary asymptotic expansion of our solution by merely removing the symbols "$\sim$" from $\tilde{\epsilon}$, $\tilde{n}$, $\tilde{m}$, $\tilde{a}^j_\pm$, $\tilde{b}^j_\pm$, and $\tilde{w}^j_\pm$ everywhere except in the exponentials defining the p_j, where instead we replace $\tilde{n}$ by the expression (2.32).[10] Hereafter, when we refer to the equations defining these quantities, we shall assume that these approximations have been made.

Handelsman[5] has discussed the representation of a multipole oscillating about an arbitrary trajectory $\mathbf{X} = \mathbf{Y}(t)$ by the source function

$$\mathbf{J}(t, \mathbf{X}; \lambda) = \lambda^d \mathbf{F}[t, \lambda\{\mathbf{X} - \mathbf{Y}(t)\}]e^{-i\lambda q(t)}. \tag{2.34}$$

Here, d is a real number to be chosen for convenience[11] and for each t, $\mathbf{F}[t, \mathbf{X}]$ is assumed to vanish identically for $\mathbf{X}$ outside some sphere of finite radius in $\mathbf{X}$ space. Moreover, by (2.17) we must also stipulate that $\mathbf{F}[t, \mathbf{X}]$ vanishes for t outside the time interval $t_- \leq t \leq t_+$.

We note that, by the assumptions made on $\mathbf{F}[t, \mathbf{X}]$, it follows that, as $\lambda \to \infty$, the region in which $\mathbf{J}(t, \mathbf{X}; \lambda)$ is nonzero shrinks to

$$\mathbf{X} = \mathbf{Y}(t), \quad t_- \leq t \leq t_+. \tag{2.35}$$

Hence, it is reasonable that (2.34) represents the motion of a point source along the trajectory given by (2.35). In fact, this source is a generalization of the commonly used monopole source traveling on a straight line with constant velocity $\mathbf{V}$ which is given by

$$\begin{aligned}\mathbf{J}(t, \mathbf{X}; \lambda) &= \lambda^d e\mathbf{V}\delta[\lambda(\mathbf{X} - \mathbf{V}(t))] \\ &\equiv \lambda^{d-3} e\mathbf{V}\delta[\mathbf{X} - \mathbf{V}t].\end{aligned} \tag{2.36}$$

By using (2.34), we simultaneously treat moving multipoles and more general moving localized sources. The inclusion of the factor $e^{-i\lambda q(t)}$ enables us to account for the effect of the point source making small oscillations about the trajectory (2.35) or for the effect due to oscillation of the source strength.[5]

For the sake of being definite, we will stipulate that in (2.34)

$$d = 2, \quad y_1(t_-) < 0. \tag{2.37}$$

The latter condition means that the source starts on the left side of the interface. Furthermore, to simplify the analysis we will assume that

$$y_1(t_1) \geq y_2(t_2), \quad \text{for} \quad t_1 > t_2. \tag{2.38}$$

We can now obtain an asymptotic expansion for $\mathcal{E}$ and $\mathcal{H}$ by making use of the special form of the source function (2.34). The functions g_j defined by (2.23) now have the form

$$\begin{aligned}g_j(\tau, \xi_\nu, k_2, k_3; \lambda) &= \lambda^2 h_j[\tau, \lambda\{\xi_\nu - y_\nu(\tau)\}, k_2, k_3]e^{-i\lambda q(\tau)}, \\ &\qquad \nu, j = 1, 2, 3,\end{aligned} \tag{2.39}$$

where

$$\begin{aligned}\mathbf{H}[\tau, \mathbf{X}, k_2, k_3] &= (h_1, h_2, h_3) \\ &= \left(f_1, \frac{k_2}{\kappa^2}f_2 - \frac{k_2}{\kappa^2}f_3, \frac{k_2}{\kappa^2}f_2 + \frac{k_3}{\kappa^2}f_3\right).\end{aligned} \tag{2.40}$$

Thus, if we make the change of variables

$$\gamma_\nu = \lambda\{\xi_\nu - y_\nu(\tau)\}, \quad \nu = 1, 2, 3, \tag{2.41}$$

then (2.27) is replaced by[12]

$$\begin{aligned}p_j \sim \pm\frac{4\pi}{i\lambda^2\omega}\int_{t_-}^{t_+} d\tau \int d\gamma_2 \int d\gamma_3 \\ \times \exp\left\{-i\lambda\left[k_2y_2(\tau) + k_3y_3(\tau) - \omega\tau + q(\tau) - \frac{1}{\lambda}(k_2\gamma_2 + k_3\gamma_3)\right]\right\} \\ \times \left\{a^j_\pm \exp\left\{\pm i\lambda\left(n_\pm + \frac{i}{\lambda}n'_\pm\right)[x_1 - y_1(\tau)]\right\}\right. \\ \left.\times \int_{-\lambda y_1(t)}^{\lambda[x_1 - y_1(\tau)]} d\gamma_1 e^{\mp i n_\pm \gamma_1} h_j[\tau, \gamma_\nu, k_2, k_3]\right. \\ + a^j_\pm \exp\left\{\mp i\lambda\left(n_\pm + \frac{i}{\lambda}n'_\pm\right)[x_1 - y_1(\tau)]\right\} \\ \times \int_{\lambda[x_1 - y_1(t)]}^{\pm\infty} d\gamma_1 e^{\pm i n_\pm \gamma_1} h_j[\tau, \gamma_\nu, k_2, k_3] \\ + b^j_\pm \exp\left\{\pm i\lambda\left(n_\pm + \frac{i}{\lambda}n'_\pm\right)[x_1 + y_1(\tau)]\right\} \\ \times \int_{-\lambda y_1(t)}^{\pm\infty} d\gamma_1 e^{\pm i n_\pm \gamma_1} h_j[\tau, \gamma_\nu, k_2, k_3] \\ + \frac{1}{w^j_\pm}\exp\left\{\pm i\lambda\left[\left(n_\pm + \frac{i}{\lambda}n'_\pm\right)x_1 - \left(n_\mp + \frac{1}{\lambda}n'_\mp\right)y_1(\tau)\right]\right\} \\ \left.\times \int_{\mp\infty}^{-\lambda y_1(t)} d\gamma_1 e^{\mp i n_\mp \gamma_1} h_j[\tau, \gamma_\nu, k_2, k_3]\right\}, \\ \nu, j = 1, 2, 3.\end{aligned} \tag{2.42}$$

[10] More precisely, we define n by the first equation of (2.33) and replace $\tilde{n}$ by n everywhere except as noted above; we define $m^2 = (\omega^2/c^2)\mu\epsilon$ [cf. (2.21)] and replace $\tilde{m}$ everywhere by m; we define $a^j_\pm = (\frac{1}{2}i\lambda n_\pm)$ (1, 1, 1) and replace $\tilde{a}^j_\pm$ by $a^j_\pm$ everywhere, etc.

[11] Changing the number d in (2.34) merely has the effect of multiplying all fields by the same power of λ.

[12] Here and below we also make use of the weak dissipation expansion described in the paragraph following Eq. (2.33).

We now consider the case in which the trajectory $\mathbf{X} = \mathbf{Y}(t)$ intersects the interface $x_1 = 0$. By (2.37) and (2.38), it then follows that there exists a unique time of crossing t_c such that

$$\begin{aligned} y_1(t) &< 0, \quad \text{for} \quad t_- \leq t < t_c, \\ y_1(t) &> 0, \quad \text{for} \quad t_c < t \leq t_+. \end{aligned} \tag{2.43}$$

We introduce the Heaviside step function

$$\begin{aligned} \eta[x] &= 1, \quad x \geq 0, \\ &= 0, \quad x < 0, \end{aligned} \tag{2.44}$$

and the transform

$$\begin{aligned} \mathbf{H}^*[\tau, \mathbf{K}] &= (h_1^*, h_2^*, h_3^*) \\ &= \iiint d\gamma_1\, d\gamma_2\, d\gamma_3 e^{-i\mathbf{K}\cdot\boldsymbol{\Gamma}} \mathbf{H}[\tau, \boldsymbol{\Gamma}, k_2, k_3]. \end{aligned} \tag{2.45}$$

Then we obtain from (2.42) the further asymptotic approximation for large λ:

$$\begin{aligned} p_j \sim \pm \frac{4\pi}{i\lambda^2\omega} \Big\{ & a_\pm^j \int_{t_c}^{t_\pm} d\tau \exp\{i\lambda[\pm n_\pm[x_1 - y_1(\tau)] \\ & - k_2 y_2(\tau) - k_3 y_3(\tau) + \omega\tau - q(\tau)]\} \\ & \times e^{\mp n_\pm'[x_1 - y_1(\tau)]} \eta[\pm\{x_1 - y_1(\tau)\}] \\ & \times h_j^*[\tau, \pm n_\pm, k_2, k_3] \\ & + a_\pm^j \int_{t_c}^{t_\pm} d\tau \exp\{i\lambda[\mp n_\pm[x_1 - y_1(\tau)] \\ & - k_2 y_2(\tau) - k_3 y_3(\tau) + \omega\tau - q(\tau)]\} \\ & \times e^{\pm n_\pm'[x_1 - y_1(\tau)]} \eta[\mp\{x_1 - y_1(\tau)\}] \\ & \times h_j^*[\tau, \mp n_\pm, k_2, k_3] \\ & + b_\pm^j \int_{t_c}^{t_\pm} d\tau \exp\{i\lambda[\pm n_\pm[x_1 + y_1(\tau)] \\ & - k_2 y_2(\tau) - k_3 y_3(\tau) + \omega\tau - q(\tau)]\} \\ & \times e^{\mp n_\pm'[x_1 - y_1(\tau)]} h_j^*[\tau, \mp n_\pm, k_2, k_3] \\ & + \frac{1}{w_\pm^j} \int_{t_\mp}^{t_c} d\tau \exp\{i\lambda[\pm n_\pm x_1 \mp n_\mp y_1(\tau) \\ & - k_2 y_2(\tau) - k_3 y_3(\tau) + \omega\tau - q(\tau)]\} \\ & \times e^{\mp n_\pm' x_1 \pm n_\pm' y_1(\tau)} h_j^*[\tau, \pm n_\mp, k_2, k_3] \Big\}. \end{aligned} \tag{2.46}$$

We note that, if instead $y_1(t) < 0$ for $t_- \leq t \leq t_+$, then in the expression analogous to (2.46) the last integral vanishes for $x_1 < 0$, while for $x_1 > 0$ the first three integrals vanish.

The vanishing of these integrals is easily explained in terms of the physical meanings previously ascribed to them. For since the particle remains in the region $x_1 < 0$, there can be no transmitted wave in the region $x_1 < 0$ and no primary or reflected wave in the region $x_1 > 0$.

In the next section we shall employ the asymptotic expansion (2.46) for the p_j to obtain an asymptotic evaluation of the electromagnetic field corresponding to the phenomena of transition radiation.

3. ASYMPTOTIC THEORY OF TRANSITION RADIATION

As explained in Sec. 1, transition radiation is an effect due to the passage of a moving source through an interface. Thus, to obtain an asymptotic expression for the electromagnetic fields associated with this phenomena, we assume in this section that the trajectory (2.35) definitely intersects the interface $x_1 = 0$. For convenience, we assume that the time of crossing t_c, defined by (2.43), is $t = 0$ and that the point of crossing is the origin, that is,

$$t_- < t_c = 0 < t_+, \quad \mathbf{Y}(0) = 0. \tag{3.1}$$

In Sec. 2, we pointed out that the exact solution consisted of linear combinations of multiple integrals of four distinct types. This basic form has been preserved in the partial asymptotic expansion given by (2.25), (2.26), and (2.46). The four types of integrals referred to above may be distinguished by their phases, which (aside from a factor λ) are

$$\begin{aligned} \varphi_\pm^a = {} & \pm n_\pm[x_1 - y_1(\tau)] + k_2[x_2 - y_2(\tau)] \\ & + k_3[x_3 - y_3(\tau)] - \omega[t - \tau] - q(\tau), \end{aligned} \tag{3.2}$$

$$\begin{aligned} \varphi_\pm^i = {} & \mp n_\pm[x_1 - y_1(\tau)] + k_2[x_2 - y_2(\tau)] \\ & + k_3[x_3 - y_3(\tau)] - \omega[t - \tau] - q(\tau), \end{aligned} \tag{3.3}$$

$$\begin{aligned} \varphi_\pm^r = {} & \pm n_\pm[x_1 + y_1(\tau)] + k_2[x_2 - y_2(\tau)] \\ & + k_3[x_3 - y_3(\tau)] - \omega[t - \tau] - q(\tau), \end{aligned} \tag{3.4}$$

$$\begin{aligned} \varphi_\pm^t = {} & \pm n_\pm x_1 \mp n_\mp y_1(\tau) + k_2[x_2 - y_2(\tau)] \\ & + k_3[x_3 - y_3(\tau)] - \omega[t - \tau] - q(\tau). \end{aligned} \tag{3.5}$$

Here we have superscripted the four phases in a manner which is suggestive of the interpretations ascribed to the corresponding integrals. Thus, $\varphi_\pm^a$ is the phase we have associated with the wave moving *away* from the interface and similarly $\varphi_\pm^i$, $\varphi_\pm^r$, and $\varphi_\pm^t$ are, respectively, the phases we have associated with the waves which are *incident* on, *reflected* from, and *transmitted* through the interface.

The primary tool that will be employed in obtaining the asymptotic evaluation of the integrals that define the fields $\mathcal{E}$ and $\mathcal{H}$ for large λ is the method of multiple stationary phase.[4]

There are two equivalent viewpoints which may be taken. The *explicit* viewpoint is that for each fixed $(t, \mathbf{X})$ we find the values of the variables of integration for which the phase is stationary; then the given integral is asymptotic to a sum of contributions from these values. The *parametric* viewpoint is that for each fixed set of values of the integration variable we find the locus of all points $(t, \mathbf{X})$ such that this set of values makes the phase stationary; then the given integral is asymptotic to the sum of contributions from those loci which contain $(t, \mathbf{X})$.

We will find that, in general, the locus corresponding to each set of values of the integration variables is a straight line in $(t, \mathbf{X})$ space. We shall refer to this line as the *ray*[13] associated with the given set of values of the integration variables. The form of the rays associated with each contribution to the asymptotic expansion of the electromagnetic field will allow us to ascribe an appropriate physical interpretation to each of these contributions. Thus, the rays assume the highest importance in our theory.

In order to carry out the stationary phase evaluation of the integrals [(2.25) and (2.26)] defining $\mathcal{E}$ and $\mathcal{H}$, we deform the contour c to the real ω axis. From (2.46) and Lemma 5 of the Appendix, we can see that the solution is exponentially small in λ for real ω, k_2, k_3 such that

$$\omega^2\mu_\pm\epsilon_\pm/c^2 < \kappa^2. \tag{3.6}$$

Furthermore, the contributions from the real branch points $(\omega^2\mu_\pm\epsilon_\pm/c^2 = \kappa)$, are of lower order than those we will consider.[14] Thus, in our stationary phase evaluations we may confine our attention to the "parameter spaces" $\mathcal{T}_\pm$, defined as the set of all real triples (ω, k_2, k_3) such that

$$\omega^2\mu_\pm\epsilon_\pm/c^2 > \kappa^2. \tag{3.7}$$

The leading term of the electromagnetic field defined by (2.25), (2.26), and (2.46) and (3.1) is obtained by performing multiple stationary phase with respect to ω, k_2, k_3, τ. We adopt the parametric viewpoint and obtain the loci (rays) which, for each fixed set ω, k_2, k_3, τ, make the phases $\varphi^a_\pm$, $\varphi^i_\pm$, $\varphi^r_\pm$, $\varphi^t_\pm$, defined by (3.2)–(3.5), stationary. We consider only values of ω, k_2, k_3 which lie in $\mathcal{T}_\pm$ and only those values of τ which lie within the limits of the respective integrations over τ in (2.46).

Consider, for example, an integral defining the electromagnetic field for $x_1 > 0$ which contains the phase φ^a_+. These integrals have the form

$$\psi(t, \mathbf{X}; \lambda) = \int d\omega \int dk_2 \int dk_3 \int_0^{t_+} d\tau A(\omega, k_2, k_3, \tau) e^{i\lambda\varphi_+{}^a} \tag{3.8}$$

The stationary phase relations

$$\frac{\partial}{\partial\omega}\varphi^a_+ = \frac{\partial}{\partial k_2}\varphi^a_+ = \frac{\partial}{\partial k_3}\varphi^a_+ = \frac{\partial}{\partial\tau}\varphi^a_+ = 0$$

yield the equations

$$\mathbf{X} = \mathbf{Y}(\tau) + [n_+(n_+)_\omega]^{-1}\mathbf{K}(t-\tau)\big|_{k_1=n_+}, \tag{3.9}$$

$$\omega = \dot{q}(\tau) + \mathbf{K}\cdot\mathbf{Y}(\tau)\big|_{k_1=n_+}, \tag{3.10}$$

where the dot denotes differentiation w.r.t. τ and where we have

$$(\omega, k_2, k_3) \text{ in } \mathcal{T}_\pm, \quad 0 \le \tau \le t_+. \tag{3.11}$$

Here we have introduced the vector

$$\mathbf{K} = (k_1, k_2, k_3) \tag{3.12}$$

and have used the notation

$$f(k_1, k_2, k_3)\big|_{k_1=n} = f(n, k_2, k_3). \tag{3.13}$$

The rays defined by (3.9) evidently emerge from the particle trajectory $\mathbf{X} = \mathbf{Y}(t)$ at the time $t = \tau$, with the velocity vector $[n_+(n_+)_\omega]^{-1}\mathbf{K}|_{k_1=n_+}$. Moreover, (3.10) is just the familiar "Doppler–Cerenkov law"[5] governing the emitted frequencies on these rays. We can thus conclude that the leading term of the expansion of the integrals of the form (3.8) corresponds to Doppler–Cerenkov radiation. Moreover, in the parameter space $\mathcal{T}_+$, we have, by Lemma (6) of the Appendix, that

$$\frac{k_1}{n_+(n_+)_\omega}\bigg|_{k_1=n_+} = \frac{1}{(n_+)_\omega} > 0.$$

Thus, the ray (3.9) propagates in the direction of increasing x_1, that is, away from the interface. This bears out the significance given to the integrals containing the phase φ^a_+ in Sec. 2.

In a similar manner we could investigate the leading contributions from the other types of integrals which occur in our partial expansion [(2.25), (2.26), and (2.46)]. For each of them we could obtain equations analogous to (3.9) and (3.10), and in each case we would find that this leading term corresponds to Doppler–Cerenkov radiation and that the associated rays propagated in a way consistent with the interpretation ascribed in Sec. 2 to the various integrals. The multiple stationary phase formula[4] would then allow us to obtain a parametric representation of the electromagnetic fields due to Doppler–Cerenkov radiation. Since these effects are discussed in great

[13] In the cases treated here, these rays are identical to those predicted by the ray theory presented in Ref. 1.

[14] It can be shown that the branch-point contributions are at most $O(\lambda^{-1})$. We shall only investigate terms which are $O(1)$ and $O(\lambda^{-\frac{1}{2}})$.

detail in Ref. 5, we shall not examine them further here. However, we remark that these terms are all $O(1)$ with respect to λ.

The next highest order contributions to the fields **E** and **H** come from "edge-critical points" on the boundary of the domain of integration.[15] They can be obtained by integrating by parts with respect to τ and then performing a stationary phase evaluation with respect to the remaining integration variables, ω, k_2, and k_3. We observe from (2.46) that there are three such end-point contributions in τ, namely, $\tau = t_-$, $\tau = t_c = 0$, and $\tau = t_+$. The first and third of these give rise to *transient* effects.[16] Here we will assume that the point source starts smoothly from rest at $t = t_-$ and goes smoothly to rest again at $t = t_+$.[17] In this case the contributions from $\tau = t_\pm$ are of lower order in λ than the contribution for $\tau = 0$, and we shall neglect them.

Denoting the contribution from $\tau = 0$ by a superscript T (for "transition radiation"), by integration by parts in (2.46) we obtain

$$\begin{aligned} p_j^T \sim \pm \frac{4\pi}{\lambda^3\omega} e^{-i\lambda q(0)} & \times \Big\{ a_\pm^j \exp[\pm i\lambda n_\pm x_1] \exp[\mp n'_\pm x_1]\eta[\pm x_1] \\ & \times \frac{h_j^*[0, \pm n_\pm, k_2, k_3]}{\omega - \dot q(0) \mp n_\pm \dot Y_1(0) - k_2 \dot Y_2(0) - k_3 \dot Y_3(0)} \\ & + a_\pm^j \exp[\mp i\lambda n_\pm x_1] \exp[\pm n'_\pm x_1]\eta[\mp x_1] \\ & \times \frac{h_j^*[0, \mp n_\pm, k_2, k_3]}{\omega - \dot q(0) \pm n_\pm \dot Y_1(0) - k_2 \dot Y_2(0) - k_3 \dot Y_3(0)} \\ & + b_\pm^j \exp[\pm i\lambda n_\pm x_1] \exp[\mp n'_\pm x_1] \\ & \times \frac{h_j^*[0, \mp n_\pm, k_2, k_3]}{\omega - \dot q(0) \pm n_\pm \dot Y_1(0) - k_2 \dot Y_2(0) - k_3 \dot Y_3(0)} \\ & - \frac{1}{w_\pm^j} \exp[\pm i\lambda n_\pm x_1] \exp[\mp n'_\pm x_1] \\ & \times \frac{h_j^*[0, \pm n_\mp, k_2, k_3]}{\omega - \dot q(0) \mp n_\mp \dot Y_1(0) - k_2 \dot Y_2(0) - k_3 \dot Y_3(0)} \Big\}. \end{aligned} \tag{3.14}$$

Here we have used the values $t_c = 0$, $\mathbf{Y}(0) = 0$ given by Eq. (3.1).

Since the upper (lower) sign refers to $x_1 > 0$ ($x_1 < 0$), we see that $\eta[\pm x_1] = 1$, $\eta[\mp x_1] = 0$. Thus, the second term in (3.14) vanishes. This is the term which we have associated with a wave moving *towards* the interface. Since we expect the transition radiation to emerge *from* the interface, the vanishing of this term is entirely consistent with the meaning assigned to it. Thus, using the notation developed in (3.12) and (3.13), we have

$$\begin{aligned} p_j^T \sim \pm \frac{4\pi}{\lambda^3\omega} e^{-i\lambda q(0)} \exp[\pm i\lambda n_\pm x_1] \exp[\mp n'_\pm x_1] & \\ \times \Big\{ a_\pm^j \frac{h_j^*[0, \mathbf{K}]}{\omega - \dot q(0) - \mathbf{K}\cdot\dot{\mathbf{Y}}(0)}\Big|_{k_1 = \pm n_\pm} & \\ + b_\pm^j \frac{h_j^*[0, \mathbf{K}]}{\omega - \dot q(0) - \mathbf{K}\cdot\dot{\mathbf{Y}}(0)}\Big|_{k_1 = \mp n_\pm} & \\ - \frac{1}{w_\pm^j} \frac{h_j^*[0, \mathbf{K}]}{\omega - \dot q(0) - \mathbf{K}\cdot\dot{Y}(0)}\Big|_{k_1 = \pm n_\mp} \Big\}. & \end{aligned} \tag{3.15}$$

We notice that in (3.15) the three phases corresponding to the waves away from, reflected from, and transmitted through the interface have reduced to a single common phase. This means that the three waves corresponding to these phases can no longer be distinguished from one another. This is again consistent with the notion that the transition radiation emerges from a point on the interface.

By (3.15) we have

$$\frac{\partial}{\partial x_1} p_j^T \sim i\lambda n_\pm p_j^T, \quad \frac{\partial^2}{\partial x_1^2} p_j^T \sim -\lambda^2 n_\pm^2 p_j^T. \tag{3.16}$$

Thus, by (2.25) and (2.26), we have

$$\begin{aligned} \mathcal{E}^T(t, \mathbf{X}) \sim \frac{\lambda^5}{(2\pi)^3} \int d\omega \int dk_2 \int dk_3 & \\ \times \exp[i\lambda(k_2 x_2 + k_3 x_3 - \omega t)] & \\ \times \frac{1}{\epsilon_\pm} [(p_1^T - n_\pm p_3^T)\kappa^2 \mathbf{N} + m_\pm^2 p_2^T(\boldsymbol{\varkappa} \times \mathbf{N}) & \\ - (p_1^T - n_\pm p_3^T) n_\pm \boldsymbol{\varkappa}], & \end{aligned} \tag{3.17}$$

$$\begin{aligned} \mathcal{H}^T(t, \mathbf{X}) \sim \frac{\lambda^5}{(2\pi)^3} \int d\omega \int dk_2 \int dk_3 & \\ \times \exp[i\lambda(k_2 x_2 + k_3 x_3 - \omega t)] & \\ \times \frac{\omega}{c} [-p_2^T \kappa^2 \mathbf{N} + (p_1^T - n_\pm p_3^T) & \\ \times (\boldsymbol{\varkappa} \times \mathbf{N}) + p_2^T n_\pm \boldsymbol{\kappa}]. & \end{aligned} \tag{3.18}$$

[Here we have also used the weak dissipation assumption (2.31). See the discussion following (2.33).] We now perform stationary phase evaluation of (3.17) and (3.18) with respect to ω, k_2, k_3. We adopt the parametric viewpoint and obtain the loci (rays) which, for

[15] D. S. Jones and M. Kline, J. Math. & Phys. 37, 1 (1958).

[16] R. M. Lewis, "Asymptotic Theory of Transients," in *Proceedings of the Delft Symposium, 1965 (Electromagnetic Wave Theory)* (Pergamon Press, New York, 1967).

[17] That is, the source function (2.34) is supposed to vanish smoothly at $t = t_\pm$.

each fixed set ω, k_2, k_3 in $\mathfrak{T}_\pm$, make stationary the phase

$$\varphi_\pm = \pm n_\pm x_1 + k_2 x_2 + k_3 x_3 - \omega t - q(0) \quad (3.19)$$

appearing in the expansion given by (3.15), (3.17), and (3.18).

The stationary phase relations

$$(\partial/\partial\omega)\varphi_\pm = (\partial/\partial k_2)\varphi_\pm = (\partial/\partial k_3)\varphi_\pm = 0$$

yield the equations

$$\mathbf{X} = \frac{1}{n_\pm(n_\pm)_\omega}\,\mathbf{K}t\big|_{k_1=\pm n_\pm}, \quad (\omega, k_2, k_3) \text{ in } \mathfrak{T}_\pm,\ t > 0. \quad (3.20)$$

From the stationary phase formula[4] we obtain from (3.15), (3.17), and (3.18) a parametric representation, with parameters ω, k_2, k_3, of the transition radiation fields given by (3.20) and

$$\begin{aligned}\mathcal{E}^T \sim{}& \left(\frac{2}{\lambda\pi}\right)^{\frac{1}{2}} \exp\left[i\lambda\varphi^T - \frac{n'_\pm}{(n_\pm)_\omega}\,t + \frac{i\pi}{4}\,\operatorname{sig}\varphi^T_{vj}\right] \\ &\times [|\det \varphi^T_{vj}|]^{-\frac{1}{2}} \\ &\times \frac{1}{\omega\epsilon_\pm}\,[(r_1^T - n_\pm r_3^T)\kappa^2\mathbf{N} + m_\pm^2 r_2^T(\boldsymbol{\varkappa}\times\mathbf{N}) \\ &- (r_1^T - n_\pm r_3^T)n_\pm\boldsymbol{\varkappa}], \end{aligned} \quad (3.21)$$

$$\begin{aligned}\mathcal{H}^T \sim{}& \left(\frac{2}{\lambda\pi}\right)^{\frac{1}{2}} \exp\left[i\lambda\varphi^T - \frac{n'_\pm}{(n_\pm)_\omega}\,t + \frac{i\pi}{4}\,\operatorname{sig}\varphi^T_{vj}\right] \\ &\times [|\det \varphi^T_{vj}|]^{-\frac{1}{2}} \\ &\times \frac{1}{c}\,[-r_2^T\kappa^2\mathbf{N} + (r_1^T - n_\pm r_3^T) \\ &\times (\boldsymbol{\varkappa}\times\mathbf{N}) + r_2^T n_\pm\boldsymbol{\varkappa}]. \end{aligned} \quad (3.22)$$

Here and below the upper (lower) sign refers to $x_1 > 0$ ($x_1 < 0$) and

$$\varphi^T = \left(\frac{m_\pm}{(m_\pm)_\omega} - \omega\right)t - q(0), \quad (3.23)$$

$$\det \varphi_{vj} = \frac{(m_\pm)_{\omega\omega}}{n_\pm^2(m_\pm)_\omega^3}\,t^3, \quad (3.24)$$

$$\operatorname{sig}\varphi_{vj} = \operatorname{sgn}(m_\pm)_{\omega\omega} - 2\operatorname{sgn}\omega, \quad (3.25)$$

and

$$\begin{aligned} r_j^T = \pm\lambda\Bigg[& a_\pm^j \frac{h_j^*[0,\mathbf{K}]}{\omega - \dot q(0) - \mathbf{K}\cdot\dot{\mathbf{Y}}(0)}\bigg|_{k_1=\pm n_\pm} \\ & + b_j^\pm \frac{h_j^*[0,\mathbf{K}]}{\omega - \dot q(0) - \mathbf{K}\cdot\mathbf{Y}(0)}\bigg|_{k_1=\mp n_\pm} \\ & - \frac{1}{W_\pm^j}\,\frac{h_j^*[0,\mathbf{K}]}{\omega - \dot q(0) - \mathbf{K}\cdot\dot{\mathbf{Y}}(0)}\bigg|_{k_1=\pm n_\mp}\Bigg]. \end{aligned} \quad (3.26)$$

The quantities $m_\pm^2$, $n_\pm^2$, $a_\pm^j$, $b_\pm^j$, $1/W_\pm^j$ are defined by (2.21), (2.22), (2.28), (2.29), and (2.30), respectively,[18] and the $h_j^*[0,\mathbf{K}]$ are defined by (2.34), (2.40), and (2.45). Since the quantities $a_\pm^j$, $b_\pm^j$, $1/W_\pm^j$ all contain the factor $1/\lambda$, the r_j^T are $O(1)$ with respect to λ and, thus, the fields $\mathcal{E}^T$ and $\mathcal{H}^T$ given by (3.21) and (3.22) are $O(\lambda^{-\frac{1}{2}})$, a half-order lower than the Doppler–Cerenkov fields.

The rays defined by (3.20) emerge from the point $\mathbf{X} = 0$ at the "time of transition" $t = 0$, with the velocity vector $[n_\pm(n_\pm)_\omega]^{-1}\mathbf{K}|_{k_1=\pm n_\pm}$. Thus, we conclude that the contribution $\mathcal{E}^T$, $\mathcal{H}^T$ to the asymptotic expansion of the electromagnetic field is, indeed, transition radiation. We also note that in the parameter spaces $\mathfrak{T}_\pm$ Lemma 6 of the Appendix shows that

$$\frac{k_1}{n_\pm(n_\pm)_\omega}\bigg|_{k_1=\pm n_\pm} = \pm\frac{1}{(n_\pm)_\omega}$$

is positive for $x_1 > 0$ and negative for $x_1 < 0$. Thus, in both media we have propagation away from the interface.

It is a general feature of asymptotic solutions to partial differential equations that, on the common ray between contributions of different orders, the lower-order contribution becomes singular. This is exemplified in the expansion just derived—for by (3.26) our expansion fails when

$$\omega = \dot q(0) + \mathbf{K}\cdot\dot{\mathbf{Y}}(0), \quad k_1 = n_+, n_-, -n_+, \text{ or } -n_-, \quad (3.27)$$

and (3.27) is just the Doppler–Cerenkov law[19] for the various waves of the zeroth-order field evaluated at the time when the source trajectory crosses the interface.

The expansions obtained so far for the transition radiation fields have been derived under the assumption that both media are actually electrically dispersive. An expansion which is valid for the case when one medium is nondispersive is of obvious interest for comparison with experiments. In a nondispersive medium, however, ϵ is not a function of ω, and this has the consequence that $\det \varphi^T_{jv}$ vanishes.[20] Thus, the expansions (3.21) and (3.22) are not valid in this case.

Since the expansions (3.15), (3.17), and (3.18) are valid for all weakly dispersive media, we will use them to obtain asymptotic expansions for the transition radiation fields in the case where the region $x_1 > 0$ is

[18] It is understood that we are to remove the symbols "~" in these expressions.

[19] Compare (3.27) with (3.10).

[20] For a nondispersive medium, $m = (\omega/c)(\epsilon\mu)^{\frac{1}{2}}$, where ϵ and μ are independent of ω. Hence, $m_{\omega\omega} = 0$.

a vacuum.[21] We then have

$$\tilde{\epsilon}_+ = \epsilon_+ = 1, \quad \mu_+ = 1, \quad n_+^2 = (\omega^2/c^2) = \kappa^2, \quad m_+^2 = (\omega^2/c^2), \quad n'_+ = 0. \tag{3.28}$$

In the vacuum we perform stationary-phase evaluation only with respect to k_2, k_3 in $\mathfrak{I}_\pm$. By (3.19) the stationary-phase relations yield the equations

$$x_2 - (k_2/n_+)x_1 = 0, \quad x_3 - (k_3/n_+)x_1 = 0, \tag{3.29}$$

which, when solved for k_2 and k_3, yield

$$k_2 = \frac{\omega x_2}{c(x_1^2 + x_2^2 + x_3^2)}, \quad k_3 = \frac{\omega x_3}{c(x_1^2 + x_2^2 + x_3^2)}. \tag{3.30}$$

We define

$$\rho = (x_2^2 + x_3^2)^{\frac{1}{2}}, \quad R = (x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}}.$$

Then, at the stationary point, we have the relations

$$\kappa^2 = (\omega^2/c^2)(\rho^2/R^2), \quad n_+ = \omega x_1/cR. \tag{3.31}$$

Thus, by (3.19),

$$\varphi_+ = \varphi(R/c - t) - q(0). \tag{3.32}$$

If we define the vectors

$$\mathbf{N}_1 = \mathbf{N} = (1, 0, 0), \quad \mathbf{N}_2 = (0, 1, 0), \quad \mathbf{N}_3 = (0, 0, 1), \tag{3.33}$$

then, from (3.15), (3.17), and (3.18), we obtain the following representation of the fields in the vacuum:

$$\begin{aligned}\mathcal{E}^V \sim{}& \frac{\lambda}{\pi c^2}\int d\omega \\ &\times \exp\left[i\lambda\left[\omega\left(\frac{R}{c} - t\right) - q(0)\right] + \frac{i\pi}{4}\operatorname{sig}\varphi_{vj}^V\right] \\ &\times [|\det \varphi_{vj}^V|]^{-\frac{1}{2}}\omega\left[\left(r_1^V - \frac{\omega x_1}{cR}r_3^V\right)\frac{\rho^2}{R^2}\mathbf{N}_1\right. \\ &+ r_2^V\frac{\omega}{cR}(x_2\mathbf{N}_3 - x_3\mathbf{N}_2) \\ &\left. - \left(r_1^V - \frac{\omega x_1}{cR}r_3^V\right)\frac{x_1}{R^2}(x_2\mathbf{N}_2 + x_3\mathbf{N}_3)\right], \end{aligned} \tag{3.34}$$

$$\begin{aligned}\mathcal{H}^V \sim{}& \frac{\lambda}{\pi c^2}\int d\omega \\ &\times \exp\left[i\lambda\left[\omega\left(\frac{R}{c} - t\right) - q(0)\right] + \frac{i\pi}{4}\operatorname{sig}\varphi_{vj}^V\right] \\ &\times [|\det \varphi_{vj}^V|]^{-\frac{1}{2}}\omega\left[-r_2^V\frac{\omega\rho^2}{cR^2}\mathbf{N}_1 + \left(r_1^V - \frac{\omega x_1}{cR}r_3^V\right)\right. \\ &\left. \times \frac{1}{R}(x_2\mathbf{N}_3 - x_3\mathbf{N}_2) + r_2^V\frac{\omega x_1}{cR^2}(x_2\mathbf{N}_2 + x_3\mathbf{N}_3)\right]. \end{aligned} \tag{3.35}$$

[21] Similarly, we could deal with the case where $x_1 < 0$ is a vacuum.

Here

$$\det \varphi_{vj}^V = c^2R^4/\omega^2 x_1, \quad \operatorname{sig}\varphi_{vj}^V = -2\operatorname{sgn}\omega, \tag{3.36}$$

and the r_j^V arise from the r_j^T given by (3.26) if we make the substitutions given by (3.30).

We note that the expansions (3.34) and (3.35) have the form of an outgoing spherical wave from the point $t = 0$, $\mathbf{X} = 0$, which is entirely consonant with our intuitive notions about transition radiation.

Finally, we wish to show that our results agree with those given in the literature for certain special cases of (2.34). For this comparison the paper of Garibian[2] is the most convenient to consider, since he obtains a preliminary expansion analogous to that given by (3.15), (3.17), and (3.18). In this paper, Garibian considers the special case of (2.34) given by

$$q(t) \equiv 0, \quad \mathbf{Y} = (vt, 0, 0), \quad \mathbf{F} = (f_1, 0, 0), \tag{3.37}$$

where

$$\begin{aligned} f_1 &= \frac{1}{\lambda^2}ve\delta(x_1 - vt)\delta(x_2)\delta(x_3) \\ &= \lambda ve\delta[\lambda(x_1 - vt)]\delta[\lambda x_3], \end{aligned} \tag{3.38}$$

in media where dissipation may be entirely ignored.[22] It then follows from (2.40) and (2.45) that

$$h_1^*[0, \mathbf{K}] = \frac{1}{\lambda^2}ve, \quad h_2^*[0, \mathbf{K}] = 0, \quad h_3^*[0, \mathbf{K}] = 0. \tag{3.39}$$

Hence, by (3.15), $p_2^T = p_3^T = 0$ and

$$\begin{aligned} p_1^T &= \pm\frac{4\pi ve}{\lambda^5\omega}\exp[\pm i\lambda n_\pm x_1] \\ &\quad\times\left[\frac{a_\pm^1}{\omega \pm n_\pm v} + \frac{b_\pm^1}{\omega \pm n_\pm v} - \frac{1}{W^1}\frac{1}{\omega \mp n_\mp v}\right] \\ &= \pm\frac{4\pi e\epsilon_\pm}{i\lambda^6 v}\cdot\frac{\exp[\pm i\lambda n_\pm x_1]}{n_\pm\epsilon_\mp + n_\mp\epsilon_\pm}\eta, \end{aligned} \tag{3.40}$$

where

$$\begin{aligned}\eta = {}& \left(\frac{\epsilon_\mp}{\epsilon_\pm} \pm \frac{v}{\omega}n_\mp\right)\Big/\left(\frac{\omega^2}{v^2} - n_\pm^2\right) \\ &+ \left(-1 \mp \frac{v}{\omega}n_\mp\right)\Big/\left(\frac{\omega^2}{v^2} - n_\mp^2\right). \end{aligned} \tag{3.41}$$

Thus, by (3.17) and (3.18), for the special case under

[22] That is, we set the $\delta(\omega)$ appearing in (2.31) equal to zero. Consequently, the quantity $n'_\pm = 0$.

consideration we have

$$\mathcal{E}^T(t, \mathbf{X}) \sim \pm \frac{e}{2\pi^2\lambda i v} \int d\omega \int dk_2 \int dk_3 \times \exp\{i\lambda[\pm n_\pm x_1 + k_2 x_3 - \omega t]\} \times \frac{\eta}{n_\pm \epsilon_\mp + n_\mp \epsilon_\pm} [\kappa^2 \mathbf{N} - n_\pm \boldsymbol{\varkappa}], \tag{3.42}$$

$$\mathcal{H}^T(t, \mathbf{X}) \sim \pm \frac{e}{2\pi^2\lambda i v} \int d\omega \int dk_2 \int dk_3 \times \exp\{i\lambda[\pm n_\pm x_1 + k_2 x_2 + k_3 x_3 - \omega t]\} \times \frac{\omega \epsilon_\pm}{c} \frac{\eta}{n_\pm \epsilon_\mp + n_\mp \epsilon_\pm} \boldsymbol{\varkappa} \times \mathbf{N}. \tag{3.43}$$

The expressions (3.42) and (3.43) agree with Eqs. (4) and (9) of Ref. 2, except for the appearance here of the expansion parameter λ and changes in notation.[23]

APPENDIX: LEMMAS CONCERNING $\tilde{n}$ AND RELATED QUANTITIES

In Sec. 2 we defined $\hat{\epsilon}(\hat{\omega})$ by

$$\hat{\epsilon}(\hat{\omega}) = 1 + \int e^{i\hat{\omega}t} f_1(t)\, dt. \tag{A1}$$

Here we assume that $f_1(t)$ is real valued, integrable, and that it vanishes for $t < 0$. These conditions on $f_1(t)$ imply (a) that $\hat{\epsilon}(\hat{\omega})$ is analytic for $\operatorname{Im} \omega > 0$, (b) that $\hat{\epsilon}(\hat{\omega})$ is free of singularities for real ω, (c) that $\hat{\epsilon}(-\hat{\omega}^*) = \hat{\epsilon}^*(\hat{\omega})$, and (d) that $\hat{\epsilon}(0)$ is real. Landau and Lifshitz[24] show that further appropriate physical assumptions about $\hat{\epsilon}(\hat{\omega})$ are (e) $\hat{\epsilon}(\infty) = 1$ and (f) $\operatorname{Im} \hat{\epsilon}(\hat{\omega}) > 0$ for $\hat{\omega}$ real and positive. In order that $\hat{\epsilon}(\hat{\omega})$ as defined by (A1) should satisfy properties (e) and (f), it is sufficient (although not necessary) to assume that

$$f_1(t) \to 0, \quad t \to \infty, \quad f_1(t_1) \geq f_1(t_2) > 0, \quad t_1 < t_2. \tag{A2}$$

Elsewhere[25] Landau and Lifshitz develop the notion of "generalized susceptibility." Conditions (a)–(f) are sufficient to show that $\hat{\epsilon}(\hat{\omega})$ is a generalized susceptibility in this sense. It then follows easily that $\hat{\epsilon}(\omega)$ and

$$\tilde{n}^2(\omega) = (\omega^2/c^2)\mu\tilde{\epsilon} - \kappa^2 \tag{A3}$$

are also generalized susceptibilities.

From a general property of generalized susceptibilities proved in Ref. 25, we have the following:

Lemma 1: For $\operatorname{Im} \omega > 0$, $\tilde{n}^2(\omega)$ is real only on the positive imaginary axis, where it decreases monotonically from $-\kappa^2$ at $\omega = i0$ to $-\infty$ at $\omega = i\infty$.

In Sec. 2 we defined $\tilde{n}$ as that square root of $\tilde{n}^2$ which takes on the value $i\kappa$ at $\omega = 0$. As consequences of Lemma 1, we have:

Lemma 2: $\operatorname{Im} \tilde{n}(\omega) > 0$ for $\operatorname{Im} \omega > 0$.

Also we have:

Lemma 3: $\operatorname{Im} \tilde{n}(\omega) > 0$ and $\operatorname{sgn} \operatorname{Re} \tilde{n}(\omega) = \operatorname{sgn} \omega$ for ω real.

We now obtain some related results which follow from the "weak dissipation" condition

$$\tilde{\epsilon}(\omega) \sim \epsilon(\omega) + \frac{i}{\lambda}\delta(\omega), \quad \lambda \to \infty. \tag{A4}$$

Here $\epsilon(\omega)$ and $\delta(\omega)$ are real valued for real ω. From (A3) and (A4) it follows that, for $\lambda \to \infty$,[26]

$$\tilde{n} \sim n + \frac{i}{\lambda} n', \quad n = \left(\frac{\omega^2\mu\epsilon}{c^2} - \kappa^2\right)^{\frac{1}{2}}, \quad n' = \frac{\omega^2\mu\delta}{2c^2 n}. \tag{A5}$$

Using (A5) and the Lemma 3 we can establish the following:

Lemma 4: If $\omega^2\mu\epsilon/c^2 > \kappa^2$, then $n' \geq 0$, $\operatorname{sgn} n = \operatorname{sgn} \omega$.

Lemma 5: If $\omega^2\mu\epsilon/c^2 < \kappa^2$, then $\operatorname{Im} n \geq 0$.

In Ref. 24 it is shown by physical arguments that, whenever $\operatorname{Re} \hat{\epsilon}(\hat{\omega}) \gg \operatorname{Im} \hat{\epsilon}(\hat{\omega})$ is a real frequency range, then the inequality

$$\frac{d}{d\hat{\omega}} [\hat{\omega} \operatorname{Re} \hat{\epsilon}(\hat{\omega})] > 0 \tag{A6}$$

must hold in that frequency range. We will assume that $\hat{\epsilon}(\hat{\omega})$ actually has that property. It then follows at once that, whenever $\epsilon(\omega) \gg \lambda^{-1}\delta(\omega)$ in a real ω interval, the inequality

$$\frac{d}{d\omega} [\omega\epsilon(\omega)] > 0$$

is satisfied. We now can establish the following:

Lemma 6: For $\omega^2\mu\epsilon/c^2 > \kappa^2$, $n_\omega \geq 0$.

[23] Garibian's notation is related to ours by the following relations: $\omega/v = k_z$, $n_- = -\lambda_1$, $n_+ = \lambda_2$, $n_\pm \epsilon_\mp + n_\mp \epsilon_\pm = -\zeta$, $(\omega^2/v^2) + \kappa^2 = k^2$. Also, he replaces our ω integration by a k_z integration using the relation $d\omega = v\, dk_z$.

[24] L. D. Landau and E. M. Lifshitz, Ref. 6.

[25] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon Press, New York, 1958).

[26] Again, we choose the square root such that $n(0) = i\kappa$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

On the Translation Invariance of Local Internal Symmetries

LAWRENCE J. LANDAU AND EYVIND H. WICHMANN
Department of Physics, University of California, Berkeley, California* 94720

(Received 29 May 1969)

The notion of a local internal symmetry is discussed with reference to local physical theories and tentative definitions are presented. It is shown that under fairly general assumptions commonly made about local theories a local internal symmetry commutes with all translations.

I. INTRODUCTION

In this paper we shall be concerned with a general property of *local* fundamental theories, such as the theory of local von Neumann algebras[1] or orthodox quantum field theory.[2] It is a characteristic assumption of such theories that all physical phenomena can be described in terms of *local* variables: i.e., physical variables associated with arbitrarily small regions in space–time, and such that the variables associated with any two spacelike separated regions are physically independent.

Consider a bounded region R in space–time. We might define a *local internal symmetry* G as a unitary transformation G which maps, by conjugation, the variables associated with the region R onto itself. It might then be conjectured that G will commute with all translations in the Poincaré group. We shall prove that this conjecture is indeed correct under some fairly general assumptions commonly made about local theories.

We shall first discuss this question within the framework of the theory of local von Neumann algebras and prove a theorem which expresses the translation invariance of a local internal symmetry. In order to show what the conclusion really depends on, we shall, at first, not assume more about the local theory than is necessary for the proof of the theorem. After that, we shall discuss additional conditions on local internal symmetries, which seem to us to be well motivated on physical grounds. Finally, we shall show how our methods can be modified to prove the translation invariance of a local internal symmetry within the framework of orthodox quantum field theory.

* Research supported partially by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant No. AF-AFOSR-68-1471.

[1] A. S. Wightman, Ann. Inst. Henri Poincaré 1, 403 (1964).

[2] R. F. Streater and A. S. Wightman, *PCT, Spin & Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964).

II. LOCAL INTERNAL SYMMETRIES IN THE THEORY OF LOCAL VON NEUMANN ALGEBRAS

We base our discussion on the following assumptions:

(i) Local algebras of operators: The physical states are described by vectors in a separable Hilbert space $\mathcal{H}$. With every bounded open set R in four-dimensional configuration space is associated a von Neumann algebra $B(R)$ of bounded operators on $\mathcal{H}$, i.e., $B(R)$ is a weakly closed *-algebra of bounded operators such that $B(R) = B(R)''$, where $B(R)''$ is the commutant of $B(R)'$, which is the commutant of $B(R)$. [The physical interpretation of $B(R)$ is that it is the von Neumann algebra generated by the operators which describe physical observables associated with R.]

(ii) Translation invariance: The Hilbert space $\mathcal{H}$ carries a strongly continuous unitary representation of the translation group, with elements

$$T(x) = \int e^{ix\cdot p}\, dE(p). \tag{1}$$

The set of local algebras transforms onto itself under conjugations by translations, i.e., for any bounded open set R,

$$T(x)B(R)T(x)^{-1} = B(R + x), \tag{2}$$

where $(R + x)$ denotes the set of all points $x + y$ with $y \in R$.

(iii) Spectral condition: The support of the spectral measure $E(p)$ in the common spectral resolution (1) of the translations is contained in the closed forward cone $\bar{V}_+$ in momentum space,

$$\text{supp}\,\{E(p)\} \subset \bar{V}_+ = \{p \mid p \cdot p \geq 0, p_4 \geq 0\}. \tag{3}$$

(iv) Cyclic vacuum state: There exists one and only one vector, $|\text{vac}\rangle$, which is invariant under translations. For every bounded open set R, $|\text{vac}\rangle$ is cyclic with respect to $B(R)$, i.e., the set of vectors $B(R)\,|\text{vac}\rangle$ is dense in $\mathcal{H}$.

(v) Locality: For every pair of bounded open sets R_1 and R_2 in 4-space which are spacelike separated from each other, all operators in $B(R_1)$ commute with all operators in $B(R_2)$, i.e.,

$$B(R_1) \subset B(R_2)', \quad \text{whenever } R_1 \subset R_2', \tag{4}$$

where we employ the notation R' to denote the interior of the set of all points which are spacelike with respect to R.

We shall now state and prove a theorem concerning the translation invariance of a unitary transformation G under two assumptions which we feel are appropriate for local internal symmetries.

Theorem 1: If G is a unitary operator on $\mathfrak{H}$ such that

$$G\,|\text{vac}\rangle = |\text{vac}\rangle \tag{5a}$$

and such that, for every pair of bounded open sets R_1 and R_2 in 4-space which are spacelike separated from each other, all operators in $GB(R_1)G^{-1}$ commute with all operators in $B(R_2)$, i.e., if

$$GB(R_1)G^{-1} \subset B(R_2)', \quad \text{whenever } R_1 \subset R_2', \tag{5b}$$

then G commutes with all translations

$$T(x)GT(x)^{-1} = G, \quad \text{for all } x. \tag{5c}$$

Proof: We give the proof in several steps.

(i) In our proof we will actually need condition (5b) only for regions R_1 and R_2 which are in the set of all regions obtained by translations of a fixed region R_0. The choice of R_0 is in fact immaterial, but for simplicity we shall make a definite choice as follows.

(ii) For any point x in 4-space we denote by $V_+(x)$ the interior of the forward light-cone with x as apex and by $V_-(x)$ the interior of the backward light-cone with x as apex. Let x_0 be a point such that $x_0 \in V_+(-x_0)$: i.e., the four-vector x_0 is forward timelike. We choose R_0 to be the nonempty bounded open set

$$R_0 = V_+(-x_0) \cap V_-(x_0), \tag{6a}$$

and we also define a larger bounded open set R_e by

$$R_e = V_+(-2x_0) \cap V_-(2x_0). \tag{6b}$$

With these definitions it follows that R_0 is spacelike separated from $(R_0 + x)$ whenever $x \in R_e'$.

(iii) Let a unitary operator G satisfy the premises of the theorem and let b_1 and b_2 be any two elements of $B(R_0)$. With the notation

$$G(x) = T(x)GT(x)^{-1}, \quad b(x) = T(x)bT(x)^{-1}, \tag{7}$$

we then have

$$[G(y)b_1G(y)^{-1}, b_2(x)] = 0, \quad \text{for any } x \in R_e', \quad \text{all } y. \tag{8}$$

(iv) We introduce the function

$$\begin{aligned} F(x, y) &= \langle \text{vac}|\ [G(y)b_1G(y)^{-1}, b_2(x)]\ |\text{vac}\rangle \\ &= \langle \text{vac}|\ b_1G(y)^{-1}b_2(x)\ |\text{vac}\rangle \\ &\quad - \langle \text{vac}|\ b_2(x)G(y)b_1\ |\text{vac}\rangle, \end{aligned} \tag{9}$$

which is bounded and continuous in both variables. The condition (8) implies that

$$F(x, y) = 0, \quad \text{for any } x \in R_e', \quad \text{all } y. \tag{10}$$

(v) As a motivation for what follows, let us state that the crucial step in the proof is the demonstration that the Fourier transform $\hat{F}(x, q)$ of $F(x, y)$ with respect to the variable y vanishes identically as a function of x whenever $q \neq 0$. That this is so will be seen to follow from the nature of the support of $\hat{F}(x, q)$ in x space [which is essentially determined by the local character of the transformation G as expressed by (10)] and the nature of the support of the Fourier transform (with respect to x) of $\hat{F}(x, q)$ in momentum space [which is essentially determined by q and by the spectral condition (3)]. Once this property of $\hat{F}(x, q)$ has been established, it readily follows that each term in the extreme right-hand side of (9) is independent of y for every x, which, in effect, expresses the translation invariance of G.

(vi) Let $q' \neq 0$ be a point in 4-space. Let $\eta(q')$ be a (nonzero) forward timelike 4-vector such that the origin in q space is not included in the set

$$V_+(q' - 2\eta(q')) \cap V_-(q' + 2\eta(q')).$$

Such an $\eta(q')$ can obviously always be found, given q'. In the following we shall abbreviate $\eta(q')$ by η, the above constraint being understood:

Let $K(q')$ be the neighborhood of q' defined by

$$K(q') = V_+(q' - \eta) \cap V_-(q' + \eta). \tag{11}$$

Let $h(q; q')$ be an infinitely differentiable function of q with support in $K(q')$. We denote its Fourier transform by $\tilde{h}(y; q')$, that is,

$$\tilde{h}(y; q') = \int d^4(q)h(q; q') \exp(iy \cdot q), \tag{12a}$$

and we define the function $F(x; h)$ by

$$F(x; h) = \int d^4(y)F(x, y)\tilde{h}(y; q'). \tag{12b}$$

We then have

$$F(x; h) = 0, \quad \text{for any } x \in R_e', \tag{12c}$$

which we can express in terms of the support of F as

$$\text{supp}\,\{F(x; h)\} \subset \bar{V}_+(-2x_0) \cup \bar{V}_-(2x_0), \tag{13}$$

where here and in the following $\bar{V}_+(x)$ denotes the closure of $V_+(x)$ and $\bar{V}_-(x)$ denotes the closure of $V_-(x)$.

(vii) Let us write

$$F(x;h) = F_1(x;h) - F_2(x;h), \tag{14a}$$

where

$$F_1(x;h) = \int d^4(y) \langle \text{vac}| b_1 G(y)^{-1} b_2(x) |\text{vac}\rangle \tilde{h}(y;q'), \tag{14b}$$

$$F_2(x;h) = \int d^4(y) \langle \text{vac}| b_2(x) G(y) b_1 |\text{vac}\rangle \tilde{h}(y;q'). \tag{14c}$$

The functions $F(x;h)$, $F_1(x;h)$, and $F_2(x;h)$ are continuous bounded functions of x and, hence, tempered distributions. Their Fourier transforms

$$\tilde{F}_{1,2}(p;h) = \int d^4(x) F_{1,2}(x;h) \exp(-ix\cdot p), \tag{15a}$$

$$\tilde{F}(p;h) = \tilde{F}_1(p;h) - \tilde{F}_2(p;h) \tag{15b}$$

will accordingly also be tempered distributions. Because of the spectral condition (3) and because the support of h is confined to $K(q')$, we obtain

$$\text{supp}\{\tilde{F}_1(p;h)\} \subset \bar{V}_+ \cap \bar{V}_+(q' - \eta), \tag{16a}$$

$$\text{supp}\{\tilde{F}_2(p;h)\} \subset \bar{V}_- \cap \bar{V}_-(q' + \eta), \tag{16b}$$

from which it readily follows that

$$\text{supp}\{\tilde{F}(p;h)\} \subset (\bar{V}_+ \cup \bar{V}_-(q' + \eta)) \cap (\bar{V}_- \cup \bar{V}_+(q' - \eta)). \tag{17}$$

(viii) We note that the condition (17) is, in general, less restrictive than the conditions (16). For our purpose it will be convenient to replace the condition (17) by a still weaker but simpler condition as follows: for every q' there exist two points q_+ and q_- such that

$$\text{supp}\{\tilde{F}(p;h)\} \subset \bar{V}_+(q_+) \cup \bar{V}_-(q_-) \tag{18a}$$

and such that

$$\bar{V}_+(q_+) \cap \bar{V}_-(q_-) = 0. \tag{18b}$$

To see that the support of $\tilde{F}(p;h)$ satisfies such a condition, we note that, by the selection of η in (vi) above, the origin in p space is not contained in the set $\bar{V}_+(q' - \eta) \cap \bar{V}_-(q' + \eta)$; this means that at least one of the sets $\bar{V}_+ \cap \bar{V}_-(q' + \eta)$ and $\bar{V}_+(q' - \eta) \cap \bar{V}_-$ is empty. If $\bar{V}_+ \cap \bar{V}_-(q' + \eta)$ is empty, we select $q_+ = 0$, $q_- = q' + \eta$; otherwise, we select $q_+ = q' - \eta$, $q_- = 0$. In both cases the condition (17) implies condition (18a) and condition (18b) follows from the choice of q_+ and q_-.

(ix) Let $\mathfrak{B}$ denote the complement of the set $\bar{V}_+(q_+) \cup \bar{V}_-(q_-)$. As a tempered distribution $\tilde{F}(p;h)$ accordingly vanishes in $\mathfrak{B}$. We shall now make use of a theorem proved by Araki,[3] which asserts that, given a region $\mathfrak{B}$ in which a tempered distribution [in our case $\tilde{F}(p;h)$] vanishes and given that the support of its Fourier transform $F(x;h)$ satisfies the condition (13), there exists a region $\hat{\mathfrak{B}}$, in general larger than $\mathfrak{B}$, in which the distribution $\tilde{F}(p;h)$ will also vanish. We shall show that on the basis of the principles presented in the quoted paper for the determination of $\hat{\mathfrak{B}}$ the region $\hat{\mathfrak{B}}$ will, in fact, be all of p space, i.e., the distribution $\tilde{F}(p;h) = 0$.

The ideas which underlie the discussion in Araki's paper can be briefly described as follows: 4-dimensional p space is interpreted as the plane $s = 0$ in a 5-dimensional pseudo-Euclidean space in which the Cartesian coordinates are (s, p) and in such a way that the s axis is of a spacelike character and orthogonal to the plane $s = 0$. Araki then shows that there is associated with every tempered distribution (in p space) which vanishes in an open region $\mathfrak{B}$ and whose Fourier transform satisfies the condition (13), an infinitely differentiable solution $f(s,p)$ of the wave equation in the 5-dimensional space, such that $f(s,p)$ and its normal derivative vanish on the plane $s = 0$ in a region $\mathfrak{B}_1$ which is contained in $\mathfrak{B}$ and which can be chosen to approximate $\mathfrak{B}$ arbitrarily closely. Furthermore: the given tempered distribution will vanish in every region on the plane $s = 0$ in which $f(s,p)$ and its normal derivative vanish. In view of the above, known uniqueness theorems for hyperbolic differential equations with constant coefficients can be exploited in the determination of the region $\hat{\mathfrak{B}}$: The region $\hat{\mathfrak{B}}$ is the region on the plane $s = 0$ in which every solution of the wave equation will vanish together with its normal derivative, provided it vanishes together with its normal derivative in the region $\mathfrak{B}$ on the plane $s = 0$. In particular, $\hat{\mathfrak{B}}$ will be all of p space if there exists a spacelike plane $\mathfrak{T}_5$ in the 5-dimensional (s, p) space, such that every solution of the wave equation vanishes together with its normal derivative on $\mathfrak{T}_5$ whenever the solution vanishes together with its normal derivative in the region $\mathfrak{B}$ on the plane $s = 0$. This is, in fact, the situation in our case, as can be seen as follows:

(x) In view of the conditions (18) there exists a spacelike plane $\mathfrak{T}$ in p space such that the cone $\bar{V}_+(q_+)$ lies on one side of $\mathfrak{T}$ and the cone $\bar{V}_-(q_-)$ lies

[3] H. Araki, Helv. Phys. Acta **36**, 132 (1963). We refer specifically to Lemma 8. It may be noted that in our discussion the roles of the position and momentum coordinates are reversed relative to their roles in Araki's paper.

on the other side and such that the plane $\mathfrak{P}$ passes through the point $\frac{1}{2}(q_+ + q_-)$. Let $\mathfrak{P}^5$ be the plane

$$\mathfrak{P}_5 = \{(s,p) \mid p \in \mathfrak{P}\}, \tag{19}$$

and let n be the unit normal to $\mathfrak{P}_5$. Let p' be any point of $\mathfrak{P}$ and let $\lambda > 0$ be such that the line segment $[p' - \lambda n, p' + \lambda n]$ lies in $\mathfrak{B}$. Now if a solution of the wave equation in (s,p) space vanishes together with its normal derivative in $\mathfrak{B}$ on $s = 0$, then the solution will vanish together with its normal derivative in the region

$$s^2 - (p - p')^2 < \lambda^2, \quad (s,p) \in \mathfrak{P}_5, \tag{20}$$

in the plane $\mathfrak{P}_5$. A trivial computation now shows that for every $(s,p) \in \mathfrak{P}_5$ there exist a $p' \in \mathfrak{P}$ and a corresponding λ, such that the inequality (20) is satisfied. [We can select p' such that it is "sufficiently large" and perpendicular to p and to the projection of $(q_+ - q_-)$ into $\mathfrak{P}$.] It follows that the solution of the wave equation will vanish together with its normal derivative everywhere on $\mathfrak{P}_5$ and, hence, everywhere in (s,p) space. The region $\hat{\mathfrak{B}}$ is thus all of p space, and the distribution $\tilde{F}(p;h)$ accordingly vanishes for all p.[4]

(xi) According to the relations (16) the supports of the two distributions $\tilde{F}_1(p;h)$ and $\tilde{F}_2(p;h)$ are disjoint in p space. Since we have shown that their difference $\tilde{F}(p;h)$ vanishes identically, it follows that each distribution separately must vanish identically. We thus have

$$\tilde{F}_2(p;h) = 0, \quad F_2(x;h) = 0, \quad \text{for all } x. \tag{21}$$

In view of the definition (14c) and the second relation (21), we then have

$$F_2(0;h) = \int d^4(y) \langle \text{vac} | \, b_2 G(y) b_1 \, | \text{vac} \rangle \, \tilde{h}(y;q') = 0. \tag{22}$$

(xii) Since the tempered distribution

$$\langle \text{vac} | \, b_2 G(y) b_1 \, | \text{vac} \rangle$$

is actually a continuous bounded function of y, it follows immediately from the validity of (22) for all $q' \neq 0$ that

$$\langle \text{vac} | \, b_2 G(y) b_1 \, | \text{vac} \rangle = \langle \text{vac} | \, b_2 G b_1 \, | \text{vac} \rangle. \tag{23}$$

Since b_1 and b_2 are arbitrary elements of $B(R_0)$ and since we have assumed that the set $B(R_0) \, | \text{vac} \rangle$ is dense in the Hilbert space $\mathfrak{H}$, it follows that

$$T(y) G T(y)^{-1} = G(y) = G, \tag{24}$$

which thus establishes the theorem.

III. DISCUSSION

It is clear from the proof of this theorem that the assumptions can be relaxed. It suffices, for instance, that for a fixed region R_0 the vectors $B(R_0) \, | \text{vac} \rangle$ are dense in $\mathfrak{H}$ and that the condition (5b) holds only for all regions R_1 and R_2 obtained by translations of R_0. Such a sharpening of the theorem does not seem to increase its physical interest.

The condition (5a) that the vacuum state vector is invariant under G cannot, however, be omitted. A simple counterexample is provided by the case of the system of algebras generated by a single Hermitian free scalar field $\phi(x)$: for a real test function $r(x)$ the unitary transformation $\exp(i\phi[r])$ (which does not preserve the vacuum state) satisfies the condition (5b), but it obviously does not commute with all translations.

In a realistic local theory we require invariance under (the identity component of) the entire Poincaré group $\bar{\mathcal{G}}$ rather than under the subgroup of translations. We should therefore replace the postulate (ii) by the following:

(ii)′ *Poincaré invariance:* The Hilbert space $\mathfrak{H}$ carries a strongly continuous unitary representation $U(\lambda)$ of the Poincare group $\bar{\mathcal{G}}$. The set of local algebras transforms onto itself under conjugation by $U(\lambda)$ as follows:

$$U(\lambda) B(R) U(\lambda)^{-1} = B(R_\lambda) \tag{25}$$

for every $\lambda \in \bar{\mathcal{G}}$, for every bounded open set R, and where R_λ is the image of R under the action of λ.

The conditions on G in the theorem are minimum conditions which a "local (internal)" symmetry ought to satisfy. We shall here present a definition.

Tentative definition: A local internal symmetry G is a unitary operator G on the Hilbert space $\mathfrak{H}$ such that

$$G \, | \text{vac} \rangle = | \text{vac} \rangle \tag{26a}$$

and such that

$$G B(R) G^{-1} = B(R) \tag{26b}$$

for any bounded open set R.

The conditions (26) are apparently stronger in general than the assumptions in Theorem 1. With this definition of a local internal symmetry we have another theorem.

[4] A very similar situation concerning the supports of a tempered distribution and its Fourier transform has been discussed by one of us elsewhere: L. J. Landau, Commun. Math. Phys. 13, 246 (1969).

Theorem 2:

(a) The set $\mathfrak{G}$ of all local internal symmetries form a group. If G is in $\mathfrak{G}$, so is $U(\lambda)GU(\lambda)^{-1}$ for all $\lambda \in \bar{\mathcal{G}}$: the group $\mathfrak{G}$ is mapped onto itself under conjugation by any element of the Poincaré group $\bar{\mathcal{G}}$.

(b) Every local internal symmetry commutes with every translation, that is,

$$T(x)GT(x)^{-1} = G, \quad \text{for all } G \in \mathfrak{G} \text{ and all } x.$$

The first assertion of this theorem is trivial; the second corresponds to our earlier Theorem 1. In view of these theorems, we feel that our definition above is a reasonable mathematical expression of the physical meaning of the phrase "local internal symmetry." [5]

The ideas presented so far must be modified somewhat if we deal with a system of local algebras in the presence of superselection principles. In this case, the vacuum vector will not be cyclic with respect to $B(R)$, if $B(R)$ is the von Neumann algebra of local *observables* associated with R. One approach to this situation,[6] which is in the spirit of a local basic theory, is to consider a system $\{F(R)\}$ of "local field algebras" $F(R)$. Here $F(R)$ will contain the von Neumann algebra of local observables $B(R)$, but $F(R)$ will also contain elements which change the superselection quantum numbers, such that the vacuum state vector is cyclic with respect to $F(R)$. The field algebra $F(R)$ is assumed to be generated by elements each one of which has either a boson or a fermion character, subject to the familiar constraints by which locality is expressed, namely, that commutators or anticommutators vanish for spacelike separations. It should be clear that, under appropriate assumptions about the local internal symmetry G, theorems analogous to our Theorems 1 and 2 can be proved with minor modifications of our procedure. We shall not pursue the matter further here, but instead consider the case of a local symmetry in a conventional field theory.

IV. LOCAL INTERNAL SYMMETRIES IN QUANTUM FIELD THEORY

We shall here discuss an analog to our Theorem 1 within the framework of an orthodox quantum field theory. Since the method of proof in the case of a field theory is essentially the same as in the case of the theory of local von Neumann algebras, it will suffice to outline the procedure.

We shall base our discussion of field theory on the well-known axioms formulated by Wightman.[7] In such a theory we have a set of operator-valued tempered distributions $\varphi_i(x)$, $i = 1, 2, \ldots, n$, where we assume, for simplicity and without loss of generality, that the adjoint of each member of the set is contained in the set. The set of all physical states is, as before, associated with the vectors in a separable Hilbert space $\mathcal{H}$, which carries a strongly continuous unitary representation of the Poincaré group. There exists a dense set D of vectors in $\mathcal{H}$ such that $\langle\alpha|\ \varphi_i(x)\ |\beta\rangle$ is a tempered distribution for each $i = 1, \cdots, n$ and for all vectors $\alpha, \beta \in D$. Furthermore, $\varphi_i[f]\ D \subset D$ for every $f = f(x)$ in the set of test functions $\mathcal{S}$, where $\varphi_i[f]$ denotes the quantum field $\varphi_i(x)$ averaged by f. As before, there exists a unique vacuum state invariant under all transformations in the Poincaré group. The vacuum state $|\text{vac}\rangle$ is assumed to be contained in D, and it is further assumed that $|\text{vac}\rangle$ is a cyclic vector with respect to the polynomial algebra generated by the operators $\varphi_i[f]$. The spectral condition (3) is assumed for the translations, which satisfy the condition

$$T(y)\varphi_i(x)T(y)^{-1} = \varphi_i(x + y). \tag{27}$$

The condition of locality is expressed through the relations

$$\varphi_i(x)\varphi_j(y) = \sigma_{ij}\varphi_j(y)\varphi_i(x), \quad \text{for } (x - y)^2 < 0, \tag{28}$$

where the constants σ_{ij} have the values $+1$ or -1 depending on the boson or fermion character of the fields involved.

Within this framework we state the result.

Theorem 3: If a unitary transformation G on $\mathcal{H}$ satisfies the conditions

$$G\,|\text{vac}\rangle = |\text{vac}\rangle, \tag{29a}$$

$$GD = D, \tag{29b}$$

and

$$G\varphi_i(x)G^{-1}\varphi_j(y) = \sigma_{ij}\varphi_j(y)G\varphi_i(x)G^{-1}, \quad \text{whenever } (x - y)^2 < 0, \tag{29c}$$

then G commutes with all translations, that is,

$$T(x)GT(x)^{-1} = G, \quad \text{for all } x.$$

[5] For related discussions of internal symmetries we refer to R. F. Streater, "Spontaneous Breakdown of Symmetry," in *Mathematical Theory of Elementary Particles*, R. Goodman and I. Segal, Eds. (MIT Press, Cambridge, 1966) and to D. W. Robinson, "Symmetries, Broken Symmetries, Currents and Charges," in *Symmetry Principles and Fundamental Particles*, B. Kursunoglu and A. Perlmutter, Eds. (W. H. Freeman and Company, San Francisco, 1967).

[6] For discussions of this approach see H. J. Borchers, Commun. Math. Phys. 1, 281 (1965) and also R. F. Streater, Ref. 5.

[7] Specifically, we shall adhere to the formulation in Chapter 3 of Ref. 2.

Proof:

(i) Let R_0 be the open region defined in (6a). We consider arbitrary monomials b in the averaged fields of the form

$$b = \varphi_{i(1)}[f_1]\varphi_{i(2)}[f_2]\cdots\varphi_{i(k)}[f_k], \tag{31}$$

where the indices $i(r)$ assume values in the set $\{1, 2, \cdots, n\}$ and where each one of the test functions f_r has its support confined to R_0. Let b_1 and b_2 be two such monomials. With the notation in (7) we then have, in place of the condition (8),

$$G(y)b_1G(y)^{-1}b_2(x) = \sigma b_2(x)G(y)b_1G(y)^{-1}, \quad \text{for any } x \in R_e' \text{ and all } y, \tag{32}$$

where the factor σ is equal to either $+1$ or -1, depending on the fields which appear in the monomials b_1 and b_2 and on the basis sign factors σ_{ij}.

(ii) The appearance of the sign factor σ does not affect the essential arguments in the proof of Theorem 1, and we immediately arrive at the analog of the conclusion (23), namely

$$\langle \text{vac}|\, b_2G(y)b_1\,|\text{vac}\rangle = \langle \text{vac}|\, b_2Gb_1\,|\text{vac}\rangle, \tag{33}$$

for arbitrary monomials b_1 and b_2 constructed as above. We now refer to a theorem of Reeh and Schlieder[8] which asserts that the vacuum state vector is cyclic with respect to the polynomial algebra generated by all operators $\varphi_i[f]$, where the test functions f have their support confined to an arbitrary open set, which in our case will be the set R_0. The linear span of the vector $|\text{vac}\rangle$ and all vectors $b\,|\text{vac}\rangle$, where b is any monomial of the form (31), is accordingly dense in $\mathcal{H}$, and it follows immediately from (33) that

$$G(y) = T(y)GT(y)^{-1} = G,$$

as asserted.

ACKNOWLEDGMENTS

It is a pleasure to thank I. Herbst and P. Mohr for helpful discussions of the issues in this paper.

[8] H. Reeh and S. Schlieder, Nuovo Cimento **22**, 1051 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Wavepacket Volkov Solutions*

JAMES W. MEYER
Department of Physics, Purdue University, Lafayette, Indiana 47907

(Received 7 April 1969)

The problem of a scalar particle interacting with a planewave external field is solved in terms of the complete set of normalized in-states which are present in the asymptotic past. By expressing the field operator in terms of the annihilation and creation operators for the in-states, we show that the normal LSZ asymptotic condition holds allowing an unambiguous interpretation of the wavefunctions, i.e., in the sharp-momentum limit of the wavepacket Volkov solutions, the momentum parameter appearing in the solution is the momentum of the particle before the particle entered the beam. A brief discussion on how to include other interactions is also presented.

There has been some discussion[1-3] in the literature as to the correct interpretation of the Volkov-type solutions[4] which occur in the problem of a charged particle interacting with an external planewave electromagnetic field. In this paper we introduce the asymptotic in- and out-states and solve for the field operator in terms of the annihilation and creation operators of these asymptotic states. In order to simplify the problem, we work with a scalar charged particle and a scalar "photon," since the vector nature of the photon is not relevant. For this field operator we prove an LSZ-type asymptotic condition and this will give us a clear interpretation of the solution since the asymptotically free states are physical states before and after the charged particle interacts with the external field.

The method used in this paper was motivated by the work of Capri, as reported by Wightman.[5] Capri's work was directed toward constructing the nonperturbative solution for higher-spin fields in the presence of an external electromagnetic field which vanishes rapidly in both spacelike and timelike directions. Besides studying only a scalar field in the presence of a scalar external field, we demand only that the external field vanish rapidly in a lightlike direction. The cutoff in both the timelike and spacelike directions is then obtained by studying only the matrix elements of field operator between normalized states. This procedure gives us only weak convergence for the field operator instead of the strong convergence obtained by Capri, but has the advantage that the formal operator obtained is essentially covariant.

The equation of motion for a charged particle interacting with a planewave external electromagnetic field is

$$(-\partial^2 + M^2)\varphi(x) = -2ieA_\mu\partial_\mu\varphi - e^2A_\mu^2\varphi.$$

The term quadratic in A_μ creates an apparent mass shift, for, if we take $A_\mu \sim \cos k \cdot x$, then $A_\mu^2 \sim \frac{1}{2}(1 + \cos 2k \cdot x)$ and the constant term has the form of a mass term. The same apparent mass shift occurs if a scalar external field interacts with a charged particle with the equation of motion[6]

$$[-\partial^2 + m^2 + e^2A^2(x)]\varphi(x) = 0. \tag{1}$$

Our model will have the equation of motion given in (1), and in order to solve it we take the external field A to be a function of $n \cdot x$ alone, where

$$n_\mu = (\mathbf{n}, \mathrm{i}n_0) \quad \text{and} \quad n^2 = 0.$$

Normally we take $\mathbf{n}^2 = n_0^2 = 1$, since this condition just specifies the frequency in the particular reference frame in which we are working. A is further required to be in the form of a wavepacket, i.e., A and all of its derivatives approach zero faster than any inverse power of $n \cdot x$ as $|n \cdot x|$ goes to infinity. The Fourier transform of A will be peaked about a frequency ω_0. One example of an A satisfying these properties is

$$A(n \cdot x) = ae^{i\omega_0 n\cdot x}e^{-\epsilon^2(n\cdot x)^2},$$

but, since we will nowhere need an explicit form for A, we will not choose one.

In order to solve (1), we introduce a complete set of in-states, obtained from the vacuum using the

* Work supported by the U.S. Atomic Energy Commission.

[1] T. W. B. Kibble, Phys. Rev. **138**, B740 (1965); Lee M. Frantz, Phys. Rev. **139**, 1326 (1965).

[2] Oldwig von Roos, Phys. Rev. **150**, 1112 (1966); Z. Fried, A. Baker, and D. Korff, Phys. Rev. **151**, 1040 (1966).

[3] P. Stehle and P. G. DeBaryshe, Phys. Rev. **152**, 1135 (1966).

[4] D. M. Volkov, Z. Physik **94**, 250 (1935).

[5] A. Capri, thesis, Princeton University, 1967; A. S. Wightman, in *Proceedings of the 1967 International Conference of Particles and Fields*, C. R. Hagen, C. Guralink, and V. S. Mathur, Eds., (Interscience Publishers, Inc., New York, 1967), p. 208. See also A. Capri, J. Math. Phys. **10**, 575 (1969).

[6] Our metric is such that $p \cdot x = \mathbf{p} \cdot \mathbf{x} - p_0x_0$, $p_\mu = (\mathbf{p}, ip_0)$, $\partial^2 = \nabla^2 - \partial_t^2$.

creation operators[7]

$$a_\alpha^\dagger = -i\int d^3x f_\alpha(x)\overleftrightarrow{\partial_t}\varphi_{\text{in}}^\dagger(x), \tag{2}$$

where f_α's are a complete set of normalized solutions to the free-field Klein–Gordon equation

$$(-\partial^2 + m^2)f_\alpha = 0 \quad \text{and} \quad i\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t} f_\beta(x) = \delta_{\alpha\beta}.$$

The a's satisfy the comutation relation

$$[a_\alpha, a_\beta^\dagger] = \delta_{\alpha\beta}.$$

The state $|\alpha\rangle = a_\alpha^\dagger|0\rangle$ represents the state which in the remote past is a free particle with momentum-space wavefunction $f_\alpha(\mathbf{p})$, where

$$f_\alpha(x) = (2\pi)^{-\frac{3}{2}}\int d^4p\theta(p_0)\delta(p^2 + m^2)f_\alpha(\mathbf{p})e^{ip\cdot x}$$

and

$$\int d^4p\theta(p_0)\delta(p^2 + m^2)f_\alpha^*(\mathbf{p})f_\beta(\mathbf{p}) = \delta_{\alpha\beta}.$$

In terms of the a's, φ_{in} can be written as

$$\varphi_{\text{in}}(x) = \sum_\alpha \{a_\alpha f_\alpha(x) + b_\alpha^\dagger f_\alpha^*(x)\}, \tag{3}$$

where b_α annihilates the antiparticles and the sum over α in this expression for φ_{in} is a formal sum with the interpretation that for any two normalized states $|1\rangle$ and $|2\rangle$ we have

$$\langle 1|\sum_\alpha O_\alpha|2\rangle \equiv \sum_\alpha \langle 1|O_\alpha|2\rangle. \tag{4}$$

In particular, if the states $|1\rangle$ and $|2\rangle$ are members of a complete set of states formed by applying the $a_\alpha^\dagger$'s to the vacuum, the sum over α in (4) will be over a finite number of terms.

We now assert that the solution to (1) satisfying the proper boundary conditions is, in terms of these in operators,

$$\varphi(x) = \sum_\alpha \{F_\alpha(x)a_\alpha + F_\alpha^*(x)b_\alpha^\dagger\}, \tag{5}$$

with

$$F_\alpha(x) \equiv f_\alpha(x) - \Lambda_\alpha(x) \tag{6}$$

$$= (2\pi)^{-\frac{3}{2}}\int d^4p\delta(p^2 + m^2)\theta(p_0)f_\alpha(\mathbf{p}) \times \exp\left(ip\cdot x - i\int_\infty^v dy' J_p(y')\right), \tag{7}$$

where

$$\Lambda_\alpha(x) \equiv e^2\int d_4x' f_\alpha(x')A^2(y') \times \Delta_R(x - x', m^2 + M^2(y, y')), \tag{8}$$

$$J_p(n\cdot x) \equiv (e^2/2n\cdot p)A^2(n\cdot x), \tag{9}$$

and $y = n\cdot x$, $y' = n\cdot x'$. The sum over α is interpreted as in (4), and Δ_R is the usual free-field retarded commutator but with a mass $m^2 + M^2$ where

$$M^2(y, y') \equiv \frac{e^2}{y - y'}\int_{v'}^{v} dy'' A^2(y'') \geq 0. \tag{10}$$

We note that M^2 is not only positive, but also bounded above by $\max[e^2A^2] \geq M^2$. In Appendix B we show the equivalence of the forms (6) and (7) for F.[8] That (5) is indeed a solution to (1) can be obtained directly from (7). Here we will use the form (6) for F since it is not familiar in the literature. In order to do so we will first show that $\Delta_R(x - x', m^2 + M^2)$ is the retarded Green's function for (1), namely

$$(-\partial^2 + m^2 + e^2A^2)\Delta_R(x - x', m^2 + M^2) = \delta^4(x - x'). \tag{11}$$

This type of solution for the Green's function was obtained by Brown and Kibble.[9] Here we will prove it directly. The usual Fourier transform for the free-field retarded commutator gives the integral representation for Δ_R (but not its Fourier transform since M^2 still depends upon y, y'):

$$\Delta_R(x - x', m^2 + M^2) = \frac{1}{(2\pi)^4}\int_{C_R} d^4p \frac{e^{ip\cdot(x-x')}}{p^2 + m^2 + M^2}. \tag{12}$$

Equation (12) gives

$$-\partial^2\Delta_R = \frac{1}{(2\pi)^4}\int d^4p \frac{e^{ip\cdot(x-x')}}{p^2 + m^2 + M^2} \times \left\{p^2 - \frac{2\,in\cdot p(\partial M^2/\partial y)}{p^2 + m^2 + M^2}\right\}, \tag{13}$$

since $\partial_\mu g(y) = n_\mu(\partial/\partial y)g(y)$. Now

$$\frac{1}{(p^2 + m^2 + M^2)^2} = -\frac{n_\mu}{2n\cdot p}\frac{\partial}{\partial p_\mu}\frac{1}{(p^2 + m^2 + M^2)}, \tag{14}$$

so that an integration by parts in the second term of (13) gives

$$-\partial^2\Delta_R = \frac{1}{(2\pi)^4}\int d^4p \frac{e^{ip\cdot(x-x')}}{p^2 + m^2 + M^2} \times \left\{p^2 - (y - y')\frac{\partial M^2}{\partial y}\right\}. \tag{15}$$

From (10) we have

$$\frac{\partial M^2}{\partial y} = \frac{1}{y - y'}[e^2A^2(y) - M^2]$$

[7] For the present we will use only in operators and will not label them as "in." When we later use out-states they will be explictly labeled as a^{out}, etc.

[8] If in (7) we let $f \to \delta^{(3)}(p - q)(2\pi)^{\frac{3}{2}}/2p_0$, we obtain the normal Volkov-type solution.

[9] L. S. Brown and T. W. B. Kibble, Phys. Rev. 133, A705 (1964) (especially the appendix).

so that (15) becomes

$$-\partial^2\Delta_R = \frac{1}{(2\pi)^4}\int d^4p\,\frac{e^{ip\cdot(x-x')}}{p^2+m^2+M^2} \times [p^2+m^2+M^2-m^2-e^2A^2(y)] = \delta^4(x-x') - (m^2+e^2A^2)\Delta_R,$$

which proves (11). Since C_R in (12) avoids the poles in the p_0 plane by going above them, we have

$$\Delta_R(x-x', m^2+M^2) = 0, \quad \text{for } x_0 - x_0' < 0,$$

so that Δ_R is, indeed, the retarded Green's function for (1).

Since $\varphi(x)$ given by (5) is to be used only between normalized states and this effectively makes the sum over α finite or absolutely convergent, we need only show that each term in (5) satisfies (1). Using the form (6) for F_α we have from (11) that

$$[-\partial^2 + m^2 + e^2A^2(y)]F_\alpha(x) = (-\partial^2 + m^2 + e^2A^2)f_\alpha - e^2A^2f_\alpha = 0, \quad (16)$$

since $(-\partial^2 + m^2)f_\alpha = 0$.

We are left with showing that the boundary conditions are satisfied. In Appendix A we show that the product $f_\alpha(x)A^2(y)$ is a "good" function in the sense that it and all of its derivatives decrease faster than any inverse power of $\|x\| \equiv \mathbf{x}^2 + x_0^2$, as $\|x\| \to \infty$.[10] In particular, since Δ_R is a distribution, this means that the integral defining Λ exists in the distribution sense. Further, since Δ_R is nonzero as a function of x' only in the past light cone about the point x, if we let $x_0 \to -\infty$, then the only contribution to the integral comes from large negative x_0' where the product fA^2 decreases faster than $C_N/(x_0')^N$, N arbitrary. Hence we have $\Lambda_\alpha(x) \to 0$, as $x_0 \to -\infty$, and

$$F_\alpha(x) \xrightarrow[x_0\to-\infty]{} f_\alpha(x). \quad (17)$$

The limit in (17) is obtained faster than any inverse power of x_0. Since derivatives of Δ_R with respect to x are zero for $x_0 - x_0' < 0$, similar limits as in (17) will hold for any derivative of F_α. In particular, this gives

$$i\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t}F_\beta(x) = \delta_{\alpha\beta} - ie^2\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t}\Lambda_\alpha(x) \xrightarrow[x_0\to-\infty]{} \delta_{\alpha\beta}, \quad (18)$$

since f_α^* is bounded by $C(\mathbf{x})\,|x_0|^{\frac{3}{2}}$, as $|x_0| \to \infty$. Thus remembering that we are using φ only between normalized states, we have

$$\lim_{x_0\to-\infty} i\int d^3x F_\alpha^*(x)\overleftrightarrow{\partial_t}\varphi(x) = a_\alpha, \quad (19)$$

which is the usual LSZ asymptotic condition.[11]

Another useful formula that can be obtained is

$$\Sigma_{\alpha\beta} \equiv i\int d^3x F_\alpha^*(x)\overleftrightarrow{\partial_t}F_\beta(x) = \delta_{\alpha\beta}, \quad x_0 \text{ arbitrary}. \quad (20)$$

If we show that $\Sigma_{\alpha\beta}$ is independent of time then the result follows immediately by taking $x_0 \to -\infty$. Since F_α satisfies (16), we have

$$\partial_t\Sigma_{\alpha\beta} = i\int d^3x[(-\partial_t^2F_\alpha^*)F_\beta + F_\alpha^*\partial_t^2F_\beta] = i\int d^3x[(-\nabla^2 + m^2 + e^2A^2)F_\alpha^*F_\beta - F_\alpha^*(-\nabla^2 + m^2 + e^2A^2)F_\beta] = 0$$

by integrating by parts, where this is possible since both f_α and Λ decrease strongly in a spacelike direction, the spacelike decrease for Λ following in the same manner as for $x_0 \to -\infty$. This proves (20).

So far we have only used a complete set of in-states. We now introduce a complete set of out-states with the annihilation operators

$$a_\alpha^{\text{out}} = i\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t}\varphi_{\text{out}}(x). \quad (21)$$

Then we can write the solution for $\varphi(x)$ of (1) as

$$\varphi(x) = \sum_\alpha \{G_\alpha(x)a_\alpha^{\text{out}} + G_\alpha^*(x)b^{\dagger\text{out}}\} \quad (22)$$

with

$$G_\alpha(x) = f_\alpha(x) - \Lambda_\alpha^{\text{out}}(x) \quad (23)$$

$$= (2\pi)^{-\frac{3}{2}}\int d^4p\,\delta(p^2+m^2)\theta(p_0)f_\alpha(\mathbf{p}) \times \exp\left(ip\cdot x - i\int_{-\infty}^{y} dy' J_{p'}(y')\right), \quad (24)$$

where

$$\Lambda_\alpha^{\text{out}}(x) = e^2\int d^4x' A^2(y')f_\alpha(x') \times \Delta_A[x-x', m^2+M^2(y,y')], \quad (25)$$

where Δ_A is the free-field advanced commutator [Eq. (12) with C_R replaced by C_A]. Then, in the same manner as above, we have

$$G_\alpha(x) \xrightarrow[x_0\to\infty]{} f_\alpha(x), \quad (26)$$

$$\lim_{x_0\to\infty} i\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t}\varphi(x) = a_\alpha^{\text{out}}, \quad (27)$$

[10] In mathematical language, $fA^2 \in S_4$.

[11] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento 1, 205 (1955).

and

$$i\int d^3x G_\alpha^*(x)\overleftrightarrow{\partial_t} G_\beta(x) = \delta_{\alpha\beta}. \tag{28}$$

Equation (28) gives us a_α^{out} in terms of the a's ($= a^{\text{in}}$'s):

$$\begin{aligned} a_\alpha^{\text{out}} &= i\int d^3x G_\alpha^*(x)\overleftrightarrow{\partial_t}\varphi(x) \\ &= \sum_\beta a_\beta i\int d^3x G_\alpha^*(x)\overleftrightarrow{\partial_t} F_\beta(x) \qquad (29) \\ &\equiv \sum_\beta a_\beta S_{\alpha\beta}. \qquad (30) \end{aligned}$$

The $S_{\alpha\beta}$ defined by (29) and (30) are independent of x_0 in the same way that $\Sigma_{\alpha\beta}$ above was. Since $G_\alpha^* \to f_\alpha^*$ as $x_0 \to \infty$, in order to evaluate $S_{\alpha\beta}$ we need

$$\Gamma_{\alpha\beta} \equiv \lim_{x_0\to\infty} i\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t} F_\beta(x). \tag{31}$$

This is done in Appendix C, with the result

$$\begin{aligned} \Gamma_{\alpha\beta} = S_{\alpha\beta} = &\int d^4p\,\delta(p^2 + m^2)\theta(p_0) f_\alpha^*(\mathbf{p}) f_\beta(\mathbf{p}) \\ &\times \exp\left(-i\int_\infty^{-\infty} dy' J_p(y')\right). \end{aligned} \tag{32}$$

This result would follow immediately if we could take $y \to -\infty$ ($x_0 \to \infty$) in the upper limit of the integral over $J_p(y')$ in the expression (7) for F. Knowing the constants $S_{\alpha\beta}$ then allows us to easily calculate S-matrix elements. For example, the single-particle-to-single-particle S-matrix element is given by

$$\begin{aligned} \langle \alpha \text{ out} \,|\, \beta \text{ in}\rangle &= \langle 0|\, a_\alpha^{\text{out}}\,|\beta \text{ in}\rangle \\ &= \sum_{\beta'} S_{\alpha\beta'}\langle 0|\, a_{\beta'}\,|\beta \text{ in}\rangle = S_{\alpha\beta}. \end{aligned} \tag{33}$$

Since a_{out} is linear in the a's, a more complicated S-matrix element will be a simple composition of the above $S_{\alpha\beta}$'s. This then completely solves the dynamical part of the theory.

To extend the above to the problem to vector photons instead of scalar photons, the work of Brown and Kibble[9] shows us that we have only to make the replacement

$$J_p(y) \to (2n\cdot p)^{-1}[-2ep_\mu A_\mu + e^2A_\mu^2] \equiv I_p(y)$$

in (7) for F_α. The other formulas such as (8) and (20) become more complicated but retain the same structure. With this replacement the ordinary Volkov solution to the problem of an external planewave electromagnetic field interacting with a charged particle can be properly interpreted, namely that the wavefunction $f_\alpha(\mathbf{p})$ which appears in $F_\alpha(x)$ as in (7) is the momentum-space wavefunction of the particle before it entered the external field, and, similarly, the f in G as in (24) is the momentum-space wavefunction of the particle after it leaves the external field. This is where the momentum of a particle is usually measured in a scattering experiment. In particular, the vacuum-to-one-particle expectation values of φ are just

$$\langle 0|\,\varphi(x)\,|\alpha \text{ in}\rangle = F_\alpha(x) \xrightarrow[x_0\to-\infty]{} f_\alpha(x)$$

and

$$\langle 0|\,\varphi(x)\,|\alpha \text{ out}\rangle = G_\alpha(x) \xrightarrow[x_0\to\infty]{} f_\alpha(x),$$

in agreement with the results of Brown and Kibble, if we further take the sharp-momentum limit

$$f_\alpha(\mathbf{p}) \to 2p_0\delta^3(\mathbf{p} - \mathbf{q}).$$

One more generalization that is of interest, since the problem above is rather trivial, is to a field satisfying the equation of motion

$$(-\partial^2 + m^2 + e^2A^2)\Psi = j. \tag{34}$$

The usual LSZ asymptotic condition is

$$\lim_{x_0\to-\infty} -i\int d^3x f_\alpha(x)\overleftrightarrow{\partial_t}\Psi^\dagger(x) = a_\alpha^{\dagger\text{in}}. \tag{35}$$

Equation (17) lets us write this as

$$\lim_{x_0\to-\infty} -i\int d^3x F_\alpha(x)\overleftrightarrow{\partial_t}\Psi^\dagger(x) = a_\alpha^{\dagger\text{in}}. \tag{36}$$

Then the usual LSZ reduction procedure leads to

$$\begin{aligned} a_\alpha^{\dagger\text{in}} = &\lim_{x_0\to\infty} -i\int d^3x F_\alpha(x)\overleftrightarrow{\partial_t}\Psi^\dagger(x) \\ &+ i\int d^4x F_\alpha(x)[-\partial^2 + m^2 + e^2A^2]\Psi^\dagger(x). \end{aligned} \tag{37}$$

The first term in (37) is easily found to be

$$\sum_\beta a_\beta^{\dagger\text{out}} S_{\beta\alpha}$$

and represents the scattering which would take place if j were equal to zero. Equation (37) is now in a form that allows the external field to be taken into account exactly and the extra interaction described by j to be treated as a perturbation, since (37) can be written

$$a_\alpha^{\dagger\text{in}} = \sum_\beta a_\beta^{\dagger\text{out}} S_{\beta\alpha} + i\int d^4x F_\alpha(x) j^\dagger(x). \tag{38}$$

Note that in (38) $a_\alpha^{\dagger\text{out}}$ is not the out operator introduced earlier but is the out-state resulting from the total interaction described by (34). Also note that (38) does not totally separate out the external field interaction since j will in general still depend upon the external field.

APPENDIX A

In this appendix we wish to show that $A^2(y)f_\alpha(x)$ is a "good" function ($\in S_4$). Dropping the α we have,

for the Fourier transform,

$$\begin{aligned}H(p) &\equiv \frac{(2\pi)^{\frac{3}{2}}}{(2\pi)^4}\int d^4x e^{-ip\cdot x}f(x)A^2(y)\\ &= [(2\pi)^4]^{-1}\int d^4x e^{-ip\cdot x}\\ &\quad\times\int d^4p'\delta(p'^2+m^2)\theta(p_0')f(\mathbf{p}')e^{ip'\cdot x}\\ &\quad\times\int d\omega I(\omega)e^{i\omega n\cdot x}\\ &=\int d\omega\delta[(p-\omega n)^2+m^2]\\ &\quad\times\theta(p_0-\omega)I(\omega)f(\mathbf{p}-\omega n)\\ &=\theta(p_0-\gamma)I(\gamma)f(\mathbf{p}-\gamma\mathbf{n})/2\,|n\cdot p|,\end{aligned}$$

where

$$A^2(y)=\int d\omega e^{i\omega v}I(\omega)$$

and

$$\gamma=(p^2+m^2)/2n\cdot p.$$

Let $p=p_\perp+\alpha n+\beta n^*$, $n^*=(\mathbf{n},-in_0)$, and $\mathbf{p}_\perp=\mathbf{p}-(\mathbf{n}/n_0^2)\mathbf{n}\cdot\mathbf{p}$; then

$$\gamma=\alpha+(p_\perp^2+m^2)/4\beta$$

and

$$H(p)=\theta(-\beta)I\left(\alpha+\frac{p_\perp^2+m^2}{4\beta}\right)\times f\left(\mathbf{p}_\perp,\beta-\frac{p_\perp^2+m^2}{4\beta}\right)\Big/(-4\beta).$$

Thus, since both I and f are "good" functions we need to show that H has left derivatives with respect to β of any order at $\beta=0$ and that

$$\lim_{\beta\to 0-}\left(\frac{\partial}{\partial\beta}\right)^n H=0,$$

so that they will match continuously onto $H(p)\equiv 0$ for $\beta>0$. But this follows immediately since both I and f and all their derivatives have infinite zeros at $\beta=0$ (their respective arguments large), since, for example, $|I(\omega)|<(C_N/\omega^N)$, ω large. Thus, since the Fourier transform of a "good" function is "good," we have our desired result.

APPENDIX B

In this appendix we wish to show the equivalence of the two forms for $F_\alpha(x)$ used in the text. We do this by calculating $\Lambda_\alpha(x)$:

$$\begin{aligned}\Lambda(x)&=e^2\int d^4x'A^2(y')\Delta_R(x-x',m^2+M^2)f(x')\\ &=\int\frac{d^4p'}{(2\pi)^{\frac{3}{2}}}\theta(p_0')\delta(p'^2+m^2)f(\mathbf{p}')\Gamma(p',x),\end{aligned}$$

with

$$\Gamma(p',x)=e^2\int\frac{d^4p}{(2\pi)^4}e^{ip\cdot x}\times\int d^4x'A^2(y')\frac{e^{i(p'-p)\cdot x'}}{[p^2+m^2+M^2-i\epsilon\epsilon(p_0)]}.$$

Let

$$x'=x_\perp'+\sigma n+\tfrac{1}{2}\rho n^*,\quad p=p_\perp+\alpha n+\beta n^*,$$

then

$$d^4x'=d^2x_\perp'\,d\sigma\,d\rho\quad\text{and}\quad d^4p=2\,d^2p_\perp\,d\alpha\,d\beta.$$

Γ becomes, after doing the $d^2x_\perp'$ and $d^2p_\perp$ integrations,

$$\Gamma'\equiv e^{-ip_\perp'\cdot x_\perp}\Gamma=\frac{e^2}{(2\pi)^2}\int 2\,d\alpha\,d\beta\,d\rho\,d\sigma\frac{A^2(\rho)\exp[i(\alpha y+\beta n^*\cdot x-\alpha\rho-2\beta\sigma+\sigma n\cdot p'+\frac{1}{2}\rho n^*\cdot p')]}{p_\perp^2+2\alpha\beta+m^2+M^2(y,\rho)-i\epsilon\epsilon(\alpha-\beta)]}.$$

The σ integration gives $\delta(2\beta-n\cdot p')$ so that the β integration can be done, which gives

$$\Gamma'=\frac{e^2}{2\pi}\int d\alpha\,d\rho A^2(\rho)\frac{\exp[i(\alpha y+\frac{1}{2}n\cdot p'n^*\cdot x+\frac{1}{2}\rho n^*\cdot p'-\alpha\rho)]}{[p_\perp'^2+2\alpha n\cdot p'+m^2+M^2-i\epsilon\epsilon(2\alpha-n\cdot p')]}.$$

In the α integration there is a pole at

$$\alpha=-(p_\perp'^2+m^2+M^2)/2n\cdot p'-i\epsilon.$$

Thus closing the α contour in the upper (lower) half-plane for $y-\rho>0$ (<0), Γ' becomes, using $p'^2+m^2=0$,

$$\begin{aligned}\Gamma'&=\frac{-ie^2}{2n\cdot p'}\int_v^\infty d\rho A^2(\rho)\\ &\quad\times\exp\left[-i\frac{e^2}{2n\cdot p'}\int_\rho^v dy'A^2(y')+ip_{\prime\prime}'\cdot x_{\prime\prime}\right]\\ &=-\int_v^\infty d\rho\frac{\partial}{\partial\rho}\\ &\quad\times\exp\left[-i\frac{e^2}{2n\cdot p'}\int_\rho^v dy'A^2(y')+ip_{\prime\prime}'\cdot x_{\prime\prime}\right]\\ &=\exp(ip_{\prime\prime}'\cdot x_{\prime\prime})\left[1-\exp\left(-i\int_\infty^v dy'J_{p'}(y')\right)\right].\end{aligned}$$

Thus

$$F=f-\Lambda=(2\pi)^{-\frac{3}{2}}\int d^4p\delta(p^2+m^2)\theta(p_0)f(\mathbf{p})\times\exp(ip\cdot x)\exp\left(-i\int_\infty^v dy'Jp(y')\right),$$

which is Eq. (7) of the text.

APPENDIX C

In this appendix we evaluate (31) for $\Gamma_{\alpha\beta}$. We note that, if instead of letting $x_0 \to \infty$, we let $x_0 \to -\infty$, we obtain from (18) just $\delta_{\alpha\beta}$. We consider the expression

$$\Gamma_{\alpha\beta}(x_0) \equiv i\int d^3x f_\alpha^*(x)\overleftrightarrow{\partial_t} F_\beta(x) \xrightarrow[x_0\to\infty]{} \Gamma_{\alpha\beta} \xrightarrow[x_0\to-\infty]{} \delta_{\alpha\beta}.$$

Now,

$$\partial_t\Gamma_{\alpha\beta}(x_0) = i\int d^3x\{-\partial_t^2 f_\alpha^* F_\beta + f_\alpha^*\partial_t^2 F_\beta\}$$
$$= -e^2\int d^3x A^2(y) f_\alpha^*(x) F_\beta(x).$$

Hence,

$$\Gamma_{\alpha\beta} = \delta_{\alpha\beta} + \int_{-\infty}^{\infty} dx_0 \partial_t \Gamma_{\alpha\beta}(t)$$
$$= \delta_{\alpha\beta} - ie^2\int d^4x A^2(y) f_\alpha^*(x) F_\beta(x)$$
$$\equiv \delta_{\alpha\beta} - R_{\alpha\beta}.$$

Using the form (7) for F, we have

$$R_{\alpha\beta} = \frac{ie^2}{(2\pi)^3}\int d^4p'\delta(p'^2 + m^2)\theta(p_0')f_\alpha^*(\mathbf{p}') \times \int d^4x\, d^4p\delta(p^2 + m^2)\theta(p_0)f_\beta(\mathbf{p})A^2(y) \times \exp\left[i(p - p')\cdot x - i\int_\infty^y dy' J_{p'}(y')\right]$$

Letting $p = p_\perp + \alpha n + \beta n^*$, $x = x_\perp + \sigma n + \frac{1}{2}\rho n^*$, as in Appendix B, we have

$$R_{\alpha\beta} = \int d^4p'\delta(p'^2 + m^2)\theta(p_0')f_\alpha^*(\mathbf{p}')I$$

with

$$I = \frac{ie^2}{2\pi}\int d\rho\, d\sigma 2\, d\alpha\, d\beta\theta(\alpha - \beta) \times \delta(p_\perp'^2 + 4\alpha\beta + m^2)f(\mathbf{p}_\perp, \alpha + \beta)A^2(\rho) \times \exp\left[i(\alpha\rho + 2\sigma\beta - \sigma n\cdot p' - \tfrac{1}{2}\rho n^*\cdot p') - i\frac{e^2}{4\beta}\int_\infty^\rho dy' A^2(y')\right]$$
$$= ie^2\int d\rho 2\, d\alpha\, d\beta\theta(\alpha - \beta)\delta(p_\perp'^2 + m^2 + 4\alpha\beta) \times \delta(2\beta - n\cdot p')A^2(\rho)f(\mathbf{p}_\perp, \alpha + \beta) \times \exp\left[i(\alpha\rho - \tfrac{1}{2}\rho n^*\cdot p') - i\int_\infty^\rho dy' J_{p'}(y')\right].$$

Using $p'^2 + m^2 = 0$ and the δ functions, we find

$$\beta = \tfrac{1}{2}n\cdot p' < 0, \quad \alpha = \tfrac{1}{2}n^*\cdot p',$$
$$\alpha + \beta = \mathbf{n}\cdot\mathbf{p}', \quad \alpha - \beta = p_0' > 0.$$

Hence,

$$I = \frac{-ie^2 f_\beta(\mathbf{p}')}{2n\cdot p'}\int d\rho A^2(\rho)\exp\left[\left(\frac{-ie^2}{2n\cdot p'}\right)\int_\infty^\rho dy' A^2(y')\right]$$
$$= f_\beta(\mathbf{p}')\int d\rho\frac{\partial}{\partial\rho}\exp(-i[\ ])$$
$$= f_\beta(\mathbf{p}')\left[1 - \exp\left(-i\int_{+\infty}^{-\infty} dy' J_{p'}(y')\right)\right].$$

Putting I into R and then into Γ, we obtain the result given in (32), since

$$\int d^4p\delta(p^2 + m^2)\theta(p_0)f_\alpha^*(\mathbf{p})f_\beta(\mathbf{p}) = \delta_{\alpha\beta}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Relativistically Dynamic Elastic Dielectrics

LINCOLN E. BRAGG
Mathematics Department, University of Kentucky, Lexington, Kentucky 40506

(Received 29 May 1969)

A relativistically exact theory of elastic dielectrics is developed, plus incidentals useful in a wider context. The most general form of a relativistically objective elastic response function is found in terms of Bressan's deformation gradient and of arbitrary tensors on the body manifold depending on the Born strain tensor. A form for constitutive relations is postulated for dielectrics which involves objective response functions, physical laws are stated in terms of integrals of exterior forms on the world manifold, and the resultant differential equations of motion are derived. Facts about algebra and calculus of exterior forms are developed which will be useful in continuum physics whenever one introduces laws involving exterior forms on manifolds. The relation to historical theories of dielectrics is indicated.

1. INTRODUCTION

The present paper develops a theory of elastic dielectrics which is based on essentially the same premises as the theory developed by Toupin[1] in 1963 (Secs. 1, 2, 4, and 5). The present theory, however, is relativistically exact. Toupin's Eqs. (5.1), the constitutive equations of his elastic dielectric, can be summarized by a statement which also characterizes the present theory: The polarization vector and the deformation gradient are the independent variables on which depend the state of the material and, hence, on which depend the energy and stress.

We describe motions in relativistic space–times in terms of the function which assigns to each event the material particle which experiences that event. We describe kinematical, dynamical, and electromagnetic variables in terms of tensors and exterior forms over the world manifold. These modes of description are designed to be efficient and intuitive when one is postulating or discussing the foundations of relativistic theories of materials.

We derive the most general form which an elastic response function can take if it is to be objective relative to relativistic changes of observer. Postulating constitutive relations for elastic dielectrics in terms of objective response functions and inserting these relations into the commonly accepted laws of motion, we derive the forms which these laws of motion take for elastic dielectrics.

The reader wishing to construct relativistic response functions governing the behavior of particular materials for which nonrelativistic response functions are known may read Secs. 2, 3, and 5. The reader interested in the most general form objective-response functions can take may read Secs. 2 and 3. He who is interested only in seeing how concepts of kinematics, dynamics, and electromagnetics can be treated in the terms which we have chosen, without regard to how the present theory of a particular class of materials is constructed, may read Secs. 2, 4, and 6. A reader who does not see readily how the preceding six sections relate to theories which he knows by similar names may read Sec. 8 in conjunction with the first five sections of Toupin,[1] Secs. 283 and 284 on polarization and magnetization and on electromagnetic energy, momentum, stress, and energy flux of Ref. 2, and pp. 392–405 of Ref. 3.

The primary mathematical tool used is the theory of exterior differential forms. The particular concrete realization of these abstract entities which we employ is the realization as skew-symmetric multilinear functions of k-tuples of tangent vectors. We treat metric tensors, electromagnetic field tensors and such, not as bilinear forms, but as linear transformations of the tangent vector space into the dual space of real-valued linear one-forms. To indicate the real number which is the value of any one-form λ at any vector $\mathbf{V}$, we write the names of the one-form and the vector with a centered dot between them, $\lambda \cdot \mathbf{V}$. In particular, if $\mathbf{G}$ is a metric tensor and $\mathbf{U}$ and $\mathbf{V}$ are vectors, then $\mathbf{GU}$ is a one-form and $\mathbf{V} \cdot (\mathbf{GU})$ is a real number, the inner product of $\mathbf{V}$ and $\mathbf{U}$. Other similar centered dot notations are introduced ad hoc to indicate the values of various types of functions whose arguments are one-forms or vectors.

2. MOTION, ENERGY-MOMENTUM, AND ELECTROMAGNETISM

Consider a three-dimensional manifold S^*, which we will call "the body," and consider a relativistic universe $\mathcal{E}$. By "relativistic universe," we mean any four-dimensional manifold supplied with a metric tensor having signature plus or minus two, supplied with a future direction choice among the timelike vectors, and supplied with an orientation. In particular, the common variety of special relativistic

[1] R. A. Toupin, Intern. J. Eng. Sci. 1, 101 (1963).

[2] C. Truesdell and R. A. Toupin, *The Classical Field Theories; Handbuch der Physik*, III/1 (Springer, Berlin, 1960).

[3] E. T. Whittaker, *A History of the Theories of Aether and Electricity, Volume 1, The Classical Theories* (Nelson, London, 1951; and The Philosophical Library, New York, 1951).

universe, with a t, x, y, z coordinate system singled out, is included; this is probably the most interesting type of universe in connection with the present theory. The time direction and the orientation are implicitly determined, of course, by the choice of the specially significant coordinate system.

Consider the class $\mathcal{M}$ of all smooth functions ζ of $\mathcal{E}$ onto S^* which have the following properties: At every point p, the rank of the gradient $\nabla\zeta(p)$ is the maximum possible, three. The (one-dimensional) null space of $\nabla\zeta(p)$ is timelike. Given a function ζ having at least these two properties, then, for each X in S^*, the set of all points p in $\mathcal{E}$ such that $\zeta(p) = X$ will be called "the world line of the particle X." It is possible for such a function ζ that some particles X can have "world lines" which consist of many separate curves in $\mathcal{E}$. To eliminate this possibility, we add a final property to the list of membership requirements which define the class $\mathcal{M}$. We insist that if a function ζ is to belong, $\zeta^{-1}(X)$ must be a connected set for every X in S^*. On a local scale, this last property is no real restriction on the class of functions.

The functions ζ which belong to $\mathcal{M}$ will be called "motions." The interpretation of such a function is that, for every event-point p in the universe $\mathcal{E}$, $X = \zeta(p)$ is the material particle which experiences the event. It is somewhat customary—perhaps largely for historical reasons—to consider the inverse relation. To each material particle X, we associate all of the events p which X experiences. To express this one-to-many relation in terms of a single-valued function, one embeds S^* into a four-dimensional manifold, with an absolute-time or a proper-time parameter representing the added dimension. One then introduces invariance requirements which express the nonrelevance of the four-dimensional ambient manifold. This procedure is natural when the body S^* consists of a finite number of points, and is fairly natural in general when $\mathcal{E}$ has a classical structure, with an absolute time function. For dealing with relativistic universes, we prefer the simpler procedure of dealing with the function which associates to each event the particle which experiences the event.

For each given motion ζ and given point p in $\mathcal{E}$, denote by $\mathbf{W} = \mathbf{W}(\zeta; p)$ the unique future-directed timelike unit vector in the null space of $\nabla\zeta(p)$. We will call $\mathbf{W}$ the world velocity of the motion. We note that the worldlines of ζ are the loci of the solutions $\hat{p}$ of the following differential equation:

$$\frac{d\hat{p}(s)}{ds} = \mathbf{W}(\zeta; \hat{p}(s)). \tag{2.1}$$

The parameter s along each worldline, which is determined by Eq. (2.1) up to the addition of a constant, will be called the proper time.

Consider a collection of functions consisting of a motion ζ, two scalar-valued skew-symmetric bilinear forms ϕ and β on $\mathcal{E}$, and a scalar-valued multilinear form t on $\mathcal{E}$ of the following type: t has four vector arguments and is skew-symmetric as a function of its last three arguments. We will call such a collection a "dynamic motion with dielectric polarization" if the equations which we give in Sec. 6 are satisfied as physical laws. We will call ϕ the electromagnetic field, and call β the dielectric polarization field. We will call t the energy–momentum flux field. Denote by E, and also by $\phi\mathbf{W}$, the one-form whose value $E \cdot \mathbf{X} = (\phi\mathbf{W}) \cdot \mathbf{X}$ at every vector $\mathbf{X}$ is the value $\phi(\mathbf{W}, \mathbf{X})$ of ϕ. We will call E the electromotive intensity.

The interpretations of these fields in terms of intuitive concepts of flow of charge, polarization charge, and magnetic induction, and concepts of work and flow of energy, will be given in Sec. 6 as the physical laws are introduced. The representations of the electromagnetic entities ϕ and β in terms of the historically primary electric, magnetic, and polarization vectors will be given in Sec. 8. The representation of t in terms of mass and energy density and a stress tensor will be developed in Secs. 3 and 4 as we develop forms for response functions whose values are energy–momentum tensors t. In Sec. 6 this will be summarized and expressed in terms of classical concepts. It is more efficient to work with the fields as we have introduced them here, while postulating the theory.

3. OBJECTIVITY OF RESPONSE FUNCTIONS: STRAIN MEASURES

A function $\tilde{R}$ will be called an elastic response function if it is of the following type. Its domain $\mathcal{D}$ consists of all fourtuples $(p, X, \mathbf{H}, \alpha)$, where p is any point in $\mathcal{E}$, X is any point in the body S^*, $\mathbf{H}$ is any linear transformation whose null space is timelike and which maps the tangent space $t\mathcal{E}(p)$ of $\mathcal{E}$ at p onto the tangent space $tS^*(X)$ of S^* at X, and α is a tensor or a collection of tensors of a specified type on the space $tS^*(X)$. The values $\tilde{R}(p, X, \mathbf{H}, \alpha)$ are tensors on the space $t\mathcal{E}(p)$.

An elastic response-function $\tilde{R}$ will be called "objective" if, for every $(p, X, \mathbf{H}, \alpha)$ in $\mathcal{D}$ and for every time-sense-preserving Lorentz transformation $\mathbf{S}$ of the tangent space $t\mathcal{E}(p')$ at any p' onto $t\mathcal{E}(p)$, the following identity holds: Let us denote by $\mathbf{S}[\cdots]$ the linear operator taking tensors at p' into tensors at p which canonically extends the operator $\mathbf{S}$ on vectors at p'. The identity is

$$\mathbf{S}[\tilde{R}(p', X, \mathbf{HS}, \alpha)] = \tilde{R}(p, X, \mathbf{H}, \alpha). \tag{3.1}$$

By way of justification for this definition, consider the case when $\mathcal{E}$ is a *special* relativistic universe. Let ζ be any motion, p be any point in $\mathcal{E}$, X be $\zeta(p)$, and $\mathbf{H}$ be $\nabla\zeta(p)$. For any point p' on $\mathcal{E}$ and any (time-sense-preserving) Lorentz transformation $\mathbf{S}$ of $t\mathcal{E}(p')$ onto $t\mathcal{E}(p)$, let f be the affine transformation of $\mathcal{E}$ such that $f(p') = p$ and $\nabla f(p') = \mathbf{S}$. Then $\zeta' = \zeta \circ f$ is another motion, which we regard as equivalent to ζ, differing only in its location and orientation relative to $\mathcal{E}$. We have $\zeta'(p') = X$ and $\nabla\zeta'(p') = \mathbf{HS}$. Hence we can regard $\tilde{R}(p', X, \mathbf{HS}, \alpha)$ as the value of some dynamical tensor variable at p' related to the motion ζ', and $\tilde{R}(p, X, \mathbf{H}, \alpha)$ as the value at p of this same tensor variable for the motion ζ. If $\tilde{R}$ is objective in the sense which we have postulated, then these values of the dynamical variable are related by the operator $\mathbf{S}[\cdots]$ which embodies the change of location and orientation due to f.

In case $\mathcal{E}$ is not a special, but a general relativistic universe, it may be that automorphisms f of $\mathcal{E}$ do not exist. One can be content to let the justification for the definition of objectivity rest on the special relativistic case. Alternatively, one can be explicit about the role of the metric tensor as a dynamical variable. One would have to include at least the value $\mathbf{G}(p)$ of the metric tensor at p as an argument of a response function $\tilde{R}$, in order that it is clear just which tensors $\mathbf{H}$ are to be included in fourtuples $(p, X, \mathbf{H}, \alpha)$ in the domain $\mathfrak{D}$ of $\tilde{R}$, since the null space of $\mathbf{H}$ is to be timelike. One might, or might not, want to include a dependence of $\tilde{R}$ on the values at p of some derivatives of $\mathbf{G}$, such as the curvature tensor of $\mathbf{G}$, calling the dependence a gravitational interaction. An automorphism f of $\mathcal{E}$ would then be any smooth homeomorphism; the metric tensor $\mathbf{G}(p)$ and any of its derivatives which appear as arguments would be transformed like $\mathbf{H}$ by the appropriate transformation law relative to f. We leave implicit the dependence of $\tilde{R}$ on $\mathbf{G}(p)$ which is implied in the nature of $\tilde{R}$'s domain; we include no further dependence now upon the metric tensor field $\mathbf{G}$ and regard the special relativistic case as adequate justification.

For convenience, let us adjoin an abstract vector $\mathbf{W}_0$ to each of the tangent spaces $tS^*(X)$ of S^*. For each X, let $TS^*(X)$ be the vector space of all linear combinations $(a\mathbf{W}_0 + \mathbf{Z})$ of $\mathbf{W}_0$ and vectors $\mathbf{Z}$ in $tS^*(X)$. Suppose that we are given a linear transformation $\mathbf{H}$ which, for some p and X, maps $t\mathcal{E}(p)$ onto $tS^*(X)$ and has a timelike null space. Let $\mathbf{W}$ be the unique future-directed timelike unit vector in this null space, and let ω be the linear form on $t\mathcal{E}(p)$ whose null space is the space $\mathcal{S}$ which is orthogonal to $\mathbf{W}$ and for which $\omega \cdot \mathbf{W} = 1$. That is, ω is $\mathbf{GW}$ or $-\mathbf{GW}$, depending on the signature of $\mathbf{G}$. Let us denote by δ the factor of plus or minus one such that $\omega = -\delta\mathbf{GW}$, hence such that $\delta\mathbf{U} \cdot \mathbf{GU}$ is positive for spacelike $\mathbf{U}$.

In terms of the given $\mathbf{H}$, define as follows a non-singular linear transformation $\mathbf{H}^*$ of $t\mathcal{E}(p)$ onto $TS^*(X)$. Denote by $\mathbf{W}_0 \otimes \omega$ the operator which to each vector $\mathbf{U}$ assigns the product $(\omega \cdot \mathbf{U})\mathbf{W}_0$ of $(\omega \cdot \mathbf{U})$ and $\mathbf{W}_0$. Then set

$$\mathbf{H}^* = \mathbf{H} + \mathbf{W}_0 \otimes \omega. \tag{3.2}$$

The restriction of $\mathbf{H}^*$ to the orthogonal complement $\mathcal{S}$ of $\mathbf{W}$ is clearly the same as the restriction of $\mathbf{H}$ to $\mathcal{S}$. Since the restriction of any linear transformation $\mathbf{H}$ to any direct-sum complement of its null space has the same range as $\mathbf{H}$, the image of $\mathcal{S}$ by $\mathbf{H}$, and hence the image of $\mathcal{S}$ by $\mathbf{H}^*$, is exactly $tS^*(X)$. Since $\mathbf{H}^*\mathbf{W} = \mathbf{W}_0$, it follows in particular that the range of $\mathbf{H}^*$ is all of $TS^*(X)$; hence $\mathbf{H}^*$ is invertible, as we claimed.

Denote by $\mathbf{A}$ the linear transformation of $tS^*(X)$ onto $\mathcal{S}$ which is the inverse of the restriction of $\mathbf{H}$ to $\mathcal{S}$. Call $\mathbf{A}$ the "Bressan deformation gradient"[4] associated with $\mathbf{H}$. Extend $\mathbf{A}$ from $tS^*(X)$ to a linear transformation $\mathbf{A}^*$ of $TS^*(X)$ onto $t\mathcal{E}(p)$ by setting $\mathbf{A}^*\mathbf{W}_0 = \mathbf{W}$. Note that $\mathbf{A}^*$ is in fact the inverse of $\mathbf{H}^*$. Define, as follows, linear transformations $\mathbf{C}^*$ and $\mathbf{C}$ taking $TS^*(X)$ and $tS^*(X)$, respectively, onto their dual spaces. Recall that $\mathbf{U} \cdot \lambda$ denotes the value at a vector $\mathbf{U}$ of a linear form λ in the dual space. Note that the factor δ introduced above is such that $\delta\mathbf{G}$ is positive-definite on spacelike subspaces, hence on $\mathcal{S}$. Then $\mathbf{C}^*$ is the operator such that, for $\mathbf{U}$ and $\mathbf{V}$ in $TS^*(X)$,

$$\begin{aligned}\mathbf{U} \cdot \mathbf{C}^*\mathbf{V} &= \delta(\mathbf{A}^*\mathbf{U}) \cdot \mathbf{G}(\mathbf{A}^*\mathbf{V}) \\ &= \mathbf{U} \cdot (\delta\mathbf{A}^{*t}\mathbf{G}\mathbf{A}^*)\mathbf{V}\end{aligned} \tag{3.3}$$

The same equation, without the asterisks and with $\mathbf{U}$ and $\mathbf{V}$ restricted to $tS^*(X)$, defines $\mathbf{C}$. Note that both $\mathbf{C}$ and $\mathbf{C}^*$ are symmetric, since $\mathbf{G}$ is symmetric. $\mathbf{C}$ is positive-definite, since, on the range of $\mathbf{A}$, $\delta\mathbf{G}$ is positive-definite. Once the restriction $\mathbf{C}$ of $\mathbf{C}^*$ is known, then $\mathbf{C}^*$ is determined, since, for all $\mathbf{U}$ and $\mathbf{V}$ in $tS^*(X)$,

$$\mathbf{U} \cdot \mathbf{C}^*\mathbf{V} = \mathbf{U} \cdot \mathbf{C}\mathbf{V}, \quad \mathbf{U} \cdot \mathbf{C}^*\mathbf{W}_0 = \mathbf{W}_0 \cdot \mathbf{C}^*\mathbf{V} = 0, \quad \mathbf{W}_0 \cdot \mathbf{C}^*\mathbf{W}_0 = -1. \tag{3.4}$$

The components of $\mathbf{C}$ are commonly called Born's measures of strain[5]; we shall consequently call $\mathbf{C}$ the Born tensor associated with $\mathbf{H}$.

Thus far, we have introduced elastic response functions and objectivity of such functions, have motivated these definitions, and have introduced

[4] Aldo Bressan, Ann. Mat. Pura Appl.,Ser. 4, **62**, 99 (1963).
[5] Max Born, Ann. Physik, Ser. 4, **33**, 1 (1909).

Bressan deformation gradients **A** and Born tensors **C** associated with the "motion-function-gradient-" **H** part of the argument of response functions, as well as formal extensions **A*** and **C*** of such tensors **A** and **C**. Let us consider the most general form which an objective response function can take. Let $\tilde{R}$ be any elastic response function. Noting that, for each $(p, X, \mathbf{H}, \alpha)$ in the domain, the operator $\mathbf{H}^*[\cdots]$ takes the space of all tensors over $t\mathcal{E}(p)$, in which the tensor $\tilde{R}(p, X, \mathbf{H}, \alpha)$ lies, into the space of tensors over $TS^*(X)$, define a function $\tilde{R}^*$ whose value at each $(p, X, \mathbf{H}, \alpha)$ is the following tensor over $TS^*(X)$:

$$\tilde{R}^*(p, X, \mathbf{H}, \alpha) = \mathbf{H}^*[\tilde{R}(p, X, \mathbf{H}, \alpha)]. \quad (3.5)$$

Suppose for instance that the values of the given response function $\tilde{R}$ are covariant tensors, as are the values of all response functions in which we shall be interested. For definiteness, suppose that two is the number of vector arguments that $\tilde{R}(p, X, \mathbf{H}, \alpha)$ has. To find the value of $\tilde{R}^*(p, X, \mathbf{H}, \alpha)$ at a pair $(\mathbf{U}, \mathbf{V})$ of vectors in $TS^*(X)$, one substitutes the pair $(\mathbf{A}^*\mathbf{U}, \mathbf{A}^*\mathbf{V})$ of vectors in $t\mathcal{E}(p)$ into $\tilde{R}(p, X, \mathbf{H}, \alpha)$. Here $\mathbf{A}^*$ is the inverse of the extension $\mathbf{H}^*$ of whatever linear transformation **H** appears as the third argument. That is,

$$\tilde{R}^*(\mathbf{U}, \mathbf{V}, p, X, \mathbf{H}, \alpha) = \tilde{R}(\mathbf{A}^*\mathbf{U}, \mathbf{A}^*\mathbf{V}, X, \mathbf{H}, \alpha). \quad (3.6)$$

Let $(p, X, \mathbf{H}, \alpha)$ be an arbitrary fixed element in $\mathfrak{D}$, and let $\mathbf{S}$ be an arbitrary (time-sense-preserving) Lorentz transformation of $t\mathcal{E}(p')$ onto $t\mathcal{E}(p)$ for an arbitrary point p'. Define tensors $\mathbf{R}'$ at p', **R** at p, **M** at X and $\mathbf{M}'$ at X by

$$\begin{aligned} \mathbf{R}' &= \tilde{R}(p', X, \mathbf{HS}, \alpha), \\ \mathbf{R} &= \tilde{R}(p, X, \mathbf{H}, \alpha), \quad \mathbf{M} = \mathbf{H}^*[\mathbf{R}] \\ &= \tilde{R}^*(p, X, \mathbf{H}, \alpha), \\ \mathbf{M}' &= (\mathbf{HS})^*[\mathbf{R}'] \\ &= \tilde{R}^*(p', X, \mathbf{HS}, \alpha). \end{aligned} \quad (3.7)$$

Note that **R** being equal to $\mathbf{S}[\mathbf{R}']$ identically in the arbitrary arguments is equivalent to objectivity of $\tilde{R}$. The "square bracket" operation of linear transformations on tensors is an associative operation, and $(\mathbf{HS})^*$ equals the product $\mathbf{H}^*\mathbf{S}$. Therefore, we have

$$(\mathbf{HS})^*[\mathbf{R}'] = \mathbf{H}^*[\mathbf{S}[\mathbf{R}']]. \quad (3.8)$$

The operation $\mathbf{H}^*[\cdots]$ is invertible; hence the $\mathbf{H}^*[\mathbf{S}[\mathbf{R}']]$ which appears in the right member is equal to $\mathbf{H}^*[\mathbf{R}]$ if and only if $\mathbf{S}[\mathbf{R}']$ is equal to **R**. Since $\mathbf{S}[\mathbf{R}']$ being equal to **R** is equivalent to objectivity of $\tilde{R}$, we conclude that $\tilde{R}$ is objective if and only if $\mathbf{M}' = (\mathbf{HS})^*[\mathbf{R}']$ is equal to $\mathbf{M} = \mathbf{H}^*[\mathbf{R}]$. That is, if and only if the following identity involving the function $\tilde{R}^*$ holds:

$$\tilde{R}^*(p, X, \mathbf{H}, \alpha) = \tilde{R}^*(p', X, \mathbf{HS}, \alpha). \quad (3.9)$$

Theorem: A function $\tilde{R}^*$ satisfies the identity Eq. (3.9) if and only if there exists a function $\hat{R}$ such that the domain $\mathfrak{D}_0$ of $\hat{R}$ consists of all triples $(X, \mathbf{C}, \alpha)$, where X is a point in S^*, **C** is a positive-definite symmetric linear transformation of $tS^*(X)$ onto its dual space, and α is a tensor at X of the same type as occurs in the members $(p, X, \mathbf{H}, \alpha)$ of $\mathfrak{D}$. The values $\hat{R}(X, \mathbf{C}, \alpha)$ are tensors at X. When $(p, X, \mathbf{H}, \alpha)$ is a member of $\mathfrak{D}$ and **C** is the Born tensor associated with **H**, we have

$$\tilde{R}^*(p, X, \mathbf{H}, \alpha) = \hat{R}(X, \mathbf{C}, \alpha). \quad (3.10)$$

When it exists, the function $\hat{R}$ is determined by $\tilde{R}^*$.

Proof: Let $(p, X, \mathbf{H}, \alpha)$ be any element in $\mathfrak{D}$, let q be any point in $\mathcal{E}$, and let **K** be any linear transformation of $t\mathcal{E}(q)$ onto $tS^*(X)$ which has a timelike null space. Note that $(q, X, \mathbf{K}, \alpha)$ is also in $\mathfrak{D}$. Let **S** be the linear transformation of $t\mathcal{E}(q)$ onto $t\mathcal{E}(p)$ defined by

$$\mathbf{S} = \mathbf{H}^{*-1}\mathbf{K}^*. \quad (3.11)$$

We shall show that **H** and **K** have the same Born tensor if and only if **S** is a time-sense-preserving Lorentz transformation. Note that $\mathbf{K}^*$ followed by $\mathbf{H}^{*-1}$ takes a future-directed vector **W** in the null space of **K** into $\mathbf{W}_0$ and thence into a vector $\mathbf{W}'$ in the null space of **H**, which is a future-directed vector. Thus, if **S** is a Lorentz transformation, it is a time-sense-preserving one.

Consider the linear transformation $\mathbf{S}^t\mathbf{G}(p)\mathbf{S}$. The three factors in turn take vectors **U** in $t\mathcal{E}(q)$ into vectors **SU** in $t\mathcal{E}(p)$ into covectors $\mathbf{G}(p)\mathbf{SU}$ in the dual of $t\mathcal{E}(p)$ into covectors $\mathbf{S}^t\mathbf{G}(p)\mathbf{SU}$ in the dual of $t\mathcal{E}(q)$. Thus its domain and range are the same as $\mathbf{G}(q)$; **S** is a Lorentz transformation by definition if $\mathbf{S}^t\mathbf{G}(p)\mathbf{S}$ equals $\mathbf{G}(q)$. Since $\mathbf{K}^*$ and its transpose $\mathbf{K}^{*t}$ are invertible, this is equivalent to

$$(\mathbf{K}^*)^{t-1}\mathbf{S}^t\mathbf{G}(p)\mathbf{S}\mathbf{K}^{*-1} = (\mathbf{K}^*)^{t-1}\mathbf{G}(q)\mathbf{K}^{*-1}. \quad (3.12)$$

The right member is the extended Born tensor $\mathbf{D}^*$ of **K**, whose restriction to $tS^*(X)$ is **K**'s Born tensor **D**. After substitution for **S** of the expression in Eq. (3.11), the left member reduces to the extended Born tensor $\mathbf{C}^*$ of **H**, whose restriction to $tS^*(X)$ is **H**'s Born tensor **C**. Thus **H** and **K** have the same Born tensor if and only if Eq. (3.12) holds, which, in turn, holds if

and only if $\mathbf{S}$ is a time-sense-preserving Lorentz transformation.

It follows from what has been shown that if a function $\tilde{R}^*$ satisfies the objectivity identity, Eq. (3.9), then, for each $(p, X, \mathbf{H}, \alpha)$ in its domain, $\tilde{R}^*$ takes the same value $\tilde{R}^*(p, X, \mathbf{H}, \alpha) = \mathbf{R}$ on all of the members of its domain of the form $(q, X, \mathbf{K}, \alpha)$ for which the Born tensor $\mathbf{D}$ of $\mathbf{K}$ equals the Born tensor $\mathbf{C}$ of $\mathbf{H}$. When this is true, we can define a function $\hat{R}$ on the set of all triples $(X, \mathbf{C}, \alpha)$ as follows: For each $(X, \mathbf{C}, \alpha)$, set $\hat{R}(X, \mathbf{C}, \alpha)$ equal to the tensor $\mathbf{R}^*$ which is the common value of $\tilde{R}^*$ on all of the fourtuples $(p, X, \mathbf{H}, \alpha)$ for which the Born tensor of $\mathbf{H}$ is $\mathbf{C}$. Conversely, given any function $\hat{R}$, we can define a function $\tilde{R}^*$ by Eq. (3.10), that is, by composition with $\hat{R}$ of the function which to each fourtuple $(p, X, \mathbf{H}, \alpha)$ assigns the triple $(X, \mathbf{C}, \alpha)$. Since the uniqueness of $\hat{R}$ corresponding to a given $\tilde{R}^*$ is clear, the proof of the theorem is complete.

Now refer back to Eq. (3.5) and to the paragraph containing it, in which we introduced an arbitrary elastic response-function $\tilde{R}$. We have shown in the intervening paragraphs that $\tilde{R}$ is objective if and only if there exists a function $\hat{R}$ such that

$$\mathbf{H}^*[\tilde{R}(p, X, \mathbf{H}, \alpha)] = \hat{R}(X, \mathbf{C}, \alpha), \qquad (3.13)$$

where $\mathbf{C}$ is the Born tensor of $\mathbf{H}$. We have noted that $\hat{R}$ is unique when it exists.

Let $\tilde{R}$ be an objective elastic response-function, whose values are covariant tensors, and denote by k the order of the tensors $\tilde{R}(p, X, \mathbf{H}, \alpha)$. The space $TS^*(X)$ was somewhat artificially manufactured out of the tangent space $tS^*(X)$ of S^* in order to render more efficient and clear the derivation of the representation of $\tilde{R}$ in terms of the arbitrary function $\hat{R}$, as given in Eq. (3.10). Let us now reduce this representation to a more natural form by replacing the function $\hat{R}$ by a collection of functions whose values are tensors over the tangent space $tS^*(X)$ instead of over $TS^*(X)$.

Let λ be any value $\hat{R}(X, \mathbf{C}, \alpha)$ of $\hat{R}$. Or, more generally, let λ be any covariant tensor of order k over the space $TS^*(X)$, for any X. For every subset J of the set K of integers from one to k, define a real-valued function $\lambda(J; \cdots)$ of $k$tuples of vectors in $tS^*(X)$ by the following prescription: To obtain the value $\lambda(J; \mathbf{U}_1, \mathbf{U}_2, \cdots)$ at any $k$tuple $(\mathbf{U}_1, \mathbf{U}_2, \cdots)$, first determine for each index i in K whether i is in J or i is not in J. If i is in J, substitute the vector $\mathbf{W}_0$ into λ as the ith argument. If i is not in J, substitute the given tangent vector $\mathbf{U}_i$ into λ as the ith argument. A scalar value is thus determined, a vector of $TS^*(X)$ having been substituted in for each argument of λ. This value is $\lambda(J; \mathbf{U}_1, \mathbf{U}_2, \cdots)$. By an abuse of language, we may refer to each such function $\lambda(J; \cdots)$ as a covariant tensor, although it is not a multilinear function of its $k$tuple argument. The function is equivalent to the composition of a covariant tensor, whose order m is k minus the number of indices in J, with the prescription that tells which of the vectors are to be deleted from a given $k$tuple $(\mathbf{U}_1, \mathbf{U}_2, \cdots)$ to obtain an $m$tuple for substitution into the tensor.

Since the values of $\tilde{R}$ and the values of $\hat{R}$ are purely covariant tensors, the identity, Eq. (3.10), is equivalent to the following: For every $k$tuple $(\mathbf{U}_1, \mathbf{U}_2, \cdots)$ of vectors in $t\mathcal{E}(p)$,

$$\tilde{R}(\mathbf{U}_1, \mathbf{U}_2, \cdots; p, X, \mathbf{H}, \alpha) = \hat{R}(\mathbf{H}^*\mathbf{U}_1, \mathbf{H}^*\mathbf{U}_2, \cdots; X, \mathbf{C}, \alpha). \qquad (3.14)$$

For any vector $\mathbf{U}$ in $t\mathcal{E}(p)$, $\mathbf{H}^*\mathbf{U} = \mathbf{HU} + (\omega \cdot \mathbf{U})\mathbf{W}_0$, in which the first vector $\mathbf{HU}$ lies in $tS^*(X)$ and the second is a multiple of $\mathbf{W}_0$. The reader who is skilled in the expansion of multilinear functions of unspecified order k, each of whose arguments is a linear combination of two vectors, will recognize that

$$\tilde{R}(\mathbf{U}_1, \mathbf{U}_2, \cdots; p, X, \mathbf{H}, \alpha) = \sum_J \left[\prod_i (\omega \cdot \mathbf{U}_i)\right] \hat{R}(J; \mathbf{HU}_1, \mathbf{HU}_2, \cdots; X, \mathbf{C}, \alpha), \qquad (3.15)$$

where the sum is over all subsets J of K and the product is over all i in J. For illustration and for later use, let us consider special cases of the general expansion formula, Eq. (3.15). Let us now suppress the variables $(p, X, \mathbf{H}, \alpha)$, upon which all of the functions that have tildes depend, and suppress the variables $(X, \mathbf{C}, \alpha)$ upon which all functions that have carets depend.

Suppose $\tilde{\theta}$ is a response function for which the order k is one. There are two subsets of K—namely, K itself and the empty set ϕ. Denote by $\hat{g}$ the scalar-valued function of $(X, \mathbf{C}, \alpha)$ obtained by substitution of $\mathbf{W}_0$ in $\hat{\theta}$; $\hat{g} = \hat{\theta}(K)$. The function $\hat{\theta}(\phi; \cdots)$ is the restriction of the given $\hat{\theta}$ to tS^*; denote it simply by $\hat{\theta}$. Then, for every $\mathbf{U}$ in $t\mathcal{E}(p)$,

$$\tilde{\theta}(\mathbf{U}) = (\omega \cdot \mathbf{U})\hat{g} + \hat{\theta}(\mathbf{HU}). \qquad (3.16)$$

Suppose $\tilde{\beta}$ is a response function for which the order k is two, and suppose that $\tilde{\beta}$ is skew symmetric. Then $\hat{\beta}(\mathbf{W}_0, \mathbf{W}_0)$ vanishes identically, and $\hat{\beta}(\mathbf{Z}, \mathbf{W}_0) = -\hat{\beta}(\mathbf{W}_0, \mathbf{Z})$ for all $\mathbf{Z}$. Denote by $\hat{\lambda}$ the one-form on $tS^*(X)$, and denote by $\tilde{\lambda}$ the one-form on $t\mathcal{E}(p)$, which are given by

$$\hat{\lambda}(\mathbf{Z}) = \hat{\beta}(\mathbf{W}_0, \mathbf{Z}), \quad \tilde{\lambda}(\mathbf{V}) = \hat{\lambda}(\mathbf{HV}) = \hat{\beta}(\mathbf{W}_0, \mathbf{HV}). \qquad (3.17)$$

Note that by definition of the exterior product, $(\omega \cdot \mathbf{U})(\tilde{\lambda} \cdot \mathbf{V}) - (\omega \cdot \mathbf{V})(\tilde{\lambda} \cdot \mathbf{U})$ is the value of the function $\omega \wedge \tilde{\lambda}$ at $(\mathbf{U}, \mathbf{V})$. Therefore, upon expanding $\hat{\beta}(\mathbf{H}^*\mathbf{U}, \mathbf{H}^*\mathbf{V})$ and making substitutions for the various terms, we find that

$$\tilde{\beta}(\mathbf{U}, \mathbf{V}) = (\omega \wedge \tilde{\lambda})(\mathbf{U}, \mathbf{V}) + \hat{\beta}(\mathbf{HU}, \mathbf{HV}). \quad (3.18)$$

For a final example, suppose $\hat{t}$ is a response function for which the order k is four, and suppose that $\hat{t}$ has the symmetries suitable for an energy–momentum flux. That is, $\hat{t}$ is skew symmetric in its last three arguments. There being 16 subsets J of the set $K = (1, 2, 3, 4)$, the direct derivation of the relevant special case of the formula, Eq. (3.15), would be so difficult that it is worthwhile to use the general formula. In view of the skew symmetry of $\hat{t}$ in its arguments number 2, 3, and 4, every term in the right member of Eq. (3.15) vanishes in which J contains two or more of the integers 2, 3, 4. Furthermore, there are relations among the terms in which J contains exactly one of these integers 2, 3, 4.

Define $\hat{p}$, $\hat{\rho}$, $\hat{\sigma}$, and $\hat{h}$ to be the forms which one obtains by substituting $\mathbf{W}_0$ into $\hat{t}$ for the arguments whose indices are, respectively, (none), (1), (2), and (1, 2), and restrict the resulting functions of the remaining arguments to $t\mathcal{S}^*(X)$. Denote by $\tilde{p}$, $\tilde{\rho}$, $\tilde{\sigma}$, and $\tilde{h}$, the forms on $t\mathcal{E}(p)$ which one obtains by composing the similarly named forms on $t\mathcal{S}^*(X)$ with $\mathbf{H}$. The forms $\tilde{\rho}$ and $\tilde{h}$ are skew symmetric in all of their arguments, while the forms $\tilde{p}$ and $\tilde{\sigma}$ are, like $\tilde{t}$, skew symmetric in all except the first one of their arguments. Denote by $\tilde{t}\mathbf{U}$, $\tilde{p}\mathbf{U}$, and $\tilde{\sigma}\mathbf{U}$ the skew-symmetric forms one obtains by substituting a vector $\mathbf{U}$ in as the first argument. The general representation formula, Eq. (3.15), can be expressed for $\tilde{t}$ as follows:

$$\tilde{t}\mathbf{U} = \tilde{p}\mathbf{U} + (\omega \cdot \mathbf{U})\tilde{\rho} + \omega \wedge (\tilde{\sigma}\mathbf{U}) + (\omega \cdot \mathbf{U})\omega \wedge \tilde{h}. \quad (3.19)$$

Note that $(\omega \cdot \mathbf{W}) = 1$ and $\tilde{p}\mathbf{W}$ and $\tilde{\sigma}\mathbf{W}$ both vanish identically. Thus, when $\mathbf{U}$ is $\mathbf{W}$, the right member reduces to $\tilde{\rho} + \omega \wedge \tilde{h}$. Thinking of $\tilde{t}$ as an energy–momentum flux, we call $\tilde{p}$ the momentum flux, $\tilde{\rho}$ the mass and internal energy flux, $\tilde{\sigma}$ the stress flux, and $\tilde{h}$ the heat flux. We are not prepared to defend the choice of name for $\tilde{h}$ as well as the other choices. As we will show that $\tilde{h}$ must vanish identically in our theory, it is perhaps unimportant what name we give $\tilde{h}$.

We have shown that any given objective response functions $\tilde{\theta}$, $\tilde{\beta}$, and $\tilde{t}$ can be expressed in terms of compositions with $\mathbf{H}$ of functions $\hat{g}$, $\hat{\theta}$, $\hat{\lambda}$, $\hat{\beta}$, $\hat{p}$, $\hat{\rho}$, $\hat{\sigma}$, and $\hat{h}$ which depend on $(X, \mathbf{C}, \alpha)$ and sets of vectors in the tangent space of $\mathcal{S}^*$ at X. These representation functions are uniquely determined by the given response functions $\tilde{\theta}$, $\tilde{\beta}$, and $\tilde{t}$. In fact, one can obtain each careted function by substituting $\mathbf{W}$ for some of the arguments in the response function and substituting for the remaining arguments vectors $\mathbf{AZ}$ [which are images of vectors $\mathbf{Z}$ in $t\mathcal{S}^*(X)$ by the Bressan deformation gradient $\mathbf{A}$]. For instance,

$$\hat{h}(\mathbf{X}, \mathbf{Y}) = \hat{t}(\mathbf{W}, \mathbf{W}, \mathbf{AX}, \mathbf{AY}). \quad (3.20)$$

Conversely, suppose that arbitrary functions $\hat{g}$, $\hat{\theta}$, $\hat{\lambda}$, $\hat{\beta}$, $\hat{p}$, $\hat{\rho}$, $\hat{\sigma}$, and $\hat{h}$ are given, and that functions $\tilde{\theta}$, $\tilde{\beta}$, and $\tilde{t}$ of $(p, X, \mathbf{H}, \alpha)$ and of sets of vectors in the tangent space at p are defined by Eqs. (3.16), (3.18), and (3.19). It is clear that $\tilde{\theta}$, $\tilde{\beta}$, and $\tilde{t}$ will be objective.

4. ALGEBRA AND CALCULUS

One can state the integral laws of motion using only conventional notation drawn from two widely studied theories. The one theory encompasses vector fields, covariant derivatives of vector fields, and linear transformation fields. The other is the theory of integration of differential forms on manifolds. In order to discuss these laws in any detail, it is convenient to introduce and study some additional specific tensor operations, operations which are of interest only in the context of integral conservation laws. This section is designed in part to aid the reader in reconciling the types of tensors to which we attached electromagnetic and dynamic names in Sec. 2 with the types of tensors which he is accustomed to associating with these names. The results of this section also will be used in Sec. 7 to derive differential equations from the integral laws which we will have introduced in Sec. 6.

Let us fix our attention on an arbitrary point p of $\mathcal{E}$ during the following discussion of multilinear algebra. We shorten the name of $t\mathcal{E}(p)$ to $t\mathcal{E}$, and introduce names of other tensor spaces at p without the specificatory symbol "(p)" attached, but with its presence understood.

For each positive integer n, let $\mathcal{F}n$ be the space of all multilinear functions γ of $n$tuples of vectors from $t\mathcal{E}$, which are scalar valued and are skew symmetric in all n arguments. Let $\mathfrak{L}_0$ be the space of all multilinear functions t of fourtuples of vectors from $t\mathcal{E}$, which are scalar valued and which are skew symmetric in the last three of their arguments. It is convenient to think of substituting a vector $\mathbf{V}$ in as the first argument of a function t in $\mathfrak{L}_0$, and considering the result as a function of the three remaining arguments. Let us denote this function of threetuples by $t\mathbf{V}$. The correspondence between $\mathbf{V}$ and $t\mathbf{V}$ is linear, and for each $\mathbf{V}$, $t\mathbf{V}$ belongs to $\mathcal{F}3$; thus, we are interpreting t as a linear operator from $t\mathcal{E}$ into $\mathcal{F}3$.

The space $\mathcal{F}3$ is of special interest because in it lie the tensors which represent fluxes of scalar quantities, like charge and mass, across three-dimensional hypersurfaces. The space $\mathfrak{L}_0$ is of special interest because in it lie the tensors which represent fluxes of covector valued quantities, like energy–momentum, across three-dimensional hypersurfaces. We introduce two more spaces $\mathcal{F}$ and $\mathfrak{L}$, which we shall canonically identify with $\mathcal{F}3$ and $\mathfrak{L}_0$, respectively. The tensors $\mathbf{m}$ in $\mathcal{F}$ have directions associated with them, like vectors in $t\mathcal{E}$, but do not have scalar magnitudes. Thus one can think of a direction of flow of stuff, but the concept of rate of flow usually associated with the magnitude of the flow vector is not so natural. In a similar and related way, the elements $\mathbf{T}$ of $\mathfrak{L}$ are similar to linear transformations of $t\mathcal{E}$ into itself, which are the type of tensors usually used to represent energy–momentum flux and in connection with which standard ideas of physical intuition have been developed.

Let $\mathcal{F}$ be the space of all multilinear functions $\mathbf{m}$ of fivetuples of the following type: The values of $\mathbf{m}$ are scalars, the first argument is a covector, and the last four arguments are vectors. Finally, $\mathbf{m}$ is skew symmetric in its last four arguments. For each $\mathbf{m}$ in $\mathcal{F}$ and each covector λ in $\mathcal{F}1$, let $\lambda \cdot \mathbf{m}$ denote the four-form in $\mathcal{F}_4$ which one obtains by substituting λ into $\mathbf{m}$ as its first argument. Call $\mathbf{m}$ and λ orthogonal to each other if this dot product $\lambda \cdot \mathbf{m}$ is zero. Let us say that a vector $\mathbf{V}$ in $t\mathcal{E}$ and a tensor $\mathbf{m}$ in $\mathcal{F}$ have the same direction if the collection of all covectors λ which are orthogonal to $\mathbf{V}$ is the same as the collection of covectors which are orthogonal to $\mathbf{m}$. We note that each nonzero $\mathbf{m}$ has a unique direction, in the sense that there is a unique one-dimensional space of vectors $\mathbf{V}$ which have the same direction as $\mathbf{m}$.

Having defined $\mathcal{F}$, we let $\mathfrak{L}$ be the space of all linear transformations $\mathbf{T}$ of $t\mathcal{E}$ into $\mathcal{F}$. For $\mathbf{T}$ in $\mathfrak{L}$ and $\mathbf{V}$ in $t\mathcal{E}$, denote by $\mathbf{TV}$ the tensor in $\mathcal{F}$ which is the value at $\mathbf{V}$ of $\mathbf{T}$. Besides the spaces $t\mathcal{E}$, $\mathcal{F}n$, $\mathfrak{L}_0$, $\mathcal{F}$, and $\mathfrak{L}$, which we have thus far named, one other space will occur in out theory. Let $\mathcal{F}^*$ be the space of all linear operators α taking vectors $\mathbf{V}$ in $t\mathcal{E}$ into forms γ in the one-dimensional space $\mathcal{F}4$. Such operators α are similar to one-forms λ in $\mathcal{F}1$, in that they map vectors $\mathbf{V}$ into a one-dimensional space. In view of the similarity with one-forms, we denote the value γ of an operator α at a vector $\mathbf{V}$ by $\alpha \cdot \mathbf{V}$.

We shall employ some multilinear functions defined on various Cartesian products of the spaces which we have introduced. These functions will have the requisite properties of nonsingularity that enable us to call them tensor-product functions. Since, for each Cartesian product of spaces, there is only one function whose domain is that Cartesian product, it will be unambiguous to denote the values of any one of the tensor-product functions by the names of its arguments with the multiplication sign $\otimes$ between them.

For each pair (θ, τ) in the Cartesian product $\mathcal{F}1 \times \mathcal{F}3$, denote by $\theta \otimes \tau$ the operator t in $\mathfrak{L}_0$ whose value $t\mathbf{V}$ at any vector $\mathbf{V}$ is the product $(\theta \cdot \mathbf{V})\tau$ of the scalar $\theta \cdot \mathbf{V}$ and the three-form τ. Similarly, define a tensor product on the Cartesian product $t\mathcal{E} \times \mathcal{F}4$ having values in $\mathcal{F}$, and in turn define a tensor product on $t\mathcal{E} \times \mathcal{F}1 \times \mathcal{F}4$ having values in $\mathfrak{L}$, by the following formulas: For $\mathbf{U}$ in $t\mathcal{E}$, θ in $\mathcal{F}1$, and γ in $\mathcal{F}4$,

$$\mathbf{U} \otimes \gamma = \mathbf{m}, \text{ for all } \lambda, \quad \lambda \cdot \mathbf{m} = (\lambda \cdot \mathbf{U})\gamma, \tag{4.1}$$

$$\mathbf{U} \otimes \theta \otimes \gamma = \mathbf{T}, \text{ for all } \mathbf{V}, \quad \mathbf{TV} = (\theta \cdot \mathbf{V})(\mathbf{U} \otimes \gamma). \tag{4.2}$$

For a fourth and final tensor product, for pairs (θ, γ) in $\mathcal{F}1 \times \mathcal{F}4$, let $\theta \otimes \gamma$ be the operator α in $\mathcal{F}^*$ which is defined as follows:

$$\theta \otimes \gamma = \alpha, \text{ for all } \mathbf{V}, \quad \alpha \cdot \mathbf{V} = (\theta \cdot \mathbf{V})\gamma. \tag{4.3}$$

We introduce a function on $\mathfrak{L}$ which takes values in the one-dimensional space $\mathcal{F}4$, which we call the "trace" because of its close analogy with the function of the same name on the space of linear transformations of $t\mathcal{E}$ into $t\mathcal{E}$, which takes values in the one-dimensional space of scalars. We denote the value of the trace of an operator $\mathbf{T}$ in $\mathfrak{L}$ by $\mathrm{Tr}\mathbf{T}$. The trace is defined by the requirements that it be linear and that the following equation hold whenever its argument $\mathbf{T}$ is a tensor product:

$$\mathbf{T} = \mathbf{U} \otimes \theta \otimes \gamma \quad \text{implies} \quad \mathrm{Tr}\mathbf{T} = (\theta \cdot \mathbf{U})\gamma. \tag{4.4}$$

Note for future use that for any $\mathbf{T}$ in $\mathfrak{L}$, and for any linear transformation $\mathbf{L}$ of $t\mathcal{E}$ into $t\mathcal{E}$, the composition $\mathbf{TL}$ is in $\mathfrak{L}$. Hence, $\mathrm{Tr}(\mathbf{TL})$ is well-defined.

Besides the trace, there are four other linear operations on $\mathfrak{L}$ of the type commonly called tensor contractions. These four take values in $\mathfrak{L}_0$, and they differ one from another merely by a multiplicative factor of plus or minus unity. We choose one, call it "the contraction," and denote its value at an operator $\mathbf{T}$ by $\mathrm{Ct}\mathbf{T}$. Making a similar choice among equivalent functions on $\mathcal{F}$, we define a closely related function taking $\mathcal{F}$ into $\mathcal{F}3$, which we also call "the contraction," and we denote its value at a tensor $\mathbf{m}$ by $\mathrm{Ct}\mathbf{m}$. Each contraction function is defined by the requirements that it be linear and that the appropriate one of the following equations hold whenever its argument is a tensor product. For any form γ in $\mathcal{F}4$ and any vector $\mathbf{U}$, denote by $\gamma(\mathbf{U})$ the three-form which is obtained by

substituting **U** into γ as its first argument:

$$\mathbf{m} = \mathbf{U} \otimes \gamma \quad \text{implies} \quad \mathrm{Ct}\mathbf{m} = \gamma(\mathbf{U}), \tag{4.5}$$

$$\mathbf{T} = \mathbf{U} \otimes \theta \otimes \gamma \quad \text{implies} \quad \mathrm{Ct}\mathbf{T} = \theta \otimes \gamma(\mathbf{U}). \tag{4.6}$$

It is apparent from these equations and from the definition of tensor products in Eqs. (4.2) and (4.3) that, at least when an operator **T** in $\mathfrak{L}$ is a tensor product $\mathbf{U} \otimes \theta \otimes \gamma$, one gets the same result $(\theta \cdot \mathbf{V})\gamma(\mathbf{U})$ whether he contracts **T** and then operates on a vector **V**, or first operates on **V** and then contracts. That is, for every vector **V**,

$$\mathrm{Ct}(\mathbf{TV}) = (\mathrm{Ct}\mathbf{T})\mathbf{V}. \tag{4.7}$$

Since tensor products span the space $\mathfrak{L}$, and the members of Eq. (4.7) are linear functions of **T**, the fact that it holds when **T** is a tensor product implies that it holds for all **T** in $\mathfrak{L}$.

Each contraction function is invertible and its range is the whole of $\mathcal{F}3$ or $\mathfrak{L}_0$, respectively, as we shall presently show. Thus, the contraction functions will serve as canonical identifications of their domain spaces $\mathfrak{L}$ and $\mathcal{F}$ with their range spaces $\mathfrak{L}_0$ and $\mathcal{F}3$. We begin the demonstration of invertibility by defining the operation which is the inverse. We call the operation "the expansion"; we denote its value at any τ in $\mathcal{F}3$ by $\mathrm{Ep}\tau$, and denote its value at any t in $\mathfrak{L}_0$ by $\mathrm{Ep}t$. The expansion function is defined by the following identities in its arguments τ and t and in one-forms λ and vectors **V**:

$$\mathrm{Ep}\tau = \mathbf{m}, \text{ for all } \lambda, \quad \lambda \cdot \mathbf{m} = \lambda \wedge \tau, \tag{4.8}$$

$$\mathrm{Ep}t = \mathbf{T}, \text{ for all } \lambda \text{ and } \mathbf{V}, \quad \lambda \cdot (\mathbf{TV}) = \lambda \wedge (t\mathbf{V}). \tag{4.9}$$

Note that, as in contraction, expansion commutes with operation on vectors **V**: If for given t and **V**, we set $\tau = t\mathbf{V}$, then, from the definition of $\mathrm{Ep}t$, it is clear that

$$(\mathrm{Ep}t)\mathbf{V} = \mathrm{Ep}\tau = \mathrm{Ep}(t\mathbf{V}). \tag{4.10}$$

Theorem: The expansion and the contraction are each the inverse of the other. Specifically, for every **m** in $\mathcal{F}$ and for every τ in $\mathcal{F}3$, $\mathrm{Ep}(\mathrm{Ct}\mathbf{m}) = \mathbf{m}$ and $\mathrm{Ct}(\mathrm{Ep}\tau) = \tau$. Hence, for every **T** in $\mathfrak{L}$ and for every t in $\mathfrak{L}_0$, $\mathrm{Ep}(\mathrm{Ct}\mathbf{T}) = \mathbf{T}$ and $\mathrm{Ct}(\mathrm{Ep}t) = t$.

Proof: We show that the composition $\mathrm{Ep}(\mathrm{Ct}\cdot\cdot\cdot)$ is the identity on the set of tensor products in $\mathcal{F}$. Since the composition is linear, it is then the identity on all of $\mathcal{F}$. We show that all of $\mathcal{F}3$ is included in the range of Ct. Since, by the previous result, $\mathrm{Ct}(\mathrm{Ep}(\mathrm{Ct}\cdot\cdot\cdot))$ is Ct, it follows that $\mathrm{Ct}(\mathrm{Ep}\cdot\cdot\cdot)$ is the identity on the range of Ct, hence on $\mathcal{F}3$. To prove the statement about operators **T** and t, we note that the results about elements of $\mathcal{F}$ and $\mathcal{F}3$ so far imply that, for any vector **V**, $\mathrm{Ep}(\mathrm{Ct}(\mathbf{TV})) = \mathbf{TV}$ and $\mathrm{Ct}(\mathrm{Ep}(t\mathbf{V})) = t\mathbf{V}$. Since operation on **V** commutes with both Ep and Ct, it follows that $\mathrm{Ep}(\mathrm{Ct}\mathbf{T}) = \mathbf{T}$ and $\mathrm{Ct}(\mathrm{Ep}t) = t$. The details which support these statements follow.

Let $\mathbf{m} = \mathbf{U} \otimes \gamma$ be any nonzero tensor product in $\mathcal{F}$. Since **m** is nonzero, both **U** and γ are nonzero. Hence every four-form in $\mathcal{F}4$ is a multiple of γ, and the result $\gamma(\mathbf{U})$ of substituting **U** into γ as first argument is a nonzero three-form. Therefore, for every one-form λ there is a scalar factor c such that

$$\lambda \wedge \gamma(\mathbf{U}) = c\gamma. \tag{4.11}$$

Substitution of **U** into the second factor $\gamma(\mathbf{U})$ of the left member amounts to substitution of **U** in for two of the arguments of γ; since γ is skew symmetric, the result is zero. It follows that substitution of **U** into the product $\lambda \wedge \gamma(\mathbf{U})$ as its first argument yields $(\lambda \cdot \mathbf{U})\gamma(\mathbf{U})$. This must equal the result $c\gamma(\mathbf{U})$ of substitution of **U** into the right member of the equation. Therefore $c = (\lambda \cdot \mathbf{U})$. Since then the right member $c\gamma$ is $(\lambda \cdot \mathbf{U})\gamma$, which equals $\lambda \cdot (\mathbf{U} \otimes \gamma)$, we have established that, for every λ and for every tensor product $\mathbf{m} = \mathbf{U} \otimes \gamma$,

$$\lambda \cdot (\mathrm{Ep}(\mathrm{Ct}\mathbf{m})) = \lambda \cdot \mathbf{m}. \tag{4.12}$$

Since this equation holds for every λ, we have $(\mathrm{Ep}(\mathrm{Ct}\mathbf{m})) = \mathbf{m}$ for every tensor product **m**, which is one of the results we set out to prove.

Let τ be any form in $\mathcal{F}3$. In the null space of τ choose any nonzero vector **U**, that is, a vector such that if it be substituted into τ as one of its arguments, then τ becomes identically zero as a function of the remaining arguments. Then choose a one-form λ such that $\lambda \cdot \mathbf{U} = 1$. Define a four-form γ and in turn a tensor **m** in $\mathcal{F}$ by

$$\gamma = \lambda \wedge \tau, \quad \mathbf{m} = \mathbf{U} \otimes \gamma = \mathbf{U} \otimes (\lambda \wedge \tau). \tag{4.13}$$

Since substitution of **U** into the second factor τ of $\gamma = \lambda \wedge \tau$ reduces τ to the zero function of its remaining arguments, $\gamma(\mathbf{U}) = (\lambda \cdot \mathbf{U})\tau = \tau$. Therefore, since $\gamma = \mathrm{Ct}\mathbf{m}$ by definition of the contraction of a tensor product, we have shown that there exists an **m** in $\mathcal{F}$ such that $\tau = \mathrm{Ct}\mathbf{m}$. Thus, the range of Ct on $\mathcal{F}$ is all of $\mathcal{F}3$. This completes the proof.

One definition and one resulting formula similar to the above will be useful, relative to the three-dimensional space tS^* of vectors tangent at some

point X to the body manifold S^*. For use in Sec. 5 (in discussing symmetry of the "stress-tensor" part of an energy–momentum tensor-valued objective response function $\hat{t}$), consider a multilinear three-form σ on tS^* which is skew symmetric as a function of its last two arguments. Thinking of σ as an operator taking vectors $\mathbf{V}$ into two-forms $\sigma\mathbf{V}$ obtained by substitution of $\mathbf{V}$ in σ as its first argument, we define its "expansion" $\mathbf{S} = \mathrm{Ep}\sigma$ as follows as an operator on vectors $\mathbf{V}$: The value $\mathbf{SV}$ for any $\mathbf{V}$ is a linear function of one-forms λ, with values denoted by $\lambda \cdot \mathbf{SV}$. For given λ and $\mathbf{V}$, the value $\lambda \cdot \mathbf{SV}$ lies in the (one-dimensional) space of skew-symmetric three-forms. The definition of $\mathbf{S}$ and the explicit formula for its values which is of interest are

$$\lambda \cdot \mathbf{SV} = \lambda \wedge (\sigma\mathbf{V}), \tag{4.14}$$

$$\begin{aligned}(\lambda \cdot \mathbf{SV})(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = {} & (\lambda \cdot \mathbf{X})(\sigma(\mathbf{V}, \mathbf{Y}, \mathbf{Z})) \\ & + (\lambda \cdot \mathbf{Y})(\sigma(\mathbf{V}, \mathbf{Z}, \mathbf{X})) \\ & + (\lambda \cdot \mathbf{Z})(\sigma(\mathbf{V}, \mathbf{X}, \mathbf{Y})).\end{aligned} \tag{4.15}$$

Let us apply the above algebraic results. For any point p such that within some neighborhood of p the energy–momentum field t and the charge-flux potential β are continuously differentiable and the electromagnetic field ϕ is continuous, the momentum equation [which we will postulate in Sec. 6 as Eq. (6.8)] implies that, for every smooth vector field $\mathbf{V}$, the following equation holds at p:

$$s(\nabla\mathbf{V}) = d(t\mathbf{V}) + (\phi\mathbf{V}) \wedge d\beta. \tag{4.16}$$

Herein, $d(t\mathbf{V})$ denotes the exterior derivative of the three-form $t\mathbf{V}$. We shall refer to the second term on the right as the work density due to interaction of ϕ, $\mathbf{V}$, and β, and we shall call the first term the work density due to energy–momentum transfer by interaction of t and $\mathbf{V}$. It is of interest to make the dependence of these two terms on $\mathbf{V}$ more explicit.

Consider arbitrary fields ϕ, $\mathbf{V}$, and β. Since the expansion $\mathrm{Ep}(d\beta)$ is the unique tensor $\mathbf{m}$ such that $\mathrm{Ct}\mathbf{m} = d\beta$, it is clear from the last paragraph of the above proof that $\mathrm{Ep}(d\beta)$ is a tensor product $\mathbf{U} \otimes \gamma$—indeed, that every tensor in $\mathcal{T}$ is a tensor product. We may think of the tensor $\mathrm{Ep}(d\beta)$ as a semiclassical representation of the charge-flux form $d\beta$. For any choice of factors $\mathbf{U}$ and γ in the representation of $\mathrm{Ep}(d\beta)$ as a tensor product, we may think in the classical manner of $\mathbf{U}$ as a flow velocity vector. We can keep in mind that the factor γ is needed to account for volume density and time-rate effects, and that it is only the product of $\mathbf{U}$ and γ that is canonically determined by $d\beta$.

The value $\mathrm{Ep}(d\beta) = \mathbf{U} \otimes \gamma$ of a tensor product does not determine its factors $\mathbf{U}$ and γ. However, if one chooses any favorite nonzero four-form η, then, for any given factors $\mathbf{U}$ and γ, the second factor γ is some scalar multiple $g\eta$ of η. By removing the factor g from γ and adding it to $\mathbf{U}$, we do not alter the value of the product, but do obtain η as the second factor. For a chosen value η of the second factor, the value $\mathbf{J} \otimes \eta$ of the product does determine the first factor $\mathbf{J}$. The vector $\mathbf{J}$ is the classical charge flow world-vector, whose time component in any frame of reference is the charge density and whose spatial part is the current vector.

Laying aside the question of interpretation of the factors, let us choose a four-form η and let $\mathbf{J}$ be the unique vector such that $d\beta = \mathbf{J} \otimes \eta$. Since, by definition of the one-form $\phi\mathbf{V}$, $(\phi\mathbf{V}) \cdot \mathbf{J} = \phi(\mathbf{V}, \mathbf{J})$, the work density is given by

$$(\phi\mathbf{V}) \wedge d\beta = (\phi\mathbf{V}) \cdot (\mathbf{J} \otimes \eta) = \phi(\mathbf{V}, \mathbf{J})\eta. \tag{4.17}$$

We may replace $\phi(\mathbf{V}, \mathbf{J})$ in the right member by $-\phi(\mathbf{J}, \mathbf{V}) = -(\phi\mathbf{J}) \cdot \mathbf{V}$. Therefore, by way of isolating the dependence on $\mathbf{V}$, we may express the work density as follows in terms of the operator $(\phi\mathbf{J}) \otimes \eta$ in $\mathcal{T}^*$:

$$(\phi\mathbf{V}) \wedge d\beta = -\mathbf{V} \cdot ((\phi\mathbf{J}) \otimes \eta). \tag{4.18}$$

We remark that this tensor $(\phi\mathbf{J}) \otimes \eta$ which we have introduced does not depend on the choice of vector $\mathbf{J}$ and four-form η used in factoring $\mathrm{Ep}(d\beta)$. If we remove a scalar factor g from $\mathbf{J} = g\mathbf{U}$ and add it in $g\eta = \gamma$, then $(\phi\mathbf{J}) \otimes \eta = (\phi\mathbf{U}) \otimes \gamma$ is unaltered in value.

We have need of a formula for the exterior derivative $d(t\mathbf{V})$ of the three-form $t\mathbf{V}$ which is obtained by substituting a vector field $\mathbf{V}$ into a tensor field t with values in $\mathfrak{L}_0$, to use in the right member of Eq. (4.16). Let $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ be any ordered fourtuple of vectors at p. Choose four vector fields near p and denote them by the same names $\mathbf{U}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$; the choice to be arbitrary except for the following restrictions: The value at p of each vector field must be the given vector, and the value of the covariant derivative of each field must be zero. The restriction to fields whose covariant derivatives are zero is for convenience, to avoid a clutter of irrelevant terms in identities involving the fields. We note that if the covariant derivatives of all of the fields vanish at p, then the Lie bracket of each pair of fields vanishes at p also. Lest the reader wonder about the possibility of the choice, we note that it is well known that, for any symmetric covariant derivative ∇ and any point p, there exists a coordinate system near p such that the gamma symbols of ∇

vanish at p. Given a vector $\mathbf{Z}$ at p, with components Z^i relative to such a coordinate system, the field whose components are constant and equal to Z^i will reduce to $\mathbf{Z}$ at p and will have zero covariant derivative at p.

Let $(\mathbf{U}', \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ be any ordered fourtuple consisting of the four fields $\mathbf{U}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, but not necessarily in that order. Substitute the last three fields into $t\mathbf{V}$, obtaining a scalar field $g = t(\mathbf{V}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$. Denote the derivative $\mathbf{U}' \cdot dg$ of this scalar field at p in the direction of $\mathbf{U}'$ by f. By this we mean the derivative of the values of g with respect to the parameter s along a curve $\hat{p}(s)$ whose tangent at p is $\mathbf{U}'$, and we mean the sum of the products with the components of $\mathbf{U}'$ of the partial derivatives of g with respect to coordinates in any coordinate system, etc. However one defines the right member, we set

$$f = \mathbf{U}' \cdot d[t(\mathbf{V}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')]. \tag{4.19}$$

Consider these four special cases of this formula: Let f_1, f_2, f_3, and f_4 denote the numbers which one obtains by substituting in for $(\mathbf{U}', \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ the four cyclic permutations $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$, $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{U})$, $(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{X})$, and $(\mathbf{Z}, \mathbf{U}, \mathbf{X}, \mathbf{Y})$ of vector fields relative to positions in the fourtuple $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Since in this collection of permutations each vector field gets exactly one turn in position number one, we obtain the exterior derivative $d(t\mathbf{V})$ by multiplying each number f_i by the signum of the related permutation and adding the results:

$$d(t\mathbf{V})(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = (+1)(f_1) + (-1)(f_2) + (+1)(f_3) + (-1)(f_4). \tag{4.20}$$

There would in general be twelve more terms, each involving a Lie bracket of some pair of the vector fields. Since all such Lie brackets are zero, these terms vanish, and we may ignore them.

The derivative $f = \mathbf{U}' \cdot dg$ in Eq. (4.19) of the value g of the tensor t evaluated at the fourtuple $(\mathbf{V}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ is equal to the sum of five terms, each term involving the covariant derivative of one of t, $\mathbf{V}$, $\mathbf{X}'$, $\mathbf{Y}'$, or $\mathbf{Z}'$. Since the last three have zero derivatives at p, the sum reduces to one term involving ∇t and one term involving $\nabla\mathbf{V}$. Denote by $\nabla t(\mathbf{V}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}', \mathbf{U}')$ the covariant derivative of t in the direction of $\mathbf{U}'$, evaluated at $(\mathbf{V}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, it being customary with covariant derivatives to add the direction $\mathbf{U}'$ of differentiation as a new argument in the far right-hand position, rather than in the left-hand position (as with exterior derivatives). Denote by $\nabla\mathbf{V} \cdot \mathbf{U}'$ the covariant derivative of $\mathbf{V}$ in the direction of $\mathbf{U}'$. Then we have

$$f = \nabla t(\mathbf{V}, \mathbf{X}', \mathbf{Y}', \mathbf{Z}', \mathbf{U}') + t(\nabla\mathbf{V} \cdot \mathbf{U}', \mathbf{X}', \mathbf{Y}', \mathbf{Z}'). \tag{4.21}$$

Substituting the vector fields into this equation in various orders, the reader may evaluate f_1, f_2, f_3, and f_4 and hence obtain a formula for $d(t\mathbf{V})(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Four terms will involve ∇t, and four terms will involve $\nabla\mathbf{V}$. These two groups of four terms are similar to each other; we now prove a theorem relating such expressions to the trace and expansion functions, and then write down a formula for $d(t\mathbf{V})$.

Theorem: For every r in $\mathfrak{L}_0$, the following identity in fourtuples $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ of vectors is valid:

$$\begin{aligned}(\mathrm{Tr}(\mathrm{Ep}r))(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = {} & r(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \\ & - r(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{U}) \\ & + r(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{X}) \\ & - r(\mathbf{Z}, \mathbf{U}, \mathbf{X}, \mathbf{Y}).\end{aligned} \tag{4.22}$$

Proof: Note that if Eq. (4.22) holds whenever $\mathbf{R} = \mathrm{Ep}r$ is a tensor product $\mathbf{U} \otimes \theta \otimes \gamma$, then it holds for all r, since each member of the equation is a linear function of r, and the set of contractions $\theta \otimes \gamma(\mathbf{U})$ of tensor products $\mathbf{U} \otimes \theta \otimes \gamma$ spans $\mathfrak{L}_0$. Since r is skew symmetric in its last three arguments, the particular combination of values of r which appears in the right-hand side of Eq. (4.22) is skew symmetric in all four arguments. Therefore, for a given r, if there exists a single fourtuple $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ of linearly independent vectors for which Eq. (4.22) is valid, then it is valid for all fourtuples. We conclude that it is sufficient to prove that Eq. (4.22) is valid in the case where $\mathbf{R} = \mathrm{Ep}r$ is a tensor product and $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is any one fourtuple of linearly independent fourtuples.

Given any tensor product $\mathbf{R} = \mathbf{U} \otimes \theta \otimes \gamma$ in $\mathfrak{L}$, choose any linearly independent fourtuple $(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ whose first vector $\mathbf{U}$ is the same as the first vector in $\mathbf{R}$. Since $r = \mathrm{Ct}\mathbf{R}$ is $\theta \otimes \gamma(\mathbf{U})$, the right member of Eq. (4.22) is

$$\begin{aligned}&(\theta \cdot \mathbf{U})\gamma(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - (\theta \cdot \mathbf{X})\gamma(\mathbf{U}, \mathbf{Y}, \mathbf{Z}, \mathbf{U}) \\ &\quad + (\theta \cdot \mathbf{Y})\gamma(\mathbf{U}, \mathbf{Z}, \mathbf{U}, \mathbf{X}) - (\theta \cdot \mathbf{Z})(\gamma(\mathbf{U}, \mathbf{U}, \mathbf{X}, \mathbf{Y}) \\ &\qquad = (\theta \cdot \mathbf{U})(\gamma(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) + 0 + 0 + 0.\end{aligned} \tag{4.23}$$

Since $\mathrm{Tr}\mathbf{R} = (\theta \cdot \mathbf{U})\gamma$, the left member $\mathrm{Tr}\mathbf{R}(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ also equals $(\theta \cdot \mathbf{U})\gamma(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Thus Eq. (4.22) is valid in this case, hence in every case.

To deal with the four terms which involve $\nabla\mathbf{V}$ in the formula for $d(t\mathbf{V})$ (which the reader was invited to write down just before this theorem was stated), we set $\mathbf{T} = \mathrm{Ep}t$, and then let $\mathbf{R}$ be the composition $\mathbf{T}\nabla\mathbf{V}$ of $\mathbf{T}$ and $\nabla\mathbf{V}$. Applying the theorem to $r = \mathrm{Ct}\mathbf{R}$, we find that the four terms which involve $\nabla\mathbf{V}$ add up to $\mathrm{Tr}(\mathbf{T}\nabla\mathbf{V})$.

For each fixed vector $\mathbf{V}$ consider the function r defined by

$$r(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \nabla t(\mathbf{V}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{U}). \quad (4.24)$$

Since the right member is skew symmetric as a function of $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, r belongs to $\mathfrak{L}_0$. Thus $\mathrm{Tr}(\mathrm{Ep}r)$ is well defined and belongs to $\mathcal{F}4$. Since the function r is a linear function of $\mathbf{V}$, the four-form $\mathrm{Tr}(\mathrm{Ep}r)$ is also a linear function of $\mathbf{V}$. Thus the operator which takes $\mathbf{V}$ into $\mathrm{Tr}(\mathrm{Ep}r)$ is a member of $\mathcal{F}^*$. Let us denote the latter operator by "$\mathrm{div}\mathbf{T}^t$," which we read as "the divergence of $\mathbf{T}$ transposed." Let us use a dot product notation when indicating the dependence on $\mathbf{V}$ but suppressing the other variables. That is, we set

$$(\mathrm{div}\mathbf{T}^t)(\mathbf{V}, \mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = (\mathrm{Tr}(\mathrm{Ep}r))(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}), \quad (4.25)$$

$$(\mathrm{div}\mathbf{T}^t) \cdot \mathbf{V} = \mathrm{Tr}(\mathrm{Ep}r). \quad (4.26)$$

The name $\mathrm{div}\mathbf{T}^t$ can be considered only as mnemonic device referring to a well-known theory in which the symbol $\mathbf{T}$ denotes a field of linear transformations of the tangent space $t\mathcal{E}$ into itself, rather than into the space $\mathcal{F}$ as in the present theory. In that case, both transposition and divergence-taking are well-known operations. Alternatively, one can justify the name $\mathrm{div}\mathbf{T}^t$ as follows: $\mathbf{T}$ is an operator on $t\mathcal{E}$, whose values $\mathbf{TV}$ are operators on $\mathcal{F}1$, whose values $\lambda \cdot (\mathbf{TV})$ are, in turn, members of $\mathcal{F}4$. Althogether, $\mathbf{T}$ is a real valued function of six arguments $\mathbf{V}$, λ, and the four vector arguments of $\lambda \cdot (\mathbf{TV})$, which we think of as being substituted into $\mathbf{T}$ in stages. It we think of substituting λ in first, we obtain an operator which we call $\mathbf{T}^t$ which takes a one-form λ into the function whose value at a fivetuple $(\mathbf{V}, \mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is

$$(\mathbf{T}^t\lambda)(\mathbf{V}, \mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = (\lambda \cdot (\mathbf{TV}))(\mathbf{U}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}).$$

For any operator such as $\mathbf{T}^t$, which takes $\mathcal{F}1$ into any tensor space, such as here the space of real-valued functions of five vectors, one can define the divergence as follows: The covariant derivative $\nabla(\mathbf{T}^t)$ is an operator taking pairs $(\lambda, \mathbf{K})$ into the same tensor space, where λ is the original one-form argument and $\mathbf{K}$ is the direction of differentiation vector. The contraction of $\nabla(\mathbf{T}^t)$ on the two primary arguments $(\lambda, \mathbf{K})$ is the divergence $\mathrm{div}\mathbf{T}^t$ of the operator $\mathbf{T}^t$. This tensor $\mathrm{div}\mathbf{T}^t$ is an element of the range space of $\mathbf{T}^t$, in this case the space of functions of five vectors.

The following is the formula which we have derived for the exterior derivative $d(t\mathbf{V})$:

$$d(t\mathbf{V}) = \mathrm{Tr}(\mathbf{T}\nabla\mathbf{V}) + (\mathrm{div}\mathbf{T}^t) \cdot \mathbf{V}, \quad \mathbf{T} = \mathrm{Ep}t. \quad (4.27)$$

The right member is a familiar expression to the student of continuum mechanics, except that the meanings of the symbols are not the familiar meanings. In particular, each term in this equation is a skew-symmetric four-form.

5. CONSTITUTIVE RELATIONS

Suppose that a body manifold S^* and a relativistic universe $\mathcal{E}$ are given. What we shall call a "material state" is a collection $(p, X, \mathbf{H}, \alpha, \cdots)$ of things, one thing corresponding to each entry in a list of interpretations. The first thing is a point p in $\mathcal{E}$, and the second is a point X in S^*. The third is a linear transformation $\mathbf{H}$ of $t\mathcal{E}(p)$ onto $tS^*(X)$, whose null space is timelike. The remaining objects $(\alpha, \cdots)$ are to be tensors of various specified types, over $tS^*(X)$. Their interpretations and the types depend on the kind of material being described; we shall refer to them as "the nonkinematical variables." Given a motion function ζ and given tensor fields α on tS^* to play the role of nonkinematical variables, for each p we call $(p, \zeta(p), \nabla\zeta(p), \alpha(\zeta(p)), \cdots)$ the "state of the material at p during the motion ζ."

To every state $(p, X, \mathbf{H}, \alpha)$ we associate what we shall call an "internal state": $(X, \mathbf{C}, \alpha)$, where X and α are the same as in the given material state, and $\mathbf{C}$ is the Born tensor associated with the $\mathbf{H}$ which appears in the given state. Note that if an internal state $(X, \mathbf{C}, \alpha)$ be given, then each of the material states $(p', X, \mathbf{H}', \alpha)$ which are associated with $(X, \mathbf{C}, \alpha)$ can be obtained from any one $(p, X, \mathbf{H}, \alpha)$ as follows: Let p' be an arbitrary point of $\mathcal{E}$ and let $\mathbf{S}$ be an arbitrary (time-sense-preserving) Lorentz transformation of $t\mathcal{E}(p')$ onto $t\mathcal{E}(p)$. Set $\mathbf{H}' = \mathbf{HS}$. Thus the internal state determines the material state up to location in $\mathcal{E}$ and up to what may be intuitively called the orientation of the material relative to directions in the tangent space.

To "define a type of material" means to us to specify the following things: One must specify $\mathcal{E}$, S^*, and the nature of the nonkinematical variables. One must specify the natures of a number of fields on $\mathcal{E}$, these fields to be called the dynamical variables. One must give a number of objective response functions, defined on the set of all material states. Finally, one must specify a number of "constitutive equations," equations which relate the values of the dynamical variables and the values of the response functions. Given a definition of a type of material, a set of fields on $\mathcal{E}$ will be called an "admissible dynamical motion of the material" if the following is true: The set consists of a motion function ζ, appropriate fields to play the roles of nonkinematical state variables, and appropriate fields to play the roles of dynamical

variables. At each point p on $\mathcal{S}$, the constitutive equations are satisfied by the values at p of the dynamical variables and by the values at the state of the motion $(p, \zeta(p), \nabla\zeta(p), \alpha(\zeta(p)))$ of the response functions.

To do all that is required for definition of a material except specification of the response functions (merely placing restrictions on the choice of response functions) is to define a class of materials. We shall define a class, which we shall call the class of polarization-responsive elastic dielectrics. These materials have just one nonkinematical state variable α, and it is a scalar-valued skew-symmetric two-form over the tangent space tS^* of the body. The dynamical variables include an electromagnetic field ϕ, a dielectric polarization charge-flux potential β, and electromotive intensity E, and an energy–momentum flux t.

The value of the charge-flux potential β is given by a specific response function, for polarization-responsive elastic dielectrics. We have noted that, in general, a skew-symmetric two-form-valued response function $\tilde{\beta}$ which is objective is a function of the following form:

$$\begin{aligned}\tilde{\beta}(\mathbf{U}, \mathbf{V}, p, X, \mathbf{H}, \alpha)\cdots = {} & \hat{\beta}(\mathbf{HU}, \mathbf{HV}, X, \mathbf{C}, \alpha) \\ & + (\omega\cdot\mathbf{U})\hat{\lambda}(\mathbf{HV}, X, \mathbf{C}, \alpha) \\ & - (\omega\cdot\mathbf{V})\hat{\lambda}(\mathbf{HU}, X, \mathbf{C}, \alpha),\end{aligned} \tag{5.1}$$

in which $\hat{\beta}$ and $\hat{\lambda}$ are arbitrary functions of their arguments, except that they must be multilinear and skew symmetric in the vectors. For the specific response function in which we are interested, we set $\hat{\lambda}$ equal to zero, and we set $\hat{\beta}$ equal to α, in the following sense:

$$\tilde{\beta}(\mathbf{U}, \mathbf{V}, p, X, \mathbf{H}, \alpha) = \alpha(\mathbf{HU}, \mathbf{HV}). \tag{5.2}$$

We note that since $\mathbf{HW} = 0$, if $\mathbf{W}$ be substituted into $\tilde{\beta}$ as one argument, the resulting function of the remaining argument vanishes identically. Conversely, given a partial state $(p, X, \mathbf{H}, \cdots)$ with α unspecified, and given any skew two-form β at p such that $\beta(\mathbf{W}, \cdots)$ vanishes identically, there exists a unique α at X such that $\beta(\mathbf{U}, \mathbf{V}) = \alpha(\mathbf{HU}, \mathbf{HV})$ for all $\mathbf{U}$ and $\mathbf{V}$. It follows that if we arbitrarily specify a motion ζ and a field β on $\mathcal{S}$ whose null space at each point contains the world velocity $\mathbf{W}$ of ζ, then for each point p there will exist a unique α and hence a unique material state $(p, \zeta(p), \nabla\zeta(p), \alpha)$, at which the value of the response function $\tilde{\beta}$ is the given field β. Thus, we may equally well think of α or β as the independent variable.

As a second and a third constitutive equation, we adopt the standard relation among the electromagnetic field ϕ, the world velocity $\mathbf{W}$, and the electromotive intensity E, and then postulate that the value of E must be given by a response function. We have noted that a general response function $\tilde{E}$ whose values are one-forms is a function of the following form:

$$\begin{aligned}\tilde{E}(\mathbf{U}, p, X, \mathbf{H}, \alpha) = {} & (\omega\cdot\mathbf{U})\hat{g}(X, \mathbf{C}, \alpha) \\ & + \hat{E}(\mathbf{HU}, X, \mathbf{C}, \alpha).\end{aligned} \tag{5.3}$$

In order that it be possible that there exist some skew-symmetric two-form ϕ such that $\phi(\mathbf{W}, \mathbf{U}) = \tilde{E}(\mathbf{U}, p, X, \mathbf{H}, \alpha)$ for all $\mathbf{U}$, it is necessary that $\tilde{E}(\mathbf{U}, \cdots)$ vanish whenever $\mathbf{U} = \mathbf{W}$. Since $\omega\cdot\mathbf{W}$ equals 1 and $\mathbf{HW}$ equals 0, $\tilde{E}(\mathbf{W}, \cdots) = \hat{g}(\cdots)$. Thus, it is necessary that $\hat{g}$ vanish identically. Conversely, if $\hat{g}$ is identically zero, then the collection of such two-forms ϕ is not empty: $(\omega\wedge\tilde{E})$ is a two-form and

$$(\omega\wedge\tilde{E})(\mathbf{W}, \mathbf{U}) = \tilde{E}(\mathbf{U}, \cdots).$$

Therefore, it is reasonable to restrict the choice of the response function $\tilde{E}$ by insisting that $\hat{g}$ vanish.

Let $\hat{E}$ be any scalar-valued function which is linear as a function of its first argument. Define a response function $\tilde{E}$ as in the right-hand equality below. Then introduce the remaining equations relating E, ϕ, $\mathbf{W}$, and the values of the response function, whenever $(p, X, \mathbf{H}, \alpha)$ is the actual state of the material:

$$\begin{aligned}E\cdot\mathbf{U} &= \tilde{E}(\mathbf{U}, p, X, \mathbf{H}, \alpha) \\ &= (\hat{E}\mathbf{HU}, X, \mathbf{C}, \alpha),\end{aligned} \tag{5.4}$$

$$E\cdot\mathbf{U} = \phi(\mathbf{W}, \mathbf{U}). \tag{5.5}$$

As a fourth and final constitutive equation, we postulate that the value of the energy–momentum flux t be given by a response function. We have noted that a general objective response function $\tilde{t}$ whose values lie in the space $\mathcal{L}_0$ of energy–momentum flux tensors is a function of the following form: For every vector $\mathbf{U}$,

$$\tilde{t}\mathbf{U} = \tilde{p}\mathbf{U} + (\omega\cdot\mathbf{U})\tilde{\rho} + \omega\wedge(\tilde{\sigma}\mathbf{U}) + (\omega\cdot\mathbf{U})\omega\wedge\tilde{h}, \tag{5.6}$$

where $\tilde{p}$, $\tilde{\rho}$, $\tilde{\sigma}$, and $\tilde{h}$ are compositions with $\mathbf{H}$ of multilinear functions $\hat{p}$, $\hat{\rho}$, $\hat{\sigma}$, and $\hat{h}$ on $tS^*(X)$, which depend furthermore on the internal state $(X, \mathbf{C}, \alpha)$. The functions $\hat{p}$, $\hat{\rho}$, $\hat{\sigma}$, and $\hat{h}$ are arbitrary except that $\hat{p}$ is skew symmetric in the last three of its four arguments and $\hat{\sigma}$ is skew symmetric in the last two of its three arguments; $\hat{\rho}$ is skew symmetric in all three of its arguments, and $\hat{h}$ is skew symmetric in its two arguments. We shall restrict the choices of these functions by requiring that, in a sense which we shall make

precise, $\tilde{t}$ is a symmetric tensor and the world velocity vector **W** is an eigenvector of $\tilde{t}$.

One of the consequences of the physical laws to be postulated in Sec. 6 is that the value t of the response function $\tilde{t}$ at a state $(p, X, \mathbf{H}, \alpha)$ must be symmetric if that state is to occur in any dynamic motion with dielectric polarization. We are making the usual requirement that this symmetry condition constitute no restriction on the class of states which can occur. We say that a tensor t in $\mathfrak{L}_0$ is symmetric if for all pairs of vectors **V** and **U** in the tangent space $t\mathcal{E}(p)$,

$$(\mathbf{GV}) \wedge (t\mathbf{U}) = (\mathbf{GU}) \wedge (t\mathbf{V}). \tag{5.7}$$

We call a vector **V** an eigenvector of a tensor t in $\mathfrak{L}_0$ if the three-form $t\mathbf{V}$ has the same direction as **V**, in the sense that every one-form λ which is orthogonal to **V** is also orthogonal to $t\mathbf{V}$. An equivalent condition not involving λ is that **V** be in the null space of $t\mathbf{V}$, that is, that $t(\mathbf{V}, \mathbf{V}, \cdots, \cdots)$ vanish identically as a function of its remaining arguments. A motivation for insisting that **W** be an eigenvector of the energy–momentum tensor is that in most of the known reasonable theories, **W** is an eigenvector, and that in certain simple theories involving clouds of particles, the fact that **W** is an eigenvector is related to highly plausible properties of the flow of particles. One feels that it is normal for **W** to be an eigenvector, and that one needs a definite reason for proposing a theory in which **W** is not.

Theorem: The world-velocity vector **W** is an eigenvector of every tensor value t of an energy–momentum response function $\tilde{t}$ if and only if the "heat flux" function $\tilde{h}$ vanishes identically. Every value t of $\tilde{t}$ is symmetric if and only if these two conditions hold: The "Piola–Kirchoff tensor" **S**, which is defined below and which involves only $\hat{\sigma}$, is symmetric relative to the Born tensor **C**; $\hat{p}$ is a certain linear function of $\hat{h}$. In particular, **W** is an eigenvector of every t and every t is symmetric if and only if both $\hat{p}$ and $\hat{h}$ vanish and **S** is symmetric.

Definition: Given an objective response function $\tilde{t}$, and hence given $\hat{\sigma}$, let **S** be the operator on vectors **U** in $tS^*(X)$ whose value **SU** is a function of one-forms λ over $tS^*(X)$, the value $\lambda \cdot \mathbf{SU}$ of **SU** in turn being a skew-symmetric three-form over $tS^*(X)$—namely, the following three-form:

$$\lambda \cdot \mathbf{SU} = \lambda \wedge (\hat{\sigma}\mathbf{U}). \tag{5.8}$$

[See the discussion of such operators in Sec. 4 and Eqs. (4.14) and (4.15) there.] We call **S** the Piola–Kirchoff tensor. We call **S** symmetric, relative to the Born tensor **C** which occurs as part of the internal state $(X, \mathbf{C}, \alpha)$ on which **S** depends, if the bilinear form obtained by substituting one-forms $\mathbf{CV}$ for the λ above is a symmetric function of the two vectors **V** and **U**.

The symmetry of **S** is the only hard property to ensure when trying to formulate an objective response function $\tilde{t}$ whose values are symmetric and have **W** as eigenvector. Aside from this problem, one can simply use Eq. (3.19), with $\tilde{p}$ and $\tilde{h}$ replaced by zero. For every vector **U**,

$$\tilde{t}U = (\omega \cdot \mathbf{U})\tilde{\rho} + \omega \wedge (\tilde{\sigma}\mathbf{U}), \tag{5.9}$$

where $\tilde{\rho}$ is the composition with **H** of an arbitrary skew-symmetric three-form $\hat{\rho}$ on $tS^*(X)$ which depends on the internal state $(X, \mathbf{C}, \alpha)$, and where $\tilde{\sigma}$ is the composition with **H** of an arbitrary three-form $\hat{\sigma}$, which is skew symmetric in its last two variables, which depends on the internal state.

We shall indicate how to choose the function $\hat{\sigma}$ so that **S** is symmetric, and in the process indicate how to use the elastic response function of a material which describes behavior of the material in terms of classical space–time to write down the corresponding relativistic response function $\tilde{t}$. To define a $\hat{\sigma}$, for each internal state $(X, \mathbf{C}, \alpha)$ and for each pair of vectors **U** and **V** in $tS^*(X)$ we specify the three-form $(\mathbf{CU}) \wedge (\hat{\sigma}\mathbf{V})$. Since this form is to be skew symmetric, it is enough to specify its value at any basis $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$ and then extend the definition to all triples of vectors using the skew-symmetric multilinearity. Let $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$ be any basis; let λ_i be the one-forms in the dual basis, which is to say $\lambda_i \cdot \mathbf{V}_j = \delta_{ij}$; and let $\mathbf{U}_i$ be the vectors such that $\lambda_i = \mathbf{CU}_i$. We arbitrarily choose a three-by-three matrix of real numbers S_{ij}, and set

$$S_{ij} = (\lambda_i \wedge (\hat{\sigma}\mathbf{U}_j))(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3). \tag{5.10}$$

Brief arguments will show that each matrix determines a unique function $\hat{\sigma}$, and the corresponding **S** is symmetric if and only if the matrix is symmetric. Because the $\mathbf{V}_k$ are dual to the λ_i, the expression for the exterior product reduces to one term: If (i, p, q) is the cyclic permutation of (1, 2, 3) which starts with i, the right member is $\hat{\sigma}(\mathbf{U}_j, \mathbf{V}_p, \mathbf{V}_q)$. Thus, Eq. (5.10) assigns a value for $\hat{\sigma}$ to each triple $(\mathbf{U}_j, \mathbf{V}_p, \mathbf{V}_q)$ in which the first vector is any one of the **U**'s and the second and third are from the **V** basis and have their indices (p, q) in natural cyclic order. One can obtain a value for any triple $(\mathbf{U}_j, \mathbf{V}_q, \mathbf{V}_p)$ in which (q, p) is in reverse of cyclic order by introducing a minus sign and reversing the vectors $(\mathbf{V}_q, \mathbf{V}_p)$, and then using Eq. (5.10). Extending $\hat{\sigma}$ by multilinearity, we obtain a unique function of all triples of vectors.

Finally, since $\lambda_i = \mathbf{C}\mathbf{U}_i$, the three-form $(C\mathbf{U}_i) \wedge (\hat{\sigma}\mathbf{U}_j)$ depends symmetrically on the basis vectors $\mathbf{U}_i$ and $\mathbf{U}_j$ if and only if its value S_{ij} at the $\mathbf{V}$ basis depends symmetrically on i and j, and a function is symmetric on pairs of basis vectors if and only if it is symmetric on all pairs of vectors.

By definition of $\mathbf{S}$, S_{ij} is the value of $(\mathbf{C}\mathbf{U}_i) \cdot (\mathbf{S}\mathbf{U}_j)$ at the chosen basis triple $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$. Suppose that one has a Piola–Kirchoff stress tensor $\mathbf{S}_0$ which describes a material relative to classical space–time and relative to some "reference configuration" of S^* in space. $\mathbf{S}_0$ will depend on position X in S^*, on a positive-definite symmetric linear transformation $\mathbf{C}$ of $tS^*(X)$ into its dual space, and on a tensor α over $tS^*(X)$. The tensor $\mathbf{C}$ that is to be substituted into this function $\mathbf{S}_0$, when $\mathbf{S}_0$ is used to find the stress during a motion in classical space–time, is the right Cauchy–Green tensor. One can define a Piola–Kirchoff tensor $\mathbf{S}$ for use during relativistic motions as follows: Choose a basis $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$ of $tS^*(X)$ which corresponds to a right-handed orthonormal basis in the reference configuration. Find the related basis $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$. Then for any internal state $(X, \mathbf{C}, \alpha)$ define the matrix entries S_{ij} to be equal to $(\mathbf{C}\mathbf{U}_i) \cdot (\mathbf{S}_0(X, \mathbf{C}, \alpha)\mathbf{U}_i)$ and then define $\mathbf{S}$ in terms of the S_{ij}. It makes sense to substitute the Born tensor $\mathbf{C}$ into the function $\mathbf{S}_0$ where the right Cauchy–Green tensor belongs, since both tensors are positive-definite symmetric linear transformations of $tS^*(\mathbf{X})$ into its dual.

It is reasonable to identify the Born tensor with the right Cauchy–Green tensor, in that the two kinds of tensors describe in the same way the distortion of the material from the abstract body S^* to the actual world $\mathcal{E}$ during a motion, the right Cauchy–Green tensor when the structure of $\mathcal{E}$ is classical and the Born tensor when the structure of $\mathcal{E}$ is relativistic. In fact, if one has both a classical and a relativistic structure for $\mathcal{E}$, and if the material particle X is at rest relative to the classical structure at some particular event during some motion, then the Born tensor equals the right Cauchy–Green tensor at that event; if the whole body S^* is nearly at rest relative to the classical structure during the whole motion, then the Born tensor nearly equals the right Cauchy–Green tensor during the whole motion. Finally, note that the reason for choosing $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$ in this definition to be a right-handed orthonormal basis relative to the reference configuration is that $\mathbf{S}_0$ depends implicitly on the unit of volume of the reference configuration, whereas the corresponding dependence for $\mathbf{S}$ is the explicit dependence on the triple $(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$. The values of the two functions are related only when the volume associated with the triple is unity.

Proof: Each of the functions $\tilde{\rho}$, $\tilde{p}$, $\tilde{\sigma}$, and $\tilde{h}$ vanishes when any one of its arguments equals $\mathbf{W}$. From this, from the fact that $\omega \cdot \mathbf{W} = 1$, and from Eq. (5.6), we conclude that $\tilde{t}\mathbf{W} = \tilde{\rho} + \omega \wedge \tilde{h}$. Since

$$\tilde{\rho}(\mathbf{W}, \cdots, \cdots) = 0,$$

$\tilde{\rho}$ has the same direction as $\mathbf{W}$. Since

$$(\omega \wedge \tilde{h})(\mathbf{W}, \cdots, \cdots) = \tilde{h}(\cdots, \cdots),$$

$\omega \wedge \tilde{h}$ does not have the same direction as $\mathbf{W}$ unless $\tilde{h}$ vanishes identically. Thus the condition that $\tilde{t}\mathbf{W}$ have the same direction as $\mathbf{W}$ is equivalent to the condition that $\tilde{h}$ vanish.

Equation (5.7), which expresses the condition that $\tilde{t}$ be symmetric, will hold for all $\mathbf{U}$ and $\mathbf{V}$ if it holds in these two special cases: when $\mathbf{U} = \mathbf{W}$ and $\omega \cdot \mathbf{V} = 0$, and when both $\omega \cdot \mathbf{U} = 0$ and $\omega \cdot \mathbf{V} = 0$. Consider the right member of Eq. (5.7) in the case when $\mathbf{U} = \mathbf{W}$. The one-form ω and the numerical factor δ are defined so that $\omega = -\delta(\mathbf{G}\mathbf{W})$. Therefore, for any vector $\mathbf{V}$ such that $\omega \cdot \mathbf{V} = 0$, $-\delta(\mathbf{G}\mathbf{W}) \wedge (\tilde{t}\mathbf{V})$ equals $\omega \wedge (\tilde{p}\mathbf{V})$, plus two terms which drop out because they are proportional to $\omega \cdot \mathbf{V}$, plus $\omega \wedge \omega \wedge (\tilde{\sigma}\mathbf{V})$, which drops out because $\omega \wedge \omega$ is zero. Thus, it equals $\omega \wedge (\tilde{p}\mathbf{V})$ alone. Consider next the left member of Eq. (5.7). Since $(\mathbf{G}\mathbf{V}) \cdot \mathbf{W} = 0$ and $\tilde{\rho}(\mathbf{W}, \cdots, \cdots) = 0$, the form $(\mathbf{G}\mathbf{V}) \wedge \tilde{\rho}$ has a nonzero vector in its null space—namely, $\mathbf{W}$. Since $(\mathbf{G}\mathbf{V}) \wedge \tilde{\rho}$ is a skew-symmetric four-form on a four-dimensional space, it must vanish identically. Since $\tilde{t}\mathbf{W} = \tilde{\rho} + \omega \wedge \tilde{h}$ and we have just noted that the product of $(\mathbf{G}\mathbf{V})$ with the $\tilde{\rho}$ in the right member must be zero, we have $-\delta(\mathbf{G}\mathbf{V}) \wedge (\tilde{t}\mathbf{W}) = -\delta(\mathbf{G}\mathbf{V}) \wedge \omega \wedge \tilde{h} = \delta\omega \wedge (\mathbf{G}\mathbf{V}) \wedge \tilde{h}$. We have shown that the symmetry condition involving pairs $\mathbf{W}$, $\mathbf{V}$ such that $\omega \cdot \mathbf{V} = 0$ is that the exterior product with ω of the left member of the following equation equals the exterior product with ω of the right member:

$$\tilde{p}\mathbf{V} = \delta(\mathbf{G}\mathbf{V}) \wedge \tilde{h}. \tag{5.11}$$

Since $\mathbf{W}$ is in the null space of each member of this equation and $\omega \cdot \mathbf{W}$ is nonzero, this equation holds if and only if the equation holds which one would obtain by multiplying each member by ω. Thus this Eq. (5.11) is equivalent to the symmetry condition involving pairs $\mathbf{W}$, $\mathbf{V}$.

For any two vectors $\mathbf{U}$ and $\mathbf{V}$ such that $\omega \cdot \mathbf{U} = \omega \cdot \mathbf{V} = 0$, consider the exterior product of $\mathbf{G}\mathbf{V}$ with the expression for $\tilde{t}\mathbf{U}$ which is displayed in Eq. (5.6). The first term $(\mathbf{G}\mathbf{V}) \wedge (\tilde{p}\mathbf{V})$ is a skew-symmetric four-form whose null space contains a nonzero vector $\mathbf{W}$; hence this term vanishes. The second and third terms contain the zero factor $\omega \cdot \mathbf{U}$; hence these terms drop

out also. Thus

$$(\mathbf{GV}) \wedge (\tilde{t}\mathbf{U}) = (\mathbf{GV}) \wedge \omega \wedge (\tilde{\sigma}\mathbf{U}) = -\omega \wedge (\mathbf{GV}) \wedge (\tilde{\sigma}\mathbf{U}). \quad (5.12)$$

As a function of **U** and **V** orthogonal to ω, the four-form on the right is a symmetric function if and only if the three-form factor $(\mathbf{GV}) \wedge (\tilde{\sigma}\mathbf{U})$ is symmetric; in fact, if and only if the restriction of $(\mathbf{GV}) \wedge (\tilde{\sigma}\mathbf{U})$ to the space of vectors orthogonal to ω depends symmetrically on **U** and **V**. This is because $\omega \cdot \mathbf{W}$ is nonzero and **W** lies in the null space of all one-forms **GV** and **GU** and all two-forms $\tilde{\sigma}\mathbf{V}$ and $\tilde{\sigma}\mathbf{U}$ for which **U** and **V** are orthogonal to ω.

The space of vectors in $t\mathcal{E}(p)$ orthogonal to ω is the range of the Bressan deformation gradient **A**. The three-form $(\mathbf{GV}) \wedge (\tilde{\sigma}\mathbf{U})$ restricted to the range of **A** is a symmetric function of pairs **V**, **U** in the range of **A** if and only if the composition with **A** in all three arguments of the three-form $(\mathbf{GAV}) \wedge (\tilde{\sigma}\mathbf{AU})$ depends symmetrically on pairs **V**, **U** of vectors in the domain $tS^*(X)$ of **A**. The composition with **A** of the one-form **GAV** is $\mathbf{A}^t\mathbf{GAV} = \mathbf{CV}$. The composition with **A** of the two-form $\tilde{\sigma}\mathbf{AU}$ is $\hat{\sigma}\mathbf{U}$, this being essentially the definition of $\tilde{\sigma}$. We conclude that the symmetry condition on $\tilde{t}$ involving pairs of vectors orthogonal to ω is equivalent to symmetric dependence on pairs **V**, **U** in $tS^*(X)$ of

$$(\mathbf{CV}) \wedge (\hat{\sigma}\mathbf{U}) = (\mathbf{CV}) \cdot (\mathbf{SU}). \quad (5.13)$$

The proof is complete.

6. PHYSICAL LAWS

The Maxwell–Lorentz equations involving the electromagnetic field ϕ and the polarization field β can be expressed as in Eq. (6.1) below, the meaning of which we discuss in the intervening paragraphs. Any skew-symmetric bilinear form γ can be regarded as descriptive of the flux of some scalar quantity in either of two ways, depending on our intuitive conception of the nature of the flow. Our theory would be incomplete without some of the intuitive ideas which we discuss below, concerning flowing stuff associated with the integrals of the two-forms γ on the four-dimensional space $\mathcal{E}$ which occur in our theory.

In the analogous situation when all dimensions are lower by one and visualization is easier, consider a one-form γ defined on a three-dimensional space $\mathcal{E}$. For any directed piece of curve $\mathcal{A}$, we call the value of the integral of γ along $\mathcal{A}$ "the flux across $\mathcal{A}$," and γ itself we call "the density" of the flux. But one does not have a good idea of which direction is "across" $\mathcal{A}$. If we imagine some piece of oriented surface $\mathcal{S}$ of which $\mathcal{A}$ forms a part of the boundary $\partial\mathcal{S}$, then we can imagine stuff flowing in the surface and escaping through the edge. For any such surface $\mathcal{S}$, we can interpret the line integral of γ along $\mathcal{A}$ as the rate of escape of stuff from the interior of $\mathcal{S}$ through the part $\mathcal{A}$ of $\partial\mathcal{S}$. This is admittedly a somewhat sophisticated concept of flow, since the flux $\int_{\mathcal{A}} \gamma$ across $\mathcal{A}$ is determined by γ and $\mathcal{A}$ alone, and hence is the same for all of the many surfaces $\mathcal{S}$ which contain $\mathcal{A}$ as part of their boundaries. As we shall see, however, this is the way the electromagnetic field is commonly treated, and the approach has advantages.

The validity of any laws binding on γ which one makes about conservation of stuff, etc., is independent of how we visualize the flow, so we may in fact use several different visualizations. For each piece of surface $\mathcal{S}$ we have said that the integral of γ along $\partial\mathcal{S}$ represents the total rate of escape of stuff from $\mathcal{S}$. We may think of this escape as occurring from the front face of $\mathcal{S}$, rather than out the edges. In this case, γ is called a "potential" for the flow. Because the dimension of $\mathcal{S}$ is one lower than that of the ambient space $\mathcal{E}$, $\mathcal{S}$ divides the space within a small neighborhood N of any point on $\mathcal{S}$ into two separate regions, with $\mathcal{S}$ between the regions. In each such neighborhood N, we can visualize stuff from one region flowing into the one side of $\mathcal{S}$ and then escaping out the other side. This is a relatively unsophisticated concept of flow, directly related to the original concept, that of a collection of particles moving about in $\mathcal{E}$, with the flux from any surface $\mathcal{S}$ being the number of particles crossing $\mathcal{S}$ per unit time. However, this concept of flow across $\mathcal{S}$, rather than within $\mathcal{S}$ and across its boundary, has the disadvantage that only integrals of γ around complete boundaries $\partial\mathcal{S}$ can be given intuitive meaning, while integrals of γ along pieces $\mathcal{A}$ of curve are unrelated to the physical ideas. Thus the concept of flow across surfaces is by itself likely to be not entirely adequate for relating physical concepts to forms γ.

In four-dimensional space $\mathcal{E}$, three-dimensional hypersurfaces $\mathcal{S}$ locally divided $\mathcal{E}$ into regions on the back side and regions on the front side of $\mathcal{S}$, just as surfaces divide three-dimensional spaces. Thus, when γ is a two-form on a four-dimensional space $\mathcal{E}$, one can visualize stuff flowing from regions on the back side, across $\mathcal{S}$, into regions on the front side, the total rate of flow being $\int \partial\mathcal{S}\gamma$. If one prefers the alternate picture of stuff flowing across pieces $\mathcal{A}$ of surface, one might have trouble deciding what "across" means. As in the lower-dimensional case, one can consider $\mathcal{A}$ to be part of the boundary of an unspecified hypersurface $\mathcal{S}$, and the flow in question to be occurring within $\mathcal{S}$ and out the edge.

The first of the Maxwell–Lorentz equations can be interpreted to mean that if ϕ be regarded as a potential for a flow across hypersurfaces $\mathcal{S}$, the total rate being $\int \partial\mathcal{S}\gamma$, then the flux is zero across every $\mathcal{S}$. Clearly, this mode of visualization of the flow is devoid of intuitive content, so it is better to think of flow across pieces $\mathcal{A}$ of edges of hypersurface. Within hypersurfaces, there is a nontrivial flow in general, although the net outflux from every piece $\mathcal{S}$ is zero.

The usual practice is to consider coordinate systems, in which one of the coordinates is interpreted to be the time seen by some observer. The set of events at which the coordinate time takes a given fixed value, which is interpreted as an instantaneous space by the observer, is a three-dimensional hypersurface. The electromagnetic field ϕ is regarded as the density of fluxes within such instantaneous spaces, across surfaces $\mathcal{A}$ which form parts of boundaries of pieces $\mathcal{S}$ of the instantaneous spaces. The name of the stuff which is considered to be flowing is "magnetic induction." A somewhat-less-frequent practice is to consider in addition some hypersurface pieces $\mathcal{S}$ which do not lie in instantaneous spaces, but rather are timelike in the sense that there are timelike vectors tangent to them. The stuff flowing in such hypersurface pieces is considered to be electrical rather than magnetic in nature and named accordingly. We favor what seems to be the practice in situations like the present, when it is unnatural to separately consider instantaneous spaces and time like hypersurfaces, to use the magnetic names. For arbitrary pieces of hypersurface $\mathcal{S}$, we call the stuff magnetic induction for which ϕ is the flux density.

We regard β as a potential for the flux of a scalar quantity across three-dimensional pieces $\mathcal{S}$ of hypersurface. The name of the flowing stuff is polarization charge. As part of the structure of the universe $\mathcal{E}$, there is a field of linear transformations $\mathbf{K}$ of the space $\mathcal{F}2$ of all skew-symmetric bilinear forms into itself. For a given electromagnetic field ϕ, $\mathbf{K}\phi$ is regarded as a potential for a flux, called the "total charge flux." The second of the Maxwell–Lorentz equations states that the total charge flux equals whatever charge flux we recognize as being present and ascribable to various causes. We are only going to recognize charge flux due to polarization.

In summary, the Maxwell–Lorentz equations are these: For every orientable piece of hypersurface $\mathcal{S}$ with boundary $\partial\mathcal{S}$,

$$\oint_{\partial\mathcal{S}} \phi = 0, \quad \oint_{\partial\mathcal{S}} \mathbf{K}\phi = \oint_{\partial\mathcal{S}} \beta. \tag{6.1}$$

We can describe the linear transformation $\mathbf{K}$ by listing two of its properties, or by giving its components in a coordinate system. We define a multilinear form $\hat{K}$ which is canonically associated with $\mathbf{K}$: For an arbitrary ordered fourtuple $U^* = (\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4)$ of vectors, consider the two-form $\gamma = (\mathbf{GU}_1) \wedge (\mathbf{GU}_2)$ defined as follows in terms of the metric tensor $\mathbf{G}$ and the first two vectors in U^*:

$$\gamma(\mathbf{X}, \mathbf{Y}) = (\mathbf{X} \cdot \mathbf{GU}_1)(\mathbf{Y} \cdot \mathbf{GU}_2) - (\mathbf{Y} \cdot \mathbf{GU}_1)(\mathbf{X} \cdot \mathbf{GU}_2). \tag{6.2}$$

Apply $\mathbf{K}$ to the two-form γ, and then substitute the remaining vectors $(\mathbf{U}_3, \mathbf{U}_4)$ into the result $\mathbf{K}\gamma$. The resulting number is the value of $\hat{K}$ at U^*. That is,

$$\hat{K}(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4) = \mathbf{K}[(\mathbf{GU}_1) \wedge (\mathbf{GU}_2)](\mathbf{U}_3, \mathbf{U}_4). \tag{6.3}$$

The first of the two properties by which we describe $\mathbf{K}$ is that the form $\hat{K}$ is skew symmetric in all four of its variables. It follows from this that in every coordinate system the components of $\hat{K}$ are a multiple of the permutation symbols ϵ_{ijkl}. The second property of $\mathbf{K}$ is that it is covariant constant, relative to the differentiation operator for which $\mathbf{G}$ is covariant constant. Equivalently, $\hat{K}$ is covariant constant. It follows that there is a constant a_0 such that the components of $\hat{K}$ are $a_0(|g|)^{\frac{1}{2}}$ times ϵ_{ijkl} in right-handed systems, and minus these values in left-handed systems, where g is the determinant of the matrix of components of $\mathbf{G}$. We may presume a_0 to be positive, either by making this the definition of which coordinate systems are to be called right handed, or by listing positivity of a_0 as a third property of $\mathbf{K}$. From Eqs. (6.2) and (6.3), we find that the components of $\mathbf{K}$ are

$$K^{ij}_{pq} = \tfrac{1}{2} g^{ri} g^{sj} \hat{K}_{rspq} = \tfrac{1}{2} a_0 (|g|)^{\frac{1}{2}} g^{ri} g^{sj} \epsilon_{rspq} \tag{6.4}$$

in right-handed systems, where the g^{ij} are the components of the contravariant metric tensor $\mathbf{G}^{-1}$, and where the factor $\frac{1}{2}$ arises because there are two terms in the right member of Eq. (6.2). The value of a_0 and the name which one gives it depend on the choice of a system of units. We note that what is involved are the dimensional constants ϵ_0, μ_0, and c, and numerical factors.

In addition to the Maxwell–Lorentz equations, we impose equations of conservation of energy, momentum, and angular momentum as physical laws, to be binding on collections (ζ, ϕ, β, t) of fields which make up dynamical processes with dielectric polarization. Let $\mathbf{W}$ denote the world-velocity vector field of the motion ζ, and let $E = \phi\mathbf{W}$ denote the electromotive intensity one-form.

We avoid introduction of individual forces, energy densities and such at this point, introducing instead the functionals of an arbitrary world-velocity field **U** whose values are the energies that would be acquired in an arbitrary region $\mathfrak{W}$ of $\mathcal{E}$ by a hypothetical material moving with velocity **U**, acquired because of the action of these forces and energy densities and such upon the material. At a point where the following of the world line of a particle of the hypothetical material would take one from the outside into $\mathfrak{W}$, any energy possessed by that particle is being carried into $\mathfrak{W}$. Pursuing this idea, one sees that if one were to introduce an energy density, this density would correspond to a transfer of energy to the material in $\mathfrak{W}$, for it to possess while it is in $\mathfrak{W}$. The amount of transfer would depend on the direction of **U**, into, out of, or along the boundary of $\mathfrak{W}$. The idea of acquisition of energy by a material through interaction with a given force field, the amount acquired being dependent on the velocity, is more conventional than the idea of a material acquiring energy during a certain time period by inheritance from itself of an earlier time. This idea for force fields goes back to the classical mechanics of a particle moving through an applied force field, where the work done on the particle is the force times the velocity.

There is an ambiguity possible in exhibiting a functional and saying that its values represent energy generated by interaction of a velocity field **U** and some other field. When it is possible to think of the other field possessing a store of energy, as one does think of an electromagnetic field, then one has to specify whether the generated energy is acquired by the material or by the other field, acquired by both, or gained by one and lost by the other. The functionals we will introduce all represent energy acquired by the material. Although it may be meaningful and worthwhile to keep accounts of the energies stored in various fields in addition to accounting for the energy possessed by the material, we have no occasion to do so.

Let t be an energy–momentum flux-type of field, a scalar-valued multilinear function of four vectors which is skew symmetric in its last three arguments. For every oriented four-dimensional region $\mathfrak{W}$, with smooth boundary $\partial\mathfrak{W}$, and for every smooth field of timelike future-directed unit vectors **U**, we say that "the work done on the material in $\mathfrak{W}$," or, "the net energy transferred into $\mathfrak{W}$ by the interaction of **U** and t," is the value w given by

$$w = \int_{\partial\mathfrak{W}} t\mathbf{U}, \tag{6.5}$$

where $t\mathbf{U}$ denotes the three-form which one obtains by substituting **U** in as the first argument of t.

Let ϕ be any electromagnetic-field type of field ($\phi \in \mathcal{F}2$), and let γ be any charge-flux density-type of field ($\gamma \in \mathcal{F}3$). For every oriented four-dimensional region $\mathfrak{W}$ and for every world-velocity field **U**, we say that "the work w' done on the material in $\mathfrak{W}$," or, "the energy generated in $\mathfrak{W}$ by the interaction of ϕ, γ, and **U**," is the value w' given by

$$w' = \int_{\mathfrak{W}} (\phi\mathbf{U}) \wedge \gamma, \tag{6.6}$$

where $(\phi\mathbf{U})$ denotes the one-form which results from substitution of **U** into ϕ as the first argument. For a given charge-flux potential β, we put $\gamma = d\beta$ into this integral to obtain the energy generated, it being presumed that $d\beta$ is defined, if not everywhere, at least "almost everywhere," at enough points to use in integration over four-dimensional regions.

The equation which we call "conservation of energy" states that the total energy transferred in and generated within because of interactions of t, ϕ, and β with the actual world velocity **W** must be zero for every oriented region $\mathfrak{W}$:

$$0 = \int_{\partial\mathfrak{W}} t\mathbf{W} + \int_{\mathfrak{W}} E \wedge d\beta. \tag{6.7}$$

The equation which we call "conservation of momentum" states that, as a functional of an arbitrary world velocity **U** and of a region $\mathfrak{W}$, the total work due to interaction with t, ϕ, and β has the form of an integral over the interior of $\mathfrak{W}$ of a fixed linear function of the covariant derivative $\nabla\mathbf{U}$ of the velocity. That is, for given t, ϕ, and β there exists a tensor s which is a linear function of linear transformations **L** and whose values $s(\mathbf{L})$ are four-forms. For all **U** and all $\mathfrak{W}$,

$$\int_{\mathfrak{W}} s(\nabla\mathbf{U}) = \int_{\partial\mathfrak{W}} t\mathbf{U} + \int_{\mathfrak{W}} (\phi\mathbf{U}) \wedge d\beta. \tag{6.8}$$

Postulating the existence of an s satisfying this equation is equivalent to postulating the existence of suitable limits of the right member as the region $\mathfrak{W}$ is allowed to shrink down on points and postulating that the values of these limits depend on **U** only through a linear dependence on $\nabla\mathbf{U}$. In any case, it is clear that when a tensor s exists, it is determined by t, ϕ, and β. The equation which we call "conservation of angular momentum" states that the tensor s depends only on the symmetric part of its linear transformation argument **L**. That is,

$$s(\mathbf{L}) = 0 \quad \text{if} \quad (\mathbf{GL})^t + (\mathbf{GL}) = 0, \tag{6.9}$$

the expression in the second equation being what we mean by the symmetric part of **L** relative to the metric

tensor $\mathbf{G}$, aside from inconsequential factors, like $\frac{1}{2}$ and $\mathbf{G}^{-1}$.

7. DIFFERENTIAL EQUATIONS OF MOTION

Suppose we have a collection of fields (ζ, ϕ, β, t) which satisfy the integral equations introduced in Sec. 6 as physical laws, and hence form a dynamical motion with dielectric polarization. For any region $\mathcal{W}$ throughout which all of the integrands of the integrals over $\partial\mathcal{W}$ are continuously differentiable, these integrals can be replaced by the integrals over $\mathcal{W}$ itself of the exterior derivatives of the original integrands. An equation among integrals over $\mathcal{W}$ being valid for arbitrary regions $\mathcal{W}$ containing a point p at which all of the integrands are continuous implies that the integrands at p satisfy the same equation.

The conservation of momentum equation, Eq. (6.8), thus implies the following equation, in which we use the expression for $d(t\mathbf{V})$ given in Eq. (4.27) and the expression for $(\phi\mathbf{V}) \wedge d\beta$ given in Eq. (4.18):

$$s(\nabla\mathbf{V}) = \mathrm{Tr}(\mathbf{T}\nabla\mathbf{V}) + \mathbf{V} \cdot \mathrm{div}\mathbf{T}^t - \mathbf{V} \cdot (\phi\mathbf{J}) \otimes \eta. \tag{7.1}$$

The energy equation, Eq. (6.7), implies that when $\mathbf{V}$ equals the world velocity $\mathbf{W}$ of ζ, the right member of this equation vanishes. We shall show that these differential forms of the equations of conservation of energy and momentum, together with the fact of conservation of angular momentum [that $s(\mathbf{L})$ vanishes whenever $\mathbf{GL}$ is skew symmetric], are equivalent to the following equations:

$$0 = \mathrm{div}\mathbf{T}^t - (\phi\mathbf{J}) \otimes \eta, \tag{7.2}$$

$$(\mathbf{GU}) \cdot (\mathbf{TV}) = (\mathbf{GV}) \cdot (\mathbf{TU}), \tag{7.3}$$

$$\mathrm{Tr}(\mathbf{T}\nabla\mathbf{W}) = 0, \tag{7.4}$$

$$\int_{\mathcal{W}} \mathrm{Tr}(\mathbf{T}\nabla\mathbf{V}) = \int_{\partial\mathcal{W}} t\mathbf{V} - \int_{\mathcal{W}} (\mathbf{V} \cdot \phi\mathbf{J})\eta. \tag{7.5}$$

The second equation expresses what we mean in saying that $\mathbf{T}$ is symmetric. It can be expressed in terms of the forms $t\mathbf{V}$ and $t\mathbf{U}$ in $\mathcal{F}3$ instead of $\mathbf{TV}$ and $\mathbf{TU}$ in $\mathcal{F}$, at the cost of using exterior multiplication instead of the dot product: $(\mathbf{GU}) \wedge (t\mathbf{V}) = (\mathbf{GV}) \wedge (t\mathbf{U})$. In addition, the following differential equations are clearly implied by the electromagnetic integral equations, Eq. (6.1):

$$d\phi = 0, \quad d(\mathbf{K}\phi) = d\beta. \tag{7.6}$$

The equivalence of Eqs. (7.2)–(7.5) with Eq. (7.1) follows from the fact that the values of $\nabla\mathbf{V}$ and $\mathbf{V}$ at a given point p can be chosen essentially independent of each other. That is, one can show that, given any future-directed timelike vector $\mathbf{V}_0$ at a point p and any linear transformation $\mathbf{L}_0$ of $t\mathcal{E}(p)$ into itself such that $(\mathbf{GL}_0)^t\mathbf{V}_0 = 0$, there exists a smooth field $\mathbf{V}$ defined on some neighborhood of p such that the values at p of $\mathbf{V}$ and $\nabla\mathbf{V}$ are $\mathbf{V}_0$ and $\mathbf{L}_0$, and such that the value of $\mathbf{V}$ at every point is a future-directed unit vector. One constructs first any vector field taking the required value $\mathbf{V}_0$ and value $\mathbf{L}_0$ of the derivative. One shows that normalizing this field to unit magnitude does not affect the values of itself and its derivative at p. Furthermore, if desired, one can easily replace a timelike future-directed unit vector-field $\mathbf{V}$ defined only on a neighborhood of p by a field defined on all of $\mathcal{E}$, which takes the same values as $\mathbf{V}$ throughout some smaller neighborhood of p. Thus a field $\mathbf{V}$ exists on $\mathcal{E}$ having prescribed values of $\mathbf{V}$ and $\nabla\mathbf{V}$ at p.

Choosing to set $\nabla\mathbf{V} = 0$ in Eq. (7.1), we find that the dot product of $\mathbf{V}$ with the right member of Eq. (7.2) must be zero. Its inner product with every $\mathbf{V}$ being zero, the tensor itself must be zero. Thus, Eq. (7.2) is valid, and our original equation reduces to the terms which do not depend on $\mathbf{V}$: $s(\nabla\mathbf{V}) = \mathrm{Tr}(\mathbf{T}\nabla\mathbf{V})$. Since this equation depends linearly on $\nabla\mathbf{V}$, and since every linear transformation $\mathbf{L}$ can be expressed as a linear combination of $\mathbf{L}$'s for which $(\mathbf{GL})^t\mathbf{V} = 0$ for some timelike vector $\mathbf{V}$, the equation $s(\mathbf{L}) = \mathrm{Tr}(\mathbf{TL})$ holds for all $\mathbf{L}$. In particular, for any pair of vectors $\mathbf{U}$ and $\mathbf{V}$, this is true for the $\mathbf{L}$ defined by

$$\mathbf{LX} = (\mathbf{GU} \cdot \mathbf{X})\mathbf{V} - (\mathbf{GV} \cdot \mathbf{X})\mathbf{U}.$$

Since this $\mathbf{L}$ is skew symmetric, $s(\mathbf{L})$ vanishes; hence $\mathbf{TL}$ is traceless. The reader will recognize and be able to prove for himself that $\mathrm{Tr}(\mathbf{TL})$ is the difference between the left and the right members of Eq. (7.3); hence that equation is valid. Conversely, since every skew-symmetric $\mathbf{L}$ is a linear combination of the kind defined here in terms of $\mathbf{U}$ and $\mathbf{V}$, if Eq. (7.3) holds and hence $s(\mathbf{L}) = 0$ for these $\mathbf{L}$, then $s(\mathbf{L}) = 0$ for all skew-symmetric $\mathbf{L}$.

We have shown that the momentum and angular momentum equations are equivalent to Eqs. (7.2) and (7.3). Given Eq. (7.2), the last two terms in Eq. (7.1) add up to zero for any field $\mathbf{V}$, hence in particular for $\mathbf{V} = \mathbf{W}$. The energy equation requires that the sum of all three terms vanish, hence the first term vanishes; this is Eq. (7.4). Finally, Eq. (7.5) follows from the fact that $\mathrm{Tr}(\mathbf{T}\nabla\mathbf{V})$ is equal to, and hence may be substituted for, $s(\nabla\mathbf{V})$ in the original integral equation, Eq. (6.8).

To derive differential equations of motion for a given material, one would have to substitute into

Eqs. (7.2)–(7.5) the constitutive relations which were postulated in Sec. 5, and then carry out some of the indicated differentiations to obtain explicit functional equations relating the values and values of derivatives of the unknown functions. For $\mathrm{div}\mathbf{T}^t$ in Eq. (7.2), one would have to examine the definition in Eq. (4.25) to work out a formula in terms of partial derivatives of components of $\tilde{t}$ or something similar, since a formula useful for calculating $\mathrm{div}\mathbf{T}^t$ has not been derived. The symmetry condition, Eq. (7.3), would be automatically satisfied and hence need not be considered, and $\mathrm{Tr}(\mathbf{T}\nabla\mathbf{W})$ in the next equation is already in a suitable form. The treatment of the derivatives of electromagnetic variables in Eq. (7.6) would depend on exactly what functions are known and what are unknown. The integral relation in Eq. (7.5) would be used only in case jump discontinuities occur, to derive jump conditions.

8. ELECTRIC, MAGNETIC, DIELECTRIC, AND CURRENT VECTORS

The representation of dielectric polarization in terms of a skew-symmetric two-form β and of polarization charge flux in terms of its exterior derivative cannot be considered well known. Even the better-known process of representing the electromagnetic field in terms of a two-form ϕ is fraught with uncertainties due to the complexity of possible choices of sign conventions and systems of units. It will be efficient to display one mode of representing β and ϕ and related objects in terms of a polarization vector **P**, electric and magnetic vectors **E** and **B**, and related vectors. The reader can then make suitable modifications to obtain other modes of representation.

The exposition will be restricted to the case of the special relativistic universe, the vector representations being of less interest in more general universes. We choose a fixed t, x, y, z coordinate system which is right-handed, and index the coordinates 0, 1, 2, 3. We choose the signature of the metric tensor **G** so that its components g_{ij} are given by $g_{00} = -c^2$, $g_{11} = g_{22} = g_{33} = 1$, and other $g_{ij} = 0$. If the reader prefers coordinates in the order x, y, z, t, then he will have to modify some of the representation formulas, since for a given set of axes in space–time the two orders, t, x, y, z and x, y, z, t yield coordinate systems of opposite orientations. One can pass from a representation appropriate to one system to a representation appropriate to the other by replacing some of the vectors with their negatives, such as the polarization vector **P**. However, since our representations have been adjusted to involve the accepted classical vectors, one would want to use these same vectors and achieve the sign changes by such readjustments of the formulas as reversing the order of the vectors **W** and **P** in the representation formula for β.

Let η be the standard hypervolume four-form, and let ϵ_{ijrs} be the permutation symbols. Since in the chosen coordinate system the determinant of the components g_{ij} is $-c^2$, the components of η are given by

$$\eta_{ijrs} = c\epsilon_{ijrs}. \tag{8.1}$$

Let W^i be the components of the world-velocity vector **W**. Let $\gamma = cW^0$ and let V^i be the components of the classical world-velocity vector **V**, obtained from **W** by factoring out the positive scalar W^0. We have

$$W^i = \gamma V^i/c, \quad V^0 = 1. \tag{8.2}$$

Theorem: Given a world-velocity vector **W** and a dielectric polarization two-form β, there exists a unique vector **P** which is spacelike in the sense that $P^0 = 0$ and which satisfies the following representation identity:

$$\eta(\mathbf{P}, \mathbf{W}, \mathbf{X}, \mathbf{Y}) = \gamma\beta(\mathbf{X}, \mathbf{Y}),$$

$$\text{i.e.,} \quad P^iW^j\eta_{ijrs} = \gamma\beta_{rs}. \tag{8.3}$$

The relevant facts are that W^0 is not zero and that **W** is in the null space of β, i.e., that $W^r\beta_{rs} = 0$. The condition $P^0 = 0$ is but a normalization, which we can impose because **P** is determined by the identity only up to the addition of a multiple of **W**. Inclusion of the factor γ is but a normalization of the magnitude of **P**, imposed because it yields the historically correct polarization vector.

Proof: Assuming that such a spacelike vector **P** exists, it is easy to determine its components P^i in terms of the β_{rs}. For all pairs (r, s) of which neither index is zero, the sum over i and j in Eq. (8.3) contains only two terms in which η_{ijrs} is nonzero, of which the term having $i = 0$ vanishes because $P^0 = 0$. In the surviving term $W^j = W^0 = \gamma/c$, which cancels conveniently. Thus, making suitable choices for (r, s), we have

$$P^1 = \beta_{32}, \quad P^2 = \beta_{13}, \quad P^3 = \beta_{21}. \tag{8.4}$$

We begin proving the theorem by defining P^i by Eq. (8.4). Then Eq. (8.3) holds for (r, s) equal to (3, 2), (1, 3), and (2, 1) by definition, hence for all pairs (r, s) in which both r and s differ from zero, because both members of the equation are skew-symmetric functions of (r, s). Again by skew symmetry, we shall have shown that Eq. (8.3) holds in the remaining cases, when one or both of r and s are zero,

when we show that it holds whenever $r = 0$ and $s \neq 0$. The following equation is trivially valid because **W** is in the null space of β and η is skew symmetric, hence each member is zero:

$$P^i W^j W^r \eta_{ijrs} = \gamma W^r \beta_{rs}. \tag{8.5}$$

Since $s \neq 0$, the coefficient of each W^r on the left for which r is nonzero equals the corresponding coefficient on the right, by the previous results. Since these terms thus cancel, the single remaining term on the left, in which $r = 0$, equals its counterpart on the right. Dividing by the nonzero factor W^0, we obtain the final case of Eq. (8.3) needed to complete the proof.

The exterior derivative $d\beta$, which plays the role of a charge-flux form, can be represented like any skew-symmetric three-form as follows in terms of η and a world vector **J**:

$$(1/c)J^s \eta_{sijk} = (d\beta)_{ijk}. \tag{8.6}$$

Let Q denote the scalar J^0, and call it the polarization charge density. Since $Q\mathbf{V}$ is a multiple of the world-velocity vector **V** of the material, we can think of it as representing the flux of some charge (the polarization charge) which moves with the material. The remaining part of the actual charge flux is represented by $\mathbf{P}^*$, defined by $\mathbf{J} = Q\mathbf{V} + \mathbf{P}^*$. Note that $\mathbf{P}^*$ is spacelike: $P^{*0} = 0$.

To determine J^l for each l and hence determine Q and $\mathbf{P}^*$ in terms of **P** and **V** and their partial derivatives, let *ijk* be the remaining integers in any order such that *lijk* is an even permutation. Since $\epsilon_{lijk} = 1$, for each such *ijk* we have the first equation below, the second being a standard formula:

$$(P^l V^i - P^i V^l) = \beta_{jk}. \tag{8.7}$$

$$(d\beta)_{ijk} = \partial_i \beta_{jk} + \partial_j \beta_{ki} + \partial_k \beta_{ij}, \tag{8.8}$$

wherein ∂_s denotes partial differentiation with respect the sth coordinate. We obtain the term $\partial_i \beta_{jk}$ for the second equation by applying ∂_i to the first equation. Since any cyclic permutation of the positions of i, j, k in the fourtuple *lijk* will yield a permutation which is even, we obtain valid equations from Eq. (8.7) by replacing *ijk* by *jki* and applying ∂_j, and by replacing *ijk* by *kij* and applying ∂_k. We thus obtain three of the four terms in the following sum over s, the fourth term being trivially zero because it is a derivative of $(P^l V^l - P^i V^l)$:

$$(d\beta)_{ijk} = \partial_s (P^l V^s - P^s V^l). \tag{8.9}$$

When *lijk* is even, the sum over s in Eq. (8.6) has only one nonzero term, the term when $s = l$, and the value of η_{lijk} is c. Thus, the sum equals J^l; hence $(d\beta)_{ijk}$ and hence the right member of Eq. (8.9) equals J^l. Setting $l = 0$ and noting that $P^0 = 0$ and $V^0 = 1$, we find that

$$Q = -\partial_s P^s = -\mathrm{div}\mathbf{P}, \tag{8.10}$$

wherein we may restrict the sum over s to the three spacelike indices, treating **P** essentially as a spatial vector rather than a world vector. Setting $l = i$ and thinking of it as a spacelike index, carrying out the differentiation indicated in Eq. (8.9), removing the term which equals QV^i, and restricting the range of summation to the spacelike indices while separately including the terms with timelike index, we obtain

$$P^{*i} = \partial P^i/\partial t + (\partial_j P^i)V^j - P^j(\partial_j V^i) + P^i(\partial_j V^j). \tag{8.11}$$

This may be compared term by term with Eq. (4.4) of Toupin,[1] and the conclusion drawn that the vector $\mathbf{P}^*$ coincides with his vector of the similar name.

The connection between the electromagnetic field ϕ and electric and magnetic vector fields **E** and **B** derives from the following formula for the Lorentz force per unit charge:

$$\mathbf{F} = \mathbf{E} + \mathbf{V} \times \mathbf{B}. \tag{8.12}$$

The vectors **F**, **E**, and **B** are spacelike vectors; we can take them to be four-dimensional vectors with their first components equal to zero. **V** is the classical world velocity of the charged particle on which the force acts, the vector whose components are the derivatives with respect to coordinate time of the spatial coordinates of the position. In accordance with the discussion in Sec. 6, we consider the work which this force would do on the particle if the velocity were not **V** but an arbitrary vector **U**. Let u be the positive scalar such that $(u/c)\mathbf{U}$ is a unit vector. The rate of working would be the product of **F** and **U**, multiplied by the relativistic correction factor u. In terms of components, and without the correction factor, this is

$$F_i U^i = E_i U^i + \epsilon_{0ijk} U^i V^j B^k \tag{8.13}$$

The reader is invited to check that the second term is the triple scalar product of the space parts of **U** and **V** and of **B**, and insert any constant factors required by his system of units.

The comparable energy source postulated in Sec. 6 is the integral of a four-form which depends upon the electromagnetic field ϕ, an arbitrary world-velocity **V**, and a charge-flux form γ. We put $(u/c)\mathbf{U}$ in place of **V** as the name of the arbitrary world-velocity

vector. If q denotes the charge, then the current vector corresponding to the actual velocity vector $\mathbf{V}$ is $q\mathbf{V}$, and the charge-flux form is $\gamma = \mathrm{Ct}(q\mathbf{V} \otimes \eta)$. Thus the form τ whose integral is the energy source is

$$\tau = (qu/c)(U^i \phi_{ij} V^j)\eta. \tag{8.14}$$

To compare the two expressions, we remove the charge q and the relativistic correction factor u. The integral of τ is the integral of its 0123 component with respect to the coordinates; hence we obtain the energy source per unit volume per unit time by replacing η in the right member by its component $\eta_{0123} = c$, cancelling the $(1/c)$ factor. In the resulting expression $U^i \phi_{ij} V^j$, the term in which $i = 0$ has no counterpart in Eq. (6.6). Since $U^0 = 1$, this term is independent of $\mathbf{U}$ and remains even when the spatial components of $\mathbf{U}$ vanish. Therefore this term represents an energy which is acquired by the material by inheritance from itself at earlier times, rather than as a result of work being done on the material. We omit the term with $i = 0$ and identify the remaining with the Lorentz rate of working in Eq. (6.6).

Upon equating coefficients of U^i, we obtain two expressions for the Lorentz force, for $i \neq 0$:

$$F_i = \phi_{ij} V^j = E_i + \epsilon_{lijk} V^j B^k. \tag{8.15}$$

For given ϕ, this is satisfied for $V^0 = 1$ and V^1, V^2, V^3 arbitrary. Thus, for i and j not equal to 0, we have

$$\phi_{i0} = E_i, \quad \phi_{ij} = \epsilon_{0ijk} B^k. \tag{8.16}$$

The remaining components of ϕ are determined by these, since $\phi_{ji} = -\phi_{ij}$. Aside from constant factors related to choices of systems of units, these expressions may be the negatives of what the reader is accustomed to using. To correct the theory which appears in this paper, the reader should change the definitions of electromotive intensity and of energy due to interaction with an electromagnetic field in Secs. 3 and 6. Instead of putting the world-velocity vector $\mathbf{W}$ or $\mathbf{V}$ into ϕ as its first argument, he should put it in second position. The expression for the Lorentz force in Eq. (8.15) will become $V^j \phi_{ji} = -\phi_{ij} V^j$, and minus signs will appear in Eqs. (8.16). Various authors differ in this, so that, for given $\mathbf{E}$ and $\mathbf{B}$, one man's ϕ is another man's $-\phi$.

ACKNOWLEDGMENTS

This research was supported in part by the U.S. Air Force Office of Scientific Research under grant AFOSR 68-1538 to Mellon Institute. The background reading and development of the theory was done during June and July of 1968 at Mellon Institute. The paper was written during the following academic year at the Mathematics Department of the University of Kentucky.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Entropy Principle for the Derivation of a New Cluster Expansion

G. V. RAMANATHAN,* J. M. DAWSON, AND M. D. KRUSKAL
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey

(Received 22 September 1966; Revised Manuscript Received 11 July 1969)

A product expansion for the s-particle distribution function (s arbitrary) is derived by means of a variational principle involving the entropy associated with the $(s+1)$-particle distribution function.

1. INTRODUCTION

In an earlier paper,[1] a new method of defining correlations in terms of a product expansion of distribution functions was introduced to describe the statistical mechanics of systems of interacting particles, and some of its advantages were demonstrated. It was pointed out that in an equilibrium system these correlations can be most simply viewed as potentials of average forces.

In the present paper we wish to show that this product expansion can be derived from a variational principle for an entropy function. Given an $(s-1)$-particle distribution function, we seek that s-particle distribution function which is consistent with it and maximizes the entropy. We then view the correlations produced by particle interactions as deviations from this maximum-entropy distribution.

The basic idea of the method can be illustrated by the following simple example. Suppose that the one-body distribution $f_1^{(1)}$ is given and it is required to determine the two-body distribution $\hat{f}_{12}^{(2)}$ which maximizes the entropy. For the entropy we shall use the negative of the "H" function

$$H^{[2]} = \frac{1}{V^2}\int d^3x_1\, d^3p_1\, d^3x_2\, d^3p_2 \hat{f}_{12}^{(2)} \log \hat{f}_{12}^{(2)}. \quad (1.1)$$

The problem is to find that $\hat{f}_{12}^{(2)}$ which minimizes $H^{[2]}$ under the constraint

$$f_1^{(1)} = \frac{1}{V}\int d^3x_2\, d^3p_2 \hat{f}_{12}^{(2)}. \quad (1.2)$$

It can readily be seen that

$$\hat{f}_{12}^{(2)} = f_1^{(1)} f_2^{(1)}, \quad (1.3)$$

corresponding to complete absence of correlation (and indeed to statistical independence). We introduce a two-body correlation $\alpha_{12}^{(2)}$ by relating $\hat{f}_{12}^{(2)}$ to the actual two-body distribution $f_{12}^{(2)}$ governed by the dynamics of the system:

$$f_{12}^{(2)} = \hat{f}_{12}^{(2)}(1 + \alpha_{12}^{(2)}). \quad (1.4)$$

Alternatively, we can write this in the form

$$f_{12}^{(2)} = \hat{f}_{12}^{(2)} \exp \psi_{12}^{(2)}, \quad (1.5)$$

where $\psi_{12}^{(2)}$ can be interpreted, in an equilibrium system,[2] as a potential of average force.

Analogously, given the $(s-1)$-body distribution function $f^{(s-1)}$, we seek the s-body function $\hat{f}^{(s)}$ that minimizes the quantity

$$H^{[s]} = \frac{1}{V^s}\int \prod_{i=1}^{s} d^3x_i\, d^3p_i \hat{f}^{(s)} \log \hat{f}^{(s)}. \quad (1.6)$$

(The computation of $\hat{f}^{(s)}$ for $s > 2$ is presented in Sec. 4.) Then we relate $\hat{f}^{(s)}$ to the actual $f^{(s)}$ through an s-particle correlation $\alpha^{(s)}$ or an s-body potential of average force $\psi^{(s)}$:

$$f^{(s)} = \hat{f}^{s}(1 + \alpha^{(s)}) = \hat{f}^{(s)} \exp \psi^{(s)}. \quad (1.7)$$

Here $\hat{f}^{(s)}$ can be viewed as the minimal correlation function at the s level. In this paper an explicit solution is obtained for $\hat{f}^{(s)}$ in the limit of infinite volume and finite correlation length. At the three-particle level the maximum-entropy distribution turns out to be the same as that given by the Kirkwood superposition approximation, and the form of (1.7) to be the same as the one proposed in Ref. 1. This last can, therefore, be viewed as a generalization of the superposition approximation.[3]

A few remarks about the method are in order at this point. First of all, what we present here is an *ad hoc* procedure enabling one to define certain functions, which, from an intuitive point of view, can be called correlations.

Secondly, it may be questioned whether (1.6) is the best form for the negative "entropy." Such a question is irrelevant to our problem. The purpose of this procedure is merely to define certain functions and this cannot lead to any loss of generality.

Thirdly, the whole procedure could be carried out without essential change even in the absence of symmetry, i.e., in effect with a variety of species of particles.

* Present address: The Courant Institute for Mathematical Sciences, New York University.

[1] G. V. Ramanathan, J. Math. Phys. 7, 1507 (1966).

[2] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), p. 324.

[3] Reference 2, p. 328.

2. NOTATION AND DEFINITIONS

We consider a system of N identical particles in a region of volume V. We denote by $\{p\}_r^s$ a generic set of p particles chosen from the collection of $(s-r+1)$ particles with labels $r, r+1, \cdots, s$. The functions $f^{(p)}_{\{p\}_r^s}$, $\alpha^{(p)}_{\{p\}_r^s}$, $\psi^{(p)}_{\{p\}_r^s}$, etc., are assumed to be symmetric functions of the positions and momenta of the chosen p particles. By $f^{(p+n)}_{\{p\}_r^s,a_1,a_2,\cdots,a_n}$ we denote a symmetric function of the positions and momenta of $(p+n)$ particles with the indicated labels. Sometimes when there is no ambiguity, we may drop either the superscripts or the subscripts in these functions.

The N-particle distribution function $D^{(N)}_{\{N\}_1^N}$ is normalized to unity by

$$\int \prod_{i=1}^{N} d^3x_i\, d^3p_i D^{(N)}_{\{N\}_1^N} = 1. \tag{2.1}$$

The reduced distribution function $f^{(s)}_{\{s\}_1^s}$ $(1 \leq s \leq N)$ is, as usual, defined by

$$f^{(s)}_{\{s\}_1^s} = V^s \int \prod_{i=s+1}^{N} d^3x_i\, d^3p_i D^{(N)}_{\{N\}_1^N}. \tag{2.2}$$

3. THE VARIATIONAL PRINCIPLE

Let $f_{\{s-1\}_1^{s-1}}$ be given. We seek $\hat{f}_{\{s\}_1^s}$ minimizing

$$H^{[s]} = \frac{1}{V^s} \int \prod_{i=1}^{s} d^3x_s\, d^3p_s \hat{f}_{\{s\}_1^s} \log \hat{f}_{\{s\}_1^s},$$

subject to the constraints

$$f_{\{s-1\}_1^s} = \frac{1}{V} \int d^3x_k\, d^3p_k f_{\{s\}_1^s}, \tag{3.1}$$

where k is the index omitted on the left. These conditions (3.1) are all equivalent because of the assumed symmetry. Using a common symmetric Lagrange multiplier function $l_{\{s-1\}_1^s}$ for all these conditions in order to preserve the formal symmetry, we obtain the variational condition

$$1 + \log \hat{f}_{\{s\}_1^s} + \sum_{\text{all } \{s-1\}_1^s} l_{\{s-1\}_1^s} = 0. \tag{3.2}$$

Setting

$$q_{\{s-1\}_1^s} = -s^{-1} - l_{\{s-1\}_1^s}, \tag{3.3}$$

we can write (3.2) as

$$\hat{f}_{\{s\}_1^s} = \exp \sum_{\text{all } \{s-1\}_1^s} q_{\{s-1\}_1^s}. \tag{3.4}$$

To determine $q_{\{s-1\}_1^s}$, we substitute (3.4) into any one of the constraint conditions given by (3.1). Setting $k = s$, we have

$$f_{\{s-1\}_1^{s-1}} = \frac{1}{V} \int d^3x_s\, d^3p_s \exp\{\textstyle\sum q_{\{s-1\}_1^s}\}. \tag{3.5}$$

For orientation, it may be noted that, in the case of noninteracting particles in thermal equilibrium, we have

$$f_1 = \frac{1}{V}\left(\frac{\beta}{\pi}\right)^{\frac{3}{2}} \exp(-\beta v_1^2),$$

where $v = p/m$ (m is the mass of the particle), $\psi_1 = \log f_1$, $\psi_{12} = \psi_{123} = \cdots = 0$, $q_{\{s\}_1^s} = s^{-1} \sum_{r=1}^{s} \psi_r$, $\lambda_{12} = \lambda_{123} = \cdots = 0$ (λ quantities defined later).

Although it does not seem possible to solve (3.5) for $q_{\{s-1\}_1^s}$ explicitly, it would not seem unreasonable to conjecture that a unique solution exists.[4] Here, however, we merely solve explicitly in the limit of infinite volume with the correlation length remaining finite.

4. SOLUTION IN THE LIMIT OF INFINITE VOLUME

A. Three-Body Case

We shall first solve (3.5) for the special case $s = 3$ and recover Kirkwood's superposition approximation. It is convenient to introduce functions ψ given by

$$f_1 \equiv \exp \psi_1, \tag{4.1}$$

so that

$$\frac{1}{V} \int d^3x_1\, d^3p_1 \exp \psi_1 = 1. \tag{4.2}$$

Then f_{12} is given by

$$f_{12} \equiv \exp(\psi_1 + \psi_2 + \psi_{12}). \tag{4.3}$$

Now (3.4) is

$$\hat{f}_{123} = \exp(q_{12} + q_{13} + q_{23}) \tag{4.4}$$

and, taking the logarithm, (3.5) becomes

$$q_{12} = \psi_1 + \psi_2 + \psi_{12} - \log \frac{1}{V} \int d^3x_3\, d^3p_3 \times \exp(q_{13} + q_{23}). \tag{4.5}$$

It was mentioned earlier that ψ_{12} as defined by (4.3) can be interpreted as a potential of average force when the system is at thermal equilibrium. Its range and strength will be governed by the density of the system and the range and strength of the interparticle potential. It is reasonable to assume that as we let the volume of the system become infinite, keeping the density constant, the range R of ψ_{12} will remain finite. We shall solve for q in terms of ψ in the limit of R^3/V tending to zero.

[4] H. Grad, Commun. Pure Appl. Math. 14, 323 (1961).

It is convenient to write q in terms of a one-body function and a two-body function,

$$q_{12} \equiv \lambda_1 + \lambda_2 + \lambda_{12}. \tag{4.6}$$

Here we have the freedom to choose λ_1 as we please and (4.6) can then be viewed as a definition of λ_{12}. Substituting for q in (4.5),we obtain

$$\begin{aligned} &2(\lambda_1 + \lambda_2) + \lambda_{12} \\ &\quad = \psi_1 + \psi_2 + \psi_{12} \\ &\qquad - \log\left[\frac{1}{V}\int d^3x_3\, d^3p_3 \exp\{2\lambda_3 + \lambda_{13} + \lambda_{23}\}\right]. \end{aligned} \tag{4.7}$$

Making the (seemingly simplest) choice $\lambda_1 = \frac{1}{2}\psi_1$, this becomes

$$\lambda_{12} = \psi_{12} - \log\left\{\frac{1}{V}\int d^3x_3\, d^3p_3 \exp(\psi_3 + \lambda_{13} + \lambda_{23})\right\}. \tag{4.8}$$

The choice made for λ_1 is also natural because, in the noninteracting thermal-equilibrium case, the solution of (4.8) is $\lambda_{12} = 0$. It can be verified at once that

$$\lambda_{12} = \psi_{12} + O(R^3/V) \tag{4.9}$$

is a solution of (4.8) by substituting and noting that finite regions around particles 1 and 2 do not contribute significantly, while

$$\frac{1}{V}\int d^3x_3\, d^3p_3 \exp\psi_3 = 1.$$

Then

$$\begin{aligned} \hat{f}_{123} &= \exp(\psi_1 + \psi_2 + \psi_3 + \psi_{12} + \psi_{13} + \psi_{23}) \\ &= f_{12}f_{13}f_{23}/f_1f_2f_3. \end{aligned} \tag{4.10}$$

This is just the Kirkwood superposition approximation.

The actual three-body distribution f_{123} can be written

$$f_{123} \equiv \hat{f}_{123} \exp\psi_{123}, \tag{4.11}$$

which defines the new three-body function ψ_{123}.

B. s-Body Case

Let us suppose that $f^{(s-1)}$ has been written in the form

$$f^{(s-1)} = \exp\left\{\sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^{s-1}} \psi^{(r)}_{\{r\}_1^{s-1}}\right\}, \tag{4.12}$$

where every $\psi^{(r)}$ with $r \geq 2$ has a finite range. When we take the logarithm, (3.5) becomes

$$\begin{aligned} q^{(s-1)}_{\{s-1\}_1^{s-1}} &= \sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^{s-1}} \psi^{(r)}_{\{r\}_1^{s-1}} \\ &\quad - \log\frac{1}{V}\int d^3x_s\, d^3p_s \\ &\quad \times \exp\left\{\sum_{\text{all}\{s-2\}_1^{s-1}} q^{(s-1)}_{\{s-2\}_1^{s-1},s}\right\}. \end{aligned} \tag{4.13}$$

Write

$$q_{\{s-1\}_1^{s-1}} = \sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^{s-1}} \lambda_{\{r\}_1^{s-1}};$$

substituting this in (4.13) and collecting the various terms, we get

$$\begin{aligned} &\sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^{s-1}} (s-r)\lambda^{(r)}_{\{r\}_1^{s-1}} \\ &\quad = \sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^{s-1}} \psi^{(r)}_{\{r\}_1^{s-1}} \\ &\qquad - \log\frac{1}{V}\int d^3x_s\, d^3p_s \\ &\qquad \times \exp\left\{\sum_{r=0}^{s-2}\sum_{\text{all}\{r\}_1^{s-1}} (s-1-r)\lambda^{(r+1)}_{\{r\}_1^{s-1},s}\right\}. \end{aligned}$$

We are free to and do choose

$$\lambda^{(r)} = (s-r)^{-1}\psi^{(r)} \tag{4.14}$$

for all r, $1 \leq r \leq s-2$. Then we have

$$\begin{aligned} \lambda^{(s-1)}_{\{s-1\}_1^{s-1}} &= \psi^{(s-1)}_{\{s-1\}_1^{s-1}} - \log\frac{1}{V}\int d^3x_s\, d^3p_s \\ &\quad \times \exp\left\{\psi^{(1)}_s + \sum_{r=1}^{s-3}\sum_{\text{all}\{r\}_1^{s-1}} \psi^{(r+1)}_{\{r\}_1^{s-1},s}\right. \\ &\quad \left. + \sum_{\text{all}\{s-2\}_1^{s-1}} \lambda^{(s-1)}_{\{s-2\}_1^{s-1},s}\right\}. \end{aligned} \tag{4.15}$$

From the assumption of finite range and the normalization condition

$$\frac{1}{V}\int d^3x_s\, d^3p_s \exp\psi_s = 1,$$

we obtain

$$\lambda_{\{s-1\}_1^{s-1}} = \psi_{\{s-1\}_1^{s-1}} + O(R^3/V) \tag{4.16}$$

or

$$q_{\{s-1\}_1^{s-1}} = \sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^{s-1}} (s-r)^{-1}\psi_{\{r\}_1^{s-1}} + O(R^3/V). \tag{4.17}$$

Substituting this into (3.4),we obtain

$$\hat{f}_{\{s\}_1^s} = \exp\left\{\sum_{r=1}^{s-1}\sum_{\text{all}\{r\}_1^s} \psi_{\{r\}_1^s}\right\} + O(R^3/V), \tag{4.18}$$

and now we write

$$f_{\{s\}_1^s} = \hat{f}_{\{s\}_1^s} \exp \psi_{\{s\}_1^s}, \tag{4.19}$$

introducing the new s-body function $\psi_{\{s\}_1^s}$ through a relation between the actual distribution and the maximum entropy distribution. This is nothing but (4.12) with s replacing $s-1$. It can be readily seen that (4.19) is in the form suggested in Ref. 1.

5. FINITE-VOLUME CORRECTIONS

Equation (4.15) is in a convenient form for computing the finite-volume corrections by iteration. To next order in R^3/V we have

$$\lambda_{\{s-1\}_1^{s-1}} = \psi_{\{s-1\}_1^{s-1}} - \log\left[\frac{1}{V}\int d^3x_s\, d^3p_s \sum_{r=0}^{s-2} \sum_{\text{all } \{r\}_1^{s-1}} \psi^{(s+1)}_{\{r\}_1^{s-1},s}\right] + O((R^3/V)^2). \tag{5.1}$$

In particular, for the three-body case [see Eq. (4.8)],

$$\lambda_{12} = \psi_{12} - \log\left[\frac{1}{V}\int d^3x_3\, d^3p_3 \exp(\psi_3 + \psi_{13} + \psi_{23})\right] + O((R^3/V)^2). \tag{5.2}$$

6. REMARKS

We have shown that the Kirkwood superposition approximation and its generalization to higher distribution functions can be derived through an entropy principle, in the limit that the volume becomes infinitely large while the correlation lengths remain finite.

We may expect this generalization of the Kirkwood superposition approximation to give a rapidly improving description of the system with increasing s. This approximation takes the potential of the average force for s particles to be a sum of potentials involving less than s particles. The s-particle potential, $\psi_{\{s\}_1^s}$, may be thought of as resulting from a distortion of the average potentials of all the other particles due to the introduction of s. This distortion should be large only if the interaction with all the $(s-1)$ particles is strong. However, since only a few particles can get close enough so that all interact strongly at the same time, $\psi_{\{s\}_1^s}$ should drop off very rapidly for s greater than 3 or 4 even in dense gases and liquids. Unfortunately, the equations determining the ψ are hard to analyze and the verification of this conjecture probably requires extensive numerical calculations. Some calculations have been made of the equation of state for hard spheres using the generalized formula up to ψ_{123}.[5] It has been found that this approximation gives better results for the two-particle correlation function (as compared with Monte Carlo data) than other commonly used approximations.

ACKNOWLEDGMENTS

This work was performed under the auspices of the Air Force Office of Scientific Research of the Office of Aerospace Research, under Contract No. AF 49(638)-1555.

One of the authors (G. V. R.) wishes to acknowledge partial support by AEC Contract No. AT(30-1)1480.

[5] Y. T. Lee, F. H. Ru, and J. Ru, J. Chem. Phys. **48**, 3506 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 1 JANUARY 1970

Spinor Structure of Space-Times in General Relativity. II

ROBERT GEROCH*
Department of Mathematics, Birkbeck College, London, England

(Received 9 June 1969)

Spinor fields can only be defined on a space-time which has been given a spinor structure. A number of conditions (some sufficient, others necessary and sufficient) for the existence of a spinor structure are derived. By applying one or another of these conditions, it is shown that many well-known solutions of Einstein's equations do have spinor structure. The question of the existence of spinor structure depends only on the topology of the underlying manifold, not on the (time- and space-oriented) metric. It is shown that, nonetheless, a certain "threshold" of curvature must be exceeded before there can be even the possibility of a space-time's having no spinor structure.

INTRODUCTION

Before spinor fields may be defined on a space-time M, it is first necessary to endow M with some further structure—called the spinor structure—in addition to the Lorentz metric. Certain space-times cannot be given any spinor structure at all, and, furthermore, even when there exists a spinor structure it will not, in general, be unique.[1] However, there is a gedanken experiment,[2] based on the quantum-mechanical properties of spin-$\frac{1}{2}$ particles, by which one could, in principle, determine a definite spinor structure for our own universe. Thus, the existence of a spinor structure appears, on physical grounds, to be a reasonable condition to impose on any cosmological model in general relativity. In the present paper we shall investigate the global restrictions this condition places on possible cosmological models.

It was shown in I that a necessary and sufficient condition that a noncompact space-time M have spinor structure is that M may be given a global system of orthonormal tetrads. While this criterion represents a strong condition to be satisfied by cosmological models, it is not always the most convenient way to decide whether or not a given space-time has spinor structure. The Schwarzschild solution, for example, has spinor structure, but it is not immediately clear that this space-time may be given a system of tetrads. [It would not do, for example, to choose two of the vectors to lie along the r and t axes (in the usual coordinates), for then the other two vectors would have to lie in the 2-spheres $r = \text{const}, t = \text{const}$, which is impossible.] In Sec. I we develop some criteria for the existence of spinor structure, based on the neighborhoods of certain 2-spheres in M. With each such 2-sphere S we associate an index, defined as the number of times that S intersects a surface obtained by slightly deforming S. That this index be even for each S in M is a necessary and sufficient condition that M have spinor structure. We may also characterize the existence of spinor structure in terms of the type of the Weyl tensor. It is shown that a space-time whose Weyl tensor is everywhere type [1, 1, 1, 1], [2, 1, 1], [3, 1], or [4] necessarily has spinor structure. In the "generic" case, however, we would expect that the Weyl tensor would be algebraically general everywhere except in certain regions of lower dimensionality. It turns out that the behavior of these regions in the large is sufficient to determine whether or not the space-time has spinor structure. Finally, we show that every space-time which arises from initial-value data—that is, every space-time which has a Cauchy surface—also has spinor structure. The goal of Sec. I is to obtain a list of criteria which may be used to test in practice whether a given space-time has spinor structure (or, equivalently, a global system of tetrads). Most common solutions of Einstein's equations satisfy one or another of the conditions in Sec. I.

In Sec. II we discuss the relation of spinor structure to the amount of curvature present in the space-time. That there should be any relation at all is somewhat surprising, for the existence or nonexistence of a spinor structure is a property only of the underlying manifold, independent of the metric (provided only that the metric is time and space oriented). We display a curvature integral over 2-spheres: that this integral be less than a certain value is a sufficient (but not necessary) condition for the existence of spinor structure. Intuitively, we may think of the integral as representing a threshold condition on the curvature: the threshold must be exceeded before there is even the possibility that the space-time have no spinor structure. This integral represents one of the few

* U.S. Air Force Office of Scientific Research postdoctoral fellow. Present address: Dept. of Physics, Syracuse University, Syracuse, N.Y.

[1] See, for example, R. Penrose, "The Structure of Space-Time," in *Battelle Recontres in Mathematics and Physics: Seattle, 1967*, C. DeWitt and J. A. Wheeler, Eds. (W. A. Benjamin, Inc., New York, 1968); R. Geroch, J. Math. Phys. 9, 1739 (1968); this paper will be referred to hereafter as I.

[2] Y. Aharonov and L. Susskind, Phys. Rev. **158**, 1237 (1967); Ref. 1.

situations in which, without imposing symmetries, the curvature of a space with an indefinite metric is known to have a bearing on the global structure of the space.

I. CRITERIA FOR THE EXISTENCE OF SPINOR STRUCTURE

By a *space–time* M we understand a 4-dimensional manifold with a metric of signature $(+, -, -, -)$. We shall assume, furthermore, that M is noncompact and both space and time oriented.[3] The noncompactness assumption is reasonable from the physical point of view because every compact space–time is known[4,5] to have closed timelike curves. If, on the other hand, M were not space and time oriented, then we could always find a covering space of M—representing exactly the same physical universe—which is.[5]

Let S be a fixed (not necessarily spacelike) 2-surface in M which is topologically equivalent to a 2-sphere, but which may cross itself at isolated points.[6] By slightly deforming S, we obtain another 2-sphere S' in M. In general, two 2-dimensional surfaces in a 4-dimensional manifold will intersect in a region of dimension zero, i.e., in points. Choose S' so that it intersects S only at isolated, nondegenerate points $p_k, k = 1, 2, \cdots, m$, that is, so that S and S' have no common tangent vectors at their points of intersection. Let us assign an orientation to S, whence S', which is obtained by deforming S, is also assigned a definite orientation. Now at each intersection point p_k, the vectors tangent to the oriented surfaces S and S' span the set of all vectors at p_k, and so define an orientation of this 4-dimensional vector space. Define $\iota(p_k)$ to be $+1$ if this orientation is he same as that of M, and -1 it it is the opposite. The *index* of the surface S is defined by[7]

$$I(S) \equiv \sum_{k=1}^{m} \iota(p_k).$$

If we reverse the orientation originally assigned to S, the orientation of S' is also reversed, and so the index is unchanged. Furthermore, the index of S is independent of the distortion by which S' is obtained from S, for under any further deformation of S' new points of intersection are created in pairs whose ι-values are -1 and $+1$. Finally, the index must be continuous under deformations of S, and so, since $I(S)$ takes only integral values, the index is invariant under such deformations. Note also that the definition of the index does not involve the metric on M.[8]

Alternatively, the index of S may be characterized in the following way. Consider any vector field ξ^a which is nonvanishing in a neighborhood of S and which is tangent to S only at nondegenerate points. Define a new 2-surface S', obtained from S by moving a small distance along the trajectories of ξ^a. The intersection points of S and S' now correspond precisely to the points at which ξ^a is tangent to S. Thus, the index of S is equal to the number of times (properly counted with regard to sign) that ξ^a is tangent to S.

The relationship between the index and spinor structure is given by the following result: *A necessary and sufficient condition that M have spinor structure is that the index I(S) be even for every 2-surface S, topologically a 2-sphere, in M.* To prove this statement, we make use of the fact (see I) that there exists a neighborhood of S which is topologically equivalent to just one of a certain collection of 4-dimensional manifolds M_m, $m = 0, 1, 2, \cdots$. (If S happens to cross itself, then any neighborhood of S will overlap itself. We must then count each point in the overlap region twice, that is, we treat the neighborhood as though it did not intersect itself.) The M_m are defined as follows. Let A_1 and A_2 be two copies of the Cartesian product of a closed unit 2-disc (polar coordinates θ_i and r_i, $i = 1, 2$) and a 2-dimensional plane (Cartesian coordinates x_i and y_i). Join A_1 and A_2 across their boundaries ($r_1 = 1$ and $r_2 = 1$, respectively) by identifying the point $(\theta_1, r_1, x_1, y_1)$ of A_1 and the point $(\theta_2, r_2, x_2, y_2)$ of A_2 whenever[9]

$$r_1 = r_2 = 1, \quad x_1 = x_2 \cos(m\theta_1) - y_2 \sin(m\theta_1),$$
$$\theta_1 = \theta_2, \quad y_1 = y_2 \cos(m\theta_1) + x_2 \sin(m\theta_1).$$

The 2-sphere S in M is to correspond to the 2-sphere S_m in M_m given by $x_i = y_i = 0$. It was shown in I that

[3] A space–time M is said to be *time oriented* if the light cones of M are divided into two systems, past and future, and *space oriented* if any collection of three independent spacelike vectors at a point which are all orthogonal to a single timelike vector is assigned a definite parity, $+1$ or -1.

[4] E. Kronheimer and R. Penrose, Proc. Cambridge Phil. Soc. **63**, 481 (1967).

[5] R. Geroch, J. Math. Phys. **8**, 782 (1967).

[6] We shall regard such a crossing point as representing two distinct points of S, each of which must be treated independently of the other. More precisely, S represents a mapping from a 2-sphere into M, not just the image of such a mapping.

[7] This index is well known in homology theory. See, for example, P. S. Aleksandrov, *Combinatorial Topology* (Graylock Press, Albany, N.Y., 1960), Vol. 3, p. 73.

[8] The various properties of the index are most easily understood by considering a lower-dimensional case: closed curves on a 2-dimensional manifold. There is, however, a curious property of intersections of 2-surfaces in a 4-dimensional manifold which is not present in the 2-dimensional case. The surfaces in Minkowski space given (in the usual coordinates) by $y = z^2 - t^2$, $x = tz$ and $y = -z^2 + t^2$, $x = -zt$ intersect only at the origin. If, however, we distort either surface slightly, the number of intersections increases to two: there is no way—as there would be in the analogous situation in two dimensions—slightly to distort the surfaces so that they do not intersect.

[9] The corresponding expressions were given incorrectly in I: the term $m\varphi_1$ there should have been $m\theta_1$.

the space-time M has spinor structure if and only if, for every S in M, the corresponding M_m has an even value for m. Consider the vector field in M_m whose components in the coordinate patches A_1 and A_2 are

$$\xi^{r_1} = 0, \quad \xi^{x_1} = (r_1)^2 \cos(m\theta_1) + [1 - (r_1)^2],$$

$$\xi^{\theta_1} = (r_1)^2[1 - (r_1)^2], \quad \xi^{y_1} = (r_1)^2 \sin(m\theta_1),$$

and

$$\xi^{r_2} = 0, \quad \xi^{x_2} = 1,$$

$$\xi^{\theta_2} = 0, \quad \xi^{y_2} = 0,$$

respectively. The field ξ^a is tangent to S_m at the m (nondegenerate) points $r_1 = 2^{-\frac{1}{2}}$, $x_1 = y_1 = 0$, $\theta_1 = \pi/m, 3\pi/m, \cdots, (2m-1)\pi/m$, and so the index of S_m is simply m (or $-m$, depending on the orientation chosen for M_m). Thus, the index of S is even precisely when m is even, and our theorem follows.

Choosing ξ^a to be everywhere timelike, we see that the index of S is determined by the number of times that the light cone "cuts across" S. In particular, if S is spacelike everywhere, then its index is necessarily zero, because a timelike ξ^a will never be tangent to S. Consider, for example, the Schwarzschild solution. The 2-spheres $r = \text{cost}$, $t = \text{const}$, in the usual coordinates, are spacelike. Furthermore, every 2-sphere S in the Schwarzschild solution may be continuously deformed to one of these 2-spheres, perhaps described several times. (In other words, the second homotopy group has a set of spacelike generators.) Therefore, the index of every S is even, and so the Schwarzschild solution has a spinor structure. As another application of our theorem, we see that any space-time in which every S may be contracted to a point (i.e., whose second homotopy group vanishes) necessarily has spinor structure,[10] for the index, which is invariant under continuous deformations of S, clearly vanishes when S is a small sphere in some coordinate patch. In particular, the Robertson-Walker models, the Gödel solution, the fluid-ball solutions, and the plane waves all have spinor structure.[11]

A further criterion for the existence of spinor structure can be obtained in terms of the algebraic properties of the Weyl tensor. It is known that if the Weyl tensor C_{abcd} is type[12] [1, 1, 1, 1], [2, 1, 1], or [3, 1] at a point p, then C_{abcd} determines four orthonormal vectors at p. (For types [1, 1, 1, 1] and [2, 1, 1], the vectors are completely determined by the principal null directions, but this is not the case for type [3, 1].) These vectors are uniquely defined up to sign, but are not assigned any particular ordering. That is to say, if we select one of the vectors to be labeled η^a and keep track of η^a around some closed curve, then, on our return to the starting point, it may be that η^a now coincides with one of the other vectors of the tetrad. Furthermore, the vectors may change discontinuously when the type of the Weyl tensor changes. If, for example, C_{abcd} is type [1, 1, 1, 1] in some region R_1 and type [2, 1, 1] in some other region R_2, then the tetrad defined by the Weyl tensor will not be continuous across the boundary region between R_1 and R_2.

We first show that the space-time M must have spinor structure provided its Weyl tensor is *everywhere* type [1, 1, 1, 1], [2, 1, 1], or [3, 1]. We may take M to be simply connected (i.e., such that every closed curve in M may be contracted to a point) because the universal covering space of M, which is always simply connected, has spinor structure if and only if M does (see I). At every point of M we have an orthonormal tetrad of vectors defined by the Weyl tensor. If we carry any vector of this tetrad continuously around some closed curve γ in M, then we must return with the same element of the tetrad. This property certainly holds in the limit as γ is contracted to a point, and is invariant under continuous deformations of γ, and so, since M is simply connected, it holds for all closed curves γ. The directions defined by the Weyl tensor therefore constitute a global system of orthonormal tetrads, and so M must have a spinor structure.

In particular, certain of the Robinson-Trautman solutions[13] are everywhere type [2, 1, 1] or [3, 1]. We conclude that these space-times have a spinor structure.

A similar, but slightly more complicated, argument suffices to show that a space-time must also have spinor structure if its Weyl tensor is everywhere type [4]. In this case the Weyl tensor defines, at each point, a null direction l^a and a pair of orthogonal null 2-planes containing l^a. We may therefore choose a triad of independent vector fields x^a, y^a, and l^a

[10] This result follows directly from the usual characterization of spinor structure in terms of Stiefel-Whitney classes. See, for example, J. Milnor, L'Enseignement Math. **9**, 198 (1963); K. Bichteler, J. Math. Phys. **9**, 813 (1968).

[11] Each of these spaces have topology R^4, except for certain of the Robertson-Walker models (those with positive spatial curvature) whose topology is $S^3 \times R$.

[12] We shall adopt Penrose's notation for the six types of Weyl tensors. The relation to Petrov's notation is given by [1, 1, 1, 1] = I, [2, 1, 1] = II, [3, 1] = III, [2, 2] = 0, [4] = N, [–] = 0. For a discussion of the classification of the Weyl tensor, see, for example, R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960); R. Penrose, W. Rindler, and R. Geroch, *The Spinor Approach to Space-Time* (Cambridge University Press, to be published).

[13] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **265A**, 463, 1962.

on M, where x^a lies in one of the 2-planes defined by the Weyl tensor and y^a lies in the other. To construct an orthonormal tetrad on M, we first choose an arbitrary unit timelike vector field t^a. Two of the spacelike vectors of the tetrad are obtained by projecting x^a and y^a into the space orthogonal to t^a at each point, while the third spacelike vector is that vector orthogonal to the other three. Thus, a type [4] space–time must have spinor structure.

It is not possible to construct an orthonormal tetrad in this way from a type [2, 2] or [–] Weyl tensor. Of course, a similar analysis could be carried out for the stress–energy tensor—or, indeed, for any tensor which has been defined on M. However, the Weyl tensor is perhaps the most convenient object for this purpose because it is always available and, except in very special cases, is nonzero.

In the case of a "generic" space–time, we would not expect that the Weyl tensor would be the same type throughout the entire space. The collection of all tensors at a point which have both the symmetries of the Weyl tensor and vanishing contractions is ten-dimensional. The tensors of type [1, 1, 1, 1] form a ten-dimensional subset of this collection, those of type [2, 1, 1] an eight-dimensional subset, and those of type [3, 1], [2, 2], [4], and [–] subsets of six dimensions or fewer. Hence, in the generic case, we would expect that the Weyl tensor C_{abcd} would be type [1, 1, 1, 1] everywhere except in some 2-dimensional region D. Let S be any 2-sphere in M, chosen to intersect D in nondegenerate isolated points. If we move in a small circle on S about one such point, the tetrad defined by the Weyl tensor will undergo a rotation (relative to a fixed tetrad at the intersection point) which, in the generic case, is through an angle of 2π. Adding these rotations over all of S, we obtain the total rotation of the tetrad. It was shown in I that a necessary and sufficient condition that the index of S be even (i.e., that there exist a tetrad in a neighborhood of S) is that this total rotation be an even multiple of 2π. Thus, M will have spinor structure if and only if D intersects every 2-sphere S in M an even number of times (with nondegenerate intersection points).

Consider, for example, the algebraically general Weyl solutions.[14] The Weyl tensor is type [1, 1, 1, 1] everywhere except on the symmetry axis ($r = 0$ in the usual coordinates), where C_{abcd} is type [2, 2]. In the full four-dimensional space–time, the symmetry axis defines a 2-surface. The tetrad defined by C_{abcd}, on being taken around the axis, undergoes one complete rotation. (This property follows from the axial symmetry: it is not necessary to calculate the Riemann tensor explicitly.) But every 2-sphere S in the Weyl solution which intersects the symmetry axis at nondegenerate points does so an even number of times. (The sphere about the singular region, for example, intersects $r = 0$ at one positive and one negative value of z.) Thus, the Weyl solutions have spinor structure.

Finally, consider a space–time M which may be expressed in the form of a topological product of a 3-surface with the real line such that each of the embedded 3-surfaces is spacelike. That is to say, suppose that there is some scalar field t on M such that (i) $\nabla_a t$ is timelike (and nonzero) everywhere, and (ii) the spacelike 3-surfaces Q_t, given by $t = \text{const}$, are topologically identical where, for any t and t', Q_t is mapped onto $Q_{t'}$ by following along the trajectories of $\nabla^a t$. Since M is space oriented, the surface Q_0 is oriented. But every oriented 3-manifold may be given a global system of triads.[15] Choose the elements of this triad to be covariant vectors in M (lowering indices with the metric on M) and extend the vector fields to all of M by imposing the requirement that the Lie derivative of each vector of the triad with respect to $\nabla^a t$ be zero. We thus obtain four independent vector fields on M such that, at each point, three of the vectors are spacelike and orthogonal to the timelike vector $\nabla_a t$. Finally, orthogonalize the tetrad (e.g., by the Gram–Schmidt orthogonalization procedure). Thus, every space–time which satisfies the above "topological-product" condition has spinor structure. In particular, every space–time which has a Cauchy surface, which is globally hyperbolic, or which is asymptotically simple with null $\mathfrak{I}$ is known[16] to satisfy the topological-product condition. We conclude that any one of these three conditions implies the existence of a spinor structure.

Suppose we begin with initial-value data for Einstein's equations (with or without sources) on a spacelike 3-surface Q. This initial data defines some space–time M. It may be that M can be further extended (as is the case, for example, in the Reissner–Nordström solution), but the metric in the extension will not be determined uniquely by the data on Q. That is, Q is necessarily a Cauchy surface for the space–time M, and so M must have a spinor structure. Thus, we see that the initial data on Q will never

[14] See, for example, J. L. Synge, *Relativity: The General Theory* (North-Holland Publ. Co., Amsterdam, 1960), p. 312.

[15] See, for example, N. Steenrod, *The Topology of Fibre Bundles* (Princeton University Press, Princeton, N.J., 1951), p. 204.

[16] In fact, every space–time with a Cauchy surface is a topological product (R. Penrose, "An Analysis of the Structure of Space–Time," preprint; S. W. Hawking, "Singularities and the Geometry of Space–Time," preprint; R. Geroch, "The Domain of Dependence," to be published in J. Math. Phys.); global hyperbolicity is completely equivalent to the existence of a Cauchy surface ("The Domain of Dependence"); and asymptotic simplicity with null $\mathfrak{I}$ is sufficient (but not necessary) for global hyperbolicity.

suffice to "predict" that at some later time no spinor structure will be possible. The absence of spinor structure can occur only in a space-time which has been extended in some way not defined by initial data.

II. SPINOR STRUCTURE AND CURVATURE

In our discussion of spinor structure so far we have been concerned only with the topological properties of the space-time M. The relationship of spinor structure to the Weyl tensor, for example, did not directly involve the amount of curvature present, but rather merely the qualitative behavior of certain directions constructed from the metric. That the existence of spinor structure should be related to the topological properties of tensor fields is not particularly suprising, for it is only the topology of the underlying manifold—not the choice of metric—which determines whether or not there will be a spinor structure. We shall now show that the degree of curvature is also related to the question of the existence of spinor structure. A certain minimum amount of curvature—expressed in the form of a surface integral—is necessary if there is to be even the possibility of a space-time's having no spinor structure.

The integral expression itself, while too complicated to be of much use in deciding whether or not a space-time has spinor structure, is of interest primarily because it provides a definite link between spinor structure and curvature. Thus, certain manifolds—the M_{2n+1} of Sec. I for example—are themselves sufficiently "twisted" already that any metric defined on them must contain at least a certain minimum amount of curvature. It may even turn out to be true that all exact source-free solutions of Einstein's equations have spinor structure. (As far as the author is aware, all known solutions whose global structure has been analyzed do have spinor structure.)

Consider a (noncompact, space- and time-oriented) space-time M, and let S be a 2-surface, topologically a 2-sphere, in M. The idea is to attempt to construct a tetrad on S using a certain kind of "parallel transport" in M. Such a construction must succeed if the curvature is sufficiently small in a neighborhood of S. Choose a point p of S and a one-parameter family $\gamma_s(w)$ of curves on S, where $s \in [0, 1]$ labels the individual curves and $w \in [0, 1]$ is a parameter along each curve. The curves are to all begin and end at p, and each point of S (except p) is to lie on exactly one curve. The curves $s = 0$ and $s = 1$ correspond to "zero curves" which remain at p (Fig. 1). Let ω^a and σ^a denote the vectors tangent to the lines $s = \text{const}$ and $w = \text{const}$, respectively, where ω^a and σ^a are

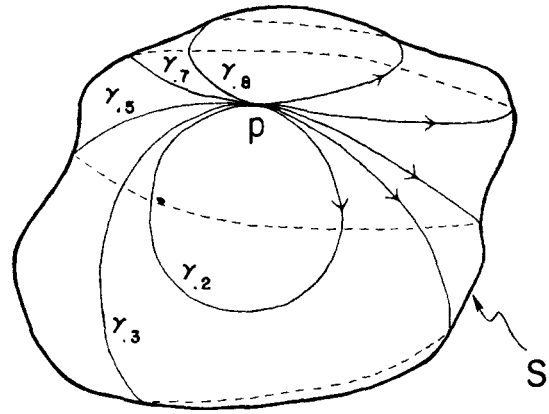


FIG. 1. A one-parameter family of curves covering the 2-sphere S. Each curve begins and ends at p, and each point of S (except p) lies on exactly one curve.

normalized by the conditions

$$\omega^a \nabla_a w = \sigma^a \nabla_a s = 1.$$

Our construction proceeds as follows. Choose an arbitrary unit timelike vector field t^a on M. At p, choose a triad[17] x_α^a of spacelike vectors which, together with t^a, form an orthonormal tetrad at p. For each value of s, we transport the triad x_α^a along the curve γ_s according to the equation

$$\omega^b \nabla_b x_\alpha^a = -t^a (x_\alpha^c \omega^b \nabla_b t_c). \tag{1}$$

Under the transport (1) the x_α^a remain orthogonal to t^a and orthogonal to each other. On returning to p, we have a new tetrad whose timelike vector coincides with t^a, but whose spacelike vectors will in general be different from our original x_α^a Let $\mathcal{R}_\alpha^\beta(s)$ denote the corresponding rotation matrix:

$$x_\alpha^a|_{w=1;s} = \mathcal{R}_\alpha^\beta(s)(x_\beta^a|_{w=0;s=0}),$$
$$\mathcal{R}_\alpha^\gamma(s)\mathcal{R}_\gamma^\beta(s) = \delta_\alpha^\beta. \tag{2}$$

Thus, for each value of s, we obtain a rotation at p, and so we define a curve $\mathcal{R}_\alpha^\beta(s)$ in the rotation group. For $s = 0$ and $s = 1$, the rotation is just the identity, and so $\mathcal{R}_\alpha^\beta(s)$ represents a closed curve, beginning and ending at the identity element of the rotation group. The tangent to this curve is obtained by taking the derivative of the first equation (2) with respect to s:

$$\mathcal{R}_{\alpha\gamma} \frac{d}{ds} \mathcal{R}_\beta^\gamma = (x_\alpha^a \sigma^b \nabla_b x_{a\beta})_{w=1;s} = \int_{\gamma_s} dw P_{\alpha\beta}, \tag{3}$$

[17] Greek letters are triad indices (range 1, 2, 3) which label individual spacelike vectors, while Latin letters are tensor indices. The Greek indices are raised and lowered using the unit 3×3 matrix.

where

$$P_{\alpha\beta} = P_{[\alpha\beta]} = \omega^c \nabla_c (x^a_\alpha \sigma^b \nabla_b x_{a\beta}). \tag{4}$$

Expanding (4) and using the transport equation (1) and the fact (which follows from our construction) that the Lie derivative of σ^a with respect to ω^a vanishes, we have

$$P_{\alpha\beta} = 2x^a_\alpha x^b_\beta \sigma^{[c}\omega^{d]}[(\nabla_c t_b)(\nabla_d t_a) + R_{abcd}]. \tag{5}$$

The closed curves in the rotation group are divided into two classes: those (such as a small loop, or a rotation through angle 4π) which may be contracted to a point, and those (such as a rotation through 2π) which cannot. If our curve $\mathcal{R}^\beta_\alpha(s)$ is of the latter type, then there is an essential 2π twist in our tetrad system on S; in this case, it will not be possible to find a tetrad in a neighborhood of S, and so M will not have spinor structure. We may characterize the curve $\mathcal{R}^\beta_\alpha(s)$ in terms of a length using the standard invariant metric on the rotation group:

$$L = 2^{-\frac{1}{2}} \int_0^1 ds \left[\left(\frac{d}{ds} \mathcal{R}_{\alpha\beta} \right) \left(\frac{d}{ds} \mathcal{R}^{\alpha\beta} \right) \right]^{\frac{1}{2}}. \tag{6}$$

Whenever the length L is less than 2π, the curve $\mathcal{R}^\beta_\alpha(s)$ may always be contracted to a point. (The length is exactly 2π for a 360° rotation about a single axis.) We conclude that there will necessarily be a tetrad system in a neighborhood of S provided $L < 2\pi$.

To obtain an upper bound for L which is independent of the coordinate grid (s, w), we first substitute Eq. (3) into (6):

$$\begin{aligned} L &= 2^{-\frac{1}{2}} \int_0^1 ds \left[\left(\mathcal{R}_{\alpha\mu} \frac{d}{ds} \mathcal{R}^{\beta\mu} \right) \left(\mathcal{R}^{\alpha\nu} \frac{d}{ds} \mathcal{R}_{\beta\nu} \right) \right]^{\frac{1}{2}} \\ &= 2^{-\frac{1}{2}} \int_0^1 ds \left[\left(\int_0^1 dw P_{\alpha\beta} \right) \left(\int_0^1 dw P^{\alpha\beta} \right) \right]^{\frac{1}{2}} \\ &\leq 2^{-\frac{1}{2}} \int_0^1 ds \int_0^1 dw [P_{\alpha\beta} P^{\alpha\beta}]^{\frac{1}{2}}. \end{aligned} \tag{7}$$

But

$$\begin{aligned} P_{\alpha\beta} P^{\alpha\beta} &= 4(g^{ap} - t^a t^p)(g^{bq} - t^b t^q) \sigma^{[c}\omega^{d]} \sigma^{[r}\omega^{s]} \\ &\quad \times [(\nabla_c t_b)(\nabla_d t_a) + R_{abcd}][(\nabla_r t_q)(\nabla_s t_p) + R_{pqrs}] \\ &\leq 8\sigma^{[c}\omega^{d]}\sigma^{[r}\omega^{s]}[(\nabla_c t_b)(\nabla_d t_a)(\nabla_r t^b)(\nabla_s t^a) \\ &\quad + R_{abcd} R^{ab}{}_{rs} - 2t^b t^p R_{abcd} R^a{}_{prs}]. \end{aligned} \tag{8}$$

Finally, substituting (8) into (7) and introducing the surface element of S, $dS^{ab} = \sigma^{[a}\omega^{b]}\, ds\, dw$, we obtain

$$\begin{aligned} L \leq &\int_S 2[(R_{abcd} R^{ab}{}_{rs} - 2t^b t^p R_{abcd} R^a{}_{prs})\, dS^{cd}\, dS^{rs}]^{\frac{1}{2}} \\ &+ \int_S 2[(\nabla_c t_b)(\nabla_r t^b)(\nabla_d t_a)(\nabla_s t^a)\, dS^{cd}\, dS^{rs}]^{\frac{1}{2}}. \end{aligned} \tag{9}$$

Equation (9) still depends on the arbitrary unit timelike vector field t^a.[18] We may eliminate this dependence, at least formally, by defining $\Psi(S)$ to be the minimum value of the right side of (9) over all possible choices of t^a. (Note that the right side is always nonnegative.) Unfortunately, the quantity Ψ depends on S in a nonlocal way because of the second integral on the right in (9): it appears, in fact, that Ψ cannot be expressed as a single integral over S. Thus, our relationship between spinor structure and curvature is the following: A sufficient (but not necessary) condition that M have spinor structure is that every S may be so deformed that $\Psi(S) < 2\pi$.

It follows immediately that every flat space-time has spinor structure. (By "flat" we mean only that the Riemann tensor vanishes: M could still have a quite complicated topology.) We may, without loss of generality, take M to be simply connected. Choosing t^a to be (covariantly) constant over M, we see from (9) that $\Psi(S) = 0$ for every S, and, therefore, that M has spinor structure. In particular, none of the 4-manifolds M_{2n+1} of Sec. I can be given a flat metric.

[18] If the metric of space-time were positive-definite, then we could parallel transport the entire tetrad around each curve γ_s, and so obtain a formula which does not involve an arbitrary vector field. This procedure does not work, however, in the indefinite case because there is no invariant positive-definite length defined for curves in the Lorentz group.