JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Mandelstam Representation in Potential Scattering*

JULIAN CHELA-FLORES
Physics Department, Queen Mary College, Mile End Road, London E1, England

(Received 26 November 1969)

A simpler and more general proof of the Mandelstam representation based on Hartog's theorem is presented within the framework of potential scattering; the potential is chosen to be a superposition of the Yukawa type with no further restrictions.

I. INTRODUCTION

The proof of the Mandelstam representation in potential scattering has been considered on various occasions in the past decade.[1–5] Recently, a new simplified proof was presented by Cheung.[6] In the present paper, we focus our attention on two questions concerning the classical proof of Blankenbecler, Goldberger, Khuri, and Treiman.[4] First, we must emphasize that finding the double analyticity properties of the scattering amplitude for a superposition of Yukawa potentials is a very well-defined mathematical problem. Bessis, in a paper published a few years ago,[5] closed a loophole in the classical proof; for potentials $V(z)$, holomorphic in $\mathrm{Re}\, z > 0$ and bounded by

$$|V(z)| < K\,|z|^{-\rho}, \quad \rho < 2 \quad \text{for} \quad |z| \leq 1, \tag{1}$$

$$|V(z)| < K\,|z|^{-\nu} \exp(-\mu_0 \,\mathrm{Re}\, z) \quad \text{for} \quad |z| \geq 1, \quad \nu > \tfrac{7}{4}, \tag{2}$$

the double spectral function $\rho_{st}(s, t)$ was shown to be a continuous function of s and t in $\{s > 0, t > 0\}$ and Bessis also showed that

$$|\rho_{st}(s, t)| \leq C(V)[1/(st)^{\frac{1}{4}} + (2 + t)^{L_0}/s^{\frac{1}{2}}], \tag{3}$$

where $C(V)$ is some constant depending only on the potential and L_0 is the number of subtractions in t. Therefore, the above inequalities imply that, for a superposition of Yukawa potentials with the further restrictions (1) and (2), the double spectral function is bounded as shown in Eq. (3). Then, and only then, it follows that the double spectral integral, in which, originally, two single integrations are to be performed, is, indeed, equivalent to a double integral, in view of the Courant theorem,[7] as required by the Mandelstam representation. The double dispersion relation thus obtained (explicitly written out in Sec. II) implies that the scattering amplitude is an analytic function of the two complex variables s and t in the topological product of the s and t planes appropriately cut.

In Sec. III, we appeal to a well-known theorem of the theory of several complex variables and present a new proof of the Mandelstam representation in potential scattering by replacing the Bessis work on the upper bound for the double spectral function. This, we hasten to emphasize, is not just mathematically more appealing, but has very definite physical content: namely, we do not require any constraint on the potential in question besides the requirement of it being of the Yukawa type. Therefore, our proof is more simple and general, since Eqs. (1) and (2) are unnecessary in our analysis.

II. THE CLASSICAL PROOF

In the present section, we review the combined work of Regge,[2] Blankenbecler *et al.*,[4] and Bessis,[5] which provide the backbone of the orthodox approach to the problem. This serves a dual purpose: First, this presentation emphasizes the necessity and relevance of the work of Ref. 5 and, second, at the same time, we introduce the notation used in the next section. We begin by restricting our attention to the Khuri dispersion relation[8]

$$f(s, t) = f_B(t) + \sum_{i=1}^{N} \frac{\Gamma_i(s_i, t)}{s + s_i} + \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\, f(s', t)}{s' - s}\, ds', \tag{4}$$

where the s_i are the negative energies of the bound states and the Γ_i are polynomials in t of degree l_i, where l_i represents the angular momentum of the ith bound state. The aspect of Ref. 4 which concerns us here is the possibility of extending the domain of analyticity of the scattering amplitude onto the whole complex plane in the t variable, except for a cut along the real positive axis beginning at $4m^2$, where m denotes the inverse of the range of the force.

On the other hand, Regge has shown,[2] by representing the scattering amplitude by means of a Sommerfeld–Watson transform, that the absorptive part of the amplitude A_s has a domain of analyticity which coincides with that of the full amplitude. Thus, we may write a "superconvergent" dispersion relation (namely, one with no subtractions)

$$A_s(s', t) = \mathrm{Im}\, f(s', t) = \frac{1}{\pi}\int_{4m^2}^\infty \frac{\rho_{st}(s', t')}{t' - t}\, dt'. \tag{5}$$

For simplicity, we assume no bound states in Eq. (4); therefore, we write

$$f(s,t) = f_B(t) + \frac{1}{\pi}\int_0^\infty \frac{A_s(s',t)}{s'-s}\,ds'. \qquad (6)$$

We now proceed to insert Eq. (5) into (6) in order to obtain

$$f(s,t) = f_B(t) + \frac{1}{\pi^2}\lim_{\epsilon\to 0}\int_0^\infty ds'\int_{4m^2}^\infty \frac{\rho_{st}(s',t')\,dt'}{(s'-s-i\epsilon)(t'+t)}. \qquad (7)$$

The crux of the matter is that the Eq. (7) is not the Mandelstam representation, since we still require a double integral on the right-hand side. The two single integrals will combine into a double integral if we succeed in showing that ρ_{st} is a continuous function of s and t and bounded in some domain.

This brings us back to our remarks in the introductory section; for, in 1965, Bessis showed both continuity and boundedness of the double spectral function for restricted Yukawa potentials [cf. Eqs. (1) and (2)]. As $\rho_{st}(s,t)$ is continuous and bounded, we may rewrite Eq. (7) as

$$f(s,t) = f_B(t) + \frac{1}{\pi^2}\lim_{\epsilon\to 0}\int_0^\infty\int_{4m^2}^\infty \frac{\rho_{st}(s',t')\,ds'\,dt'}{(s'-s-i\epsilon)(t'+t)}; \qquad (8)$$

i.e., the two single integrals give a double integral, as required by the Mandelstam representation. Equation (8) is the simplest form of the double dispersion relation in potential scattering for Bessis–Yukawa potentials.

We now remark that the superconvergency assumption of Eq. (5) is too naive, for, in writing such a relation, we have indeed embodied the correct analyticity property of the absorptive part. But there is more to it; that is, we must include the correct asymptotic behavior, or, in other words, we must study the correct number of subtractions. It follows from the work of Regge[2] that the number of subtractions is given by the large-t behavior of the scattering amplitude; if the leading Regge pole has a real part denoted by M, then the smallest integer greater than M is denoted by L_0. The number L_0 leads to the number of subtractions required in Eq. (5). However, neither this question nor the inclusion of possible bound states into the Khuri dispersion relation leads to any essential new steps in the line of argument of the classical proof. It is sufficient, therefore, to discuss the simplified Eq. (8).

III. AN ALTERNATIVE PROOF

Because we have obtained the analyticity properties of the scattering amplitude, with respect to s, by means of the Khuri dispersion relation and, with respect to t, by the work of Regge using the Sommerfeld–Watson transform, we can deduce the Mandelstam representation without necessarily analyzing the very complicated properties of the double spectral function and *a fortiori* restricting ourselves to Yukawa potentials only.

In Ref. 9 Bochner and Martin prove the following theorem (Hartog's):

Theorem: The space of two complex variables $s = x_1 + iy_1$ and $t = x_2 + iy_2$ is the ordinary Euclidean space E_4 of the four real variables x_1, x_2, y_1, and y_2. Let $f(s,t)$ be a function defined in a domain $\mathfrak{D}$ in the space E_4 and let f have the property that at every point (s_1,t_1) of each of the functions $f(s_1,t)$ and $f(s,t_1)$ it is analytic in the single variable in the neighborhood of s_1 and t_1, respectively. Then f is analytic in s and t in $\mathfrak{D}$.

With Hartog's theorem, we can produce a more direct proof of double analyticity than the one presented in Sec. II. Essentially, all we need is the analyticity embodied in the Khuri dispersion relation and that derivable from the Sommerfeld–Watson transform. Thus, we have a domain in E_4 of the two complex variables s and t in which f is analytic in both variables; this is, in fact, the domain required by the Mandelstam representation.

IV. CONCLUSION

To conclude this article, the spotlight should bring out some features of the present work. In the realm of potential scattering, once we have chosen a given potential, we are able to formulate rigorous problems (which in some cases will only accept approximate solutions) based essentially on the Schrödinger equation. The present case is yet another example which calls for rigor and, while the orthodox treatment of the subject does provide a completely rigorous proof, the question of the potential chosen has needed further elucidation. Another and perhaps more fruitful way of viewing potential scattering is as a model for the strong interactions; indeed, this is the reason why we bother to look at the question of analytic properties of scattering amplitudes from the point of view of this paper. It is generally agreed[10] that potential scattering may be considered as a nonrelativistic (i.e., Schrödinger) limit of field theory only if superpositions of Yukawa potentials are chosen. On the

other hand, the orthodox proof of our problem is based on the Bessis–Yukawa potentials. It is the main feature of our (Hartog's) proof of the Mandelstam representation that only Yukawa potentials are required.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. R. Chela and Dr. R. K. Colegrave for their encouragement and Professor A. Ashmore and Professor J. G. Valatin for the hospitality I have enjoyed at Queen Mary College (University of London).

* This work was supported by a postdoctoral grant from the Instituto Venezolano de Investigaciones Cientificas, Apartado 1827, Caracas, Venezuela.

[1] J. Bowcock and A. Martin, Nuovo Cimento **14**, 516 (1959).
[2] T. Regge, Nuovo Cimento **14**, 951 (1959).
[3] A. Klein, J. Math. Phys. **1**, 41 (1960).
[4] R. Blankenbecler, M. L. Goldberger, N. N. Khuri, and S. B. Treiman, Ann. Phys. (N.Y.) **10**, 62 (1960).
[5] D. Bessis, J. Math. Phys. **6**, 637 (1965).
[6] F. F. K. Cheung, Phys. Letters **30B**, 257 (1969).
[7] R. Courant, *Differential and Integral Calculus* (Blackie, London, 1961), Vol. II, p. 311.
[8] N. N. Khuri, Phys. Rev. **107**, 1148 (1957).
[9] S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, N.J., 1946), p. 140.
[10] V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965), p. 48.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

De Sitter Symmetric Field Theory. II. Tensorial and Spinorial Realizations

M. M. BAKRI*
International Centre for Theoretical Physics, Trieste, Italy
AND
A. F. GHALEB AND M. I. HESSEIN
Applied Mathematics Department, Faculty of Science, Cairo University, Giza, United Arab Republic

(Received 2 June 1969)

Plane-wave solutions with definite helicity are obtained for the field equations $(S_{\mu\nu}p_\nu + m\gamma_\mu)\psi = p_\mu\psi$. These equations are then rewritten in spinorial form. The relations between spinors, 4- and 5-dimensional tensors and spin-tensors, and the components of ψ are derived.

1. INTRODUCTION

In a previous paper[1] a de Sitter symmetric field theory was developed based on the equations[2]

$$S_{AB}p_B\psi = p_A\psi, \tag{1.1}$$

where the S_{AB} are the infinitesimal generators of the group $SO(4, 1)$ and $p_A = (p_\mu, p_5 = m)$ is the energy–momentum–mass 5-vector. It is interesting to note that these equations are completely equivalent[3] to Salam's *et al.* generalization[4] of the Bargmann–Wigner equations.[5] Ten Kate,[6] using the de Vos–Hilgevoord implicit form[2] of Eqs. (1.1), wrote these equations in 5-dimensional tensorial and spin-tensorial forms for all finite-dimensional representations $R_5(\lambda_1, \lambda_2)$ of the de Sitter group $O(4, 1)$. He also wrote these equations in spinorial form for the particular representation $R_5(\lambda_1, \lambda_1)$. However, he did not try to connect these tensors and spinors to the components of ψ. These connections are particularly useful when one studies the quantization of the field and also when one tries to introduce a self-consistent interaction into Eqs. (1.1), which is by no means trivial.

The purpose of the present paper is to write Eqs. (1.1) in spinorial form for all representations $R_5(\lambda_1, \lambda_2)$ and to establish the connections between the components of ψ, 5-tensors, spin-tensors, and spinors. The present paper is a continuation of the previous paper,[1] hereafter called Paper I. We use the same notation as in Paper I. Formulas of that part will be denoted, for example, by (I.2.5) for Eq. (2.5) of Paper I.

In Sec. 2 we express the Poincaré components $\psi(s_1, s_2)$ of ψ in terms of Wigner's components $\psi(\lambda_2, 0)$ and $\psi(0, \lambda_2)$, which transform according to the unitary representations of the Poincaré group.[7] We evaluate the matrix elements of $\psi(s_1, s_2)$ for plane-wave solutions with definite helicity along the z axis.

In Sec. 3 we find the relation between $\psi(s_1, s_2)$ and the corresponding spinor and derive the spinorial field equations. In Sec. 4 we find the relations between the spinors and the corresponding tensors or spin-tensors of $SO(3, 1)$.

In Secs. 5 and 6 we express the Poincaré components in terms of the 5-tensor or the spin-tensor carrying the representation $R_5(\lambda_1, \lambda_2)$. The inverse relation is complicated. We derive it in the particular representations $R_5(l, 0)$, $R_5(l + \frac{1}{2}, \frac{1}{2})$, and $R_5(\lambda_1, \lambda_1)$.

other hand, the orthodox proof of our problem is based on the Bessis–Yukawa potentials. It is the main feature of our (Hartog's) proof of the Mandelstam representation that only Yukawa potentials are required.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. R. Chela and Dr. R. K. Colegrave for their encouragement and Professor A. Ashmore and Professor J. G. Valatin for the hospitality I have enjoyed at Queen Mary College (University of London).

* This work was supported by a postdoctoral grant from the Instituto Venezolano de Investigaciones Cientificas, Apartado 1827, Caracas, Venezuela.

[1] J. Bowcock and A. Martin, Nuovo Cimento **14**, 516 (1959).
[2] T. Regge, Nuovo Cimento **14**, 951 (1959).
[3] A. Klein, J. Math. Phys. **1**, 41 (1960).
[4] R. Blankenbecler, M. L. Goldberger, N. N. Khuri, and S. B. Treiman, Ann. Phys. (N.Y.) **10**, 62 (1960).
[5] D. Bessis, J. Math. Phys. **6**, 637 (1965).
[6] F. F. K. Cheung, Phys. Letters **30B**, 257 (1969).
[7] R. Courant, *Differential and Integral Calculus* (Blackie, London, 1961), Vol. II, p. 311.
[8] N. N. Khuri, Phys. Rev. **107**, 1148 (1957).
[9] S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, N.J., 1946), p. 140.
[10] V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965), p. 48.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

De Sitter Symmetric Field Theory. II. Tensorial and Spinorial Realizations

M. M. BAKRI*
International Centre for Theoretical Physics, Trieste, Italy
AND
A. F. GHALEB AND M. I. HESSEIN
Applied Mathematics Department, Faculty of Science, Cairo University, Giza, United Arab Republic

(Received 2 June 1969)

Plane-wave solutions with definite helicity are obtained for the field equations $(S_{\mu\nu}p_\nu + m\gamma_\mu)\psi = p_\mu\psi$. These equations are then rewritten in spinorial form. The relations between spinors, 4- and 5-dimensional tensors and spin-tensors, and the components of ψ are derived.

1. INTRODUCTION

In a previous paper[1] a de Sitter symmetric field theory was developed based on the equations[2]

$$S_{AB}p_B\psi = p_A\psi, \tag{1.1}$$

where the S_{AB} are the infinitesimal generators of the group $SO(4, 1)$ and $p_A = (p_\mu, p_5 = m)$ is the energy–momentum–mass 5-vector. It is interesting to note that these equations are completely equivalent[3] to Salam's *et al.* generalization[4] of the Bargmann–Wigner equations.[5] Ten Kate,[6] using the de Vos–Hilgevoord implicit form[2] of Eqs. (1.1), wrote these equations in 5-dimensional tensorial and spin-tensorial forms for all finite-dimensional representations $R_5(\lambda_1, \lambda_2)$ of the de Sitter group $O(4, 1)$. He also wrote these equations in spinorial form for the particular representation $R_5(\lambda_1, \lambda_1)$. However, he did not try to connect these tensors and spinors to the components of ψ. These connections are particularly useful when one studies the quantization of the field and also when one tries to introduce a self-consistent interaction into Eqs. (1.1), which is by no means trivial.

The purpose of the present paper is to write Eqs. (1.1) in spinorial form for all representations $R_5(\lambda_1, \lambda_2)$ and to establish the connections between the components of ψ, 5-tensors, spin-tensors, and spinors. The present paper is a continuation of the previous paper,[1] hereafter called Paper I. We use the same notation as in Paper I. Formulas of that part will be denoted, for example, by (I.2.5) for Eq. (2.5) of Paper I.

In Sec. 2 we express the Poincaré components $\psi(s_1, s_2)$ of ψ in terms of Wigner's components $\psi(\lambda_2, 0)$ and $\psi(0, \lambda_2)$, which transform according to the unitary representations of the Poincaré group.[7] We evaluate the matrix elements of $\psi(s_1, s_2)$ for plane-wave solutions with definite helicity along the z axis.

In Sec. 3 we find the relation between $\psi(s_1, s_2)$ and the corresponding spinor and derive the spinorial field equations. In Sec. 4 we find the relations between the spinors and the corresponding tensors or spin-tensors of $SO(3, 1)$.

In Secs. 5 and 6 we express the Poincaré components in terms of the 5-tensor or the spin-tensor carrying the representation $R_5(\lambda_1, \lambda_2)$. The inverse relation is complicated. We derive it in the particular representations $R_5(l, 0)$, $R_5(l + \frac{1}{2}, \frac{1}{2})$, and $R_5(\lambda_1, \lambda_1)$.

In the Appendix we give several useful relations between the fundamental spinor-vectors and spinor-tensors.

2. PLANE-WAVE SOLUTIONS

A. Relations Between the Poincaré Components

For convenience we denote the Poincaré component by

$$|s_1, s_2\rangle = \psi(s_1, s_2).$$

Taking the matrix elements (i.e., the Poincaré block $\langle s_1 s_2 | \psi \rangle$) of the field equations in the forms (I.3.62) and (I.3.63), using Eqs. (I.3.23), (I.3.26), and (I.3.27), choosing $d_1 = -i$, $d_2 = 1$, and then contracting the resulting equations with $v_a(s_1)$, $u_a(s_1 + \frac{1}{2})$, $v_{\dot b}(s_2)$ and $u_{\dot b}(s_2 + \frac{1}{2})$, using Eqs. (I.3.28), one easily obtains the following equations:

$$\begin{aligned}(\lambda_1 + 1)(1 + x)(1 + y)v_a(s_1)p^{a\dot b}\,|s_1, s_2\rangle \\ = m(2s_1 + 1)[(1 + x)\theta_{--}u^{\dot b}(s_2)\,|s_1 - \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle \\ - (1 + y)\theta_{-+}v^{\dot b}(s_2 + \tfrac{1}{2})\,|s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2}\rangle],\end{aligned} \quad (2.1a)$$

$$\begin{aligned}(\lambda_1 + 1)(1 - x)(1 - y)u_a(s_1 + \tfrac{1}{2})p^{a\dot b}\,|s_1, s_2\rangle \\ = -m(2s_1 + 1)[(1 - y)\theta_{+-}u^{\dot b}(s_2)\,|s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle \\ + (1 - x)\theta_{++}v^{\dot b}(s_2 + \tfrac{1}{2})\,|s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle],\end{aligned} \quad (2.1b)$$

$$\begin{aligned}(\lambda_1 + 1)(1 - x)(1 + y)v^{\dot b}(s_2)p^{a\dot b}\,|s_1, s_2\rangle \\ = m(2s_2 + 1)[(1 - x)\theta_{--}u^{a}(s_1)\,|s_1 - \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle \\ + (1 - y)\theta_{+-}v^{a}(s_1 + \tfrac{1}{2})\,|s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle],\end{aligned} \quad (2.1c)$$

$$\begin{aligned}(\lambda_1 + 1)(1 + x)(1 - y)u^{\dot b}(s_2 + \tfrac{1}{2})p^{a\dot b}\,|s_1, s_2\rangle \\ = m(2s_2 + 1)[(1 - y)\theta_{-+}u^{a}(s_1)\,|s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2}\rangle \\ - (1 + x)\theta_{++}v^{a}(s_1 + \tfrac{1}{2})\,|s_1 + \tfrac{1}{2}, s_2 + \tfrac{1}{2}\rangle].\end{aligned} \quad (2.1d)$$

Here $x = \mu/(\lambda_1 + 1)$, $y = (\lambda + 1)/(\lambda_1 + 1)$, $\mu = s_1 - s_2$, $\lambda = s_1 + s_2$, and $\theta_{--} = (c_{--})^{\frac{1}{2}}$, etc., where the c's are given by Eqs. (I.3.40) and (I.3.41).

By suitable contractions of Eqs. (2.1) we obtain the following equations:

$$\begin{aligned}v_a(s_1)v_{\dot b}(s_2)p^{a\dot b}\,|s_1, s_2\rangle \\ = [m(2s_1 + 1)(2s_2 + 1)\theta_{--}/(\lambda_1 + \lambda + 2)] \\ \times |s_1 - \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle,\end{aligned} \quad (2.2a)$$

$$\begin{aligned}v_a(s_1)u_{\dot b}(s_2 + \tfrac{1}{2})p^{a\dot b}\,|s_1, s_2\rangle \\ = [m(2s_1 + 1)(2s_2 + 1)\theta_{-+}/(\lambda_1 + \mu + 1)] \\ \times |s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2}\rangle,\end{aligned} \quad (2.2b)$$

$$\begin{aligned}u_a(s_1 + \tfrac{1}{2})v_{\dot b}(s_2)p^{a\dot b}\,|s_1, s_2\rangle \\ = -[m(2s_1 + 1)(2s_2 + 1)\theta_{+-}/(\lambda_1 - \mu + 1)] \\ \times |s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle,\end{aligned} \quad (2.2c)$$

$$\begin{aligned}u_a(s_1 + \tfrac{1}{2})u_{\dot b}(s_2 + \tfrac{1}{2})p^{a\dot b}\,|s_1, s_2\rangle \\ = [m(2s_1 + 1)(2s_2 + 1)\theta_{++}/(\lambda_1 - \lambda)] \\ \times |s_1 + \tfrac{1}{2}, s_2 + \tfrac{1}{2}\rangle.\end{aligned} \quad (2.2d)$$

For convenience, we introduce

$$\begin{aligned}P_{a_1 a_2 \cdots a_{2s}}(s) &= u_{a_1}(s)u_{a_2}(s - \tfrac{1}{2}) \cdots u_{a_{2s}}(\tfrac{1}{2}), \\ R_{a_1 a_2 \cdots a_{2s}}(s) &= v_{a_1}(\tfrac{1}{2})v_{a_2}(1) \cdots v_{a_{2s}}(s),\end{aligned} \quad (2.3)$$

and

$$\begin{aligned}T_{a_1 \cdots a_{2s_2}}(s_1, \mu, \lambda_2) \\ = u_{a_1}(s_1)u_{a_2}(s_1 - \tfrac{1}{2}) \cdots u_{a_{\lambda - \lambda_2}}(\tfrac{1}{2}(\lambda_2 + \mu + 1)) \\ \times v_{a_{\lambda - \lambda_2 + 1}}(\tfrac{1}{2}(\lambda_2 + \mu + 1)) \\ \times v_{a_{\lambda - \lambda_2 + 2}}(\tfrac{1}{2}(\lambda_2 + \mu + 2)) \cdots v_{a_{2s_2}}(\lambda_2).\end{aligned} \quad (2.4)$$

$P_{a_1 \cdots a_{2s}}(s)$ and $R_{a_1 \cdots a_{2s}}(s)$ are $(2s + 1)$-column and -row matrices, respectively, which are symmetric in the spinor indices. $T_{a_1 \cdots a_{2s_2}}(s_1, \mu, \lambda_2)$ is a rectangular matrix of $(2s_1 + 1)(2s_2 + 1)$ rows and $(2\lambda_2 + 1)$ columns; it is symmetric in the indices belonging to the u's and in those belonging to the v's separately.

By successive application of Eqs. (2.2b) and (2.2d), we get

$$|s_1, s_2\rangle = m^{-2s_2}\frac{A(\lambda, \mu)}{(\lambda_2 - \mu)!}\prod_{k=1}^{2s_2} p^{a_k \dot b_k} T_{a_1 \cdots a_{2s_2}}(s_1, \mu, \lambda_2) \otimes P_{\dot b_1 \cdots \dot b_{2s_2}}(s_2)\,|\lambda_2, 0\rangle, \quad (2.5)$$

where

$$\begin{aligned}A(\lambda, \mu) = (\lambda_1 - \lambda_2)!\,(\lambda_1 + \lambda_2 + 1)! \\ \times \left(\frac{(\lambda_1 - \lambda_2 + 1)(\lambda_1 + \lambda_2 + 2)(\lambda + \mu + 1)(\lambda - \mu + 1) \cdot (\lambda_2 + \mu)!\,(\lambda_2 - \mu)!}{(\lambda_1 - \mu + 1)!\,(\lambda_1 + \mu + 1)!\,(\lambda - \lambda_2)!\,(\lambda_1 - \lambda)!\,(\lambda + \lambda_2 + 1)!\,(\lambda + \lambda_1 + 2)!}\right)^{\frac{1}{2}}.\end{aligned} \quad (2.6)$$

In particular,

$$\begin{aligned}|0, \lambda_2\rangle = \frac{1}{m^{2\lambda_2} \cdot (2\lambda_2)!}\prod_{k=1}^{2\lambda_2} p^{a_k \dot b_k} \\ \times R_{a_1 \cdots a_{2\lambda_2}}(\lambda_2) \otimes P_{\dot b_1 \cdots \dot b_{2\lambda_2}}(\lambda_2)\,|\lambda_2, 0\rangle.\end{aligned} \quad (2.7)$$

Further, by successive application of Eqs. (2.2a) and (2.2c), we get

$$\begin{aligned}|s_1, s_2\rangle = (-1)^{\lambda_2 + \mu} m^{-2s_1}\left(\frac{A(\lambda, \mu)}{(\lambda_2 + \mu)!}\right)\prod_{k=1}^{2s_1} p^{a_k \dot b_k} \\ \times P_{a_1 \cdots a_{2s_1}}(s_1) \otimes T_{\dot b_1 \cdots \dot b_{2s_1}}(s_2, -\mu, \lambda_2)\,|0, \lambda_2\rangle.\end{aligned} \quad (2.8)$$

In particular,

$$|\lambda_2, 0\rangle = \frac{1}{(-m)^{2\lambda_2}\cdot(2\lambda_2)!}\prod_{k=1}^{2\lambda_2} p^{a_k\dot{b}_k} \times P_{a_1\cdots a_{2\lambda_2}}(\lambda_2)\otimes R_{\dot{b}_1\cdots\dot{b}_{2\lambda_2}}(\lambda_2)\,|0, \lambda_2\rangle. \quad (2.9)$$

We have thus related all Poincaré components to Wigner's components $|\lambda_2, 0\rangle$ and $|0, \lambda_2\rangle$, which belong to the unitary representations of the Poincaré group.[7]

B. Solutions with Definite Helicity

Consider the representation

$$\langle m|\, u^1(s)\,|m'+\tfrac{1}{2}\rangle = \langle m'+\tfrac{1}{2}|\, v_1(s)\,|m\rangle = (s-m)^{\frac{1}{2}}\delta_{mm'},$$
$$\langle m|\, u^2(s)\,|m'-\tfrac{1}{2}\rangle = \langle m'-\tfrac{1}{2}|\, v_2(s)\,|m\rangle = -(s+m)^{\frac{1}{2}}\delta_{mm'}, \quad (2.10)$$

where $m = s, s-1, \cdots, -s$. The other matrix elements are obtained from the unitarity conditions[8]

$$u^a(s)^+ = v_a(s), \qquad u^{\dot{a}}(s)^+ = v_{\dot{a}}(s),$$
$$u^{\dot{a}}(s) = u_a(s), \qquad v^{\dot{a}}(s) = v_a(s), \quad (2.11)$$
$$u_{\dot{a}}(s) = -u^a(s), \qquad v_{\dot{a}}(s) = -v^a(s).$$

In this representation

$$\langle m|\,\Sigma_3^{(1)}(s)\,|m'\rangle = m\delta_{mm'},$$
$$\langle m|\,\Sigma_1^{(1)} \mp i\Sigma_2^{(1)}\,|m'\rangle = [(s\mp m)(s\pm m+1)]^{\frac{1}{2}}\delta_{m',m\pm1}, \quad (2.12)$$

where $\Sigma_1^{(1)}$, $\Sigma_2^{(1)}$, and $\Sigma_3^{(1)}$ are given by (I.3.7). Denoting

$$\bar{P}_m^t = \langle m \mid P^{11\cdots1\,22\cdots2}(s)\rangle,$$
$$\underline{P}_m^t = \langle m \mid P_{11\cdots1\,22\cdots2}(s)\rangle,$$
$$\dot{\bar{P}}_m^t = \langle m \mid P^{\dot1\dot1\cdots\dot1\,\dot2\dot2\cdots\dot2}(s)\rangle,$$
$$\dot{\underline{P}}_m^t = \langle m \mid P_{\dot1\dot1\cdots\dot1\,\dot2\dot2\cdots\dot2}(s)\rangle,$$

where the indices 1 and 2 occur t and $2s-t$ times, respectively, and there is corresponding notation for R, we find that

$$\underline{R}_m^{s-m'} = \dot{\bar{R}}_m^{s-m'} = \bar{P}_m^{s-m'} = (-1)^{2s}\dot{\underline{P}}_m^{s-m'} = (-1)^{s+m}\delta_{mm'}[(s-m)!\,(s+m)!]^{\frac{1}{2}}, \quad (2.13a)$$

$$\dot{\underline{R}}_m^{s+m'} = (-1)^{2s}\bar{R}_m^{s+m'} = \underline{P}_m^{s+m'} = \dot{\bar{P}}_m^{s+m'} = (-1)^{2s}\delta_{mm'}[(s-m)!\,(s+m)!]^{\frac{1}{2}}. \quad (2.13b)$$

Further, denoting

$$T^{rt}_{mM}(s_1, \mu, \lambda_2) = \langle m|\, T_{11\cdots1\,22\cdots2\,11\cdots1\,22\cdots2}(s_1, \mu, \lambda_2)\,|M\rangle, \quad (2.14)$$

where $0 \le r \le \lambda - \lambda_2$ and $0 \le t \le \lambda_2 - \mu$ and the successive indices 1, 2, 1, 2 occur r, $\lambda - \lambda_2 - r$, t, and $\lambda_2 - \mu - t$ times, respectively, we find that

$$T^{rt}_{mM}(s_1, \mu, \lambda_2) = (-1)^{2s_2-t}\delta_{r+t,\,m+s_2-M} \times \{[(s_1-m)!\,(s_1+m)!\,(\lambda_2-M)!\,(\lambda_2+M)!]^{\frac{1}{2}}/ (s_1+m-r)!\,(\lambda_2-M-t)!\}. \quad (2.15)$$

The matrix elements $|s_1, m_1; s_2, m_2\rangle$ of the Poincaré component can be evaluated in general with the help of (2.8), (2.13), and (2.15). These matrix elements become particularly simple if we consider the propagation along the z axis, such that the components of momentum are $p_1 = p_2 = 0$, $p_3 = p$, and

$$p^{1\dot1} = p_0 + p, \quad p^{2\dot2} = p_0 - p, \quad p^{1\dot2} = p^{2\dot1} = 0, \quad (2.16)$$

where $p_0 = \pm(p^2+m^2)^{\frac{1}{2}}$. In this case,

$$\langle m_1, m_2|\prod_{k=1}^{2s_2} p^{a_k\dot{b}_k} T_{a_1\cdots a_{2s_2}}(s_1, \mu, \lambda_2)\otimes\bar{P}_{\dot{b}_1\cdots\dot{b}_{2s_2}}(s_2)\,|M, 0\rangle = \sum_{rt}\frac{(\lambda-\lambda_2)!\,(\lambda_2-\mu)!\,(p_0+p)^\sigma(p_0-p)^\tau}{r!\,t!\,(\lambda-\lambda_2-r)!\,(\lambda_2-\mu-t)!} \times T^{rt}_{m_1M}(s_1, \mu, \lambda_2)\dot{\underline{P}}^\sigma_{m_2}(s_2), \quad (2.17)$$

where $\sigma = r + t$ and $\tau = 2s_2 - r - t$. Now, the nonvanishing matrix elements of $\dot{\underline{P}}^{rt}_{m_2}(s_2)$ are those with $r + t = s_2 - m_2$, and those of $T^{rt}_{m_1M}(s_1, \mu, \lambda_2)$ have $r + t = s_2 + m_1 - M$. Thus, the nonvanishing matrix elements of (2.17) are those with

$$M = m_1 + m_2, \qquad r + t = s_2 - m_2. \quad (2.18)$$

Using (2.15), (2.17), and (2.18), we finally get

$$|s_1, m_1; s_2, m_2\rangle = (-1)^{\lambda_2-\mu}\left(\frac{(\lambda_1-\lambda_2+1)(\lambda_1+\lambda_2+2)(\lambda+\mu+1)(\lambda-\mu+1)}{(2\lambda_1+1)\cdot(\lambda_1-\mu+1)!\,(\lambda_1+\mu+1)!\,(\lambda_1-\lambda)!\,(\lambda+\lambda_1+2)!}\right)^{\frac{1}{2}} \times \left(\frac{p_0-p}{p_0+p}\right)^{m_2} C(s_1, s_2, \lambda_2; m_1, m_2, M)\,|\lambda_2, M; 0, 0\rangle. \quad (2.19)$$

Here $C(s_1, s_2, \lambda_2; m_1, m_2, M)$ is a Clebsch–Gordan coefficient.

With the choice (2.16) of the momentum, the helicity is $\boldsymbol{\Sigma}\cdot\mathbf{p}/p = \Sigma_3 = \Sigma_3^{(1)} + \Sigma_3^{(2)}$, where $\Sigma_3^{(1)}$ and $\Sigma_3^{(2)}$ have the eigenvalues m_1 and m_2. For a state of definite helicity $\Sigma_3 = S$, $|S| \le \lambda_2$, the nonvanishing matrix

elements (2.19) are only those with

$$M = m_1 + m_2 = S. \tag{2.20}$$

3. SPINORIAL FIELD EQUATIONS

The decomposition of ψ into Poincaré components $|s_1, s_2\rangle$ corresponds to the decomposition of the de Sitter group $O(4, 1)$ restricted to the proper Lorentz group $SO(3, 1)$, or actually to its universal covering group $SL(2, C)$. The two fundamental representations of the latter group $D(\frac{1}{2}, 0)$ and $D(0, \frac{1}{2})$ are carried by the 2-dimensional spinors ϕ_a and $\phi^{\dot b}$, respectively. Lowering and raising the indices is effected by the Levi-Civita symbols $\epsilon^{ab} = \epsilon_{ab} = -\epsilon_{ba} = -\epsilon^{ba}$, $\epsilon_{12} = 1$, such that

$$\phi^a = \epsilon^{ba}\phi_b, \quad \phi_a = \epsilon_{ab}\phi^b,$$

with similar relations for the dotted indices. The most general spinor will be denoted by

$$\begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1} \end{pmatrix} = \phi_{a_1 \cdots a_{2s_1}}^{\dot b_1 \cdots \dot b_{2s_2}}. \tag{3.1}$$

It is symmetric in the a's and the b's separately, and a_k and b_k take on the two values 1 and 2. This spinor transforms irreducibly according to the representation $D(s_1, s_2)$ of $SL(2, C)$; it has $(2s_1 + 1)(2s_2 + 1)$ independent components. It can be related to the Poincaré component $|s_1, s_2\rangle$ as follows[8]:

$$\begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1} \end{pmatrix} = R_{a_1 \cdots a_{2s_1}}(s_1) \otimes R^{\dot b_1 \cdots \dot b_{2s_2}}(s_2)\,|s_1, s_2\rangle, \tag{3.2}$$

$$|s_1, s_2\rangle = \frac{1}{(2s_1)!\,(2s_1)!} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1} \end{pmatrix} \times P^{a_1 \cdots a_{2s_1}}(s_1) \otimes P_{\dot b_1 \cdots \dot b_{2s_2}}(s_2). \tag{3.3}$$

R and P are given by (2.3) and satisfy the following relations:

$$P^{a_1 \cdots a_{2s}}(s) R_{a_1 \cdots a_{2s}}(s) = (2s)!, \tag{3.4}$$

$$R_{a_1 \cdots a_{2s}}(s) P^{b_1 \cdots b_{2s}}(s) = (-1)^{2s} R^{b_1 \cdots b_{2s}}(s) P_{a_1 \cdots a_{2s}}(s) = \sum \prod_{k=1}^{2s} \delta_{ak}^{bk} I, \tag{3.5}$$

where the summation extends over all permutations of the a's and the b's and I is the $(2s + 1) \times (2s + 1)$ unit matrix. One easily verifies that they also satisfy the following relations:

$$R^{a_1 \cdots a_{2s}}(s) u^a(s) = \sum_{k=1}^{2s} \epsilon^{aa_k} R^{a_1 \cdots a_{k-1} a_{k+1} \cdots a_{2s}}(s - \tfrac{1}{2}), \tag{3.6}$$

$$v_a(s) P_{a_1 \cdots a_{2s}}(s) = \sum_{k=1}^{2s} \epsilon_{aa_k} P_{a_1 \cdots a_{k-1} a_{k+1} \cdots a_{2s}}(s - \tfrac{1}{2}), \tag{3.7}$$

and similar relations for the dotted indices.

For convenience, we introduce the following lowering and raising of indices of the energy–momentum spinor $p^{a\dot b} = s_\mu{}^{a\dot b} p_\mu$:

$$p_a^{\dot b} = \epsilon_{ac} p^{c\dot b}, \quad p_{\dot b}^{a} = \epsilon_{\dot b \dot c} p^{a\dot c},$$

$$p_{\dot b a} = \epsilon_{ac} p_{\dot b}^{c} = \epsilon_{\dot b \dot c} p_a^{\dot c} = s_{\mu \dot b a} p_\mu. \tag{3.8}$$

The last relation follows from the identity

$$s_{\mu \dot b a} = \epsilon_{ac} \epsilon_{\dot b \dot d} s_\mu^{c\dot d}, \tag{3.9}$$

where $s_\mu{}^{a\dot b}$ and $s_{\mu \dot b a}$ are given by (I.3.1) and (I.3.2), respectively. (See also the Appendix.) Multiplying the field equations (2.1) by adequate R's, using (3.6), we obtain directly the following four spinorial equations:

$$(\lambda_1 + 1)(1 + x)(1 + y) p^{a\dot b} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2} \\ a a_1 \cdots a_{2s_1 - 1} \end{pmatrix} = m(2s_1 + 1)\Big[-(1 + y)\theta_{-+} \begin{pmatrix} \dot b \dot b_1 \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1 - 1} \end{pmatrix} + (1 + x)\theta_{--} \sum_{k=1}^{2s_2} \epsilon^{\dot b \dot b_k} \begin{pmatrix} \dot b_1 \cdots \dot b_{k-1} \dot b_{k+1} \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1 - 1} \end{pmatrix} \Big], \tag{3.10a}$$

$$(\lambda_1 + 1)(1 + x)(1 - y) \times \sum_{k=1}^{2s_1 + 1} p_{a_k}^{\dot b} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2} \\ a_1 \cdots a_{k-1} a_{k+1} \cdots a_{2s_1 + 1} \end{pmatrix} = m(2s_1 + 1)\Big[(1 - x)\theta_{++} \begin{pmatrix} \dot b \dot b_1 \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1 + 1} \end{pmatrix} + (1 - y)\theta_{+-} \sum_{k=1}^{2s_2} \epsilon^{\dot b \dot b_k} \begin{pmatrix} \dot b_1 \cdots \dot b_{k-1} \dot b_{k+1} \cdots \dot b_{2s_2} \\ a_1 \cdots a_{2s_1 + 1} \end{pmatrix} \Big], \tag{3.10b}$$

$$(\lambda_1 + 1)(1 - x)(1 + y) p_{\dot b a} \begin{pmatrix} \dot b \dot b_1 \cdots \dot b_{2s_2 - 1} \\ a_1 \cdots a_{2s_1} \end{pmatrix} = -m(2s_2 + 1)\Big[(1 + y)\theta_{+-} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2 - 1} \\ a a_1 \cdots a_{2s_1} \end{pmatrix} + (1 - x)\theta_{--} \sum_{k=1}^{2s_1} \epsilon_{a a_k} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2 - 1} \\ a_1 \cdots a_{k-1} a_{k+1} \cdots a_{2s_1} \end{pmatrix} \Big], \tag{3.10c}$$

$$(\lambda_1 + 1)(1 + x)(1 - y) \times \sum_{k=1}^{2s_2 + 1} p_a^{\dot b_k} \begin{pmatrix} \dot b_1 \cdots \dot b_{k-1} \dot b_{k+1} \cdots \dot b_{2s_2 + 1} \\ a_1 \cdots a_{2s_1} \end{pmatrix} = m(2s_2 + 1)\Big[(1 + x)\theta_{++} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2 + 1} \\ a a_1 \cdots a_{2s_1} \end{pmatrix} - (1 - y)\theta_{-+} \sum_{k=1}^{2s_1} \epsilon_{a a_k} \begin{pmatrix} \dot b_1 \cdots \dot b_{2s_2 + 1} \\ a_1 \cdots a_{k-1} a_{k+1} \cdots a_{2s_1} \end{pmatrix} \Big]. \tag{3.10d}$$

We get a hierarchy of spinorial equations (3.10) for all spinors with $\lambda_1 \geq s_1 + s_2 \geq \lambda_2 \geq |s_1 - s_2|$.

In the particular representation $R_5(\lambda_1, \lambda_1), s_1 + s_2 = \lambda$, $y = 1$, $\theta_{++} = \theta_{--} = 0$, and $\theta_{+-} = \theta_{-+} = 1$ [Eqs. (I.3.42) and (I.3.43)]. The field equations reduce to those of Dirac,[9] as readily shown by ten Kate[6]:

$$p^{a\dot{b}}\begin{pmatrix} \dot{b}_1 \cdots \dot{b}_{\lambda_1-\mu} \\ aa_1 \cdots a_{\lambda_1+\mu-1} \end{pmatrix} = -m\begin{pmatrix} b\dot{b}_1 \cdots \dot{b}_{\lambda_1-\mu} \\ a_1 \cdots a_{\lambda_1+\mu-1} \end{pmatrix},$$
$$p_{\dot{b}a}\begin{pmatrix} b\dot{b}_1 \cdots \dot{b}_{\lambda_1-\mu+1} \\ a_1 \cdots a_{\lambda_1+\mu} \end{pmatrix} = -m\begin{pmatrix} \dot{b}_1 \cdots \dot{b}_{\lambda_1-\mu+1} \\ aa_1 \cdots a_{\lambda_1+\mu} \end{pmatrix}, \quad (3.11)$$

for all $\mu = \lambda_1, \lambda_1 - 1, \cdots, -\lambda_1$.

Wigner's components $|\lambda_2, 0\rangle$ and $|0, \lambda_2\rangle$ are directly related to the spinors

$$(a_1 \cdots a_{2\lambda_2}) = \phi_{a_1 \cdots a_{2\lambda_2}} = R_{a_1 \cdots a_{2\lambda_2}}(\lambda_2)\,|\lambda_2, 0\rangle,$$
$$(\dot{b}_1 \cdots \dot{b}_{2\lambda_2}) = \phi^{\dot{b}_1 \cdots \dot{b}_{2\lambda_2}} = R^{\dot{b}_1 \cdots \dot{b}_{2\lambda_2}}(\lambda_2)\,|0, \lambda_2\rangle. \quad (3.12)$$

Multiplying Eqs. (2.7) and (2.9) by $R^{\dot{c}_1 \cdots \dot{c}_{2\lambda_2}}(\lambda_2)$ and $R^{c_1 \cdots c_{2\lambda_2}}(\lambda_2)$, respectively, using (3.5), it is easily verified that these spinors satisfy Weinberg's equations[10] in all representations $R_5(\lambda_1, \lambda_2)$:

$$\prod_{k=1}^{2\lambda_2} p^{a_k \dot{b}_k}(a_1 \cdots a_{2\lambda_2}) = (-m)^{2\lambda_2}(\dot{b}_1 \cdots \dot{b}_{2\lambda_2}),$$
$$\prod_{k=1}^{2\lambda_2} p_{a_k \dot{b}_k}(\dot{b}_1 \cdots \dot{b}_{2\lambda_2}) = (-m)^{2\lambda_2}(a_1 \cdots a_{2\lambda_2}). \quad (3.13)$$

4. RELATION BETWEEN TENSORS AND SPINORS

A. Pseudotensors

The irreducible representations $R_4(\lambda, \mu)$ of the proper Lorentz group $SO(3, 1)$ are characterized by two integers or two half-integers, λ and μ, where $\lambda \geq |\mu|$ and μ may be positive or negative. The corresponding equivalent representation $D(s_1, s_2)$ of the universal covering group $SL(2, C)$ is that with $\lambda = s_1 + s_2$ and $\mu = s_1 - s_2$. The double-valued representations of $SO(3, 1)$ are those with λ and μ half-integral; they actually belong to $SL(2, C)$. $SO(3, 1)$ possesses three fundamental representations

$$\begin{aligned} R_4(1, 0) &= D(\tfrac{1}{2}, \tfrac{1}{2}) \sim f_\mu \quad \sim \phi_a^{\dot{b}}, \\ R_4(1, 1) &= D(1, 0) \sim f_{(\mu\nu)_+} \sim \phi_{ab}, \\ R_4(1, -1) &= D(0, 1) \sim f_{(\mu\nu)_-} \sim \phi^{\dot{a}\dot{b}}, \end{aligned} \quad (4.1)$$

where f_μ is a 4-vector and $f_{(\mu\nu)_+}$ and $f_{(\mu\nu)_-}$ are antisymmetric self-dual and anti-self-dual tensors, each of three independent components, such that

$$f_{(\mu\nu)_\pm} = \pm\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta} f_{(\alpha\beta)_\pm}. \quad (4.2)$$

These are pseudotensors, transforming into each other under spatial inversions. The connection between these tensors and the corresponding spinors of $SL(2, C)$ are as follows[8]:

$$\begin{aligned} f_\mu &= s_{\mu\dot{b}}^{a}\phi_a^{\dot{b}} = s_{\mu\dot{b}}^{a} v_a(\tfrac{1}{2}) v^{\dot{b}}(\tfrac{1}{2})\,|\tfrac{1}{2}, \tfrac{1}{2}\rangle, \\ f_{(\mu\nu)_+} &= s_{\mu\nu}{}^{ab}\phi_{ab} = s_{\mu\nu}{}^{ab} v_a(\tfrac{1}{2}) v_b(1)\,|1, 0\rangle, \\ f_{(\mu\nu)_-} &= \bar{s}_{\mu\nu\dot{a}\dot{b}}\phi^{\dot{a}\dot{b}} = \bar{s}_{\mu\nu\dot{a}\dot{b}} v^{\dot{a}}(\tfrac{1}{2}) v^{\dot{b}}(1)\,|0, 1\rangle. \end{aligned} \quad (4.3)$$

The inverse relations are

$$\begin{aligned} \phi_a^{\dot{b}} &= \tfrac{1}{2} s_\mu{}^{\dot{b}}{}_a f_\mu, \\ \phi_{ab} &= \tfrac{1}{8} s_{\mu\nu ab} f_{(\mu\nu)_+}, \\ \phi^{\dot{a}\dot{b}} &= \tfrac{1}{8} \bar{s}_{\mu\nu}{}^{\dot{a}\dot{b}} f_{(\mu\nu)_-}. \end{aligned} \quad (4.4)$$

Here $s_\mu{}^a{}_{\dot{b}}$ and $s_a{}^{\dot{b}}{}_\mu$ are obtained by lowering indices of $s_\mu{}^{a\dot{b}}$. We define $s_{\mu\nu}{}^{ab}$ and $\bar{s}_{\mu\nu}{}^{\dot{a}\dot{b}}$ in the Appendix. They are self-dual and anti-self-dual antisymmetric tensors in the indices $\mu\nu$, respectively, and symmetric in the spinor indices ab and $\dot{a}\dot{b}$.

The irreducible tensor, transforming according to $R_4(l + n, \pm n)$, where l and n are nonnegative integers, is of the form

$$f_{[\alpha_1 \cdots \alpha_l][(\mu_1\nu_1) \cdots (\mu_n\nu_n)]_\pm}. \quad (4.5)$$

It is symmetric in the indices α_i and also in the subbrackets $(\mu_i\nu_i)$. Each subbracket $(\mu\nu)$ is antisymmetric. All subbrackets are self-dual in the representation $R_4(l + n, n)$ and anti-self-dual in the representation $R_4(l + n, n)$. They are denoted correspondingly by $[(\mu_1\nu_1) \cdots (\mu_n\nu_n)]_+$ and $[(\mu_1\nu_1) \cdots (\mu_n\nu_n)]_-$. The tensor (4.5) is traceless with respect to contractions of any two indices.

We note, however, that the irreducible tensors can also be of the form

$$f_{[\alpha_1 \cdots \alpha_K][(\mu_1\nu_1) \cdots (\mu_L\nu_L)]_+[(\sigma_1\tau_1) \cdots (\sigma_M\tau_M)]_-}, \quad (4.6)$$

which transforms according to $R_5(K + L + M, L - M)$. This tensor can be rewritten in the form (4.5) with $n = |L - M|$ and $l = K + L + M - |L - M|$, with signature $\pm = (L - M)/|L - M|$. Using (4.3), we obtain the connection between the tensor (4.6) and the corresponding spinor as follows:

$$f_{[\alpha_1 \cdots \alpha_K][(\mu_1\nu_1) \cdots (\mu_L\nu_L)]_+[(\sigma_1\tau_1) \cdots (\sigma_M\tau_M)]_-} = \prod_{i=1}^{K} s_{\alpha_i}{}^{a_i}{}_{\dot{b}_i} \prod_{j=1}^{L} s_{\mu_j\nu_j}{}^{e_j f_j} \prod_{k=1}^{M} \bar{s}_{\sigma_k\tau_k\dot{c}_k\dot{d}_k} \times \begin{pmatrix} \dot{b}_1 \cdots \dot{b}_K & \dot{c}_1\dot{d}_1 \cdots \dot{c}_M\dot{d}_M \\ a_1 \cdots a_K & e_1 f_1 \cdots e_L f_L \end{pmatrix}. \quad (4.7)$$

In the previous notation

$$\begin{aligned} \lambda &= K + L + M, \quad \mu = L - M, \\ s_1 &= \tfrac{1}{2}K + L, \qquad s_2 = \tfrac{1}{2}K + M. \end{aligned} \quad (4.8)$$

For the same representation $D(s_1, s_2)$, we can construct several tensors of the form (4.7) by taking all possible values of K, L, and M which are consistent with (4.8). Using the contraction relations between $s_\mu{}^a{}_{\dot{b}}$, $s_{\mu\nu}{}^{ab}$, and $\bar{s}_{\mu\nu\dot{a}\dot{b}}$, given in the Appendix, one can

easily prove the following tracelessness conditions:

$$f_{[\alpha\alpha\alpha_3\cdots\alpha_K](\mu_i\nu_i)\}+\{(\sigma_j\tau_j)\}_-} = 0,$$
$$f_{\{\alpha_i\}[(\mu\nu_1)(\mu\nu_2)(\mu_3\nu_3)\cdots]_\pm\{(\sigma_j\tau_j)\}_\mp} = 0, \quad (4.9)$$
$$f_{[\alpha\alpha_2\cdots\alpha_K][(\alpha\nu_1)(\mu_2\nu_2)\cdots]_\pm\{(\sigma_i\tau_i)\}_\mp} = 0.$$

Besides, one establishes the following connection between equivalent tensors:

$$f_{[\alpha_1\cdots\alpha_K][(\mu\nu_1)(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_+[(\mu\tau_1)(\sigma_2\tau_2)\cdots(\sigma_M\tau_M)]_-}$$
$$= 2f_{[\alpha_1\cdots\alpha_K\nu_1\tau_1][(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_+[(\sigma_2\tau_2)\cdots(\sigma_M\tau_M)]_-}. \quad (4.10)$$

Further, using (3.10a), we get

$$p_\alpha f_{[\alpha\alpha_2\cdots\alpha_K][(\mu_i\nu_i)\}+\{(\sigma_j\tau_j)\}_-}$$
$$= -m[(2s_1+1)(2s_2+1)\theta_{--}/(\lambda_1+\lambda+2)]$$
$$\times f_{[\alpha_2\cdots\alpha_K][(\mu_i\nu_i)\}+\{(\sigma_j\tau_j)\}_-}, \quad (4.11)$$

where s_1 and s_2 are given by (4.8).

B. Spin-Tensors

For half-integral spin the representations $R_4(\lambda,\mu) = D(s_1, s_2)$ are characterized by half-integral values of λ and μ, such that s_1 is integral and s_2 is half-integral, or the contrary. The corresponding spinor has an odd number of indices. We construct in this case the spin-tensors

$$f_{a[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_+[(\sigma_1\tau_1)\cdots(\sigma_M\tau_M)]_-}$$
$$= \prod_{i=1}^{K} s_{\alpha_i}{}^{a_i}\dot{b}_i \prod_{j=1}^{L} s_{\mu_j\nu_j}{}^{e_jf_j} \prod_{k=1}^{M} \bar{s}_{\sigma_k\tau_k}\dot{c}_k\dot{d}_k$$
$$\times \begin{pmatrix} \dot{b}_1\cdots\dot{b}_K & \dot{c}_1\dot{d}_1\cdots\dot{c}_M\dot{d}_M \\ aa_1\cdots a_K & e_1f_1\cdots e_Lf_L \end{pmatrix}, \quad (4.12)$$

transforming according to $R_4(K+L+M+\frac{1}{2}, L-M+\frac{1}{2})$, and

$$f^{\dot{b}}_{[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_+[(\sigma_1\tau_1)\cdots(\sigma_M\tau_M)]_-}$$
$$= \prod_{i=1}^{K} s_{\alpha_i}{}^{a_i}\dot{b}_i \prod_{j=1}^{L} s_{\mu_j\nu_j}{}^{e_jf_j} \prod_{k=1}^{M} \bar{s}_{\sigma_k\tau_k}\dot{c}_k\dot{d}_k$$
$$\times \begin{pmatrix} \dot{b}\dot{b}_1\cdots\dot{b}_K & \dot{c}_1\dot{d}_1\cdots\dot{c}_M\dot{d}_M \\ a_1\cdots a_K & e_1f_1\cdots e_Lf_L \end{pmatrix}, \quad (4.13)$$

transforming according to $R_4(K+L+M+\frac{1}{2}, L-M-\frac{1}{2})$. These spin-tensors transform as 4-dimensional tensors in the Greek indices and as 2-dimensional spinors in the indices a and $\dot{b}$. They satisfy the irreducibility conditions (4.9), as well as relation (4.10), for each spinor component. One easily verifies that they also satisfy the following relations:

$$s^{a\dot{b}}_{\alpha_1} f_{a[\alpha_1\alpha_2\cdots\alpha_K](\mu_i\nu_i)\}+\{(\sigma_j\tau_j)\}_-} = 0, \quad (4.14a)$$
$$s_{\alpha_1\dot{b}a} f^{\dot{b}}_{[\alpha_1\alpha_2\cdots\alpha_K](\mu_i\nu_i)\}+\{(\sigma_j\tau_j)\}_-} = 0, \quad (4.14b)$$
$$s_{\mu_1}{}^{ab} f_{a\{\alpha_i\}[(\mu_1\nu_1)(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_+\{(\sigma_j\tau_j)\}_-} = 0, \quad (4.14c)$$
$$s_{\sigma_1\dot{b}a} f^{\dot{b}}_{\{\alpha_i\}(\mu_j\nu_j)\}+[(\sigma_1\tau_1)(\sigma_2\tau_2)\cdots(\sigma_M\tau_M)]_-} = 0, \quad (4.14d)$$
$$s_{\sigma_1}{}^{ab} f_{a[\alpha_1\cdots\alpha_K](\mu_i\nu_i)\}+[(\sigma_1\tau_1)(\sigma_2\tau_2)\cdots(\sigma_M\tau_M)]_-}$$
$$= 2f^{\dot{b}}_{[\alpha_1\cdots\alpha_K\tau_1](\mu_j\nu_j)\}+[(\sigma_2\tau_2)\cdots(\sigma_M\tau_M)]_-}, \quad (4.14e)$$
$$s_{\mu_1\dot{b}a} f^{\dot{b}}_{[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_+\{(\sigma_i\tau_i)\}_-}$$
$$= -2f_{a[\alpha_1\cdots\alpha_K\nu_1][(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_+\{(\sigma_i\tau_i)\}_-}. \quad (4.14f)$$

Further, it is easily verified that each spinor component satisfies Eq. (4.11). Also,

$$p^{a\dot{b}} f_{a[\alpha_1\cdots\alpha_K](\mu_i\nu_i)\}+[(\sigma_1\tau_1)\cdots(\sigma_M\tau_M)]_-}$$
$$= -m[(2s_1+1)\theta_{-+}/(\lambda_1+\mu+1)] f^{\dot{b}}_{[\alpha_1\cdots\alpha_K](\mu_i\nu_i)\}+[(\sigma_1\tau_1)\cdots(\sigma_M\tau_M)]_-}$$
$$- m[(2s_1+1)\theta_{--}/(\lambda_1+\lambda+2)]\Big(\sum_{k=1}^{K} s^{a\dot{b}}_{\alpha_k} f_{a[\alpha_1\cdots\alpha_{k-1}\alpha_{k+1}\cdots\alpha_K](\mu_i\nu_i)\}+[(\sigma_1\tau_1)\cdots(\sigma_M\tau_M)]_-}$$
$$+ 2\sum_{k=1}^{M} \bar{s}_{\sigma_k\tau_k}{}^{\dot{b}}{}_{\dot{c}} f^{\dot{c}}_{[\alpha_1\cdots\alpha_K](\mu_i\nu_i)\}+[(\sigma_1\tau_1)\cdots(\sigma_{k-1}\tau_{k-1})(\sigma_{k+1}\tau_{k+1})\cdots(\sigma_M\tau_M)]_-}\Big), \quad (4.15)$$

$$p_{\dot{b}a} f^{\dot{b}}_{[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_+\{(\sigma_i\tau_i)\}_-}$$
$$= -m[(2s_2+1)\theta_{+-}/(\lambda_1-\mu+1)] f_{a[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_+\{(\sigma_i\tau_i)\}_-}$$
$$- m[(2s_2+1)\theta_{--}/(\lambda_1+\lambda+2)]\Big(\sum_{k=1}^{K} s_{\alpha_k\dot{b}a} f^{\dot{b}}_{[\alpha_1\cdots\alpha_{k-1}\alpha_{k+1}\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_+\{(\sigma_i\tau_i)\}_-}$$
$$+ 2\sum_{k=1}^{L} s_{\mu_k\nu_k}{}^{c}{}_{a} f_{c[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_{k-1}\nu_{k-1})(\mu_{k+1}\nu_{k+1})\cdots(\mu_L\nu_L)]_+\{(\sigma_i\tau_i)\}_-}\Big). \quad (4.16)$$

We note that, with each two-valued representation $R_4(\lambda,\mu)$ of the proper Lorentz group, we can associate a spin-tensor of the form

$$f_{a[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_+}, \quad (4.17a)$$

with $\lambda = K+L+\frac{1}{2}$ and $\mu = L+\frac{1}{2}$, or a spin-tensor of the form

$$f^{\dot{b}}_{[\alpha_1\cdots\alpha_K][(\mu_1\nu_1)\cdots(\mu_L\nu_L)]_-}, \quad (4.17b)$$

with $\lambda = K+L+\frac{1}{2}$ and $\mu = -(L+\frac{1}{2})$.

5. 5-TENSORS

A. The Poincaré Components in Terms of the 5-Tensor

The 5-dimensional tensor carrying the representation $R_5(l+n, n)$ of $O(4, 1)$, where l and n are two nonnegative integers, is of the form

$$T_{[A_1\cdots A_l][(B_1C_1)\cdots(B_nC_n)]}. \tag{5.1}$$

It satisfies the irreducibility conditions

$$T_{[AAA_3\cdots A_l][(B_iC_i)]} = 0, \tag{5.2a}$$
$$T_{\{A_i\}[(BC_1)(BC_2)\cdots(B_nC_n)]} = 0, \tag{5.2b}$$
$$\epsilon_{B_1C_1B_2DE}T_{\{A_i\}[(B_1C_1)(B_2C_2)\cdots(B_nC_n)]} = 0, \tag{5.2c}$$
$$T_{[AA_2\cdots A_l][(AC_1)(B_2C_2)]} = 0, \tag{5.2d}$$
$$\epsilon_{A_1B_1C_1DE}T_{[A_1A_2\cdots A_l][(B_1C_1)(B_2C_2)\cdots(B_nC_n)]} = 0. \tag{5.2e}$$

The field equations, corresponding to Eqs. (1.1) and satisfied by the tensor (5.1), are[6]

$$p_{A_1}T_{[A_1A_2\cdots A_l][(B_iC_i)]} = 0, \tag{5.3a}$$
$$p_{B_1}T_{\{A_i\}[(B_1C_1)(B_2C_2)\cdots(B_nC_n)]} = 0, \tag{5.3b}$$
$$\epsilon_{B_1C_1DEF}p_DT_{\{A_i\}[(B_1C_1)(B_2C_2)\cdots(B_nC_n)]} = 0, \tag{5.3c}$$
$$p_AT_{[A_1A_2\cdots A_l][(B_iC_i)]} = p_{A_1}T_{[AA_2\cdots A_l][(B_iC_i)]}. \tag{5.3d}$$

For convenience we denote the different 4-dimensional components of the 5-tensor (5.1) as follows:

$$T^{(l-t)}_{[\alpha_1\cdots\alpha_t][\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = T_{[\alpha_1\cdots\alpha_t 55\cdots 5][(5\beta_1)(5\beta_2)\cdots(5\beta_{n-k})(\mu_1\nu_1)\cdots(\mu_k\nu_k)]}. \tag{5.4}$$

The number of indices 5 in the first completely symmetric bracket is $l-t$. This 4-dimensional tensor is reducible under $SO(3, 1)$. From this tensor we can construct the irreducible traceless tensor

$$f_{[\alpha_1\cdots\alpha_t\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]_\pm}, \tag{5.5}$$

which carries the representation $R_4(n+t, \pm k)$. The totality of the tensors (5.5) with $0 \le t \le l$ and $0 \le k \le n$ afford a basis for the representation $R_5(l+n, n)$. In fact, this corresponds to the decomposition

$$R_5(l+n, n) = \sum_{t=0}^{l}\sum_{k=-n}^{n} \oplus R_4(n+t, k). \tag{5.6}$$

Since (5.5) can be related to the Poincaré component $|s_1, s_2\rangle$ via the corresponding spinor, we can relate $|s_1, s_2\rangle$ direct to (5.4) as follows:

If $s_1 \ge s_2$, we take

$$2s_1 = n+t+k, \quad 2s_2 = n+t-k, \quad \lambda = n+t, \quad \mu = k. \tag{5.7}$$

Then, in favor of (3.3) and (4.7), we have

$$\begin{aligned}|s_1, s_2\rangle = 2^{-\mu}a(\lambda,\mu)&\prod_{i=1}^{t} s_{\alpha_i}{}^{a_i}{}_{b_i}\prod_{j=1}^{n-k} s_{\beta_j}{}^{c_j}{}_{d_j}\prod_{r=1}^{k} s_{\mu_r\nu_r}{}^{e_rf_r}\\ &\times T^{(l-t)}_{[\alpha_1\cdots\alpha_t][\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]}\\ &\times P_{a_1\cdots a_t\,c_1\cdots c_{n-k}\,e_1f_1\cdots e_kf_k}(s_1)\\ &\otimes P^{b_1\cdots b_t d_1\cdots d_{n-k}}(s_2).\end{aligned} \tag{5.8}$$

On the other hand, if $s_1 \le s_2$, we take

$$2s_1 = n+t-k, \quad 2s_2 = n+t+k, \quad \lambda = n+t, \quad \mu = -k. \tag{5.9}$$

Then

$$\begin{aligned}|s_1, s_2\rangle = 2^{\mu}b(\lambda,\mu)&\prod_{i=1}^{t} s_{\alpha_i}{}^{a_i}{}_{b_i}\prod_{j=1}^{n-k} s_{\beta_j}{}^{c_j}{}_{d_j}\prod_{r=1}^{k} \bar{s}_{\mu_r\nu_r}{}^{\dot e_r\dot f_r}\\ &\times T^{(l-t)}_{[\alpha_1\cdots\alpha_t][\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]}\\ &\times P_{a_1\cdots a_t\,c_1\cdots c_{n-k}}(s_1)\\ &\otimes P^{b_1\cdots b_t d_1\cdots d_{n-k}\,\dot e_1\dot f_1\cdots\dot e_k\dot f_k}(s_2),\end{aligned} \tag{5.10}$$

where $a(\lambda, \mu)$ and $b(\lambda, \mu)$ are coupling coefficients which depend on the field equations satisfied by ψ.

To determine the coupling coefficients for the free field, we use Eqs. (2.2) and (5.3). With the help of relations (3.7) we get, after some lengthy manipulations, the following relations. Applying Eq. (2.2a) for (5.8) and (5.10), we get

$$\frac{a(\lambda,\mu)}{a(\lambda-1,\mu)} = \frac{b(\lambda,\mu)}{b(\lambda-1,\mu)} = \frac{(\lambda+\mu+1)(\lambda-\mu+1)\theta_{--}}{(\lambda_1+\lambda+2)(\lambda-\lambda_2)(\lambda+\lambda_2)}, \tag{5.11}$$

where $\lambda_1 = l+n$ and $\lambda_2 = n$; by applying Eq. (2.2c) for (5.8), we get

$$\frac{a(\lambda,\mu)}{a(\lambda,\mu+1)} = \frac{-(\lambda+\mu+1)(\lambda-\mu+1)\theta_{+-}}{(\lambda_2-\mu)(\lambda_1-\mu+1)}; \tag{5.12}$$

finally, applying Eq. (2.2b) for (5.10), we get

$$\frac{b(\lambda,\mu)}{b(\lambda,\mu-1)} = \frac{(\lambda+\mu+1)(\lambda-\mu+1)\theta_{-+}}{(\lambda_2+\mu)(\lambda_1+\mu+1)} \tag{5.13}$$

From these recurrence relations, using (I.3.40) and (I.3.41), we get

$$\begin{aligned}b(\lambda,\mu) &= (-1)^{\lambda_2-\mu}a(\lambda,\mu)\\ &= \{[(\lambda+\lambda_2+1)\cdot(2\lambda_2)!]/[(2\lambda_2+1)\\ &\quad\cdot(\lambda_2-\mu)!\,(\lambda_2+\mu)!]\}\\ &\quad\times A(\lambda,\mu)a(\lambda_2,\lambda_2),\end{aligned} \tag{5.14}$$

where $a(\lambda_2, \lambda_2)$ is an arbitrary normalization constant and $A(\lambda, \mu)$ is given by (2.6). We note that $a(\lambda, \mu)$ and $b(\lambda, \mu)$ are defined for $\mu \ge 0$ and $\mu \le 0$, respectively.

The inverse relation for expressing the components (5.4) in terms of the Poincaré components is complicated in the general case $R_5(l+n, n)$. We next consider the two particular representations $R_5(l, 0)$ and $R_5(n, n)$.

B. The Representation $R_5(l, 0)$

The tensor carrying this representation is $T_{[A_1 \cdots A_l]}$. Its 4-dimensional components $T^{(l-k)}_{[\alpha_1 \cdots \alpha_k]}$, $0 \leq k \leq l$, satisfy the condition

$$T^{(l-k)}_{[\alpha\alpha\alpha_3 \cdots \alpha_k]} = -T^{(l-k+2)}_{[\alpha_3 \cdots \alpha_k]}. \tag{5.15}$$

These 4-tensors are not traceless. They can be expressed in terms of traceless tensors $F_{[\alpha_1 \cdots \alpha_s]}$ as follows:

$$T^{(l-k)}_{[\alpha_1 \cdots \alpha_k]} = \sum_{t=0}^{[\frac{1}{2}k]} a_t(k) F^{(2t)}_{[\alpha_1 \cdots \alpha_k]}, \tag{5.16}$$

where $[\frac{1}{2}k]$ is the integral part of $\frac{1}{2}k$ and

$$F^{(2t)}_{[\alpha_1 \cdots \alpha_k]} = \sum \delta_{\alpha_1\alpha_2} \cdots \delta_{\alpha_{2t-1}\alpha_{2t}} F_{[\alpha_{2t+1} \cdots \alpha_k]}; \tag{5.17}$$

the summation extends over all permutations of the α's. The coefficients $a_t(k)$ are determined partially by the condition (5.15). Noting that

$$F^{(2t)}_{[\alpha\alpha\alpha_3 \cdots \alpha_k]} = 2(k-t+1) F^{(2t-2)}_{[\alpha_3 \cdots \alpha_k]}, \tag{5.18}$$

we obtain the recurrence relation

$$a_{t-1}(k-2) = -2(k-t+1) a_t(k). \tag{5.19}$$

Hence,

$$a_t(k) = (-\tfrac{1}{2})^t [(k-2t+1)!/(k-t+1)!] a_0(k-2t), \tag{5.20}$$

where $a_0(k)$ remains arbitrary and depends on the field equations satisfied by the tensor $T_{[A_1 \cdots A_l]}$.

To express $T^{(l-k)}_{[\alpha_1 \cdots \alpha_k]}$ in terms of Poincaré components $|\frac{1}{2}k, \frac{1}{2}k\rangle$ of the representation with $0 \leq k \leq l$, we need only express the irreducible tensors $F_{[\alpha_1 \cdots \alpha_k]}$ in terms of $|\frac{1}{2}k, \frac{1}{2}k\rangle$. For convenience, we write

$$F_{[\alpha_1 \cdots \alpha_k]} = \frac{1}{(k+1)!} \left(\frac{(l-k)!\,(l+k+2)!}{k+1} \right)^{\frac{1}{2}} f_{[\alpha_1 \cdots \alpha_k]}, \tag{5.21}$$

where $f_{[\alpha_1 \cdots \alpha_k]}$ is defined by (4.7). Using (3.2), we get

$$\begin{aligned} f_{[\alpha_1 \cdots \alpha_k]} &= \prod_{i=1}^{k} s_{\alpha_i}{}^{a_i} {}_{b_i} \begin{pmatrix} b_1 \cdots b_k \\ a_1 \cdots a_k \end{pmatrix} \\ &= \prod_{i=1}^{k} s_{\alpha_i}{}^{a_i} {}_{b_i} R_{a_1 \cdots a_k}(\tfrac{1}{2}k) \otimes R^{b_1 \cdots b_k}(\tfrac{1}{2}k)\, |\tfrac{1}{2}k, \tfrac{1}{2}k\rangle. \end{aligned} \tag{5.22}$$

In order to determine $a_0(k)$ for the free field satisfying Eqs. (5.3a) and (5.3d), it is more convenient to consider the inverse relation

$$F_{[\alpha_1 \cdots \alpha_k]} = \sum_{t=0}^{[\frac{1}{2}k]} b_t(k) T^{(l-k+2t,2t)}_{[\alpha_1 \cdots \alpha_k]}, \tag{5.23}$$

where

$$T^{(l-k+2t,2t)}_{[\alpha_1 \cdots \alpha_k]} = \sum \delta_{\alpha_1\alpha_2} \cdots \delta_{\alpha_{2t-1}\alpha_{2t}} T^{(l-k+2t)}_{[\alpha_{2t+1} \cdots \alpha_k]}. \tag{5.24}$$

Comparing the leading terms in (5.16) and (5.23), we get

$$a_0(k) = 1/b_0(k). \tag{5.25}$$

Now, in favor of (5.15), we have

$$\begin{aligned} &T^{(l-k+2t,2t)}_{[\alpha\alpha\alpha_3 \cdots \alpha_k]} \\ &\quad = 2(k-t+1) T^{(l-k+2t,2t-2)}_{[\alpha_3 \cdots \alpha_k]} - T^{(l-k+2t+2,2t)}_{[\alpha_3 \cdots \alpha_k]}. \end{aligned} \tag{5.26}$$

Thus, from the tracelessness condition $F_{[\alpha\alpha\alpha_3 \cdots \alpha_k]} = 0$, we get

$$b_{t-1}(k) = 2(k-t+1) b_t(k). \tag{5.27}$$

Hence,

$$b_t(k) = [(k-t)!/(2^t \cdot k!)] b_0(k). \tag{5.28}$$

Now, from the field equation (5.3d), we get

$$T^{(l-k)}_{[\alpha_1 \cdots \alpha_k]} = m^{-k} p_{\alpha_1} \cdots p_{\alpha_k} T^{(l)}, \tag{5.29}$$

where $T^{(l)} = T_{55 \ldots 5}$ is a Lorentz invariant. Noting that $p_\mu p_\mu = -m^2$, we get

$$\begin{aligned} \frac{p_\alpha}{m} T^{(l-k+2t,2t)}_{[\alpha\alpha_2 \cdots \alpha_k]} &= (k-2t+1) T^{(l-k+2t-1,2t-2)}_{[\alpha_2 \cdots \alpha_k]} \\ &\quad - T^{(l-k+2t+1,2t)}_{[\alpha_2 \cdots \alpha_k]} \end{aligned} \tag{5.30}$$

Further, from (4.11) and (I.3.49), we get

$$p_\alpha F_{[\alpha\alpha_2 \cdots \alpha_k]} = -m F_{[\alpha_2 \cdots \alpha_k]}. \tag{5.31}$$

Using (5.23), (5.30), and (5.31), we obtain the recurrence relation

$$b_t(k) = b_t(k-1) + (k-2t+1) b_{t+1}(k). \tag{5.32}$$

Combining (5.28) and (5.32), we get

$$b_0(k) = 1/a_0(k) = 2^k/(k+1), \tag{5.33}$$

where, for convenience, we choose $a_0(0) = b_0(0) = 1$. Hence,

$$b_t(k) = 2^{k-t} \cdot (k-t)!/(k+1)! \tag{5.34}$$

and

$$\begin{aligned} a_t(k) = (-1)^t [(k-2t+1) \cdot (k-2t+1)!]/ \\ [2^{k-t} \cdot (k-t+1)!]. \end{aligned} \tag{5.35}$$

C. The Representation $R_5(n, n)$

The carrier of this representation is the tensor $T_{[(B_1C_1) \cdots (B_nC_n)]}$. We denote its 4-dimensional

components as follows:

$$\phi_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = T_{[(5\beta_1)\cdots(5\beta_{n-k})(\mu_1\nu_1)\cdots(\mu_k\nu_k)]}, \quad (5.36)$$

for all $0 \leq k \leq n$. For convenience, we denote the dual subbracket as follows:

$$(\overline{\mu\nu}) = \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}(\alpha\beta).$$

The irreducibility conditions (5.2b) and (5.2c) read in the 4-dimensional notation as follows:

$$\phi_{[\beta\beta\beta_3\cdots\beta_{n-k}][(\mu_i\nu_i)\}} = 0, \quad (5.37a)$$

$$\phi_{[\beta\beta_2\cdots\beta_{n-k}][(\beta\nu_1)(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} = 0, \quad (5.37b)$$

$$\phi_{[\beta\beta_2\cdots\beta_{n-k}][(\overline{\beta\nu_1})(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} = 0, \quad (5.37c)$$

$$\phi_{\{\beta_i\}[(\overline{\mu\nu_1})(\mu\nu_2)(\mu_3\nu_3)\cdots(\mu_k\nu_k)]} = 0, \quad (5.37d)$$

$$\phi_{[\beta_1\cdots\beta_{n-k}][(\mu\nu_1)(\mu\nu_2)(\mu_3\nu_3)\cdots(\mu_k\nu_k)]} = -\phi_{[\beta_1\cdots\beta_{n-k}\nu_1\nu_2][(\mu_3\nu_3)\cdots(\mu_k\nu_k)]}. \quad (5.37e)$$

The 4-dimensional irreducible tensors constituting a basis for this representation are

$$f_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_t\nu_t)]_+[(\mu_{t+1}\nu_{t+1})\cdots(\mu_k\nu_k)]_-},$$

for all $0 \leq t \leq k \leq n$. They are defined by (4.7) and satisfy (4.9). They are self-dual and anti-self-dual in the subbrackets with subscripts + and −, respectively. We note that the conditions (5.37a)–(5.37d) are common for the tensors ϕ and f. This suggests that ϕ may be written in the form

$$\phi_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = \sum_{t=0}^{k} c_t(k) F^{(t,k-t)}_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]}, \quad (5.38)$$

where

$$F^{(t,k-t)}_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = \sum f_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_t\nu_t)]_+[(\mu_{t+1}\nu_{t+1})\cdots(\mu_k\nu_k)]_-}. \quad (5.39)$$

Here t and $k - t$ are the numbers of self-dual and anti-self-dual subbrackets, respectively. The summation implies symmetrization in all subbrackets $(\mu_i\nu_i)$. From (4.10) we have

$$F^{(t,k-t)}_{[\beta_1\cdots\beta_{n-k}][(\mu\nu_1)(\mu\nu_2)(\mu_3\nu_3)\cdots(\mu_k\nu_k)]} = 4F^{(t-1,k-t-1)}_{[\beta_1\cdots\beta_{n-k}\nu_1\nu_2][(\mu_3\nu_3)\cdots(\mu_k\nu_k)]}. \quad (5.40)$$

Using (5.37e), (5.38), and (5.40), we obtain the recurrence relation

$$c_{t-1}(k-2) = -4c_t(k). \quad (5.41)$$

Thus

$$c_t(k) = (-\tfrac{1}{4})^t c_0(k-2t). \quad (5.42)$$

This relation holds irrespective of the field equations satisfied by the tensors.

To determine $c_t(k)$ for the free field, we use the field equations

$$p_\beta\phi_{[\beta\beta_2\cdots\beta_{n-k}][(\mu_i\nu_i)\}} = 0, \quad (5.43a)$$

$$p_\mu\phi_{\{\beta_i\}[(\overline{\mu\nu_1})(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} = 0, \quad (5.43b)$$

$$p_\mu\phi_{[\beta_1\cdots\beta_{n-k}][(\mu\nu_1)(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} = -m\phi_{[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]}. \quad (5.43c)$$

Further, from (4.11) we get

$$p_\beta f_{[\beta\beta_2\cdots\beta_{n-k}][(\mu_i\nu_i)\}_+\{(\sigma_i\tau_i)\}_-} = 0, \quad (5.44)$$

since in this representation $\theta_{--} = \theta_{++} = 0$, $\theta_{+-} = \theta_{-+} = 1$, $\lambda = s_1 + s_2 = n$, and $\mu = s_1 - s_2 = \pm k$. Moreover, using (3.10a), (3.10c), and (4.7), we get

$$p_\mu f_{[\beta_1\cdots\beta_K][(\mu\nu_1)(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_\pm\{(\sigma_i\tau_i)\}_\mp} = \mp m f_{[\beta_1\cdots\beta_K\nu_1][(\mu_2\nu_2)\cdots(\mu_L\nu_L)]_\pm\{(\sigma_i\tau_i)\}_\mp}. \quad (5.45)$$

Thus

$$(p_\mu/m)F^{(t,k-t)}_{[\beta_1\cdots\beta_{n-k}][(\mu\nu_1)(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} = F^{(t,k-t-1)}_{[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} - F^{(t-1,k-t)}_{[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]}, \quad (5.46)$$

and

$$-(p_\mu/m)F^{(t,k-t)}_{[\beta_1\cdots\beta_{n-k}][(\overline{\mu\nu_1})(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} = F^{(t,k-t-1)}_{[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]} + F^{(t-1,k-t)}_{[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]}. \quad (5.47)$$

From (5.38), (5.43b), and (5.46) we get

$$c_t(k) = -c_{t-1}(k), \quad (5.48)$$

and from (5.43c) and (5.47) we get

$$c_t(k-1) = c_{t+1}(k) - c_t(k). \quad (5.49)$$

Hence

$$c_t(k) = (-1)^{t+k}2^{-k}c(0), \quad (5.50)$$

which is consistent with (5.41).

6. 5-SPIN-TENSORS

A. The Poincaré Components in Terms of the 5-Spin-Tensor

The representations $R_5(\lambda_1, \lambda_2)$ with half-integral λ_1 and λ_2 actually belong to $Sp(2,2)$, the universal covering group of $SO(4,1)$. The carrier of the representation $R_5(l + n + \frac{1}{2}, n + \frac{1}{2})$ is the spin-tensor[6]

$$T_{[A_1\cdots A_l][(B_1C_1)\cdots(B_nC_n)]} = \begin{pmatrix} T^a_{[A_1\cdots A_l][(B_1C_1)\cdots(B_nC_n)]} \\ T_{\dot b[A_1\cdots A_l][(B_1C_1)\cdots(B_nC_n)]} \end{pmatrix}. \quad (6.1)$$

It is a 4-column matrix, transforming as a tensor in the indices A_i and (B_iC_i) and as a spinor in the indices

a and $\dot{b}$. In particular,

$$T = \begin{pmatrix} T^a \\ T_{\dot{b}} \end{pmatrix} \tag{6.2}$$

is a Dirac 4-spinor, transforming according to the fundamental representation $R_5(\frac{1}{2}, \frac{1}{2})$ of $Sp(2, 2)$. In this representation the pseudovector Γ_A, defined by (I.2.15), is expressed in terms of the Dirac matrices γ_μ as follows:

$$\Gamma_\mu = \gamma_5\gamma_\mu, \quad \Gamma_5 = \gamma_5 = -\gamma_1\gamma_2\gamma_3\gamma_4. \tag{6.3}$$

The spin-tensor satisfies the irreducibility conditions (5.2) together with

$$\Gamma_A T_{[AA_2\cdots A_t](B_iC_i)} = 0, \tag{6.4a}$$

$$\Gamma_B T_{\{A_i\}[(BC_1)(B_2C_2)\cdots(B_nC_n)]} = 0. \tag{6.4b}$$

It satisfies the field equations (5.3), together with[6]

$$\Gamma_A p_A T_{\{A_i\}(B_jC_j)\}} = 0, \tag{6.5}$$

which is just the Dirac equation (multiplied by Γ_5)

$$(\gamma_\mu p_\mu + m)T_{\{A_i\}(B_jC_j)\}} = 0. \tag{6.6}$$

We denote the components of (6.1) in the same way as in (5.4), except for adding an upper suffix a or a lower suffix $\dot{b}$.

Consider the following representation of the Dirac matrices:

$$\gamma_k = \begin{pmatrix} 0 & -s_k \\ s_k & 0 \end{pmatrix}, \quad \gamma_4 = i\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \quad \gamma_5 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix},$$

where the s_k are the Pauli matrices. The matrix elements of γ_μ are then

$$\gamma_\mu = \begin{pmatrix} 0 & -s_\mu^{ab} \\ -s_{\mu\dot{b}a} & 0 \end{pmatrix}. \tag{6.7}$$

Conditions (6.4) become

$$s_{\alpha_1}^{ab} T^{(l-t)}_{\dot{b}[\alpha_1\alpha_2\cdots\alpha_t](\beta_i)\{(\mu_j\nu_j)\}} = T^{a(l-t+1)}_{[\alpha_2\cdots\alpha_t](\beta_i)\{(\mu_j\nu_j)\}}, \tag{6.8a}$$

$$s_{\alpha_1\dot{b}a} T^{a(l-t)}_{[\alpha_1\alpha_2\cdots\alpha_t](\beta_i)\{(\mu_j\nu_j)\}} = T^{(l-t+1)}_{\dot{b}[\alpha_2\cdots\alpha_t](\beta_i)\{(\mu_j\nu_j)\}}, \tag{6.8b}$$

$$s_{\beta_1}^{ab} T^{(l-t)}_{\dot{b}\{\alpha_i\}[\beta_1\beta_2\cdots](\mu_j\nu_j)\}} = 0, \tag{6.8c}$$

$$s_{\beta_1\dot{b}a} T^{a(l-t)}_{\{\alpha_i\}[\beta_1\beta_2\cdots](\mu_j\nu_j)\}} = 0, \tag{6.8d}$$

$$s_{\mu_1}^{a\dot{b}} T^{(l-t)}_{\dot{b}\{\alpha_i\}[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)(\mu_2\nu_2)\cdots]} = T^{a(l-t)}_{\{\alpha_i\}[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots]}, \tag{6.8e}$$

$$s_{\mu_1\dot{b}a} T^{a(l-t)}_{\{\alpha_i\}[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)(\mu_2\nu_2)\cdots]} = T_{\dot{b}\{\alpha_i\}[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots]}. \tag{6.8f}$$

Also, Dirac's equation (6.6) reads

$$p^{a\dot{b}} T_{\dot{b}\{\}} = mT^a_{\{\}}, \qquad p_{\dot{b}a} T^a_{\{\}} = mT_{\dot{b}\{\}}, \tag{6.9}$$

where $\{\}$ stands for all tensorial indices.

For each spinor component of T we can construct a spin-tensor of the form (4.17). Thus the Poincaré components can be written as follows:

For $s_1 > s_2$, we take

$$\lambda = n + t + \tfrac{1}{2}, \qquad \mu = k + \tfrac{1}{2},$$
$$2s_1 = n + t + k + 1, \quad 2s_2 = n + t - k.$$

Then

$$|s_1, s_2\rangle = 2^{-\mu+\frac{1}{2}} a(\lambda, \mu) \prod_{i=1}^{t} s_{\alpha_i}^{a_i\dot{b}_i} \prod_{j=1}^{n-k} s_{\beta_j}^{c_j\dot{d}_j} \prod_{r=1}^{k} s_{\mu_r\nu_r}^{e_r f_r} \times T^{a(l-t)}_{[\alpha_1\cdots\alpha_t][\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} \times P_{aa_1\cdots a_t\, c_1\cdots c_{n-k}\, e_1 f_1\cdots e_k f_k}(s_1) \otimes P^{\dot{b}_1\cdots\dot{b}_t\dot{d}_1\cdots\dot{d}_{n-k}}(s_2). \tag{6.10}$$

On the other hand, for $s_1 < s_2$, we take

$$\lambda = n + t + \tfrac{1}{2}, \qquad \mu = -(k + \tfrac{1}{2}),$$
$$2s_1 = n + t - k, \quad 2s_2 = n + t + k + 1.$$

Then

$$|s_1, s_2\rangle = 2^{\mu+\frac{1}{2}} b(\lambda, \mu) \prod_{i=1}^{t} s_{\alpha_i}^{a_i\dot{b}_i} \prod_{j=1}^{n-k} s_{\beta_j}^{c_j\dot{d}_j} \prod_{r=1}^{k} \bar{s}_{\mu_r\nu_r\dot{e}_r\dot{f}_r} \times T^{(l-t)}_{\dot{b}[\alpha_1\cdots\alpha_t][\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} \times P_{a_1\cdots a_t\, c_1\cdots c_{n-k}}(s_1) \otimes P^{\dot{b}\dot{b}_1\cdots\dot{b}_t\dot{d}_1\cdots\dot{d}_{n-k}\dot{e}_1\dot{f}_1\cdots\dot{e}_k\dot{f}_k}(s_2). \tag{6.11}$$

Following the same steps as in Sec. 5A, using (6.8) and (6.9), one verifies, after some lengthy manipulations, that $a(\lambda, \mu)$ and $b(\lambda, \mu)$ are given by (5.14), where $\lambda_1 = l + n + \frac{1}{2}$ and $\lambda_2 = n + \frac{1}{2}$.

B. The Representation $R_5(l + \frac{1}{2}, \frac{1}{2})$

The carrier of this representation is the spin-tensor $T_{[A_1\cdots A_l]}$. Its 4-dimensional components are the spin-tensors $T^{(l-k)}_{[\alpha_1\cdots\alpha_k]}$, $0 \le k \le l$. We construct the spin-tensors of the Rarita–Schwinger type[11]

$$F_{[\alpha_1\cdots\alpha_k]} = \begin{pmatrix} f^a_{[\alpha_1\cdots\alpha_k]} \\ f_{\dot{b}[\alpha_1\cdots\alpha_k]} \end{pmatrix}, \tag{6.12}$$

where $f^a_{\{\alpha_i\}}$ and $f_{\dot{b}\{\alpha_i\}}$ are given by (4.13) and (4.14), by raising and lowering the spinor indices. Using the representation (6.7) for γ_μ, one verifies that the spin-tensor (6.12) satisfies the conditions

$$F_{[\alpha\alpha\alpha_3\cdots\alpha_k]} = 0, \tag{6.13a}$$

$$\gamma_\alpha F_{[\alpha\alpha_2\cdots\alpha_k]} = 0, \tag{6.13b}$$

in favor of (4.14). Further, we introduce the spin-tensor

$$G_{[\alpha_1\cdots\alpha_k]} = \sum_{i=1}^{k} \gamma_{\alpha_i} F_{[\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_k]} = \begin{pmatrix} g^a_{[\alpha_1\cdots\alpha_k]} \\ g_{\dot{b}[\alpha_1\cdots\alpha_k]} \end{pmatrix}, \tag{6.14a}$$

where

$$g^{a}_{[\alpha_1\cdots\alpha_k]} = -\sum_{i=1}^{k} s^{ab} f_{b[\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_k]},$$

$$g_{b[\alpha_1\cdots\alpha_k]} = -\sum_{i=1}^{k} s_{\alpha_i ba} f^{a}_{[\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_k]}. \quad (6.14b)$$

One easily verifies that G satisfies the following conditions:

$$G_{[\alpha\alpha\alpha_3\cdots\alpha_k]} = 0, \quad (6.15a)$$

$$\gamma_\alpha G_{[\alpha\alpha_2\cdots\alpha_k]} = -2(k+1)F_{[\alpha_2\cdots\alpha_k]}. \quad (6.15b)$$

F and G are related to the Poincaré components $|\frac{1}{2}(k+1), \frac{1}{2}k\rangle$ and $|\frac{1}{2}k, \frac{1}{2}(k+1)\rangle$ in the usual way.

The components of T can be written in the form

$$T^{(l-k)}_{[\alpha_1\cdots\alpha_k]} = \sum_{t=0}^{[\frac{1}{2}k]} A_t(k) F^{(2t)}_{[\alpha_1\cdots\alpha_k]} + \sum_{t=1}^{[\frac{1}{2}k]} B_t(k) G^{(2t)}_{[\alpha_1\cdots\alpha_k]}, \quad (6.16)$$

where $F^{(2t)}$ and $G^{(2t)}$ are defined by (5.17). They satisfy Eq. (5.18). In addition, they satisfy the following equations:

$$\gamma_{\alpha_1} F^{(2t)}_{[\alpha_1\alpha_2\cdots\alpha_k]} = G^{(2t-2)}_{[\alpha_2\cdots\alpha_k]}, \quad (6.17a)$$

$$\gamma_{\alpha_1} G^{(2t)}_{[\alpha_1\alpha_2\cdots\alpha_k]} = -2(k-t+1)F^{(2t)}_{[\alpha_2\cdots\alpha_k]}. \quad (6.17b)$$

T satisfies (5.15), together with

$$\gamma_{\alpha_1} T^{(l-k)}_{[\alpha_1\alpha_2\cdots\alpha_k]} = -T^{(l-k+1)}_{[\alpha_2\cdots\alpha_k]}. \quad (6.18)$$

Using Eqs. (5.15) and (5.18), it is easily verified that $A_t(k)$ and $B_t(k)$ satisfy both relations (5.19) and (5.20). Further, using (6.17) and (6.18), we get

$$A_t(k) = -B_{t-1}(k-1), \quad (6.19a)$$

$$A_t(k-1) = 2(k-t+1)B_t(k). \quad (6.19b)$$

Equation (6.19b) is derivable from Eq. (6.19a), in favor of Eq. (5.20). Thus we get one relation between $A_t(k)$ and $B_t(k)$.

To determine $A_t(k)$ and $B_t(k)$ completely, we use the field equations. From (4.15), (4.16), (I.3.48), and (I.3.49) we get

$$(k+1)(\gamma_\mu p_\mu/m)F_{[\alpha_1\cdots\alpha_k]} + F_{[\alpha_1\cdots\alpha_k]} = -(k+2)[(l-k+1)/(l+k+3)]^{\frac{1}{2}} G_{[\alpha_1\cdots\alpha_k]}. \quad (6.20)$$

Further, using (3.10), (4.12), and (4.13), we obtain, after some lengthy manipulations,

$$(k+1)(\gamma_\mu p_\mu/m)G_{[\alpha_1\cdots\alpha_k]} - G_{[\alpha_1\cdots\alpha_k]} = -k[(l+k+3)/(l-k+1)]^{\frac{1}{2}} F_{[\alpha_1\cdots\alpha_k]}. \quad (6.21)$$

Using Eqs. (6.20) and (6.21) together with the Dirac equation

$$(\gamma_\mu p_\mu + m)T^{(l-k)}_{[\alpha_1\cdots\alpha_k]} = 0, \quad (6.22)$$

we obtain the relation

$$B_t(k) = [(l-k+2t+1)/(l+k-2t+3)]^{\frac{1}{2}} A_t(k). \quad (6.23)$$

Using (6.19b), we get

$$2(k-t+1)A_t(k) = [l+k-2t+3)/(l-k+2t+1)]^{\frac{1}{2}} A_t(k-1). \quad (6.24)$$

Using (5.20) and (6.24), we get

$$A_t(k) = (-1)^t\{1/[2^{k-t}\cdot(k-t+1)!]\} \times [(l+k-2t+3)!\,(l-k+2t)!/(l+3)!\,l!]^{\frac{1}{2}} A_0(0). \quad (6.25)$$

$B_t(k)$ is then determined from (6.23).

C. The Representation $R_5(n+\frac{1}{2}, n+\frac{1}{2})$

The components of the spin-tensor $T_{[(B_1C_1)\cdots(B_nC_n)]}$, carrying this representation, may be written in the form

$$T^{a}_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = \sum_{t=0}^{k} C_t(k) F^{a(t,k-t)}_{\{\beta\}\{(\mu\nu)\}}, \quad (6.26)$$

$$T_{b[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = \sum_{t=0}^{k} D_t(k) F^{(t,k-t)}_{b\{\beta\}\{(\mu\nu)\}}, \quad (6.27)$$

$F^{a(t,k-t)}_{\{\ \}}$ and $F^{t,k-t)}_{b\{\ \}}$ are defined by (5.39) in terms of the corresponding spin-tensors. They satisfy the same contraction conditions and field equations as (5.39). Thus $C_t(k)$ and $D_t(k)$ satisfy (5.41) irrespective of the field equations. For the free field they are both given by (5.50).

Using Eq. (4.14), one easily verifies that

$$s_{\mu_1 ba} F^{a(t,k-t)}_{[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = -2F^{(t,k-t-1)}_{b[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]}, \quad (6.28)$$

$$s^{ab}_{\mu_1} F^{(t,k-t)}_{b[\beta_1\cdots\beta_{n-k}][(\mu_1\nu_1)\cdots(\mu_k\nu_k)]} = 2F^{a(t-1,k-t)}_{[\beta_1\cdots\beta_{n-k}\nu_1][(\mu_2\nu_2)\cdots(\mu_k\nu_k)]}. \quad (6.29)$$

These relations, together with (6.8c) and (6.8d), give

$$D_t(k-1) = -2C_t(k),$$
$$C_t(k-1) = 2D_{t+1}(k), \quad (6.30)$$

which are consistent with (5.41).

Further, from (4.15) and (5.16) we get

$$p_{ba} F^{a(t,k-t)}_{\{\ \}} = mF^{(t,k-t)}_{b\{\ \}},$$
$$p^{ab} F^{(t,k-t)}_{b\{\ \}} = mF^{a(t,k-t)}_{\{\ \}}. \quad (6.31)$$

Using Dirac's equation (6.9) together with (6.31), we finally get

$$D_t(k) = C_t(k). \quad (6.32)$$

Equations (6.30) and (6.32) lead to (5.40), proving the self-consistency of the choice (6.26) and (6.27).

ACKNOWLEDGMENTS

One of the authors (M. M. B.) would like to thank Professor A. Salam and Professor P. Budini as well as the International Atomic Energy Agency for hospitality at the International Centre for Theoretical Physics, Trieste, where this work was completed.

APPENDIX

We give here some contraction relations between the spinor-vectors $s_\mu^{a\dot{b}}$, which are the matrix elements in the row a and the column b of the matrix s_μ, defined by

$$s_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad s_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$
$$s_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad s_4 = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}. \tag{A1}$$

We construct further

$$s_\mu^{\ a}{}_{\dot{b}} = \epsilon_{\dot{b}\dot{c}} s_\mu^{a\dot{c}}, \quad s_\mu^{\ \dot{b}}{}_a = \epsilon_{ac} s_\mu^{c\dot{b}}, \quad s_{\mu\dot{a}\dot{b}} = \epsilon_{\dot{a}\dot{c}}\epsilon_{b\dot{d}} s_\mu^{\dot{d}\dot{c}}. \tag{A2}$$

One easily verifies that

$$s_{k\dot{a}\dot{b}} = -s_k^{a\dot{b}}, \quad s_{4\dot{a}\dot{b}} = s_4^{a\dot{b}}, \quad k = 1, 2, 3, \tag{A3}$$

in accordance with the definition (I.3.2). The following relations hold:

$$s_\mu^{a\dot{b}} s_\mu^{\dot{c}\dot{d}} = -2\epsilon^{ac}\epsilon^{\dot{b}\dot{d}}, \tag{A4}$$
$$s_\mu^{a\dot{b}} s_{\nu\dot{b}a} = -2\delta_{\mu\nu}, \tag{A5}$$
$$s_\mu^{a\dot{c}} s_\nu^{\ b}{}_{\dot{c}} - s_\mu^{\ b\dot{c}} s_\nu^{\ a}{}_{\dot{c}} = -2\epsilon^{ab}\delta_{\mu\nu}, \tag{A6}$$
$$s_\mu^{c\dot{a}} s_\nu^{\ \dot{b}}{}_c - s_\mu^{\ \dot{b}} s_\nu^{c\dot{a}}{}_c = -2\epsilon^{\dot{a}\dot{b}}\delta_{\mu\nu}, \tag{A7}$$
$$s_\mu^{a\dot{b}} s_{\nu\dot{b}c} s_\alpha^{\ c}{}_{\dot{d}} = \epsilon_{\mu\nu\alpha\beta} s_\beta^{\ a}{}_{\dot{d}} + \delta_{\mu\alpha} s_\nu^{\ a}{}_{\dot{d}} - \delta_{\nu\alpha} s_\mu^{\ a}{}_{\dot{d}} - \delta_{\mu\nu} s_\alpha^{\ a}{}_{\dot{d}}. \tag{A8}$$

Further, we introduce the fundamental spinor-tensors

$$\begin{aligned} s_{\mu\nu}^{\ ab} &= s_{\mu\nu}^{\ ba} = -s_{\nu\mu}^{\ ab} = \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta} s_\alpha^{\ ab} \\ &= \tfrac{1}{2}(s_\mu^{a\dot{c}} s_\nu^{\ b}{}_{\dot{c}} + s_\mu^{b\dot{c}} s_\nu^{\ a}{}_{\dot{c}}) \\ &= \begin{cases} s_\mu^{a\dot{c}} s_\nu^{\ b}{}_{\dot{c}}, & \text{if } \mu \neq \nu, \\ 0, & \text{if } \mu = \nu. \end{cases} \end{aligned} \tag{A9}$$

and

$$\begin{aligned} \bar{s}_{\mu\nu}^{\ \dot{a}\dot{b}} &= \bar{s}_{\mu\nu}^{\ \dot{b}\dot{a}} = -\bar{s}_{\nu\mu}^{\ \dot{a}\dot{b}} = -\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}\bar{s}_\alpha^{\ \dot{a}\dot{b}} \\ &= \tfrac{1}{2}(s_\mu^{c\dot{a}} s_\nu^{\ \dot{b}}{}_c + s_\mu^{c\dot{b}} s_\nu^{\ \dot{a}}{}_c) \\ &= \begin{cases} s_\mu^{c\dot{a}} s_\nu^{\ \dot{b}}{}_c, & \text{if } \mu \neq \nu, \\ 0, & \text{if } \mu = \nu. \end{cases} \end{aligned} \tag{A10}$$

They satisfy the following relations:

$$s_{\mu\nu}^{\ a}{}_c s_\alpha^{\ c}{}_b = \epsilon_{\mu\nu\alpha\beta} s_\beta^{\ a}{}_b + \delta_{\mu\alpha} s_\nu^{\ a}{}_b - \delta_{\nu\alpha} s_\mu^{\ a}{}_b, \tag{A11}$$
$$\bar{s}_{\mu\nu}^{\ \dot{a}}{}_c s_\alpha^{\ c}{}_b = -\epsilon_{\mu\nu\alpha\beta} s_\beta^{\ \dot{a}}{}_b + \delta_{\mu\alpha} s_\nu^{\ \dot{a}}{}_b - \delta_{\nu\alpha} s_\mu^{\ \dot{a}}{}_b, \tag{A12}$$
$$s_{\mu\nu}^{\ ab} s_{\alpha\beta ab} = 2(\epsilon_{\mu\nu\alpha\beta} + \delta_{\mu\alpha}\delta_{\nu\beta} - \delta_{\nu\alpha}\delta_{\mu\beta}), \tag{A13}$$
$$\bar{s}_{\mu\nu}^{\ \dot{a}\dot{b}} \bar{s}_{\alpha\beta\dot{a}\dot{b}} = 2(-\epsilon_{\mu\nu\alpha\beta} + \delta_{\mu\alpha}\delta_{\nu\beta} - \delta_{\nu\alpha}\delta_{\mu\beta}). \tag{A14}$$

Further,

$$s_\mu^{\ a}{}_{\dot{b}} s_{\mu\nu}^{\ c\dot{d}} = \epsilon^{ac} s_\nu^{\ \dot{d}}{}_{\dot{b}} + \epsilon^{a\dot{d}} s_\nu^{\ c}{}_{\dot{b}}, \tag{A15}$$
$$s_\mu^{\ \dot{b}}{}_a \bar{s}_{\mu\nu}^{\ c\dot{d}} = \epsilon^{\dot{b}c} s_\nu^{\ \dot{d}}{}_a + \epsilon^{\dot{b}\dot{d}} s_\nu^{\ c}{}_a, \tag{A16}$$
$$\begin{aligned} s_{\mu\alpha}^{\ ab} s_{\alpha\nu}^{\ c\dot{d}} = \tfrac{1}{2}(&\epsilon^{ac} s_\mu^{\ b\dot{e}} s_\nu^{\ \dot{d}}{}_{\dot{e}} + \epsilon^{a\dot{d}} s_\mu^{\ b\dot{e}} s_\nu^{\ c}{}_{\dot{e}} \\ &+ \epsilon^{bc} s_\mu^{\ a\dot{e}} s_\nu^{\ \dot{d}}{}_{\dot{e}} + \epsilon^{b\dot{d}} s_\mu^{\ a\dot{e}} s_\nu^{\ c}{}_{\dot{e}}), \end{aligned} \tag{A17}$$
$$\begin{aligned} \bar{s}_{\mu\alpha\dot{a}\dot{b}} \bar{s}_{\alpha\nu c\dot{d}} = \tfrac{1}{2}(&\epsilon_{\dot{a}c} s_\mu^{\ e}{}_{\dot{b}} s_{\nu\dot{d}e} + \epsilon_{\dot{a}\dot{d}} s_\mu^{\ e}{}_{\dot{b}} s_{\nu ec} \\ &+ \epsilon_{\dot{b}c} s_\mu^{\ e}{}_{\dot{a}} s_{\nu\dot{d}e} + \epsilon_{\dot{b}\dot{d}} s_\mu^{\ e}{}_{\dot{a}} s_{\nu\dot{c}e}), \end{aligned} \tag{A18}$$
$$\begin{aligned} s_{\mu\alpha}^{\ ab} \bar{s}_{\alpha\nu c\dot{d}} = -\tfrac{1}{2}(&s_\mu^{\ a}{}_c s_\nu^{\ b}{}_{\dot{d}} + s_\mu^{\ a}{}_{\dot{d}} s_\nu^{\ b}{}_c \\ &+ s_\mu^{\ b}{}_{\dot{c}} s_\nu^{\ a}{}_{\dot{d}} + s_\mu^{\ a}{}_{\dot{d}} s_\nu^{\ a}{}_{\dot{c}}), \end{aligned} \tag{A19}$$
$$s_{\mu\nu}^{\ ab} s_{\mu\nu}^{\ c\dot{d}} = 4(\epsilon^{ac}\epsilon^{b\dot{d}} + \epsilon^{a\dot{d}}\epsilon^{bc}), \tag{A20}$$
$$\bar{s}_{\mu\nu\dot{a}\dot{b}} \bar{s}_{\mu\nu\dot{c}\dot{d}} = 4(\epsilon_{\dot{a}c}\epsilon_{\dot{b}\dot{d}} + \epsilon_{\dot{a}\dot{d}}\epsilon_{\dot{b}c}). \tag{A21}$$

* On leave of absence from Cairo University, Giza, U.A.R.

[1] M. M. Bakri, J. Math. Phys. **10**, 298 (1969).

[2] By assuming the same 5-dimensional symmetry, equivalent equations were obtained by J. A. de Vos and J. Hilgevoord, Nucl. Phys. **B1**, 494 (1967).

[3] M. M. Bakri, Nucl. Phys. **B12**, 341 (1969).

[4] A. Salam, R. Delbourgo, and J. Strathdee, Proc. Roy. Soc. (London) **A284**, 146 (1965).

[5] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **34**, 211 (1948).

[6] A. ten Kate, Nucl. Phys. **B8**, 73 (1968).

[7] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[8] See E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave Equations* (Blackie, London, 1954).

[9] P. A. M. Dirac, Proc. Roy. Soc. (London) **A155**, 447 (1936).

[10] S. Weinberg, Phys. Rev. **133B**, 1318 (1964).

[11] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

De Sitter Symmetric Field Theory. III. Quantization in the Tung–Weinberg Basis

M. M. BAKRI*
International Centre for Theoretical Physics, Trieste, Italy

(Received 17 October 1969)

The de Sitter symmetric fields, satisfying the field equations $(S_{\mu\nu}p_\nu + \kappa\gamma_\mu)\psi = p_\mu\psi$, are quantized in the Tung–Weinberg basis. The properties of the associated Green's and causal functions are studied. Transformations of the Hilbert space under *CPT* and inhomogeneous Lorentz transformations are given. A simple closed expression is obtained for the Shaw–Weinberg matrix tensor $t_{m_1m_2}^{\alpha_1\cdots\alpha_{2s}}$.

1. INTRODUCTION

In the present paper we quantize the free de Sitter symmetric fields studied before,[1,2] which are the solutions of the field equations

$$(S_{\mu\nu}p_\nu + \kappa\gamma_\mu) = p_\mu\psi.$$

We follow the quantization scheme proposed by Tung[3] and studied extensively recently by Weinberg,[4] in which the Hilbert space is constructed from the linearly independent states corresponding to the "physical" spin multiplicity. This means that, for each representation $R_5(\lambda_1, \lambda_2)$ of the homogeneous de Sitter group, we introduce $2(2\lambda_2 + 1)$ creation operators for particles and antiparticles. We study the quantized field in two bases:

(i) The observable spin basis, which was studied for more general fields by Pursey [5];

(ii) The helicity basis, which was studied for general fields by Tung[3] and Weinberg.[4]

We define the associated Green's and causal functions and study their properties. Inhomogeneous Lorentz and *CPT* transformations of the Hilbert space are derived.

Although our equations are not derived from a Lagrangian and are not even equivalent to a single equation, all results obtained bear close similarity to the schemes following from Schwinger's action principle or from one matrix equation, as in the Umezawa–Visconti–Takabashi formalism.[6]

In Appendix A we prove the independence of the Lorentz-invariant scalar product of the hypersurface on which it is defined. In Appendix B we obtain a very simple closed expression for t-matrix tensor introduced by Weinberg[7,4] and Shaw.[8] We use in this derivation a relation obtained in Paper II[2] by a direct solution of the wave equations. This shows the power of wave equations in studying the free fields, since these equations serve as identities among the field components.

We have carried out the quantization in terms of the compact matrix form ψ of the field. These results can be rewritten in multispinor forms, using the connections between multispinors and the components of ψ given in Ref. 9. This helps us to compare our formalism with those of Guralnik and Kibble[10] and Chang.[11] Also, one can relate our quantization of ψ to Chang's quantization[12] of the Pauli–Fierz and Rarita–Schwinger fields. This can easily be done using the connections between 5-tensors and spin tensors and the components of ψ established in Paper II.[2]

2. QUANTIZATION IN WEINBERG'S BASIS

A. Plane-wave Solutions

For the finite-dimensional irreducible representations $R_5(\lambda_1, \lambda_2)$ of the de Sitter group $O(4, 1)$, the field equations[1]

$$(S_{\mu\nu}\partial_\nu + i\kappa\gamma_\mu)\psi = \partial_\mu\psi \tag{2.1}$$

possess $2(2\lambda_2 + 1)$ linearly independent solutions of the form

$$\psi_M(x; \mathbf{p}, \epsilon) = (2\pi)^{-\frac{3}{2}} \exp\,[i(\mathbf{r}\cdot\mathbf{p} - \epsilon E_p t)]w_M(\mathbf{p}, p_0). \tag{2.2}$$

Here $p_0 = \epsilon E_p$, $\epsilon = \pm 1$, and $E_p = (\mathbf{p}^2 + \kappa^2)^{\frac{1}{2}}$. M is the eigenvalue of the helicity operator

$$\boldsymbol{\Sigma}\cdot\mathbf{p}w_M(\mathbf{p}, p_0) = Mpw_M(\mathbf{p}, p_0), \tag{2.3}$$

where $\boldsymbol{\Sigma} = \lambda_1\boldsymbol{\sigma}$ is the "mathematical" spin operator and $p = (\mathbf{p}^2)^{\frac{1}{2}}$, or is of the third component of the observable spin $\mathbf{S}$:

$$S_3w_M(\mathbf{p}, p_0) = Mw_M(\mathbf{p}, p_0). \tag{2.4}$$

The observable spin is defined in terms of the Pauli–Lubanski pseudovector $W_\mu = \frac{1}{2}\lambda_1\epsilon_{\mu\nu\alpha\beta}S_{\alpha\beta}p_\nu$ as [cf. (I.4.25)]

$$\mathbf{S} = (\epsilon/\kappa)\mathbf{W} - [\kappa(E_p + \kappa)]^{-1}W_0\mathbf{p}. \tag{2.5}$$

It satisfies the relation

$$\mathbf{S}^2\psi = \kappa^{-2}W_\mu W_\mu\psi = \lambda_2(\lambda_2 + 1)\psi.$$

The observable spin is λ_2, and its component M takes the values $-\lambda_2, -\lambda_2+1, \cdots, \lambda_2$.

$w_M(\mathbf{p}, p_0)$ is related to the rest system wavefunction $\phi_M(\epsilon)$ as follows:

$$w_M(\mathbf{p}, p_0) = B(\mathbf{p}, \epsilon)\phi_M(\epsilon). \tag{2.6}$$

$B(\mathbf{p}, \epsilon)$ is the Lorentz boost which takes the momentum $(0, 0, 0, \epsilon\kappa)$ into $(\mathbf{p}, p_0)$. It is given by

$$B(\mathbf{p}, \epsilon) = \exp(\lambda_1\theta_1\boldsymbol{\alpha}\cdot\mathbf{p}/p), \tag{2.7}$$

where

$$\cosh\theta_1 = E_p/\kappa, \quad \sinh\theta_1 = \epsilon p/\kappa. \tag{2.8}$$

$B(\mathbf{p}, \epsilon) = V^{-1}(\mathbf{p}, \epsilon)$, where $V(\mathbf{p}, \epsilon)$ is the Chakrabarti transformation [cf. (I.5.1)]. The observable spin is related to the "mathematical spin" $\boldsymbol{\Sigma}$ as follows:

$$\mathbf{S} = B\boldsymbol{\Sigma}B^{-1}. \tag{2.9}$$

The state of definite observable spin is characterized by

$$\Sigma_3\phi_M = M\phi_M. \tag{2.10}$$

On the other hand, the helicity states are characterized by

$$\boldsymbol{\Sigma}\cdot\mathbf{p}\phi_M = Mp\phi_M, \tag{2.11}$$

since $[\boldsymbol{\Sigma}\cdot\mathbf{p}, B] = 0$.

Normalizing ϕ_M such that

$$\phi_M^+\phi_{M'} = \delta_{MM'}, \tag{2.12}$$

we then normalize $\psi_M(x, \mathbf{p}, \epsilon)$ as follows:

$$(\psi_M(x, \mathbf{p}, \epsilon), \psi_{M'}(x, \mathbf{p}', \epsilon')) = c_\epsilon(E_p/\kappa)\delta_{MM'}\delta_{\epsilon\epsilon'}\delta^3(\mathbf{p}-\mathbf{p}'), \tag{2.13}$$

where

$$c_\epsilon = \epsilon\beta_\epsilon = 1, \quad \text{for fermions, } \lambda_2 \text{ half-integral,}$$
$$= \epsilon, \quad \text{for bosons, } \lambda_2 \text{ integral.} \tag{2.14}$$

The Lorentz-invariant scalar product of two functions ψ_1 and ψ_2 is defined as

$$(\psi_1, \psi_2) = \int \psi_1\gamma_\alpha\psi_2\, d\sigma_\alpha, \tag{2.15}$$

where the integration is carried over a spacelike hypersurface σ. Relation (2.13) was proved in Part I for a flat surface $t = \text{const}$, such that

$$(\psi_1, \psi_2) = \int \bar{\psi}_1\gamma_0\psi_2 d^3x. \tag{2.16}$$

In Appendix A we prove that the scalar product (2.15) is independent of the choice of the hypersurface σ. This proves the Lorentz covariance of (2.13). It proves also that the Lorentz-covariant expectation value of a dynamical variable A is given by

$$\langle A\rangle = \tfrac{1}{2}[(\psi, A\psi) + (A\psi, \psi)]$$
$$= \tfrac{1}{2}\int(\bar{\psi}\gamma_0 A\psi + \overline{A\psi}\gamma_0\psi)d^3x. \tag{2.17}$$

In particular, the total charge is

$$Q = e\int\bar{\psi}\gamma_0\psi d^3x, \tag{2.18}$$

and the total momentum

$$P_\mu = \frac{1}{2}\int(\psi\gamma_0\hat{P}_\mu\psi + \overline{\hat{P}_\mu\psi}\gamma_0\psi)d^3x$$
$$= \frac{1}{2i}\int[\psi\gamma_0\partial_\mu\psi - \overline{(\partial_\mu\psi)}\gamma_0\psi]d^3x. \tag{2.19}$$

$\hat{P}_\mu = -iS_{\mu\nu}\partial_\nu + m\gamma_\mu$ is the single-particle energy–momentum operator. The second line in (2.19) follows from the field equations (2.1). Expressions (2.18) and (2.19) (the second line) can be derived from a variational principle for the Lagrangian

$$\mathfrak{L} = -i\bar{\psi}\gamma_\mu\partial_\mu\psi + \kappa\bar{\psi}\psi.$$

B. Quantization

The solutions of the free field equations (2.1) do not span the carrier space of the representation $R_5(\lambda_1, \lambda_2)$ except in the case of Dirac's theory ($\lambda_1 = \lambda_2 = \frac{1}{2}$). In fact, the carrier space decomposes under the proper Lorentz group as follows:

$$\psi = \sum\oplus\psi(s_1, s_2), \tag{2.20}$$

where $\psi(s_1, s_2)$ transforms according to $D(s_1, s_2)$ and the summation extends over all

$$\lambda_1 \geq s_1 + s_2 \geq \lambda_2 \geq |s_1 - s_2|. \tag{2.21}$$

Each inequivalent representation $D(s_1, s_2)$ occurs only once. For the solutions of the field equations (2.1), each component $\psi(s_1, s_2)$ is the carrier of an irreducible representation of the Poincaré group with spin λ_2 and mass κ; it satisfies

$$(p_\mu p_\mu + \kappa^2)\psi(s_1, s_2) = 0,$$
$$W_\mu W_\mu\psi(s_1, s_2) = \kappa^2\lambda_2(\lambda_2+1)\psi(s_1, s_2). \tag{2.22}$$

All Poincaré components $\psi(s_1, s_2)$ of ψ are related uniquely to one independent component $\psi(\lambda_2, 0)$ for each energy sign by means of the field equations. $\psi(\lambda_2, 0)$ is Wigner's unitary representation of the Poincaré group. The connection between $\psi(s_1, s_2)$ and $\psi(\lambda_2, 0)$ is derived in Paper II in matrix form [cf. (II.2.5)]. $\psi(\lambda_2, 0)$ is determined up to an arbitrary normalization constant by spin or helicity assignments.

Recently, Weinberg[4] quantized the Poincaré components $\psi(s_1, s_2)$ corresponding to a definite mass κ and definite spin $s = \lambda_2$ for states with definite helicity,

by introducing a Hilbert space based on the physical multiplicity $2(2\lambda_2 + 1)$ of the states described by the component $\psi(s_1, s_2)$. He did not make use of field equations to evaluate $\psi(s_1, s_2)$. Our Poincaré components are proportional to Weinberg's components; the proportionality coefficients are determined from the field equations. In fact, these coefficients follow from the $O(4, 1)$ symmetry. We adopt here Weinberg's quantization scheme. We carry out the quantization in a unified way, both for helicity and observable spin states. The new features in our approach are the existence of field energy–momentum and charge operators and the existence of a uniquely defined adjoint wavefunction $\bar{\psi}$.

We introduce annihilation and creation operators $a_M(\mathbf{p})$, $a^+_M(\mathbf{p})$, and $b_M(\mathbf{p})$, $b^+_M(\mathbf{p})$ for particles and antiparticles, respectively, with definite momentum $\mathbf{p}$ and definite helicity or spin component $S_3 = M$. They satisfy the usual statistics relations:

$$[a_M(\mathbf{p}), a^+_{M'}(\mathbf{p}')]_\pm = [b_M(\mathbf{p}), b^+_{M'}(\mathbf{p}')]_\pm = \delta_{MM'}\delta^3(\mathbf{p} - \mathbf{p}'). \quad (2.23)$$

All other commutators or anticommutators vanish. Commutators are used for bosons with λ_1 and λ_2 both integral; anticommutators are used for fermions with λ_1 and λ_2 both half-integral. The quantized fields are given by

$$\psi(x) = \int d^3p \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} \sum_{M=-\lambda_2}^{\lambda_2} [a_M(\mathbf{p})\psi_M(x; \mathbf{p}, +) + b^+_{\sigma M}(-\mathbf{p})\psi_M(x; \mathbf{p}, -)],$$
$$\bar{\psi}(x) = \int d^3p \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} \sum_{M=-\lambda_2}^{\lambda_2} [a^+_M(\mathbf{p})\bar{\psi}_M(x; \mathbf{p}, +) + b_{\sigma M}(-\mathbf{p})\bar{\psi}_M(x; \mathbf{p}, -)]. \quad (2.24)$$

$\psi_M(x; \mathbf{p}, \pm)$ are the positive- and negative-energy plane-wave solutions (2.2). $\sigma M = M$ or $-M$, according to whether M stands for the eigenvalues of the observable spin S_3 or the helicity $\boldsymbol{\Sigma} \cdot \mathbf{p}/p$, respectively. We note that, as $p_\mu \to -p_\mu$, $\mathbf{S} \to \mathbf{S}$. Thus $\mathbf{S}$ is defined by (2.5) for $\epsilon = +1$; explicitly,

$$\mathbf{S} = (\lambda_1/\kappa)[E_p\boldsymbol{\sigma} + i\boldsymbol{\alpha} \wedge \mathbf{p} - (E_p + \kappa)^{-1}(\boldsymbol{\sigma} \cdot \mathbf{p})\mathbf{p}]. \quad (2.25)$$

We have used the notation (2.24) to facilitate the calculations in what follows.

If we normalize $\psi_M(x; \mathbf{p}, \epsilon)$ according to (2.13), the total charge operator (2.18) becomes

$$Q = e\int d^3p \sum_{M=-\lambda_2}^{\lambda_2} [a^+_M(\mathbf{p})a_M(\mathbf{p}) + c_- b_M(\mathbf{p})b^+_M(\mathbf{p})]. \quad (2.26)$$

The occurrence of c_-, defined by (2.14), is a manifestation of the well-known spin–statistics theorem; it is given by

$$c_- = (-1)^{2\lambda_2} = +1, \text{ for fermions}, \quad = -1, \text{ for bosons}. \quad (2.27)$$

Eliminating the vacuum expectation value $\langle 0|\, Q\, |0\rangle$, using (2.23), we get

$$Q - \langle 0|\, Q\, |0\rangle = e\int d^3p \sum_{M=-\lambda_2}^{\lambda_2} [a^+_M(\mathbf{p})a_M(\mathbf{p}) - b^+_M(\mathbf{p})b_M(\mathbf{p})]. \quad (2.28)$$

The helicity operator is given by

$$\langle \boldsymbol{\Sigma} \cdot \mathbf{p}/p \rangle = \int d^3p \sum_{M=-\lambda_1}^{\lambda_2} M[a^+_M(\mathbf{p})a_M(\mathbf{p}) - c_- b_{-M}(\mathbf{p})b^+_{-M}(\mathbf{p})] = \int d^3p \sum_{M=-\lambda_2}^{\lambda_2} M[a^+_M(\mathbf{p})a_M(\mathbf{p}) + b^+_M(\mathbf{p})b_M(\mathbf{p})]. \quad (2.29)$$

If M labels the eigenvalue of S_3 instead of the helicity, then

$$\langle S_3 \rangle = \int d^3p \sum_{M=-\lambda_2}^{\lambda_2} M[a^+_M(\mathbf{p})a_M(\mathbf{p}) + c_- b_M(\mathbf{p})b^+_M(\mathbf{p})] = \int d^3p \sum_{M=-\lambda_2}^{\lambda_2} M[a^+_M(\mathbf{p})a_M(\mathbf{p}) + b^+_M(\mathbf{p})b_M(\mathbf{p})]. \quad (2.30)$$

Similarly, using (2.19), we find the total momentum operator to be

$$\mathbf{P} = \int d^3p \sum_{M=-\lambda_2}^{\lambda_2} \mathbf{p}[a^+_M(\mathbf{p})a_M(\mathbf{p}) + b^+_M(\mathbf{p})b_M(\mathbf{p})], \quad (2.31)$$

and the Hamiltonian $H = P_0$ is given by

$$H - \langle 0|\, H\, |0\rangle = \int d^3p E_p \sum_{M=-\lambda_2}^{\lambda_2} [a^+_M(\mathbf{p})a_M(\mathbf{p}) + b^+_M(\mathbf{p})b_M(\mathbf{p})]. \quad (2.32)$$

The Hamiltonian is semi-positive-definite. Expressions (2.28)–(2.32) show that $a^+_M(\mathbf{p})$ and $b^+_M(\mathbf{p})$ are creation operators of particles and antiparticles, respectively, of definite momentum $\mathbf{p}$ and spin or helicity M. For neutral particles we take $a_M(\mathbf{p}) = b_M(\mathbf{p}) = c_M(\mathbf{p})/\sqrt{2}$, where $c_M(\mathbf{p})$ is the annihilation operator of the neutral particle.

In general, using the previous relations, we obtain Heisenberg's equations

$$\partial_\mu\psi(x) = i[\psi(x), P_\mu], \quad \partial_\mu\bar{\psi}(x) = i[\bar{\psi}(x), P_\mu], \quad (2.33)$$

and

$$[\psi(x), Q] = e\psi(x),$$
$$[\bar{\psi}(x), Q] = -e\bar{\psi}(x). \tag{2.34}$$

Also

$$\begin{aligned}[P_\mu, P_\nu] &= [Q, P_\mu] \\ &= [\langle \boldsymbol{\Sigma} \cdot \mathbf{p}/p\rangle, P_\mu] = [\langle \boldsymbol{\Sigma} \cdot \mathbf{p}/p\rangle, Q] \\ &= [\langle S_3\rangle, P_\mu] = [\langle S_3\rangle, Q] = 0. \end{aligned} \tag{2.35}$$

We note, however, that $\langle S_3\rangle$ and $\langle \boldsymbol{\Sigma} \cdot \mathbf{p}/p\rangle$ do not commute with each other, since in their definitions (2.29) and (2.30) different a's and b's are used.

3. GREEN'S FUNCTIONS

A. States with Definite Spin

The field equations (2.1) are a special case of those constructed by Pursey[5] for wavefunctions describing states of definite mass κ and definite spin $s = \lambda_2$, which are the direct sum of several Poincaré components. Actually, Pursey starts from the relation (2.6) between the wavefunction ϕ_M in the canonical basis and $w(\mathbf{p}, p_0)$. Then he derives the equations which express the supplementary conditions required to make the spin unique $s = \lambda_2$. Due to the de Sitter symmetry of our fields, the components of ϕ_M are all given, except for a normalization constant which we now fix by the condition (2.13).

The components of $w_M(\mathbf{p}, p_0)$ were evaluated in Paper I [cf. (I.5.23)]. In Le Couteur's representation, they are given by

$$\begin{aligned}&w_M(\mathbf{p}, p_0 \mid s_1 s_2 s m) \\ &\quad = \epsilon^{2s_1} a(s_1 s_2) \langle s_1 s_2 s m \mid B(\mathbf{p}, \epsilon) \mid s_1 s_2 \lambda_2 M\rangle. \end{aligned} \tag{3.1}$$

Here w_M is the eigenvalue of observable spin $S_3 w_M = M w_M$. The coupling coefficients $a(s_1 s_2)$ are given by [cf. Eq. (I.5.63)]

$$\begin{aligned}a(s_1 s_2) = Z^{-1}\zeta_{s_1 s_2}\Bigg[&(2s_1+1)(2s_2+1)\binom{2\lambda_1+2}{\lambda_1 - s_1 - s_2} \\ &\times \binom{2\lambda_1+2}{\lambda_1 - s_1 + s_2 + 1}\Bigg]^{\frac{1}{2}} \end{aligned} \tag{3.2a}$$

Here

$$\begin{aligned}\zeta_{s_1 s_2} &= 1, \qquad \text{for} \quad s_1 \geq s_2, \\ &= (-1)^{[\lambda_2]}, \quad \text{for} \quad s_1 < s_2, \end{aligned} \tag{3.2b}$$

where $[\lambda_2]$ is the integral part of λ_2 and Z is a normalization constant given by

$$\begin{aligned}Z^2 = {}&[2^{2\lambda_1}(2\lambda_1+2)!\,(\lambda_1 - \lambda_2 + 1)!\,(\lambda_1 + \lambda_2 + 2)!] \\ &\times [2(\lambda_1 - \lambda_2 + 1)(\lambda_1 + \lambda_2 + 2)^2 \\ &\times F(-\lambda_1 - \lambda_2, 1; \lambda_1 - \lambda_2 + 1; -1) \\ &- (\lambda_1 + 1)(2\lambda_1 + 1)(\lambda_1 + \lambda_2 + 3) \\ &\times F(-\lambda_1 - \lambda_2 - 1, 1; \lambda_1 - \lambda_2 + 2; -1)]. \end{aligned} \tag{3.2c}$$

$F(a, b; c; x)$ is a hypergeometric series.

Further, recalling that $\gamma_0 \phi_M(\epsilon) = \epsilon \phi_M(\epsilon)$, we get

$$\bar{w}_M(\mathbf{p}, p_0) = \bar{\phi}_M B^{-1} = \beta_\epsilon \phi_M^+ B^{-1}(\mathbf{p}, \epsilon), \tag{3.3}$$

where $\beta_\epsilon = 1$ and ϵ for bosons and fermions, respectively. Thus the components of $\bar{w}_M$ are

$$\begin{aligned}&\bar{w}_M(\mathbf{p}, p_0 \mid s_1 s_2 s m) \\ &\quad = \beta_\epsilon \epsilon^{2s_1} a(s_1 s_2) \langle s_1 s_2 \lambda_2 M \mid B^{-1}(\mathbf{p}, \epsilon) \mid s_1 s_2 s m\rangle. \end{aligned} \tag{3.4}$$

The components of w_M in Bhabha's direct-product representation, $O(4) \approx O(3) \times O(3)$, are related to those in Le Couteur's representation by the usual Clebsch–Gordan decomposition:

$$\begin{aligned}w_M(s_1 s_2 s m) &= \sum_{m_1 m_2} c(s_1 s_2 s,\, m_1 m_2 m) w_M(s_1 m_1,\, s_2 m_2), \\ w_M(s_1 m_1,\, s_2 m_2) &= \sum_s c(s_1 s_2 s,\, m_1 m_2 m) w_M(s_1 s_2 s m). \end{aligned} \tag{3.5}$$

Recalling that

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}^{(1)} + \boldsymbol{\Sigma}^{(2)}, \\ \lambda_1 \boldsymbol{\alpha} &= \boldsymbol{\Sigma}^{(1)} - \boldsymbol{\Sigma}^{(2)}, \end{aligned} \tag{3.6}$$

where $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$ are the infinitesimal generators of the two rotation subgroups of $O(4) \approx O(3) \times O(3)$, the boost is written as

$$B(\mathbf{p}, \epsilon) = L^{(1)}(\mathbf{p}, \epsilon) \times L^{(2)}(-\mathbf{p}, \epsilon), \tag{3.7}$$

where

$$L^{(i)}(\mathbf{p}, \epsilon) = \exp(\theta_1 \boldsymbol{\Sigma}^{(i)} \cdot \mathbf{p}/p). \tag{3.8}$$

θ_1 is given by (2.8). $L(\mathbf{p}, \epsilon)$ and $L(-\mathbf{p}, \epsilon)$ are the boosts in the representation $D(s_1, 0)$ and $D(0, s_2)$, respectively.

The matrix elements of $B(\mathbf{p}, \epsilon)$ is Bhabha's, and Le Couteur's representations are related as follows:

$$\begin{aligned}&\langle s_1 s_2 s m \mid B(\mathbf{p}, \epsilon) \mid s_1 s_2 \lambda_2 M\rangle \\ &\quad = \sum_{m_1 m_2} \sum_{\mu_1 \mu_2} c(s_1 s_2 s,\, m_1 m_2 m) c(s_1 s_2 \lambda_2,\, \mu_1 \mu_2 M) \\ &\qquad \times \langle s_1 m_1,\, s_2 m_2 \mid B(\mathbf{p}, \epsilon) \mid s_1 \mu_1,\, s_2 \mu_2\rangle \\ &\quad = \sum_{m_1 m_2} \sum_{\mu_1 \mu_2} c(s_1 s_2 s,\, m_1 m_2 m) c(s_1 s_2 \lambda_2,\, \mu_1 \mu_2 M) \\ &\qquad \times L^{s_1}_{m_1 \mu_1}(\mathbf{p}, \epsilon) L^{s_2}_{m_2 \mu_2}(-\mathbf{p}, \epsilon), \end{aligned} \tag{3.9}$$

$$\begin{aligned}&\langle s_1 s_2 \lambda_2 M \mid B^{-1}(\mathbf{p}, \epsilon) \mid s_1 s_2 s m\rangle \\ &\quad = \sum_{m_1 m_2} \sum_{\mu_1 \mu_2} c(s_1 s_2 \lambda_2,\, \mu_1 \mu_2 M) c(s_1 s_2 s,\, m_1 m_2 m) \\ &\qquad \times L^{s_1}_{\mu_1 m_1}(-\mathbf{p}, \epsilon) L^{s_2}_{\mu_2 m_2}(\mathbf{p}, \epsilon). \end{aligned} \tag{3.10}$$

Hence

$$\begin{aligned}&w_M(\mathbf{p}, p_0 \mid s_1 m_1,\, s_2 m_2) \\ &\quad = \epsilon^{2s_1} a(s_1 s_2) \sum_{\mu_1 \mu_2} c(s_1 s_2 \lambda_2,\, \mu_1 \mu_2 M) \\ &\qquad \times L^{s_1}_{m_1 \mu_1}(\mathbf{p}, \epsilon) L^{s_2}_{m_2 \mu_2}(-\mathbf{p}, \epsilon), \end{aligned} \tag{3.11}$$

and

$$\bar{w}_M(\mathbf{p}, p_0 \mid s_1m_1, s_2m_2) = \epsilon^{2s_1}\beta_\epsilon a(s_1s_2)\sum_{\mu_1\mu_2} c(s_1s_2\lambda_2, \mu_1\mu_2M) \times L^{s_1}_{\mu_1m_1}(-\mathbf{p}, \epsilon)L^{s_2}_{\mu_2m_2}(\mathbf{p}, \epsilon). \quad (3.12)$$

Denoting the positive- and negative-energy solutions with definite observable spin by

$$U_M(\mathbf{p}) = w_M(\mathbf{p}, E_p), \quad V_M(\mathbf{p}) = w_M(-\mathbf{p}, -E_p), \quad (3.13)$$

and noting that

$$B(\mathbf{p}, E_p) = B(-\mathbf{p}, -E_p), \quad (3.14)$$

$$\mathbf{S}(\mathbf{p}, E_p) = \mathbf{S}(-\mathbf{p}, -E_p), \quad (3.15)$$

we find that

$$S_3U_M(\mathbf{p}) = MU_M(\mathbf{p}), \quad S_3V_M(\mathbf{p}) = MV_M(\mathbf{p}), \quad (3.16)$$

and

$$V_M(\mathbf{p}; s_1m_1, s_2m_2) = (-1)^{2s_1}U_M(\mathbf{p}; s_1m_1, s_2m_2), \quad (3.17)$$

$$\bar{V}_M(\mathbf{p}; s_1m_1, s_2m_2) = \beta_-(-1)^{2s_1}\bar{U}_M(\mathbf{p}; s_1m_1, s_2m_2), \quad (3.18)$$

$$U_M(\mathbf{p}; s_1m_1, s_2m_2) = a(s_1s_2)F^{-1(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p}), \quad (3.19)$$

$$\bar{U}_M(\mathbf{p}; s_1m_1, s_2m_2) = a(s_1s_2)F^{(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p}). \quad (3.20)$$

Here $\beta_- = (-1)^{2\lambda_2} = 1$ and -1 for bosons and fermions, respectively. F^{-1} and F are mixed matrix elements of the boosts $B(\mathbf{p}, E_p)$ and $B^{-1}(\mathbf{p}, E_p)$, respectively, between Bhabha's and Le Couteur's bases:

$$F^{-1(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p}) = \langle s_1m_1, s_2m_2 \mid B \mid s_1s_2\lambda_2M\rangle = \sum_{\mu_1\mu_2} c(s_1s_2\lambda_2, \mu_1\mu_2M)L^{s_1}_{m_1\mu_1}(\mathbf{p})L^{s_2}_{m_2\mu_2}(-\mathbf{p}), \quad (3.21)$$

$$F^{(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p}) = \langle s_1s_2\lambda_2M \mid B^{-1} \mid s_1m_1, s_2m_2\rangle = \sum_{\mu_1\mu_2} c(s_1s_2\lambda_2, \mu_1\mu_2M)L^{s_1}_{\mu_1m_1}(-\mathbf{p})L^{s_2}_{\mu_2m_2}(\mathbf{p}). \quad (3.22)$$

The coefficients F and F^{-1} were introduced by Pursey.[5] There is a minor difference of notation; what we call here $L(\mathbf{p})$, he calls $L^{-1}(\mathbf{p})$ and $L(-\mathbf{p})$. However, expressions (3.21) and (3.22) are identical to his. Pursey studied the properties of the F-coefficients. He proved the following useful identity:

$$\sum_M F^{-1(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p})F^{(s_1's_2'\lambda_2)}_{m_1'm_2'M}(\mathbf{p}) = (2\lambda_2 + 1)\sum_{smm'}(-1)^{s_1+s_2'+s+\lambda_2} \times \left(\frac{(2s+1)^3}{(2s_1+1)(2s_2+1)}\right)^{\frac{1}{2}}\begin{Bmatrix} s_1' & s_2' & \lambda_2 \\ s_2 & s_1 & s \end{Bmatrix} \times c(ss_1's_1, mm_1'm_1)c(ss_2's_2, m'm_2'm_2) \times \kappa^{-2s}p^{(ss)}_{mm'}(\mathbf{p}, E_p), \quad (3.23)$$

where $\begin{Bmatrix} s_1's_2'\lambda_2 \\ s_2s_1s \end{Bmatrix}$ is a 6-j symbol and $p^{(ss)}_{mm'}(\mathbf{p}, E_p)$ is the mass-shell $[p_0 = E_p = (p^2 + \kappa^2)^{\frac{1}{2}}]$ value of the standard traceless symmetric tensor $p^{(ss)}_{mm'}$ constructed from the 4-vector p_μ, which transforms irreducibly under the Lorentz group according to $D(s, s)$. McKerrel[13] has shown that

$$p^{(ss)}_{mm'}(\mathbf{p}, E_p) = \kappa^{2s}F^{-1(ss0)}_{mm'0}(\mathbf{p}) = \sum_\mu c(ss_0, \mu -\mu\, 0)\langle sm, sm' \mid D \mid s\mu, s -\mu\rangle \times (E_p + p)^{s-\mu}(E_p - p)^{s+\mu} = \kappa^{2s}\sum_\mu (-1)^{s-\mu}(2s+1)^{-\frac{1}{2}}D^s_{m\mu}(\varphi, \theta, \gamma) \times D^s_{m',-\mu}(\varphi, \theta, \gamma)\left(\frac{E_p + p}{\kappa}\right)^{-2\mu} \quad (3.24)$$

D is the rotation $D(\varphi, \theta, \gamma)$ defined in Sec. 4A which takes the momentum $(0, 0, p)$ to $\mathbf{p}$. In the last line of (3.24) we have made use of the relation

$$c(ss0, \mu -\mu\, 0) = (-1)^{s-\mu}(2s+1)^{-\frac{1}{2}}.$$

B. Off-the-Mass-Shell Relations

The off-the-mass-shell tensor is given by

$$p^{(ss)}_{mm'} = (2s+1)^{-\frac{1}{2}}\sum_\mu (-1)^{s-\mu}D^s_{m\mu}(\varphi, \theta, \gamma) \times D^s_{m',-\mu}(\varphi, \theta, \gamma)(p_0 + p)^{s-\mu}(p_0 - p)^{s+\mu} = (2s+1)^{-\frac{1}{2}}e^{-i(m+m')\varphi}\sum_\mu (-1)^{s-\mu}d^s_{m\mu}(\theta)d^s_{m',-\mu}(\theta) \times (p_0 + p)^{s-\mu}(p_0 - p)^{s+\mu}, \quad (3.25)$$

where (θ, φ) are the polar angles of the momentum vector

$$p_1 = p\sin\theta\cos\varphi, \quad p_2 = p\sin\theta\sin\varphi, \quad p_3 = p\cos\theta$$

and $d^s_{m\mu}(\theta) = \langle sm \mid e^{-i\theta\Sigma_2} \mid s\mu\rangle$ is the usual rotation matrix. We note that $p^{(ss)}_{mm'}$ is independent of Euler's angle γ, which is an arbitrary rotation about the direction of $\mathbf{p}$.

To compare with Weinberg's[4] notation, we note that his tensor $\Pi^{(s)}_{mm'}(p)$ is related to $p^{(ss)}_{mm'}$ as follows:

$$\Pi^{(s)}_{mm'}(p) = (-1)^{s-m'}(2s+1)^{\frac{1}{2}}p^{(ss)}_{m,-m'} = \kappa^{2s}\langle sm \mid \exp(-2\theta_1\boldsymbol{\Sigma}\cdot\mathbf{p}/p) \mid sm'\rangle = (-1)^{2s}t^{\mu_1\cdots\mu_{2s}}_{mm'}p_{\mu_1}\cdots p_{\mu_{2s}}. \quad (3.26)$$

The $\Pi^{(s)}_{mm'}(p)$ are the off-the-mass-shell matrix elements of $[L(-\mathbf{p})]^2$ in the representation $D(s)$ of the 3-dimensional rotation group, and t is a traceless symmetric constant tensor defined by Weinberg[7] and Shaw.[8] In Appendix B we evaluate this tensor using a simple relation obtained in Paper II.

Using the symmetry properties of the Clebsch–Gordan coefficients,[14] we easily verify that the expression (3.23) is related to Weinberg's[4] analogous expression as follows:

$$\sum_M F^{-1(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p})F^{(s_1's_2'\lambda_2)}_{m_1'm_2'M}(\mathbf{p}) = (-1)^{s_1'-s_2'-m_1'-m_2'}\pi^{(s_1's_2',s_1s_2)}_{-m_1'-m_2',m_1m_2}(\mathbf{p}/\kappa;\lambda_2), \quad (3.27)$$

where

$$\pi^{(s_1's_2',s_1s_2)}_{m_1'm_2',m_1m_2}(\mathbf{p}/(-p_\mu p_\mu)^{\frac{1}{2}};j) = (2s+1)\sum_{sv'v''} W(s_1's_2's_1s_2;js)(-1)^{s_1'-s_2-j+v''}(-p_\mu p_\mu)^{-s} \times c(s_1's_1s, m_1'm_1v')c(s_2's_2s, m_2'm_2v'')\Pi^{(s)}_{v',-v''}(p). \quad (3.28)$$

Here $W(s_1's_2's_1s_1;js)$ is a Racah coefficient.

The off-the-mass-shell expression for the relation (3.23) is obtained by substituting $p^{(ss)}_{mm'}(\mathbf{p}, E_p)$ by its off-the-mass-shell expression (3.25). We denote the resulting expression by $\langle s_1m_1, s_2m_2|\, G(p)\,|s_1'm_1', s_2'm_2'\rangle$ and consider it as a matrix element of a matrix $G(p)$. It is related to the off-the-mass-shell expression of Weinberg in the following way:

$$\langle s_1m_1, s_2m_2|\, G(p)\,|s_1'm_1', s_2'm_2'\rangle = (-1)^{s_1'-s_2'-m_1'-m_2'}N^{(s_1's_2',s_1s_2)}_{-m_1'-m_2',m_1m_2}(p;\lambda_2,\kappa). \quad (3.29)$$

Its mass-shell value is (3.27). Weinberg[4] has shown that

$$N^{(s_1's_2',s_1s_2)}_{m_1'm_2',m_1m_2}(p;j,\kappa) = (2j+1)\sum_{sv'v''} W(s_1's_2's_1s_2;js)(-1)^{s_1'-s_2-j+v''}\kappa^{-2s} \times c(s_1's_1s, m_1'm_1v')c(s_2's_2s, m_2'm_2v'')\Pi^{(s)}_{v',-v''}(p) = \sum_{j'} F_{jj'}\left(\frac{-p_\mu p_\mu}{\kappa^2}; s_1's_2's_1s_2\right) \times \pi^{(s_1's_2',s_1s_2)}_{m_1'm_2',m_1m_2}(\mathbf{p}/(-p_\mu p_\mu)^{\frac{1}{2}};j'), \quad (3.30)$$

where

$$F_{jj'}(x, s_1's_2's_1s_2) = (-1)^{j'-j}(2j+1)\sum_s (2s+1)x^s \times W(s_1's_2's_1s_2;js)W(s_1's_2's_1s_2;j's). \quad (3.31)$$

As we shall see, N defines the Green's function. The different matrix elements correspond to all $\lambda_1 \geq s_1 + s_2 \geq j' \geq |s_1 - s_2| \geq \delta$, where $\delta = 0$ and $\frac{1}{2}$ for bosons and fermions, respectively. Thus, in the expansions (3.30), various spins j' contribute with $\delta \leq j' \leq \lambda_1$.

C. The Green's Function

The quantized field (2.24) can be written in the form

$$\psi(x) = (2\pi)^{-\frac{3}{2}}\int\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M [a_M(\mathbf{p})U_M(\mathbf{p})e^{ipx} + b^+_M(\mathbf{p})V_M(\mathbf{p})e^{-ipx}],$$
$$\bar\psi(x) = (2\pi)^{-\frac{3}{2}}\int\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M [b_M(\mathbf{p})\bar V_M(\mathbf{p})e^{ipx} + a^+_M(\mathbf{p})\bar U_M(\mathbf{p})e^{-ipx}], \quad (3.32)$$

where $px = \mathbf{p}\cdot\mathbf{r} - E_pt$ and the U and V are defined by (3.13). Using (3.17)–(3.20) and denoting the components of $\psi(x)$ in the Bhabha representation by $\psi^{s_1s_2}_{m_1m_2}(x)$, in accordance with Weinberg's notation, we get

$$\psi^{s_1s_2}_{m_1m_2}(x) = (2\pi)^{-\frac{3}{2}}a(s_1s_2)\int\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M F^{-1(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p}) \times [a_M(\mathbf{p})e^{ipx} + (-1)^{2s_1}b^+_M(\mathbf{p})e^{-ipx}],$$
$$\bar\psi^{s_1s_2}_{m_1m_2}(x) = (2\pi)^{-\frac{3}{2}}a(s_1s_2)\int\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M F^{(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p}) \times [a^+_M(\mathbf{p})e^{-ipx} + (-1)^{2\lambda_2-2s_1}b_M(\mathbf{p})e^{ipx}]. \quad (3.33)$$

Using the commutation relations (2.23), we get

$$[\psi^{s_1s_2}_{m_1m_2}(x), \psi^{s_1's_2'}_{m_1'm_2'}(y)]_\pm = [\bar\psi^{s_1s_2}_{m_1m_2}(x), \bar\psi^{s_1's_2'}_{m_1'm_2'}(y)]_\pm = 0. \quad (3.34)$$

Also,

$$[\psi^{s_1s_2}_{m_1m_2}(x), \bar\psi^{s_1's_2'}_{m_1'm_2'}(y)]_\pm = \kappa(2\pi)^{-3}\int\frac{d^3p}{E_p}\sum_M F^{-1(s_1s_2\lambda_2)}_{m_1m_2M}(\mathbf{p})F^{(s_1's_2'\lambda_2)}_{m_1'm_2'M}(\mathbf{p}) \times (e^{ip(x-y)} - (-1)^{2s_1+2s_1'}e^{-ip(x-y)}).$$

Now, using the symmetry property

$$N^{(s_1's_2',s_1s_2)}_{m_1'm_2',m_1m_2}(-p;\lambda_2,\kappa) = (-1)^{2s_1+2s_1'}N^{(s_1's_2',s_1s_2)}_{m_1'm_2',m_1m_2}(p;\lambda_2,\kappa), \quad (3.35)$$

we get

$$[\psi^{s_1s_2}_{m_1m_2}(x), \bar\psi^{s_1's_2'}_{m_1'm_2'}(y)]_\pm = 2i\kappa a(s_1s_2)a(s_1's_2') \times \langle s_1m_1, s_2m_2|\, G(-i\partial)\,|s_1'm_1', s_2'm_2'\rangle\,\Delta(x-y). \quad (3.36)$$

Here

$$\Delta(x) = -i(2\pi)^{-3}\int d^4p\,\epsilon(p_0)\delta(p_\mu p_\mu + \kappa^2)e^{ipx}, \quad (3.37)$$

and $G(-i\partial)$ implies the substitution $p_\mu = -i\partial/\partial x_\mu$ in the off-the-mass-shell matrix (3.29). The Wightman

two-point functions are

$$\langle \psi^{s_1 s_2}_{m_1 m_2}(x)\bar{\psi}^{s_1' s_2'}_{m_1' m_2'}(y)\rangle_0 = \langle \psi^{(+)s_1 s_2}_{m_1 m_2}(x)\bar{\psi}^{(-)s_1' s_2'}_{m_1' m_2'}(y)\rangle_0 = 2i\kappa a(s_1 s_2)a(s_1' s_2') \times \langle s_1 m_1, s_2 m_2|\, G(-i\partial)\,|s_1' m_1', s_2' m_2'\rangle\, \Delta_+(x-y), \tag{3.38}$$

$$\langle \bar{\psi}^{s_1' s_2'}_{m_1' m_2'}(y)\psi^{s_1 s_2}_{m_1 m_2}(x)\rangle_0 = \langle \bar{\psi}^{(+)s_1' s_2'}_{m_1' m_2'}(y)\psi^{(-)s_1 s_2}_{m_1 m_2}(x)\rangle_0 = (-1)^{2\lambda_2} 2i\kappa a(s_1 s_2)a(s_1' s_2') \times \langle s_1 m_1, s_2 m_2|\, G(-i\partial)\,|s_1' m_1', s_2' m_2'\rangle\, \Delta_-(x-y). \tag{3.39}$$

Here $\psi^{(+)}$ and $\psi^{(-)}$ are the positive- and negative-frequency parts of ψ, and

$$\Delta_+(x) = \Delta_-^*(x) = \frac{-i}{2(2\pi)^3}\int \frac{d^3p}{E_p}\, e^{ipx}.$$

The Green's function

$$\langle T\{\psi^{s_1 s_2}_{m_1 m_2}(x)\bar{\psi}^{s_1' s_2'}_{m_1' m_2'}(y)\}\rangle_0 = \langle \theta(x_0-y_0)\psi^{(+)s_1 s_2}_{m_1 m_2}(x)\bar{\psi}^{(-)s_1' s_2'}_{m_1' m_2'}(y) + (-1)^{2\lambda_2}\theta(y_0-x_0)\bar{\psi}^{(+)s_1' s_2'}_{m_1' m_2'}(y)\psi^{(-)s_1 s_2}_{m_1 m_2}(x)\rangle_0 = 2i\kappa a(s_1 s_2)a(s_1' s_2')[\theta(x_0-y_0) \times \langle s_1 m_1, s_2 m_2|\, G(-i\partial)\,|s_1' m_1', s_2' m_2'\rangle \times \Delta_+(x-y) - \theta(y_0-x_0) \times \langle s_1 m_1, s_2 m_2|\, G(-i\partial)\,|s_1' m_1', s_2' m_2'\rangle\, \Delta_-(x-y)]. \tag{3.40}$$

For convenience, we introduce the matrices $S^{(i)}(x-y)$ with matrix elements

$$S^{(i)}_{\rho\rho'}(x-y) = d_{\rho\rho'}(-i\partial)\Delta_i(x-y) \tag{3.41}$$

in any representation. The index i specifies the type of $\Delta_i(x)$ involved; for example, in the Bhabha representation, $\rho = (s_1 m_1, s_2 m_2)$ and

$$\langle s_1 m_1, s_2 m_2|\, d(-i\partial)\,|s_1' m_1', s_2' m_2'\rangle = 2\kappa a(s_1 s_2)a(s_1' s_2') \times \langle s_1 m_1, s_2 m_2|\, G(-i\partial)\,|s_1' m_1', s_2' m_2'\rangle. \tag{3.42}$$

Then, we write, in general,

$$[\psi_\rho(x), \psi_{\rho'}(y)]_\pm = [\bar{\psi}_\rho(x), \bar{\psi}_{\rho'}(y)]_\pm = 0, \quad [\psi_\rho(x), \bar{\psi}_{\rho'}(y)]_\pm = iS_{\rho\rho'}(x-y), \tag{3.43}$$

$$\langle T\{\psi_\rho(x)\bar{\psi}_{\rho'}(y)\}\rangle_0 = i[\theta(x_0-y_0)S^{(+)}_{\rho\rho'}(x-y) - \theta(y_0-x_0)S^{(-)}_{\rho\rho'}(x-y)]. \tag{3.44}$$

The time-ordered product is not, in general, Lorentz covariant. A Lorentz-covariant time-ordered product is obtained by interchanging $\theta(\pm(x_0-y_0))$ and $d(-i\partial)$ in (3.44), as suggested by Weinberg.[4] We get

$$\langle T^*\{\psi_\rho(x)\bar{\psi}_{\rho'}(y)\}\rangle = iS^c_{\rho\rho'}(x-y), \tag{3.45}$$

where $S^c(x-y)$ is the causal Green's function obtained by using

$$\Delta_c(x) = \theta(x)\Delta_+(x) - \theta(-x)\Delta_-(x)$$

in (3.41). The covariant propagator in momentum space is

$$\Delta_{\rho\rho'}(p) = i\int d^4x e^{-ip(x-y)}S^c_{\rho\rho'}(x-y) = \frac{d_{\rho\rho'}(p)}{(p_\mu p_\mu + \kappa^2 - i\epsilon)}. \tag{3.46}$$

However, the definition of the causal Green's function is not unique and is very much dependent on the type of interaction. As is well known,[6] all functions of the form

$$G'(-i\partial) = G(-i\partial) + g(-i\partial)(\Box - \kappa^2),$$

with arbitrary $g(-i\partial)$, lead to the same relation (3.40), but they lead to different causal Green's functions.

4. HELICITY STATES

A. The Rotation Matrix

We prove in what follows that, if ψ_M is the solution of the field equations with definite helicity for an arbitrary direction of the momentum and ψ^0_M is the corresponding state if the momentum is parallel to the z axis, then $\psi_M = D\psi^0_M$, where D is the rotation matrix taking the momentum $(0, 0, p)$ into $\mathbf{p}$. We restrict ourselves to the momentum space.

In general, ψ_M satisfies the following equations:

$$P_k\psi_M = p_k\psi_M, \tag{4.1}$$

$$H\psi_M = p_0\psi_M, \tag{4.2}$$

$$S\psi_M = M\psi_M, \tag{4.3}$$

where

$$\mathbf{P} = \boldsymbol{\alpha}p_0 + i\boldsymbol{\sigma}\wedge\mathbf{p} + \kappa\boldsymbol{\gamma} \tag{4.4}$$

is the field momentum operator,

$$H = \boldsymbol{\alpha}\cdot\mathbf{p} + \kappa\gamma_0 \tag{4.5}$$

is the Hamiltonian, and

$$S = \boldsymbol{\Sigma}\cdot\mathbf{p}/p \tag{4.6}$$

is the helicity.

On the other hand, if the momentum is in the z direction, $p_1 = p_2 = 0$, $p_3 = p$, then ψ^0_M satisfies the equations

$$P^0_1\psi^0_M = P^0_2\psi^0_M = 0, \quad P^0_3\psi^0_M = p\psi^0_M, \tag{4.7}$$

$$H^0\psi^0_M = p_0\psi^0_M, \tag{4.8}$$

$$S^0\psi^0_M = \Sigma_3\psi^0_M = M\psi^0_M, \tag{4.9}$$

where

$$P_1^0 = \alpha_1 p_0 + i\sigma_2 p + \kappa\gamma_1,$$
$$P_2^0 = \alpha_2 p_0 - i\sigma_1 p + \kappa\gamma_2,$$
$$P_3^0 = \alpha_3 p_0 + \kappa\gamma_3, \tag{4.10}$$
$$H^0 = \alpha_3 p + \kappa\gamma_0, \tag{4.11}$$
$$S^0 = \Sigma_3. \tag{4.12}$$

We now consider an arbitrary 3-dimensional rotation

$$x_k = a_{kl}x_l^0,$$
$$a_{ki}a_{li} = a_{ik}a_{il} = \delta_{kl}. \tag{4.13}$$

The corresponding rotation matrix D of the space carrying the representation $R_5(\lambda_1, \lambda_2)$ is determined from the relations

$$D^{-1}\Sigma_k D = a_{kl}\Sigma_l,$$
$$D^{-1}\alpha_k D = a_{kl}\alpha_l, \tag{4.14}$$
$$D^{-1}\gamma_k D = a_{kl}\gamma_l,$$

and

$$D\gamma_0 = \gamma_0 D. \tag{4.15}$$

Due to the orthogonality of the transformation (4.13), $(a^{-1})_{kl} = (a^{\mathrm{T}})_{kl} = a_{lk}$,and the group property

$$D(a^{-1}) = D(a^{\mathrm{T}}) = D^{-1}(a),$$

the inverse relations hold:

$$D\Sigma_k D^{-1} = a_{lk}\Sigma_l,$$
$$D\alpha_k D^{-1} = a_{lk}\alpha_l, \tag{4.16}$$
$$D\gamma_k D^{-1} = a_{lk}\gamma_l.$$

Now, if

$$p_k = a_{kl}p_l^0, \tag{4.17}$$

where $p_k^0 = (0, 0, p)$, then [cf. Eq. (I.2.56)]

$$D^{-1}P_k D = a_{kl}P_l^0, \tag{4.18}$$

$$D^{-1}HD = H^0. \tag{4.19}$$

Thus, if

$$\psi_M = D\psi_M^0, \tag{4.20}$$

where ψ_M^0 satisfies Eqs. (4.7) and (4.8), then ψ_M will satisfy Eqs. (4.1) and (4.2).

We note, however, that (4.17) and (4.13) do not determine a_{kl} completely. In fact, using $p_1^0 = p_2^0 = 0$ and $p_3^0 = p$, we get

$$a_{k3} = p_k/p. \tag{4.21}$$

This determines only two of the three parameters of $O(3)$. However, (4.21) suffices to ensure that (4.20) is a state of definite helicity. Using (4.16) and (4.21), we get

$$D\Sigma_3 D^{-1} = a_{k3}\Sigma_k = \boldsymbol{\Sigma}\cdot\mathbf{p}/p = S. \tag{4.22}$$

Thus

$$S\psi_M = S\Sigma_3 D^{-1}\psi_M = D\Sigma_3\psi_M^0 = M\psi_M. \tag{4.23}$$

This completes the proof that $\psi_M = D\psi_M^0$ is a solution of the field equations with definite helicity.

Since D is a 3-dimensional rotation, it can be written in the usual form

$$D(\varphi, \theta, \gamma) = e^{-i\varphi\Sigma_3}e^{-i\theta\Sigma_2}e^{-i\gamma\Sigma_3}, \tag{4.24}$$

where φ, θ, and γ are Euler angles which are determined in terms of a_{kl}. From (4.15), (4.19), and (4.22) and the definitions of H and H^0, it is required that

$$D\alpha_3 D^{-1} = \boldsymbol{\alpha}\cdot\mathbf{p}/p, \tag{4.25}$$

$$D\Sigma_3 D^{-1} = \boldsymbol{\Sigma}\cdot\mathbf{p}/p. \tag{4.26}$$

Using the commutation relations (I.2.11), both requirements lead to

$$p_1 = p\sin\theta\cos\varphi,$$
$$p_2 = p\sin\theta\sin\varphi, \tag{4.27}$$
$$p_3 = p\cos\theta,$$

while γ remains arbitrary. The two Euler angles θ and φ are just the polar angles of the momentum vector $\mathbf{p}$. Any choice of γ determines the rest of a_{kl} in such a way that Eqs. (4.18) and, hence, also Eqs. (4.1) are automatically satisfied. Applying (4.24) to a helicity state, we get

$$\psi_M = D\psi_M^0 = e^{-iM\gamma}(e^{-i\varphi\Sigma_3}e^{-i\theta\Sigma_2}\psi_M^0). \tag{4.28}$$

The arbitrariness of γ results in a phase factor $e^{-iM\gamma}$, which can be absorbed in the definition of ψ_M. It is of no physical consequence. In fact, γ is a rotation about the direction $\mathbf{p}$, which is irrelevant for our purpose.

B. The Helicity States

If the momentum is in the z direction, the helicity and the z component of the observable spin coincide:

$$S_3 = \boldsymbol{\Sigma}\cdot\mathbf{p}/p = \Sigma_3 = \Sigma_3^{(1)} + \Sigma_3^{(2)}, \tag{4.29}$$

where $\Sigma_3^{(1)}$ and $\Sigma_3^{(2)}$ are defined by (3.6). Noting that m_1 and m_2 are the eigenvalues of $\Sigma_3^{(1)}$ and $\Sigma_3^{(2)}$, respectively, in the Bhabha representation, we see that the matrix elements of the boosts (3.8) reduce, for $p_1 = p_2 = 0$ and $p_3 = p$, to

$$L^{s_1}_{m_1\mu_1}(\mathbf{p}, \epsilon) = \delta_{m_1\mu_1}e^{m_1\theta_1} = \delta_{m_1\mu_1}(E_p + \epsilon p/\kappa)^{m_1},$$
$$L^{s_2}_{m_2\mu_2}(-\mathbf{p}, \epsilon) = \delta_{m_2\mu_2}(E_p + \epsilon p/\kappa)^{-m_2}. \tag{4.30}$$

w_M and $\bar{w}_M$, given by (3.11) and (3.12), reduce to

$$w^0_M(\epsilon \mid s_1 m_1, s_2 m_2) = \epsilon^{2s_1} a(s_1 s_2)(E_p + p/\kappa)^{\epsilon(m_1-m_2)} c(s_1 s_2 \lambda_2, m_1 m_2 M), \tag{4.31}$$

$$\bar{w}^0_M(\epsilon \mid s_1 m_1, s_2 m_2) = \beta_\epsilon \epsilon^{2s_1} a(s_1 s_2)(E_p - p/\kappa)^{\epsilon(m_1-m_2)} c(s_1 s_2 \lambda_2, m_1 m_2 M). \tag{4.32}$$

Relation (4.31) was obtained in Part II [Eq. (II.2.19)] except for an over-all proportionality factor which is fixed here by the normalization condition (2.13).

In Le Couteur's representation

$$w^0_M(\epsilon \mid s_1 s_2 s m) = \delta_{mM} \epsilon^{2s_1} a(s_1 s_2) \sum_{m_1 m_2} c(s_1 s_2 s, m_1 m_2 M) \times c(s_1 s_2 \lambda_2, m_1 m_2 M) \left(\frac{E_p + p}{\kappa}\right)^{\epsilon(m_1-m_2)}, \tag{4.33}$$

$$\bar{w}^0_M(\epsilon \mid s_1 s_2 s m) = \delta_{mM} \beta_\epsilon \epsilon^{2s_1} a(s_1 s_2) \sum_{m_1 m_2} c(s_1 s_2 s, m_1 m_2 M) \times c(s_1 s_2 \lambda_2, m_1 m_2 M) \left(\frac{E_p - p}{\kappa}\right)^{\epsilon(m_1-m_2)}. \tag{4.34}$$

Although helicity and the observable spin component S_3 coincide in the frame $\mathbf{p}^0 = (0, 0, p)$, they differ in any arbitrary frame $\mathbf{p}$. The state w_M of definite spin S_3 and that of definite helicity, which we denote by $\chi_M(\mathbf{p}, p_0)$, do not coincide. From the preceding section we have

$$\chi_M(\mathbf{p}, p_0) = D(\varphi, \theta, \gamma) w^0_M(\epsilon). \tag{4.35}$$

From the unitarity of $D: D^+ = D^{-1}$ we get

$$\bar{\chi}_M(\mathbf{p}, p_0) = \bar{w}^0_M(\epsilon) D^+(\varphi, \theta, \gamma). \tag{4.36}$$

Thus

$$\chi_M(\mathbf{p}, p_0 \mid s_1 s_2 s m) = D^s_{mM}(\varphi, \theta, \gamma) w^0_M(\epsilon \mid s_1 s_2 s M) = e^{-iM\gamma - im\varphi} d^s_{mM}(\theta) \epsilon^{2s_1} a(s_1 s_2) \times \sum_{m_1 m_2} c(s_1 s_2 s, m_1 m_2 M) \times c(s_1 s_2 \lambda_2, m_1 m_2 M) \left(\frac{E_p + p}{\kappa}\right)^{\epsilon(m_1-m_2)}, \tag{4.37}$$

$$\bar{\chi}_M(\mathbf{p}, p_0 \mid s_1 s_2 s m) = \bar{w}^0_M(\epsilon \mid s_1 s_2 s M) D^s_{mM}(\varphi, \theta, \gamma)^* = e^{iM\gamma + im\varphi} d^s_{mM}(\theta) \beta_\epsilon \epsilon^{2s_1} a(s_1 s_2) \times \sum_{m_1 m_2} c(s_1 s_2 s, m_1 m_2 M) \times c(s_1 s_2 \lambda_2, m_1 m_2 M) \left(\frac{E_p - p}{\kappa}\right)^{\epsilon(m_1-m_2)}. \tag{4.38}$$

Using (3.6), we write $D = D^{(1)} D^{(2)}$, where $D^{(i)}$ is defined in terms of the corresponding $\Sigma^{(i)}_k$. Then, in the Bhabha representation, we get

$$\chi_M(\epsilon \mid s_1 m_1, s_2 m_2) = e^{-iM\gamma - i(m_1+m_2)\varphi} \epsilon^{2s_1} a(s_1 s_2) \times \sum_{\mu_1 \mu_2} d^s_{m_1\mu_1}(\theta) d^{s_2}_{m_2\mu_2}(\theta) \left(\frac{E_p + p}{\kappa}\right)^{\epsilon(\mu_1-\mu_2)} \times c(s_1 s_2 \lambda_2, \mu_1 \mu_2 M), \tag{4.39}$$

$$\bar{\chi}_M(\epsilon \mid s_1 m_1, s_2 m_2) = e^{iM\gamma + i(m_1+m_2)\varphi} \beta_\epsilon \epsilon^{2s_1} a(s_1 s_2) \times \sum_{\mu_1 \mu_2} d^s_{m_1\mu_1}(\theta) d^s_{m_2\mu_2}(\theta) \left(\frac{E_p - p}{\kappa}\right)^{\epsilon(\mu_1-\mu_2)} \times c(s_1 s_2 \lambda_2, \mu_1 \mu_2 M). \tag{4.40}$$

Denoting

$$u_M(\mathbf{p}) = \chi_M(\mathbf{p}, E_p), \qquad v_M(\mathbf{p}) = \chi_{-M}(-\mathbf{p}, -E_p), \tag{4.41}$$

we read the quantized field (2.24) as

$$\psi(x) = (2\pi)^{-\frac{3}{2}} \int d^3p \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} \sum_M [a_M(\mathbf{p}) u_M(\mathbf{p}) e^{ipx} + b^+_M(\mathbf{p}) v_M(\mathbf{p}) e^{-ipx}], \tag{4.42}$$

$$\bar{\psi}(x) = (2\pi)^{-\frac{3}{2}} \int d^3p \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} \sum_M [b_M(\mathbf{p}) \bar{v}_M(\mathbf{p}) e^{ipx} + a^+_M(\mathbf{p}) \bar{u}_M(\mathbf{p}) e^{-ipx}]. \tag{4.43}$$

Noting that, as $\mathbf{p} \to -\mathbf{p}$, $\varphi \to \pi + \varphi$, $\theta \to \pi - \theta$, $\gamma \to \gamma$, and using the relations

$$c(s_1 s_2 s, -\mu_1 - \mu_2 - M) = (-1)^{s_1+s_2-s} c(s_1 s_2 s, \mu_1 \mu_2 M),$$

$$D^s_{m,-\mu}(\pi + \varphi, \pi - \theta, \gamma) = e^{2i\mu\gamma + i\pi s} D^s_{m\mu}(\varphi, \theta, \gamma),$$

we obtain

$$v_M(\mathbf{p}; s_1 m_1, s_2 m_2) = (-1)^{2s_1} e^{i\pi\lambda_2 + 2iM\gamma} u_M(\mathbf{p}; s_1 m_1, s_2 m_2), \tag{4.44}$$

$$\bar{v}_M(\mathbf{p}; s_1 m_1, s_2 m_2) = (-1)^{2s_1} e^{i\pi\lambda_2 - 2iM\gamma} \bar{u}_M(\mathbf{p}; s_1 m_1, s_2 m_2), \tag{4.45}$$

$$u_M(\mathbf{p}; s_1 m_1, s_2 m_2) = e^{-iM\gamma - i(m_1+m_2)\varphi} a(s_1 s_2) \times \sum_{\mu_1 \mu_2} d^{s_1}_{m_1\mu_1}(\theta) d^{s_2}_{m_2\mu_2}(\theta) \left(\frac{E_p + p}{\kappa}\right)^{\mu_1-\mu_2} \times c(s_1 s_2 \lambda_2, \mu_1 \mu_2 M) = a(s_1 s_2) \sum_{\mu_1 \mu_2} D^{s_1}_{m_1\mu_1} D^{s_2}_{m_2\mu_2} \left(\frac{E_p + p}{\kappa}\right)^{\mu_1-\mu_2} \times c(s_1 s_2 \lambda_2, \mu_1 \mu_2 M), \tag{4.46}$$

$$\bar{u}_M(\mathbf{p}; s_1 m_1, s_2 m_2) = e^{iM\gamma + i(m_1+m_2)\varphi} a(s_1 s_2) \times \sum_{\mu_1 \mu_2} d^{s_1}_{m_1\mu_1}(\theta) d^{s_2}_{m_2\mu_2}(\theta) \left(\frac{E_p - p}{\kappa}\right)^{\mu_1-\mu_2} \times c(s_1 s_2 \lambda_2, \mu_1 \mu_2 M). \tag{4.47}$$

Further, using the symmetry property

$$d^s_{-m-\mu}(\theta) = (-1)^{\mu-m} d^s_{m\mu}(\theta),$$

we get

$$\bar{u}_{-M}(\mathbf{p}; s_1 - m_1, s_2 - m_2) = (-1)^{s_1+s_2-m_1-m_2+\lambda_2-M} u_M(\mathbf{p}; s_1 m_1, s_2 m_2). \quad (4.48)$$

To compare with Weinberg's[4] expressions for u_M and v_M, which can be obtained by group-theoretical considerations, we choose $\gamma = \frac{1}{2}\pi$. Then

$$\psi^{s_1 s_2}_{m_1 m_2}(x) = (2\pi)^{-\frac{3}{2}} \int \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M u_M(\mathbf{p}; s_1 m_1, s_2 m_2) \times [a_M(\mathbf{p}) e^{ipx} + (-1)^{2s_2+\lambda_2-M} b^+_M(\mathbf{p}) e^{-ipx}], \quad (4.49)$$

$$\bar{\psi}^{s_1 s_2}_{m_1 m_2}(x) = (2\pi)^{-\frac{3}{2}} \int \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M \bar{u}_M(\mathbf{p}; s_1 m_1, s_2 m_2) \times [a^+_M(\mathbf{p}) e^{-ipx} + (-1)^{2s_1+\lambda_2-M} b_M(\mathbf{p}) e^{ipx}], \quad (4.50)$$

and

$$\bar{\psi}^{s_1 s_2}_{-m_1 -m_2}(x) = (-1)^{s_1-s_2-m_1-m_2} \int \left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} d^3p \sum_M u_M(\mathbf{p}; s_1 m_1, s_2 m_2) \times [b_{-M}(\mathbf{p}) e^{ipx} + (-1)^{2s_2+\lambda_2-M} a^+_{-M}(\mathbf{p}) e^{-ipx}]. \quad (4.51)$$

Expressions (4.49) and (4.51) are the same as Weinberg's expressions for $\psi^{s_1 s_2}_{m_1 m_2}(x)$ and $(-1)^{s_2-s_1+m_1+m_2} \times \bar{\psi}^{s_1 s_2}_{m_1 m_2}(x)$, which explains the signs and the factors in relation (3.29). There are two differences of convention between Weinberg's expressions and ours. The first is that instead of $b_M(\mathbf{p})$ he uses $b_{-M}(\mathbf{p})$, which means that he defines the helicity of the antiparticle opposite to the direction of its momentum. The second difference is that in his corresponding expression for $u_M(\mathbf{p})$, a factor $[(E_p - p)/\kappa]^{m_1-m_2}$ enters in place of the factor $[(E_p + p)/\kappa]^{m_1-m_2}$ in our expression (4.46). This is due to Weinberg's use of the boost as $L(-\mathbf{p})$, while we use $L(\mathbf{p})$. Actually, this means that his definition of the particle amounts to our definition of the antiparticle. If we define

$$u_M(\mathbf{p}) = \chi_M(\mathbf{p}, -E_p)$$

instead of $\chi_M(\mathbf{p}, E_p)$, we arrive at Weinberg's expression.

5. LORENTZ AND CPT TRANSFORMATIONS

A. Properties of the Causal Functions

From what preceded, it is easily verified that

$$S^{(\epsilon)}_{\rho\rho'}(x - y) = -i\kappa \int \frac{d^3p}{E_p} \sum_M \epsilon \beta_\epsilon \psi_M(x, \mathbf{p}, \epsilon) \bar{\psi}_M(y, \mathbf{p}, \epsilon). \quad (5.1)$$

$\epsilon = \pm$ stands for the energy sign and $\psi_{M\rho}(x, \mathbf{p}, \epsilon)$ is given by (2.2). Hence $S^{(\pm)}(x - y)$ and also $S(x - y)$ satisfy the homogeneous field equations

$$\left(S_{\mu\nu} \frac{\partial}{\partial x_\nu} + i\kappa\gamma_\mu - \frac{\partial}{\partial x_\mu}\right) S(x - y) = 0, \quad (5.2)$$

$$\frac{\partial}{\partial y_\nu} S(x - y) S_{\mu\nu} + i\kappa S(x - y)\gamma_\mu + \frac{\partial}{\partial y_\mu} S(x - y) = 0, \quad (5.3)$$

$$\gamma_\mu \frac{\partial}{\partial x_\mu} S(x - y) + i\kappa S(x - y) = 0, \quad (5.4)$$

$$\frac{\partial}{\partial y_\mu} S(x - y)\gamma_\mu - i\kappa S(x - y) = 0. \quad (5.5)$$

These equations follow directly from Eqs. (I.4.32) and (I.4.33). For convenience we write

$$S_{\rho\rho'}(x - y) = S^{(+)}_{\rho\rho'}(x - y) + S^{(-)}_{\rho\rho'}(x - y) = -i\kappa \sum_{M,\epsilon} \epsilon \beta_\epsilon \int \frac{d^3p}{E_p} \psi_{M\rho}(x, \mathbf{p}, \epsilon) \bar{\psi}_{M\rho'}(y, \mathbf{p}, \epsilon). \quad (5.6)$$

We prove now that

$$\psi(x) = i \int S(x - y)\gamma_\alpha \psi(y)\, d\sigma_\alpha(y) = i \int S(x - y)\gamma_0 \psi(y) d^3y, \quad (5.7)$$

$$\bar{\psi}(x) = i \int \bar{\psi}(y)\gamma_\alpha S(y - x)\, d\sigma_\alpha(y) = i \int \bar{\psi}(y)\gamma_0 S(y - x) d^3y. \quad (5.8)$$

We note first that the manifest covariant integrals are independent of the spacelike hypersurface σ on which the integral is supposed to be performed. This is due to the fact that

$$\frac{\partial}{\partial y_\alpha} [S(x - y)\gamma_\alpha \psi(y)] = \frac{\partial}{\partial y_\alpha} [\bar{\psi}(y)\gamma_\alpha S(y - x)] = 0.$$

Writing

$$\psi_\rho(y) = \sum_{M\epsilon} \int d^3p c_M(\mathbf{p}, \epsilon) \psi_{M\rho}(y, \mathbf{p}, \epsilon),$$

where

$$c_M(\mathbf{p}, +) = (\kappa/E_p)^{\frac{1}{2}} a_M(\mathbf{p}),$$
$$c_M(\mathbf{p}, -) = (\kappa/E_p)^{\frac{1}{2}} b^+_M(-\mathbf{p}),$$

and making use of the expansion (5.6), we get

$$i \int d^3y [S(x - y)\gamma_0 \psi(y)]_\rho = \kappa \iint \frac{d^3p d^3p'}{E_p} \sum_{M\epsilon} \sum_{M'\epsilon'} \epsilon \beta_\epsilon c_{M'}(\mathbf{p}', \epsilon') \psi_{M\rho}(x, \mathbf{p}, \epsilon) \times \sum_{\rho'} \int \bar{\psi}_{M\rho'}(y, \mathbf{p}, \epsilon)[\gamma_0 \psi_{M'}(y, \mathbf{p}', \epsilon')]_{\rho'} d^3y = \psi(x).$$

The last step follows from the orthonormality relation (2.13). Relation (5.8) can be proved in the same way.

We note that, if A is an observable, then $A\psi$ is a solution of the field equations satisfied by ψ. Thus, Eqs. (5.7) and (5.8) hold for $A\psi$:

$$A(x)\psi(x) = i\int S(x-y)\gamma_0 A(y)\psi(y)d^3y = i\int S(x-y)\gamma_\alpha A(y)\psi(y)\,d\sigma_\alpha(y); \quad (5.9)$$

$$\overline{A(x)\psi(x)} = i\int \overline{A(y)\psi(y)}\gamma_0 S(y-x)d^3y = i\int \overline{A(y)\psi(y)}\gamma_\alpha S(y-x)\,d\sigma_\alpha(y). \quad (5.10)$$

Further, we prove the invariance of $S(x-y)$ under inhomogeneous Lorentz transformations. First, we note that it is invariant under translations. Moreover, under rotations,

$$\psi'(x,\Lambda p) = L(\Lambda,x)\psi(x,p),$$
$$\bar\psi'(x,\Lambda p) = \bar\psi(x,p)\overleftarrow{L}^{-1}(\Lambda,x),$$

where $L(\Lambda,x)$ is the homogeneous 4-dimensional rotation and $\overleftarrow{L}^{-1}$ acts on $\bar\psi$ by differentiation. For the infinitesimal transformations,

$$L(x) = I - \tfrac{1}{2}i\epsilon_{\mu\nu}\left[i\left(x_\nu\frac{\partial}{\partial x_\mu} - x_\mu\frac{\partial}{\partial x_\nu}\right) + \Sigma_{\mu\nu}\right].$$

Noting that $d^3(\Lambda p)/E_{\Lambda p} = d^3p/E_p$, we write

$$S_{\rho\rho'}(x-y) = \kappa\sum_{M\epsilon}\epsilon\beta_\epsilon\int\psi_{M\rho}(x,\mathbf{\Lambda p},\epsilon)\bar\psi_{M\rho'}(y,\mathbf{\Lambda p},\epsilon)\frac{d^3p}{E_p}$$
$$= \kappa\sum_{\rho_1\rho_2}\sum_{M\epsilon}\epsilon\beta_\epsilon L_{\rho\rho_1}(x)L^{-1}_{\rho_2\rho'}(y) \times\int\psi_{M\rho_1}(x,\mathbf{p},\epsilon)\bar\psi_{M\rho_2}(y,\mathbf{p},\epsilon)\frac{d^3p}{E_p}.$$

Hence

$$S(x-y) = L(x)S(x-y)\overleftarrow{L}^{-1}(y). \quad (5.11)$$

This relation holds for the inhomogeneous transformations as well. It follows that, for any generator A of the rotations and the translations,

$$A(x)S(x-y) = S(x-y)\overleftarrow{A}(y). \quad (5.12)$$

B. The Inhomogeneous Lorentz Transformation

For our purpose it is more convenient to define the Hilbert space operator of a dynamical variable $A(x)$ as

$$\langle A\rangle = \int\bar\psi(x)\gamma_0 A(x)\psi(x)d^3x$$

instead of the definition (2.17). It satisfies

$$[\langle A\rangle,\langle B\rangle] = \iint d^3xd^3y\sum_{\rho_1\rho_1'}\sum_{\rho_2\rho_2'}\{\bar\psi_{\rho_1}(x)[\gamma_0A(x)]_{\rho_1\rho_1'} \times\psi_{\rho_1'}(x),\bar\psi_{\rho_2}(y)[\gamma_0B(y)]_{\rho_2\rho_2'}\psi_{\rho_2'}(y)\}$$
$$= i\iint d^3xd^3y[\bar\psi(x)\gamma_0A(x)S(x-y)\gamma_0B(y)\psi(y) - \bar\psi(y)\gamma_0B(y)S(y-x)\gamma_0A(x)\psi(x)].$$

If A and B are two observables, we use (5.9) and get

$$[\langle A\rangle,\langle B\rangle] = \langle[A,B]\rangle. \quad (5.13)$$

Thus, the energy, momentum, and angular-momentum operators

$$P_\mu = \langle -iS_{\mu\nu}\partial_\nu + \kappa\gamma_\mu\rangle,$$
$$M_{\mu\nu} = \langle -i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}\rangle \quad (5.14)$$

satisfy the usual commutation relations

$$[P_\mu,P_\nu] = 0, \quad (5.15)$$

$$[M_{\mu\nu},P_\alpha] = i(\delta_{\mu\alpha}P_\nu - \delta_{\nu\alpha}P_\mu), \quad (5.16)$$

$$[M_{\mu\nu},M_{\alpha\beta}] = i(\delta_{\mu\alpha}M_{\nu\beta} + \delta_{\nu\beta}M_{\mu\alpha} - \delta_{\mu\beta}M_{\nu\alpha} - \delta_{\nu\alpha}M_{\mu\beta}). \quad (5.17)$$

To make these operators finite, we subtract from them their vacuum expectation values. Also, we have

$$[\langle A\rangle,\psi_\sigma(x)] = \int d^3y\sum_{\rho\rho'}[\bar\psi_\rho(y)\psi_{\rho'}(y),\psi_\sigma(x)](\gamma_0A(y))_{\rho\rho'}$$
$$= -i\int[S(x-y)\gamma_0A(y)\psi(y)]_\sigma d^3y.$$

Thus, for observables,

$$[\langle A\rangle,\psi(x)] = -A(x)\psi(x). \quad (5.18)$$

Accordingly, the infinitesimal inhomogeneous Lorentz transformation of the Hilbert space is

$$U = I - \tfrac{1}{2}i\epsilon_{\mu\nu}M_{\mu\nu} + a_\mu P_\mu \quad (5.19)$$

such that

$$U^{-1}\psi(x)U = (I - \tfrac{1}{2}i\epsilon_{\mu\nu}J_{\mu\nu} + ia_\mu\hat P_\mu)\psi(x)$$
$$= (I - \tfrac{1}{2}i\epsilon_{\mu\nu}J_{\mu\nu} + a_\mu\partial_\mu)\psi(x). \quad (5.20)$$

Here

$$J_{\mu\nu} = -i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu},$$
$$\hat P_\mu = -iS_{\mu\nu}\partial_\nu + \kappa\gamma_\mu. \quad (5.21)$$

The second line of (5.20) follows from the field equations $\hat{P}_\mu\psi = -i\partial_\mu\psi$.

From (5.20) and the integrability conditions (5.15)–(5.17) the corresponding relation for finite proper inhomogeneous transformation follows:

$$U^{-1}(\Lambda, a)\psi(x)U(\Lambda, a) = \psi'(\Lambda x + a) = L(\Lambda, a)\psi(x). \quad (5.22)$$

Further, for an observable,

$$[\langle A\rangle, \bar{\psi}(x)] = i\int \bar{\psi}(y)\gamma_0 A(y)S(y-x)d^3y = i\int \bar{\psi}(y)\gamma_0 S(y-x)\overleftarrow{A}(x)d^3y = \bar{\psi}(x)\overleftarrow{A}(x), \quad (5.23)$$

which proves that

$$U^{-1}(\Lambda, a)\bar{\psi}(x)U(\Lambda, a) = \bar{\psi}'(\Lambda x, a) = \bar{\psi}(x)L^{-1}(\Lambda, a). \quad (5.24)$$

This shows the unitarity of the transformation U of the Hilbert space.

C. The Parity Operator

In discussing the discrete transformations, we shall restrict ourselves to the expansion in terms of the eigenstates of the observable spin

$$\psi(\mathbf{r}, t) = (2\pi)^{-\frac{3}{2}}\sum_M \int d^3p\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}} \times [e^{ipx}a_M(\mathbf{p})U_M(\mathbf{p}) + e^{-ipx}b_M^+(\mathbf{p})V_M(\mathbf{p})].$$

Noting that the boost $B(\mathbf{p}, \epsilon)$ satisfies the relations

$$B(-\mathbf{p}, \epsilon) = B^{-1}(\mathbf{p}, \epsilon) = \beta B(\mathbf{p}, \epsilon)\beta,$$

we obtain the following relation:

$$\beta w_M(-\mathbf{p}, p_0) = \beta B(-\mathbf{p}, \epsilon)\phi_M(\epsilon) = B(\mathbf{p}, \epsilon)\beta\phi_M(\epsilon) = \beta_\epsilon w_M(\mathbf{p}, p_0), \quad (5.25)$$

since $\beta\phi_M(\epsilon) = \beta_\epsilon\phi_M(\epsilon)$. Hence

$$\beta U_M(-\mathbf{p}) = U_M(\mathbf{p}), \qquad \beta V_M(-\mathbf{p}) = \beta_- V_M(\mathbf{p}) = (-1)^{2\lambda_2}V_M(\mathbf{p}). \quad (5.26)$$

Under spatial inversion,

$$\psi(\mathbf{r}, t) \to \beta\psi(-\mathbf{r}, t) = (2\pi)^{-\frac{3}{2}}\int d^3p\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}}\sum_M [e^{ipx}U_M(\mathbf{p})a_M(-\mathbf{p}) + (-1)^{2\lambda_2}e^{-ipx}V_M(\mathbf{p})b_M^+(-\mathbf{p})], \quad (5.27)$$

and

$$\bar{\psi}(\mathbf{r}, t) \to \psi^+(-\mathbf{r}, t) = (2\pi)^{-\frac{3}{2}}\int d^3p\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}}\sum_M [e^{-ipx}\bar{U}_M(\mathbf{p})a_M^+(-\mathbf{p}) + (-1)^{2\lambda_2}e^{ipx}\bar{V}_M(\mathbf{p})b_M(-\mathbf{p})]. \quad (5.28)$$

The corresponding transformation of the Hilbert space is

$$P = \exp\left\{\tfrac{1}{2}i\pi\sum_M\int d^3p[b_M(\mathbf{p})b_M^+(-\mathbf{p}) - a_M^+(\mathbf{p})a_M(-\mathbf{p})]\right\}. \quad (5.29)$$

It is unitary: $P^+ = P^{-1}$. It satisfies

$$\begin{aligned} Pa_M(\mathbf{p})P^{-1} &= ia_M(-\mathbf{p}),\\ Pa_M^+(\mathbf{p})P^{-1} &= -ia_M^+(-\mathbf{p}),\\ Pb_M^+(\mathbf{p})P^{-1} &= (-1)^{2\lambda_2}ib_M^+(-\mathbf{p}),\\ Pb_M(\mathbf{p})P^{-1} &= (-1)^{2\lambda_2+1}ib_M(-\mathbf{p}), \end{aligned} \quad (5.30)$$

such that

$$\begin{aligned} P\psi(x)P^{-1} &= i\beta\psi(-\mathbf{r}, t),\\ P\bar{\psi}(x)P^{-1} &= -i\bar{\psi}(-\mathbf{r}, t)\beta. \end{aligned} \quad (5.31)$$

Hence

$$[P, Q] = [P, H] = [P, \langle S_k\rangle] = 0, \qquad PP_kP^{-1} = -P_k. \quad (5.32)$$

D. The Charge-Conjugation Operator

Under charge conjugation, $\psi \to \psi^c = \beta C\psi^\dagger$, where C is the complex-conjugation matrix [cf. (I.2.45)]. The boost commutes with the charge-conjugation operator. Hence

$$w_M^c(\mathbf{p}, p_0) = \beta C w_M^*(\mathbf{p}, p_0) = B(\mathbf{p}, \epsilon)\phi_M^c(\epsilon).$$

Further, under charge conjugation, $\Sigma_k \to -\Sigma_k$ and $\gamma_0 \to -\gamma_0$. Since $\gamma_0\phi_M(\epsilon) = \epsilon\phi_M(\epsilon)$ and $\Sigma_3\phi_M(\epsilon) = M\phi_M(\epsilon)$, then $\phi_M^c(\epsilon) = \xi\phi_{-M}(-\epsilon)$. From the normalization condition $\phi_M^+\phi_M = 1$, it follows that $|\xi|^2 = 1$. Using the property $B(-\mathbf{p}, -\epsilon) = B(\mathbf{p}, \epsilon)$, we get

$$w_M^c(\mathbf{p}, p_0) = \xi w_{-M}(-\mathbf{p}, -p_0). \quad (5.33)$$

Hence

$$U_M^c(\mathbf{p}) = \xi V_{-M}(\mathbf{p}), \qquad V_M^c(\mathbf{p}) = \xi U_{-M}(\mathbf{p}),$$

and

$$\psi^c(x) = \beta C\psi^+(x) = \xi(2\pi)^{-\frac{3}{2}}\int d^3p\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}}\sum_M [e^{ipx}U_M(\mathbf{p})b_{-M}(\mathbf{p}) + e^{-ipx}V_M(\mathbf{p})a_{-M}^+(\mathbf{p})]. \quad (5.34)$$

The corresponding transformation of the Hilbert space,

$$\psi^c(x) = \mathcal{C}\psi(x)\mathcal{C}^{-1},$$
$$\bar{\psi}^c(x) = \mathcal{C}\bar{\psi}(x)\mathcal{C}^{-1}, \tag{5.35}$$

is unitary and is given by $\mathcal{C} = \mathcal{C}_1\mathcal{C}_2$, where[15]

$$\mathcal{C}_1 = \exp\left[i\pi\int d^3p \sum_M b_M^+(\mathbf{p})b_M(\mathbf{p})\right],$$
$$\mathcal{C}_2 = \exp\left[\tfrac{1}{2}\pi\int d^3p \sum_M \xi^* b_M^+(\mathbf{p})a_{-M}(\mathbf{p}) - \xi a_M^+(\mathbf{p})b_{-M}(\mathbf{p})\right], \tag{5.36}$$

such that

$$\mathcal{C}a_M(\mathbf{p})\mathcal{C}^{-1} = \xi b_{-M}(\mathbf{p}),$$
$$\mathcal{C}b_M(\mathbf{p})\mathcal{C}^{-1} = \xi^* a_{-M}(\mathbf{p}),$$
$$\mathcal{C}a_M^+(\mathbf{p})\mathcal{C}^{-1} = \xi^* b_{-M}^+(\mathbf{p}),$$
$$\mathcal{C}b_M^+(\mathbf{p})\mathcal{C}^{-1} = \xi a_{-M}^+(\mathbf{p}). \tag{5.37}$$

Hence

$$\mathcal{C}Q\mathcal{C}^{-1} = -Q,$$
$$\mathcal{C}P_\mu\mathcal{C}^{-1} = P_\mu, \tag{5.38}$$
$$\mathcal{C}\langle S_3\rangle\mathcal{C}^{-1} = -\langle S_3\rangle.$$

E. The Time-Reversal Operator

The canonical time-reversal operation is, as is well known,[16]

$$\psi^W(x) = \rho C\psi^*(\mathbf{r}, -t)$$
$$= (2\pi)^{-\frac{3}{2}}\int d^3p\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}}\sum_M [a_M^*(-\mathbf{p})\rho C U_M^*(-\mathbf{p})e^{ipx} + b_M^{\mathrm{T}}(-\mathbf{p})\rho C V_M^*(-\mathbf{p})e^{-ipx}], \tag{5.39}$$

where a^* and b^{T} are the complex conjugate and the transposed operators in the Hilbert space. ρ is the space–time reversal operator: $\rho\gamma_\mu + \gamma_\mu\rho = 0$,

$$[\rho, S_{\mu\nu}] = 0.$$

Using the properties [cf. Eqs. (I.2.45)–(I.2.48)]

$$(\rho C)^+\gamma_0\rho C = \gamma_0^*,$$
$$(\rho C)^+\gamma_k\rho C = -\gamma_k^*,$$
$$(\rho C)^+\Sigma_k\rho C = -\Sigma_k^*,$$
$$(\rho C)^+\alpha_k\rho C = -\alpha_k^*,$$

one concludes that

$$\rho C\phi_M^*(\epsilon) = \eta\phi_{-M}(\epsilon),$$

where $|\eta|^2 = 1$. Further, $[\rho, B(\mathbf{p}, \epsilon)] = 0$ and

$$B^*(\mathbf{p}, \epsilon) = C^+B^{-1}(\mathbf{p}, \epsilon)C = C^+B(-\mathbf{p}, \epsilon)C.$$

Hence

$$\rho Cw_M^*(-\mathbf{p}, p_0) = B(\mathbf{p}, \epsilon)\rho C\phi_M^*(\epsilon) = \eta w_{-M}(\mathbf{p}, p_0). \tag{5.40}$$

Accordingly,

$$\psi^W(x) = \eta(2\pi)^{-\frac{3}{2}}\int d^3p\left(\frac{\kappa}{E_p}\right)^{\frac{1}{2}}\sum_M [a_{-M}^*(-\mathbf{p})U_M(\mathbf{p})e^{ipx} + b_{-M}^{\mathrm{T}}(-\mathbf{p})V_M(\mathbf{p})e^{-ipx}]. \tag{5.41}$$

The corresponding Hilbert space unitary operator W satisfies the usual relations

$$\psi^W(x) = W\psi(x)W^{-1},$$
$$\bar{\psi}^W(x) = W\bar{\psi}(x)W^{-1}, \tag{5.42}$$

$$Wa_M(\mathbf{p})W^{-1} = \eta a_{-M}^*(-\mathbf{p}),$$
$$Wb_M^+(\mathbf{p})W^{-1} = \eta b_{-M}^{\mathrm{T}}(-\mathbf{p}),$$
$$Wa_M^+(\mathbf{p})W^{-1} = \eta^* a_{-M}^{\mathrm{T}}(-\mathbf{p}),$$
$$Wb_M(\mathbf{p})W^{-1} = \eta^* b_{-M}^*(-\mathbf{p}). \tag{5.43}$$

ACKNOWLEDGMENTS

The author is grateful to the referee, who looked at the paper before its revision, for drawing his attention to the paper by Tung (Ref. 3). The author would also like to thank Professor A. Salam and Professor P. Budini as well as the International Atomic Energy Agency for hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX A: THE SCALAR PRODUCT

In Paper I the Lorentz-invariant scalar product of two vectors ψ_1 and ψ_2 is defined by [cf. (I.4.91)]

$$(\psi_1, \psi_2) = \int \bar{\psi}_1(x)\gamma_\alpha\psi_2(x)\, d\sigma_\alpha, \tag{A1}$$

where the integration is carried out over a spacelike hypersurface and

$$d\sigma_\alpha = (dx_2\, dx_3\, dt, dx_3\, dx_1\, dt, dx_1\, dx_2\, dt, -id^3x).$$

The dependence of the scalar product on the hypersurface σ is given by the functional derivative

$$\frac{\delta}{\delta\sigma}(\psi_1, \psi_2) = \partial_\alpha(\psi_1\gamma_\alpha\psi_2). \tag{A2}$$

If ψ_1 and ψ_2 are two solutions of the field equations, they satisfy Bhabha's equations

$$\gamma_\alpha\partial_\alpha\psi = -i\kappa\psi,$$
$$\partial_\alpha\bar{\psi}\gamma_\alpha = i\kappa\psi, \tag{A3}$$

such that

$$\partial_\alpha(\bar{\psi}_1\gamma_\alpha\psi_2) = 0. \tag{A4}$$

Thus the scalar product of two solutions of the field equations is independent of the hypersurface σ, and we can use a flat surface $t = \text{const}$ to evaluate this

scalar product:

$$(\psi_1, \psi_2) = \int \bar{\psi}_1 \gamma_0 \psi_2 d^3x. \tag{A5}$$

The expectation value of a dynamical variable A is defined as

$$\langle A \rangle = (\psi, A\psi). \tag{A6}$$

Since, for an observable, $\phi = A\psi$ is a solution of the same field equations satisfied by ψ, it follows that

$$\frac{\delta}{\delta\sigma} \langle A \rangle = 0, \tag{A7}$$

and we can simply calculate $\langle A \rangle$ on a flat surface:

$$\langle A \rangle = \int \bar{\psi}(x) \gamma_0 A(x) \psi(x) d^3x. \tag{A8}$$

APPENDIX B: THE TENSOR $t^{\mu_1 \cdots \mu_{2j}}_{m_1 m_2}$

Let ϕ and χ be two $(2j+1)$-column matrices carrying the representations $D(0,j)$ and $D(j,0)$ of the Poincaré group with mass κ, respectively. To fix the notation, they transform under Lorentz transformations as follows:

$$\begin{aligned} \phi'(p') &= L^{-1}(\mathbf{p})\phi(p) = L(-\mathbf{p})\phi(p), \\ \chi'(p') &= L(\mathbf{p})\chi(p), \end{aligned} \tag{B1}$$

where

$$L(\mathbf{p}) = \exp\,(\theta_1 \boldsymbol{\Sigma} \cdot \mathbf{p}/p), \tag{B2}$$

$\sinh \theta_1 = p/\kappa$, and the Σ_k are the spin matrices in the irreducible representation $D(j)$ of $O(3)$.

Weinberg[7] has shown that, with appropriate normalization, ϕ and χ are related as follows:

$$\begin{aligned} \phi &= (-\kappa)^{-2j} \Pi^{(j)}(\mathbf{p})\chi, \\ \chi &= (-\kappa)^{-2j} \overline{\Pi}^{(j)}(\mathbf{p})\phi, \end{aligned} \tag{B3}$$

where $\Pi^{(j)}(\mathbf{p})$ is the mass-shell value of the expression (3.26). Here

$$\begin{aligned} \Pi^{(j)}(\mathbf{p}) &= \kappa^{2j} L^{-2}(\mathbf{p}), \\ \overline{\Pi}^{(j)}(\mathbf{p}) &= \kappa^{2j} L^{2}(\mathbf{p}). \end{aligned} \tag{B4}$$

By solving the field equations (2.1), we have obtained the connection (B3) in a most suitable form. In fact, taking $\lambda_2 = j$, we see that the Wigner components are $\psi(\lambda_2, 0) = \chi$ and $\psi(0, \lambda_2) = \phi$. Equations (II.2.7) and (II.2.9) lead directly to

$$\begin{aligned} \Pi^{(j)}(p) = [(2j)!]^{-1} & p^{a_1 \dot{b}_1} \cdots p^{a_{2j} \dot{b}_{2j}} \\ & \times P_{a_1 \cdots a_{2j}}(j) \otimes R_{\dot{b}_1 \cdots \dot{b}_{2j}}(j), \end{aligned} \tag{B5}$$

$$\begin{aligned} \overline{\Pi}^{(j)}(p) = [(-1)^{2j}/(2j)!] & p^{a_1 \dot{b}_1} \cdots p^{a_{2j} \dot{b}_{2j}} \\ & \times R_{a_1 \cdots a_{2j}}(j) \otimes P_{\dot{b}_1 \cdots \dot{b}_{2j}}(j). \end{aligned} \tag{B6}$$

Here a_s and $\dot{b}_s$ are spinor indices taking the values 1, 2 with

$$\begin{aligned} p^{1\dot{2}} &= p_1 - ip_2, \quad p^{2\dot{1}} = p_1 + ip_2, \\ p^{1\dot{1}} &= p_0 + p_3, \quad p^{2\dot{2}} = p_0 - p_3. \end{aligned} \tag{B7}$$

$R_{a_1 \cdots a_{2j}}(j)$ and $P_{a_1 \cdots a_{2j}}(j)$ are $(2j+1)$-row and -column matrices, respectively, symmetric in the spinor indices. These matrix elements are given by [cf. (II.2.13)]

$$\begin{aligned} \underline{R}^{j-m'}_{m} &= (-1)^{2j} \underline{\dot{P}}^{j-m'}_{m} \\ &= (-1)^{j+m} \delta_{mm'} [(j-m)!\,(j+m)!]^{\frac{1}{2}}, \\ \underline{\dot{R}}^{j+m'}_{m} &= \underline{P}^{j+m'}_{m} \\ &= (-1)^{2j} \delta_{mm'} [(j-m)!\,(j+m)!]^{\frac{1}{2}}. \end{aligned} \tag{B8}$$

Here $\underline{P}^{\sigma}_{m}$ and $\underline{\dot{P}}_{m}$ are the matrix elements of $P_{a_1 \cdots a_{2j}}$ and $P_{\dot{b}_1 \cdots \dot{b}_{2j}}$, having σ of the indices equal to 1. Hence,

$$\begin{aligned} \Pi^{(j)}_{mm'}(p) = & [(j-m)!\,(j+m)!\,(j-m')!\,(j+m')!]^{\frac{1}{2}} \\ & \times \sum [n_1!\, n_2!\, n_3!\, n_4!]^{-1} \\ & \times (p^{1\dot{1}})^{n_1} (p^{1\dot{2}})^{n_2} (p^{2\dot{1}})^{n_3} (p^{2\dot{2}})^{n_4}, \end{aligned} \tag{B9}$$

where the summation is subject to the restrictions

$$\begin{aligned} n_1 + n_2 = j + m, \quad & n_1 + n_3 = j + m', \\ n_1 + n_2 + n_3 + n_4 = 2j. & \end{aligned} \tag{B10}$$

Expanding in powers of p_μ, we finally get

$$\Pi^{(j)}_{mm'}(p) = \sum_{l_1 l_2 l_3 l_0} \tau^{(l_1 l_2 l_3 l_0)}_{mm'}(j) p_1^{l_1} p_2^{l_2} p_3^{l_3} p_0^{l_0}, \tag{B11}$$

where

$$\begin{aligned} &\tau^{(l_1 l_2 l_3 l_0)}_{mm'}(j) \\ &\quad = 2^{s+s'} (-1)^{l_3} i^{l_2} d^{s}_{m_1 m_2}(\tfrac{1}{2}\pi) d^{s'}_{m_1' m_2'}(\tfrac{1}{2}\pi) \\ &\qquad \times \left[\frac{(j+m)!\,(j-m)!\,(j-m')!\,(j+m')!}{n_1!\, n_2!\, n_3!\, n_4!\, l_1!\, l_2!\, l_3!\, l_0!} \right]^{\frac{1}{2}} \end{aligned} \tag{B12}$$

Here

$$\begin{aligned} l_0 + l_1 + l_2 + l_3 &= 2j, \\ 2s = l_0 + l_3, \quad & 2s' = l_1 + l_2, \\ 2m_1 = m + m', \quad & 2m_1' = m - m', \\ 2m_2 = l_0 - l_3, \quad & 2m_2' = l_1 - l_2, \end{aligned} \tag{B13}$$

and

$$\begin{aligned} 2n_1 &= l_0 + l_3 + m + m', \\ 2n_4 &= l_0 + l_3 - m - m', \\ 2n_2 &= l_1 + l_2 + m - m', \\ 2n_3 &= l_1 + l_2 - m + m'. \end{aligned} \tag{B14}$$

We note that $l_0 + l_3 - |m + m'|$ and $l_1 + l_2 - |m - m'|$ should both be even and nonnegative. This restricts the admissible l's for a given m and m'. The $d^{s}_{m_1 m_2}(\frac{1}{2}\pi)$ in (B12) is the usual rotation matrix.

We note also that

$$\overline{\Pi}^{(j)}(\mathbf{p}, p_0) = \Pi^{(j)}(-\mathbf{p}, p_0),$$

such that

$$\overline{\Pi}^{(j)}_{mm'}(p) = \sum (-1)^{l_1+l_2+l_3} \tau^{(l_1 l_2 l_3 l_0)}_{mm'}(j) p_1^{l_1} p_2^{l_2} p_3^{l_3} p_0^{l_0}, \quad \text{(B15)}$$

as already pointed out by Weinberg.

To make sure that there is no phase difference between expressions (B4) and (B5), we note that in the rest system $p_k = 0$ and $p_0 = \kappa$. Expression (B4) leads to $\pi^{(j)}(\mathbf{0}, \kappa) = \kappa^{2j} I$. Using (B9), we get, in favor of (B10),

$$\Pi^{(j)}_{mm'}(\mathbf{0}, \kappa) = \kappa^{2j} \delta_{mm'},$$

proving the equality of (B4) and (B5).

* On leave of absence from Cairo University, Giza, U.A.R.

[1] M. M. Bakri, J. Math. Phys. **10**, 298 (1969). This is Paper I.
[2] M. M. Bakri, A. F. Ghaleb, and M. I. Hessein, J. Math. Phys. **11**, 2015 (1970). This is Paper II.
[3] Wu-Ki Tung, Phys. Rev. **156**, 1385 (1967).
[4] S. Weinberg, Phys. Rev. **181**, 1893 (1969).
[5] D. L. Pursey, Ann. Phys. (N.Y.) **32**, 157 (1965).
[6] Cf. Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, Oxford, 1969).
[7] S. Weinberg, Phys. Rev. **133B**, 1318 (1964).
[8] R. Shaw, Nuovo Cimento **33**, 1074 (1964).
[9] M. M. Bakri, Nucl. Phys. **B12**, 341 (1969).
[10] G. S. Guralnik and T. W. B. Kibble, Phys. Rev. **139B**, 712 (1965).
[11] S. J. Chang, Phys. Rev. **161**, 1316 (1967).
[12] S. J. Chang, Phys. Rev. **161**, 1308 (1967).
[13] A. McKerrel, Ann. Phys. (N.Y.) **40**, 237 (1966).
[14] Cf. M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[15] The charge-conjugation operator was taken by Takahashi (Ref. 6, p. 167) to be C_2 only. However, this would lead to the wrong sign in the transformation of $b_M(\mathbf{p})$ and $b^\dagger_M(\mathbf{p})$.
[16] Cf. Ref. 6, p. 181.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Peratization of a Class of Exponentially Singular Potentials*

WILLIAM M. FRANK AND DAVID J. LAND†
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland 20910

(Received 5 September 1969)

First, second, and third peratizations are performed for evaluation of the scattering length for repulsive potentials of the form $V(r) = gr^{-m} \exp(\lambda/r)$ with m an integer ≥ 4. For $m = 4$ a nontrivial result is found in second and higher peratization which has the same weak-coupling limit as the exact answer and agrees with the result of Calogero and Cassandro. For $m > 4$ one finds a zero result in both first, second, and third peratization, suggesting the inadequacy of the procedure as generally applied. The procedure of separate peratization of numerator and denominator of a ratio which is employed in this work is discussed at length. The optimum n method is introduced in the course of this discussion as a powerful and convenient method of evaluating the scattering length without summing the appropriate series. It also clarifies the limitations of the peratization approximation. A modification of the peratization procedure is suggested which has promise of being a more successful approximation scheme.

I. INTRODUCTION

In this article a procedure is presented for calculating the scattering length by successive peratizations for the class of repulsive singular potentials of the form

$$gV(r) = ge^{\lambda/r}/r^m, \quad (1)$$

where m is an integer greater than or equal to four. The calculation with $m = 4$ has been performed by Calogero and Cassandro[1] (CC) with one form of regularization and is significant in that it, and a very closely related potential[1,2] $V'(r) = r^{-4}(ge^{\lambda/r} + g')$, are the only potentials for which the peratization prescription seems to provide a successive approximation procedure in the leading coupling constant. This is to be distinguished from potentials, having an additive weaker singularity, which can be peratized in the coupling constant of the weaker part.[3,4] It is therefore of particular interest to ascertain whether such a result might perhaps be characteristic of other exponentially singular potentials, or whether the CC result is only a fluke. We find the latter to be the case. In the present calculation we do not know the exact expression for the scattering length (except when $m = 4$), but we do have a tractable procedure for calculating peratization approximations. The calculation is performed up to third peratization. The fact that the first three peratizations give zero is an indication that the procedure is not practical if it converges at all, which is doubtful. For the $m = 4$ case, we find that a higher peratization, which has not been calculated previously, gives an improved approximation to the correct answer.

In Sec. II the calculation of the scattering length

We note also that

$$\overline{\Pi}^{(j)}(\mathbf{p}, p_0) = \Pi^{(j)}(-\mathbf{p}, p_0),$$

such that

$$\overline{\Pi}^{(j)}_{mm'}(p) = \sum (-1)^{l_1+l_2+l_3} \tau^{(l_1 l_2 l_3 l_0)}_{mm'}(j) p_1^{l_1} p_2^{l_2} p_3^{l_3} p_0^{l_0}, \quad \text{(B15)}$$

as already pointed out by Weinberg.

To make sure that there is no phase difference between expressions (B4) and (B5), we note that in the rest system $p_k = 0$ and $p_0 = \kappa$. Expression (B4) leads to $\pi^{(j)}(\mathbf{0}, \kappa) = \kappa^{2j} I$. Using (B9), we get, in favor of (B10),

$$\Pi^{(j)}_{mm'}(\mathbf{0}, \kappa) = \kappa^{2j} \delta_{mm'},$$

proving the equality of (B4) and (B5).

* On leave of absence from Cairo University, Giza, U.A.R.

[1] M. M. Bakri, J. Math. Phys. **10**, 298 (1969). This is Paper I.
[2] M. M. Bakri, A. F. Ghaleb, and M. I. Hessein, J. Math. Phys. **11**, 2015 (1970). This is Paper II.
[3] Wu-Ki Tung, Phys. Rev. **156**, 1385 (1967).
[4] S. Weinberg, Phys. Rev. **181**, 1893 (1969).
[5] D. L. Pursey, Ann. Phys. (N.Y.) **32**, 157 (1965).
[6] Cf. Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, Oxford, 1969).
[7] S. Weinberg, Phys. Rev. **133B**, 1318 (1964).
[8] R. Shaw, Nuovo Cimento **33**, 1074 (1964).
[9] M. M. Bakri, Nucl. Phys. **B12**, 341 (1969).
[10] G. S. Guralnik and T. W. B. Kibble, Phys. Rev. **139B**, 712 (1965).
[11] S. J. Chang, Phys. Rev. **161**, 1316 (1967).
[12] S. J. Chang, Phys. Rev. **161**, 1308 (1967).
[13] A. McKerrel, Ann. Phys. (N.Y.) **40**, 237 (1966).
[14] Cf. M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[15] The charge-conjugation operator was taken by Takahashi (Ref. 6, p. 167) to be C_2 only. However, this would lead to the wrong sign in the transformation of $b_M(\mathbf{p})$ and $b^\dagger_M(\mathbf{p})$.
[16] Cf. Ref. 6, p. 181.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Peratization of a Class of Exponentially Singular Potentials*

WILLIAM M. FRANK AND DAVID J. LAND†
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland 20910

(Received 5 September 1969)

First, second, and third peratizations are performed for evaluation of the scattering length for repulsive potentials of the form $V(r) = gr^{-m} \exp(\lambda/r)$ with m an integer ≥ 4. For $m = 4$ a nontrivial result is found in second and higher peratization which has the same weak-coupling limit as the exact answer and agrees with the result of Calogero and Cassandro. For $m > 4$ one finds a zero result in both first, second, and third peratization, suggesting the inadequacy of the procedure as generally applied. The procedure of separate peratization of numerator and denominator of a ratio which is employed in this work is discussed at length. The optimum n method is introduced in the course of this discussion as a powerful and convenient method of evaluating the scattering length without summing the appropriate series. It also clarifies the limitations of the peratization approximation. A modification of the peratization procedure is suggested which has promise of being a more successful approximation scheme.

I. INTRODUCTION

In this article a procedure is presented for calculating the scattering length by successive peratizations for the class of repulsive singular potentials of the form

$$gV(r) = ge^{\lambda/r}/r^m, \quad (1)$$

where m is an integer greater than or equal to four. The calculation with $m = 4$ has been performed by Calogero and Cassandro[1] (CC) with one form of regularization and is significant in that it, and a very closely related potential[1,2] $V'(r) = r^{-4}(ge^{\lambda/r} + g')$, are the only potentials for which the peratization prescription seems to provide a successive approximation procedure in the leading coupling constant. This is to be distinguished from potentials, having an additive weaker singularity, which can be peratized in the coupling constant of the weaker part.[3,4] It is therefore of particular interest to ascertain whether such a result might perhaps be characteristic of other exponentially singular potentials, or whether the CC result is only a fluke. We find the latter to be the case. In the present calculation we do not know the exact expression for the scattering length (except when $m = 4$), but we do have a tractable procedure for calculating peratization approximations. The calculation is performed up to third peratization. The fact that the first three peratizations give zero is an indication that the procedure is not practical if it converges at all, which is doubtful. For the $m = 4$ case, we find that a higher peratization, which has not been calculated previously, gives an improved approximation to the correct answer.

In Sec. II the calculation of the scattering length

in two different regularizations is expressed in the form of a ratio N/D, and the coefficients of the power series are expressible in terms of certain multiple integrals. The peratization procedure employed in this work consists of the separate expansion of numerator and denominator into leading singularities rather than the direct expansion of the Born series. In Sec. III the peratization of N and D is performed and the first peratized scattering length is calculated. In Sec. IV the "optimum-n" point of view is introduced as a convenient means of analyzing the separate peratization of numerator and denominator and as a neat, quick way of summing the power series for N and D. Section V deals specifically with the peratization of a ratio. Some aspects of this question are also discussed in Appendices A and D. In Sec. VI higher-order peratizations of the scattering length are computed. The optimum-n method is used to show why any peratization procedure which is purely mechanical should not be expected to yield significant results. Some general discussion is presented in Sec. VII.

II. CALCULATION OF THE SCATTERING LENGTH

We follow the prescription of calculating the scattering length for a repulsive singular potential as the limit of the scattering length of a regularized sequence of potentials[5] $V(r, \alpha)$, where $V(r, 0) = V(r)$ and $V(r, \alpha)$ is nonsingular for $\alpha \neq 0$. While there are examples where such a limiting procedure does not apply,[6] it can be shown[7a,b] to be legitimate for the regularizations used in the present article. We consider two alternative regularizations, the "θ regularization" where the singular potential $V(r)$ is considered as the limit as $\alpha \to 0^+$ of the potential sequence

$$V_\theta(r, \alpha) = V(r)\theta(r - \alpha) \tag{2}$$

[$\theta(x)$ denotes the step function which is unity for nonnegative values of x and zero otherwise], and the "+ regularization" where $V(r)$ is considered as the limit as $\alpha \to 0^+$ of the sequence

$$V_+(r, \alpha) \equiv V(r + \alpha). \tag{3}$$

The negative of the scattering length $-A(g)$ is expressible as the ratio of the two power series[8]

$$-A(g) = \frac{N(g)}{D(g)} = \frac{\sum_{n=1}^{\infty} g^n \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n r_1(r_1 - r_2)\cdots(r_{n-1} - r_n)r_n V(r_1)\cdots V(r_n)}{\sum_{n=0}^{\infty} g^n \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n (r_1 - r_2)\cdots(r_{n-1} - r_n)r_n V(r_1)\cdots V(r_n)}. \tag{4}$$

The regularized scattering lengths are denoted, respectively, by $A^\theta(g, \alpha)$ and $A^+(g, \alpha)$. In an obvious notation we write

$$N^\gamma(g, \alpha) = \sum_{n=1}^{\infty} g^n N_n^\gamma(\alpha), \quad D^\gamma(g, \alpha) = \sum_{n=0}^{\infty} g^n D_n^\gamma(\alpha), \tag{5}$$

where $\gamma = \theta, +$. The dependence on the variable g, α is frequently suppressed in the notation.

With $V(r)$ as in Eq. (1), one finds

$$\begin{aligned} D_n^\theta(\alpha) &= \int_\alpha^\infty dr_1 \cdots \int_\alpha^{r_{n-1}} dr_n (r_1 - r_2)\cdots r_n \\ &\quad \times e^{\lambda \Sigma r_j^{-1}} \prod_{j=1}^n r_j^{-m} \\ &= \int_0^{1/\alpha} dy_n \cdots \int_0^{y_2} dy_1 (y_n - y_{n-1})\cdots(y_2 - y_1)y_1 \\ &\quad \times e^{\lambda \Sigma y_j} \prod_{j=1}^n (y_1 \cdots y_n)^{m-4}, \end{aligned} \tag{6}$$

where the variable changes $r_j = y_j^{-1}$ $(j = 1, \cdots, n)$ have been made. We introduce the parametrized integral

$$J(\beta_1, \cdots, \beta_n) \equiv \int_0^{1/\alpha} dy_n \cdots \int_0^{y_2} dy_1 \exp \sum_{j=1}^n (\beta_j y_j), \tag{7}$$

Then

$$\begin{aligned} D_n^\theta(\alpha) &= (\partial_1^\mu \cdots \partial_n^\mu)\partial_1(\partial_2 - \partial_1)\cdots(\partial_n - \partial_{n-1}) \\ &\quad \times J(\beta_1, \cdots, \beta_n)|_{\beta=\lambda}, \end{aligned} \tag{8a}$$

where $\mu \equiv m - 4$, $\partial_j \equiv \partial/\partial\beta_j$, and one eventually sets all the $\beta_j = \lambda$. One finds similarly that

$$\begin{aligned} N_n^\theta(\alpha) &= (\partial_1^\mu \cdots \partial_n^\mu)(\partial_2 - \partial_1)\cdots(\partial_n - \partial_{n-1}) \\ &\quad \times J(\beta_1, \cdots, \beta_n)|_{\beta=\lambda}. \end{aligned} \tag{8b}$$

For the + regularization one finds

$$\begin{aligned} D_n^+(\alpha) &= \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n (r_1 - r_2)\cdots(r_{n-1} - r_n) \\ &\quad \times r_n V(r_1 + \alpha)\cdots V(r_n + \alpha) \\ &= \int_\alpha^\infty dx_1 \cdots \int_\alpha^{x_{n-1}} dx_n (x_1 - x_2)\cdots(x_{n-1} - x_n) \\ &\quad \times (x_n - \alpha)V(x_1)\cdots V(x_n), \end{aligned} \tag{9}$$

where the variable changes $x_j = r_j + \alpha$ $(j = 1, \cdots, n)$ have been made. Following through with the variable changes $x_j = y_j^{-1}$ $(j = 1 \cdots n)$, one readily finds

$$\begin{aligned} D_n^+(\alpha) &= (\partial_1^\mu \cdots \partial_n^\mu)\partial_1(\partial_2 - \partial_1)\cdots(\partial_n - \partial_{n-1}) \\ &\quad \times (1 - \alpha\partial_n)J(\beta_1, \cdots, \beta_n)|_{\beta=\lambda} \end{aligned} \tag{10a}$$

and

$$N_n^+(\alpha) = (\partial_1^\mu \cdots \partial_n^\mu)(1 - \alpha\partial_1)(\partial_2 - \partial_1)\cdots(\partial_n - \partial_{n-1}) \times (1 - \alpha\partial_n)J(\beta_1, \cdots, \beta_n)|_{\beta=\lambda}. \quad (10b)$$

We work with quantities $D_n^\gamma(\alpha:\beta_1, \cdots, \beta_n) \equiv D_n^\gamma(\alpha:\beta)$ and $N_n^\gamma(\alpha:\beta_1, \cdots, \beta_n) \equiv N_n^\gamma(\alpha:\beta)$, which are merely the appropriate expressions on the right side of Eqs. (8a, b) and (10a, b) without the $\beta_j = \lambda$ substitution.

III. PERATIZATION OF $N(g, \alpha)$ AND $D(g, \alpha)$

We proceed to peratize $A(g, \alpha)$ by peratizing $N(g, \alpha)$ and $D(g, \alpha)$ separately and evaluating the ratio. The appropriate prescription for doing this so that it is equivalent to the peratization of the Born series is discussed in the course of the article.

One readily recognizes that the integral $J(\beta_1, \cdots, \beta_n)$ of Eq. (7), when evaluated at the point $\beta_1 = \beta_2 = \cdots \beta_n = \lambda$, results in an nth-degree polynomial in the variable $e^{\lambda/\alpha}$. The same set of terms appear in the $D_n^\gamma(\alpha)$ and $N_n^\gamma(\alpha)$ ($\gamma \equiv \theta, +$) with each power of $e^{\lambda/\alpha}$ multiplied by some polynomial (with positive and negative exponents) in $1/\alpha$. The most singular exponential dependence contains the term $e^{n\lambda/\alpha}$, and throughout the larger portion of this work our peratization retains only terms connected with this factor. We cast out all powers of $e^{\lambda/\alpha}$ of degree less than n through the simple device of setting all the lower limits in the integrals in Eq. (7) equal to $-\infty$. We denote this "exponentially peratized" integral of Eq. (7) by $\bar{J}(\beta_1, \cdots, \beta_n)$:

$$\bar{J}(\beta_1, \cdots, \beta_n) = \int_{-\infty}^{1/\alpha} dy_n \cdots \int_{-\infty}^{y_2} dy_1 \exp \sum_{j=1}^{n} (\beta_j y_j) = \frac{1}{\beta_1(\beta_1 + \beta_2)\cdots(\beta_1 + \beta_2 + \cdots \beta_n)} e^{1/\alpha \Sigma \beta_j} \equiv P(\beta_1, \cdots, \beta_n) \exp\left(\frac{1}{\alpha}\sum \beta_j\right). \quad (11)$$

The quantities $D_n^\gamma(\alpha)$ and $N_n^\gamma(\alpha)$ ($\gamma = \theta, +1$) are to be understood as being "exponentially peratized" for the remainder of this work, unless it is otherwise stated. They are calculated through Eqs. (8a, b) and (10a, b) from $\bar{J}(\beta_1, \cdots, \beta_n)$. One can now proceed to peratize the polynomial (in $1/\alpha$) factor of $\exp(n\lambda/\alpha)$ in the $D_n^\gamma(\alpha)$ and $N_n^\gamma(\alpha)$, through selective retention of terms of the higher degrees in $1/\alpha$. We define "first, second, etc., peratization" as the procedure for retaining the most, second most, etc., singular terms in the polynomial factor. These peratizations are algebraically feasible, and we explicitly perform the first three.

The following identity is helpful:

$$(\partial_2 - \partial_1)\cdots(\partial_n - \partial_{n-1})P(\beta_1, \cdots, \beta_n)e^{1/\alpha\Sigma\beta_j} = \left(\sum_{j=1}^{n} \beta_j\right) P^2(\beta_1, \cdots, \beta_n)e^{1/\alpha\Sigma\beta_j}. \quad (12)$$

The sth-order peratization contribution to the quantities $D_n^\gamma(\alpha:\beta)$ and $N_n^\gamma(\alpha:\beta)$ appropriate to the exponential potential $r^{-\mu-4}\exp(\lambda/r)$ are denoted in a changed notation respectively as $D_{\gamma\mu}^{(s)}(n, \alpha:\beta)$, $N_{\gamma\mu}^{(s)}(n, \alpha:\beta)$. The absence of the superscript (s) denotes the quantities with no further peratization than the "exponential peratization" which has been done in all cases [corresponding to setting the lower integration limits on Eq. (7) equal to $-\infty$]. Indices or arguments will frequently be dropped when confusion is unlikely. Simple calculation gives for $D_{\gamma\mu}^{(s)}(n, \alpha:\beta) = D_{\gamma\mu}^{(s)}(\beta)$, $N_{\gamma\mu}^{(s)}(n, \alpha:\beta) = N_{\gamma\mu}^{(s)}(\beta)$,

$$D_{\theta 0}(\beta) = [\alpha^{-1}\textstyle\sum\beta + 1 - 2(\sum\beta)H(\beta)]M(\beta), \quad (13a)$$

$$N_{\theta 0}(\beta) = (\textstyle\sum\beta)M(\beta), \quad (13b)$$

$$D_{+0}(\beta) = [1 - 2\alpha H(\beta)]M(\beta), \quad (14a)$$

$$N_{+0}(\beta) = 2\alpha^2 H(\beta)M(\beta), \quad (14b)$$

where $\sum\beta = \sum_{j=1}^{n}\beta_j$,

$$M(\beta) = P^2(\beta)e^{1/\alpha(\Sigma\beta)} = \frac{1}{\beta_1^2(\beta_1 + \beta_2)^2\cdots(\beta_1 + \beta_2 + \cdots \beta_n)^2} \times \exp\frac{1}{\alpha}(\textstyle\sum\beta), \quad (15a)$$

$$H(\beta) = \frac{1}{\beta_1} + \frac{1}{\beta_1 + \beta_2} + \cdots \frac{1}{\beta_1 + \beta_2 + \cdots \beta_n}. \quad (15b)$$

Proceeding from the relations

$$D_{\gamma\mu}(\beta) = (\partial_1^\mu \cdots \partial_n^\mu)D_{\gamma 0}(\beta), \quad N_{\gamma\mu}(\beta) = (\partial_1^\mu \cdots \partial_n^\mu)N_{\gamma 0}(\beta), \quad (16)$$

we find

$$D_{\theta\mu}^{(1)}(\beta) = \alpha^{-n\mu-1}(\textstyle\sum\beta)M(\beta), \quad (17a)$$

$$D_{\theta\mu}^{(2)}(\beta) = \alpha^{-n\mu}[n\mu + 1 - 2(\textstyle\sum\beta)H(\beta) - 2\mu(\sum\beta)K(\beta)]M(\beta), \quad (17b)$$

$$N_{\theta\mu}^{(1)}(\beta) = \alpha^{-n\mu}(\textstyle\sum\beta)M(\beta), \quad (17c)$$

$$N_{\theta\mu}^{(2)}(\beta) = \mu\alpha^{-n\mu+1}[n - 2(\textstyle\sum\beta)K(\beta)]M(\beta), \quad (17d)$$

$$D_{+\mu}^{(1)}(\beta) = \alpha^{-n\mu}M(\beta), \quad (17e)$$

$$D_{+\mu}^{(2)}(\beta) = -2\alpha^{-n\mu+1}[H(\beta) + \mu K(\beta)]M(\beta), \quad (17f)$$

$$N_{+\mu}^{(1)}(\beta) = 2\alpha^{-n\mu+2}H(\beta)M(\beta), \quad (17g)$$

$$N_{+\mu}^{(2)}(\beta) = -2\mu\alpha^{-n\mu+3}[L(\beta) + 2H(\beta)K(\beta)]M(\beta), \quad (17h)$$

where

$$K(\beta) \equiv \sum_{j=1}^{n} \frac{j}{\beta_1 + \beta_2 + \cdots + \beta_j}, \quad L(\beta) \equiv \sum_{j=1}^{n} \frac{j}{(\beta_1 + \beta_2 + \cdots + \beta_j)^2}. \quad (18)$$

The quantities

$$D_{\gamma\mu}^{(s)} \equiv D_{\gamma\mu}^{(s)}(\beta)|_{\beta=\lambda}$$

found from the substitution $\beta_j = \lambda$ yield the desired coefficients for the appropriately peratized power series from which one calculates the scattering length. The calculations of $N^{(3)}_{\gamma\mu}(\beta)$ and $D^{(3)}_{\gamma\mu}(\beta)$ are outlined in Appendix C, and are not listed owing to their unwieldy form. Using the notation

$$H(\beta)|_{\beta=\lambda} = \lambda L(\beta)|_{\beta=\lambda} = \frac{1}{\lambda}\sum_{j=1}^{n}\frac{1}{j} \equiv \frac{1}{\lambda}H_n \tag{19}$$

and the results in Appendix C, one readily finds

$$[x_0 \equiv 2\lambda^{-1}\alpha^{-\mu/2}\exp(\lambda/2\alpha)]:$$

$$D^{(1)}_{\theta\mu} = \frac{\lambda n}{\alpha}\frac{(x_0/2)^{2n}}{n!^2}, \tag{20a}$$

$$D^{(2)}_{\theta\mu} = (1 + n\mu - 2nH_n - 2\mu n^2)\frac{(x_0/2)^{2n}}{n!^2}, \tag{20b}$$

$$D^{(3)}_{\theta\mu} = \frac{2\mu\alpha}{\lambda}n[\mu n^2 - (\tfrac{1}{2}\mu + 2)n + (2n + \tfrac{1}{2})H_n]\frac{(x_0/2)^{2n}}{n!^2}, \tag{20c}$$

$$N^{(1)}_{\theta\mu} = \lambda n\frac{(x_0/2)^{2n}}{n!^2}, \tag{20d}$$

$$N^{(2)}_{\theta\mu} = \mu\alpha n(1 - 2n)\frac{(x_0/2)^{2n}}{n!^2}, \tag{20e}$$

$$N^{(3)}_{\theta\mu} = \frac{2\mu\alpha^2}{\lambda}n[\mu n^2 - (\tfrac{1}{2}\mu + 2)n + \tfrac{1}{2}H_n + 1]\frac{(x_0/2)^{2n}}{n!^2}, \tag{20f}$$

$$D^{(1)}_{+\mu} = \frac{(x_0/2)^{2n}}{n!^2}, \tag{20g}$$

$$D^{(2)}_{+\mu} = -\frac{2\alpha}{\lambda}(H_n + \mu n)\frac{(x_0/2)^{2n}}{n!^2}, \tag{20h}$$

$$D^{(3)}_{+\mu} = \frac{2\mu\alpha^2}{\lambda^2}[\mu n^2 + 2nH_n + (\tfrac{1}{2}\mu - 2)n + \tfrac{3}{2}H_n]\frac{(x_0/2)^{2n}}{n!^2}, \tag{20i}$$

$$N^{(1)}_{+\mu} = \frac{2\alpha^2}{\lambda}H_n\frac{(x_0/2)^{2n}}{n!^2}, \tag{20j}$$

$$N^{(2)}_{+\mu} = -\frac{2\mu\alpha^3}{\lambda^2}(2n + 1)H_n\frac{(x_0/2)^{2n}}{n!^2}, \tag{20k}$$

$$N^{(3)}_{+\mu} = \frac{4\mu\alpha^4}{\lambda^3}H_n[\mu n^2 + (\tfrac{3}{2}\mu - 2)n + \tfrac{1}{2}\mu - 1]\frac{(x_0/2)^{2n}}{n!^2}. \tag{20l}$$

The ubiquitous $(n!)^{-2}$ is to be noted. Its appearance follows from very general principles, and its significance is indicated in the following section. Note also that some of the second peratized and all of the third peratized quantities vanish when $\mu = 0$. This is due to the small number of terms in the polynomial coefficient of the leading exponential factor for $m = 4$, reflecting the especially simplifying features of that potential.

The first peratized scattering length can be found readily by summing the respective $N^{(1)}_{\gamma\mu}(n, \alpha)$ and $D^{(1)}_{\gamma\mu}(n, \alpha)$. These summations give rise to the Bessel functions of imaginary argument,[9] $I_0(z)$ and $K_0(z)$, where $I_0(z)$ is the regular solution normalized by $I_0(0) = 1$, and $K_0(z)$ has exponential decay for large positive z and is normalized by $K_0(z) \approx (\pi/2z)^{\frac{1}{2}} \times \exp(-z)$ for large z. One finds ($x = g^{\frac{1}{2}}x_0$)

$$A^{(1)}_{\theta\mu}(\alpha) = -\alpha, \tag{21a}$$

$$A^{(1)}_{+\mu}(\alpha) = -\frac{2\alpha^2}{\lambda}\frac{K_0(x) + (\ln\tfrac{1}{2}x + \gamma)I_0(x)}{I_0(x)}, \tag{21b}$$

where $\gamma = 0.5772$ is the Euler–Mascheroni constant. For both $\mu = 0$ and $\mu > 0$, $A^{(1)}$ vanishes as $\alpha \to 0$.

It is of interest to note that Cornille[7a] has calculated the first peratized scattering length for the general class of potentials

$$V(r) = gr^{-m}e^{\lambda/r^s},$$
$$m > 3, \quad s, \lambda > 0,$$

and has obtained zero in all cases.

One might be tempted to calculate the higher-order peratized scattering lengths in a similar way. However, the separate expansion of the numerator and denominator to a given order of peratization results in a Born series which would contain the correct contribution to that order of peratization plus some additional contributions from a higher order of peratization. One might expect that the presence of these additional terms of higher order would not be problematic if the sequence of peratizations of the Born series converges. However, such an expectation would not be vindicated if the additional higher-order terms were large. This is in fact frequently the case with these additional higher-order terms. This situation can be analyzed if we introduce the optimum-n method, which is a simplifying and labor-saving device in calculations and which also sheds light on the analytic structure of the scattering length itself.

IV. OPTIMUM-n METHOD

Some analytical labor can be dispensed with, and much insight into both the mechanisms and the flaws of peratization can be discerned, if we employ a convenient procedure which we call the "optimum-n" method. This method is capable of giving order of

magnitude estimates in the asymptotic limit for functions of known exponential order and type[10]; in the case of a quotient of two such functions where all coefficients of the power series are positive, the method gives the exact asymptotic limit. The scattering length in fact has been expressed as the ratio of two entire functions of g, $N(g)$ and $D(g)$, with positive coefficients. In the case of scattering by regular potentials, the exponential order and type[10] of these functions are known.[11–13] The optimum-n method will be described in detail in a subsequent article. We presently outline the leading ideas.

We consider an entire function $f(z)$ of finite order. A theorem in the theory of entire functions due to Borel[14] states that, for large $r = |z|$,

$$\ln \mu(r) \sim \ln M(r), \tag{22}$$

where

$$M(r) \equiv \max_{|z|=r} |f(r)|, \tag{23a}$$

and

$$\mu(r) \equiv \max_{n} |c_n r^n| \equiv c_{n(r)} r^{n(r)}, \tag{23b}$$

with $f(z) = \sum c_n z^n$. Qualitatively, this theorem states that the maximum modulus of $f(z)$ on a circle of sufficiently large radius r has the same growth (measured in a logarithmic scale) as the maximum term of the series. We denote by $n(r)$ the exponent of the term which dominates for $|z| = r$. (It need not be unique, but for simplicity and with no real loss of generality we shall assume that it is.) For a given value of $r = |z|$, we term $n(r)$ the "optimum-n" of the function $f(z)$ for $|z| = r$. Finer estimates of the relation between $M(r)$, $\mu(r)$, and $n(r)$ can be found in specialized monographs.[15] In particular, if, for large r,

$$\ln M(r) \sim \tau r^{\rho} \tag{24}$$

(τ being the type), then[16] one has

$$n(r) \sim \rho \tau r^{\rho}. \tag{25}$$

This result is readily derived heuristically, if, in the spirit of Borel's theorem, one writes $M(r) \sim \mu(r)$. If all the coefficients of the power series are positive, then one finds $f(r) = M(r)$ for positive r. The additive error term in Eq. (22) corresponds to a multiplicative error factor in the ratio $M(r)/\mu(r)$. However, in evaluating the ratio of two entire power series with positive coefficients and the same exponential order for positive r, as in the case of the scattering length, N/D, or certain Green's functions, the ansatz

$$M(r) \sim \mu(r) = c_{n(r)} r^{n(r)} \tag{26}$$

for each power series turns out to give the correct asymptotic limit. We call this the "optimum-n" method.

In the case of scattering by regular potentials, the scattering length, given by Eq. (4), is expressed as the ratio of two functions of the coupling constant g, which are entire and of exponential order $\frac{1}{2}$.[11–13] The fact that $\rho = \frac{1}{2}$ for $D(g)$ originates in the $(n!)^{-2}$ factors which are characteristic of D_n and N_n. This is discussed in Appendix B, where it is shown that $\rho = \frac{1}{2}$ for $N(g)$ also. In addition, it is shown there that the type τ for both $N(g)$ and $D(g)$ is given by

$$\tau = \int_0^{\infty} dr[V(r)]^{\frac{1}{2}},$$

if $V(r)$ belongs to a fairly general class of purely repulsive potentials. The coefficients in the N and D power series are all positive. Hence, the optimum-n method prescribes the evaluation of the scattering length for these potentials in the limit of large g to be

$$-A(g) \approx N_{n(g)}/D_{n(g)},$$

where $n(g)$ is found from Eq. (25) to be $n(g) = \rho\tau g^{\rho}$.

The foregoing considerations are also applicable in the case of scattering by singular potentials. One introduces the regularized potential $V(r, \alpha)$ and calculates the N and D functions, now denoted as $N(g, \alpha)$ and $D(g, \alpha)$, in terms of this potential. The exponential order of $N(g, \alpha)$ and $D(g, \alpha)$ as a function of g is still $\frac{1}{2}$. This explains the $g^{\frac{1}{2}}$ singularity of singular potential wavefunctions. The type is given by

$$\tau \equiv \tau(\alpha) = \int_0^{\infty} dr V^{\frac{1}{2}}(r, \alpha) = \int_{\alpha}^{\infty} dr V^{\frac{1}{2}}(r). \tag{27}$$

The last expression on the right holds for either + or θ regularization. If we let $\bar{\tau}(\alpha)$ be the α-dependent part of $\tau(\alpha)$ [note that, in general, $\tau(\alpha)$ contains an additive α-independent part arising from the upper limit of integration in Eq. (27)], $N(g, \alpha)$ and $D(g, \alpha)$ can be expanded as

$$N(g, \alpha) = \sum_{1}^{\infty} \frac{g^n \bar{\tau}(\alpha)^{2n}}{(2n)!} q(n, \alpha), \tag{28a}$$

$$D(g, \alpha) = \sum_{0}^{\infty} \frac{g^n \bar{\tau}(\alpha)^{2n}}{(2n)!} p(n, \alpha), \tag{28b}$$

where $q(n, \alpha)$ and $p(n, \alpha)$ generally have polynomial growth in n. These expressions can be shown to hold for several solvable potentials and can be inferred from the following argument. It can be shown[17] that upper and lower bounds for the Jost solution evaluated at $r = \alpha$ for monotonically decreasing repulsive potentials are given by

$$D(g, \alpha) < \exp g^{\frac{1}{2}}\tau(\alpha),$$

$$D(g, \alpha) > 1 + \frac{V^{\frac{1}{2}}(R)}{V^{\frac{1}{2}}(\alpha)} \sinh^2 \frac{g^{\frac{1}{2}}\tau(\alpha, R)}{2},$$

where R is the value of a cutoff imposed on the

potential in obtaining the lower bound and

$$\tau(\alpha, R) = \int_{\alpha}^{R} dr V^{\frac{1}{2}}(r).$$

One notes that $\tau(\alpha) = \bar{\tau}(\alpha)[1 - \tau_1/\bar{\tau}(\alpha)]$ and $\tau(\alpha, R) = \bar{\tau}(\alpha)[1 - \tau_2/\bar{\tau}(\alpha)]$, where τ_1 and τ_2 denote the α-independent parts of $\tau(\alpha)$ and $\tau(\alpha, R)$, respectively. Hence the product $g^{\frac{1}{2}}\bar{\tau}(\alpha)$ is common for both upper and lower bounds. The discussion of Appendix B shows that similar results hold for $N(g, \alpha)$.

The optimum-n method can now be applied if one chooses as the expansion parameter of $N(g, \alpha)$ and $D(g, \alpha)$ not simply g, but rather $g\bar{\tau}^2(\alpha)$. Then for α near zero, one is effectively in the strong coupling regime, where the estimates of Eqs. (22), (23), (25), and (26) are appropriate. The exponential order of N and D as functions of the new variable $g\bar{\tau}^2(\alpha)$ is still $\rho = \frac{1}{2}$ and the type can be found from the expression[18]

$$\tau = \limsup_{n\to\infty} [(2n)!\, c_n]^{1/(2n)}, \tag{29}$$

where c_n represents either coefficient $q(n, \alpha)$ or $p(n, \alpha)$ in the expansion of $N(g, \alpha)$ or $D(g, \alpha)$. The optimum-n method thus prescribes the evaluation of A by

$$-A = \lim_{\alpha\to 0} q[n(g, \alpha), \alpha]/p[n(g, \alpha), \alpha], \tag{30}$$

where the optimum-n dependence $n(g, \alpha)$ has been explicitly indicated in the argument of $p(n, \alpha)$ and $q(n, \alpha)$ and can be found from Eq. (25) with $\rho = \frac{1}{2}$ and τ as given by Eq. (29). One notes that a common factor $g^n\bar{\tau}^{2n}$ in the numerator and denominator of Eq. (30) has been cancelled.

We illustrate the method by calculating the first peratized scattering length for the potential $V(r) = gr^{-m}e^{\lambda/r}$. From Eqs. (20) it is clear that the proper expansion variable $g\bar{\tau}^2(\alpha)$ is $gx_0^2(\alpha) \equiv x^2(\alpha)$. In addition we find from Eqs. (20) the type of N and D as a function of $x^2(\alpha)$ to be unity. The optimum-n value $n(g, \alpha)$ is therefore

$$\frac{g^{\frac{1}{2}}}{\lambda}\alpha^{-n/2}e^{\lambda/2\alpha}.$$

Noting that[19]

$$H_n \equiv \sum_{j=1}^{n} j^{-1} = \psi(n + 1) + \gamma \approx \ln n + \gamma$$

for large n, we find the first peratized scattering lengths to be

$$-A_{+\mu}^{(1)}(\alpha) = \frac{N_{+\mu}^{(1)}[n(g, \alpha), \alpha]}{D_{+\mu}^{(1)}[n(g, \alpha), \alpha]} = \frac{2\alpha^2}{\lambda} H_{n(g,\alpha)}$$
$$= \frac{2\alpha^2}{\lambda}\left(\ln \frac{g^{\frac{1}{2}}}{\lambda}\alpha^{-\mu/2} + \frac{\lambda}{2\alpha} + \gamma\right), \tag{31a}$$

$$-A_{\theta\mu}^{(1)}(\alpha) = \frac{N_{\theta\mu}^{(1)}[n(g, \alpha), \alpha]}{D_{\theta\mu}^{(1)}[n(g, \alpha), \alpha]} = \alpha, \tag{31b}$$

and in the $\alpha \to 0$ limit both of these vanish, in accordance with the results obtained above by explicitly summing the series.

V. PERATIZATION OF A RATIO

We now use the optimum-n method to analyze the separate peratization of numerator and denominator of a ratio. According to the prescriptions of this paper, one expands both numerator and denominator in terms of singular basis functions, keeps the first s leading singularities, and then takes the ratio to obtain the sth peratized scattering length. An important consideration for the peratization of a ratio is that the numerator and denominator contain no "spurious" singular basis functions. This point is discussed at length in Appendix A.

It is generally the case that the singular basis functions which appear in calculations are to be found from the peratization of the factor $\tau(\alpha)$ or $\bar{\tau}(\alpha)$. Thus, from Eq. (28) we see that the sth peratization of N or D may be obtained from the product

$$g^n(\bar{\tau}_1 + \bar{\tau}_2 + \cdots + \bar{\tau}_s)^{2n}(q_1 + q_2 + \cdots + q_s), \tag{32}$$

where the summation $\sum_1^s \bar{\tau}_i \equiv \bar{\tau}^{(s)}$ is simply $\bar{\tau}$ s-peratized and the summation $\sum q_i$ is an expansion of of the factor $q(n, \alpha)$ in appropriate basis functions (in α). Note that $\bar{\tau}_k$ and q_k are functions of α and q_k is a function of n (these arguments will frequently be suppressed in the interest of simplification of notation). In addition, the q_k are generally not singular in α. However, in the actual calculation of the sth peratized N and D, one does not obtain the complete product of Eq. (32) which contains higher-order terms in α, but only the first s leading singularities. For example, $D^{(1)}$ is written

$$D^{(1)} = g^n\bar{\tau}_1^{2n}p_1, \tag{33a}$$

while $D^{(2)}$ so obtained is given by

$$D^{(2)} = g^n(2n\bar{\tau}_1^{2n-1}\bar{\tau}_2 p_1 + \bar{\tau}_1^{2n}p_2). \tag{33b}$$

Similar expressions hold for $N^{(s)}$ also. The first peratized scattering length is given by the optimum-n method as

$$-A^{(1)} = \lim_{\alpha\to 0} q_1[n(g, \alpha), \alpha]/p_1[n(g, \alpha), \alpha], \tag{34}$$

where the explicit n dependence of the p and q factors has been exhibited. Note that the common factor $g^n\bar{\tau}_1^{2n}(\alpha)$ has been cancelled in Eq. (34).

Using Eqs. (33a) and (33b) for $D^{(1)}$ and $D^{(2)}$ and analogous expressions for $N^{(1)}$ and $N^{(2)}$, one might try to calculate the second peratized scattering

length $A^{(2)}$ from

$$-A^{(2)}(g) = \lim_{\alpha\to 0} \frac{N^{(1)}[n(g,\alpha),\alpha] + N^{(2)}[n(g,\alpha),\alpha]}{D^{(1)}[n(g,\alpha),\alpha] + D^{(2)}[n(g,\alpha),\alpha]} = \lim_{\alpha\to 0} \frac{(g\bar{\tau}_1^2)^n(q_1 + 2n\bar{\tau}_1^{-1}\bar{\tau}_2 q_1 + q_2)}{(g\bar{\tau}_1^2)^n(p_1 + 2n\bar{\tau}_1^{-1}\bar{\tau}_2 p_1 + p_2)}. \quad (35)$$

The exponential order of N and D peratized s times as a function of $g\bar{\tau}_1^2$ is $\frac{1}{2}$ and the type is generally unity, so that the optimum-n value computed from Eq. (25) is simply $n(g,\alpha) = \frac{1}{2}g^{\frac{1}{2}}\bar{\tau}_1(\alpha)$. Thus we have

$$-A^{(2)} = \lim_{\alpha\to 0} (q_1 + g^{\frac{1}{2}}\bar{\tau}_2 q_1 + q_2)/(p_1 + g^{\frac{1}{2}}\bar{\tau}_2 p_1 + p_2). \quad (36)$$

It is usually true that the $\bar{\tau}_k$ do not approach a constant as $\alpha \to 0$; i.e., $\bar{\tau}_k$ is either infinite, as will be seen for the potential $gr^{-m}e^{\lambda/r}$ with $m > 4$, or approaches zero as for the case with $m = 4$ ($\bar{\tau}_1$, of course, is infinite). If $\bar{\tau}_2 \to 0$, then $A^{(2)}$ becomes

$$-A^{(2)} = \lim_{\alpha\to 0} \frac{q_1[n(g,\alpha),\alpha] + q_2[n(g,\alpha),\alpha]}{p_1[n(g,\alpha),\alpha] + p_2[n(g,\alpha),\alpha]}. \quad (37)$$

However, if $\bar{\tau}_2 \to \infty$, one finds that

$$-A^{(2)} = \lim_{\alpha\to 0} \frac{q_1[n(g,\alpha),\alpha]}{p_1[n(g,\alpha),\alpha]}, \quad (38)$$

which is the same as $A^{(1)}$, so that second peratization in this case appears to give no improvement. This result, however, is in contradiction with a solvable example presented in Appendix D, for which second peratization in terms of the Born series does indeed give an improved value for the scattering length.[4]

The root of this discrepancy lies in the method just employed of the separate peratization of the numerator and denominator of a ratio. Upon forming the quotient to obtain the Born series, one finds that the higher-order terms that are present in the quotient can be effectively large. This can be illustrated easily by Eq. (36). If peratization does succeed, one would expect higher peratizations to be in some sense corrections to first peratization. Hence one should be able to expand the denominator of Eq. (36) to obtain

$$-A^{(2)} = \lim_{\alpha\to 0} \frac{q_1}{p_1}\left[1 + \frac{q_2}{q_1} - \frac{p_2}{p_1} - g^{\frac{1}{2}}\bar{\tau}_2\left(\frac{p_2}{p_1} + \frac{q_2}{q_1}\right) + \cdots\right]$$

As $\alpha \to 0$, $A^{(2)}$ becomes infinite because of the effect of the divergent quantity $\bar{\tau}_2$. However a careful scrutiny of Eqs. (28) and (32) suggests a flaw in our calculation of $A^{(2)}$. The exact N and D functions corresponding to the regularized potential should be expanded in terms of the variable $g\bar{\tau}^2(\alpha)$ as appropriate for optimum n and it was subsequently suggested that the sth peratized N_n and D_n can be obtained from the product $g^n(\sum_1^s \bar{\tau}_k)^{2n}(\sum_1^s q_k)$. This situation is a strong indication that the appropriate expansion parameter for the peratized N_n and D_n ought to be $g[\bar{\tau}^{(s)}(\alpha)]^2$ rather than $g\bar{\tau}_1^2 = g(\bar{\tau}^{(1)})^2$, which has just been used. Thus, in order to introduce this variable in second peratization, it is necessary to add to the numerator and denominator of Eq. (35) the terms $2n\bar{\tau}_1^{2n-1}\bar{\tau}_2 q_2$ and $2n\bar{\tau}_1^{2n-1}\bar{\tau}_2 p_2$, respectively. These terms are less singular in α than the terms which have been retained, but they would appear together with other terms of the same degree of singularity in higher-order peratizations. Thus we have, in place of Eq. (35), the expression

$$-A^{(2)} = \lim_{\alpha\to 0} \frac{g^n(\bar{\tau}_1^{2n} + 2n\bar{\tau}_1^{2n-1}\bar{\tau}_2)(q_1 + q_2)}{g^n(\bar{\tau}_1^{2n} + 2n\bar{\tau}_1^{2n-1}\bar{\tau}_2)(p_1 + p_2)}, \quad (39)$$

and the undermining effect of the term $\bar{\tau}_2$ discussed above is eliminated. We note that the common factor in this expression can be written

$$\bar{\tau}_1^{2n} + 2n\bar{\tau}_1^{2n-1}\bar{\tau}_2 = [(\bar{\tau}_1 + \bar{\tau}_2)^{2n}]^{(2)} = (\bar{\tau}^{(2)})^{2n},$$

and hence the variable of expansion is effectively $g(\bar{\tau}^{(s)})^2$ with $s = 2$ in this case. Note that the second and third members of the above expression are equal only to second peratization. We call the above procedure the "completion of the τ factor."

The foregoing development was carried out within the context of the optimum-n method. However, it is clear at this point that the completion of the τ factor is also required and follows through readily when the series in the numerator and denominator are explicitly summed. When one completes the τ factor to a given order of peratization, the numerator, denominator, and Born series are automatically expressed in terms of the quantity $g[\tau^{(s)}(\alpha)]^2$ as the expansion parameter which diverges in the $\alpha \to 0$ limit. The ratio of the numerator and denominator so peratized contains additional terms of higher order of peratization in the Born series, but these are finite for the classes of potentials of interest. These would be expected to converge if the sequence of successively peratized Born approximations does. The general method described here and applied to second peratization also applies in higher orders of peratization. These considerations are well illustrated in a nontrivial way in Appendix D with an exactly solvable potential.

We note in passing that the procedure of completing the τ factor is automatic in first peratization inasmuch as the first peratized $\bar{\tau}(\alpha)$ is used anyway.

Thus the preceding calculations of Secs. III and IV for the first peratized scattering length for the potential $V(r) = gr^{-m}e^{\lambda/r}$ stand unchanged.

VI. HIGHER-ORDER PERATIZATIONS

The procedure of completing the τ factor now makes it possible to calculate the higher-order peratizations for the potential $V(r) = gr^{-m}e^{\lambda/r}$. The primary quantity of interest, $\bar{\tau}(\alpha)$, is obtained from $\tau(\alpha)$ and from Eq. (27) we find

$$\tau(\alpha) = \int_\alpha^\infty dr\, \frac{e^{\lambda/2\alpha}}{r^{m/2}} = \frac{2}{\lambda}\begin{cases}(e^{\lambda/2\alpha} - 1), & \mu = 0,\\ \alpha^{-\mu/2}e^{\lambda/2\alpha}[1 + O(\alpha)], & \mu > 0.\end{cases} \tag{40}$$

For $m = 4$ ($\mu = m - 4$), $\bar{\tau}_1(\alpha)$ alone is nonvanishing and peratization does not require the completion of the τ factor, but, for $m > 4$, $\bar{\tau}_2(\alpha)$ certainly diverges and hence in this case it is necessary to complete the τ factor.

We first calculate the second peratized scattering length for the case $m = 4$, the potential discussed by CC. The optimum-n value is readily found to be

$$n(g, \alpha) = \frac{g^{\frac{1}{2}}}{\lambda}\, e^{\lambda/2\alpha}.$$

Denoting the contribution of the first two peratizations by $A^{(2)}$ (N.B.: $A^{(s)}$ denotes the cumulative contribution of the first s peratizations, unlike the parallel definition for $N^{(s)}$ and $D^{(s)}$), one finds from Eqs. (20)

$$-A^{(2)}_{+0}(\alpha) \sim \frac{(2\alpha^2/\lambda)H_{n(g,\alpha)}}{1 - (2\alpha/\lambda)H_{n(g,\alpha)}} = \frac{(2\alpha^2/\lambda)[\ln (g^{\frac{1}{2}}/\lambda)e^{\lambda/2\alpha} + \gamma]}{1 - (2\alpha/\lambda)[\ln (g^{\frac{1}{2}}/\lambda)e^{\lambda/2\alpha} + \gamma]}, \tag{41a}$$

$$-A^{(2)}_{\theta 0}(\alpha) \sim \frac{\lambda n(g, \alpha)}{[\lambda n(g, \alpha)/\alpha] + 1 - 2n(g, \alpha)H_{n(g,\alpha)}} = \lambda\left[\frac{\lambda}{\alpha} + \frac{\lambda}{g^{\frac{1}{2}}}e^{-\lambda/2\alpha} - 2\left(\ln\frac{g^{\frac{1}{2}}}{\lambda}e^{\lambda/2\alpha} + \gamma\right)\right]^{-1}. \tag{41b}$$

In the $\alpha \to 0$ limit, one recovers from both regularizations the CC result[20] (which they derived only for the + regularization) that

$$A^{(2)} = \frac{\frac{1}{2}\lambda}{\frac{1}{2}\ln (g/\lambda^2) + \gamma}, \tag{42}$$

which, though not the exact answer, gives the correct weak-coupling limit.

This result also follows from the more elementary method of summing the series for N and D. One can readily verify that

$$A^{(2)}_{+0}(\alpha) = \alpha\frac{K_0(x) + (\ln \tfrac{1}{2}x + \gamma)I_0(x)}{[\ln \tfrac{1}{2}x - (\lambda/2\alpha) + \gamma]I_0(x) + K_0(x)}, \tag{43a}$$

$$A^{(2)}_{\theta 0}(\alpha) = \frac{-\frac{1}{2}\lambda x I_0'(x)}{[(\lambda/2\alpha) - \gamma - \ln \tfrac{1}{2}x]xI_0'(x) - xK_0'(x)}, \tag{43b}$$

where $x = g^{\frac{1}{2}}\lambda^{-1}e^{\lambda/2\alpha}$. In the $\alpha \to 0$ limit one again obtains Eq. (42).

Next we calculate both the second and third peratized scattering lengths for the case with $\mu > 0$. Since $\bar{\tau}_2(\alpha)$ was shown above to diverge as $\alpha \to 0$, we must complete the τ factor. Up to the third order of peratization, one finds that

$$\bar{\tau}^{(3)}(\alpha) = \frac{2}{\lambda}\alpha^{-\mu/2}e^{\lambda/2\alpha}\left(1 - \frac{\mu\alpha}{\lambda} + \frac{\mu(\mu - 2)}{\lambda^2}\alpha^2\right). \tag{44}$$

The second peratized scattering length is to be computed from the expression (in, say, + regularization)

$$-A^{(2)}_{+\mu}(\alpha) = \frac{[g\bar{\tau}_1(\alpha)^2]^n 2\alpha^2\lambda^{-1}\{H_{n(g,\alpha)} - \mu\alpha\lambda^{-1}[2n(g, \alpha) + 1]H_{n(g,\alpha)}\}}{[g\bar{\tau}_1(\alpha)^2]^n\{1 - 2\alpha\lambda^{-1}[H_{n(g,\alpha)} + \mu n(g, \alpha)]\}}, \tag{45}$$

where $n(g, \alpha)$, the optimum-n value, is given by

$$n(g, \alpha) = (g^{\frac{1}{2}}/\lambda)\alpha^{-\mu/2}e^{\lambda/2\alpha} + \text{"higher peratization terms."}$$

One notes in the following that the higher peratization terms in $n(g, \alpha)$ do not alter the results. To complete the τ factor, it is necessary to take for the expansion parameter in Eq. (45) $g(\bar{\tau}^{(2)})^2 = g(\bar{\tau}_1 + \bar{\tau}_2)^2$. Noting from Eq. (44) that $(\bar{\tau}^{(2)})^{2n} = \bar{\tau}_1^{2n}(1 - 2n\lambda^{-1}\mu\alpha)$ in second peratization, we may rewrite Eq. (45) in the form

$$-A^{(2)}_{+\mu}(\alpha) = \frac{2\alpha^2\lambda^{-1}[1 - 2n(g, \alpha)\lambda^{-1}\mu\alpha](1 - \lambda^{-1}\mu\alpha)H_{n(g,\alpha)}}{[1 - 2n(g, \alpha)\lambda^{-1}\mu\alpha](1 - 2\alpha\lambda^{-1}H_{n(g,\alpha)})} \tag{46}$$

in which the higher-order terms $(4n\mu^2/\lambda^3)H_n\alpha^4$ and $(4n\mu/\lambda^2)H_n\alpha^2$ were added to N_n and D_n, respectively, to complete the τ factor. After cancelling the common

factor $(1 - 2n\lambda^{-1}\mu\alpha)$, we obtain

$$-A^{(2)}_{+\mu}(\alpha)$$
$$= \frac{2\alpha^2\lambda^{-1}(1 - \lambda^{-1}\mu\alpha)H_{n(g,\alpha)}}{1 - 2\alpha\lambda^{-1}H_{n(g,\alpha)}}$$
$$\approx \frac{2\alpha^2}{\lambda}\frac{(1 - \lambda^{-1}\mu\alpha)[\ln(g^{\frac{1}{2}}/\lambda) - \frac{1}{2}\mu\ln\alpha + (\lambda/2\alpha) + \gamma]}{1 - 2\alpha\lambda^{-1}[\ln(g^{\frac{1}{2}}/\lambda) - \frac{1}{2}\mu\ln\alpha + (\lambda/2\alpha) + \gamma]}$$
$$\underset{\alpha\to 0}{\longrightarrow} 0.$$

In θ regularization we have similarly

$$-A^{(2)}_{\theta\mu} = \frac{\lambda n + \mu\alpha n(1 - 2n)}{\lambda n\alpha^{-1} + 1 + n\mu - 2nH_n - 2\mu n^2},$$

and after completing the τ factor we find

$$-A^{(2)}_{\theta\mu}$$
$$= \frac{(1 - 2n\lambda^{-1}\mu\alpha)\lambda n(1 + \lambda^{-1}\alpha\mu)}{\alpha^{-1}\lambda n(1 - 2n\lambda^{-1}\mu\alpha)[1 + (\alpha/\lambda n) + \lambda^{-1}\mu\alpha - 2\alpha\lambda^{-1}H_n]}$$
$$\underset{\alpha\to 0}{\longrightarrow} 0.$$

Thus the second peratized scattering length for $\mu > 0$ is exactly zero.[21] One concludes that the success of peratization for the potential $V(r) = gr^{-4}e^{\lambda/r}$ and presumably for its companion $V'(r) = r^{-4}(ge^{\lambda/r} + g')$ seems to be a fluke and not attributable to the exponential character of the singularity.

The foregoing result can also be obtained by summing the power series. The nth term of N and D can be found from Eqs. (20) after completion of the τ factor which yields the appropriate expansion variable, given by $g(\bar{\tau}^{(2)})^2 = 4g\lambda^{-2}\alpha^{-\mu}e^{\lambda/\alpha}(1 - \lambda^{-1}\alpha\mu)^2$. If we set $x/2 = g^{\frac{1}{2}}\bar{\tau}^{(2)}$, one readily finds (in second peratization and, say, + regularization) that

$$N^+(g,\alpha)$$
$$= \frac{2\alpha^2}{\lambda}\sum\frac{(x/2)^{2n}}{n!^2}(1 - \lambda^{-1}\mu\alpha)[\psi(n+1) - \psi(1)]$$
$$= \frac{2\alpha^2}{\lambda}(1 - \lambda^{-1}\mu\alpha)[K_0(x) + I_0(x)(\ln\tfrac{1}{2}x + \gamma)],$$

$$D^+(g,\alpha)$$
$$= \sum\frac{(x/2)^{2n}}{n!^2}[1 - 2\alpha\lambda^{-1}(\psi(n+1) - \psi(1))]$$
$$= (1 - 2\alpha\lambda^{-1}\gamma - 2\alpha\lambda^{-1}\ln\tfrac{1}{2}x)I_0(x) - 2\alpha\lambda^{-1}K_0(x).$$

The second peratized scattering length is given by

$$A^{(2)}_{+\mu} = -\lim_{\alpha\to 0}\frac{2\alpha^2}{\lambda}\left(1 - \frac{\mu\alpha}{\lambda}\right)\frac{\ln\frac{1}{2}x + \gamma}{1 + 2\alpha\lambda^{-1}\gamma - 2\alpha\lambda^{-1}\ln\frac{1}{2}x},$$

and for $\mu > 0$, this becomes zero.

The calculation of the third peratized scattering length for $\mu > 0$ is similar. We consider + regularization using the optimum-n method. We note that

$$[\bar{\tau}^{(3)}(\alpha)]^{2n} = \frac{e^{n\lambda/\alpha}}{(\frac{1}{4}\lambda^2\alpha^\mu)^n}\left(1 - 2n\frac{\mu\alpha}{\lambda} + n\frac{\mu^2\alpha^2}{\lambda^2} - n\frac{4\mu\alpha^2}{\lambda^2} + n^2\frac{2\mu^2\alpha^2}{\lambda^2}\right).$$

We find from Eqs. (20) after the completion of the τ factor that

$$N^{(1)+(2)+(3)}_{+\mu} = [\bar{\tau}^{(3)}(\alpha)]^{2n}\frac{2\alpha^2}{\lambda}H_n\left(1 - \frac{\mu\alpha}{\lambda} + \frac{\alpha^2\mu^2}{\lambda^2} - \frac{2\alpha^2\mu}{\lambda}\right),$$

$$D^{(1)+(2)+(3)}_{+\mu} = [\bar{\tau}^{(3)}(\alpha)]^{2n}\left(1 - \frac{2\alpha}{\lambda}H_n + \frac{3\alpha^2\mu}{\lambda^2}H_n\right).$$

Thus we have

$$A^{(3)}_{+\mu} = -\frac{N^{(1)+(2)+(3)}_{+\mu}[n(g,\alpha),\alpha]}{D^{(1)+(2)+(3)}_{+\mu}[n(g,\alpha),\alpha]}$$
$$= -\frac{1 + 2\lambda^{-1}\alpha[\ln(g^{\frac{1}{2}}/\lambda) - \frac{1}{2}\mu\ln\alpha + \gamma]}{2\lambda^{-1}[\ln(g^{\frac{1}{2}}/\lambda) - \frac{1}{2}\mu\ln\alpha + \gamma] + O(\alpha)}$$
$$\underset{\alpha\to 0}{\longrightarrow} 0.$$

One can also show easily that $A^{(3)}_{\theta\mu} = 0$ by the optimum-n method. Thus we find that for $\mu > 0$ third peratization again gives zero for the scattering length.

We note that if one had not completed the τ factor in these cases, the second and third peratized scattering lengths would also have been zero for $\mu > 0$. The present example is thus not very instructive as to the appropriateness of the procedure. However, it is shown by an example in Appendix D that only by completing the τ factor do higher peratizations give corrections to first peratization, as the explicit peratization of the Born series shows to be the case.[4]

For $\mu = 0$, we calculate higher peratizations by the optimum-n method for the + regularization case. We note that, in the present instance, we are going beyond the exponential peratization which has been employed until now. Referring to Eq. (C2) in Appendix C, we find in "third" peratization by keeping only the leading singular (in α) part of the coefficient of $e^{\lambda(n-1)/\alpha}$ that $A^{(3)}_0 = +2$. On the other hand, if we keep the complete polynomial (in α) factor of $e^{\lambda(n-1)/\alpha}$ in N_n and D_n, we obtain (in "fifth" peratization) by the optimum-n method

$$A^{(5)}_{+0} = +\frac{\frac{1}{2}\lambda}{\frac{1}{2}\ln(g/\lambda^2) + \gamma} \times \left\{1 + \left(\frac{g}{\lambda^2}\right)\left[\ln\left(\frac{g}{\lambda^2}\right) - 2 + 2\gamma\right]\right\}, \quad (47)$$

which is the first two terms of a convergent expansion[22] of the exact scattering length[23] in the neighborhood of $g = 0$. This example hints at the interesting idea that peratization might prove more successful if one peratizes in terms of appropriate "singular units," such as orders of $e^{\lambda/\alpha}$ in the present example, retaining the complete polynomial factor which accompanies the given order of $e^{\lambda/\alpha}$.[24] This idea is discussed further in Sec. VII and will be elaborated in further work.

The present example is instructive as to the unsoundness of simple-minded peratization as an approximation procedure. A glance at the first expression on the right of Eq. (41a) or (41b) would suggest naively in the spirit of the peratization approach that the term 1 or $\lambda n/\alpha$, respectively, should be the dominant term in the denominator. Retaining only this term gives, of course, the first peratization result $A^{(1)} = 0$. By recognizing that in the summation of the series the n emerges effectively as a function of α through the optimum-n relation, $n(g, \alpha) = g^{\frac{1}{2}}\lambda^{-1}e^{\lambda/2\alpha}$, it is clear that the indexing of "most singular" terms by their explicit α dependence is a fallacious procedure. In fact,one finds that the expression

$$H_{n(g,\alpha)} \approx \ln n(g) + \gamma \approx \lambda/2\alpha + \tfrac{1}{2}\ln(g/\lambda^2)$$

exactly cancels this "most singular" term in both cases and presents an entirely different origin for the dominant term to this order of approximation. The same has been found for higher-order peratizations for $\mu = 0$. Indeed, with this in mind, it would be surprising that peratization should work at all, and in fact it rarely does, with the CC potentials the only ones in the literature which seem to work in some sense.

VII. DISCUSSION

The peratizations of the exponentially singular potentials considered in this article lead to some worthwhile observations. Foremost is that the $m = 4$ case for which peratization succeeds in some sense is exceptional. This success, moreover, has its qualifications. Sensible answers result when one peratizes in units of the exponential factor $e^{\lambda/\alpha}$ together with its complete polynomial factor. Peratization according to the degree of the polynomial terms within each exponential order does not yield sensible answers. For $m = 4$ the polynomial factors contain a small number of terms, and peratization in units of the exponential singularity is feasible. An examination of the exact solution, in fact, confirms that successive peratization in units of the exponential singularity provides a converging procedure to the exact answer, where successive terms contribute smaller-order corrections. For the case $m > 4$, we do not know how to peratize fully to a given order in units of the exponential singularity, and perhaps have therefore not correspondingly succeeded. This suggests the point of view that one may possibly salvage something of the peratization philosophy in some problems, if one were to collect singularities according to appropriate groupings other than the simple-minded ones that have been tried. The prognosis for the development of a general technique, however, does not seem very bright.

Another qualification of the success of the $m = 4$ case is that it yields a converging successive approximation procedure only in the weak coupling limit $g \approx 0$. This is not difficult to understand. The scattering length is determined by the whole potential and not only by the singularity near $r = 0$. For strong coupling, the tail of the repulsive potential plays an important role in suppressing approach to the singular region, and therefore the effect of the singularity on the scattering length should be small. This is confirmed in the $m = 4$ case where the leading term in the strong-coupling behavior of the scattering length agrees exactly with that of the g/r^4 potential. It is a general property of potential scattering that the leading strong-coupling behavior depends only on the behavior of the tail.[13] It should be only in the weak-coupling regime that the singularity near $r = 0$ would determine the dominant behavior of the scattering length (and the phase shift). Peratization is a device for determining the effect of the singularity on the scattering, and if it should succeed, it should succeed only in the weak-coupling regime. Otherwise put, since peratization is only a device for summing the Born series, itself a weak-coupling expansion, its success, if any, would be expected only in this limit.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. H. Cornille for a helpful correspondence.

APPENDIX A: SINGULAR BASIS FUNCTIONS

The customary peratization procedure entails the selection of the leading singular terms in each order of the Born series for the amplitude or scattering length. The procedure of the present article is to select leading singular terms in the numerator and denominator of a ratio which represents the scattering length. It is well known and easily demonstrated that these two procedures are by no means necessarily equivalent. For example, the numerator and denominator could both be multiplied by a common g-independent factor which expands into a series with stronger

singularities than possessed by the original numerator and denominator. One would then find that peratizations up to a certain order, or even to all finite orders, would yield a ratio of unity. If the numerator and denominator were to involve entirely different bases of singular functions, then the definition of order of peratization would be ambiguous. Similar difficulties apply in fact to the scheme of directly peratizing the Born series, since the addition of a damped exponential of a very singular function would lead to a zero or artificial result in successive peratizations.

We illustrate the problems that arise from an extraneous common factor with a simple example. Suppose a function $F(g, \alpha)$, written as a ratio $N(g, \alpha)/D(g, \alpha)$, can be peratized in the sense that first peratization gives a first approximation to the exact value of the function and higher peratizations give correction terms. Suppose that $D(g, \alpha)$ has the singularity expansion

$$D(g, \alpha) = \sum_{n=0}^{\infty} g^n \sum_{m=0}^{n} c_{nm} e^{m\lambda/\alpha}.$$

The first and second peratized contributions to $D(g, \alpha)$ are respectively given by

$$D^{(1)}(g, \alpha) = \sum g^n c_{nn} e^{n\lambda/\alpha},$$
$$D^{(2)}(g, \alpha) = \sum g^n c_{n,n-1} e^{(n-1)\lambda/\alpha}.$$

Similar expressions hold for $N(g, \alpha)$.

We consider the effect of multiplying N and D by the extraneous common factor

$$e^{\nu/\alpha^2} E(\alpha) = e^{\nu/\alpha^2} \sum_{r=0}^{\infty} e_r \alpha^r,$$

where the function $E(\alpha)$ is assumed to be analytic about $\alpha = 0$. Writing

$$\bar{N}(g, \alpha) = e^{\nu/\alpha^2} E(\alpha) N(g, \alpha),$$
$$\bar{D}(g, \alpha) = e^{\nu/\alpha^2} E(\alpha) D(g, \alpha),$$

one finds the first and second peratized contributions to $\bar{D}(g, \alpha)$ to be given by

$$\bar{D}^{(1)}(g, \alpha) = \sum_n g^n c_{nn} e_0 e^{n\lambda/\alpha + \nu/\alpha^2},$$
$$\bar{D}^{(2)}(g, \alpha) = \sum_n g^n c_{nn} e_1 \alpha e^{n\lambda/\alpha + \nu/\alpha^2}.$$

Again similar expressions hold for $\bar{N}(g, \alpha)$. Clearly neither second peratization nor any finite number of peratizations will lead to any improvement to first peratization because of the dominance of the singular factor $e^{\nu/\alpha^2} E(\alpha)$ over the remaining singularities $e^{n\lambda/\alpha}$.

Difficulties of a different nature arise if an extraneous common factor is g dependent. Consider $\bar{\bar{N}}(g, \alpha)$ and $\bar{\bar{D}}(g, \alpha)$, given by

$$\bar{\bar{N}}(g, \alpha) = \exp(g e^{\nu/\alpha^2}) N(g, \alpha),$$
$$\bar{\bar{D}}(g, \alpha) = \exp(g e^{\nu/\alpha^2}) D(g, \alpha).$$

The first and second peratized contributions to $\bar{\bar{D}}(g, \alpha)$ become

$$\bar{\bar{D}}^{(1)}(g, \alpha) = \sum_{n=0}^{\infty} g^n \frac{e^{n\nu/\alpha^2}}{n!} c_{00} = c_{00} \exp(g e^{\nu/\alpha^2}),$$
$$\bar{\bar{D}}^{(2)}(g, \alpha) = \sum_{n=1}^{\infty} g^n \frac{e^{(n-1)\nu/\alpha^2}}{(n-1)!} c_{11} e^{\lambda/\alpha} = g c_{11} \exp(g e^{\nu/\alpha^2 + \lambda/\alpha}).$$

In this case we see that peratization fails completely.

Despite the arbitrariness of such procedures, one can formulate a criterion for when separate peratization of the numerator and denominator would be effectively equivalent to peratization of the ratio (Born series). We find that by requiring the denominator function $D(g, \alpha)$ to be the zero-energy Jost function, one eliminates the arbitrariness due to extraneous factors in the numerator and denominator of the type mentioned above. We express $N_n(\alpha)$ and $D_n(\alpha)$, the coefficients in the respective coupling constant expansions of $N(g, \alpha)$ and $D(g, \alpha)$, as a finite or infinite expansion in terms of decreasing orders of singularity in α, i.e.,

$$N_n(\alpha) = \omega(\alpha) \sum_l a_{nl} y_{nl}(\alpha), \tag{A1}$$

$$D_n(\alpha) = \omega(\alpha) \sum_l b_{nl} y_{nl}(\alpha), \tag{A2}$$

where the l summation begins with $l = 1$ and $y_{n,l+1}(\alpha)/y_{nl}(\alpha) \to 0$ as $\alpha \to 0$. We have assumed in these expressions that the same set of singular basis functions $\{y_{nl}(\alpha)\}$ appear in both N_n and D_n. This turns out to be the case in practice and, in any event, this assumption can be verified easily. Of course one may have some vanishing coefficients among the a_{nl} and b_{nl}. We also note that, because of dimensional considerations, an over-all function of α, $\omega(\alpha)$, usually α itself, appears as a factor in $N(g, \alpha)$. The sth peratization (or the sth order of peratization) entails the prescription of retaining in $N_n(\alpha)$ and $D_n(\alpha)$ the terms of Eqs. (A1) and (A2) from $l = 1$ to $l = s$. Other aspects related to the retention of these terms are described later in this appendix and in the body of the article. One wishes at this point to be sure that the series obtained by formally dividing the sth peratized N and D is the same, to within an error term of higher order of peratization, as what would be obtained by peratizing the Born series itself. In order to insure this, we at least demand that the set of

singular basis functions $\{y_{nl}(\alpha)\}$ used in the expansion of $N_n(\alpha)$ and $D_n(\alpha)$ is the same as would occur in an appropriate expansion of the Born series in singularities, or more generally that an appropriate nonvanishing subset of the singular basis functions occurring in $N_n(\alpha)$ or $D_n(\alpha)$ also occur in the nth term of the Born series. Thus it is conceivable that some singular basis functions occurring in N and D may be absent in the Born series. A set of singular basis functions in N and D is said to be spurious if no nonempty subset of these terms appears in the corresponding order of the Born series.[25] For instance, in the example presented above, the Born series has singular basis functions of the form $e^{n\lambda/\alpha}$, while the functions $\bar{D}$ and $\bar{\bar{D}}$ have the spurious singular basis functions of the form $\alpha e^{n\lambda/\alpha+\nu/\alpha^2}$ or $e^{n\nu/\alpha^2}e^{\beta\lambda/\alpha}$, respectively. If no set of the basis functions $y_{nl}(\alpha)$ is spurious, then peratization of N and D to a given order s should correspond to peratization of the ratio to that order plus an error term of higher order. If peratization of the Born series were a valid successive approximation procedure, then the presence of an error term at higher order ought not detract from the efficacy of the procedure. However, it is frequently the case that the "error" term is not small, but rather is very large and dominates the series. This circumstance was discussed in the article and led to the prescription of the completion of the τ factor.

We now show in two ways that, by choosing $D(g, \alpha)$ to be the zero-energy Jost function, no set of spurious basis functions should occur in the expansion at $N_n(\alpha)$ or $D_n(\alpha)$. This is done by showing that N and D so chosen can have no common factor and by showing more explicitly that the singular basis functions in the numerator and denominator functions and in the Born series ought to be the same.

The fact that the denominator function $D(g, \alpha)$ is identical with the zero-energy Jost function permits one to rule out cancelling factors in the numerator and denominator. Thus the numerator and denominator may not possess a common zero. This follows from the expression for the function for $N(g)$ in the form[26]

$$N(g) = \lim_{k\to 0} \frac{i}{k}[f(k, g) - f(0, g)] = i\frac{\partial}{\partial k} f(k, g)\Big|_{k=0}, \tag{A3}$$

with $f(k, g)$ the Jost function, and $f(0, g) \equiv D(g)$. In general, the point $k = 0$ need not be a point of analyticity of $f(k, g)$, but the limit indicated in Eq. (A3) exists if the point $k = 0$ is approached from the lower half-plane.[27] A common zero to $D(g)$ and $N(g)$ would imply a double zero of $f(k, g)$ at $k = 0$, which one knows for $l = 0$ partial waves not to be possible.[26] Nor can $N(g)$ and $D(g)$ have a nonvanishing entire function as a common factor, since $N(g)$ and $D(g)$ both have exponential order $\frac{1}{2}$ (see Appendix B), and an entire factor which is a function of g would raise the exponential order to at least unity. Nor can there be a g-independent factor in common since $D(g) = 1$ at $g = 0$. The quantities $N(g)$ and $D(g)$ are therefore in their lowest terms.

We can, however, be somewhat more specific in showing why spurious basis functions ought not to occur with the choice of the Jost function as the denominator function. The following argument is restricted to certain classes of potentials of interest. These include the potentials $e^{\lambda/r}/r^m$ discussed in the present article, as well as a generalization of this class, given by $gR(r)/r^m$, $m \geq 4$, where $R(r)$ is a "rapidly varying function" defined by the property[21]

$$R'(r)/R(r) = 1/r\zeta(r)$$

and $\zeta(r) \to 0$ as $r \to 0$. An article describing the peratization of these potentials is in preparation. Another class of potentials is of the form $L(r)/r^m$, where $L(r)$ is a tempered slowly varying function. The peratization of these potentials is described in the following article. A third class consists of the sum of two inverse power singular potentials, an example of which is discussed in Appendix D.

The s-wave scattering length A is related to the asymptotic form of the zero-energy s-wave physical wavefunction by

$$\psi(r) \underset{r\to\infty}{\approx} r + A + o(1). \tag{A4}$$

The wavefunction $\psi(r)$ satisfies the integral equation[28]

$$\begin{aligned}\psi(r) &= r + 2mg\int_0^\infty dr' G(r, r')V(r')\psi(r') \\ &= r - 2mg\int_0^r dr' r' V(r')\psi(r') \\ &\quad - 2mgr\int_r^\infty dr' V(r')\psi(r').\end{aligned} \tag{A5}$$

It can be shown from Newton's article[28] that the physical wavefunction $\psi(r)$ may be written as the quotient

$$\psi(r) = \varphi(r)/f(0). \tag{A6}$$

Here $\varphi(r)$ is a regular solution of the zero-energy, s-wave Schrödinger equation satisfying the boundary conditions $\varphi(0) = 0$ and $\varphi'(0) = 1$, and $f(r)$ is the zero-energy s-wave Jost solution of the Schrödinger equation satisfying the boundary condition $\lim f(r) = 1$, as $r \to \infty$. The s-wave Jost function is $f(0)$. $\varphi(r)$

and $f(r)$ satisfy, respectively, the integral equations[28]

$$\varphi(r) = r + 2mg\int_0^r dr'(r - r')V(r')\varphi(r'),$$

$$f(r) = 1 - 2mg\int_r^\infty dr'(r - r')V(r')f(r').$$

Peratization of the scattering length in terms of the Born series directly is effectively accomplished by solving Eq. (A5) for $\psi(r)$ by iteration and by performing the resulting integrations by parts to obtain the expansion in singular basis functions. In the present paper, however, peratization is performed by peratizing numerator and denominator of a ratio separately. From Eqs. (A4) and (A6), it is clear that this involves obtaining an expansion in g of $\varphi(r)$ and $f(r)$ by iterating their integral equations and by integrating by parts to obtain the singularity expansion. One sees from the similarity of the integral equations for $\psi(r)$ and $f(r)$ that, by iterating and integrating by parts, the singular basis functions which appear in ψ are included in the basis functions that appear in either φ or f. The terms that appear in the series for ψ are obtained by formally dividing φ and f to obtain the scattering length, according to Eqs. (A4) and (A6). It is moreover true that the nth term of the Born series is nonvanishing for each n. This is a consequence of the fact that the total contribution to the nth iteration of ψ is a quantity of definite sign, since the Green's function $G(r, r')$ in Eq. (A5) is a negative quantity. Therefore that contribution cannot vanish.

We conclude this Appendix by applying the foregoing considerations concerning the presence or absence of spurious basis functions to the problems which lead to the prescription at the completion of the τ factor, when studied by the optimum-n method. We refer to Eq. (35) of the text,which we write as

$$A^{(2)} = \lim_{\alpha\to 0} \frac{q_1(\alpha)}{p_1(\alpha)} \times \frac{1 + [2n(g,\alpha)\bar{\tau}_2(\alpha)/\bar{\tau}_1(\alpha)] + q_2(\alpha)/q_1(\alpha)}{1 + [2n(g,\alpha)\bar{\tau}_2(\alpha)/\bar{\tau}_1(\alpha)] + p_2(\alpha)/p_1(\alpha)}. \quad \text{(A7)}$$

We may regard both the numerator and denominator of this expression as the first two contributions of an expansion in singularities in α of the complete $N_{n(g,\alpha)}(\alpha)$ and $D_{n(g,\alpha)}(\alpha)$. We particularly note that the index n is a function of g and α according to the precepts of the optimum-n method. On dividing the ratio of Eq. (A7), one finds that the singular basis function $2n(g, \alpha)\bar{\tau}_2(\alpha)/\bar{\tau}_1(\alpha)$ does not in fact appear in the quotient and hence might be regarded as a spurious basis function. That this is indeed the case can be seen by recalling that this singular function arises from the extraneous common factor $[\bar{\tau}(\alpha)]^{2n}$, second peratized. It is the purpose of the prescription of completing the τ factor to remove this factor, as discussed in the main portion of the text.

APPENDIX B: ANALYTIC PROPERTIES OF $N(g)$

The determination that $D^\nu(g, \alpha)$ is entire and of exponential order $\frac{1}{2}$ has been shown by three methods.[11–13] One method,[12] based on the existence of an upper bound to the potential in the term of a monotonically decreasing $L^{(\frac{1}{2})}$ function, is readily adapted to the function $N(g, \alpha)$ in either regularization. This method, however, does not specify the type of $D^\nu(g, \alpha)$, nor would it for $N^\nu(g, \alpha)$. Another method[11] would entail the representability of $|V(r, \alpha)|^{\frac{1}{2}}$ as an absolutely convergent Laplace transform (ACLT). The potential $|V(r)|^{\frac{1}{2}}$ for $V(r)$ of Eq. (1) is so expressible:

$$|V(r)|^{\frac{1}{2}} = \frac{g^{\frac{1}{2}}e^{\lambda/2r}}{r^{\frac{1}{2}m}} = g^{\frac{1}{2}}\int_0^\infty dt e^{-rt}\left(\frac{2t}{\lambda}\right)^{(m-2)/4} I_{\frac{1}{2}m-1}(2\lambda t)^{\frac{1}{2}}. \quad \text{(B1)}$$

Quite clearly, if $|V(r)|^{\frac{1}{2}}$ is an ACLT, then $|V(r + \alpha)|^{\frac{1}{2}}$ is too, but $\theta(r - \alpha)|V(r)|^{\frac{1}{2}}$ is not. It is shown in Ref. 11 that if $V(r)$ is nonnegative and $V(r)^{\frac{1}{2}}$ has nonnegative Laplace weight whose support is bounded away from the origin, then the type τ is explicitly given by

$$\tau = \int_0^\infty dr[V(r)]^{\frac{1}{2}}. \quad \text{(B2)}$$

Through a limiting procedure one can extend this result to the case where the support of the Laplace weight is not bounded away from the origin. Thus one can derive Eq. (27) for the + regularized potential, and can adapt the result through minor modifications to apply also to $N^+(g, \alpha)$. While the θ-regularized potential is not an ACLT, the methods of Ref. 11 do go through for the potential $\theta(r - \alpha)V(r)$ to yield the validity of Eq. (27) for both $D^\theta(g, \alpha)$ and $N^\theta(g, \alpha)$. The third method[13] is based on the WKB solution to the scattering differential equation, and the identity of $D(g)$ with the Jost function. This method specifies the exponential order, as well as the type for $D(g)$. Since $N(g)$ is not simply interpretable in terms of the solution to the differential equation, one does not obtain any direct conclusion concerning $N(g)$. One can, however, determine the type of $N^\nu(g, \alpha)$ from that of $D^\nu(g, \alpha)$ by employing some easily

proven inequalities for the coefficients of the N and D power series:

$$\alpha D_n^\theta(\alpha) \leq N_n^\theta(\alpha) \leq \left(\int_\alpha^\infty dr r^2 V(r)\right) D_{n-1}^\theta(\alpha). \quad \text{(B3)}$$

The type is expressible as[18]

$$\tau \equiv \overline{\lim_{n\to\infty}} ((2n)!\, N_n)^{1/(2n)}. \quad \text{(B4)}$$

The inequality [Eq. (B3)] sandwiches τ between two equal quantities and this known value, which is the type $D^\theta(g, \alpha)$, must therefore be the type of $N^\theta(g, \alpha)$. For $N_n^+(\alpha)$, one easily verifies the upper bound

$$N_n^+(\alpha) \leq \left(\int_0^\infty dr r^2 V(r+\alpha)\right) D_{n-1}^\theta(\alpha). \quad \text{(B5)}$$

A lower bound can also be established:

$$\begin{aligned} N_n^+(\alpha) &\equiv \int_0^\infty dr_1 \cdots \int_0^{r_{n-1}} dr_n r_1(r_1 - r_2)\cdots(r_{n-1} - r_n) \\ &\quad \times r_n V(r_1+\alpha)\cdots V(r_n+\alpha) \\ &= \int_\alpha^\infty dx_1 \cdots \int_\alpha^{x_{n-1}} dx_n (x_1 - \alpha)(x_1 - x_2)\cdots \\ &\quad \times (x_{n-1} - x_n)(x_n - \alpha) V(x_1)\cdots V(x_n), \end{aligned} \quad \text{(B6)}$$

where the variable changes $r_j = x_j - \alpha$, $j = 1 \cdots n$, have been made. From Eq. (B6), one readily finds that

$$N_n^+(\alpha) = N_n^\theta(\alpha) - \alpha D_n^+(\alpha) - \alpha \Delta_n(\alpha), \quad \text{(B7)}$$

where

$$\begin{aligned} \Delta_n(\alpha) &\equiv \int_\alpha^\infty dx_1 \cdots \int_\alpha^{x_{n-1}} dx_n \\ &\quad \times x_1(x_1 - x_2)\cdots(x_{n-1} - x_n) V(x_1)\cdots V(x_n) \\ &\leq \frac{1}{\alpha}\left(\int_\alpha^\infty dx x^2 V(x)\right) \int_\alpha^\infty dx_2 \cdots \int_\alpha^{x_{n-1}} dx_n \\ &\quad \times (x_2 - x_3)\cdots(x_{n-1} - x_n) x_n V(x_2)\cdots V(x_n). \end{aligned} \quad \text{(B8)}$$

Equations (B7) and (B8) provide a lower bound to $N_n^+(\alpha)$, which determines the same type as does the upper bound in Eq. (B5) and therefore establishes the type of $N_n^+(\alpha)$.

APPENDIX C: HIGHER PERATIZATIONS

The higher-order terms in the expression for the scattering length in the $m = 4$ case might be obtained by the techniques of the present article, but one would have to work with the parametric integral $J(\beta_1, \cdots, \beta_n)$ of Eq. (7) prior to the exponential peratization performed in Eq. (11). Instead we shall avail ourselves of the explicit form of the scattering length[23] derived from the known exact wavefunction for the + regularized case:

$$A = g^{\frac{1}{2}} \frac{K_0(g^{\frac{1}{2}}x_0) I_0'(2g^{\frac{1}{2}}/\lambda) - I_0(g^{\frac{1}{2}}x_0) K_0'(2g^{\frac{1}{2}}/\lambda)}{K_0(g^{\frac{1}{2}}x_0) I_0(2g^{\frac{1}{2}}/\lambda) - I_0(g^{\frac{1}{2}}x_0) K_0(2g^{\frac{1}{2}}/\lambda)} + \alpha. \quad \text{(C1)}$$

Retention of the terms of order $e^{n\lambda/\alpha}$ and $e^{(n-1)\lambda/\alpha}$ in the coefficient of g^n results in

$$A_{+0}^{(5)} = \frac{\sum_{n=1}^\infty \frac{(x/2)^{2n}}{n!^2} \{\alpha H_n - \frac{1}{2}\lambda n^2 e^{-\lambda/\alpha}[\lambda\alpha^{-1} - 2(1 + \lambda^{-1}\alpha)(H_{n-1} - 1)]\}}{\sum_{n=0}^\infty \frac{(x/2)^{2n}}{n!^2} \{-(\lambda/2\alpha) + H_n + n^2 e^{-\lambda/\alpha}[-(\lambda/2\alpha) + H_{n-1} - 1]\}}. \quad \text{(C2)}$$

We have labeled this the "fifth" peratized scattering length (though one might quibble about the enumeration), since we have retained in toto five orders of singular terms. In the third peratized expression, $A_{+0}^{(3)}$ one would retain only the $1/\alpha$ terms in the factor of $e^{(n-1)\lambda/\alpha}$ in the coefficient of g^n. The results of these peratizations are presented in Sec. VI.

In the $\mu > 0$ case, there are many more singular terms in the factor of $e^{\lambda n/\alpha}$ in the coefficient of g^n, and the calculation of succeeding terms invites greater and greater complexity. We presently outline the calculation of the third peratized terms in the N and D series. We denote $N_{\gamma\mu}(\beta)$ and $D_{\gamma\mu}(\beta)$ generically by $G_{\gamma\mu}(\beta)$, dropping indices where confusion is unlikely. Generally,

$$G_{\gamma 0}(\beta) \equiv R_{\gamma 0}(\alpha; \beta) e^{1/\alpha \cdot \Sigma\beta}, \quad \text{(C3)}$$

where the $R_{\gamma 0}(\alpha; \beta)$ can be read off from Eqs. (13a, b) and (14a, b). From the relation Eq. (16),

$$G_{\gamma\mu}(\beta) = (\partial_1^\mu \cdots \partial_n^\mu) G_{\gamma 0}(\beta), \quad \text{(C4)}$$

one sees that one can rank the degree of the singular (in $1/\alpha$) terms in $G_{\gamma\mu}(\beta)$ according to the number of times the derivatives in Eq. (C4) act on the factor $\exp(1/\alpha \cdot \sum \beta)$ of Eq. (C3). For the $D_{\gamma\mu}$, the corresponding $R_{\gamma\mu}$ contain terms of two different degrees in α, and a fixed number of differentiations of the $\exp(1/\alpha \cdot \sum \beta)$ factor gives rise to a corresponding

mixing of two degrees of singularity in α. The contributions to $G_{\gamma\mu}$ from the first three singular orders in α can be expressed by

$$G_{\gamma\mu}^{(1+2+3)} = \frac{e^{1/\alpha\cdot\Sigma\beta}}{\alpha^{n\mu}}\Big[R_{\gamma\mu} + \mu\alpha\sum_j \partial_j R_{\gamma\mu} - \tfrac{1}{2}\mu\alpha^2\sum_j \partial_j^2 R_{\gamma\mu} + \tfrac{1}{2}\alpha^2\mu^2\Big(\sum_j \partial_j\Big)\Big(\sum_k \partial_k\Big)R_{\gamma\mu}\Big] \quad \text{(C5)}$$

For the $N_{\gamma\mu}$ case the first, second, and third peratization, respectively, correspond to the first, second, and last two terms in the right of Eq. (C5). For the $D_{\gamma\mu}$ case, there will be some mixing, and some higher-order terms will be present which one would discard. Explicit calculation gives

$$\begin{aligned} D_{\theta\mu}^{(3)}(\beta) &= \frac{2\mu\alpha e^{1/\alpha\cdot\Sigma\beta}}{\alpha^{\mu n}}\{-nH_n + (\textstyle\sum\beta)L - K(1 - 2(\sum\beta)H) \\ &\quad - [(\textstyle\sum\beta)(S + \tfrac{1}{2}T) - K] \\ &\quad + \tfrac{1}{2}\mu[(\textstyle\sum\beta)Q - 2nK + 2(\sum\beta)K^2]\}P^2 \\ &\to \frac{2\mu\alpha e^{n\lambda/\alpha}}{\lambda(\lambda^2\alpha^\mu)^n n!^2}\, n[\mu n^2 - (\tfrac{1}{2}\mu + 2)n + (2n + \tfrac{1}{2})H_n], \end{aligned} \quad \text{(C6)}$$

$$\begin{aligned} N_{\theta\mu}^{(3)}(\beta) &= \frac{2\mu\alpha^{2\,1/\alpha\cdot\Sigma\beta}}{\alpha^{\mu n}}\{-[(\textstyle\sum\beta)(S + \tfrac{1}{2}T) - K] \\ &\quad + \tfrac{1}{2}\mu[(\textstyle\sum\beta)Q - 2nK + 2(\sum\beta)K^2]\}P^2 \\ &\to \frac{2\mu\alpha^2 e^{n\lambda/\alpha}}{\lambda(\lambda^2\alpha^\mu)^n n!^2}\, n[\mu n^2 - (\tfrac{1}{2}\mu + 2)n + 1 + \tfrac{1}{2}H_n], \end{aligned} \quad \text{(C7)}$$

$$\begin{aligned} D_{+\mu}^{(3)}(\beta) &= \frac{\mu\alpha^2 e^{1/\alpha\cdot\Sigma\beta}}{\alpha^{\mu n}}[2L + 4HK - (2S + T) + \mu Q + 2\mu K^2]P^2 \\ &\to \frac{2\mu\alpha^2 e^{n\lambda/\alpha}}{\lambda^2(\lambda^2\alpha^\mu)^n n!^2}[\mu n^2 + (\tfrac{1}{2}\mu - 2)n + 2nH_n + \tfrac{3}{2}H_n], \end{aligned} \quad \text{(C8)}$$

$$\begin{aligned} N_{+\mu}^{(3)}(\beta) &= \frac{2\mu\alpha^4 e^{1/\alpha\cdot\Sigma\beta}}{\alpha^{\mu n}}[-(X + 2Y + 2SH + HT) \\ &\quad + \mu(2KL + 2K^2H + HQ + V)]P^2 \\ &\to \frac{4\mu\alpha^4 e^{n\lambda/\alpha}}{\lambda^3(\lambda^2\alpha^\mu)^n n!^2}\, H_n[\mu n^2 + (\tfrac{3}{2}\mu - 2)n - 1 + \tfrac{1}{2}\mu], \end{aligned} \quad \text{(C9)}$$

where the arrow denotes the result of the ansatz $\beta_j = \lambda$. The following notation has been used:

$$P = \frac{1}{\beta_1(\beta_1 + \beta_2)\cdots(\beta_1 + \beta_2 + \cdots \beta_n)} \to \frac{1}{\lambda^n n!},$$

$$H = \frac{1}{\beta_1} + \frac{1}{\beta_1 + \beta_2} + \cdots \frac{1}{\beta_1 + \cdots \beta_n} \to \frac{1}{\lambda}\Big(1 + \tfrac{1}{2} + \cdots \frac{1}{n}\Big) \equiv \frac{1}{\lambda}H_n,$$

$$L = -\sum_j \partial_j H = \sum_j \frac{j}{(\beta_1 + \cdots \beta_j)^2} \to \frac{1}{\lambda^2}H_n,$$

$$K = -\frac{1}{P}\sum_j \partial_j P = \sum_j \frac{j}{\beta_1 + \cdots \beta_j} \to \frac{n}{\lambda},$$

$$Q = -\sum_j \partial_j K = \sum_j \frac{j^2}{(\beta_1 + \cdots \beta_j)^2} \to \frac{n}{\lambda^2},$$

$$X = \tfrac{1}{2}\sum_j \partial_j^2 H = \sum_j \frac{j}{(\beta_1 + \cdots \beta_j)^3} \to \frac{1}{\lambda^3}\sum_{j=1}^n j^{-2} \equiv \frac{W_n}{\lambda^3},$$

$$Z_j = -\frac{1}{P}\partial_j P = \sum_{l\geq j} \frac{1}{\beta_1 + \cdots \beta_l} \to \frac{1}{\lambda}(H_n - H_{j-1}),$$

$$T = -\sum_j \partial_j Z_j = \sum_j \frac{j}{(\beta_1 + \cdots \beta_j)^2} = L = \frac{1}{\lambda^2}H_n,$$

$$V = -\tfrac{1}{2}\sum_j L = \sum_j \frac{j^2}{(\beta_1 + \cdots \beta_j)^3} \to \frac{1}{\lambda^3}H_n,$$

$$\begin{aligned} S = \sum_j Z_j^2 &= \sum_{j=1}^n \sum_{l\geq j}\sum_{m\geq j} \frac{1}{(\beta_1 + \cdots \beta_l)(\beta_1 + \cdots \beta_m)} \\ &= \Big(\sum_{l=1}^n \sum_{m=1}^{l-1}\sum_{j=1}^m + \sum_{m=1}^n \sum_{l=1}^m \sum_{j=1}^l\Big) \\ &\quad \times \frac{1}{(\beta_1 + \cdots \beta_l)(\beta_1 + \cdots \beta_m)} \to \frac{1}{\lambda^2}(2n - H_n), \end{aligned}$$

$$\begin{aligned} Y &= \frac{1}{2P^2}\sum_j (\partial_j H)(\partial_j P^2) \\ &= \sum_{j=1}^n \sum_{l\geq j}\sum_{m\geq j} \frac{1}{(\beta_1 + \cdots \beta_l)^2}\frac{1}{(\beta_1 + \cdots \beta_m)} \\ &\to \frac{1}{\lambda^3}(H_n - \tfrac{1}{2}W_n + \tfrac{1}{2}H_n^2). \end{aligned}$$

APPENDIX D: PERATIZATION OF THE POTENTIAL $V(r) = g\left(\frac{1}{r^5} + \frac{\varphi}{r^4}\right)$

In this appendix we outline the peratization procedure employed in this article for the potential

$$V(r) = g\Big(\frac{1}{r^5} + \frac{\varphi}{r^4}\Big) \quad \text{(D1)}$$

This potential is exactly solvable at zero energy[29] and hence can be used to calculate the peratized scattering lengths to all orders and to compare these results with the exact scattering length. Gale[4] has studied the peratization of this potential by peratizing the Born series directly and has found that first peratization gives the exact scattering length for the potential

gr^{-5}, while second peratization gives a correction term which is linear in φ.

One can readily find that the exact scattering length for the + regularized potential $V(r+\alpha)$ is given by

$$A(g,\alpha) = -N(g,\alpha)/D(g,\alpha), \tag{D2}$$

$$D(g,\alpha) = g^{-\frac{1}{3}}\pi\alpha\Big\{\mathrm{Bi}\left[g^{\frac{1}{3}}\left(\frac{1}{\alpha}+\varphi\right)\right]\mathrm{Ai}\,(g^{\frac{1}{3}}\varphi) - \mathrm{Ai}\left(g^{\frac{1}{3}}\left(\frac{1}{\alpha}+\varphi\right)\right]\mathrm{Bi}\,(g^{\frac{1}{3}}\varphi)\Big\}, \tag{D3a}$$

$$N(g,\alpha) = -\pi\alpha\Big\{\mathrm{Bi}\left[g^{\frac{1}{3}}\left(\frac{1}{\alpha}+\varphi\right)\right]\mathrm{Ai}'\,(g^{\frac{1}{3}}\varphi) - \mathrm{Ai}\left[g^{\frac{1}{3}}\left(\frac{1}{\alpha}+\varphi\right)\right]\mathrm{Bi}'\,(g^{\frac{1}{3}}\varphi)\Big\} - \alpha D(g,\alpha). \tag{D3b}$$

Here $D(g,\alpha)$ is the Jost function and Ai (x) and Bi (x) are the Airy functions having the asymptotic behavior for large z, $\frac{1}{2}\pi^{-\frac{1}{2}}z^{-\frac{1}{4}}\exp(-\frac{2}{3}z^{\frac{3}{2}})$, and $\pi^{-\frac{1}{2}}z^{-\frac{1}{4}}\exp(\frac{2}{3}z^{\frac{3}{2}})$, respectively.[30] In the $\alpha\to 0$ limit the scattering length for the potential $V(r)$ becomes

$$A(g) = g^{\frac{1}{3}}\frac{\mathrm{Ai}'\,(g^{\frac{1}{3}}\varphi)}{\mathrm{Ai}\,(g^{\frac{1}{3}}\varphi)}. \tag{D4}$$

Using the standard expansions for the Airy functions,[30] one finds

$$N(g,\alpha) = \sum_{n=1}^{\infty} g^n N_n(\alpha), \tag{D5a}$$

$$D(g,\alpha) = \sum_{n=0}^{\infty} g^n D_n(\alpha), \tag{D5b}$$

$$N_n(\alpha) = -\alpha\frac{2\pi}{3^{\frac{3}{2}}}\left[\frac{1}{3}\left(\frac{1+\alpha\varphi}{\alpha}\right)^{\frac{3}{2}}\right]^{2n} \times \sum_{m=0}^{n}\left(\frac{\alpha\varphi}{1+\alpha\varphi}\right)^{3m}\frac{1}{m!\,(n-m)!} \times\Bigg[(3m+1+\alpha\varphi)\left(\frac{1}{\Gamma(m+\frac{2}{3})\Gamma(n-m+\frac{4}{3})} - \frac{1}{\Gamma(m+\frac{4}{3})\Gamma(n-m+\frac{2}{3})}\right) + \left(\frac{\alpha\varphi}{1+\alpha\varphi}\right)^2\frac{3(n-m)}{\Gamma(m+\frac{5}{3})\Gamma(n-m+\frac{1}{3})}\Bigg], \tag{D6a}$$

$$D_n(\alpha) = \frac{2\pi}{3^{\frac{3}{2}}}\left[\frac{1}{3}\left(\frac{1+\alpha\varphi}{\alpha}\right)^{\frac{3}{2}}\right]^{2n} \times \sum_{m=0}^{n}\left(\frac{\alpha\varphi}{1+\alpha\varphi}\right)^{3m}\frac{1}{m!\,(n-m)!} \times\left((1+\alpha\varphi)\frac{1}{\Gamma(m+\frac{2}{3})\Gamma(n-m+\frac{4}{3})} - \alpha\varphi\frac{1}{\Gamma(m+\frac{4}{3})\Gamma(n-m+\frac{2}{3})}\right). \tag{D6b}$$

We proceed to calculate the peratized scattering lengths by the optimum-n method. The quantity $\tau(\alpha)$ is given by

$$\tau(\alpha) = \int_{\alpha}^{\infty} dr\left(\frac{1}{r^5}+\frac{\varphi}{r^4}\right)^{\frac{3}{2}} = \frac{2}{3}\left[\left(\frac{1+\alpha\varphi}{\alpha}\right)^{\frac{3}{2}} - \varphi^{\frac{3}{2}}\right], \tag{D7}$$

and hence $\bar{\tau}(\alpha)$ is

$$\bar{\tau}(\alpha) = \frac{2}{3}\left(\frac{1+\alpha\varphi}{\alpha}\right)^{\frac{3}{2}}. \tag{D8}$$

The appropriate expansion parameter for application of the optimum-n method and for peratization ought to be $g\bar{\tau}^2(\alpha)$ and Eqs. (D6a, b) confirms this to be the case, in accordance with Eqs. (28). One readily finds from Eqs. (D6a, b) that, as a function of $g\bar{\tau}^2(\alpha)$, the exponential order of $N(g,\alpha)$ and $D(g,\alpha)$ is $\frac{1}{2}$ and the type is unity. Hence the optimum-n value, obtained from Eq. (25), is given by

$$n(g,\alpha) = \tfrac{1}{3}g^{\frac{1}{2}}\alpha^{-\frac{3}{2}} \tag{D9}$$

We note that the Γ function satisfies the relations[31]

$$(2n)! = \pi^{-\frac{1}{2}}2^{2n}\Gamma(n+\tfrac{1}{2})n!,$$

and, for large n,

$$\Gamma(n+\alpha)/\Gamma(n+b) \approx n^{a-b}.$$

With these relations and the optimum-n value $n(g,\alpha)$ of Eq. (D9), one finds that, in the limit of small α, $N_{n(g,\alpha)}$ and $D_{n(g,\alpha)}$ may be written as

$$N_{n(g,\alpha)}(\alpha) \approx \frac{\bar{\tau}(\alpha)^{2n}}{(2n)!}\frac{2\pi^{\frac{1}{2}}}{(3n)^{\frac{5}{6}}}g^{\frac{1}{3}}[3^{\frac{1}{3}}\Gamma(\tfrac{4}{3})]^{-1} \times\left(1 - \frac{1}{3^{\frac{1}{3}}}\frac{\Gamma(\frac{4}{3})}{\Gamma(\frac{5}{3})}(g^{\frac{1}{3}}\varphi)^2 + \tfrac{1}{3}(g^{\frac{1}{3}}\varphi)^3 - \frac{1}{3^{\frac{7}{3}}}\frac{\Gamma(\frac{4}{3})}{\Gamma(\frac{8}{3})}(g^{\frac{1}{3}}\varphi)^5 + \cdots\right) = -\frac{\bar{\tau}(\alpha)^{2n}}{(2n)!}\frac{2\pi^{\frac{1}{2}}}{(3n)^{\frac{5}{6}}}g^{\frac{1}{3}}\mathrm{Ai}'\,(g^{\frac{1}{3}}\varphi),$$

$$D_{n(g,\alpha)}(\alpha) \approx \frac{\bar{\tau}(\alpha)^{2n}}{(2n)!}\frac{2\pi^{\frac{1}{2}}}{(3n)^{\frac{5}{6}}}[3^{\frac{2}{3}}\Gamma(\tfrac{2}{3})]^{-1} \times\left(1 - \frac{1}{3^{\frac{2}{3}}}\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{4}{3})}g^{\frac{1}{3}}\varphi + \frac{1}{3^2}\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{5}{3})}(g^{\frac{1}{3}}\varphi)^3 - \frac{1}{3^{\frac{8}{3}}}\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{7}{3})}(g^{\frac{1}{3}}\varphi)^4 + \cdots\right) = \frac{\bar{\tau}(\alpha)^{2n}}{(2n)!}\frac{2\pi^{\frac{1}{2}}}{(3n)^{\frac{5}{6}}}\mathrm{Ai}\,(g^{\frac{1}{3}}\varphi).$$

Hence, by taking the ratio of N and D, the exact scattering length of Eq. (D4) is reproduced. It is clear, in this case, that s peratizations give the first s terms

of a Taylor-series expansion of the scattering length in φ about $\varphi = 0$. It is also clear that the completion of the τ factor goes through automatically in this example, since the quantity $\bar{\tau}^2(\alpha)$ appears at the outset as a factor in $N_n(\alpha)$ and $D_n(\alpha)$.

It is instructive to calculate peratization for this example without the completion of the τ factor. We do this in second peratization. Noting that

$$\left[\left(\frac{1+\alpha\varphi}{\alpha}\right)^{3n}\right]^{(2)} = \frac{1}{\alpha^{3n}}[1+3n\alpha\varphi],$$

one finds from Eqs. (D6a, b), the first and second peratized contributions to $N_n(\alpha)$ and $D_n(\alpha)$ to be

$$N_n^{(1)+(2)}(\alpha) = -\alpha\frac{2\pi}{3^{\frac{3}{2}}}\frac{1}{3^{2n}\alpha^{3n}} \times \left([1+\alpha\varphi(3n+1)]\frac{1}{\Gamma(\frac{2}{3})\Gamma(n+\frac{4}{3})} - \frac{1}{\Gamma(\frac{4}{3})\Gamma(n+\frac{2}{3})}\right),$$

$$D_n^{(1)+(2)}(\alpha) = \frac{2\pi}{3^{\frac{3}{2}}}\frac{1}{3^{2n}\alpha^{3n}} \times \left([1+\alpha\varphi(3n+1)]\frac{1}{\Gamma(\frac{2}{3})\Gamma(n+\frac{4}{3})} - \alpha\varphi\frac{1}{\Gamma(\frac{4}{3})\Gamma(n+\frac{2}{3})}\right)$$

The first peratized scattering length, evaluated by the optimum-n method, becomes

$$A^{(1)} = \lim_{\alpha\to 0}\alpha\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{4}{3})}n(g,\alpha)^{\frac{2}{3}} = \frac{g^{\frac{1}{3}}}{3^{\frac{2}{3}}}\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{4}{3})} = g^{\frac{1}{3}}\frac{\mathrm{Ai}'(0)}{\mathrm{Ai}(0)}$$

In obtaining the second peratized scattering length, one readily notes that the term $3\alpha\varphi n(g,\alpha)$, which superficially is less singular than unity, is actually more singular when the α dependence of $n(g,\alpha)$ is considered and dominates all other terms. The coefficients of this term, however, are the same as those which contributed to the first peratized scattering length and hence we find that $A^{(2)} = A^{(1)}$ without completion of the τ factor.

* A preliminary report of this work was presented at the 1968 Chicago Meeting of the American Physical Society, Abstract GH-13.

† Part of this work was performed while a National Research Council Postdoctoral Research Associate.

[1] F. Calogero and M. Cassandro, Nuovo Cimento **38**, 760 (1965).

[2] H. H. Aly, Riazuddin, and A. M. Zimerman, Nuovo Cimento **35**, 324 (1965).

[3] W. M. Frank, "Relative Jost Function," and W. M. Frank and D. J. Land, "Transitionally Singular Potentials," preprints.

[4] W. A. Gale, J. Math. Phys. **8**, 1050 (1967).

[5] See e.g., N. N. Khuri and A. Pais, Rev. Mod. Phys. **36**, 590 (1964).

[6] F. Calogero, Phys. Rev. **139**, B602 (1965); N. N. Khuri and A. Pais, Ref. 5, see comment at end of Sec. II; R. Vasudevan, K. Venkatesan, and G. Jagannathan, Suppl. Nuovo Cimento **5**, 621 (1967); H. Cornille, Nuovo Cimento **43**, 786 (1966).

[7] (a) H. Cornille, Nuovo Cimento **38**, 1243 (1965); (b) H. Cornille, Nuovo Cimento, **39**, 557 (1965); **43**, 786 (1966).

[8] This expression for the scattering length can be obtained from the power-series expression for the Jost function $f(k)$ and the relation $A = -\lim_{k\to 0}[f(-k) - f(k)]/2ikf(k)$. The expression for the Jost function may be found in R. G. Newton, J. Math. Phys. **1**, 319 (1960). See also Ref. 5.

[9] See e.g., *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, Eds., (National Bureau of Standards, Applied Mathematics Series, 55, Washington, D.C., 1964), pp. 374–378.

[10] An entire function $f(z)$ has exponential order ρ if the inequality $|f(z)| \leq \exp M|z|^{\rho+\epsilon}$ can be satisfied for all $\epsilon > 0$ for some M, and for no $\epsilon < 0$. If $f(z)$ has exponential order ρ, its type τ is defined as $\overline{\lim_{|z|\to\infty}}\,|z^{-\rho}\ln f(z)|$.

[11] W. M. Frank, J. Math. Phys. **8**, 466 (1967).

[12] W. M. Frank, J. Math. Phys. **9**, 1890 (1968).

[13] W. M. Frank and A. Shaki, Bull. Am. Phys. Soc. **12**, 717 (1967). Article in preparation. See also D. Masson, J. Math. Phys. **8**, 2308 (1967), Sec. V.

[14] G. Valiron, *Integral Functions* (Chelsea Publishing Co., New York, 1949), p. 32.

[15] G. Valiron, Ref. 14, Sec. 4–7. M. A. Evgrafov, *Asymptotic Estimates and Entire Functions* (Gordon and Breach, New York, 1961), Chap. II, Secs. 1 and 2.

[16] M. A. Evgrafov, Ref. 15, p. 64 and Eq. (23).

[17] The upper bound has been derived in Refs. 11 and 12 and in K. Chadan, Nuovo Cimento **58A**, 191 (1968). The lower bound is derived in W. M. Frank and D. J. Land, "Some Inequalities for the Scattering Length," article in preparation. A similar lower bound is also presented by Chadan, loc. cit.

[18] See e.g., E. Hille, *Analytic Function Theory* (Ginn and Co., New York, 1962), Vol. II, Theorem 14.1.2.

[19] Ref. 9, pp. 258–259. Note that $\psi(1) = -\gamma$.

[20] Ref. 1, Eq. (19).

[21] See also, D. J. Land, "Peratization of a Generalized Class of Exponentially Singular Potentials," preprint.

[22] We note that, although this expansion is not analytic in g at $g = 0$, it is nevertheless convergent.

[23] Ref. 1, Eq. (8) with $\nu = 0$, and g replaced by $4g/\lambda^2$ to adjust to our potential, where we employ $e^{\lambda/r}$ in place of $e^{2/r}$. The $+\alpha$ has been omitted in this reference. The importance of this term for the high α limit has been pointed out to us by R. Spector.

[24] One notes that the concept of the appropriate singular unit may entail a differing number of terms in $N_n(\alpha)$ and $D_n(\alpha)$ to a given order of peratization. For example, for $m = 4$ the first nontrivial result originates in a single term in $N_n(\alpha)$ and two terms in $D_n(\alpha)$ [see Eqs. (20) or (C2)]. CC (Ref. 1) also found it necessary to proceed in this way in their peratization procedure.

[25] This definition of a set of spurious basis functions is not the most general and does not directly apply if an extraneous factor common to N and D involved the sum of a highly singular term with terms whose singularity is comparable with or less than the singularity of the basis functions of the Born series. This discussion of spurious basis functions is not intended to be comprehensive.

[26] This follows in the same way as does the expression given in Ref. 8.

[27] See e.g., R. G. Newton, Ref. 8; discussion leading up to Eqs. (4.25) and (4.25′).

[28] R. G. Newton, Ref. 8.

[29] Gale, Ref. 4, and Vasudevan, Venkatesan, and Jagannathan, Ref. 6.

[30] Ref. 9, pp. 446–449.

[31] Ref. 9, p. 257.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Peratization of Logarithmically Singular Potentials*

WILLIAM M. FRANK AND DAVID J. LAND†
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland 20910

(Received 5 September 1969)

First and second peratizations are performed for a wide class of repulsive potentials with singular behavior near $r = 0$ of the form $L(r)r^{-m}$ $(m > 3)$, where $L(r)$ is "logarithmically singular" at the origin. More precisely, when $L(r)$ is of the class of "slowly varying functions," i.e., functions such that $rL'(r)/L(r) \to 0$ as $r \to 0$, first peratization gives a (generally infinite) constant multiple of the result for the r^{-m} potential. Second peratization for a wide class of slowly varying function offers no substantive improvement.

I. INTRODUCTION

Many of the model calculations of the peratization procedure in scattering theory have dealt with logarithmically singular potentials,[1,2] i.e., potentials of the form (g is a coupling constant)

$$gV(r) = gL(r)/r^m, \tag{1}$$

with $L(r)$ some power of $\ln r$. (In some calculations,[3] the potential is a sum of two such terms, but we do not consider this case.) These calculations have all found a divergent result in first peratization. These results have been derived in a more general way for a larger class of logarithmically singular potential by a number of authors.[2] In the present article, the first and second peratizations are carried out for the class of potentials of the form in Eq. (1), where $L(r)$ is a "slowly varying function" which we define subsequently. Second peratization is found not to change the results of the first peratization significantly. Our results go beyond those of previous authors in the inclusion of second peratization as well as in the broadness of the results, and in the simplicity of the derivation for the first peratization. In Sec. II, we present some preliminaries. In Sec. III, we calculate the first peratization for the two regularizations, the "θ regularization"

$$V_\theta(r, \alpha) = V(r)\theta(r - \alpha) \tag{2}$$

[$\theta(x)$ is the sign function which is 0 for negative x, unity otherwise], and the "+ regularization"

$$V_+(r, \alpha) = V(r + \alpha). \tag{3}$$

In Sec. IV, we calculate the second-peratized contribution and show that it does not make any substantive change in the results of first peratization.

II. SOME PRELIMINARIES

This article concerns itself with singular potentials of the form in Eq. (1), where $L(r)$ is a "slowly varying function" (svf)[4] in the neighborhood of $r = 0$. This is defined by the condition

$$\lim_{r\to 0} r\frac{L'(r)}{L(r)} = 0 \quad \text{or} \quad \frac{L'(r)}{L(r)} = o\left(\frac{1}{r}\right) = \frac{1}{r\lambda(r)}, \tag{4}$$

where $|\lambda(0)| = \infty$. This class of functions has also been considered by Spector[2] and Gale.[2] It is assumed, of course, that the behavior for large r is such as to correspond to a finite scattering length. The significant property of the svf is the relation

$$\lim_{\alpha\to 0} \frac{L(\alpha\rho)}{L(\alpha)} = 1, \tag{5}$$

which follows from

$$\min_{\alpha\le r\le\alpha\rho} \lambda^{-1}(r) \ln \rho \le \int_\alpha^{\alpha\rho} dx \frac{L'(x)}{L(x)} = \ln\left(\frac{L(\alpha\rho)}{L(\alpha)}\right) \le \max_{\alpha\le r\le\alpha\rho} \lambda^{-1}(r) \ln \rho. \tag{6}$$

Since $\lambda^{-1}(r) \to 0$ as $r \to 0$, both bounds in Eq. (6) give zero in the limit $\alpha \to 0$, which proves Eq. (5). We can, in fact, deduce the stronger statement from Eq. (6):

$$\frac{L(\alpha\rho)}{L(\alpha)} = 1 + O\left(\frac{\ln \rho}{\lambda(\alpha\theta)}\right), \tag{7}$$

where $1 \le \theta \le \rho$, which may, in fact, be a function of ρ. The slowly varying functions include functions such as $\ln^p (1/r)$, $\ln \ln (1/r)$, $(1 - \ln r)^{-1}$, e^{r^ϵ} (with $\epsilon > 0$), etc. It is easy to show that, if $L(r)$ and $l(r)$ are svf, then so are $L[l(r)]$ and $l[L(r)]$. Moreover, if $L(r)$ is an svf and $L(0) = a$ (which may include the improper value ∞), and if

$$a\,|F'(a)/F(a)| < \infty,$$

interpreted as a limit when the quantities assume improper values, then $F[L(r)]$ is also an svf.

We are also interested in the subclass of svf, for which the $\lambda(r)$ of Eq. (4) is again an svf. This class includes functions of the form $F(\ln r^{-1})$, where $F'(x)/F(x) \to 0$ as $x \to \infty$, and, thus, is quite general. We call this class the class of "tempered slowly varying functions" (tsvf). In the notation of Eq. (4) we thus

require that

$$\lim_{r\to 0}\frac{L'(r)}{L(r)}=\frac{1}{r\lambda(r)}, \quad \text{where} \quad \frac{\lambda'(r)}{\lambda(r)}=\frac{1}{r\mu(r)}, \tag{8}$$

and $|\mu(0)|=\infty$. We find from Eq. (6), applying Eq. (7) to $\lambda(\alpha\rho)$ as an svf, that

$$\frac{L(\alpha\rho)}{L(\alpha)}=1+\frac{\ln\rho}{\lambda(\alpha)}[1+o(1)], \tag{9}$$

providing a second term in the singularity expansion of $L(\alpha\rho)$.

The negative of the scattering length A is expressible[5] as the ratio of two power series

$$-A(g)\equiv\frac{N(g)}{D(g)}=\frac{\sum_{n=1}^{\infty}g^n\int_0^\infty dr_1\cdots\int_0^{r_{n-1}}dr_n r_1(r_1-r_2)\cdots(r_{n-1}-r_n)r_nV(r_1)\cdots V(r_n)}{\sum_{n=0}^{\infty}g^n\int_0^\infty dr_1\cdots\int_0^{r_{n-1}}dr_n(r_1-r_2)\cdots(r_{n-1}-r_n)r_nV(r_1)\cdots V(r_n)}. \tag{10}$$

The θ- and $+$ regularized scattering lengths are denoted by $+A^\theta(g,\alpha)$ and $+A^+(g,\alpha)$, respectively. In an obvious notation, we write

$$N^\gamma(g,\alpha)\equiv\sum_{n=1}^{\infty}g^nN_n^\gamma(\alpha), \quad D^\gamma(g,\alpha)\equiv\sum_{n=0}^{\infty}g^nD_n^\gamma(\alpha), \tag{11}$$

with $\gamma=\theta$ or $+$. For the pure power potential $V(r)=gr^{-m}$, we write

$$\begin{aligned}D^\theta_{n,m}(\alpha)&\equiv\int_\alpha^\infty dr_1\cdots\int_\alpha^{r_{n-1}}dr_n\frac{(r_1-r_2)\cdots(r_{n-1}-r_n)r_n}{(r_1\cdots r_n)^m}\\&=\alpha^{-(m-2)n}\int_1^\infty d\rho_1\cdots\int_1^{\rho_{n-1}}d\rho_n\\&\quad\times\frac{(\rho_1-\rho_2)\cdots(\rho_{n-1}-\rho_n)\rho_n}{(\rho_1\cdots\rho_n)^m}\\&\equiv\frac{d^\theta_{n,m}}{\alpha^{(m-2)n}},\end{aligned} \tag{12a}$$

where the integration variable substitutions $r_j=\alpha\rho_j$, $j=1,\cdots,n$, have been made. Similarly, one can write

$$N^\theta_{n,m}(\alpha)=\alpha n^\theta_{n,m}/\alpha^{(m-2)n}, \tag{12b}$$

where the $d^\theta_{n,m}$ and $n^\theta_{n,m}$ are known (see the Appendix) and are independent of α. One finds, by identical considerations, that

$$D^+_{n,m}(\alpha)=d^+_{n,m}/\alpha^{(m-2)n}, \tag{13a}$$

$$N^+_{n,m}(\alpha)=\alpha n^+_{n,m}/\alpha^{(m-2)n}, \tag{13b}$$

with $d^+_{n,m}$ and $n^+_{n,m}$ defined in terms of n-fold integrals. Therefore, one finds (in an obvious notation) that

$$-A^\gamma_m(g,\alpha)=\frac{\alpha\sum_{n=1}^{\infty}(g\alpha^{(2-m)})^nn^\gamma_{n,m}}{\sum_{n=0}^{\infty}(g\alpha^{(2-m)})^nd^\gamma_{n,m}}\equiv\frac{N^\gamma_m(g,\alpha)}{D^\gamma_m(g,\alpha)}. \tag{14}$$

One knows from general considerations[6] or from specific evaluation[7] that the limiting value, as $\alpha\to 0$ of the ratio in Eq. (14), is in fact the (negative of the) scattering length for the potential $V(r)=gr^{-m}$, which is[7]

$$\begin{aligned}-A_m(g)&=(g\nu^2)^\nu[\Gamma(1-\nu)/\Gamma(1+\nu)]\\&=-(g\nu^2)^\nu[\Gamma(-\nu)/\Gamma(\nu)],\end{aligned} \tag{15}$$

where $\nu=(m-2)^{-1}$. The limit as $\alpha\to 0$ in Eq. (14) can also be written as

$$-A_m(g)=\lim_{x\to\infty}g^\nu x^{-\nu}\sum_{n=1}^{\infty}x^nn^\nu_{n,m}\Big/\sum_{n=0}^{\infty}x^nd^\nu_{n,m}, \tag{16}$$

where the substitution $x\equiv g\alpha^{-1/\nu}$ has been made.

III. FIRST PERATIZATION

We appeal to the discussion in the preceding article[8] to justify the calculation of the sth peratization of $A^\gamma(g)$ as the ratio of the sth peratized $N^\gamma(g,\alpha)$ and $D^\gamma(g,\alpha)$ in the $\alpha\to 0$ limit. The sth-peratized scattering length for the class of potentials of the form of Eq. (1) is denoted by $A^{(s)}_{\gamma m}(g)$. Now for these potentials, with $L(r)$ an svf,

$$\begin{aligned}D^\theta_n(\alpha)&=\int_\alpha^\infty dr_1\cdots\int_\alpha^{r_{n-1}}dr_n\frac{L(r_1)\cdots L(r_n)}{(r_1\cdots r_n)^m}\\&\quad\times(r_1-r_2)\cdots(r_{n-1}-r_n)r_n\\&=\frac{1}{\alpha^{(m-2)n}}\int_1^\infty d\rho_1\cdots\int_1^{\rho_{n-1}}d\rho_n\frac{L(\alpha\rho_1)\cdots L(\alpha\rho_n)}{(\rho_1\cdots\rho_n)^m}\\&\quad\times(\rho_1-\rho_2)\cdots(\rho_{n-1}-\rho_n)\rho_n.\end{aligned} \tag{17}$$

The first peratization of this quantity (which we denote by a bar over the symbol) may be obtained by noting that in the limit as $\alpha\to 0$, the leading singular behavior is found by employing Eq. (5):

$$\bar D^\theta_n(\alpha)=\{[L(\alpha)]^n/\alpha^{(m-2)n}\}d^\theta_{n,m}. \tag{18}$$

Equation (18) remains valid just as well for $\bar{D}_n^+(\alpha)$ and $\bar{N}_n^\gamma(\alpha)$. We note that $\bar{D}_n^\theta(\alpha)$ in Eq. (15) is of the same form as $D_{n,m}^\theta(\alpha)$ of Eq. (12a) with the additional factor $[L(\alpha)]^n$, which can be viewed as effectively replacing the coupling constant g in $D^\theta(g,\alpha)$ by $gL(\alpha)$. The same holds for $D^+(g,\alpha)$ and $N^\gamma(g,\alpha)$. We immediately conclude from Eq. (14) that, for $\gamma =$ both θ and $+$,

$$\begin{aligned} A_{\gamma m}^{(1)}(g) &= -[gL(0)\nu^2]^\nu[\Gamma(1-\nu)/\Gamma(1+\nu)] \\ &= L^\nu(0)A_m(g). \end{aligned} \tag{19}$$

Thus, if $L(0) = \infty$, the first peratization produces a divergent result and can be said to fail. If $L(0) = 0$ or a finite number, one might claim that first peratization has in some sense succeeded. In the next section, we study the second peratization and show that it does not change substantively the results of the first peratization, and so does not produce any better approximation.

IV. SECOND PERATIZATION

We perform the second peratization of potentials of the form of Eq. (1), with $L(r)$ a tsvf. Second peratization is calculated from the second term on the right side of Eq. (9). We now calculate $\bar{\bar{D}}_n^\gamma(\alpha)$ and $\bar{\bar{N}}_n^\gamma(\alpha)$, the second peratization contributions to $D_n^\gamma(\alpha)$ and $N_n^\gamma(\alpha)$, respectively, from this term. If one retains the two leading singular contributions derived from the substitution of Eq. (9) into Eq. (17), one finds

$$\begin{aligned} \bar{\bar{D}}_n^\theta(\alpha) &= \frac{[L(\alpha)]^n}{\lambda(\alpha)\alpha^{(m-2)n}} \int_1^\infty d\rho_1 \cdots \int_1^{\rho_{n-1}} d\rho_n \\ &\quad \times \frac{\left(\sum_{i=1}^n \ln \rho_i\right)(\rho_1-\rho_2)\cdots(\rho_{n-1}-\rho_n)\rho_n}{(\rho_1\cdots\rho_n)^m} \\ &= -\frac{[L(\alpha)]^n}{\lambda(\alpha)\alpha^{(m-2)n}} \frac{\partial d_{n,m}^\theta}{\partial m}, \end{aligned} \tag{20}$$

and completely parallel expressions for $\bar{\bar{D}}_n^+(\alpha)$ and $\bar{\bar{N}}_n^\gamma(\alpha)$. From the expression [see Eq. (A5) of the Appendix]

$$\begin{aligned} d_{n,m}^\theta &= [\Gamma(1+\nu)\nu^{2n-1}]/[n!\,\Gamma(n+\nu)] \\ &= [\Gamma(\nu)\nu^{2n}]/[n!\,\Gamma(n+\nu)], \end{aligned} \tag{21}$$

with $\nu \equiv (m-2)^{-1}$, one finds

$$\bar{\bar{D}}_n^\theta(\alpha) = [\nu^2/\lambda(\alpha)][2n/\nu + \psi(\nu) - \psi(n+\nu)]\bar{D}_n^\theta(\alpha), \tag{22a}$$

where $\psi(z)$ is the logarithmic derivative of the Γ function. From Eqs. (A1), (A5), (A13), and (A15), one similarly finds

$$\begin{aligned} \bar{\bar{N}}_n^\theta(\alpha) &= [\nu^2/\lambda(\alpha)][2n/\nu - \psi(-\nu) \\ &\quad + \psi(n+1-\nu)]\bar{N}_n^\theta(\alpha), \end{aligned} \tag{22b}$$

$$\begin{aligned} \bar{\bar{D}}_n^+(\alpha) &= [\nu^2/\lambda(\alpha)][2n/\nu + \psi(1+\nu) \\ &\quad - \psi(n+1+\nu)]\bar{D}_n^+(\alpha), \end{aligned} \tag{22c}$$

$$\begin{aligned} \bar{\bar{N}}_n^+(\alpha) &= \frac{\nu^2}{\lambda(\alpha)}\Bigg[\frac{2n}{\nu} + [\psi(1+\nu) - \psi(n+1+\nu)] \\ &\quad \times \left(1 - \frac{\Gamma(1-\nu)}{\Gamma(1+\nu)}\frac{\Gamma(n+1+\nu)}{\Gamma(n+1-\nu)}\right)^{-1} \\ &\quad - [\psi(1-\nu) - \psi(n+1-\nu)] \\ &\quad \times \left(1 - \frac{\Gamma(1+\nu)}{\Gamma(1-\nu)}\frac{\Gamma(n+1-\nu)}{\Gamma(n+1+\nu)}\right)^{-1}\Bigg]\bar{N}_n^+(\alpha). \end{aligned} \tag{22d}$$

We now calculate the second-peratized scattering length by the optimum-n method and according to the prescription of the completion of the τ factor. Both of these notions have been described in detail in the preceding paper.[8] The quantity $\tau(\alpha)$ is given by

$$\int_\alpha^\infty dr V^{\frac{1}{2}}(r)$$

and $\bar{\tau}(\alpha)$, which is the α-dependent part of $\tau(\alpha)$ and whose significance was discussed in the preceding article, is given to second peratization for the potentials of Eq. (1) by

$$\bar{\tau}^{(2)}(\alpha) = \bar{\tau}_1(\alpha) + \bar{\tau}_2(\alpha) = [L^{\frac{1}{2}}(\alpha)/\alpha^{1/2\nu}]2\nu[1+\nu/\lambda(\alpha)].$$

Since $\bar{\tau}_2(\alpha)$ becomes infinite in the $\alpha \to 0$ limit, it is necessary to complete the τ factor. The negative of the θ-regularized scattering length in second peratization is to be obtained by the optimum-n method from the expression

$$-A_{\theta m}^{(2)}(g,\alpha) = \frac{1 + [\nu/\lambda(\alpha)][2n(g,\alpha) - \nu\psi(-\nu) + \nu\psi(n(g,\alpha)+1-\nu)]}{1 + [\nu/\lambda(\alpha)][2n(g,\alpha) + \nu\psi(\nu) - \nu\psi(n(g,\alpha)+\nu)]} \frac{\bar{N}_{n(g,\alpha)}^\theta(\alpha)}{\bar{D}_{n(g,\alpha)}^\theta(\alpha)} \tag{23}$$

and, upon completion of the τ factor, one finds

$$-A_{\theta m}^{(2)}(g,\alpha) = \frac{[1+\nu/\lambda(\alpha)]^{2n}\{1 - [\nu^2/\lambda(\alpha)]\psi(-\nu) + [\nu^2/\lambda(\alpha)]\psi(n(g,\alpha)+1-\nu)\}}{[1+\nu/\lambda(\alpha)]^{2n}\{1 + [\nu^2/\lambda(\alpha)]\psi(\nu) - [\nu^2/\lambda(\alpha)]\psi(n(g,\alpha)+\nu)\}} \frac{\bar{N}_{n(g,\alpha)}^\theta(\alpha)}{\bar{D}_{n(g,\alpha)}^\theta(\alpha)}. \tag{24}$$

The optimum-n value $n(g, \alpha)$ for the present class of potentials is given by[9]

$$n(g, \alpha) = \nu g^{\frac{1}{2}} L^{\frac{1}{2}}(\alpha)/\alpha^{1/2\nu}. \tag{25}$$

Recalling that, for large n, $\psi(n + a)$ has the behavior $\psi(n + a) \approx \ln n$, we find that the second-peratized scattering length $A^{(2)}_{\theta m}(g)$ is given by the expression

$$-A^{(2)}_{\theta m}(g) = -[g\nu^2 L(0)]^{\nu} \frac{\Gamma(-\nu)}{\Gamma(\nu)} \lim_{\alpha \to 0} \frac{1 - \frac{1}{2}\nu \ln(\alpha)/\lambda(\alpha)}{1 + \frac{1}{2}\nu \ln(\alpha)/\lambda(\alpha)}. \tag{26}$$

Here, we have used the result, which is essentially a calculation of the scattering length for the potential gr^{-m} by the optimum-n method, that

$$\begin{aligned} \lim_{\alpha \to 0} \frac{\bar{N}^{\theta}_{n(g,\alpha)}(\alpha)}{\bar{D}^{\theta}_{n(g,\alpha)}(\alpha)} &\approx -\alpha \frac{n(g, \alpha)\Gamma(n(g, \alpha) + \nu)}{\Gamma(n(g, \alpha) + 1 - \nu)} \frac{\Gamma(-\nu)}{\Gamma(\nu)} \\ &\approx -\alpha n(g, \alpha)^{2\nu} \frac{\Gamma(-\nu)}{\Gamma(\nu)} \\ &\approx -[g\nu^2 L(0)]^{\nu} \frac{\Gamma(-\nu)}{\Gamma(\nu)} = -L^{\nu}(0) A_m. \end{aligned}$$

Similar results hold for the + regularized potential.

It thus appears from Eq. (26) that the second-peratized scattering length can differ from the first peratized scattering length at most by a constant. Whether the constant is unity or not depends on the function $\lambda(\alpha)$. There are three distinct situations. Since $\psi(x) \sim \ln x$ for large x, we see from the behavior of $n(g, \alpha)$ of Eq. (25) that $\psi(n(g, \alpha))$ behaves as $\ln \alpha^{-1}$ for small α. Thus, Eq. (26) dictates the following behaviors of the second-peratized scattering length with respect to the first-peratized scattering length as a function of $\lambda(\alpha)$:

$$\lim_{\alpha \to 0} \frac{\ln \alpha}{\lambda(\alpha)} \qquad A^{(2)}_{\gamma m},$$

$$0 \qquad +A^{(1)}_{\gamma m}, \tag{27a}$$

$$\beta^{-1} \qquad \text{const } A^{(1)}_{\gamma m}, \tag{27b}$$

$$\infty \qquad -A^{(1)}_{\gamma m}. \tag{27c}$$

Functions $\lambda(\alpha)$, such that $\ln(\alpha)/\lambda(\alpha) \to 0$ as $\alpha \to 0$, correspond, for example, to tsvf $L(r)$ of the form $a + b \ln^{-1}(1/r)$, which are constant at $r = 0$, as well as to functions like $\ln \ln (1/r)$, which are very slowly varying at the origin. The present peratization scheme gives no improvement in second peratization for these potentials.

Functions $\lambda(\alpha) = \beta^{-1} \ln \alpha$ correspond to the interesting and frequently studied class $L(r) = \ln^{\beta}(1/r)$, $\beta > 0$. We term this class of tsvf the purely logarithmic class. To evaluate the constant in Eq. (27b), we proceed according to the notion that the quantity relevant in the $\alpha \to 0$ limit, namely,

$$\frac{1 + [\nu^2/\lambda(\alpha)]\psi(n + 1 - \nu)}{1 - [\nu^2/\lambda(\alpha)]\psi(n + \nu)} \approx \frac{1 - \frac{1}{2}\nu \ln(\alpha)/\lambda(\alpha)}{1 + \frac{1}{2}\nu \ln(\alpha)/\lambda(\alpha)}, \tag{28}$$

would effectively appear in a Born series with positive powers of $\lambda^{-1}(\alpha)$ as the singular basis functions. We wish to retain in the ratio (Born series) only the contribution up to $O(\lambda^{-1}(\alpha))$. By expanding the ratio in Eq. (28), we find for the purely logarithmic potentials $V(r) = gr^{-m} \ln^{\beta}(1/r)$, for which $\lambda(\alpha)$ is given by $\beta^{-1} \ln \alpha$, that

$$A^{(2)}_{\gamma m} = A^{(1)}_{\gamma m}(1 - \nu/\beta).$$

For the potential $V(r) = gr^{-4} \ln(1/r)$, studied by Wu[1] in second peratization, we have

$$A^{(2)}_{\gamma 4} = \tfrac{1}{2} A^{(1)}_{\gamma 4},$$

a result in agreement with that obtained after a lengthy calculation by Wu.

The third class of functions $\lambda(\alpha)$, which become infinite more slowly than $\ln \alpha$, corresponds, for example, to the class of tsvf $L(r) = \exp[A \ln^{p}(1/r)]$, where $0 < p < 1$. These tsvf become infinite at $r = 0$ faster than $\ln^{\beta}(1/r)$. Second peratization of these potentials, when evaluated as a ratio, gives straightforwardly the negative at the first-peratized length. Thus, $A^{(2)}_{\gamma m}$ is a positive quantity for all values of g which is not reasonable for a potential which is purely repulsive at the origin. If the argument used above in obtaining the Born series is invoked now, the factor

$$\psi(n(g, \alpha))\lambda^{-1}(\alpha)$$

dominates in the ratio and the second-peratized scattering length would appear to diverge faster than the first as $\alpha \to 0$. We conjecture that this is the case, although further study is needed to resolve this question unambiguously. In any event, it hardly seems likely that second peratization gives any improvement to first peratization which, we recall, is itself infinite whenever $L(0)$ is.

The results of the present calculation may be said to demonstrate that second peratization "does not work" for the class of potentials considered, in the sense that it yields no improvement over the first peratized results, *even when these results are finite* (as when $|L(0)| < \infty$). Therefore, the present examples clearly erode one's confidence in the straightforward application of peratization as a calculational technique.

APPENDIX: N AND D FOR THE θ AND + REGULARIZATIONS

The quantities $D_n^\gamma(\alpha)$ and $N_n^\gamma(\alpha)$, $\gamma = \theta, +$, have been defined in Eq. (11). For potentials of the form

$$V(r) = g/r^m,$$

with $m > 3$, the quantities $d_{n,m}^\gamma$ and $n_{n,m}^\gamma$ are defined by

$$D_{n,m}^\gamma(\alpha) \equiv (1/\alpha^{(m-2)n}) d_{n,m}^\gamma, \tag{A1a}$$

$$N_{n,m}^\gamma(\alpha) \equiv (\alpha/\alpha^{(m-2)n}) n_{n,m}^\gamma, \tag{A1b}$$

as in Eqs. (12) and (13). In Eq. (12a), we see that

$$d_{n,m}^\theta = \int_1^\infty d\rho_n \cdots \int_{\rho_2}^\infty d\rho_1 \cdots \frac{(\rho_1 - \rho_2)\cdots(\rho_{n-1} - \rho_n)\rho_n}{(\rho_1 \cdots \rho_n)^m}. \tag{A2}$$

Through the succession of transformations $\rho_j = \rho_{j+1}\xi_j$, $j = 1, \cdots, n-1$, one finds

$$d_{n,m}^\theta = \beta(\mu+2)\beta(2\mu+2)\cdots\beta[(n-1)\mu+2] \times \int_1^\infty d\rho_n \rho_n^{-n\mu-1}, \tag{A3}$$

where $\mu = m - 2 = \nu^{-1}$ and

$$\beta(m) \equiv \int_1^\infty d\xi(\xi - 1)\xi^{-m} = [(m-1)(m-2)]^{-1}, \tag{A4}$$

and therefore

$$d_{n,m}^\theta = \frac{\mu\Gamma(1+\mu^{-1})}{\mu^{2n} n!\,\Gamma(n+\mu^{-1})} = \frac{\nu^{2n}\Gamma(\nu)}{n!\,\Gamma(n+\nu)} \tag{A5}$$

Similarly, one straightforwardly finds

$$n_{n,m}^\theta = \frac{\mu\Gamma(1-\mu^{-1})}{\mu^{2n}(n-1)!\,\Gamma(n+1-\mu^{-1})} = \frac{-\nu^{2n}\Gamma(-\nu)}{(n-1)!\,\Gamma(n+1-\nu)} \tag{A6}$$

The corresponding quantities for the + regularized case are evaluated with the aid of the following easily proven identities:

$$D_n^+ = D_n^\theta - \alpha\Delta_n, \tag{A7}$$

where

$$\Delta_n \equiv \int_\alpha^\infty dx_n \cdots \int_{x_2}^\infty dx_1(x_1 - x_2)\cdots(x_{n-1} - x_n)V(x_1)\cdots V(x_n), \tag{A8}$$

and[10]

$$N_n^+ = N_n^\theta - \alpha D_n^+ - \alpha\tilde{\Delta}_n, \tag{A9}$$

where

$$\tilde{\Delta}_n \equiv \int_\alpha^\infty dx_n \cdots \int_{x_2}^\infty dx_1 x_1(x_1 - x_2)\cdots(x_{n-1} - x_n)V(x_1)\cdots V(x_n). \tag{A10}$$

For the power case $V(r) = r^{-m}$, we can write

$$\Delta_n = \alpha^{-n(m-2)-1} t_{n,m}, \quad \tilde{\Delta}_n \equiv \alpha^{-n(m-2)} \tilde{t}_{n,m}, \tag{A11}$$

which explicitly exhibits the α dependence. One finds from the technique used earlier that

$$t_{n,m} = \int_1^\infty d\rho_n \cdots \int_{\rho_2}^\infty d\rho_1 \frac{(\rho_1 - \rho_2)\cdots(\rho_{n-1} - \rho_n)}{(\rho_1 \cdots \rho_n)^m} = \frac{\nu^{2n}\Gamma(\nu)}{(n-1)!\,\Gamma(n+1+\nu)}, \tag{A12}$$

so that

$$d_{n,m}^+ = d_{n,m}^\theta - t_{n,m} = [\nu^{2n}\Gamma(1+\nu)]/[n!\,\Gamma(n+1+\nu)]. \tag{A13}$$

Similarly,

$$\tilde{t}_{n,m} = -[\nu^{2n}\Gamma(-\nu)]/[n!\,\Gamma(n-\nu)], \tag{A14}$$

so that

$$n_{n,m}^+ = n_{n,m}^\theta - d_{n,m}^+ - \tilde{t}_{n,m} = -\frac{\nu^{2n}}{n!}\left(\frac{\Gamma(1+\nu)}{\Gamma(n+1+\nu)} - \frac{\Gamma(1-\nu)}{\Gamma(n+1-\nu)}\right), \tag{A15}$$

and $D_n^+(\alpha)$ and $N_n^+(\alpha)$ are determined through Eqs. (A1).

* A preliminary account of this work was presented at the winter 1968 Chicago meeting of the American Physical Society [Bull. Am. Phys. Soc., Ser. II **13**, 107 (1968)], Abstract GH-13.

† Part of this work was performed while a National Research Council Postdoctoral Research Fellow.

[1] Singular potentials of the form $g \ln^2 r/r^4$, $-g \ln r/r^4$ are respectively considered by H. H. Aly, Riazuddin, and A. H. Zimerman, Phys. Rev. **136**, B1174 (1964); T. T. Wu, *ibid.* **136**, B1176 (1964).

[2] Wider classes of logarithmically singular potentials are considered by H. Cornille [Nuovo Cimento **38**, 1243 (1965), pp. 1261–62], who works with potentials of the form $[\ln(1/r)]^\beta/r^{2n}$; also by R. M. Spector [J. Math. Phys. **7**, 2103 (1966)], who considers the same class we do, and by W. A. Gale [*ibid.* **8**, 1050 (1967)], who considers the class we do restricted to $m = 4$ in Eq. (1). Only first peratization is attempted in these articles.

[3] H. H. Aly, Riazuddin, and A. H. Zimerman, J. Math. Phys. **6**, 1115 (1965); W. A. Gale, *ibid.* **7**, 2171 (1966).

[4] M. A. Evgrafov, *Asymptotic Estimates and Entire Functions* (Gordon and Breach, New York, 1961), Chap. II, Sec. 1. They are known as "slowly increasing functions" (though they may be decreasing) in this work.

[5] W. M. Frank and D. J. Land, J. Math. Phys. **11**, 2041 (1970); Ref. 8 of preceding article.

[6] W. M. Frank and D. J. Land, J. Math. Phys. **11**, 2041 (1970); Ref. 7 of preceding article.

[7] See, e.g., N. N. Khuri and A. Pais, Rev. Mod. Phys. **36**, 590 (1964).

[8] W. M. Frank and D. J. Land, J. Math. Phys. **11**, 2041 (1970) (preceding article).

[9] W. M. Frank and D. J. Land, J. Math. Phys. **11**, 2041 (1970), Eq. (25). One can easily verify that, as a function of the expansion parameter $g[\bar{\tau}(\alpha)]^2$ or $g[\bar{\tau}^{(s)}(\alpha)]^2$, the exponential order ρ of $N^\gamma(g,\alpha)$ and $D^\gamma(g,\alpha)$ is $\frac{1}{2}$ and the type τ is unity.

[10] W. M. Frank and D. J. Land, J. Math. Phys. **11**, 2041 (1970) (preceding article), Appendix B, Eq. (B7).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

One-Dimensional Phase Transitions in the Spherical Lattice Gas from Inverse Power Law Potentials*

KENNETH H. CARPENTER†
Department of Physics, Texas Christian University, Fort Worth, Texas 76129
AND
JOSEPH L. STRECKER
Department of Physics, Wichita State University, Wichita, Kansas 67208

(Received 10 May 1966; Revised Manuscript Received 29 July 1969)

The spherical model of the lattice gas of Gersch and Berlin is studied by making an integral approximation to the Fourier coefficients of the interaction potential. A phase transition is found in one dimension for potentials $-gr^{-\alpha}$ with $1 < \alpha < 2$. A necessary condition which a potential must satisfy to allow a 1-dimensional phase transition is given. General expressions are derived for specific heat C_v and isothermal compressibility K_T. Discontinuities are found in the thermodynamic properties at the transition for α less than $\frac{3}{2}$ times the number of dimensions. Asymptotic expansions are given for C_v and K_T near the critical point.

1. INTRODUCTION

The spherical model of a ferromagnet introduced by Berlin and Kac[1] exhibits a phase transition which is mathematically due to the "sticking" of the saddle point in a steepest-descents integration in the evaluation of the partition function. This result extends at once to the spherical model of a lattice gas[2] giving a phase transition in three dimensions (but not in one or two dimensions) for short-range interactions. Gersh[3] and Strecker[4] obtained a transition in the spherical model in one dimension by using the interaction potential $-g\gamma e^{-\gamma r}$ and taking the limit as $\gamma \to 0$ at an appropriate point in the calculation. We show in the following that a transition due to a "sticking" saddle point occurs in one dimension in the spherical lattice gas for potentials $-gr^{-\alpha}$ with $1 < \alpha < 2$. In the transition region, the isotherms are flat. We also show that a necessary condition for a potential to yield a 1-dimensional transition in the model is for the large-r behavior to be between r^{-1} and r^{-2}. (Note that potentials such as $r^{-2} \ln r$ satisfy this requirement.[5]) This transition in the spherical lattice gas for potentials $-gr^{-\alpha}$ corresponds to the transition found by Joyce[6] in the spherical model of the ferromagnet for potentials of the same type, just as the transition found by Gersch and Berlin[2] in the spherical lattice gas corresponds to the transition in the spherical ferromagnet of Berlin and Kac.[1]

The critical properties of a model showing a transition are of interest due to the observations of apparent logarithmic singularities in the specific heats at constant volume of argon and oxygen[7,8] at the critical point. Yang and Yang[9] noted that the 2-dimensional lattice gas has such a singularity in C_v, while the spherical lattice gas in three dimensions shows no such singularity. We show that the general form of C_v for the spherical model yields a finite discontinuity at the transition when the potential is $-gr^{-\alpha}$ and $1 < \alpha < \frac{3}{2}$ in the 1-dimensional case. Further, the general form of the isothermal compressibility diverges at the critical point in all cases, but a finite value is retained when approaching the transition along a value $T \neq T_c$, for $1 < \alpha < \frac{3}{2}$. Similar results are found in two and three dimensions.

In Sec. 2 below, we review, following the works of Berlin, Kac, and Gersch,[1-3] the derivation of the thermodynamics of the spherical lattice gas in one dimension. In Sec. 3, the conditions which must be satisfied for a phase transition to occur in one dimension are discussed, and the potential $-gr^{-\alpha}$ is shown to provide a transition. In Sec. 4, the thermodynamic properties for the model, generalized to n dimensions, are derived and examined qualitatively at the transition; while, in Sec. 5, the properties are given asymptotic expansions on approach to the transition. Finally, the results are discussed briefly in Sec. 6.

2. THE SPHERICAL MODEL OF THE LATTICE GAS IN ONE DIMENSION

The basis of the 1-dimensional lattice gas is a 1-dimensional lattice of K sites numbered $1, 2, \cdots, j, \cdots, K$ with occupation numbers $n_j = 0, 1$ assigned to each site. The distance between sites is δ, which gives a 1-dimensional volume $V = \delta K$. The total number of particles is $N = \sum_j n_j$, which gives a specific volume $v = V/N = \delta K/N$. For convenience in calculation, the lattice is assumed to be topologically deformed into a ring such that sites 1 and $K + 1$ are equivalent. To obtain a phase transition, it is necessary to let the number of particles and sites become infinite.[10] This limit is taken in such a way as to leave the ratio $K/N = v/\delta$ constant.

For an imperfect gas in one dimension, the partition function is

$$Q_N = (2\pi m/h^2\beta)^{\frac{1}{2}N} Z_N, \tag{2.1}$$

with the configurational partition function for the lattice gas given by

$$Z_N = \delta^N \sum_{\{n_j\}} \exp\left(-\tfrac{1}{2}\beta \sum_{j,k=1}^{K} U_{jk} n_j n_k\right), \quad \sum n_j = N, \quad n_j = 0, 1, \tag{2.2}$$

and where U_{jk} is the interaction potential between sites j and k.

Next, make a change of variables to $x_j = 2n_j - 1$ to obtain a form for Z_N like that for the Ising model. Now

$$x_j = \pm 1 \quad \text{and} \quad \sum_{j=1}^{K} x_j = 2N - K.$$

The periodicity imposed on the lattice requires $U_{ij} = U_{i+l,j+l}$, where addition of indices is modulo K. Hence,

$$Z_N = \delta^N e^{-\frac{1}{2}N\beta v_1} \sum_{\{x_j\}} \exp\left(\tfrac{1}{16}\beta \sum_{j,l} U_{jl}(x_j - x_l)^2\right), \tag{2.3}$$

with

$$v_1 = \sum_{j=1}^{K} U_{1j}.$$

The spherical model of the 1-dimensional lattice gas is obtained by replacing the summation over $\{x_j\}$ configurations with x_j restricted to ± 1 in (2.3) by integration over the x_j:

$$\sum_{\{x_j\}} \to A \int_{-\infty}^{+\infty} dx_1 \cdots dx_K,$$

subject to the constraints

$$\sum_{j=1} x_j = 2N - K \quad \text{and} \quad \sum_{j=1}^{K} x_j^2 = K.$$

The latter constraint limits integration to the surface of a K-dimensional sphere and is the origin of the name "spherical model." [1] The normalization factor A is required on changing from a sum to an integral, and can be a function of K only. The choice

$$\ln A = (Nv/2\delta) \ln (2/\pi e) \tag{2.4}$$

is made to obtain the same Z_N as the ideal lattice gas with $U_{jk} = 0$ and $N = \frac{1}{2}K$. The partition function now has the form

$$Z_N = A\delta^N e^{-\frac{1}{2}N\beta v_1} \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_K \times \exp\left(\tfrac{1}{16}\beta \sum U_{jl}(x_j - x_l)^2\right), \quad \sum x_j = 2N - K, \ \sum x_j^2 = K.$$

The properties of the interaction $U_{jk} = U_{kj}$ and $U_{jk} = U_{j+m,k+m}$ allow the matrix $\{U_{jk}\}$ to be diagonalized by the orthogonal transformation

$$y_i = \sum_{j=1}^{K} t_{ij} x_j, \quad \sum_{j=1}^{K} y_j^2 = \sum_{j=1}^{K} x_j^2,$$

with

$$t_{jk} = K^{-\frac{1}{2}}\{\cos[(2\pi/K)(j-1)(k-1)] + \sin[(2\pi/K)(j-1)(k-1)]\}. \tag{2.5}$$

The eigenvalues of U_{ij} are

$$v_j = 2 \sum_{p=2}^{\frac{1}{2}(K+1)} U_{1p} \cos\left(\frac{2\pi}{K}(j-1)(p-1)\right). \tag{2.6}$$

The restriction on the integral in Z_N of

$$\sum_{j=1}^{K} x_j = 2N - K \quad \text{becomes} \quad y_1 = K^{-\frac{1}{2}}(2N - K),$$

and

$$\sum_{j=1}^{K} x_j^2 = K \quad \text{becomes} \quad \sum_{j=2}^{K} y_j^2 = 4N\left(1 - \frac{\delta}{v}\right) \equiv R^2.$$

With these substitutions, the partition function becomes

$$Z_N = A\delta^N e^{-\frac{1}{2}N\beta v_1} \int_{-\infty}^{\infty} dy_2 \cdots \int_{-\infty}^{\infty} dy_K \times \exp\left(\tfrac{1}{8}\beta \sum_{j=2}^{K} (v_1 - v_j) y_j^2\right), \quad \sum_{j=2}^{K} y_j^2 = R^2. \tag{2.7}$$

Next, the restriction on the integral is removed by inserting

$$\delta\left[R - \left(\sum_{j=2}^{K} y_j^2\right)^{\frac{1}{2}}\right]$$

in the integrand. To allow integration on the y_j, the δ function is represented by the complex integral

$$\delta\left(R - \left(\sum_{2}^{K} y_j^2\right)^{\frac{1}{2}}\right) = \frac{R}{i\pi} \int_{S_0 - i\infty}^{S_0 + i\infty} dS \exp S\left(R^2 - \sum_{2}^{K} y_j^2\right).$$

Thus, interchanging orders of integration, one obtains

$$Z_N = \frac{AR}{i\pi} \delta^N e^{-\frac{1}{2}N\beta v_1} \int_{S_0 - i\infty}^{S_0 + i\infty} dS e^{SR^2} \int_{-\infty}^{\infty} dy_2 \cdots \int_{-\infty}^{\infty} dy_K \times \exp\left(-\sum_{j=2}^{K} [S + \tfrac{1}{8}\beta(v_j - v_1)] y_j^2\right)$$

$$= \frac{AR}{i\pi} \delta^N e^{-\frac{1}{2}N\beta v_1} \int_{S_0 - i\infty}^{S_0 + i\infty} dS (e^{SR^2}) \times \prod_{j=2}^{K} \{\pi^{\frac{1}{2}}[S + \tfrac{1}{8}\beta(v_j - v_1)]^{-\frac{1}{2}}\},$$

where S_0 is a sufficiently large positive number to make each y_j integral converge.

The next step is to evaluate the complex integral on S. To begin, we write

$$\prod_{j=2}^{K}\{[S+\tfrac{1}{8}\beta(\nu_j-\nu_1)]^{-\frac{1}{2}}\} = \exp\left(-\tfrac{1}{2}\sum_{j=2}^{K}\ln[S+\tfrac{1}{8}\beta(\nu_j-\nu_1)]\right)$$

and let

$$G(S)=\lim_{K\to\infty}K^{-1}\sum_{2}^{K}\ln[S+\tfrac{1}{8}\beta(\nu_j-\nu_1)] = \frac{1}{2\pi}\int_0^{2\pi}d\omega\ln\{S+\tfrac{1}{8}\beta[\nu(\omega)-\nu(0)]\}, \quad (2.8)$$

with (2.6) replaced by

$$\nu(\omega)-\nu(0)=2\sum_{p=2}^{\frac{1}{2}(K+1)}U_{1p}\{\cos[\omega(p-1)]-1\} \quad (2.9)$$

on change to the continuous variable

$$\omega=(2\pi/K)(j-1).$$

Now $G(S)$ has branch points at values $S=-\frac{1}{8}\beta(\nu_j-\nu_1)$; therefore, in order to make $G(S)$ an analytic function, we place a branch cut along the real axis from $S=-\tilde{S}$ to $S=-\infty$, for $\tilde{S}$ the minimum of $\frac{1}{8}\beta[\nu(\omega)-\nu(0)]$ on $0\le\omega<2\pi$. (Thus, $\tilde{S}\le 0$.) Further, the singular nature of $G(S)$ is contained in the terms $\ln(S+\tilde{S})$; hence, we remove these from the summation before letting $K\to\infty$. Since the ν_j are doubly degenerate, there are two such terms. Substitution in Z_N gives

$$Z_N=C(2\pi i)^{-1}\int dS(S+\tilde{S})^{-1}e^{Ng(S)},$$

where

$$g(S)=4[1-(\delta/v)]S-[v/(2\delta)]G(S),$$
$$C=2AR\delta^N\pi^{\frac{1}{2}(K-1)}e^{-\frac{1}{2}N\beta v_1}, \quad (2.10)$$

and the S integration is on $-i\infty$ to $+i\infty$ through $S=S_0$.

The form of Z_N in Eq. (2.10) is that required for steepest-descents integration. If there exists an S_s such that $g'(S_s)=0$ and $g''(S_s)>0$ with $-\tilde{S}<S_s$, then

$$Z_N=\frac{C}{2\pi i}\int_{S_s-i\infty}^{S_s+i\infty}dS(S+\tilde{S})^{-1} \times\exp N[g(S_s)+\tfrac{1}{2}g''(S_s)(S-S_s)^2+\cdots] \xrightarrow[N\to\infty]{} C(S_s+\tilde{S})^{-1}[2\pi Ng''(S_s)]^{-\frac{1}{2}}e^{Ng(S_s)}. \quad (2.11)$$

From Eq. (2.10),

$$g'(S)=4[1-(\delta/v)]-[v/(2\delta)]I(S,T),$$

where

$$I(S,T)=\frac{1}{2\pi}\int_0^{2\pi}\{S+\tfrac{1}{8}\beta[\nu(\omega)-\nu(0)]\}^{-1}d\omega. \quad (2.12)$$

Hence, the saddle-point equation $g'(S_s)=0$ becomes

$$8(\delta/v)[1-(\delta/v)]=I(S_s,T). \quad (2.13)$$

The condition $g''(S_s)>0$ is fulfilled, since

$$g''(S)=(v/2\delta)(2\pi)^{-1}\times\int_0^{2\pi}\{S+\tfrac{1}{8}\beta[\nu(\omega)-\nu(0)]\}^{-2}d\omega>0.$$

The left-hand side of the saddle-point equation $8(\delta/v)[1-(\delta/v)]$ varies from 0 at $v=\delta$ to a maximum of 2 at $v=2\delta$ to 0 as $v\to\infty$. Hence, for a given β, an S_s exists for all v, provided $I(S,T)\gtrsim 2$ for some $S>-\tilde{S}$. Hence, if

$$\lim_{S\to-\tilde{S}}I(S,T)\equiv I(-\tilde{S},T)$$

exists, then for β sufficiently large (T sufficiently small) $I(-\tilde{S},T)<2$ and there is a range of v for which no solution for S_s exists. This is the "2-phase" region with the second phase appearing at the transition volumes v_l and v_g satisfying $8(\delta/v)[1-(\delta/v)]=I(-\tilde{S},T)$. The critical point occurs when $I(-\tilde{S},T_c)=2$. For $T>T_c$, $I(-\tilde{S},T)>2$ and S_s exists for all v. The region $v<\delta$ is excluded owing to the negative values for $1-(\delta/v)$. This arises owing to the "hard rod" of length δ constraint inherent in the 1-dimensional lattice gas.

Alternately, if

$$\lim_{S\to-\tilde{S}}I(S,T)=\infty,$$

then a solution S_s exists for all v and T, one analytic form results for all conditions, and no transition phenomenon is obtained. These conditions are illustrated in Figs. 1 and 2.

When S_s does not exist, the integral in Z_N must be evaluated by another method. The contour may not be deformed to pass through $S=-\tilde{S}$, but may be changed to encircle $S=-\tilde{S}$ at an arbitrarily small distance. Then, since $g'(-\tilde{S})>0$ and finite, for N large, the contribution to Z_N from a path near S real for $\mathrm{Re}\,S<-\tilde{S}$ is negligible, and a new steepest-descents contour is obtained having a nonzero contribution only along the circle about $-\tilde{S}$. Owing to the pole in the integrand at $S=-\tilde{S}$, this integral takes the value $2\pi i$ times the residue at $-\tilde{S}$. Hence, by following this new contour, one finds that the value of Z_N when

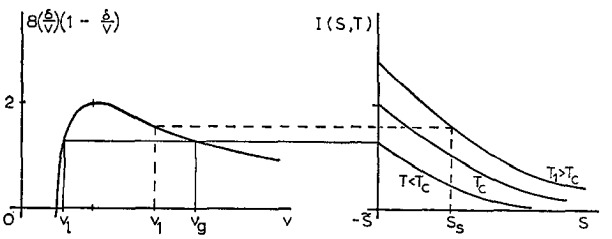


FIG. 1. Solution for T_c, v_c, v_l, v_g; and for S_s given v, T; for the case where $I(-\tilde{S},T)$ exists.

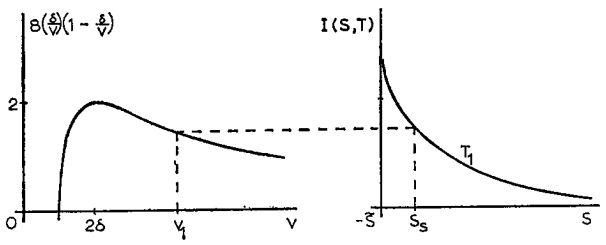


FIG. 2. Solution for S_s when $I(-\tilde{S}, T) \to \infty$.

S_s does not exist becomes

$$Z_N = Ce^{Ng(-\tilde{S})}. \tag{2.14}$$

It is interesting to note that solutions for the saddle point will never violate $S_s > -\tilde{S}$ if the saddle point is obtained before taking the limit $K \to \infty$. Then Eq. (2.8) remains a finite sum and contains a term

$$2K^{-1} \ln (S + \tilde{S}),$$

as noted previously. As a result, a term $2K^{-1}(S + \tilde{S})^{-1}$ must be added to $I(S, T)$ in (2.12) which, for K finite, causes $I(S, T)$ to diverge as $S \to -\tilde{S}$. For K large but finite, S_s takes on approximately its value in the limit $K \to \infty$, for $S_s > -\tilde{S}$. When the value of S_s, for $K \to \infty$, would be less than $-\tilde{S}$, then $S_s \to -\tilde{S}$ in the case K large but finite. The result of taking the limit $K \to \infty$, after solving for S_s, is to replace Z_N in the transition region as given in Eq. (2.14) by Z_N of Eq. (2.11) with $S_s = -\tilde{S}$. The additional factors in (2.11) do not make a contribution to the free energy per particle in the limit $N \to \infty$ and, hence, either method yields the same thermodynamic properties. Hereafter, we always take $K \to \infty$ before obtaining S_s. The result of leaving K finite was first pointed out by Lewis and Wannier.[11]

The thermodynamic properties are obtained from the free energy per particle

$$\psi = \lim_{N \to \infty} \frac{1}{N}\left(-\frac{1}{\beta}\right) \ln Q_N.$$

When S_s exists, Eq. (2.11), with Eqs. (2.1) and (2.10), gives

$$-\beta\psi_{\mathrm{I}} = \frac{v}{2\delta} \ln\left(\frac{2}{e}\right) + \ln \delta - \tfrac{1}{2}\beta\nu(0) + \tfrac{1}{2} \ln\left(\frac{2\pi m}{h^2\beta}\right) + g(S_s). \tag{2.15}$$

When S_s does not exist, Eq. (2.14) gives

$$-\beta\psi_{\mathrm{II}} = \frac{v}{2\delta} \ln\left(\frac{2}{e}\right) + \ln \delta - \tfrac{1}{2}\beta\nu(0) + \tfrac{1}{2} \ln\left(\frac{2\pi m}{h^2\beta}\right) + g(-\tilde{S}). \tag{2.16}$$

Hence, as far as thermodynamic properties are concerned, the phase transition is mathematically due to "sticking" of the saddle point S_s at the branch point $-\tilde{S}$.

3. REQUIREMENTS FOR A 1-DIMENSIONAL TRANSITION

For a transition to occur due to S_s "sticking" at $-\tilde{S}$, it is necessary and sufficient that $I(-\tilde{S}, T)$ be finite. One may write Eq. (2.12) at $S = -\tilde{S}$ as

$$I(-\tilde{S}, T) = (2\pi)^{-1} \int_0^{2\pi} [f(\omega)]^{-1}\, d\omega, \tag{3.1}$$

with

$$f(\omega) = -\tilde{S} + \tfrac{1}{8}\beta[\nu(\omega) - \nu(0)].$$

Therefore, the condition for a transition in one dimension is that the function $f(\omega)$ goes to zero as $(\omega - \omega_m)^q$ with $q < 1$ for $\omega \to \omega_m$, where ω_m is the value of ω making $f(\omega)$ a minimum. Since $\cos\theta$ is an analytic function, such a behavior for $\nu(\omega)$ requires that $U(r)$ be such that the order of summation on p in Eq. (2.9) and summation of an expansion of $\cos[\omega(p-1)]$ about ω_m may not be interchanged.

A closed form for $\nu(\omega)$ in Eq. (2.9) can be obtained only for special cases of $U(r)$. Thus, to facilitate the investigation, we replace $\nu(\omega)$ by its integral approximation[12]

$$\nu(\omega) \approx \frac{2}{\delta} \int_\delta^\infty dr U(r) \cos\left(\frac{\omega r}{\delta}\right). \tag{3.2}$$

We require that $\nu(0)$ exist. Thus, $U(r)$ must vanish faster than r^{-1} as $r \to \infty$. Further, if $[\nu(\omega) - \nu(0)] \sim (\omega - \omega_m)^q$ with $q < 1$, then $d\nu(\omega)/d\omega$ does not exist at $\omega = \omega_m$. But in (3.2), since $U(r)$ vanishes faster than r^{-1}, then, for $\omega \neq 0$,[13]

$$\frac{d\nu(\omega)}{d\omega} = -\frac{2}{\delta^2} \int_\delta^\infty dr r U(r) \sin\left(\frac{\omega r}{\delta}\right). \tag{3.3}$$

Thus, $\omega_m = 0$ is required in order to have $q < 1$. Similarly, if $U(r)$ vanishes faster than r^{-2} as $r \to \infty$, then (3.3) holds at $\omega = 0$ and $q > 1$ results. Thus, for a transition, $U(r)$ must be such that $\tilde{S} = 0$ occurs at $\omega = 0$ and must have an asymptotic behavior for large r between r^{-1} and r^{-2}.

A particular case meeting these requirements is $U(r) = -gr^{-\alpha}$, with $g > 0$ and $1 < \alpha < 2$. For $\omega = 0$, one obtains

$$\nu_\alpha(0) = 2\delta^{-1} \int_\delta^\infty -gr^{-\alpha}\, dr = -2g(\alpha - 1)^{-1}\delta^{-\alpha}, \tag{3.4}$$

while, for $\omega \neq 0$, the change of variable $x = r\omega\delta^{-1}$ gives

$$\nu_\alpha(\omega) = -2g\omega^{\alpha-1}\delta^{-\alpha} \int_\omega^\infty x^{-\alpha} \cos x\, dx. \tag{3.5}$$

Integration by parts yields

$$\nu_\alpha(\omega) = -2g\delta^{-\alpha}(\alpha-1)^{-1} \times \left(\cos\omega - \omega^{\alpha-1}\int_\omega^\infty x^{1-\alpha}\sin x\,dx\right). \quad (3.6)$$

The integral $-\omega^{\alpha-1}\int_\omega^\infty x^{1-\alpha}\sin x\,dx$ behaves for small ω as $A\omega^{\alpha-1} + \omega^2/(3-\alpha) + \cdots$, where A is a constant. Thus, the leading term in the expansion behaves as $\omega^{\alpha-1}$, and so a phase transition due to a "sticking" saddle point occurs for $1 < \alpha < 2$.

The critical temperature in the model is given by $I(0, T_c) = 2$. For $U(r) = -gr^{-\alpha}$ with g fixed, T_c varies continuously with α. Equation (2.12) gives

$$\beta_c = (kT_c)^{-1} = \frac{2}{\pi}\int_0^{2\pi}[\nu(\omega) - \nu(0)]^{-1}\,d\omega. \quad (3.7)$$

For the denominator of (3.7),[14] an evaluation of $\nu_\alpha(\omega)$ yields

$$\nu_\alpha(\omega) - \nu_\alpha(0) = \frac{2g}{\delta^\alpha}\left(\frac{\pi\Gamma(3-\alpha)}{2(\alpha-1)}\frac{\sin\frac{1}{2}\pi(2-\alpha)}{\frac{1}{2}\pi(2-\alpha)}\omega^{\alpha-1} + \sum_{l=1}^{\infty}(-1)^l\frac{\omega^{2l}}{(2l+1-\alpha)(2l)!}\right). \quad (3.8)$$

In the limit $\alpha \to 1$, the first term in the denominator of the integrand in Eq. (3.7) diverges, causing β_c to vanish as $(\alpha - 1)$. As $\alpha \to 2$, the transition ceases to occur. This result appears as a decrease of T_c to zero as $\alpha \to 2$, since, in Eq. (3.7), the first term in the denominator approaches $\frac{1}{2}\pi\omega$ and the integral for β_c diverges. Thus, T_c varies from a limit $T_c \to 0$, as $\alpha \to 2$, to $T_c \to (g/2k\delta)/(\alpha-1) \to \infty$, as $\alpha \to 1$. Thus, the more attractive the interaction, the higher the temperature required before the tendency to form a "condensed phase" is overcome.

The calculations above may be repeated for a 2-dimensional model, with the result that a transition occurs if $U(r) = -gr^{-\alpha}$ with $2 < \alpha < 4$. The transition occurs, as expected, for all $\alpha > 3$ in three dimensions.

4. THERMODYNAMIC PROPERTIES

The thermodynamic properties of the model can be obtained from the free energy per particle ψ. The generalization of Eqs. (2.15), (2.16), and (3.2) to n dimensions yields[15]

$$-\beta\psi_n = (v/2\delta)\ln(2/e) + \ln\delta - \tfrac{1}{2}\beta\nu(0) + g_n(S) + \tfrac{1}{2}n\ln(2\pi m/\beta h^2), \quad (4.1)$$

where

$$g_n(S) = 4(1-\delta/v)S - (v/2\delta)G_n(S)$$

and

$$G_n(S) = (2\pi)^{-n}\int_0^{2\pi}d\omega_1\cdots\int_0^{2\pi}d\omega_n \times \ln\{S + \tfrac{1}{8}\beta[\nu(\boldsymbol{\omega}) - \nu(0)]\},$$

with

$$\nu(\boldsymbol{\omega}) = \delta^{-1}2^{1-n}\int_{r>\delta^{1/n}}d\mathbf{r}U(r)\cos(\delta^{-1/n}\boldsymbol{\omega}\cdot\mathbf{r}).$$

In the 1-phase region, S is evaluated at S_s, while in the transition region, S is evaluated at $-\tilde{S}$.

A. Internal Energy and Specific Heat

The internal energy per particle E is given by $E = (\partial(\beta\psi)/\partial\beta)_v$. Noting that $g_n'(S_s) = 0$ is the n-dimensional form of Eq. (2.13) and that $\tilde{S}$ is proportional to β, we find

$$\left(\frac{\partial g_n(S)}{\partial\beta}\right)_v = \frac{4S}{\beta}\left(1-\frac{\delta}{v}\right) - \frac{v}{2\delta\beta}$$

and thus, from (4.1),

$$E_n = \tfrac{1}{2}nkT + \tfrac{1}{2}\nu(0) + (v/2\delta)kT - 4kT(1-\delta/v)S. \quad (4.2)$$

Again, S is to be evaluated at S_s in the 1-phase region and at $-\tilde{S}$ in the transition region.

Next, $C_v = (\partial E/\partial T)_v$ can be calculated from (4.2) with the result

$$C_v = \tfrac{1}{2}nk + \frac{v}{2\delta}k - 4k\left(1-\frac{\delta}{v}\right)\left[S + T\left(\frac{\partial S}{\partial T}\right)_v\right]. \quad (4.3)$$

Now $(\partial\tilde{S}/\partial T)_v = -\tilde{S}/T$. Thus, in the transition region $C_v = \frac{1}{2}nk + k(v/2\delta)$, a finite value independent of β. In the 1-phase region $T(\partial S_s/\partial T)_v = -\beta(\partial S_s/\partial\beta)_v$ is required. Differentiating the n-dimensional form of Eq. (2.13) with respect to β, we find

$$-\beta\left(\frac{\partial S_s}{\partial\beta}\right)_v = -S_s + \frac{\int d\omega_1\cdots\int d\omega_n\{S_s + \frac{1}{8}\beta[\nu(\omega)-\nu(0)]\}^{-1}}{\int d\omega_1\cdots\int d\omega_n\{S_s + \frac{1}{8}\beta[\nu(\omega)-\nu(0)]\}^{-2}}. \quad (4.4)$$

By using (4.4) and (2.13) in (4.3), we find that, in the 1-phase region,

$$C_v = \tfrac{1}{2}nk + k(v/2\delta) + 16k(2\delta/v)(1-\delta/v)^2[G''(S_s)]^{-1}, \quad (4.5)$$

where

$$G''(S_s) = -(2\pi)^{-n}\int d\omega_1\cdots\int d\omega_n \times \{S_s + \tfrac{1}{8}\beta[\nu(\omega)-\nu(0)]\}^{-2} < 0.$$

When the potential is such that $G''(S_s)$ diverges as $S_s \to -\tilde{S}$, the specific heat is continuous at the transition—even at the critical point. However, if $G''(-\tilde{S})$ is finite, then a discontinuity occurs at the transition. At the critical point, the discontinuity in C_v has a value $4k[G''(-\tilde{S})]^{-1}$. In the case of a 1-dimensional transition with potential $-gr^{-\alpha}$,

$$[\nu_\alpha(\omega) - \nu_\alpha(0)] \sim \omega^{\alpha-1} + O(\omega^2).$$

Hence, for $G''(0)$ to be finite, it is necessary that $0 < 2(\alpha - 1) < 1$ or that $1 < \alpha < \frac{3}{2}$. Thus, the 1-dimensional transition using the $-gr^{-\alpha}$ potential gives a continuous C_v at the transition for $\frac{3}{2} \leq \alpha < 2$ and a finite discontinuity in C_v for $1 < \alpha < \frac{3}{2}$. Similarly, the 2-dimensional transition with $U(r) = -gr^{-\alpha}$ and $2 < \alpha < 4$ gives a finite discontinuity in C_v for $2 < \alpha < 3$ and a continuous C_v for $3 \leq \alpha < 4$. In three dimensions, $U(r) = -gr^{-\alpha}$ gives a transition for all $\alpha > 3$, but the finite discontinuity in C_v at the transition occurs only for $3 < \alpha < \frac{9}{2}$.

B. Pressure and Compressibility

The pressure can be found from $P = -(\partial\psi/\partial v)_\beta$ to give

$$P_n = \frac{1}{2\delta\beta} \ln(\tfrac{1}{2}e) + \frac{4\delta}{\beta v^2} S - \frac{G_n(S)}{2\delta\beta}, \qquad (4.6)$$

where S is evaluated at S_s in the 1-phase region and at $-\tilde{S}$ in the transition region. $G_n(S)$ does not contain v explicitly and $-\tilde{S}$ does not depend on v. Hence, for cases $-\tilde{S} = 0$, the well-known property of the spherical model of constant pressure isotherms in the transition region is obtained. Now from (4.1), $G_n(0) = \ln(T_c/T) + G_n(0)|_{T=T_c}$ so for $-\tilde{S} = 0$ we have

$$P_n = \frac{kT}{2\delta}\left[\ln\left(\frac{eT}{2T_c}\right) - G_n(0)|_{T=T_c}\right], \qquad (4.7)$$

which explicitly shows the dependence of P on T in the transition region in the spherical model.

In the 1-phase region the compressibility can be obtained from

$$\left(\frac{\partial P_n}{\partial v}\right)_\beta = -\frac{8\delta}{\beta v^3} S_s + \left(\frac{\partial S_s}{\partial v}\right)_\beta \left(\frac{4\delta}{\beta v^2} - \frac{1}{2\delta\beta} G'(S_s)\right).$$

Differentiation of the saddle-point equation with respect to v gives

$$\left(\frac{\partial S_s}{\partial v}\right)_\beta = -\frac{16\delta^2}{v^3}\left(\frac{v}{2\delta} - 1\right)[G''(S_s)]^{-1},$$

so that

$$\left(\frac{\partial P_n}{\partial v}\right)_\beta = -\frac{8\delta}{\beta v^3}\left[S_s - \frac{16\delta^2}{v^2}\left(\frac{v}{2\delta} - 1\right)^2 [G''(S_s)]^{-1}\right]. \qquad (4.8)$$

Thus, for $v \neq 2\delta$ and $-\tilde{S} = 0$, $(\partial P_n/\partial v)_\beta$ approaches zero as the transition is approached for the same cases where the specific heat is continuous at the transition, and $(\partial P_n/\partial v)_\beta$ is discontinuous when C_v is discontinuous. However, $(\partial P_n/\partial v)_\beta \to 0$ as the critical point is approached in all cases (for $\tilde{S} = 0$) owing to the factor $[(v/2\delta) - 1]^2$ in Eq. (4.8). Thus, the isothermal compressibility $K_T \equiv -v^{-1}(\partial v/\partial P)_\beta$ has the value

$$K_T = \frac{\beta v^2}{8\delta}\left[S_s - \frac{16\delta^2}{v^2}\left(\frac{v}{2\delta} - 1\right)^2 \frac{1}{G''(S_s)}\right]^{-1} \qquad (4.9)$$

and, therefore, becomes infinite at the critical point in the spherical model for all potentials giving $\tilde{S} = 0$.

5. ASYMPTOTIC BEHAVIOR

The asymptotic behavior of the thermodynamic properties near the transition in the 1-phase region can be obtained in a manner similar to that used by Joyce.[6] Consider the 1-dimensional case for simplicity. Near the transition, $0 < S_s \ll 1$ and the integral in Eq. (2.12) for $I(S, T)$ can be expanded to lowest order in S as follows:

$$I(0, T) = (2\pi)^{-1}\int_0^{2\pi} \{\tfrac{1}{8}\beta[\nu(\omega) - \nu(0)]\}^{-1}\, d\omega = \frac{2\beta_c}{\beta}.$$

Hence,

$$I(S, T) = \frac{2\beta_c}{\beta} - (2\pi)^{-1}\int_0^{2\pi} \{\tfrac{1}{8}\beta[\nu(\omega) - \nu(0)]\}^{-1} \times \{1 + S^{-1}\tfrac{1}{8}\beta[\nu(\omega) - \nu(0)]\}^{-1}\, d\omega.$$

Since $I(0, T)$ exists,

$$\tfrac{1}{8}\beta[\nu(\omega) - \nu(0)] = D\omega^q + O(\omega^{q+\eta}),$$

where q, D, and η are positive constants and $q < 1$. (When the potential is $-gr^{-\alpha}$, then $q = \alpha - 1$.) A change of variable to $x = D\omega^q S^{-1}$ gives, to lowest order in S,

$$I(S, T) = \frac{2\beta_c}{\beta} - (2\pi)^{-1} D^{-1/q} S^{(1/q)-1} \times \int_0^{(2\pi)^q D/S} (1 + x)^{-1} x^{(1/q)-2}\, dx. \qquad (5.1)$$

For $q^{-1} - 2 < 0$, the integral in (5.1) converges as $S \to 0$ to the value

$$\int_0^\infty \frac{x^p}{1 + x}\, dx = \frac{\pi}{\sin(p + 1)\pi}, \quad -1 < p < 0.$$

Thus, for $S \to 0$ to lowest order in S,

$$I(S, T) - 2\beta_c/\beta \sim -\tfrac{1}{2} S^{(1/q)-1} D^{-1/q} / \sin[(q^{-1} - 1)\pi], \quad \text{for } \tfrac{1}{2} < q < 1. \qquad (5.2)$$

For $q^{-1} - 2 \geq 0$, the integral in (5.1) diverges as $S \to 0$. In that case, the identity $x(1+x)^{-1} \equiv 1 - (1+x)^{-1}$ can be used to obtain an asymptotic expansion of the integral with the result that, for $0 < q < \frac{1}{2}$ to lowest order in S as $S \to 0$,

$$I(S, T) - 2\beta_c/\beta \sim -(2\pi)^{-2q} D^{-2}(q^{-1} - 2)^{-1} S. \quad (5.3)$$

For the special case $q = \frac{1}{2}$, one finds, to lowest order in S as $S \to 0$,

$$I(S, T) - 2\beta_c/\beta \sim -(2\pi)^{-1} D^{-2} S(-\ln S). \quad (5.4)$$

Since $G''(S) = \partial I(S, T)/\partial S$, one finds from Eqs. (5.2)–(5.4) that, as $S \to 0$ to lowest order in S,

$$\begin{aligned} -\frac{1}{G''(S)} &\sim 2D^{1/q}(q^{-1} - 1)^{-1} \sin[(q^{-1} - 1)\pi] S^{2-1/q}, \\ &\qquad \tfrac{1}{2} < q < 1, \\ &\sim 2\pi D^2(-\ln S)^{-1}, \quad q = \tfrac{1}{2}, \\ &\sim (2\pi)^{2q} D^2(q^{-1} - 2), \quad 0 < q < \tfrac{1}{2}. \end{aligned} \quad (5.5)$$

The values of Eqs. (5.1)–(5.3) may be used in Eq. (2.13) to solve for S_s near the transition. Let

$$\Delta = \frac{2\beta_c}{\beta} - \frac{8\delta}{v}\left(1 - \frac{\delta}{v}\right). \quad (5.6)$$

Then, in the 1-phase region near the transition, $0 < \Delta \ll 1$. Solving for S_s, we find to lowest order in Δ

$$\begin{aligned} S_s &\sim [2 \sin(q^{-1} - 1)\pi]^{q/(1-q)} D^{1/(1-q)} \Delta^{q/(1-q)}, \\ &\qquad \tfrac{1}{2} < q < 1, \\ &\sim 2\pi D^2 \Delta(-\ln \Delta)^{-1}, \quad q = \tfrac{1}{2}, \\ &\sim (2\pi)^{2q} D^2(q^{-1} - 2)\Delta, \quad 0 < q < \tfrac{1}{2}. \end{aligned} \quad (5.7)$$

Substitution of (5.7) and (5.5) into (4.5) then gives the specific heat near the transition. For $\frac{1}{2} < q < 1$, the difference between C_v above and below the transition vanishes as $\Delta^{(2q-1)/(1-q)}$. For $q = \frac{1}{2}$, this difference vanishes as $(-\ln \Delta)^{-1}$. For $0 < q < \frac{1}{2}$, the finite discontinuity of $-16k(2\delta/v)(1 - \delta/v)(2\pi)^{2q} D^2(q^{-1} - 2)$ occurs in the specific heat on passing from the 1-phase into the transition region.

The isothermal compressibility near the transition results from substitution of Eqs. (5.5) and (5.7) into Eq. (4.9). Away from the critical volume $v_c = 2\delta$, K_T behaves as $G''(S_s)$ near the transition, i.e., for $\frac{1}{2} < q < 1$, $K_T \sim \Delta^{-(2q-1)/(1-q)}$; for $q = \frac{1}{2}$, $K_T \sim -\ln \Delta$; and for $0 < q < \frac{1}{2}$, $K_T \sim \Delta^0$—a constant, finite value. However, along the critical isochore $v = v_c$, K_T behaves as S_s^{-1}. Along the critical isochore $\Delta = 2(T/T_c - 1)$. Hence, for $v = v_c$ and $\frac{1}{2} < q < 1$, $K_T \sim (T - T_c)^{-q/(1-q)}$; for $q = \frac{1}{2}$, $K_T \sim -\ln(T - T_c)/(T - T_c)$; and for $0 < q < \frac{1}{2}$, $K_T \sim (T - T_c)^{-1}$.

The behavior of C_v and K_T for other approaches to the transition can be obtained by using the appropriate expression for Δ. For example, along an isochore $v \neq v_c$, $\Delta = 2(T/T_c - T_t/T_c)$, where T_t is the transition temperature, while along an isotherm

$$\Delta = -4\left(\frac{v_c}{v}\right)\left(1 - \frac{v_c}{v}\right) + 4\left(\frac{v_c}{v_t}\right)\left(1 - \frac{v_c}{v_t}\right),$$

where v_t is the transition volume.

For the potential $-gr^{-\alpha}$, one has, from Eq. (3.8), $q = \alpha - 1$ and

$$D = \frac{g\beta}{4\delta^\alpha} \frac{\Gamma(3 - \alpha)}{(\alpha - 1)(2 - \alpha)} \sin \tfrac{1}{2}\pi(2 - \alpha).$$

From (3.7) and (3.8), we see β_c is proportional to $g^{-1}\delta^\alpha$. Hence, at the critical point, D is independent of cell size δ and potential strength g, and so the critical properties of the model depend only on the choice of exponent α. The critical temperature depends on g and δ only through the factor $g\delta^{-\alpha}$. Hence, at least in a formal sense, the spherical lattice gas with potential $-gr^{-\alpha}$ corresponds to a continuum gas in the limit $\delta \to 0$, provided the potential strength vanishes as δ^α.

For two or three dimensions, the results are qualitatively similar. Conversion of Eq. (5.1) to a multiple integral shows that the behavior of $I(S, T) - 2\beta_c/\beta$ is as $S^{n/q-1}$ for $q > \frac{1}{2}n$, as $(-\ln S)S$ for $q = \frac{1}{2}n$, and as S for $q < \frac{1}{2}n$; n being the number of dimensions. Since in the case of the $-gr^{-\alpha}$ potential, $q = \alpha - n$ or 2 (whichever is smaller), the finite value for $G''(0)$ occurs for $\alpha < \frac{3}{2}n$ as was noted above. Along $v = v_c$, K_T generalizes to $K_T \sim (T - T_c)^{-q/(n-q)}$ for $\frac{1}{2}n < q \leq 2$, $n = 2, 3$. The other critical properties generalize in a similar manner. These critical exponents are precisely the same as those obtained by Joyce[6] for the corresponding thermodynamic properties of the spherical ferromagnet with the same interaction potential, in keeping with the general correspondence of the two systems.[2]

6. DISCUSSION

In the evaluation of the partition function made above to obtain a 1-dimensional phase transition, the approximate Fourier coefficients $\nu(\omega)$ of Eq. (3.2) were substituted for the exact Fourier coefficients of the potential given by Eq. (2.9). Although the result is not then, in general, an exact solution of the spherical lattice gas of Gersch and Berlin, the qualitative features of the general spherical model, such as constant pressure in the 2-phase region,[2] still apply. The equations resulting from this substitution are still a legitimate place to search for phase transitions,

and we anticipate that the results obtained will also be qualitatively similar to the exact solution. From Eq. (3.3), one sees readily that a necessary condition for a phase transition in this 1-dimensional model is that the potential have an asymptotic form for large r between r^{-1} and r^{-2}. We note that a potential such as $r^{-2} \ln r$ satisfies this requirement.[5]

The spherical model is such that the specific heat remains finite at the critical point in any number of dimensions for all potentials which the formalism will admit. However, for the inverse power law potentials considered above, a finite discontinuity occurs in C_v at the transition for the power α sufficiently small. Further, the isothermal compressibility, while divergent at the critical point, does retain a finite value as the transition is approached along noncritical temperatures. Asymptotic expansion of the integrals $I(S)$ and $G''(S)$ near the transition shows that properties depend linearly on the distance from the transition for the power α less than $\frac{3}{2}$ times the number of dimensions, while for larger α properties near the transition depend upon the distance from the transition raised to a power which is a function of α. The critical properties are independent of cell size and potential strength, and are determined by the power α alone. The spherical lattice gas with a nonintegral power law potential can be tailored through choice of the power, to give a desired value for the behavior of isothermal compressibility near the critical point even though a specific heat singularity near the critical point cannot be obtained.

* Part of a dissertation submitted to Texas Christian University by Kenneth H. Carpenter in partial fulfillment of the requirements for the degree Doctor of Philosophy (May 1966) while holding an N.S.F. Cooperative Graduate Fellowship.

† Present address: Department of Physics, East Tennessee State University, Johnson City, Tennessee 37601.

[1] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[2] H. A. Gersch and T. H. Berlin, Phys. Rev. **127**, 2276 (1962).
[3] H. A. Gersch, Phys. Fluids **6**, 599 (1963).
[4] J. L. Strecker, J. Math. Phys. **10**, 1541 (1969).
[5] Y. Imry, Ann. Phys. (N.Y.) **51**, 1 (1969).
[6] G. S. Joyce, Phys. Rev. **146**, 349 (1966).
[7] M. I. Bagatskii, A. V. Voronel, and V. G. Gusak, Zh. Eksp. Teor. Fiz. **43**, 728 (1962) [Sov. Phys. JETP **16**, 517 (1963)].
[8] A. V. Voronel, Yu. R. Chashkin, V. A. Popov, and V. G. Simkin, Zh. Eksp. Teor. Fiz. **45**, 828 (1963) [Sov. Phys. JETP **18**, 568 (1964)].
[9] C. N. Yang and C. P. Yang, Phys. Rev. Letters **13**, 303 (1964).
[10] Terrell L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956), p. 164ff.
[11] H. W. Lewis and G. H. Wannier, Phys. Rev. **88**, 682 (1952); **90**, 1131 (1953). Further work on this approach is in C. C. Yan and G. H. Wannier, J. Math. Phys. **6**, 1833 (1965).
[12] T. H. Berlin, L. Witten, and H. A. Gersch, Phys. Rev. **92**, 189 (1953), Eq. (4.5).
[13] Louis Brand, *Advanced Calculus* (Wiley, New York, 1955), p. 434.
[14] K. H. Carpenter, Ph.D. thesis, Texas Christian University, 1966.
[15] See Ref. 3, Eqs. (2.15), (2.19), and (2.23).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Approximate Solutions of the Unitarity Equation

C. EFTIMIU
Department of Physics, University of Missouri at St. Louis, St. Louis, Missouri 63121

(Received 29 December 1969)

The possible use of the classical Newton approximation method, as extended by Kantorovich to Banach spaces, is discussed in connection with the problem of solving the unitarity equation for the phase of a scattering amplitude, in terms of the corresponding differential cross section.

The purpose of this paper is to discuss the usefulness of a classical approximation method as applied to the problem of finding a solution of the nonlinear integral equation

$$P[\phi] \equiv \sin \phi(x) - \lambda \iint_D h(x, y, z) \cos [\phi(y) - \phi(z)]\, dy\, dz = 0. \tag{1}$$

This equation has been shown by Newton[1] to express the unitarity condition in its simplest form, and the problem of solving it arose from the question whether the phase $\phi(x)$ of a scattering amplitude

$$F(x) = G(x) \exp [i\phi(x)]$$

can be determined in terms of its modulus $G(x)$. In this context, $G(x)$ can be directly given in terms of the differential cross section which is supposed to be known at one given energy for all scattering angles θ ($x = \cos \theta$). The unitarity condition leads then to

and we anticipate that the results obtained will also be qualitatively similar to the exact solution. From Eq. (3.3), one sees readily that a necessary condition for a phase transition in this 1-dimensional model is that the potential have an asymptotic form for large r between r^{-1} and r^{-2}. We note that a potential such as $r^{-2} \ln r$ satisfies this requirement.[5]

The spherical model is such that the specific heat remains finite at the critical point in any number of dimensions for all potentials which the formalism will admit. However, for the inverse power law potentials considered above, a finite discontinuity occurs in C_v at the transition for the power α sufficiently small. Further, the isothermal compressibility, while divergent at the critical point, does retain a finite value as the transition is approached along noncritical temperatures. Asymptotic expansion of the integrals $I(S)$ and $G''(S)$ near the transition shows that properties depend linearly on the distance from the transition for the power α less than $\frac{3}{2}$ times the number of dimensions, while for larger α properties near the transition depend upon the distance from the transition raised to a power which is a function of α. The critical properties are independent of cell size and potential strength, and are determined by the power α alone. The spherical lattice gas with a nonintegral power law potential can be tailored through choice of the power, to give a desired value for the behavior of isothermal compressibility near the critical point even though a specific heat singularity near the critical point cannot be obtained.

* Part of a dissertation submitted to Texas Christian University by Kenneth H. Carpenter in partial fulfillment of the requirements for the degree Doctor of Philosophy (May 1966) while holding an N.S.F. Cooperative Graduate Fellowship.

† Present address: Department of Physics, East Tennessee State University, Johnson City, Tennessee 37601.

[1] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[2] H. A. Gersch and T. H. Berlin, Phys. Rev. **127**, 2276 (1962).
[3] H. A. Gersch, Phys. Fluids **6**, 599 (1963).
[4] J. L. Strecker, J. Math. Phys. **10**, 1541 (1969).
[5] Y. Imry, Ann. Phys. (N.Y.) **51**, 1 (1969).
[6] G. S. Joyce, Phys. Rev. **146**, 349 (1966).
[7] M. I. Bagatskii, A. V. Voronel, and V. G. Gusak, Zh. Eksp. Teor. Fiz. **43**, 728 (1962) [Sov. Phys. JETP **16**, 517 (1963)].
[8] A. V. Voronel, Yu. R. Chashkin, V. A. Popov, and V. G. Simkin, Zh. Eksp. Teor. Fiz. **45**, 828 (1963) [Sov. Phys. JETP **18**, 568 (1964)].
[9] C. N. Yang and C. P. Yang, Phys. Rev. Letters **13**, 303 (1964).
[10] Terrell L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956), p. 164ff.
[11] H. W. Lewis and G. H. Wannier, Phys. Rev. **88**, 682 (1952); **90**, 1131 (1953). Further work on this approach is in C. C. Yan and G. H. Wannier, J. Math. Phys. **6**, 1833 (1965).
[12] T. H. Berlin, L. Witten, and H. A. Gersch, Phys. Rev. **92**, 189 (1953), Eq. (4.5).
[13] Louis Brand, *Advanced Calculus* (Wiley, New York, 1955), p. 434.
[14] K. H. Carpenter, Ph.D. thesis, Texas Christian University, 1966.
[15] See Ref. 3, Eqs. (2.15), (2.19), and (2.23).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Approximate Solutions of the Unitarity Equation

C. EFTIMIU
Department of Physics, University of Missouri at St. Louis, St. Louis, Missouri 63121

(Received 29 December 1969)

The possible use of the classical Newton approximation method, as extended by Kantorovich to Banach spaces, is discussed in connection with the problem of solving the unitarity equation for the phase of a scattering amplitude, in terms of the corresponding differential cross section.

The purpose of this paper is to discuss the usefulness of a classical approximation method as applied to the problem of finding a solution of the nonlinear integral equation

$$P[\phi] \equiv \sin \phi(x) - \lambda \iint_D h(x, y, z) \cos [\phi(y) - \phi(z)]\, dy\, dz = 0. \tag{1}$$

This equation has been shown by Newton[1] to express the unitarity condition in its simplest form, and the problem of solving it arose from the question whether the phase $\phi(x)$ of a scattering amplitude

$$F(x) = G(x) \exp [i\phi(x)]$$

can be determined in terms of its modulus $G(x)$. In this context, $G(x)$ can be directly given in terms of the differential cross section which is supposed to be known at one given energy for all scattering angles θ ($x = \cos \theta$). The unitarity condition leads then to

Eq. (1), with

$$\lambda h(x, y, z) = G(y)G(z)/2\pi G(x)(1 - x^2 - y^2 - z^2 + 2xyz)^{\frac{1}{2}} \tag{2}$$

and the domain of integration is restricted to the interior of the ellipse in the (y, z) plane where the radicand in (2) is positive.

The constant λ is introduced here for later convenience, and is supposed to be chosen such that

$$\max_{-1\le x\le 1} q(x) = 1, \quad q(x) = \iint_D h(x, y, z)\, dy\, dz. \tag{3}$$

Assuming that the function $G(x)$ does not vanish in $[-1, 1]$, Newton used fixed-point theorems in the Banach space $C[-1, 1]$ of all continuous functions with the norm

$$\|\phi\| = \sup_{-1\le x\le 1} |\phi(x)| \tag{4}$$

to prove that Eq. (1) has solutions if $\lambda < 1$ and, furthermore, that, if $\lambda < 5^{-\frac{1}{2}}$, such solutions are also unique. This result was improved by Martin,[2] who showed that the uniqueness range can be pushed up to $\lambda = 0.79$ and conjectured that it could be extended even further, possibly as far as $\lambda = 1$.

In this paper we present a study of Eq. (1) by means of the Newton–Kantorovich approximation method. The motivation of this approach is twofold: Firstly, this is an alternative way of establishing the existence and uniqueness of solutions for this equation; secondly, and possibly more importantly, this method enables us to indicate an effective specific process of constructing the solution, which may be useful when the kernel $h(x, y, z)$ is explicitly known and the actual finding of a solution is undertaken.

The Newton–Kantorovich method is a direct extension to functional spaces of the classical Newton method of finding a solution of a real equation $f(x) = 0$ by successive approximations, which start with an appropriately chosen x_0 and are then produced by the formula $x_{n+1} = x_n - [f'(x_n)]^{-1}f(x_n)$. This familiar method has first been formulated in Banach spaces by Kantorovich,[3] but there exist also more or less extensive studies in spaces with partially ordered norm and in Hilbert spaces.[4] In the following we shall restrict ourselves to Banach spaces; more precisely, the domain of the operator P defined by (1) is taken to be the set K of all functions $\phi(x) \in C[-1, 1]$ such that

$$0 \le \phi(x) \le \tfrac{1}{2}\pi. \tag{5}$$

(This limitation is introduced in order to avoid the obvious ambiguity that, if ϕ is a solution of $P[\phi] = 0$, then $\pi - \phi$ is a solution as well.) The nonlinear operator P, therefore, maps $K \subset C[-1, 1]$ into $C[-1, 1]$, and we can thus make use of the following theorem.[5]

Theorem 1: Let $\phi_0(x) \in K$ be selected such that:

(i) the Frechet derivative $P'[\phi_0]$ exists and has a bounded inverse with

$$\|(P'[\phi_0])^{-1}\| \le \beta_0, \quad \|(P'[\phi_0])^{-1}(P[\phi_0])\| \le \eta_0;$$

(ii) for all $\phi(x)$ in the sphere $S(\phi_0, r_0) \subset K$ (i.e., for all $\|\phi - \phi_0\| \le r_0$ with r_0 defined below), the Frechet derivative $P''[\phi]$ exists and $\|P''[\phi]\| \le \gamma_0$;

(iii) $h = \beta_0\eta_0\gamma_0 \le \frac{1}{2}$ and r_0 is any number such that $S(\phi_0, r_0) \subset K$ and which is not less than $[1 - (1 - 2h_0)^{\frac{1}{2}}]\eta_0/h_0$. Then the sequence $\{\phi_n(x)\}$ defined by

$$\phi_{n+1} = \phi_n - (P'[\phi])^{-1}[P[\phi_n]], \quad n \ge 0,$$

exists for all n and converges to a solution $\phi \in S(\phi_0, r_0)$ of $P[\phi] = 0$ at the rate

$$\|\phi - \phi_n\| \le 2^{-n}(2h_0)^{2^n}\eta_0/h_0, \quad n = 0, 1, \cdots$$

Moreover, ϕ is unique in every closed sphere

$$\bar{S}(\phi_0, r) \subset S(\phi_0, r_0)$$

with $r < [1 + (1 - 2h_0)^{\frac{1}{2}}]\eta_0/h_0$.

For the nonlinear operator $P[\phi]$, the Frechet derivatives can be easily computed. $P'[\phi]$ is the linear operator

$$P'[\phi]\xi = \xi(x)\cos\phi(x) + \lambda\iint_D h(x, y, z)[\xi(y) - \xi(z)] \times \sin[\phi(y) - \phi(z)]\, dy\, dz, \tag{6}$$

where $\xi \in C[-1, 1]$, and $P''[\phi]$ is the bilinear operator

$$P''[\phi]\xi\eta = -\xi(x)\eta(x)\sin\phi(x) + \lambda\iint_D h(x, y, z\,[\xi(y) - \xi(z)] \times [\eta(y) - \eta(z)]\cos[\phi(y) - \phi(z)]\, dy\, dz, \tag{7}$$

with $(\xi, \eta) \in C[-1, 1] \times C[-1, 1]$. Both these operators have range in $C[-1, 1]$.

In applying this theorem, it is important to note that $S(\phi_0, r) \subset \bar{S}(\phi_0, 2\eta_0) \subset S(\phi_0, r_0)$. If $S(\phi_0, 2\eta_0) \subset K$, we may then take $r = 2\eta_0$ and estimate the least upper bound on $\|P''[\phi]\|$ for $\phi \in S(\phi_0, 2\eta_0)$. The sequence $\{\phi_n\}$ will thus converge to a solution of $P[\phi] = 0$ which is unique in $\bar{S}(\phi_0, 2\eta_0)$.

The main difficulty is readily apparent: The inverse operator $(P')^{-1}$ must be found at each step of the iteration process. This is, in general, difficult, except in one case, namely, as can be realized by simple inspection of the expression (6), if $\phi_0(x) = \text{const}$, the integral part vanishes and the operator can be trivially inverted.

In order to explore this possibility, let us then choose

$$\phi_0(x) = \sin^{-1} a, \tag{8}$$

where a is a constant. Then it follows immediately that

$$\beta_0 = (1 - a^2)^{-\frac{1}{2}} \tag{9}$$

and, since

$$P[\sin^{-1} a] = a - \lambda q(x), \tag{10}$$

that

$$\max_x |a - \lambda q(x)|\,(1 - a^2)^{-\frac{1}{2}} \leq \max\{|a - \lambda|, |a - \lambda\epsilon|\}(1 - a^2)^{-\frac{1}{2}} = \eta_0, \tag{11}$$

where $\epsilon = \min_x q(x)$. It is of course desirable that η_0 be as small a number as possible and, if $\lambda < 1$, this can be achieved by taking

$$a = \tfrac{1}{2}\lambda(1 + \epsilon). \tag{12}$$

Hence,

$$\eta_0 = \lambda(1 - \epsilon)[4 - \lambda^2(1 + \epsilon)^2]^{-\frac{1}{2}}. \tag{13}$$

From (7) it follows now that

$$\|P''[\phi]\| \leq 4\lambda + \sin\phi, \quad \phi \in S(\sin^{-1} a, 2\eta_0), \leq 4\lambda + \sin(\sin^{-1} a + 2\eta_0), \tag{14}$$

provided

$$[\sin^{-1} a - 2\eta_0, \sin^{-1} a + 2\eta_0] \subset [0, \tfrac{1}{2}\pi].$$

This clearly does not happen for all possible values of λ and ϵ. In particular if $\epsilon = 0$, $\sin^{-1} a < 2\eta_0$ for all $\lambda < 1$.

Admittedly, the possibility that $\epsilon = 0$ can hardly be thought of as corresponding to a realistic physical situation. Nevertheless, for a given λ, the condition $\sin^{-1} a \geq 2\eta_0$ imposes a lower bound on ϵ which may or may not be realized. If the latter circumstance occurs, the upper bound on $\|P''\|$ must be looked for in $S(\sin^{-1} a, 2\eta_0) \cap K$. The theorem requires now that

$$2\lambda(1 - \epsilon)[4\lambda + \sin(\sin^{-1} a + 2\eta_0)] \times [4 - \lambda^2(1 + \epsilon)^2]^{-1} \leq \tfrac{1}{2}, \tag{15}$$

and this clearly imposes a limitation upon the values of λ for which the existence of solutions is guaranteed. An estimate of the largest value of λ allowed by (15) can be obtained under the assumption that it has to be significantly less than unity to make powers higher than the fourth negligible. Then, up to terms proportional to λ^5,

$$\sin(\sin^{-1} a + 2\eta_0) = \tfrac{1}{2}\lambda(3 - \epsilon) - \tfrac{1}{12}\lambda^3(1 - \epsilon)^2(5 + \epsilon)$$

and (15) can be rewritten as

$$\lambda^4(1 - \epsilon)^3(5 + \epsilon) - 3\lambda^2(23 - 22\epsilon + 3\epsilon^2) + 12 \geq 0$$

and in the extreme situation where $\epsilon = 0$, this requires that $\lambda \leq 0.42$. It is, however, apparent that as ϵ increases, this limitation is weakened, and as ϵ approaches unity the inequality is satisfied for any $\lambda < 1$. As a matter of fact, this is hardly surprising since $\epsilon = 1$ corresponds to $q(x) = 1$ and $\eta_0 = 0$, in which case $\phi_0 = \sin^{-1}\lambda$ is actually an exact solution of (1). One can therefore draw the conclusion that the initiation of the iteration process by means of a constant makes sense if and, in fact, only if η_0 is a small enough number, which is the case if ϵ is close enough to unity or λ is small enough.

Newton's formula can now be used to compute the next iterate

$$\phi_1(x) = \sin^{-1} a - [a - \lambda q(x)](1 - a^2)^{-\frac{1}{2}}. \tag{16}$$

At this point one should first check that $\phi_1(x) \in K$, and again one can see that this is not automatically fulfilled and that it requires an adequate lower bound on $q(x)$. If this is not available, or if one insists to devise an iteration process converging independently of the existence of such a lower bound, the only possible choice one has is $a = 0$. In this case $\beta_0 = 1$, $\eta_0 = \lambda$, $\gamma_0 = 4\lambda + \sin 2\lambda$, and the condition $h_0 \leq \frac{1}{2}$ requires that $\lambda < 0.28$.

The next step of the iteration process, leading to $\phi_2(x)$, involves the explicit form of the operator $(P'[\phi_1])^{-1}$. This is no longer a trivial matter since $\phi_1(x)$ is no longer a constant. If $(P'[\phi_1])^{-1}$ cannot be found explicitly, and this might very well be the case, then one can sidestep the difficulty by using the modified formula

$$\phi_{n+1} = \phi_n - (P'[\phi_0])^{-1}(P[\phi_n]), \tag{17}$$

which gives rise to a sequence converging under the same conditions as the original one, only at the slower rate

$$\|\phi - \phi_n\| \leq [1 - (1 - 2h_0)^{\frac{1}{2}}]^{n-1} 2\eta_0. \tag{18}$$

When neither one of these possibilities is convenient, one should recognize that a more fortunate choice for $\phi_0(x)$ is needed. That such a choice is available can be shown by making use of a theorem which can be derived essentially from Theorem 1.[6]

Theorem 2: Let Eq. (1) have the form

$$P[\phi] = \Pi[\phi] + R[\phi]$$

and let $\phi_0(x) \in K$ be a solution of the equation $\Pi[\phi] = 0$. If

(i) $\Pi'[\phi_0]$ exists and has a bounded inverse with

$$\|(\Pi'[\phi_0])^{-1}\| \leq \beta_0 \quad \text{and} \quad \|(\Pi'[\phi_0])^{-1}(R[\phi_0])\| \leq \eta_0,$$

(ii) $\|(\Pi'[\phi_0])^{-1}R'[\phi_0]\| \leq \alpha < 1$,

(iii) for all $\phi(x) \in S(\phi_0, 2\eta_0/(1 - \alpha)) \subset K$,

$$\|(\Pi'[\phi_0])^{-1}\Pi''[\phi]\| \leq \gamma, \quad \|(\Pi'[\phi_0])^{-1}R''[\phi]\| \leq \delta,$$

and

(iv) $h_0 = \eta_0(\gamma + \delta)/(1 - \alpha)^2 \leq \frac{1}{2}$,

then the sequence defined by Newton's formula converges to a unique solution $\phi \in \bar{S}(\phi_0, 2\eta_0/(1 - \alpha))$ at the rate

$$\|\phi - \phi_n\| \leq 2^{-n}(2h_0)^{2^n}\eta_0/h_0(1 - \alpha).$$

One possible way of splitting Eq. (1) is by taking

$$\Pi[\phi] = \sin\phi - \sin\phi_0 \tag{19}$$

and

$$R[\phi] = \sin\phi_0 - \lambda\iint_D h(x, y, z) \times \cos[\phi(y) - \phi(z)]\,dy\,dz, \tag{20}$$

where ϕ_0 is still to be chosen. One possibly advantageous choice is

$$\phi_0(x) = \sin^{-1} A\lambda q(x), \tag{21}$$

with $0 \leq A \leq 1$. Hence

$$\beta_0 = (1 - A^2\lambda^2)^{-\frac{1}{2}} \tag{22}$$

and[7]

$$\eta_0 = \beta_0 \max\{\lambda(1 - A), \lambda[A - (1 - A^2\lambda^2)^{-\frac{1}{2}}]\}. \tag{23}$$

It appears convenient now to take

$$A = 4/(4 + \lambda^2), \tag{24}$$

in which case

$$\beta_0 = (4 + \lambda^2)(4 - \lambda^2)^{-1}, \quad \eta_0 = \lambda^3(4 - \lambda^2)^{-1}. \tag{25}$$

It follows then that, since

$$R'[\phi]\xi(x) = \lambda\iint_D h(x, y, z)[\xi(y) - \xi(z)] \times \sin[\phi(y) - \phi(z)]\,dy\,dz, \tag{26}$$

we can write

$$\|R'[\phi_0]\| \leq 8\lambda^2/(4 + \lambda^2), \tag{27}$$

so that

$$\alpha = 8\lambda^2/(4 - \lambda^2). \tag{28}$$

The condition $\alpha < 1$ implies that $\lambda < \frac{2}{3}$ but, as we shall see shortly, λ must be subjected to an even stronger limitation.

Since now

$$\Pi''[\phi]\xi\eta = -\xi\eta\sin\phi \tag{29}$$

and

$$R''[\phi]\xi\eta = \lambda\iint_D h(x, y, z)[\xi(y) - \xi(z)] \times [\eta(y) - \eta(z)]\cos[\phi(y) - \phi(z)]\,dy\,dz, \tag{30}$$

it follows that

$$\|\Pi''[\phi]\| \leq \sin[2\eta_0/(1 - \alpha)] \tag{31}$$

and

$$\|R''[\phi]\| \leq 4\lambda, \tag{32}$$

provided $2\eta_0/(1 - \alpha) < \frac{1}{2}\pi$. If this inequality is satisfied, then, to a good approximation,

$$\frac{\sin 2\eta_0}{1 - \alpha} = \frac{4\lambda}{4 + \lambda^2} + \frac{2\lambda^3}{4 - 9\lambda^2},$$

so that the condition

$$h_0 = 2\lambda^4\frac{4 + \lambda^2}{4 - 9\lambda^2}\left(2 + \frac{2}{4 + \lambda^2} + \frac{\lambda^2}{4 - 9\lambda^2}\right)$$

implies that $\lambda < 0.51$. If λ does satisfy this condition, $2\eta_0/(1 - \alpha) < 0.15$. This not only justifies *a posteriori* the estimate (31), but shows also that the solution $\phi(x) \leq 0.62 < \pi/2$, as it should.

One cannot emphasize too strongly that the successful use of the Newton–Kantorovich method rests heavily upon the implicit possibility of making a good choice for the starting point. It is therefore worthwhile to discuss a means of finding a good initial approximation, which is essentially suggested by the familiar procedure consisting in replacing the actual kernel of an integral equation by a separable kernel.

Let $\omega_i(x)$ be a set of functions orthogonal on $[-1, 1]$ (e.g., Legendre polynomials) and suppose that the function

$$g(x, y, z) = h(x, y, z)\theta[1 - x^2 - y^2 - z^2 + 2xyz] \tag{33}$$

can be expanded as

$$g(x, y, z) = \sum_{i=0}^{\infty} c_i(y, z)\omega_i(x). \tag{34}$$

Let us then define the function

$$g_N(x, y, z) = \sum_{i=0}^{N} c_i(y, z)\omega_i(x) \tag{35}$$

and recognize that

$$\sin\phi_0(x) = \psi_0(x) = \sum_{i=0}^{N} A_i\omega_i(x) \tag{36}$$

is a solution of the equation

$$\psi_0(x) - \lambda\iint g_N(x, y, z)\cos[\phi_0(y) - \phi_0(z)]\,dy\,dz = 0, \tag{37}$$

provided the A_i, $i = 0, 1, \cdots, N$, form a solution of the nonlinear system

$$A_k - \lambda\iint c_k(y, z)\cos(\sin^{-1}\textstyle\sum A_i\omega_i(y) - \sin^{-1}\sum A_i\omega_i(z))\,dy\,dz = 0, \quad k = 0, 1, \cdots, N. \tag{38}$$

If one takes then ψ_0 as the initial approximation, the convergence of the Newton sequence is guaranteed by the following theorem.[8]

Theorem 3: Let Eq. (1) be written as

$$P[\psi] = \Pi[\psi] + R[\psi] = 0,$$

where $\psi = \sin\phi$ and

$$\Pi[\psi] = \psi(x) - \lambda\iint H_N(x, y, z; \psi(y), \psi(z))\, dy\, dz,$$

$$R[\psi] = \lambda\iint [H_N(x, y, z; \psi(y), \psi(z)) - H(x, y, z; \psi(y), \psi(z))]\, dy\, dz,$$

and let $K(x, y, z; \psi_0(y), \psi_0(z))$ be the resolvent of the kernel of the linear operator $\Pi'[\psi_0]$. If

(i) $|K(x, y, z; \psi_0(y), \psi_0(z))|\, dy\, dz \leq B$,
(ii) $\|R[\psi_0]\| \leq \eta$,
(iii) $\|R'[\psi_0]\| \leq \alpha$,
(iv) $\|\Pi''[\psi]\| \leq K$, $\|R''[\psi]\| \leq L$ for $\psi \in S(\psi_0, 2\eta(1 + B)/[1 - \alpha(1 + B)])$,
(v) $\alpha(1 + B) < 1$, $h = (1 + B^2)(K + L)\eta/[1 - \alpha(1 + B)]^2 \leq \frac{1}{2}$,

then Newton's sequence starting with ψ_0 converges to a unique solution of Eq. (1) in the sphere S.

From a practical viewpoint the advantage offered by this theorem lies in the fact that condition (v) can in principle be always satisfied, provided N is taken large enough. From a theoretical viewpoint, though, it is apparent that the question of the existence of one or more solutions has been shifted from the original equation to the system (38) for the set of constants A_i. While in general the problem of establishing the existence of a solution of this system can be as hard as that for the Eq. (1), in some cases it may be easier to handle.

To illustrate this possibility, let us consider the simple case when the expansion (36) consists of just one term, namely,

$$\psi_0 = A\lambda q(x), \tag{39}$$

with $0 \leq A \leq 1$. Clearly, A must be a solution of the equation

$$f(A) - A = 0, \tag{40}$$

where

$$f(A) = \iint c(y, z) \cos[\sin^{-1} A\lambda q(y) - \sin^{-1} A\lambda q(z)]\, dy\, dz \tag{41}$$

and

$$c(y, z) = \frac{1}{\rho}\int_{-1}^{1} g(x, y, z)\, dx, \quad \rho = \int_{-1}^{1} q(x)\, dx. \tag{42}$$

Since

$$[1 - A^2\lambda^2]^{\frac{1}{2}} \leq f(A) \leq 1, \tag{43}$$

it readily follows that Eq. (40) necessarily has at least one solution in the range $((1 + \lambda^2)^{-\frac{1}{2}}, 1)$. A sufficient condition that there be just one solution can be simply derived by noticing that there could be more than one solution only if $df/dA = 1$ for values of A in the indicated range. But

$$\left|\frac{df}{dA}\right| \leq \frac{A\lambda^2}{(1 - A^2\lambda^2)^{\frac{1}{2}}} \leq \frac{\lambda^2}{(1 - \lambda^2)^{\frac{1}{2}}}, \tag{44}$$

so that $df/dA < 1$ if $\lambda^4 + \lambda^2 - 1 < 0$, i.e., if $\lambda < 0.78$. Further use of Theorem 3 could be made if Eq. (40) were actually solved for A, and this can certainly be done when the kernel is explicitly given. One could still note, however, that the condition just derived for λ guarantees the existence of an upper bound on the resolvent K. Indeed, the linear equation

$$\omega(x) + \lambda q(x)\iint c(y, z) \sin[\sin^{-1} A\lambda q(y) - \sin^{-1} A\lambda q(z)]\left(\frac{\omega(y)}{[1 - A^2\lambda^2 q^2(y)]^{\frac{1}{2}}} - \frac{\omega(z)}{[1 - A^2\lambda^2 q^2(z)]^{\frac{1}{2}}}\right) dy\, dz = \chi(x)$$

can easily be solved with respect to ω and the resolvent explicitly found. It follows then that

$$B \leq \frac{\lambda}{1 + \lambda C}\frac{2A\lambda}{(1 - A^2\lambda^2)^{\frac{1}{2}}},$$

where

$$C = \iint c(y, z) \sin[\sin^{-1} A\lambda q(y) - \sin^{-1} A\lambda q(z)] \times \left(\frac{q(y)}{[1 - A^2\lambda^2 q^2(y)]^{\frac{1}{2}}} - \frac{q(z)}{[1 - A^2\lambda^2 q^2(z)]^{\frac{1}{2}}}\right) dy\, dz,$$

so that

$$C \geq -A\lambda/(1 - A^2\lambda^2)^{\frac{1}{2}}$$

and

$$B < 2\lambda^2/[(1 - \lambda^2)^{\frac{1}{2}} - \lambda^2],$$

which exists as long as $\lambda < 0.78$.

[1] R. G. Newton, J. Math. Phys. **9**, 2050 (1968).
[2] A. Martin, Nuovo Cimento **59A**, 131 (1969).
[3] See L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Macmillan, New York, 1964), Chap. XVIII.
[4] For a review of these formulations and extensive references, see R. H. Moore, in *Nonlinear Integral Equations (Proceedings of an Advanced Seminar at the University of Wisconsin, Madison)*, P. M. Anselone, Ed. (The University of Wisconsin Press, Madison, 1964).
[5] See Ref. 3, p. 708, and also H. A. Antosiewicz and W. C. Rheinboldt, in *Survey of Numerical Analysis*, J. Todd, Ed. (McGraw-Hill, New York, 1962).
[6] See Ref. 3, p. 719.
[7] From this point on we let $\epsilon = 0$, in order to simplify the algebra. The case $\epsilon \neq 0$ can be followed through without difficulty.
[8] See Ref. 3, p. 733.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Maxwell's Equations Having a Gradient as Source

M. MURASKIN
Department of Physics, University of North Dakota, Grand Forks, North Dakota 58201

(Received 17 June 1968; Revised Manuscript Received 15 May 1969)

We study Maxwell's equations having a source given by a gradient. The source function f obeys the empty-space wave equation as a consequence of current conservation. We find that this system can be written as a mass-zero Dirac equation. Our relationship between tensor and spinor equations is an improvement over other published results, so far as simplicity is concerned.

I. INTRODUCTION

This paper has two purposes. The first is to learn a few things about Maxwell's equations having a source of a particularly simple type. We shall take the source to be a gradient of a scalar f. The second purpose is to find a relationship between simple tensor equations and spinor systems that is an improvement over the other published results, so far as simplicity is concerned.[1]

We find that the Maxwell system has an alternative description as a mass-zero Dirac equation.

II. MAXWELL EQUATIONS WITH SOURCE OBEYING THE WAVE EQUATION

We study the equations

$$\nabla \cdot \mathbf{H} = 0, \qquad \nabla \cdot \mathbf{E} = -\frac{\partial \tfrac{1}{2} f}{\partial t},$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{H}}{\partial t}, \quad \nabla \times \mathbf{H} = \frac{\partial \mathbf{E}}{\partial t} + \nabla \tfrac{1}{2} f. \tag{1}$$

The continuity equation implies that

$$\Box f = 0. \tag{2}$$

Thus, the source of the Maxwell equations (the function f) obeys the empty-space wave equation.

Defining $f_{[ij]}$ by

$$\begin{aligned} E_1 &= f_{[10]}, \quad H_1 = f_{[23]}, \\ E_2 &= f_{[20]}, \quad H_2 = f_{[31]}, \\ E_3 &= f_{[30]}, \quad H_3 = f_{[12]}, \end{aligned} \tag{3}$$

we then get

$$\Box f_{[ij]} = 0. \tag{4}$$

Thus, the source in (1) does not prevent **E** and **H** from satisfying a free second-derivative wave equation.

III. INTRODUCTION OF SPINOR FIELDS

We consider the following set of equations

$$\begin{aligned} \sigma^{kA\dot{B}} \bar{\chi}_D \sigma^l_{A\dot{F}} \sigma_\nu^{D\dot{F}} \partial_k f^\nu{}_l &= 0, \\ \sigma^{kA\dot{B}} \bar{\varphi}_D \sigma^l_{A\dot{F}} \sigma_\nu^{D\dot{F}} \partial_k f^\nu{}_l &= 0, \end{aligned} \tag{5}$$

where $\sigma^i_{A\dot{B}}$, $\sigma^{kA\dot{B}}$ are the spin matrices found in Bade and Jehle.[2] The spin indices are raised and lowered with $\epsilon_{AB} = \epsilon^{AB} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. The space–time metric is $g_{ij} = (1, -1, -1, -1)$. ∂_k is given by $(\partial/\partial t, \nabla)$. $\bar{\chi}_A$ and $\bar{\varphi}^A$ are constant 2-component spinors for which we take $\bar{\chi}_A \bar{\varphi}^A \neq 0$. Let us write $\bar{\chi} = \left(\begin{smallmatrix} A \\ B \end{smallmatrix}\right)$ and $\bar{\varphi} = \left(\begin{smallmatrix} C \\ D \end{smallmatrix}\right)$. Then, $\bar{\chi}_A \bar{\varphi}^A \neq 0$ means $AD - BC \neq 0$. We write out the sum over D in (5). This gives

$$\begin{aligned} A\sigma^{kA\dot{B}} \sigma^l_{A\dot{F}} \sigma_\nu^{1\dot{F}} \partial_k f^\nu{}_l + B\sigma^{kA\dot{B}} \sigma^l_{A\dot{F}} \sigma_\nu^{2\dot{F}} \partial_k f^\nu{}_l &= 0, \\ C\sigma^{kA\dot{B}} \sigma^l_{A\dot{F}} \sigma_\nu^{1\dot{F}} \partial_k f^\nu{}_l + D\sigma^{kA\dot{B}} \sigma^l_{A\dot{F}} \sigma_\nu^{2\dot{F}} \partial_k f^\nu{}_l &= 0. \end{aligned}$$

We multiply the top equation by D and the bottom by B and subtract. This gives

$$\sigma^{kA\dot{B}} \sigma^l_{A\dot{F}} \sigma_\nu^{1\dot{F}} \partial_k f^\nu{}_l = 0. \tag{6a}$$

The equation

$$\sigma^{kA\dot{B}} \sigma^l_{A\dot{F}} \sigma_\nu^{2\dot{F}} \partial_k f^\nu{}_l = 0 \tag{6b}$$

follows when, instead, we multiply the top by C and the bottom by A and subtract. We see that the constant spinors no longer appear in (6). We see also that (6) is a linear equation for $f^\nu{}_l$. We next expand (6) term by term using the explicit form of the spin matrices. These are four complex equations. The eight equations we obtain can be seen to be the Maxwell equations (1) with

$$f = f_1^1 + f_2^2 + f_3^3 + f_0^0 = \mathrm{Tr}\, f^\nu{}_l. \tag{7}$$

We define the spinors χ_A and φ_A by

$$\begin{aligned} \chi_A &= \pm f^\nu{}_l \sigma_\nu^{B\dot{D}} \sigma^l_{A\dot{D}} \bar{\chi}_B, \\ \varphi_A &= \pm f^\nu{}_l \sigma_\nu^{B\dot{D}} \sigma^l_{A\dot{D}} \bar{\varphi}_B. \end{aligned} \tag{8}$$

Then, from (5) we see that χ_A and φ_A obey the Dirac free-field mass-zero equations

$$\begin{aligned} \sigma^{kA\dot{B}} \partial_k \chi_A &= 0, \\ \sigma^{kA\dot{B}} \partial_k \varphi_A &= 0. \end{aligned} \tag{9}$$

We have written (5) in terms of $f^\nu{}_l$. However, in expanding (6), only $f_{[ij]}$ and f appear. This can be seen also by using the properties of the spin matrices

$$\sigma^{iA\dot{B}} \sigma^j_{C\dot{B}} + \sigma^{jA\dot{B}} \sigma^i_{C\dot{B}} = g^{ij} \delta^A_C \tag{10}$$

in conjunction with (8). χ_A and φ_A then can be written[3]

$$\chi_A = \pm\left(f_{[ij]}\sigma^{jB\dot{C}}\sigma^i_{A\dot{C}}\bar{\chi}_B + \frac{f}{2}\bar{\chi}_A\right),$$
$$\varphi_A = \pm\left(f_{[ij]}\sigma^{jB\dot{C}}\sigma^i_{A\dot{C}}\bar{\varphi}_B + \frac{f}{2}\bar{\varphi}_A\right). \quad (11)$$

This shows explicitly that only $f_{[ij]}$ and f appear in (9).

Thus, the conclusion of this section is that the Maxwell equations (1) can be written as the mass-zero Dirac equation (9), with χ_A, φ_A given in terms of $f_{[ij]}$ and f by (11).

IV. COMPARISON WITH OTHER INVESTIGATORS

The relationships between simple tensor equations with spinor equations (or simple spinor equations with tensor equations) by other authors are rather complicated. This point is emphasized by simply quoting the literature. Klauder[1] starts with mass-zero Dirac equations and gets

$$C^m(\partial_k F_{im} + \partial_i F_{mk} + \partial_m F_{ki}) + C_i\partial_m F^m{}_k - C_k\partial_m F^m{}_i = 0, \quad (12)$$

with C^m a constant vector. Penny[1] also starts with mass-zero Dirac equations. He makes use of a different relationship between spinors and tensors and then gets

$$F_{ik}\partial^m(F_{mp} + i\,{}^*F_{mp}) + F^m{}_k\partial_p(F_{mi} - i\,{}^*F_{mi}) = 0, \quad (13)$$

where ${}^*F_{mp}$ is the dual field. From Laporte and Uhlenbeck[1] we can obtain the following representation of (1) (δ_A, θ_A are the spinor fields):

$$\sigma^{kA}{}_{\dot{B}}\partial_k\theta_A = -\sigma^k_{C\dot{B}}(\partial_k f)\theta^C - \sigma^{kA}{}_{\dot{B}}\delta_A\theta^C\partial_k\theta_C - \sigma^{kA}{}_{\dot{B}}\theta_A\theta^C\partial_k\delta_C,$$
$$\sigma^{kA}{}_{\dot{B}}\partial_k\delta_A = \sigma^k_{C\dot{B}}(\partial_k f)\delta^C + \sigma^{kA}{}_{\dot{B}}\delta_A\delta^C\partial_k\theta_C + \sigma^{kA}{}_{\dot{B}}\theta_A\delta^C\partial_k\delta_C. \quad (14)$$

Thus, our equations are an improvement over the other equations so far as simplicity is concerned. The Ruse equations are extremely complicated; so we refer the reader to the original article.[1]

V. A SOLUTION OF THE MAXWELL EQUATIONS

In this section, we obtain a solution of the Maxwell's equations by using the plane-wave solutions of the $m = 0$ Dirac equations. As the plane-wave solution (17) is not the most general solution of the Dirac equation, we will not expect to get the most general Maxwell solution either.

$\bar{\chi}_A$ and $\bar{\varphi}_A$ are arbitrary so far. We shall take the simple choice

$$\bar{\chi}_A = \begin{pmatrix}1\\0\end{pmatrix}, \quad \bar{\varphi}_A = \begin{pmatrix}0\\1\end{pmatrix}. \quad (15)$$

Then, (11) becomes

$$\begin{aligned}\chi_1 &= \pm(\tfrac{1}{2}f + f_{[30]} + if_{[12]}),\\ \chi_2 &= \pm(f_{[13]} + if_{[20]} + f_{[10]} + if_{[23]}),\\ \varphi_1 &= \pm(f_{[31]} + if_{[02]} + f_{[10]} + if_{[23]}),\\ \varphi_2 &= \pm(\tfrac{1}{2}f + f_{[03]} + if_{[21]}).\end{aligned} \quad (16)$$

There are eight variables χ_A, φ_A. They are expressed in terms of seven variables $f_{[ij]}$ and f by means of eight equations. Thus, there is a relationship to be satisfied among χ_A and φ_A. From (16) we see that $\chi_1 + \varphi_2 =$ Real. We shall also study the situation where χ_A and φ_A satisfy the condition $\chi_A\varphi^A = 0$.

The following plane-wave solutions of (9) obey the conditions $\chi_1 + \varphi_2 =$ Real (we take $p_3 \neq \pm E$):

$$\chi_A = N\begin{pmatrix}1\\ \dfrac{p_1 + ip_2}{E + p_3}\end{pmatrix}e^{i(\mathbf{p}\cdot\mathbf{x} - Et)},$$
$$\varphi_A = N\begin{pmatrix}\dfrac{p_1 - ip_2}{E - p_3}\\ 1\end{pmatrix}e^{-i(\mathbf{p}\cdot\mathbf{x} - Et)}. \quad (17)$$

N is the normalization which is to be real. We also get a solution to (9) by setting $\mathbf{p} \to -\mathbf{p}$ and $E \to -E$ in (17).

By comparing (17) with (16),we get

$$\begin{aligned}f &= \pm 2N\cos(\mathbf{p}\cdot\mathbf{x} - Et),\\ E_1 &= \pm\frac{N}{E^2 - p_3^2}[p_1E\cos(\mathbf{p}\cdot\mathbf{x} - Et) - Ep_2\sin(\mathbf{p}\cdot\mathbf{x} - Et)],\\ E_2 &= \pm\frac{N}{E^2 - p_3^2}[p_2E\cos(\mathbf{p}\cdot\mathbf{x} - Et) + Ep_1\sin(\mathbf{p}\cdot\mathbf{x} - Et)],\\ E_3 &= 0,\\ H_1 &= \pm\frac{N}{E^2 - p_3^2}[-p_3p_2\cos(\mathbf{p}\cdot\mathbf{x} - Et) - p_1p_3\sin(\mathbf{p}\cdot\mathbf{x} - Et)],\\ H_2 &= \pm\frac{N}{E^2 - p_3^2}[p_3p_1\cos(\mathbf{p}\cdot\mathbf{x} - Et) - p_2p_3\sin(\mathbf{p}\cdot\mathbf{x} - Et)],\\ H_3 &= \pm N\sin(\mathbf{p}\cdot\mathbf{x} - Et).\end{aligned} \quad (18)$$

The 3 direction appears favored due to the choice of $\sigma^{3A\dot{B}}$ as the diagonal matrix. The plane-wave solutions

(17) satisfy $\chi_A \varphi^A = 0$. This implies using (16) that

$$\mathbf{E} \cdot \mathbf{H} = 0,$$
$$\mathbf{E}^2 - H^2 = \tfrac{1}{4} f^2 = N^2 \cos^2(\mathbf{p} \cdot \mathbf{x} - Et). \quad (19)$$

The condition $f = 0$ can not be satisfied unless the normalization is chosen to be zero, making $\chi_A = \varphi_A = 0$.

We can, in fact, check directly that (18) gives a solution to (1) and (19).

From (4), we obtain a plane-wave solution of the form

$$\mathbf{E} = \mathbf{E}^C(p) \cos(\mathbf{p} \cdot \mathbf{x} - Et) + \mathbf{E}^S(p) \sin(\mathbf{p} \cdot \mathbf{x} - Et),$$
$$\mathbf{H} = \mathbf{H}^C(p) \cos(\mathbf{p} \cdot \mathbf{x} - Et) + \mathbf{H}^S(p) \sin(\mathbf{p} \cdot \mathbf{x} - Et). \quad (20)$$

Since we have chosen to discuss those solutions of (9) obeying $\chi_A \varphi^A = 0$, the coefficients E^S, E^C, H^S, and H^C are subject to the condition (19). We have obtained these coefficients using the plane-wave solutions of the Dirac equation (17).

We have, thus, furnished an example of a situation when f comes out to be nonzero.

VI. CONCLUSION

We have shown that Maxwell's equations having a source of the gradient type can be expressed as a mass zero Dirac equation. We have compared our results with others and have found our equations to be simpler.

ACKNOWLEDGMENTS

I would like to thank Professor M. Dresden and Professor C. Hammer for helpful discussions.

[1] J. Klauder, J. Math. Phys. **5**, 1204 (1964); R. Penny, *ibid.* **6**, 1026 (1965); O. Laporte and G. Uhlenbeck, Phys. Rev. **37**, 1380 (1931); H. S. Ruse, Proc. Roy. Soc. Edinburgh **57**, 97 (1937).

[2] W. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953).

[3] Equation (11) is a generalization of Klauder (Ref. 1) who did not introduce the trace field.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Clebsch–Gordan Series and the Clebsch–Gordan Coefficients of $O(2, 1)$ and $SU(1, 1)$*

KUO-HSIANG WANG†
Lawrence Radiation Laboratory, University of California, Berkeley, California 94270

(Received 10 November 1969)

The Clebsch–Gordan series of the $O(2, 1)$ group and its covering group $SU(1, 1)$ are derived for all cases except that of the supplemental series. The continuable Clebsch–Gordan coefficients (or equivalently, the Wigner coefficients) are explicitly expressed in terms of the generalized hypergeometric function ${}_3F_2$. The spectra in the decomposition of the product of the two principal series are discussed. The applications to the UIR of $O(2, 2)$ are also studied.

I. INTRODUCTION

Recently, much attention has been paid to the group-theoretical analysis of the scattering amplitude at zero or negative momentum transfer.[1,2] For the latter case, the amplitude exhibits an $O(2, 1)$ symmetry. Reggeons, which take the role of forces in relativistic S-matrix theory,[3] transform under this symmetry group $O(2, 1)$ as the basis vector[4] of its unitary irreducible representation UIR. Thus, the Clebsch–Gordan coefficient (the CG coefficient) of $O(2, 1)$ plays the same role as that of $O(3)$ for physical particles.

The CG coefficient of $O(2, 1)$ for three positive discrete (or, equivalently, negative discrete) series was worked out by Andrews and Gunson[5] and Sannikov.[6] Pukanszky[7] found the multiplicity of the irreducible components resulting from decomposition of the product of two UIR's of $O(2, 1)$, but he did not work out the CG coefficients explicitly. Ferretti and Verde[8] worked out the Clebsch–Gordan series for two continuous series with some restrictions on the magnetic quantum numbers, by using the Sommerfeld–Watson transform. However, their definition of the Wigner coefficient is not normalized. They also did not investigate the relationships between the CG coefficients for various cases. Holman and Biedenharn[9] derived many CG coefficients from the difference equation of second order obtained from their recursion relations. Thus, their CG coefficient is not continuable, in the sense that it has different

(17) satisfy $\chi_A \varphi^A = 0$. This implies using (16) that

$$\mathbf{E} \cdot \mathbf{H} = 0,$$
$$\mathbf{E}^2 - H^2 = \tfrac{1}{4} f^2 = N^2 \cos^2(\mathbf{p} \cdot \mathbf{x} - Et). \quad (19)$$

The condition $f = 0$ can not be satisfied unless the normalization is chosen to be zero, making $\chi_A = \varphi_A = 0$.

We can, in fact, check directly that (18) gives a solution to (1) and (19).

From (4), we obtain a plane-wave solution of the form

$$\mathbf{E} = \mathbf{E}^C(p) \cos(\mathbf{p} \cdot \mathbf{x} - Et) + \mathbf{E}^S(p) \sin(\mathbf{p} \cdot \mathbf{x} - Et),$$
$$\mathbf{H} = \mathbf{H}^C(p) \cos(\mathbf{p} \cdot \mathbf{x} - Et) + \mathbf{H}^S(p) \sin(\mathbf{p} \cdot \mathbf{x} - Et). \quad (20)$$

Since we have chosen to discuss those solutions of (9) obeying $\chi_A \varphi^A = 0$, the coefficients E^S, E^C, H^S, and H^C are subject to the condition (19). We have obtained these coefficients using the plane-wave solutions of the Dirac equation (17).

We have, thus, furnished an example of a situation when f comes out to be nonzero.

VI. CONCLUSION

We have shown that Maxwell's equations having a source of the gradient type can be expressed as a mass zero Dirac equation. We have compared our results with others and have found our equations to be simpler.

ACKNOWLEDGMENTS

I would like to thank Professor M. Dresden and Professor C. Hammer for helpful discussions.

[1] J. Klauder, J. Math. Phys. **5**, 1204 (1964); R. Penny, *ibid.* **6**, 1026 (1965); O. Laporte and G. Uhlenbeck, Phys. Rev. **37**, 1380 (1931); H. S. Ruse, Proc. Roy. Soc. Edinburgh **57**, 97 (1937).

[2] W. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953).

[3] Equation (11) is a generalization of Klauder (Ref. 1) who did not introduce the trace field.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Clebsch–Gordan Series and the Clebsch–Gordan Coefficients of $O(2, 1)$ and $SU(1, 1)$*

KUO-HSIANG WANG†
Lawrence Radiation Laboratory, University of California, Berkeley, California 94270

(Received 10 November 1969)

The Clebsch–Gordan series of the $O(2, 1)$ group and its covering group $SU(1, 1)$ are derived for all cases except that of the supplemental series. The continuable Clebsch–Gordan coefficients (or equivalently, the Wigner coefficients) are explicitly expressed in terms of the generalized hypergeometric function ${}_3F_2$. The spectra in the decomposition of the product of the two principal series are discussed. The applications to the UIR of $O(2, 2)$ are also studied.

I. INTRODUCTION

Recently, much attention has been paid to the group-theoretical analysis of the scattering amplitude at zero or negative momentum transfer.[1,2] For the latter case, the amplitude exhibits an $O(2, 1)$ symmetry. Reggeons, which take the role of forces in relativistic S-matrix theory,[3] transform under this symmetry group $O(2, 1)$ as the basis vector[4] of its unitary irreducible representation UIR. Thus, the Clebsch–Gordan coefficient (the CG coefficient) of $O(2, 1)$ plays the same role as that of $O(3)$ for physical particles.

The CG coefficient of $O(2, 1)$ for three positive discrete (or, equivalently, negative discrete) series was worked out by Andrews and Gunson[5] and Sannikov.[6] Pukanszky[7] found the multiplicity of the irreducible components resulting from decomposition of the product of two UIR's of $O(2, 1)$, but he did not work out the CG coefficients explicitly. Ferretti and Verde[8] worked out the Clebsch–Gordan series for two continuous series with some restrictions on the magnetic quantum numbers, by using the Sommerfeld–Watson transform. However, their definition of the Wigner coefficient is not normalized. They also did not investigate the relationships between the CG coefficients for various cases. Holman and Biedenharn[9] derived many CG coefficients from the difference equation of second order obtained from their recursion relations. Thus, their CG coefficient is not continuable, in the sense that it has different

functional forms for various cases. Here, we derive two continuable CG coefficients which are orthonormal to each other for the case of three continuous series. For all other cases, one of them vanishes, and the other is identical to the CG coefficient obtained from the CG series.

In Sec. I, we introduce the definition of the CG coefficient and the conventions and notations used in this paper. We reproduce the derivation of the Clebsch–Gordan series for two continuous series with positive magnetic quantum numbers by the method initiated by Andrews and Gunson[5] and developed by Ferretti and Verde.[8] We point out the differences between our results and theirs, and explain how they occur. In Sec. III, we study the symmetry properties and asymptotic behavior of the G function, which is equivalent to the Wigner coefficient defined by Ferretti and Verde,[8] and which is simply related to the CG coefficient. In Sec. IV, we work out all other CG series and, thus, all other CG coefficients for positive magnetic quantum numbers. From them, we find two CG coefficients satisfying the properties stated at the end of the last paragraph above. In Sec. V, we calculate the CG series and, thus, the CG coefficients for all other cases. Finally, we show that the CG coefficients are also valid for the group $SU(1,1)$, the covering group of $O(2,1)$.

The scattering amplitude at vanishing momentum transfer has larger symmetry.[2] If one restricts oneself to the ordinary helicity amplitude, one has $O(4)$ symmetry[10] when the total energy is less than threshold energy, and $O(2,2)$ symmetry[11] above threshold. The explicit expression of the UIR of $O(2,2)$, suitable for this purpose, has not been worked out. In Sec. VI, we express explicitly the UIR of $O(2,2)$ group in terms of the CG coefficients of $O(2,1)$. The transformation between two UIR's of $O(2,2)$, corresponding to two different bases, is discussed. In the final section, we summarize the results obtained.

II. DECOMPOSITION OF THE PRODUCT OF TWO CONTINUOUS SERIES FOR $\nu_i > \mu_i > 0$

The CG coefficient[12] (or, equivalently, the Wigner coefficient[12]) of $O(3)$ is obtained essentially from its recursion relations[13] or from the integrals[12] involving three representation functions like $d_{\mu\nu}^{\ j}(z)$ of $O(3)$ in the integrand. The calculation of the CG coefficient of $O(2,1)$ is more complicated than that of $O(3)$, even though the representation function $d_{\nu\mu}^{\ j}(z)$ of $O(2,1)$ is a continuation in j of that of $O(3)$. The method applicable to the latter is not directly applicable to the former. The differences are (a) the group $O(2,1)$ is noncompact and has an infinite group manifold, and (b) the UIR of $O(2,1)$ has three principal series: continuous, positive discrete, and negative discrete. Each principal series has a different range for the magnetic quantum numbers. The first difference prevents one from calculating the CG coefficient directly from the integral

$$\int D^{j_1}_{\nu_1\mu_1}(g) D^{j_2}_{\nu_2\mu_2}(g) D^{j_3}_{\nu_3\mu_3}(g)\, dg,$$

since there are no general formulas for the integrals of products of three hypergeometric functions corresponding to the $O(2,1)$ representation functions. Because of the second fact, there are no simple CG coefficients for particular values of the magnetic quantum numbers, which are used as a starting point for the general case in the $O(3)$ group. Therefore, we use an indirect method, initiated by Andrews and Gunson[5] and developed by Ferretti and Verde.[8]

We begin by introducing notations and conventions. The angular momentum j is defined through the Casimir invariant Q of $O(2,1)$

$$Q = J_1^2 + J_2^2 - J_3^2 = -(j + \tfrac{1}{2})(j - \tfrac{1}{2}),$$

where J_i is the ith infinitesimal generator of $O(2,1)$. The quantum number j differs from the corresponding $O(3)$ quantum number by $\frac{1}{2}$; the definition used here has the advantage that the Legendre transformation involves replacing j_i by $-j_i$. The representation of $O(2,1)$ is given[6,14] by

$$\begin{aligned} d_{\nu\mu}^{\ j}(z) = {} & \left(\frac{\Gamma(\tfrac{1}{2} - j + \nu)\Gamma(\tfrac{1}{2} + j + \nu)}{\Gamma(\tfrac{1}{2} - j + \mu)\Gamma(\tfrac{1}{2} + j + \mu)}\right)^{\frac{1}{2}} \\ & \times \left(\frac{2}{z+1}\right)^{\frac{1}{2}(\nu+\mu)} \left(\frac{z-1}{2}\right)^{\frac{1}{2}(\nu-\mu)} \frac{1}{\Gamma(\nu - \mu + 1)} \\ & \times {}_2F_1(\tfrac{1}{2} + j - \mu, \tfrac{1}{2} - j - \mu; \nu - j - \mu; \\ & \qquad \nu - \mu + 1; \tfrac{1}{2}(1 - z)). \quad (1) \end{aligned}$$

The principal sheet in the j plane is defined by requiring that $d_{\nu\mu}^{\ j}(z)$ be positive for large and positive j. Thus, $d_{\nu\mu}^{\ j}(z)$ has cuts along the real axis whose positions depend on the relative values of ν and μ. With this convention, one has

$$d_{\nu\mu}^{\ -j}(z) = d_{\nu\mu}^{\ j}(z). \quad (2)$$

For $\nu > \mu > 0$, all the factors in (1) are finite, but some of these factors may be divergent for other cases. However, one may take a limit as j approaches an integer or a half-integer and, by using the well-known transformations of hypergeometric functions, one finds that the product on the right-hand side of

(1) is always finite. The results are

$$d^j_{-\nu-\mu}(z) = (-1)^{\nu-\mu} d_{\nu\mu}{}^j(z),$$
$$d_{\mu\nu}{}^j(z) = (-1)^{\nu-\mu} d_{\nu\mu}{}^j(z), \quad (3)$$

and

$$d^j_{-\mu-\nu}(z) = d_{\nu\mu}{}^j(z).$$

Usually, these relations are quoted for $\nu > \mu > 0$ and used to extend the definition of the representation function $d_{\nu\mu}{}^j(z)$ to other cases. In the sense of the limiting process mentioned above, the relations (3) are valid for any integral ν and μ. When some factors in (1) are zero or infinite, it is always implied that one takes the limit as j approaches an integer or a half-integer.

Following Bargmann,[14] one has, for the continuous series,

$$\mathrm{Re}\, j = 0, \quad \nu, \mu = 0, \pm 1, \pm 2, \cdots, \quad (4)$$

for the positive discrete series,

$$j = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots,$$
$$\nu, \mu = j + \tfrac{1}{2}, j + \tfrac{3}{2}, \cdots, \quad (5)$$

and for the negative discrete series,

$$j = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2}, \cdots,$$
$$\nu, \mu = -j - \tfrac{1}{2}, -j - \tfrac{3}{2}, \cdots. \quad (6)$$

The orthogonality relation is

$$\int_1^\infty dz d_{\nu\mu}{}^{j*}(z) d_{\nu\mu}{}^{j'}(z) = \frac{\delta(j,j')}{\eta(j)}, \quad (7)$$

where

$$\delta(j,j') = \delta(ij - ij'), \quad \text{for continuous series } j \text{ and } j',$$
$$= \delta_{jj'}, \quad \text{for discrete series } j \text{ and } j',$$
$$= 0, \quad \text{for one continuous series and one discrete series,}$$

and

$$\eta(j) = 2j \tan \pi(j - \mu), \quad \text{for the continuous series,}$$
$$= 2j, \quad \text{for the discrete series.} \quad (8)$$

The analytic continuation for $|1 - z| > 2$ of the representation function $d_{\nu\mu}{}^j(z)$ can be expressed as

$$d_{\nu\mu}{}^j(z) = a_{\nu\mu}{}^j(z) + a_{\nu\mu}{}^{-j}(z), \quad (9)$$

where $a_{\nu\mu}{}^j(z)$ is defined as

$$a_{\nu\mu}{}^j(z) \equiv -\frac{\pi}{\sin 2\pi j}\left(\frac{\Gamma(\frac{1}{2} - j + \nu)\Gamma(\frac{1}{2} + j + \nu)}{\Gamma(\frac{1}{2} - j + \mu)\Gamma(\frac{1}{2} + j + \mu)}\right)^{\frac{1}{2}} \times \left(\frac{z+1}{z-1}\right)^{-\frac{1}{2}(\nu+\mu)} \left(\frac{2}{z-1}\right)^{j+\frac{1}{2}} \times [\Gamma(\tfrac{1}{2} - j - \mu)\Gamma(\tfrac{1}{2} - j + \nu)\Gamma(2j+1)]^{-1} \times {}_2F_1\left(\tfrac{1}{2} + j - \nu, \tfrac{1}{2} + j - \mu; 2j + 1; \frac{2}{1-z}\right). \quad (10)$$

For the discrete series, we have

$$a_{\nu\mu}{}^{-j}(z) = a_{\nu\mu}{}^j(z),$$

i.e.,

$$d_{\nu\mu}{}^j(z) = 2a_{\nu\mu}{}^j(z). \quad (11)$$

In deriving (11), we have used the relation

$$F(a, b; c; z) = \frac{\Gamma(a - c + 1)\Gamma(b - c + 1)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(-c + 2)} z^{-c+1} \times {}_2F_1(a - c + 1, b - c + 1; -c + 2; z),$$

for negative integral c.

The CG coefficient $C(j_1, j_2, j_3; \nu_1, \nu_2)$ of $O(2,1)$, like that of $O(3)$, should satisfy the following conditions:

(a) Clebsch–Gordan series:

$$d^{j_1}_{\nu_1\mu_1}(z) d^{j_2}_{\nu_2\mu_2}(z) = \sum_{j_3} C(j_1, j_2, j_3; \nu_1, \nu_2) a^{j_3}_{\nu_3\mu_3}(z) \times C^*(j_1, j_2, j_3; \mu_1, \mu_2), \quad (12)$$

where $\sum$ means that one sums over all the discrete series and integrates over all the continuous series that occur in the reduction of the two principal series j_1 and j_2. From the conservation of magnetic quantum number, one has $\nu_3 = \nu_1 + \nu_2$ and $\mu_3 = \mu_1 + \mu_2$.

(b) Recursion relation:

$$\begin{aligned}[(\tfrac{3}{2} + j_1 + \nu_1)(\tfrac{3}{2} - j_1 + \nu_1)(\tfrac{3}{2} + j_2 - \nu_2)(\tfrac{3}{2} - j_2 - \nu_2)]^{\frac{1}{2}} &\times C(j_1, j_2, j_3; \nu_1 + 2, \nu_2 - 2) \\ + [(\tfrac{3}{2} + j_1 + \nu_1)(\tfrac{3}{2} - j_1 + \nu_1) &+ (\tfrac{1}{2} + j_2 - \nu_2)(\tfrac{1}{2} - j_2 - \nu_2) \\ - (\tfrac{1}{2} - j_3 + \nu_3)(\tfrac{1}{2} + j_3 + \nu_3)]^{\frac{1}{2}} &\times C(j_1, j_2, j_3; \nu_1 + 1, \nu_2 + 1) \\ + [(\tfrac{1}{2} + j_1 + \nu_1)(\tfrac{1}{2} - j_1 + \nu_1) &\times (\tfrac{1}{2} + j_2 - \nu_2)(\tfrac{1}{2} - j_2 - \nu_2)]^{\frac{1}{2}} \\ &\times C(j_1, j_2, j_3; \nu_1 \nu_2) = 0, \quad (13)\end{aligned}$$

for both continuous and discrete series.

(c) Orthogonality and normalization condition:

$$\sum_{\nu_1} C^*(j_1, j_2, j_3; \nu_1, \nu_2) C(j_1, j_2, j_3'; \nu_1, \nu_2) = \delta(j_3', j_3), \quad (14)$$

for fixed ν_3 and $\nu_2 = \nu_3 - \nu_1$. The summation $\sum$ for ν_1 means that one sums over all the possible values of ν_1 such that ν_1, ν_2, and ν_3 are in the spectra of the magnetic quantum numbers of the UIR's j_1, j_2, and j_3, respectively, as stated in Eqs. (4)–(6).

These three conditions are sufficient to determine the CG coefficient up to a phase factor which could be a function of the j_i, but they are not all necessary. If the CG coefficient does not have multiplicity of order

two, the first condition is enough. In order to remove those phase factors which depend on the j_i, one must introduce the continuation condition for the CG coefficient. That is to say, the CG coefficient for all cases can be expressed by one analytic function. It is because the nonconstant phase factor in the CG coefficient including the discrete (or, alternatively, the continuous) series, when it is continued to the domain corresponding to the continuous (or, alternatively, the discrete) series, is no longer a phase factor and, thus, should be omitted. The CG coefficient for three continuous series has multiplicity two; one therefore requires the third condition to obtain the individual coefficients, as is explained later. The second condition may be taken as a consistency condition. Similarly, the second and the third conditions may be used to determine the CG coefficient. We use the former method.

The CG series for two continuous series j_1 and j_2 with positive magnetic quantum numbers has been worked out by Ferretti and Verde.[8] Since our expression is somewhat different from theirs, we derive it briefly in order to show how the difference occurs.

Using the Burchnall–Chaundy formula,[15]

$$
\begin{aligned}
&{}_2F_1(a, b; c; x)\, {}_2F_1(\alpha, \beta; \gamma; x) \\
&\quad = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n (\gamma)_n}{n!\,(c)_n (c+\gamma+n-1)_n} \\
&\qquad \times {}_3F_2(\alpha, 1-c-n, -n; \gamma, 1-a-n) \\
&\qquad \times {}_3F_2(\beta, 1-c-n, -n; \gamma, 1-b-n) x^n \\
&\qquad \times {}_2F_1(a+\alpha+n, b+\beta+n; c+\gamma+2n; x),
\end{aligned}
\tag{15}
$$

where $(a)_n = \Gamma(a+n)/\Gamma(a)$, etc., one obtains, for $\nu_i > \mu_i > 0$,

$$
a^{j_1}_{\nu_1\mu_1}(z) a^{j_2}_{\nu_2\mu_2}(z) = \sum_{n=0}^{\infty} c_n \frac{\pi \tan \pi[(j_3)_n - \mu_3]}{2(j_3)_n} a^{(j_3)_n}_{\nu_3\mu_3}(z), \tag{16}
$$

where $(j_3)_n = \frac{1}{2} + j_1 + j_2 + n$, and j_1 and j_2 are in the continuous series, i.e., they are pure imaginary. The coefficient c_n is written in terms of the product of two G functions,

$$
c_n = -2(j_3)_n^2 \operatorname*{Res}_{j_3=(j_3)_n} \{G(j, \nu) G(j, -\mu)\}, \tag{17}
$$

with the G function defined as

$$
\begin{aligned}
G(j, \nu) &\equiv G(j_1, j_2, j_3; \nu_1, \nu_2) \\
&\equiv \pi^{\frac{1}{2}} \alpha(j_1, \nu_1) \alpha(j_2, \nu_2) \alpha(-j_3, \nu_3) \omega(j_1, j_2, j_3) \\
&\quad \times K(j, \nu) F_{\nu\nu}(0.45) / \Gamma(\tfrac{1}{2} - j_2 + \nu_2) \\
&\quad \times [\sin 2\pi j_2]^{\frac{1}{2}},
\end{aligned}
\tag{18}
$$

where

$$
\begin{aligned}
\alpha(j, \nu) &= [\Gamma(\tfrac{1}{2} - j + \nu)/\Gamma(\tfrac{1}{2} - j - \nu)]^{\frac{1}{2}}, \\
\omega(j_1, j_2, j_3) &= [\Gamma(\tfrac{1}{2} + j_1 + j_2 + j_3) \Gamma(\tfrac{1}{2} - j_1 + j_2 + j_3) \\
&\quad \times \Gamma(\tfrac{1}{2} + j_1 + j_2 - j_3) \\
&\quad \times \Gamma(\tfrac{1}{2} - j_1 + j_2 - j_3)]^{\frac{1}{2}},
\end{aligned}
$$

and

$$
\begin{aligned}
&K(j, \nu) \equiv K(j_1, j_2, j_3; \nu_1, \nu_2) \\
&= \left(\frac{\sin \pi(\frac{1}{2} + j_1 + \nu_1) \sin \pi(\frac{1}{2} + j_2 + \nu_2) \sin \pi(\frac{1}{2} + j_3 - \nu_3)}{\sin \pi(\frac{1}{2} + j_1) \sin \pi(\frac{1}{2} + j_2) \sin \pi(\frac{1}{2} + j_3)} \right)^{\frac{1}{2}}.
\end{aligned}
\tag{19}
$$

The Thomae–Whipple function $F_{\nu\nu}(0.45)$ is defined as

$$
\begin{aligned}
F_{\nu\nu}(0.45) &\equiv {}_3F_2(\tfrac{1}{2} + j_1 + j_2 - j_3, \tfrac{1}{2} - j_1 + j_2 - j^3, \\
&\qquad \tfrac{1}{2} + j_2 - \nu_2; 1 + j_2 - j_3 + \nu_1; 1 + 2j_2) \\
&\quad \times [\Gamma(\tfrac{1}{2} + j_3 + \nu_3) \Gamma(1 + j_2 - j_3 + \nu_1) \\
&\quad \times \Gamma(1 + 2j_2)]^{-1},
\end{aligned}
$$

where ${}_3F_2$ is the generalized hypergeometric function[16,17] with unit argument. It is invariant under exchange of j_1 to $-j_1$ or j_3 to $-j_3$, or both. The G function has a one-over-square-root singularity at $j_3 = (j_3)_n$ via $\omega(j_1, j_2, j_3)$, so the product of two G functions has simple pole, and the coefficient c_n is the residue at $j_3 = (j_3)_n$. In deriving (16), we have used the relation

$$
\begin{aligned}
F_{\nu-\mu}(0) &= (-1)^{\frac{1}{2} + j_1 + j_2 - (j_3)_n} \\
&\quad \times \frac{\Gamma(\frac{1}{2} - j_1 + \mu_1) \Gamma(\frac{1}{2} + (j_3)_n + \mu_3)}{\Gamma(\frac{1}{2} - j_1 - \mu_1) \Gamma(\frac{1}{2} + (j_3)_n - \mu_3)} F_{\nu\mu}(0),
\end{aligned}
$$

which can be proved from the definition of $F_{\nu\nu}(0)$ and the relation

$$
\begin{aligned}
&\frac{\Gamma(\frac{1}{2} + j - \mu) \Gamma(\frac{1}{2} - j + \mu)}{\Gamma(\frac{1}{2} + j) \Gamma(\frac{1}{2} - j)} \\
&\qquad = \frac{\Gamma(\frac{1}{2} + j) \Gamma(\frac{1}{2} - j)}{\Gamma(\frac{1}{2} + j + \mu) \Gamma(\frac{1}{2} - j - \mu)}.
\end{aligned}
$$

One notes that the G function here is different from the Wigner coefficient defined by Ferretti and Verde by a phase factor $K(j_1, j_3, j_3; \nu_1, \nu_2)$. The sine functions in (18) and (19) and in the rest of this paper are only symbols to represent the inverse of the product of two γ functions. Whenever one considers the phase factor for an expression involving sine functions, one must investigate the phase factor of the γ functions through the relation

$$
\sin \pi(\tfrac{1}{2} + j + n) = \pi / \Gamma(\tfrac{1}{2} + j + n) \Gamma(\tfrac{1}{2} - j - n). \tag{20}
$$

This process fixes the phase factor of the expression uniquely. In this sense, one has, for $\mathrm{Re}\, j_i \geq 0$,

$$\begin{aligned} K(j_1, j_2, j_3; \nu_1, \nu_2) &= 1, \\ K(-j_1, j_2, j_3; \nu_1, \nu_2) &= (-1)^{\nu_1}, \\ K(j_1, -j_2, j_3; \nu_1, \nu_2) &= (-1)^{\nu_2}, \end{aligned} \tag{21}$$

and

$$K(j_1, j_2, -j_3; \nu_1, \nu_2) = (-1)^{\nu}.$$

From Eqs. (18), (19), and (21), it follows that

$$\begin{aligned} G(j_1, j_2, j_3; \nu_1, \nu_2) &= G(-j_1, j_2, j_3; \nu_1, \nu_2) \\ &= G(-j_1, j_2, -j_3; \nu_1, \nu_2) \\ &= G(j_1, j_2, -j_3; \nu_1, \nu_2). \end{aligned} \tag{22}$$

These invariance properties are different by a phase factor from the Wigner coefficient in Ref. 8, because of the additional factor $K(j, \nu)$ in our definition. These properties are important to prove the positivity of the CG series, as we discuss later.

The product $G(-j, \nu)G(-j, -\mu)$ has no poles at $j_3 = (j_3)_n$; one may replace (17) by

$$c_n = -2(2j_3)_n^2 \operatorname*{Res}_{j_3=(j_3)_n} [G(j, \nu)G(j, -\mu) + G(-j, \nu)G(-j, -\mu)]. \tag{23}$$

Changing the summation in (16) into a contour integral and performing similar manipulations with $a^{-j_1}_{\nu_1\mu_1}(z)a^{j_2}_{\nu_2\mu_2}(z)$, $a^{j_1}_{\nu_1\mu_1}(z)a^{-j_2}_{\nu_2\mu_2}(z)$, and $a^{-j_1}_{\nu_1\mu_1}(z)a^{-j_2}_{\nu_2\mu_2}(z)$, one has

$$\begin{aligned} &d^{j_1}_{\nu_1\mu_1}(z)d^{j_2}_{\nu_2\mu_2}(z) \\ &= i \oint_{\gamma_1+\gamma_2+\gamma_3+\gamma_4} dj_3(2j_3) \tan \pi(j_3 - \mu_3) \\ &\quad \times [G(j, \nu)G(j, -\mu) + G(-j, \nu)G(-j, -\mu)]a^{j_3}_{\nu_3\mu_3}(z), \end{aligned} \tag{24}$$

by use of (9), where the contours enclose the poles

$$\begin{aligned} j_3 &= \tfrac{1}{2} + j_1 + j_2 + n, \\ j_3 &= \tfrac{1}{2} + j_1 - j_2 + n, \\ j_3 &= \tfrac{1}{2} - j_1 + j_2 + n, \end{aligned} \tag{25}$$

and

$$j_3 = \tfrac{1}{2} - j_1 - j_2 + n, \quad \text{for} \quad n = 0, 1, 2, \cdots,$$

as shown in Fig. 1. Investigation of the asymptotic behavior in the j_3 plane shows that a Sommerfeld–Watson transform is possible. Hence, after deforming the entire contour onto the imaginary axis and picking

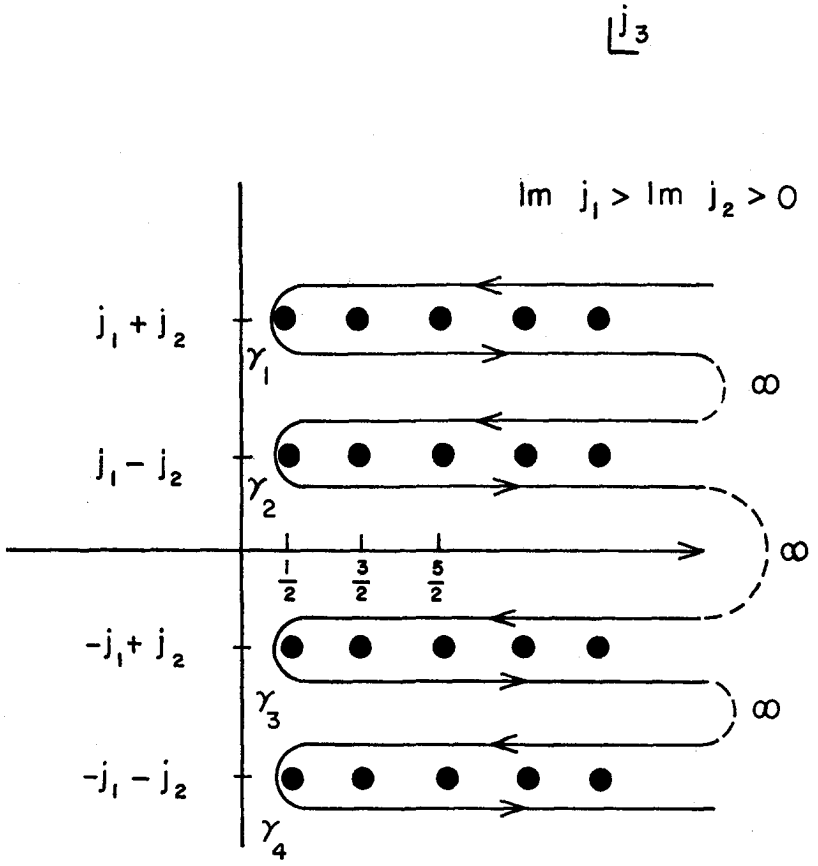


FIG. 1. The contour of the integral with $\mathrm{Im}\, j_1 > \mathrm{Im}\, j_2 > 0$ for the case in which j_1 and j_2 are both in the continuous series. The contours γ_1 and γ_2 enclose the contributing poles of $G(j, \nu)G(j, -\mu)$, and those for γ_3 and γ_4 enclose the poles of $G(-j, \nu)G(-j, -\mu)$. These four contours become a single one by being connected at infinity, as shown in this figure. Similar figures can be obtained for the cases

$$\mathrm{Im}\, j_1 > -\mathrm{Im}\, j_2 > 0,$$

$$-\mathrm{Im}\, j_1 > \pm \mathrm{Im} j_2 > 0,$$

and

$$0 < \pm \mathrm{Im}\, j_1 < \pm \mathrm{Im}\, j_2.$$

up the pole terms, one has[8]

$$d^{j_1}_{\nu_1\mu_1}(z)d^{j_2}_{\nu_2\mu_2}(z) = -\int_{0i}^{i\infty} i\,dj_3(2j_3)\tan\pi(j_3-\mu_3) \times [G(j,\nu)G(j,-\mu)+G(-j,\nu)G(-j,-\mu)]d^{j_3}_{\nu_3\mu_3}(z) + \sum_{j_3=\frac{1}{2}}^{[\mu_3-\frac{1}{2}]} 2j_3[G(j,\nu)G(j,-\mu) + G(-j,\nu)G(-j,-\mu)]d^{j_3}_{\nu_3\mu_3}(z). \quad (26)$$

This is the CG series for the product of two continuous series j_1 and j_2. The two terms in the brackets in the first term of (26) cannot be separated into two factors; one depends on the magnetic quantum numbers ν_i and the other on the μ_i. The fact that they cannot reflects that the CG coefficient for three continuous series has multiplicity of order two, as proved in the literature.[8] Thus, the j_3 spectrum consists of two continuous series and one positive discrete series. Comparing (26) with (14), one has

$$C_1(j,\nu)C_1^*(j,\mu)+C_2(j,\nu)C_2^*(j,\mu) = \eta(j_3)[G(j,\nu)G(j,-\mu)+G(-j,\nu)G(-j,-\mu)] \quad (27)$$

for the continuous series j_3, and

$$C(j,\nu)C^*(j,\mu) = \eta(j_3)[G(j,\nu)G(j,-\mu)+G(-j,\nu)G(-j,-\mu)] \quad (28)$$

for the positive-discrete series j_3. Before identifying the CG coefficient, we must study the properties of the G function.

III. THE G FUNCTION

Whipple and Thomae[18] investigated the relationships among the Thomae–Whipple functions $F_p(l; m, n)$, which are defined as

$$F_p(l;m,n) = [\Gamma(\alpha_{ghj})\Gamma(\beta_{ml})\Gamma(\beta_{nl})]^{-1} \times {}_3F_2(\alpha_{gmn},\alpha_{hmn},\alpha_{jmn};\beta_{ml},\beta_{nl}), \quad (29)$$

where g, h, j, l, m, and n take 0, 1, 2, 3, 4, and 5 permutatively. The parameters α_{lmn} and β_{mn} are defined as

$$\alpha_{lmn} = \tfrac{1}{2}+\gamma_l+\gamma_m+\gamma_n$$

and

$$\beta_{mn} = 1+\gamma_m-\gamma_n \quad (30)$$

for any γ_i, $i = 0, \cdots, 5$, with the restriction

$$\sum_{i=0}^{5}\gamma_i = 0. \quad (31)$$

The convergence condition for $F_p(l; m, n)$ is $Rl(\alpha_{ghj}) > 0$. Thomae[18] showed that $F_p(l; m, n) = F_p(l; m', n')$ for any combination of l, m, n, m', and n'. The Thomae–Whipple function $F_p(l; m, n)$ is thus independent of m and n and may be denoted by $F_p(l)$. Hence, there are ten representations for $F_p(l)$ obtained by permuting m and n; each has a different convergence domain and thus is useful for continuation. For our purposes, we express the γ_i in terms of the angular momenta j_i and the magnetic quantum numbers ν_i. These relations are

$$\begin{aligned} 3\gamma_0 &= -3j_2-2\nu_1-\nu_2,\\ 3\gamma_1 &= 3j_1+\nu_1+2\nu_2,\\ 3\gamma_2 &= -3j_1+\nu_1+2\nu_2,\\ 3\gamma_3 &= 3j_3+\nu_1-\nu_2,\\ 3\gamma_4 &= -3j_3+\nu_1-\nu_2, \end{aligned} \quad (32)$$

and

$$3\gamma_5 = 3j_2-2\nu_1-\nu_2,$$

where the j_i may be taken as complex numbers. The relationships between the set $(\alpha_{lmn}, \beta_{mn})$ and the set (j, ν) are set forth in Table I. In this connection, the Thomae–Whipple function $F_p(l)$ may be represented by $F_{p\nu}(l)$ to emphasize that it depends on the triplet (ν_1, ν_2, ν_3). By (32), we see that exchange of the indices 1 and 2 (or 3 and 4) is equivalent to the replacement of j_1 by $-j_1$ (or j_3 by $-j_3$). Hence, $F_{p\nu}(0)$ is invariant under change of sign of j_1 or j_3 or both.

Besides, there are three-term relations for $F_{p\nu}(l)$, which are collected in Bailey's and Slater's books.[18]

TABLE I. The relationships between the set $(\alpha_{l,m,n}, \beta_{mn})$ and the set (j, ν).

$\alpha_{012} = \frac{1}{2}-j_2+\nu_2$	$\alpha_{024} = \frac{1}{2}-j_1-j_2-j_3$	$\alpha_{123} = \frac{1}{2}+j_3+\nu_1+\nu_2$	$\alpha_{145} = \frac{1}{2}+j_1+j_2-j_3$
$\alpha_{013} = \frac{1}{2}+j_1-j_2+j_3$	$\alpha_{025} = \frac{1}{2}-j_1-\nu_1$	$\alpha_{124} = \frac{1}{2}-j_3+\nu_3$	$\alpha_{234} = \frac{1}{2}-j_1+\nu_1$
$\alpha_{014} = \frac{1}{2}+j_1-j_2-j_3$	$\alpha_{034} = \frac{1}{2}-j_2-\nu_2$	$\alpha_{125} = \frac{1}{2}+j_2+\nu_2$	$\alpha_{235} = \frac{1}{2}-j_1+j_2+j_3$
$\alpha_{015} = \frac{1}{2}+j_1-\nu_1$	$\alpha_{035} = \frac{1}{2}+j_3-\nu_3$	$\alpha_{134} = \frac{1}{2}+j_1+\nu_1$	$\alpha_{245} = \frac{1}{2}-j_1+j_2-j_3$
$\alpha_{023} = \frac{1}{2}-j_1-j_2+j_3$	$\alpha_{045} = \frac{1}{2}-j_3-\nu_3$	$\alpha_{135} = \frac{1}{2}+j_1+j_2+j_3$	$\alpha_{345} = \frac{1}{2}+j_2-\nu_2$
$\beta_{01} = 1-j_1-j_2-\nu_3$	$\beta_{05} = 1-2j_2$	$\beta_{15} = 1+j_1-j_2+\nu_3$	$\beta_{34} = 1+2j_3$
$\beta_{02} = 1+j_1-j_2-\nu_3$	$\beta_{12} = 1+2j_1$	$\beta_{23} = 1-j_1-j_3+\nu_2$	$\beta_{35} = 1-j_2+j_3+\nu_1$
$\beta_{03} = 1-j_2-j_3-\nu_1$	$\beta_{13} = 1+j_1-j_3+\nu_2$	$\beta_{24} = 1-j_1+j_3+\nu_2$	$\beta_{45} = 1-j_2-j_3+\nu_1$
$\beta_{04} = 1-j_2+j_3-\nu_1$	$\beta_{14} = 1+j_2+j_3+\nu_2$	$\beta_{25} = 1-j_1-j_2+\nu_3$	

Many relations between the G functions can be derived by means of them. One of the important relations[8] is

$$G(j,\nu) = a(j,\nu)G(j,-\nu) + b(j,\nu)G(-j,-\nu), \quad (33)$$

where the coefficients $a(j,\nu)$ and $b(j,\nu)$ are defined as

$$\begin{aligned} a(j,\nu) &\equiv a(j_1,j_2,j_3;\nu_1,\nu_2) \\ &\equiv [\sin\pi(\tfrac{1}{2}+j_1+j_2-j_3) \\ &\quad\times \sin\pi(\tfrac{1}{2}+j_1-j_2+j_3)\sin\pi(\tfrac{1}{2}+j_2-\nu_2) \\ &\quad+ \sin\pi(\tfrac{1}{2}+j_3+\nu_3)\sin 2\pi j_2 \\ &\quad\times \sin\pi(1-j_2+j_3-\nu_1)] \\ &\quad\times [\sin\pi(\tfrac{1}{2}+j_3-\nu_3)\sin\pi(\tfrac{1}{2}+j_1-\nu_1) \\ &\quad\times \sin\pi(2j_2+1)]^{-1}K(k,-\nu)/K(j,+\nu) \quad (34) \end{aligned}$$

and

$$\begin{aligned} b(j,\nu) &\equiv b(j_1,j_2,j_3;\nu_1,\nu_2) \\ &= -[\sin\pi(\tfrac{1}{2}+j_1+j_2+j_3) \\ &\quad\times \sin\pi(\tfrac{1}{2}+j_1+j_2-j_3) \\ &\quad\times \sin\pi(\tfrac{1}{2}-j_1+j_2+j_3) \\ &\quad\times \sin\pi(\tfrac{1}{2}-j_1+j_2-j_3)\sin\pi(\tfrac{1}{2}+j_2+\nu_2) \\ &\quad\times \sin\pi(\tfrac{1}{2}+j_2-\nu_2)]^{\frac{1}{2}} \\ &\quad\times [\sin 2\pi j_2 \sin 2\pi(1+2j_3) \\ &\quad\times \sin\pi(\tfrac{1}{2}+j_1+\nu_1)\sin\pi(\tfrac{1}{2}+j_1-\nu_1) \\ &\quad\times \sin\pi(\tfrac{1}{2}+j_3-\nu_3)\sin\pi(\tfrac{1}{2}+j_3+\nu_3)]^{-\frac{1}{2}}. \end{aligned} \quad (35)$$

For particular values of the j_i, the $a(j,\nu)$ and $b(j,\nu)$ may take simpler forms. One may easily show that

$$\begin{aligned} a(-j,\nu) &= -a(j,\nu), \quad b(-j,\nu) = b(j,\nu), \\ a(j,-\nu) &= a(j,\nu), \quad b(j,-\nu) = b(j,\nu), \end{aligned} \quad (36)$$

and

$$[a(j,\nu)]^2 + [b(j,\nu)]^2 = 1.$$

The other important relation[8] is

$$G(j_2,j_1,j_3;\nu_2,\nu_1) = c(j,\nu)G(j,\nu) + d(j,\nu)G(-j,\nu), \quad (37)$$

where

$$\begin{aligned} c(j,\nu) &\equiv c(j_1,j_2,j_3;\nu_1,\nu_2) \\ &= -\left(\frac{\sin\pi(\tfrac{1}{2}-j_1+j_2+j_3)\sin\pi(\tfrac{1}{2}-j_1+j_2-j_3)}{\sin 2\pi j_1 \sin 2\pi j_2}\right)^{\frac{1}{2}} \end{aligned}$$

and

$$\begin{aligned} d(j,\nu) &\equiv d(j_1,j_2,j_3;\nu_1,\nu_2) \\ &= -\left(\frac{\sin\pi(\tfrac{1}{2}+j_1+j_2+j_3)\sin\pi(\tfrac{1}{2}-j_1-j_2+j_3)}{\sin 2\pi j_1 \sin 2\pi(1+j_2)}\right)^{\frac{1}{2}} \frac{K(j,+\nu)}{K(j,-\nu)}, \end{aligned} \quad (38)$$

with $[c(j,\nu)]^2 + [d(j,\nu)]^2 = 1$. The relations (33) and (37) show that $G(j,\nu)G(j,-\mu) + G(-j,\nu)G(-j,-\mu)$ is invariant under the exchange of ν_i and μ_i or j_1 and j_2, or both. For particular values of the j_i, the three-term relations reduce to two-term relations. We collect some of them that will be useful later. For the case in which $\frac{1}{2} \pm j_3 - \nu_3$ equals a negative integer or zero, one has

$$a(j,\nu)G(j,-\nu) + b(j,\nu)G(-j,-\nu) = 0$$

and

$$a(-j,\nu)G(-j,-\nu) + b(-j,\nu)G(j,-\nu) = 0. \quad (39)$$

For the case in which $\frac{1}{2} + j_1 - \nu_1$ is negative integer or zero, one has

$$\begin{aligned} G(j_2,j_1,j_3;\nu_2,\nu_1) &= \left(\frac{\sin 2\pi j_2}{\sin\pi(\tfrac{1}{2}+j_1-j_2+j_3)\sin\pi(\tfrac{1}{2}-j_1+j_2+j_3)\sin 2\pi j_1}\right)^{\frac{1}{2}} \\ &\quad\times \frac{\sin\pi(\tfrac{1}{2}+j_3-\nu_3)\sin\pi(\tfrac{1}{2}+j_1+\nu_1)}{\sin\pi(\tfrac{1}{2}+j_2-\nu_2)} G(j,-\nu) \end{aligned} \quad (40)$$

and

$$\begin{aligned} G(j_2,j_1,j_3;-\nu_2,-\nu_1) &= [\sin\pi(\tfrac{1}{2}+j_1-\nu_1)\sin\pi(\tfrac{1}{2}+j_1+\nu_1)\sin\pi(\tfrac{1}{2}+j_3+\nu_3)\sin\pi(1+2j_2)]^{\frac{1}{2}} \\ &\quad\times [\sin\pi(\tfrac{1}{2}+j_1+j_2+j_3)\sin\pi(\tfrac{1}{2}+j_2+j_1-j_3)\sin\pi(\tfrac{1}{2}+j_2+\nu_2) \\ &\quad\times \sin\pi(\tfrac{1}{2}+j_2-\nu_2)\sin\pi(\tfrac{1}{2}+j_3-\nu_3)\sin 2\pi j_1]^{-\frac{1}{2}} \\ &\quad\times \sin\pi(1+j_1+j_2-\nu_3)G(-j,-\nu). \end{aligned}$$

For the case in which $\frac{1}{2} \pm j_2 - \nu_2$ equals a negative integer or zero, one has

$$G(j, \nu) = - \frac{\sin \pi(\frac{1}{2} - j_3 + \nu_3)}{\sin \pi(1 - j_1 - j_2 + \nu_3)} G(j, -\nu) \quad (41)$$

and

$$G(-j, -\nu) = e^{-\frac{1}{2}\pi i} G(j, -\nu),$$
$$G(-j, \nu) = e^{\frac{1}{2}\pi i} G(j, \nu). \quad (42)$$

From (42), we see that $G(j, \nu)G(j, -\mu)$ is invariant under change of signs of all j_i. The limits of the G functions for two or more angular momenta in discrete series can be obtained from the above relations.

We are now at a stage where we can find the asymptotic behavior of the G function in j_i or ν_i for other parameters fixed. The ν_i are always taken as integers, half integers, or zeros. To derive the asymptotic behavior for large $|j_3|$, we may take, for example, $F_{p\nu}(0, 24)$ for $F_{p\nu}(0)$ with $\mathrm{Re}\,(\frac{1}{2} + j_1 + j_2 + j_3) > 0$. The generalized hypergeometric function ${}_3F_2$ is related[19] to the hypergeometric function ${}_2F_1$, the asymptotic behavior of which can be obtained. For example, we have[20]

$${}_2F_1(\tfrac{1}{2} - j_1 + \nu_1, \tfrac{1}{2} - j_3 + \nu_3; 1 - j_1 + j_2 + \nu_3; s) \sim O((-j_3 s)^{-\frac{1}{2}+j_1-\nu_1}) \quad (43)$$

for large $|j_3|$ and $\mathrm{Re}\,(j_3 s) > 0$, and with other parameters fixed. From (43), one obtains the asymptotic behavior in j_3 of the generalized hypergeometric function

$${}_3F_2(\tfrac{1}{2} - j_3 + \nu_3, \tfrac{1}{2} - j_1 + \nu_1, \tfrac{1}{2} - j_1 + j_2 - j_3; 1 - j_1 + j_2 + \nu_3, 1 + j_2 - j_3 + \nu_1) \sim O((-j_3)^{-\frac{1}{2}+j_1-\nu_1}). \quad (44)$$

Thus, the asymptotic behavior in j_3 of the G function is

$$G(j, \nu) \sim O((j_3)^{\nu_3 - 2\nu_1 - 1}) \quad (45)$$

for larger $|j_3|$, $\mathrm{Re}\, j_3 > 0$ and $\mathrm{Re}\,(\frac{1}{2} + j_1 + j_2 + j_3) > 0$ and with other parameters fixed. Similarly, the asymptotic behaviors in j_1 or j_2 of the G function can be obtained:

$$G(j, \nu) \sim O((j_1)^{-\nu_2 - \nu_3 - 1}) \quad (46)$$

for large $|j_1|$ and $\mathrm{Re}\, j_1 > 0$ and with other parameters fixed, and

$$G(j; \nu) \sim O((j_2)^{-\nu_1 - \nu_3 - 1}) \quad (47)$$

for large $|j_2|$ and $\mathrm{Re}\, j_2 > 0$ and with other parameters fixed. In obtaining these formulas, we have used the asymptotic behavior (44) in j_3 of the G function in order to derive the CG series for two continuous series j_1 and j_2. The asymptotic behaviors in the j_i are particularly important for performing the Sommerfeld–Watson transform in the j_i plane.

The asymptotic behaviors in ν_1 (or, equivalently, ν_2) or ν_3 can be obtained in a different way. For large positive ν_1, the functions $F_{n\nu}(4.23)$ and $F_{n\nu}(3.24)$, defined respectively as $F_{p\nu}(4.23)$ and $F_{p\nu}(3.24)$, with j_i and ν_i replaced by $-j_i$ and $-\nu_i$, have the following simple asymptotic behaviors:

$$F_{n\nu}(4.23) \sim [\Gamma(\tfrac{1}{2} - j_1 + \nu_1) \times \Gamma(1 + j_1 - j_3 + \nu_1 - \nu_3)\Gamma(1 - 2j_3)]^{-1}$$

and

$$F_{n\nu}(3.24) \sim [\Gamma(\tfrac{1}{2} - j_1 + \nu_1) \times \Gamma(1 + j_1 + j_3 + \nu_1 - \nu_3)\Gamma(1 + 2j_3)]^{-1}. \quad (48)$$

Using a three-term relation[18]

$$\frac{\sin \pi\beta_{43}}{\pi\Gamma(\alpha_{043})} F_{p\nu}(0) = \frac{F_{n\nu}(4.23)}{\Gamma(\alpha_{132})\Gamma(\alpha_{135})\Gamma(\alpha_{325})} - \frac{F_{n\nu}(3.24)}{\Gamma(\alpha_{142})\Gamma(\alpha_{145})\Gamma(\alpha_{425})},$$

one can obtain, by manipulating γ functions and by taking the Stirling approximation,

$$G(j; \nu) \sim \frac{1}{(\pi)^{\frac{1}{2}}} \alpha(-j_3, \nu_3)\omega(j_1, j_2, j_3) \left(\frac{\sin \pi(\frac{1}{2} + j_1 + \nu_1) \sin \pi(\frac{1}{2} + j_2 - \nu_2)}{\sin 2\pi j_2} \right)^{\frac{1}{2}} \times \left(\frac{(\nu_1)^{-\frac{1}{2}+j_3}\Gamma(2j_3)}{\Gamma(\frac{1}{2} + j_3 + \nu_3)\Gamma(\frac{1}{2} + j_1 + j_3 + j_3)\Gamma(\frac{1}{2} - j_1 + j_2 + j_3)} + (j_3 \leftrightarrow -j_3) \right) \quad (49)$$

for large positive ν_1 and $\mathrm{Re}\,(\frac{1}{2} - j_1 + \nu_1) > 0$. Similarly, one can obtain

$$G(j, \nu) \sim \frac{1}{(\pi)^{\frac{1}{2}}} \alpha(-j_3, \nu_3)\omega(j_1, j_2, j_3) \left(\frac{\sin \pi(\frac{1}{2} + j_2 + \nu_2)}{\sin 2\pi j_2 \sin \pi(\frac{1}{2} + j_1 - \nu_1)} \right)^{\frac{1}{2}} \times \left(\frac{\sin \pi(1 + j_2 - j_3 + \nu_1)\Gamma(2j_3)(-\nu_1)^{-\frac{1}{2}+j_3}}{\Gamma(\frac{1}{2} + j_1 + j_2 + j_3)\Gamma(\frac{1}{2} - j_1 + j_2 + j_3)\Gamma(\frac{1}{2} + j_3 + \nu_3)} + (j_3 \leftrightarrow -j_3) \right) \quad (50)$$

for large negative ν_1 and Re $(\frac{1}{2} - j_1 - \nu_1) > 0$,

$$G(j, \nu) \sim \frac{1}{(\pi)^{\frac{1}{2}}} \alpha(j_1, \nu_1)\omega(j_1, j_2, j_3)\left(\frac{\sin \pi(\frac{1}{2} + j_2 - \nu_2) \sin \pi(\frac{1}{2} + j_3 - \nu_3)}{\sin 2\pi j_2}\right)^{\frac{1}{2}} \times \left(\frac{(\nu_3)^{-\frac{1}{2}+j_1}\Gamma(2j_1)}{\Gamma(\frac{1}{2} + j_1 + j_2 - j_3)\Gamma(\frac{1}{2} + j_1 + j_2 + j_3)\Gamma(\frac{1}{2} + j_1 - \nu_1)} + (j_1 \leftrightarrow -j_1)\right) \quad (51)$$

for large positive ν_3 and Re $(\frac{1}{2} + j_1 + j_2 - j_3) > 0$, and

$$G(j, \nu) \sim \frac{1}{(\pi)^{\frac{1}{2}}} \alpha(j_1, \nu_1)\omega(j_1, j_2, j_3)\left(\frac{\sin \pi(\frac{1}{2} + j_2 - \nu_2)}{\sin 2\pi j_2 \sin \pi(\frac{1}{2} + j_3 + \nu_3)}\right)^{\frac{1}{2}} \times \left(\frac{\sin \pi(1 - j_1 + j_2 + j_3)\Gamma(2j_1)(-\nu_3)^{-\frac{1}{2}+j_1}}{\Gamma(\frac{1}{2} + j_1 + j_2 - j_3)\Gamma(\frac{1}{2} + j_1 + j_2 + j_3)\Gamma(\frac{1}{2} + j_1 + \nu_1)} + (j_1 \leftrightarrow -j_1)\right) \quad (52)$$

for large negative ν_3 and Re $(\frac{1}{2} + j_2 - \nu_2) > 0$. The asymptotic behavior in ν_2 is equivalent to that in $-\nu_1$. The asymptotic behaviors of the G functions, such as $G(-j, \nu)$ and $G(-j, -\nu)$, can be obtained by proper replacements in the expressions (49)–(52).

Behaviors of the G functions, when the j_i take any of the three principal series, are tabulated in Tables II–IV.

TABLE II. Behavior of the G functions for all j_i taking the values corresponding to the UIR of $O(2, 1)$ with $\nu_1 > 0$, $\nu_2 > 0$, and $\nu_3 > 0$. The symbols F, Z, and P indicate finiteness, zero, and pole, respectively. The supercripts on Z and P represent the order of zeros and poles, respectively.

j_1	j_2	j_3	$G(j, \nu)$	$G(j, -\nu)$	$G(-j, \nu)$	$G(-j, -\nu)$
cont.	cont.	cont.	F	F	F	F
cont.	cont.	discrete	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
cont.	discrete	cont.	F	F	F	F
cont.	discrete	discrete	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
discrete	cont.	cont.	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
discrete	cont.	discrete	Z	P	Z	P
discrete	discrete	cont.	...	...	...	...
discrete	discrete	discrete	F	F	F	F

TABLE III. Behavior of the G functions for all j_k taking the values corresponding to the UIR of $O(2, 1)$ with $\nu_1 > 0$, $-\nu_2 > 0$, and $\nu_3 > 0$. The symbols and superscripts have the same meanings as in Table II.

j_1	j_2	j_3	$G(j, \nu)$	$G(j, -\nu)$	$G(-j, \nu)$	$G(-j, -\nu)$
cont.	cont.	cont.	F	F	F	F
cont.	cont.	discrete	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
cont.	discrete	cont.	F	F	F	F
cont.	discrete	discrete	$Z^{\frac{1}{2}}$	$Z^{\frac{3}{2}}$	$Z^{\frac{1}{2}}$	$Z^{\frac{3}{2}}$
discrete	cont.	cont.	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
discrete	cont.	discrete	Z	P	Z	P
discrete	discrete	cont.	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
discrete	discrete	discrete	...	...	...	...

TABLE IV. Behavior of the G functions for all j_i taking the values corresponding to the UIR of $O(2, 1)$ with $-\nu_1$, $\nu_2 > 0$, and $\nu_3 > 0$. The symbols and superscripts have the same meanings as in Table II.

j_1	j_2	j_3	$G(j, \nu)$	$G(j, -\nu)$	$G(-j, \nu)$	$G(-j, -\nu)$
cont.	cont.	cont.	F	F	F	F
cont.	cont.	discrete	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
cont.	discrete	cont.	F	F	F	F
cont.	discrete	discrete	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$
discrete	cont.	cont.	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$
discrete	cont.	discrete	F	F	F	F
discrete	discrete	cont.	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$	$P^{\frac{1}{2}}$	$Z^{\frac{1}{2}}$
discrete	discrete	discrete	...	...	...	...

IV. CLEBSCH-GORDAN COEFFICIENT FOR $\nu_i > \mu_i > 0$

Since the CG coefficient for three continuous series has multiplicity of order two, we cannot determine it uniquely from (27) for this case. However, we are able to calculate two mutually orthogonal CG coefficients.

From the recursion relation[21] between the generalized hypergeometric functions, one can prove that the G function $G(j, \nu)$ satisfies the recursion relation (13) of the CG coefficients. Observing that the coefficients in the recursion relation are even functions of the j_i, one sees that $G(-j, \nu)$ also satisfies the recursion relation. Hence, any linear combination of $G(j, \nu)$ and $G(-j, \nu)$, with its coefficients as functions of j_i and ν_3 only, satisfies this recursion relation, and so do $G(j, -\nu)$ and $G(-j, -\nu)$ by (33). This fact strongly suggests that the continuable CG coefficients are linear combinations of $G(j, \nu)$ and $G(-j, \nu)$.

We now begin to check the orthogonality and normalization. For pure imaginary j_3, the condition (14) has to be replaced by

$$\lim_{\lambda\to 0^+} \sum_{\nu_1} C^*(j_1, j_2, j_3; \nu_1, \nu_2) C(j_1, j_2, j_3'; \nu_1, \nu_2) \nu_1^{-\lambda} = \delta(ij_3 - ij_3'), \quad (53)$$

since $C^*(j, \nu)C(j, \nu)$ has oscillating terms for this case. From (49) and (50), one has

$$\lim_{\lambda\to 0^+} \sum_{\nu_1} G(j_1, j_2, j_3; \nu_1, \nu_2) \times G(j_1, j_2, j_3'; -\nu_1 - \nu_2)\eta(j_3)\nu_1^{-\lambda} = \delta(ij_3 - ij_3') + \delta(ij_3 + ij_3')$$

and

$$\lim_{\lambda\to 0^+} \sum_{\nu_1} G(j_1, j_2, j_3; \nu_1, \nu_2) \times G(-j_1, -j_2, -j_3; \nu_1, \nu_2)\eta(j_3)\nu_1^{-\lambda} = 0 \quad (54)$$

for pure imaginary j_3. In deriving these relations (54), we have used the facts that (a) the inner product of two eigenfunctions with different eigenvalues j_3 and j_3' of the difference equation of second order vanishes, (b) the ordinary Riemann ζ function $\zeta(x)$ has a pole at $x = -1$ of unit residue, and (c) the singular part of the factor $(ij_3' - ij_3 + \lambda)^{-1}$ has the same effect as $\pi\delta(ij_3' - ij_3)$. However, the G function $G(j, \nu)$ is not the complex conjugate of $G(j, -\nu)$ for the case of three continuous series. One has to introduce new expressions which are linear combinations of $G(j, -\nu)$ and $G(j, \nu)$ and which are such that the orthogonality and normalization conditions are satisfied. One set of the candidates is the pair $[C_1(j, \nu), C_2(j, \nu)]$, with $C_1(j, \nu)$ and $C_2(j, \nu)$ defined as

$$C_1(j, \nu) \equiv C_1(j_1, j_2, j_3; \nu_1, \nu_2) \equiv [\eta(j_3)/b(j, \nu)]^{\frac{1}{2}} G(j, -\nu)$$

and

$$C_2(j, \nu) \equiv C_1(j_1, j_2, j_3; \nu_1, \nu_2) \equiv [\eta(j_3)/b(j, \nu)]^{\frac{1}{2}} G(j, \nu). \quad (55)$$

From (33), (53), and (54), one can easily show that

$$\sum_{\nu_1} C_i^*(j_1, j_2, j_3; \nu_1, \nu_2) C_j(j_1, j_2, j_3'; \nu_1, \nu_2) = \delta_{ij}[\delta(ij_3' - ij_3) + \delta(ij_3' + ij_3)] \quad (56)$$

and

$$C_1(j, \nu)C_1^*(j, \mu) + C_2(j, \nu)C_2^*(j, \mu) = \eta(j_3)[G(j, \nu)G(j, -\mu) + G(-j, \nu)G(-j, -\mu). \quad (57)$$

This pair of the CG coefficients satisfies all three conditions stated in Sec. II. Nevertheless, unitary transformation in the (C_1, C_2) space preserves the orthogonality, the normalization, and the quadratic form $C_1^*(j, \nu)C_1(j, \mu) + C_2^*(j, \nu)C_2(j, \mu)$. Infinite numbers of the pairs satisfy these conditions. One needs one more condition to fix the pair of the CG coefficients. The continuation condition is just what we require. We have to find two CG coefficients which are orthogonal to each other for the case of three continuous series; one of these must give the CG coefficients when it is continued to the values of j_i corresponding to other cases while the other must vanish. This can be achieved only after one works out the CG coefficient for the other cases. In this section, we assume that ν_i and μ_i are integers or zeros with the restriction $\nu_i > \mu_i > 0$.

For positive discrete j_3, we expect that the right-hand side of expression (28) can be factorized. In this case, one has

$$G(j, \nu) \sim Z^{\frac{1}{2}}, \quad G(j, -\nu) \sim P^{\frac{1}{2}}, \quad G(-j, \nu) \sim Z^{\frac{1}{2}}, \quad G(-j, -\nu) \sim P^{\frac{1}{2}}, \quad (58)$$

where Z and P indicate zero and pole, respectively. The superscript represents the order of the pole or zero. From (39) and (36), one can derive the equation

$$\eta(j_3)[G(j, \nu)G(j, -\mu) + G(-j, \nu)G(-j, -\mu) = \eta(j_3)G(j, -\nu)G(-j, -\mu)/b(j, \nu), \quad (59)$$

where $b(j, \nu)$ can be shown to be real, i.e.,

$$b(j, \nu) = [b(j, \nu)]^*. \quad (60)$$

From (22), one can identify the CG coefficient

$$C(j, \nu) = [\eta(j_3)/b(j, \nu)]^{\frac{1}{2}} G(j, -\nu). \quad (61)$$

It is interesting to note that $[b(j, \nu)]^{-\frac{1}{2}}$ may be imaginary for some j_3 because of the factors like $[\sin \pi(\frac{1}{2} + j_3)]^{\frac{1}{2}}$. However, this factor is compensated by the factors $[\Gamma(\frac{1}{2} - j_3 - \nu_3)]^{\frac{1}{2}}$ and $K(j, \nu)$ in $G(j, \nu)$.

Thus, we have

$$C^*(j,\nu) = [\eta(j_3)/b(j,\nu)]^{\frac{1}{2}} G(-j,-\nu). \quad (62)$$

Comparing (55) with (61), we see that $C(j,\nu)$ and $C_1(j,\nu)$ have the same functional form except that the factor $\eta(j_3)$, which depends on whether the j_3 is in continuous series or in discrete series, is different in the two cases. The other CG coefficient $C_2(j,\nu)$ vanishes for the discrete j_3 case.

In the case in which j_1 and j_2 are in the continuous series and the positive discrete series respectively, one can derive the CG series by a similar method. The function $G(j,\nu)G(j,-\mu)$ has two series of poles in the right-half j_3 plane, as shown in Fig. 2, instead of four series, as above. Thus, one does not need to add a vanishing term $\operatorname{Res}_{j_3=(j_3)_n} [G(-j,\nu)G(-j,-\mu)]$ to expression (17). Performing a Sommerfeld–Watson transform, one has

$$\begin{aligned} d^{j_1}_{\nu_1\mu_1}(z) d^{j_2}_{\nu_2\mu_2}(z) &= -2\int_{0i}^{\infty i} i\, dj_3 \eta(j_3) G(j,\nu)G(j,-\mu) d^{j_3}_{\nu_3\mu_3}(z) \\ &\quad + \sum_{j_3=\frac{1}{2}}^{[\mu_3-\frac{1}{2}]} 2\eta(j_3) G(j,\nu)G(j,-\mu) d^{j_3}_{\nu_3\mu_3}(z). \end{aligned} \quad (63)$$

In deriving (63), (42) is used. The j_3 spectrum is the same as in the above case. From (12), (41), and (63), one has

$$\begin{aligned} C(j,\nu)C^*(j,\mu) &= -\frac{2\eta(j_3)\sin\pi(\frac{1}{2}-j_3+\nu_3)}{\sin\pi(1-j_1+j_2+\nu_3)} G(j,-\nu)G(j,-\mu), \end{aligned} \quad (64)$$

from which one identifies the CG coefficient

$$\begin{aligned} C(j,\nu) &= C^*(j,\nu) \\ &\equiv \left(-\frac{2\eta(j_3)\sin\pi(\frac{1}{2}+j_3+\nu_3)}{\sin\pi(\frac{1}{2}-j_1+j_2+\nu_3)}\right)^{\frac{1}{2}} G(j,-\nu). \end{aligned} \quad (65)$$

If j_3 is in the discrete series, the factors like

$$[\sin\pi(\tfrac{1}{2}+j_3+\nu_3)]^{\frac{1}{2}}$$

are compensated by $[\Gamma(\frac{1}{2}+j_3-\nu_3)]^{\frac{1}{2}}$ and $K(j,\nu)$, and the factor $(-1)^{-\frac{1}{2}+j_2+\nu_3}$ from $\sin\pi(\frac{1}{2}-j_1+j_2+\nu_3)$ is compensated by $[\Gamma(\frac{1}{2}-j_2-\nu_2)\sin 2\pi j_2]^{-\frac{1}{2}}$ and $K(j,\nu)$. It is obvious now that the presence of the factor $K(j,\nu)$ in (18) removes a phase factor that depends on the ν_i in the CG coefficient and in the CG series. The CG coefficient is different from $C_1(j,\nu)$ and $C_2(j,\nu)$ in (55). However, the extra degree of freedom which we have observed in determining the CG coefficient enables one to redefine it for the case of three continuous series so that it satisfies our continuation condition. We redefine it after working out the CG coefficient for other combinations of j_1 and j_2.

In the cases in which j_1 and j_2 are in the discrete and the continuous series, respectively, one cannot replace the summation in (16) by a contour integral, since there are two series of double poles in the j_3 plane, as shown in Fig. 3. One way to remove this difficulty is to exchange the roles of j_1 and j_2, so that one can use

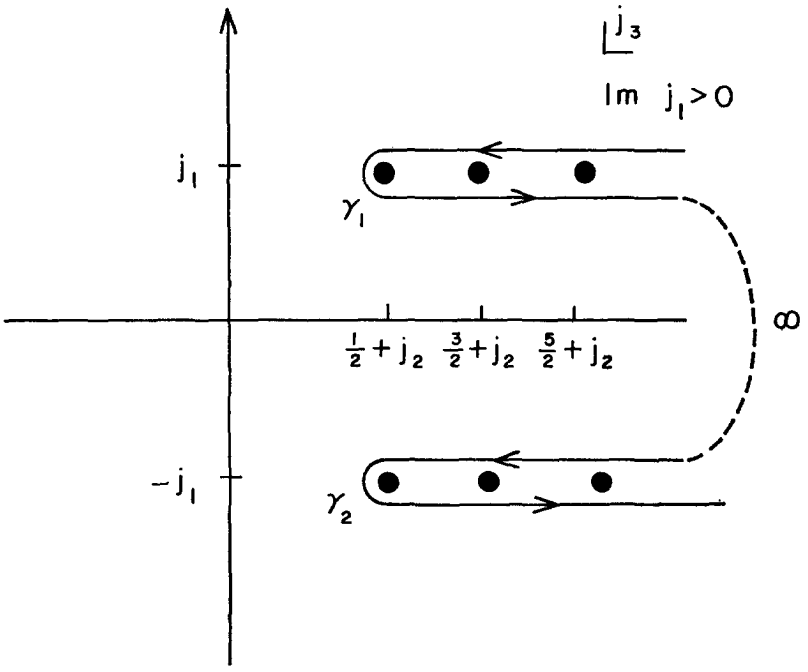


FIG. 2. The contour of the integral with $\operatorname{Im} j_1 > 0$ for the case in which j_1 and j_2 are in the continuous and the discrete series, respectively. The contours γ_1 and γ_2 enclose the contributing poles of $G(j,\nu) \times G(j,-\mu)$ in the right half-plane. These two contours become a single one by being connected at infinity, as shown in this figure. A similar figure can be obtained for $\operatorname{Im} j_1 < 0$.

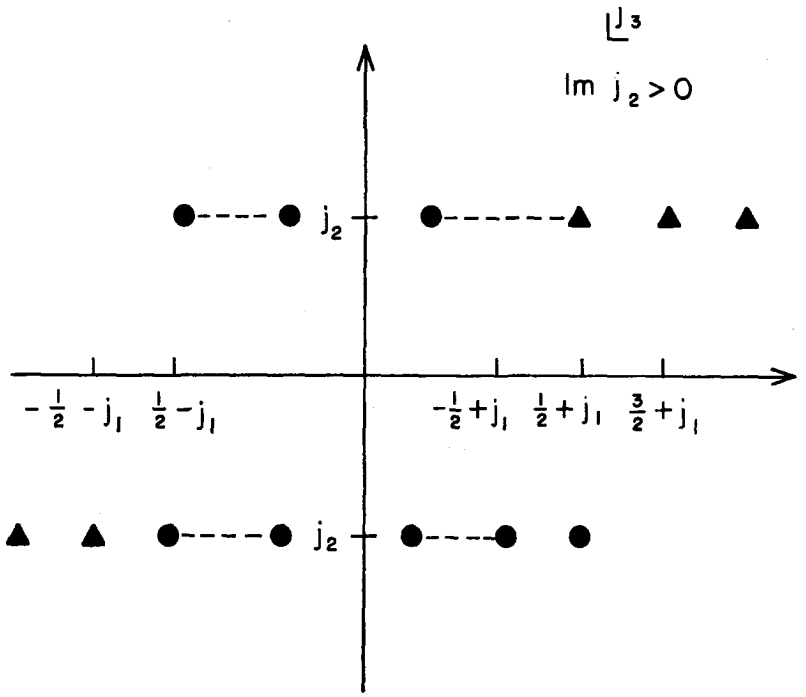


FIG. 3. The poles of $G(j, \nu)G(j, -\mu)$ with $\mathrm{Im}\, j_3 > 0$ in the j_3 plane for the case j_1 and j_2 in the discrete and the continuous series. The circle indicate a simple pole; the triangle indicates a double pole. A similar figure can be obtained for $\mathrm{Im}\, j_3 < 0$.

the previous method. One then obtains

$$d^{j_1}_{\nu_1\mu_1}(z)d^{j_2}_{\nu_2\mu_2}(z) = -2\int_{0i}^{\infty i} i\, dj_3\eta(j_3)G(j_2, j_1, j_3; \nu_2, \nu_1) \times G(j_2, j_1, j_3; -\mu_2, -\mu_1)d^{j_3}_{\nu_2\mu_3}(z) + \sum_{j_3=\frac{1}{2}}^{[\mu_3-\frac{1}{2}]} 2\eta(j_3)G(j_2, j_1, j_3; \nu_2, \nu_1) \times G(j_2, j_1, j_3; -\mu_2, -\mu_1)d^{j_3}_{\nu_3\mu_3}(z). \quad (66)$$

The j_3 spectrum is the same as in (63), as it should be. From (40), (66), and (35), one can identify the CG coefficient

$$C(j, \nu) = [\eta(j_3)/b(j, \nu)]^{\frac{1}{2}}G(j, -\nu) \quad (67)$$

with

$$b(k, \nu) = \frac{\sin \pi(j_2 + j_3) \sin \pi(-j_2 + j_3)}{2e^{\frac{1}{2}i\pi} \sin \pi(\frac{1}{2} + j_3) \sin \pi(\frac{1}{2} + j_1) \sin \pi j_2},$$

if considered in the j_3 plane. A similar expression can be obtained if considered in the j_2 plane. As above, the factors $[\sin \pi(\frac{1}{2} + j_1)]^{-\frac{1}{2}}$ and $[\sin \pi(\frac{1}{2} + j_3)]^{-\frac{1}{2}}$ in (67) are compensated by the factors in $G(j, \nu)$, if j_3 is pure imaginary. This CG coefficient is a continuation of $C_1(j, \nu)$, and $C_2(j, \nu)$ vanishes.

In the case in which both j_1 and j_2 are in the discrete series, one has, from (16) and (42),

$$d^{j_1}_{\nu_1\mu_1}(z)d^{j_2}_{\nu_2\mu_2}(z) = 2\sum_{j_3=\frac{1}{2}+j_1+j_2}^{[\mu_3-\frac{1}{2}]} \eta(j_3)G(j, \nu)G(j, -\mu)d^{j_3}_{\nu_3\mu_3}(z), \quad (68)$$

after discarding the vanishing terms. The j_3 spectrum in this case is well known.[7,9] The CG coefficient can be considered as the limiting case of (65) or (67).

In summary, we have obtained all the CG coefficients for $\nu_i > \mu_i > 0$. The CG coefficients for the three continuous series has multiplicity two. These two orthogonal CG coefficients $G(j, \nu)$ and $C_1(j, \nu)$ are defined in (55). Except for continuous j_1 and discrete j_2, the CG coefficient for other cases is the continuation of $C_1(j, \nu)$. Since the linear combination of $C_1(j, \nu)$ and $C_2(j, \nu)$ obtained by unitary transformation for the case of three continuous series is also a CG coefficient, we can find an expression such that the CG coefficient for all cases is equal to its continuation. With some calculations, we obtain the following two CG coefficients $C(j, \nu)$ and $C'(j, \nu)$:

$$C(j, \nu) = [\eta(j_3)]^{\frac{1}{2}}\{[D(j, \nu)]^{-\frac{1}{2}}G(j, -\nu) + [D(j, \nu)]^{\frac{1}{2}}G(j, \nu)\}/(\sqrt{2})[b(j, \nu)]$$

and

$$C'(j, \nu) = [\eta(j_3)]^{-\frac{1}{2}}\{-[D(j, \nu)]^{\frac{1}{2}}G(j, -\nu) + [D(j, \nu)^{-\frac{1}{2}}G(j, \nu)\}/(\sqrt{2})[b(j, \nu)], \quad (69)$$

where $D(j, \nu)$ is defined by

$$D(j, \nu) = -b(j, \nu) + \{[b(j, \nu)]^2 + 1\}^{\frac{1}{2}}. \quad (70)$$

One can see that $C(j, \nu)$ and $C'(j, \nu)$ are orthogonal for three continuous series. For all other cases, $C(j, \nu)$ reduces to the CG coefficient obtained from the CG series, and $C'(j, \nu)$ vanishes, except the two discrete j_1 and j_2 case in which $C(j, \nu)$ and $C'(j, \nu)$ are degenerate. Hence, $C(j, \nu)$ and $C'(j, \nu)$ are the required CG coefficients.

TABLE V. Behavior of $a^{j_3}_{\nu_3\mu_3}(z)$ when j_3 is taken as positive integer. The symbols and superscripts have the same meanings as in Table II.

		$\nu_3 > 0$		$-\nu_3 > 0$	
		$\nu_3 > j_3 > 0$	$j_3 > \nu_3 > 0$	$-\nu_3 > j_3 > 0$	$j_3 > -\nu_3 > 0$
$\mu_3 > 0$	$\mu_3 > j_3 > 0$	F	$Z^{\frac{1}{2}}$	F	$Z^{\frac{1}{2}}$
	$j_3 > \mu_3 > 0$	$Z^{\frac{1}{2}}$	Z	$Z^{\frac{1}{2}}$	Z
$-\mu_3 > 0$	$-\mu_3 > j_3 > 0$	F	$Z^{\frac{1}{2}}$	F	$Z^{\frac{1}{2}}$
	$j_3 > -\mu_3 > 0$	$Z^{\frac{1}{2}}$	Z	$Z^{\frac{1}{2}}$	Z

V. CLEBSCH-GORDAN COEFFICIENT FOR OTHER CASES

In the preceding section, we worked out the CG coefficients with the restriction that $\nu_i > \mu_i > 0$. In this section, we calculate the CG coefficient for arbitrary ν_i and μ_i. Finally, we extend our results to the double-valued UIR of $SU(1, 1)$.

Each representation function $d^j_{\nu\mu}(z)$ has four kinds of representations, as in (3). Hence, first of all, one must decide which one should be used in applying Burchnall–Chaundy formula (15). For convenience, we always choose the expressions (1) and (10) for the representation function of the rotation along the y axis sandwiched by the state vectors with magnetic quantum number ν_i and μ_i, irrespective of the relative values and the relative signs of ν_i and μ_i.

From (21), one can easily see that the discrete spectrum for j_3 is determined by poles of the integrand $j_3 \tan \pi j_3 G(j, \nu)G(j, -\mu)a^{j_3}_{\nu_3\mu_3}(z)$. If j_1 is in the continuous series and j_2 in the discrete series, the functions $F_{p\nu}(0)$ and $F_{p-\nu}(0)$ are finite for any ν_i and μ_ν, as they are both in the continuous series. If j_1 is in the discrete series and j_2 in the continuous series, one must derive the CG series by exchanging the roles of j_1 and j_2 in order to remove double poles which occur in the integrand as for the case $\nu_i > \mu_i > 0$. Hence, for all the cases except that of two discrete series, the functions $F_{p\nu}(0)$ and $F_{p-\nu}(0)$, as well as $\omega(j_1, j_2, j_3)$, are finite. The order of zeros or poles of the G functions can be found in Tables II–IV. For the last case, $F_{p\nu}(0)$, $F_{p-\nu}(0)$, and $\omega(j_1, j_2, j_3)$ behave differently for various relative values and signs of the j_i and ν_i. We discuss this case in more detail.

The derivation of the CG series can be carried out as for $\nu_i > \mu_i > 0$. The finiteness of the expression and the discrete spectrum for j_3 in the CG series can be determined by using Tables II–V.

Previous discussions on the normalization and orthogonality condition and unitary transformation are still valid for arbitrary ν_i and μ_i. In the following, we study the CG coefficients for any ν_i and μ_i in four cases.

A. j_1 and j_2 Continuous

The j_3 spectra for any ν_i and μ_i are given in Table VI. The two orthonormal CG coefficients for three continuous series are the same as in (55). For discrete j_3, the CG coefficient is defined as

$$C(j, \nu) \equiv C(j_1, j_2, j_3; \nu_1, \nu_2) \equiv [\eta(j_3)/b(j, \nu)]^{\frac{1}{2}} G(j, -\nu\zeta_A), \quad (71)$$

where ζ_A is determined by $\nu_3\zeta_A = |\nu_3|$. These CG coefficients are analytic continuations of one of the two CG coefficients in (55).

B. j_1 Continuous, j_2 Discrete

The j_3 spectra for any ν_i and μ_i are summarized in Table VII. Some spectra in Table VII are missing, since there is no discrete series for j_2 with two magnetic quantum numbers of different signs. We note that there is no positive-discrete series in the decomposition of the product of the continuous and negative-discrete series, and no negative-discrete series for

TABLE VI. The j_3 spectra from the decomposition of the product of two continuous series j_1 and j_2 for $\nu_3 > 0$. The positive-discrete spectrum runs from $j_3 = \frac{1}{2}$ to $j_3 = m + \frac{1}{2}$, the negative from $j_3 = -\frac{1}{2}$ to $j_3 = -\frac{1}{2} - m$, where m is the smaller of $|\mu_3|$ and $|\nu_3|$. The symbols c and d indicate continuous and discrete spectra, respectively. Similar results can be obtained for $\nu_3 < 0$.

	$\nu_1, \nu_2, \nu_3 > 0$	$-\nu_1, \nu_2, \nu_3 > 0$	$\nu_1, -\nu_2, \nu_3 > 0$
$\mu_1, \mu_2, \mu_3 > 0$	c, d	c, d	c, d
$-\mu_1, \mu_2, \mu_3 > 0$	c, d	c, d	c, d
$\mu_1, -\mu_2, \mu_3 > 0$	c, d	c, d	c, d
$-\mu_1, -\mu_2, -\mu_3 > 0$	c	c	c
$\mu_1, -\mu_2, \mu_3 > 0$	c	c	c
$-\mu_1, \mu_2, -\mu_3 > 0$	c	c	c

TABLE VII. The j_3 spectra from the decomposition of the product of one continuous series j_1 and one discrete series j_2. The range of the discrete spectrum of j_3 is the same as in Table VI. The symbols c and d indicate continuous and discrete spectra, respectively. Similar results can be obtained for $\nu_3 > 0$.

	$\nu_1, \nu_2, \nu_3 > 0$	$-\nu_1, \nu_2, \nu_3 > 0$	$\nu_1, -\nu_2, \nu_3 > 0$
$\mu_1, \mu_2, \mu_3 > 0$	c, d	c, d	$\cdots$
$-\mu_1, \mu_2, \mu_3 > 0$	c, d	c, d	$\cdots$
$\mu_1, -\mu_2, \mu_3 > 0$	$\cdots$	$\cdots$	c
$-\mu_1, -\mu_2, -\mu_3 > 0$	$\cdots$	$\cdots$	c
$+\mu_1, -\mu_1, -\mu_3 > 0$	$\cdots$	$\cdots$	c
$-\mu_1, \mu_2, -\mu_3 > 0$	c	c	$\cdots$

continuous and positive-discrete series, even though this discrete series is not forbidden by the conditions of UIR of $O(2, 1)$; one can see from Table III that in this case $G(j, \nu)G(j, -\mu)$ or $G(-j, \nu)G(-j, -\mu)$ vanishes as a double zero. This phenomenon of missing spectra also occurs in the decomposition of two positive-discrete (or negative-discrete) series; there is no continuous spectrum for j_3.

The explicit expression of the CG coefficient for any ν_i and μ_i is

$$C(j, \nu) = [-2\eta(j_3) \sin \pi(\tfrac{1}{2} + j_3 + \nu_3)/ \sin \pi(\tfrac{1}{2} - j_1 + j_2 + \nu_3)]^{\frac{1}{2}} \times G(j, -\nu\zeta_B), \quad (72)$$

where ζ_B is defined through $\nu_2\zeta_B = |\nu_2|$. This CG coefficient is identical to one of the two CG coefficients in (55), if the latter are continued in j_i to the region corresponding to this case.

C. j_1 Discrete, j_2 Continuous

The j_3 spectra are given in Table VIII. As in case (B), there is no positive-discrete spectrum of j_3 for the combination of one negative-discrete series and one continuous series (or no negative-discrete for continuous and positive-discrete). It is necessary to cope with the similar missing discrete spectrum in case (B), since by exchanging the roles of j_1 and j_2 case (C) becomes case (B). The CG coefficient is defined by

$$C(j, \nu) = [\eta(j_3)/b(j, \nu)]^{\frac{1}{2}} G(j, -\nu\zeta_c), \quad (73)$$

where ζ_c is determined by $\nu_1\zeta_c = |\nu_1|$. Once again, this CG coefficient is the analytic continuation of one of the two CG coefficients in (55).

D. j_1, j_2 Discrete

For $\nu_i > \mu_i > 0$, we have worked out the CG series (68). For many other ν_i and μ_i, there are no j_3 spectra, continuous and discrete, as shown in Table IX, because of the condition on the signs of the magnetic quantum numbers for the UIR of $O(2, 1)$. The derivations of the CG series for the present case are much more complicated for the reasons stated earlier in this section. We have divided this case into four subcases, according to the signs of ν_i and μ_i.

If j_1 is in the negative- and j_2 in the positive-discrete series, one obtains, from (11), (15), and (16),

$$d^{j_1}_{\nu_1\mu_1}(z) d^{j_2}_{\nu_2\mu_2}(z) = \sum_{j=j_1+j_2+\frac{1}{2}} G(j, \nu)G(j, -\mu)(8j_3)a^{j_3}_{\nu_3\mu_3}(z). \quad (74)$$

It is easily checked by using Table X that each term under the summation is finite except for $\nu_3 < -\frac{1}{2} - j_1 - j_2$, in which the terms with $j_3 < -\nu_3$ vanish. Equation (74) is therefore not a decomposition into UIR's. One can transform it into the required form by performing a Sommerfeld–Watson transform. By converting the summation into an integral, as shown in Fig. 4, one obtains the CG series, similar to that in (63). The determination of the discrete spectrum depends on the relative values among the j_i and the

TABLE VIII. The j_3 spectra from the decomposition of the product of one discrete series j_1 and one continuous series j_2 for $\nu_3 > 0$. The range of the discrete spectrum of j_3 is the same as in Table VI. The symbols c and d indicate continuous and discrete spectra, respectively. Similar results can be obtained for $\nu_3 < 0$.

	$\nu_1, \nu_2, \nu_3 > 0$	$-\nu_1, \nu_2, \nu_3 > 0$	$\nu_1, -\nu_2, \nu_3 > 0$
$\mu_1, \mu_2, \mu_3 > 0$	c, d	$\cdots$	c, d
$-\mu_1, \mu_2, \mu_3 > 0$	$\cdots$	c	$\cdots$
$\mu_1, -\mu_2, \mu_3 > 0$	c, d	$\cdots$	c, d
$-\mu_1, -\mu_2, -\mu_3 > 0$	$\cdots$	c	$\cdots$
$\mu_1, -\mu_2, -\mu_3 > 0$	c	$\cdots$	c
$-\mu_1, \mu_2, -\mu_3 > 0$	$\cdots$	c	$\cdots$

TABLE IX. The j_3 spectrum from the decomposition of the product of two discrete series j_1 and j_2 for $\nu_3 > 0$. The range of the positive- (negative-) discrete spectrum extends from $j_3 = \frac{1}{2}$ to $|\frac{1}{2} - j_1 + j_2|$. The symbols c and d indicate continuous and discrete spectra, respectively. The star * indicates that the negative-discrete spectrum occurs only when the angular momentum of the negative-discrete series is less than that of the positive-discrete series, and vice versa. A similar result can be obtained for $\nu_3 < 0$.

	$\nu_1, \nu_2, \nu_3 > 0$	$-\nu_1, \nu_2, \nu_3 > 0$	$\nu_1, -\nu_2, \nu_3 > 0$
$\mu_1, \mu_2, \mu_3 > 0$	d	...	...
$-\mu_1, \mu_2, \mu_3 > 0$	...	c, d^*	...
$\mu_1, -\mu_2, \mu_3 > 0$	...	...	c, d^*
$-\mu_1, -\mu_2, -\mu_3 > 0$	...	...	...
$\mu_1, -\mu_2, -\mu_3 > 0$	...	...	...
$-\mu_1, \mu_2, -\mu_3 > 0$	...	...	...

ν_i. By using Tables V and X, we can get the following results.

For $j_1 > j_2$, one has two classes:

(a) For $-\nu_3$, $-\mu_3 > 0$, one has one negative-discrete spectrum for j_3 extending from $j_3 = -\frac{1}{2}$ to $j_3 = -\frac{1}{2} - j_1 + j_2$, and one continuous series;

(b) For other cases, there are only continuous spectra. Similarly, for $j_2 > j_1$, one has one continuous series and one positive-discrete spectrum running from $j_3 = \frac{1}{2}$ to $j_3 = +\frac{1}{2} - j_1 + j_2$ for the case ν_3, $\mu_3 > 0$ and only one continuous spectrum for other cases.

From the above reasoning, one sees that there are no negative-discrete spectra in the decomposition of the product of one negative-discrete series and one positive-discrete series if the angular momentum of the latter is larger than that of the former and no positive-discrete spectrum if the angular momentum is less than the former. In a paper on the duality theorem for the $SU(1,1)$ group, Tatsuuma[22] obtained similar results.

If j_1 is in the positive and j_2 in the negative discrete series, one can obtain similar results. This can be verified directly by using Tables V and X. It is interesting to note that this subcase becomes the same as the above if one exchanges the role of j_1 and j_2 in (74) and finally in (63).

We have worked the subcase in which both j_1 and j_2 are in the positive-discrete series [see Eq. (68)]. In a similar manner, one can obtain the CG series for the two negative-discrete series. Again there is only one negative-discrete spectrum.

For all the subcases in Sec. V.D, the CG coefficients are the limits of those in Secs. V.B and V.C, as one of j_1 and j_2 becomes a half-integer.

The CG coefficients so far obtained are for the

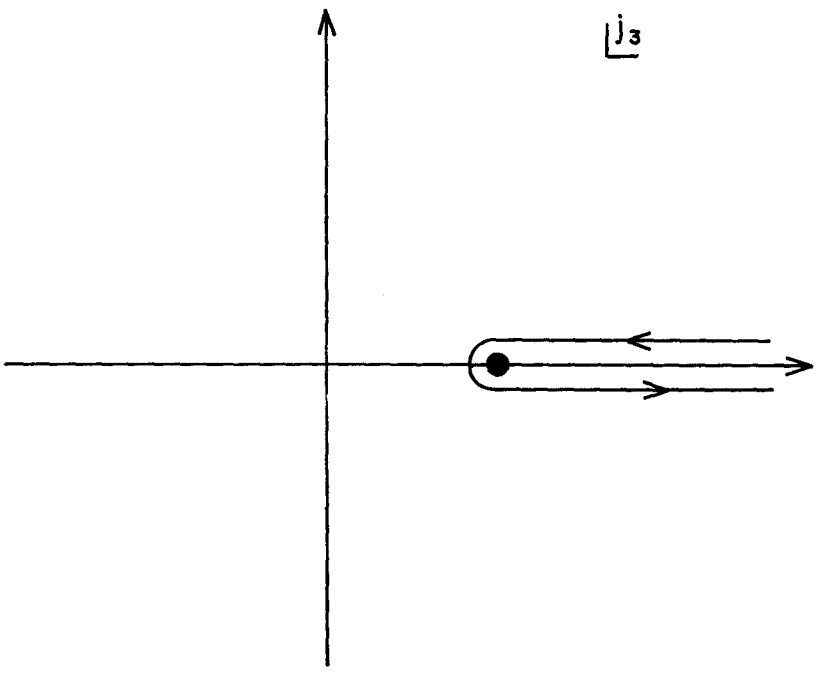


FIG. 4. The contour of the integral for the case in which j_1 and j_2 are in the negative and the positive discrete series. The starting point for the contour can be determined from Tables V and X. This figure is also valid for the case in which j_1 and j_2 are in the positive and the negative discrete series, respectively.

TABLE X. Behavior of the G functions for the case in which j_1 and j_2 are in the negative and the positive discrete series, respectively. The symbols and superscripts have the same meanings as in Table I. Behavior of the G functions for the case in which j_1 and j_2 are in the positive and the negative discrete series, respectively, can be obtained by exchanging the roles of j_1 and j_2.

	$\nu_3 > \frac{1}{2}+j_1+j_2$		$\frac{1}{2}+j_1+j_2 > \nu_3 > \lvert\frac{1}{2}-j_1+j_2\rvert$		$\lvert\frac{1}{2}-j_1+j_2\rvert > \nu_3 > -\lvert\frac{1}{2}-j_1+j_2\rvert$				$-\lvert\frac{1}{2}-j_1+j_2\rvert > \nu_3 > \frac{1}{2}-j_1-j_2$		$-\frac{1}{2}-j_1-j_2 > \nu_3$	
					$\nu_3 > 0$		$-\nu_3 > 0$					
	$j_3 > \nu_3$	$j_3 < \nu_3$	$j_3 > \nu_3$	$j_3 < \nu_3$	$j_3 > \nu_3$	$j_3 < \nu_3$	$j_3 > -\nu_3$	$j_3 < -\nu_3$	$j_3 > -\nu_3$	$j_3 < -\nu_3$	$j_3 > -\nu_3$	$j_3 < -\nu_3$
(a) $G(j, \nu)$ with $j_1 > j_2$												
$j_3 \geq \frac{1}{2}+j_1+j_2$	P^{j_3}	F	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	Z
$\frac{1}{2}+j_1+j_2 > j_3 \geq \lvert\frac{1}{2}-j_1+j_2\rvert$	$\cdots$	Z^{j_3}	F	Z^{j_3}	F	$\cdots$	F	$\cdots$	F	Z^{j_3}	$\cdots$	Z^{j_3}
$\lvert\frac{1}{2}-j_1+j_2\rvert > j_3 > -\lvert\frac{1}{2}-j_1+j_2\rvert$	$\cdots$	Z	$\cdots$	Z	Z^{j_3}	Z	Z^{j_3}	F	$\cdots$	F	$\cdots$	F
(b) $G(j, -\nu)$ with $j_1 > j_2$ and $G(j, \nu)$ with $j_1 < j_2$												
$j_3 \geq \frac{1}{2}+j_1+j_2$	P^{j_3}	F	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	F
$\frac{1}{2}+j_1+j_2 > j_3 \geq \lvert\frac{1}{2}-j_1+j_2\rvert$	$\cdots$	Z^{j_3}	F	Z^{j_3}	F	$\cdots$	F	$\cdots$	F	Z^{j_3}	$\cdots$	Z^{j_3}
$\lvert\frac{1}{2}-j_1+j_2\rvert > j_3 > -\lvert\frac{1}{2}-j_1+j_2\rvert$	$\cdots$	F	$\cdots$	F	Z^{j_3}	F	Z^{j_3}	F	$\cdots$	F	$\cdots$	F
(c) $G(j, -\nu)$ with $j_1 < j_2$												
$j_3 \geq \frac{1}{2}+j_1+j_2$	P^{j_3}	Z	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	$\cdots$	P^{j_3}	F
$\frac{1}{2}+j_1+j_2 > j_3 \geq \lvert\frac{1}{2}-j_1+j_2\rvert$	$\cdots$	Z^{j_3}	F	Z^{j_3}	F	$\cdots$	F	$\cdots$	F	Z^{j_3}	$\cdots$	Z^{j_3}
$\lvert\frac{1}{2}-j_1+j_2\rvert > j_3 > -\lvert\frac{1}{2}-j_1+j_2\rvert$	$\cdots$	F	$\cdots$	F	Z^{j_3}	F	Z^{j_3}	Z	$\cdots$	Z	$\cdots$	Z

single-valued representations, i.e., all the ν_i and μ_i take integral values or zeros. We show that one can extend them to the double-valued representations, i.e., ν_i and μ_i take half-integral values. One notes, however, that at least one pair of ν_i and μ_i are taken as integers or zeros because of conservation of magnetic quantum numbers. It is well known that double-valued representation belongs to the UIR of $SU(1, 1)$, the covering group of $O(2, 1)$.

From Bargmann's paper,[14] one has the double-valued representations of $SU(1, 1)$ for continuous series,

$$\mathrm{Re}\, j = 0, \quad \nu, \mu = \pm\tfrac{1}{2}, \pm\tfrac{3}{2}, \cdots, \tag{4'}$$

for positive-discrete series

$$\begin{aligned} &j = 1, 2, \cdots, \\ &\nu, \mu = j + \tfrac{1}{2}, j + \tfrac{3}{2}, \cdots, \end{aligned} \tag{5'}$$

and, for negative-discrete series,

$$\begin{aligned} &j = 1, 2, \cdots, \\ &\nu, \mu = -j - \tfrac{1}{2}, -j - \tfrac{3}{2}, \cdots. \end{aligned} \tag{6'}$$

We observe several facts: (a) The representation function $d_{\nu\mu}^{\,j}(z)$ defined in (1) and the G function defined in (18) are products of γ functions, hypergeometric functions, or generalized hypergeometric functions, the argument of which are quantities like $2j_i$, $\frac{1}{2} \pm j_i \pm \nu_i$, $\frac{1}{2} \pm j_1 \pm j_2 \pm j_3$, $1 \pm j_i \pm j_j \pm \nu_k$, etc. These quantities behave as integers or zeros when the corresponding j_i are in the discrete UIR's, whether they are single valued or double valued; (b) The poles and the zeros, if any, of a γ function, a hypergeometric function, or a generalized hypergeometric function occur only when its arguments take negative integral values or zeros; (c) The function $K(j, \nu)$ in the G functions is a phase factor even though some of the γ functions in it have arguments different from the quantities mentioned above. In other words, the function $K(j, \nu)$ does not contribute to the pole structure of the G function. From these four facts, one sees that one can derive the CG series and thus the CG coefficients of $SU(1, 1)$ in the same way as those of $O(2, 1)$. Only one change must be made: The range of the positive- (or negative-) discrete spectrum of j_3, if j_3 is integral, is changed to begin from 1 (or -1) instead of $\frac{1}{2}$ (or $-\frac{1}{2}$). Therefore, our results are valid for the group $SU(1, 1)$ also.

VI. UNITARY REPRESENTATION OF THE $O(2, 2)$ GROUP

In this section, we use the explicit expression of the CG coefficient to express the UIR of the $O(2, 2)$ group.

Let us define J_i and K_i as the infinitesimal generators of $O(2, 2)$, which keeps invariant the quadratic form $x_0^2 + x_3^2 - x_1^2 - x_2^2$. They satisfy the following commutation relations:

$$\begin{aligned} &[J_2, J_3] = iJ_1, \quad &&[K_2, K_3] = iJ_1, \\ &[J_3, J_1] = -iJ_2, \quad &&[K_3, K_1] = iJ_2, \\ &[J_1, J_2] = -iJ_3, \quad &&[K_1, K_2] = -iJ_3, \\ &[J_2, K_3] = iK_1, \quad &&[K_2, J_3] = iK_1, \\ &[J_3, K_1] = iK_2, \quad &&[K_3, J_1] = iK_2, \\ &[J_1, K_2] = -iK_3, \quad &&[K_1, J_2] = -iK_3. \end{aligned} \tag{75}$$

We can easily see that any three noncommuting infinitesimal generators form the Lie algebra of an $O(2, 1)$ group. The commutation relations in (75) are somewhat complicated. However, we may obtain simpler commutation by introducing new Lie algebra as linear combinations of the J_i and the K_i. Defining

$$A_i = \tfrac{1}{2}(J_i + K_i)$$

and

$$B_i = \tfrac{1}{2}(J_i - K_i), \quad \text{for} \quad i = 1, 2, 3, \tag{76}$$

one has, from (76) and (75),

$$\begin{aligned} &[A_2, A_3] = iA_1, \quad &&[B_2, B_3] = iB_1, \\ &[A_3, A_1] = iA_2, \quad &&[B_3, B_1] = iB_2, \\ &[A_1, A_2] = -iA_3, \quad &&[B_1, B_2] = -iB_3, \end{aligned} \tag{77}$$

and

$$[A_i, B_j] = 0, \quad \text{for} \quad i, j = 1, 2, 3.$$

The generators A_i and B_i separately form a Lie algebra of $O(2, 1)$. In other words, $O(2, 2)$ is the product group of two $O(2, 1)$ groups, i.e.,

$$O(2, 2) = O(2, 1) \otimes O(2, 1). \tag{78}$$

The group $O(4)$ has similar structure, i.e., $O(4) = O(3) \otimes O(3)$. This similarity is one of the reasons that lead to the conjecture that the UIR of $O(4)$ is a continuation of that of $O(2, 2)$, as well as that the UIR of $O(3)$ is that of $O(2, 1)$.

The product of the two UIR's of $O(2, 1)$ is a UIR of $O(2, 2)$. In this representation, the basis vectors $|a, b; \lambda_a, \lambda_b\rangle$ are eigenvector of A_3, B_3, and the two Casimir operators $\mathbf{A}^2$ and $\mathbf{B}^2$, where $\mathbf{A}^2$ and $\mathbf{B}^2$, are defined as

$$\mathbf{A}^2 = A_1^2 + A_2^2 - A_3^2$$

and

$$\mathbf{B}^2 = B_1^2 + B_2^2 - B_3^2. \tag{79}$$

The Casimir operators $\mathbf{A}^2$ and $\mathbf{B}^2$ have the eigenvalues $-(a+\frac{1}{2})(a-\frac{1}{2})$ and $-(b+\frac{1}{2})(b-\frac{1}{2})$. The physical interpretation of A_3 and B_3 is not clear. The basis vector $|a, b; \lambda_a\lambda_b\rangle$ is normalized by the condition

$$\langle a', b'; \lambda_a', \lambda_b' \mid a, b; \lambda_a\lambda_b\rangle = \delta(a', a)\delta(b', b)\delta_{\lambda_a'\lambda_a}\delta_{\lambda_b'\lambda_b}. \tag{80}$$

The group element g of $O(2, 2)$ can be uniquely parameterized by

$$\begin{aligned} g &= e^{-iA_3\phi_A}e^{-iA_2\theta_A}e^{-iA_3\psi_A}e^{-iB_3\phi_B}e^{-iB_2\theta_B}e^{-iB_3\psi_B} \\ &\equiv g_A g_B, \end{aligned} \tag{81}$$

with the group parameters restricted in the domains

$$0 < \phi_A, \phi_B, \psi_A, \psi_B < 2\pi \quad \text{and} \quad 0 < \theta_A, \theta_B < \infty. \tag{82}$$

The Haar measure of the group $O(2, 2)$ for this parameterization is

$$\begin{aligned} dg &= (2\pi)^{-4}\, d\phi_A d\cosh\theta_A\, d\psi_A\, d\phi_B d\cosh\theta_B\, d\psi_B \\ &\equiv dg_A\, dg_B. \end{aligned} \tag{83}$$

The corresponding UIR $D^{ab}_{\lambda_a\mu_a\lambda_b\mu_b}(g)$ defined by the equation.

$$\begin{aligned} D^{ab}_{\lambda_a\mu_a\lambda_b\mu_b}(g) &\equiv \langle a, b; \lambda_a, \lambda_b| U(g) |a, b; \mu_a, \mu_b\rangle \\ &= D^a_{\lambda_a\mu_a}(g_A)D^b_{\lambda_b\mu_b}(g_B) \end{aligned} \tag{84}$$

satisfies the normalization and orthogonality condition

$$\begin{aligned} &\int [D^{ab}_{\lambda_a\mu_a\lambda_b\mu_b}(g)]^* D^{a'b'}_{\lambda_a'\mu_a'\lambda_b'\mu_b'}(g)\, dg \\ &\quad = \eta(a)^{-1}\eta(b)^{-1}\delta(a', a)\delta(b', b)\delta_{\lambda_a'\lambda_a}\delta_{\lambda_b'\lambda_b}\delta_{\mu_a'\mu_a}\delta_{\mu_b'\mu_b}. \end{aligned} \tag{85}$$

We are particularly interested here in the UIR of $O(2, 2)$ whose basis vector $|a, b; j, \lambda\rangle$ is an eigenvector of $\mathbf{J}^2, J_3, \mathbf{A}^2$, and $\mathbf{B}^2$. The corresponding parameterization of the group element g' can be uniquely expressed as

$$\begin{aligned} g' &= e^{-i\phi J_3}e^{-i\beta K_3}e^{-i\theta J_2}e^{-i\alpha K_2}e^{-i\psi J_3}e^{-i\gamma K_3} \\ &\equiv u_z(\phi)a_z(\beta)u_y(\theta)a_y(\alpha)u_z(\psi)a_z(\gamma), \end{aligned} \tag{86}$$

where the new parameters ϕ, θ, ψ, α, β, and γ are related to the old ones by

$$\phi = \tfrac{1}{2}(\phi_A + \phi_B),\ \theta = \tfrac{1}{2}(\theta_A + \theta_B),\ \psi = \tfrac{1}{2}(\psi_A + \psi_B),$$

and

$$\alpha = \tfrac{1}{2}(\phi_A - \phi_B),\ \beta = \tfrac{1}{2}(\theta_A - \theta_B),\ \gamma = \tfrac{1}{2}(\psi_A - \psi_B). \tag{87}$$

From (87), one can calculate the Jacobian, which equals $\frac{1}{8}$. Hence, the Haar measure dg' for the second parameterization is equal to $\frac{1}{8}$ that of the first one. The domains of the second set of the parameters can be obtained from (82) and (87).

The basic vector $|a, b; j, \lambda\rangle$ is normalized by the condition

$$\langle a', b'; j', \lambda' \mid a, b; j, \lambda\rangle = \delta(a', a)\delta(b', b)\delta(j', j)\delta_{\lambda'\lambda}. \tag{88}$$

It is related to the basis vector $|a, b; \lambda_a, \lambda_b\rangle$ via the equation

$$\begin{aligned} &|a, b; j, \lambda\rangle \\ &\quad = \sum_{\lambda_a} C(a, b, j; \lambda_a, \lambda - \lambda_a)\, |a, b; \lambda_a, \lambda - \lambda_a\rangle, \end{aligned} \tag{89}$$

where the summation for λ_a has the same meaning as described in (14). From (12), one can obtain the inverse relation

$$|a, b; \lambda_a, \lambda_b\rangle = \sum_j C(a, b, j; \lambda_a, \lambda_b)\, |a, b; j, \lambda_a + \lambda_b\rangle, \tag{90}$$

where the meaning of the summation for j is specified in (12). From (89) and (90), one can relate the two UIR's of $O(2, 2)$ by the equation

$$\begin{aligned} D^{ab}_{\lambda_a'\lambda_a\lambda_b'\lambda_b}(g) = \sum_{jj'} &C(a, b, j'; \lambda_a', \lambda_b')D^{ab}_{j'\lambda'j\lambda}(g') \\ &\times C(a, b, j; \lambda_a, \lambda_b), \end{aligned} \tag{91}$$

where $\lambda = \lambda_a + \lambda_b$, and $\lambda' = \lambda_a' + \lambda_b'$. The UIR $D^{ab}_{j'\lambda j\lambda}(g')$ is defined by

$$D^{ab}_{j'\lambda' j\lambda}(g') = \langle a, b; j'\lambda'| U(g') |a, b; j\lambda\rangle.$$

From (91) and (14), one can derive the inverse relation of Eq. (91). By means of Eq. (91) or its inverse relation, one can obtain the orthogonal relation for $D^{ab}_{j'\lambda' j\lambda}(g')$ from that of $D^{ab}_{\lambda_a'\lambda_b'\lambda_a\lambda_b}(g)$. That is,

$$\begin{aligned} &\int D^{ab}_{j\lambda k\mu}(g')[D^{a'b'}_{j'\lambda'k'\mu'}(g')]^*\, dg' \\ &\quad = \eta(a)^{-1}\eta(b)^{-1}\delta(a', a)\delta(b', b)\delta(j', j)\delta(k', k)\delta_{\lambda'\lambda}\delta_{\mu'\mu}. \end{aligned}$$

From the properties of the CG coefficients and the representation function $D^{ab}_{\lambda_a\mu_a\lambda_b\mu_b}(g)$, one can calculate the orthogonality relations for the representation function of the subgroups of $O(2, 2)$. From (87), (89), and (90), one can express $D^{ab}_{j'\lambda' j\lambda}(g)$ in terms of the representation functions of its 1-parameter subgroups:

$$\begin{aligned} D^{ab}_{j'\lambda' j\lambda}(g') = \sum_{kk'\mu} &e^{-i\phi\lambda'}D^{ab}_{j\lambda' k}(a_z(\alpha))d^k_{\lambda'\mu}(\theta) \\ &\times D^{ab}_{k\mu k'\lambda}(a_y(\beta))D^{ab}_{k'\lambda g}(a_z(\gamma))e^{-i\psi\lambda}. \end{aligned}$$

The representation functions

$$D^{ab}_{j'\lambda j}(a_z(\alpha)) \quad \text{and} \quad D^{ab}_{k\mu k'\lambda}(a_y(\beta))$$

can be explicitly calculated by means of Eqs. (80), (89), and (90). The expression for $D^{ab}_{j'\lambda j}(a_z(\alpha))$ is particularly simple, i.e.,

$$D^{ab}_{j'\lambda j}(a_z(\alpha)) = \sum_{\mu} c(a, b, j'; \mu, \lambda - \mu)e^{i(\lambda-2\mu)\alpha}c^*(a, b, j; \mu, \lambda - \mu),$$

which has the same form as that of $O(4)$.

VII. CONCLUSION

The CG coefficient of $O(2, 1)$ and $SU(1, 1)$ defined in this paper, when it is continued in the j_i into the domain corresponding to the $O(3)$ group, is equal to that[12,13] of $O(3)$ except for a phase factor. Strictly speaking, the Wigner coefficient[12,23] defined by

$$W(j_1, j_2, j_3; \nu_1, \nu_2) = [\eta(j_3)]^{-\frac{1}{2}}c(j_1, j_2, j_3; \nu_1, \nu_2)$$

is a continuable quantity, rather than the CG coefficient, since the Plancherel measures of $O(2, 1)$ and $SU(1, 1)$ for the discrete and continuous series differ by a factor $\tan \pi(j_3 - \mu_3)$.

The general continuable expressions of the two CG coefficients are defined in (69). The simple expressions for particular cases are defined in (61), (65), and (67). The j_3 spectra are tabulated in Tables VI–IX. The pole structures of the related G functions for some or all of the j_i in the discrete series are collected in Tables II, III, IV, and X.

As a final remark, one notes that our $SU(1, 1)$ representation functions $d_{\nu\mu}{}^{j}(z)$ and $a_{\nu\mu}{}^{j}(z)$ of the first and the second kinds are related to Andrews and Gunson's $d_j{}^{\nu\mu}(z)$ and $e_j{}^{\nu\mu}(z)$ by the equations[5]

$$d_j{}^{\nu\mu}(z) = e^{i\pi(\nu-\mu)}d_{\nu\mu}{}^{j}(z)$$

and

$$e_j{}^{\nu\mu}(z) = \pi e^{-i\pi(\mu-\nu)} \cot \pi(\tfrac{1}{2} + j - \mu)a_{\nu\mu}{}^{j}(z).$$

ACKNOWLEDGMENT

I am grateful to Professor S. Mandelstam for his constant guidance and encouragement throughout the development of this work.

* This work was supported in part by the U.S. Atomic Energy Commission.

† Present address: Department of Physics, Pennsylvania State University, University Park, Pa. 16802.

[1] H. Joos, *Lecture in Theoretical Physics* W. E. Brittin and A. O. Barut, Eds. (U. of Colorado, Boulder, 1965), Vol. 8, p. 132.

[2] M. Toller, Nuovo Cimento **53A**, 672 (1968); **56A**, 295 (1968). Most of the earlier references are cited in these two papers.

[3] F. Chew, *S-Matrix Theory of Strong Interactions* (W. A. Benjamin, Inc., New York, 1962).

[4] Sho Tanaka, Progr. Theoret. Phys. (Kyoto) Suppl. **41**, 194 (1968).

[5] M. Andrews and J. Gunson, J. Math. Phys. **5**, 1391 (1964).

[6] S. S. Sannikov, Dokl. Akad. Nauk SSSR **7**, 1058 (1966) [Sov. Phys. Dokl. **11**, 1045 (1967)].

[7] L. Pukanszky, Trans. Am. Math. Soc. **100**, 116 (1961).

[8] I. Ferretti and M. Verde, Nuovo Cimento **55**, 110 (1968).

[9] W. J. Holman III and L. C. Biedenharn, Jr., Ann. Phys. (N.Y.) **47**, 205 (1968); **39**, 1 (1966).

[10] D. Z. Freedman and J. M. Wang, Phys. Rev. **160**, 1560 (1967).

[11] The $O(2, 2)$ symmetry of the helicity amplitude is obtained if one restricts oneself to the c.m. frame. We shall discuss this in detail in a later paper.

[12] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Chap. 27.

[13] For example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. Press, Princeton, N.J., 1957), Sec. 2.7.

[14] V. Bargmann, Ann. Math. **48**, 568 (1947).

[15] J. L. Burchnall and T. W. Chaundy, Proc. London Math. Soc. **50**, 56 (1944).

[16] W. N. Bailey, *Generalized Hypergeometric Series* (Cambridge U. Press, Cambridge, England, 1935), Sec. 2.1, p. 8.

[17] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. Press, Cambridge, England, 1966), Sec. 2.1.1, p. 40.

[18] See Ref. 16, p. 17; Ref. 14, pp. 115–120.

[19] See Ref. 17, p. 108, Eq. (4.1.2).

[20] *Higher-Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill, New York, 1953), Vol. 1, p. 77, Eq. (14).

[21] W. N. Bailey, Quart. J. Math. (Oxford) **8**, 115 (1937); Proc. Glasgow Math. Assoc. **2**, 62 (1954).

[22] Nobuhiko Tatsuuma, J. Math. Soc. Japan **17**, 313 (1965).

[23] These Wigner coefficients are different from those defined in Ref. 10.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Self-Interaction of Gravitational Radiation*

W. E. COUCH†
Department of Physics, University of Texas, Austin, Texas 78712
AND
R. J. TORRENCE†
Department of Mathematics, University of Calgary, Calgary, Alberta, Canada

(Received 14 July 1969)

A nonsingular, sourceless, first-order pulse of gravitational radiation imploding from infinity to a focus and then exploding back out to infinity is examined to second order. It is found that, contrary to what might be expected, the nonsingular second-order field contains no radiation. In space–time regions outside of the pulse, the second-order field is nonvanishing only in the region with retarded times earlier than the passing of the exploding pulse and advanced times later than the passing of the imploding pulse. In this region, the second-order field is that of a Schwarzschild mass, plus a nonradiative quadrupole, plus a nonradiative sixteen pole.

1. INTRODUCTION

The field equations of general relativity have often been studied using a perturbative approximation scheme. If one chooses coordinates properly, and assumes[1a] that

$$g_{\mu\nu} = \eta_{\mu\nu} + \epsilon \overset{1}{h}_{\mu\nu} + \epsilon^2 \overset{2}{h}_{\mu\nu} + \cdots, \tag{1.1}$$

substitution of this metric-tensor expansion[1b] into $G_{\mu\nu} = 0$ leads to[2]

$$\begin{aligned} \Box \overset{1}{\gamma}_{\mu\nu} &= 0, \\ \Box \overset{2}{\gamma}_{\mu\nu} &= \overset{2}{\delta}_{\mu\nu}, \\ \Box \overset{3}{\gamma}_{\mu\nu} &= \overset{3}{\delta}_{\mu\nu}, \\ &\cdot \\ &\cdot \\ &\cdot \\ \Box \overset{k}{\gamma}_{\mu\nu} &= \overset{k}{\delta}_{\mu\nu}, \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned} \tag{1.2}$$

where

$$\overset{k}{\gamma}_{\mu\nu} \equiv \overset{k}{h}_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}\overset{k}{h}_{\rho\sigma}, \tag{1.3}$$

$\Box$ is the d'Alembertian,

$$\Box \equiv \eta^{\mu\nu}\partial^2/\partial x^\mu \partial x^\nu,$$

and the $\overset{k}{\delta}_{\mu\nu}$ are known functions of the lower-order quantities

$$\overset{k'}{\gamma}_{\mu\nu}, k' = 1, 2, \cdots, k-1.$$

Thus, the solving of Einstein's vacuum equations in a perturbative expansion can begin with the solving of homogeneous wave equations, and then continue, at each higher order, with the solving of in homogeneous wave equations. In particular, it is easy to see that in any region of space–time where

$$\overset{1}{\gamma}_{\mu\nu} = 0, \tag{1.4}$$

we have ${}_{(2)}\delta_{\mu\nu} = 0$, and consequently

$$\Box \overset{2}{\gamma}_{\mu\nu} = 0. \tag{1.5}$$

In such a region, the ${}_{(2)}\gamma_{\mu\nu}$ may be interpreted in terms appropriate to a field in Minkowski space which satisfies the homogeneous wave equation (1.5). This analysis is always applicable to the lowest-order nonvanishing ${}_{(k)}\gamma_{\mu\nu}$ in a region, and such fields we call "formally linear," whatever their order. In this paper, we consider certain perturbative solutions of Einstein's equations, and although we shall be using null coordinates, working in a null-tetrad formalism, and treating the curvature tensor as the fundamental field rather than the metric tensor, the above remarks and definition of "formally linear" remain applicable.

The first-order solution to the homogeneous wave equation, with which we begin our perturbative expansion, is a spherical wave of gravitational quadrupole radiation imploding from past null infinity to a focus, and then exploding back out to future null infinity. Such a first-order field can be shown to be nonsingular everywhere, including at the focus of the wave. We wish in particular to consider our spherical wave to be a pulse of radiation, that is, a wave of finite temporal duration, and we sketch this first-order field on a Penrose diagram[3] in Fig. 1. The diagram represents a conformally compactified Minkowski space with I^-, I^o, and I^+ labeling past temporal, spatial, and future temporal infinity, respectively, and $\mathfrak{I}^-$ and $\mathfrak{I}^+$ labeling past and future null infinity, respectively. Two spatial dimensions have been suppressed, and points, lines, and regions bearing the same label are in fact connected, although

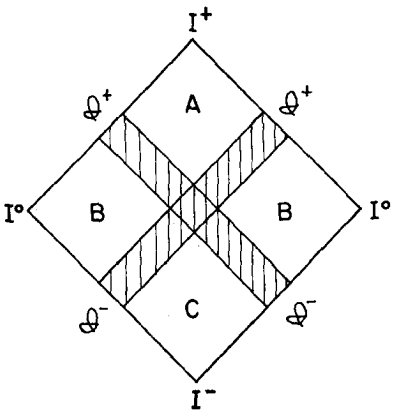


FIG. 1. The first-order field.

they appear disconnected in the diagram. Since the compactification is conformal, straight lines at a 45° angle to the horizontal represent null hypersurfaces, and the pulse of radiation results in a nonvanishing gravitational field in the region indicated by the cross hatching in Fig. 1.

The first-order solution pictured in Fig. 1 does not, of course, uniquely determine its second-order correction; one must also supply second-order initial data or boundary values. We fix the problem by making two demands. First, we require that the second-order correction is also nonsingular. The use in second order of Einstein's vacuum equations, combined with this requirement, assures us that we still have no sources present. Second, we demand that there is no second-order radiation on $\mathfrak{I}^-$, that is, that there is nothing coming in from past null infinity. By adding this requirement, we insure that any second-order radiation found at future null infinity ($\mathfrak{I}^+$) has been generated by the self-interaction of the first-order field. If we do not find radiation on $\mathfrak{I}^+$, then presumably none was generated, since there are no sources present that could have absorbed it. It is not obvious that the first demand can be satisfied, but we find that in fact it can be, and that these two demands then uniquely determine the second-order correction. The situation is depicted in Fig. 2. We then calculate the second-order Riemann tensor, and with its aid we are able to answer several interesting questions:

(1) Is there second-order radiation arriving at future null infinity ($\mathfrak{I}^+$) at points bordering regions B or A? If so, we must think of it as originating and propagating as indicated by the arrows labeled y in Fig. 2; this is because we are solving for the retarded solution of an inhomogeneous wave equation, and radiation can only arise in the region of support of the inhomogeneity and it can only propagate along the characteristics of the hyperbolic equation. Such second-order radiation might be said to originate by "backward scattering," since it propagates in a direction opposite to the propagation of the first-order wave which is its source.

(2) Is there second-order radiation arriving at future null infinity ($\mathfrak{I}^+$) at points bordering the cross-hatched region of Fig. 2? If so, it might be back-scattered radiation. If, however, we find in answer to question (1) that back scattering does not occur, we must conclude that any radiation found in answer to question (2) originated and propagated as indicated by either the arrow labeled x_1 or the arrow labeled x_2. In the first case, we might call the radiation "forward scattered"; in the second case, we might call it "mixed scattered," since it depended on the simultaneous presence of incoming and outgoing first-order radiation. Some problem might arise in distinguishing between the forward scattered and mixed scattered radiation, but we see in Sec. 6 that it does not.

(3) Regardless of the presence or absence of scattered radiation at $\mathfrak{I}^+$, one may simply ask whether there are nonvanishing second-order fields in any of regions A, B, or C, and, if so, of what type.

The answers to these questions are given in the concluding section of the paper.

In the second section, we outline the null tetrad formalism used to do the calculations. In Sec. 3, we derive the relations needed to do perturbation theory in the null tetrad formalism. In Secs. 4 and 5, we discuss the first-order solutions and the second-order corrections, respectively. In each of these two latter sections, the quantities of interest are calculated in part A, and in part B, they are examined to make sure that they are nonsingular as desired. Finally, the complicated expressions of Sec. 5 are analyzed in the concluding section, in order to answer the three questions which we have posed.

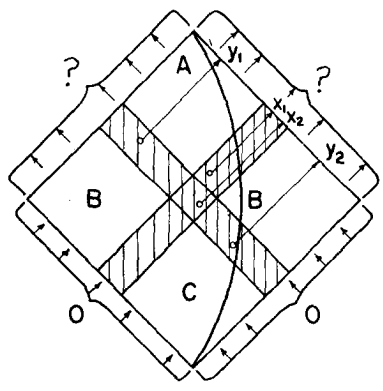


FIG. 2. The second-order field.

2. NULL-TETRAD FORMALISM

We use throughout this paper the null-tetrad formalism introduced by Newman and Penrose,[4,5] and a brief review of it is given in this section.

Consider a tetrad of basis vectors l_μ, n_μ, m_μ, and $\bar{m}_\mu$ in the 4-dimensional Riemannian space satisfying

$$l^\mu n_\mu = -m^\mu \bar{m}_\mu = 1,$$
$$l^\mu l_\mu = n^\mu n_\mu = m^\mu m_\mu = l^\mu m_\mu = n^\mu m_\mu = 0. \quad (2.1)$$

It follows from Eq. (2.1) that the metric is given in terms of the tetrad by

$$g_{\mu\nu} = l_\mu n_\nu + l_\nu n_\mu - m_\mu \bar{m}_\nu - m_\nu \bar{m}_\mu. \quad (2.2)$$

After certain definitions are made, a set of partial differential equations are derived in Ref. 4 which are equivalent to the Einstein field equations. The needed definitions are: intrinsic derivatives,

$$D\Phi \equiv \Phi_{;\mu} l^\mu, \quad \Delta\Phi \equiv \Phi_{;\mu} n^\mu, \quad \delta\Phi \equiv \Phi_{;\mu} m^\mu,$$
$$\bar{\delta}\Phi \equiv \Phi_{;\mu} \bar{m}^\mu, \quad (2.3)$$

for an arbitrary scalar Φ; combinations of Ricci rotation coefficients (called spin coefficients),

$$\kappa \equiv l_{\mu;\nu} m^\mu l^\nu, \quad (2.4a)$$
$$\pi \equiv -n_{\mu;\nu} \bar{m}^\mu l^\nu, \quad (2.4b)$$
$$\epsilon \equiv \tfrac{1}{2}(l_{\mu;\nu} n^\mu l^\nu - m_{\mu;\nu} \bar{m}^\mu l^\nu), \quad (2.4c)$$
$$\rho \equiv l_{\mu;\nu} m^\mu \bar{m}^\nu, \quad (2.4d)$$
$$\lambda \equiv -n_{\mu;\nu} \bar{m}^\mu \bar{m}^\nu, \quad (2.4e)$$
$$\alpha \equiv \tfrac{1}{2}(l_{\mu;\nu} n^\mu \bar{m}^\nu - m_{\mu;\nu} \bar{m}^\mu \bar{m}^\nu), \quad (2.4f)$$
$$\sigma \equiv l_{\mu;\nu} m^\mu m^\nu, \quad (2.4g)$$
$$\mu \equiv -n_{\mu;\nu} \bar{m}^\mu m^\nu, \quad (2.4h)$$
$$\beta \equiv \tfrac{1}{2}(l_{\mu;\nu} n^\mu m^\nu - m_{\mu;\nu} \bar{m}^\mu m^\nu), \quad (2.4i)$$
$$\nu \equiv -n_{\mu;\nu} \bar{m}^\mu n^\nu, \quad (2.4j)$$
$$\gamma \equiv \tfrac{1}{2}(l_{\mu;\nu} n^\mu n^\nu - m_{\mu;\nu} \bar{m}^\mu n^\nu), \quad (2.4k)$$
$$\tau \equiv l_{\mu;\nu} m^\mu n^\nu; \quad (2.4l)$$

and the independent components of the Weyl tensor,

$$\Psi_0 \equiv -C_{\mu\nu\rho\sigma} l^\mu m^\nu l^\rho m^\sigma, \quad (2.5a)$$
$$\Psi_1 \equiv -C_{\mu\nu\rho\sigma} l^\mu n^\nu l^\rho m^\sigma, \quad (2.5b)$$
$$\Psi_2 \equiv -C_{\mu\nu\rho\sigma} \bar{m}^\mu n^\nu l^\rho m^\sigma, \quad (2.5c)$$
$$\Psi_3 \equiv -C_{\mu\nu\rho\sigma} \bar{m}^\mu n^\nu l^\rho n^\sigma, \quad (2.5d)$$
$$\Psi_4 \equiv -C_{\mu\nu\rho\sigma} \bar{m}^\mu n^\nu \bar{m}^\rho n^\sigma. \quad (2.5e)$$

Before writing down the field equations, we impose (without loss of generality) several simplifying coordinate and tetrad conditions. A parameter u which labels null hypersurfaces of the hyperbolic Riemannian space by $u = \text{const}$ is taken as the timelike coordinate, i.e., $x^0 = u$. The null-tetrad vector l_μ is chosen as[6]

$$l_\mu = -u_{,\mu} = -\delta^0_\mu, \quad (2.6)$$

so that it is normal to the null hypersurfaces, $u = \text{const}$. It is then also geodesic. Let $x^1 = r$ denote the affine parameter along the null geodesics lying in the null surfaces, and let the x^i label the different geodesics on each hypersurface. Then we have $l^\mu = dx^\mu/dr = \delta^\mu_1$. The relation $g^{\mu\nu} l_\nu = l^\mu$ implies[6]

$$g^{01} = -1 \quad (2.7)$$

and $g^{00} = g^{0i} = 0$. This form of l^μ produces the following conditions on the spin-coefficients:

$\kappa = 0$, l^μ is geodesic,
$\epsilon + \bar{\epsilon} = 0$, l^μ in terms of an affine parameter,
$\rho = \bar{\rho}$, l^μ is hypersurface orthogonal,
$\tau = \bar{\alpha} + \beta$, l^μ is equal to a gradient.

We also parallelly propagate n_μ along l_μ, which gives $\pi = 0$, and m_μ and $\bar{m}_\mu$ along l_μ, which gives $\epsilon - \bar{\epsilon} = 0$.

In compliance with Eqs. (2.1), the vectors n^μ, m^μ, and $\bar{m}^\mu$ take the form

$$n^\mu = -\delta^\mu_0 + U\delta^\mu_1 + X^i \delta^\mu_i, \quad (2.8a)$$
$$m^\mu = \omega\delta^\mu_1 + \xi^i \delta^\mu_i, \quad (2.8b)$$

and the metric is then given by

$$g^{11} = 2(U - \omega\bar{\omega}), \quad (2.9a)$$
$$g^{1i} = X^i - (\xi^i \bar{\omega} + \bar{\xi}^i \omega), \quad (2.9b)$$
$$g^{ij} = -(\xi^i \bar{\xi}^j + \bar{\xi}^i \xi^j). \quad (2.9c)$$

In terms of the quantities introduced in Eqs. (2.8), the operators defined by Eqs. (2.3) become

$$D = \frac{\partial}{\partial r}, \quad \delta = \omega \frac{\partial}{\partial r} + \xi^i \frac{\partial}{\partial x^i},$$
$$\Delta = U \frac{\partial}{\partial r} - \frac{\partial}{\partial u} + X^i \frac{\partial}{\partial x^i}. \quad (2.10)$$

We call the set of quantities defined by Eqs. (2.4), (2.5), and (2.8) tetrad-formalism (TF) variables.

The empty-space field equations may now be written as[4]

$$D\xi^i = \rho\xi^i + \sigma\bar{\xi}^i, \quad (2.11a)$$
$$D\omega = \rho\omega + \sigma\bar{\omega} - \tau, \quad (2.11b)$$
$$DX^i = \bar{\tau}\xi^i + \tau\bar{\xi}^i, \quad (2.11c)$$
$$DU = \bar{\tau}\omega + \tau\bar{\omega} - (\gamma + \bar{\gamma}), \quad (2.11d)$$

$$D\rho = \rho^2 + \sigma\bar{\sigma}, \tag{2.12a}$$
$$D\sigma = 2\rho\sigma + \Psi_0, \tag{2.12b}$$
$$D\tau = \rho\tau + \sigma\bar{\tau} + \Psi_1, \tag{2.12c}$$
$$D\alpha = \rho\alpha + \bar{\sigma}\beta, \tag{2.12d}$$
$$D\beta = \rho\beta + \sigma\alpha + \Psi_1, \tag{2.12e}$$
$$D\gamma = \tau\alpha + \bar{\tau}\beta + \Psi_2, \tag{2.12f}$$
$$D\lambda = \rho\lambda + \bar{\sigma}\mu, \tag{2.12g}$$
$$D\mu = \rho\mu + \sigma\lambda + \Psi_2, \tag{2.12h}$$
$$D\nu = \tau\lambda + \bar{\tau}\mu + \Psi_3, \tag{2.12i}$$

$$\delta X^i - \Delta\xi^i = (\mu + \bar{\gamma} - \gamma)\xi^i + \bar{\lambda}\bar{\xi}^i, \tag{2.13a}$$
$$\delta\bar{\xi}^i - \bar{\delta}\xi^i = (\bar{\beta} - \alpha)\xi^i - (\beta - \bar{\alpha})\bar{\xi}^i, \tag{2.13b}$$
$$\delta\bar{\omega} - \bar{\delta}\omega = (\bar{\beta} - \alpha)\omega + (\bar{\alpha} - \beta)\bar{\omega} + \mu - \bar{\mu}, \tag{2.13c}$$
$$\delta U - \Delta\omega = (\mu + \bar{\gamma} - \gamma)\omega + \bar{\lambda}\bar{\omega} - \bar{\nu}, \tag{2.13d}$$

$$\Delta\lambda - \bar{\delta}\nu = 2\alpha\nu + (\bar{\gamma} - 3\gamma - \mu - \bar{\mu})\lambda - \Psi_4, \tag{2.14a}$$
$$\delta\rho - \bar{\delta}\sigma = \tau\rho + (\bar{\beta} - 3\alpha)\sigma - \Psi_1, \tag{2.14b}$$
$$\delta\alpha - \bar{\delta}\beta = \mu\rho - \lambda\sigma - 2\alpha\beta + \alpha\bar{\alpha} + \beta\bar{\beta} - \Psi_2, \tag{2.14c}$$
$$\delta\lambda - \bar{\delta}\mu = \bar{\tau}\mu + (\bar{\alpha} - 3\beta)\lambda - \Psi_3, \tag{2.14d}$$
$$\delta\nu - \Delta\mu = \gamma\mu - 2\beta\nu + \bar{\gamma}\mu + \mu^2 + \lambda\bar{\lambda}, \tag{2.14e}$$
$$\delta\gamma - \Delta\beta = \tau\mu - \sigma\nu + (\mu - \gamma + \bar{\gamma})\beta + \alpha\bar{\lambda}, \tag{2.14f}$$
$$\delta\tau - \Delta\sigma = 2\tau\beta + (\bar{\gamma} - 3\gamma + \mu)\sigma + \bar{\lambda}\rho, \tag{2.14g}$$
$$\Delta\rho - \bar{\delta}\tau = (\gamma + \bar{\gamma} - \bar{\mu})\rho - 2\alpha\tau - \lambda\sigma - \Psi_2, \tag{2.14h}$$
$$\Delta\alpha - \bar{\delta}\gamma = \rho\nu - \tau\lambda - \beta\lambda + (\bar{\gamma} - \gamma - \bar{\mu})\alpha - \Psi_3. \tag{2.14i}$$

Equations (2.11) and (2.12) are called radial equations and Eqs. (2.13) and (2.14) are called nonradial equations. The vacuum Bianchi identities in this tetrad formalism are

$$D\Psi_{A+1} - \bar{\delta}\Psi_A = (4 - A)\rho\Psi_{A+1} - 2(2 - A)\alpha\Psi_A - A\lambda\Psi_{A-1}, \tag{2.15}$$

$$\Delta\Psi_A - \delta\Psi_{A+1} = A\nu\Psi_{A-1} - 2(A - 2)\gamma\Psi_A + (A - 4)\tau\Psi_{A+1} - (A + 1)\mu\Psi_A + 2(A - 1)\beta\Psi_{A+1} - (A - 3)\sigma\Psi_{A+2}, \tag{2.16}$$

where $A = 0, 1, 2, 3$.

3. PERTURBATION THEORY

In this section, the structure of the field equations is examined in perturbation theory. We assume that each TF variable is expandable in a small parameter, and we write, for example, $\rho = {}_{(0)}\rho + {}_{(1)}\rho + {}_{(2)}\rho \cdots$, where the subscript "zero" denotes flat space. For ease of writing, in the remainder of this section, we omit the designation of the perturbative order from those TF variables which are the order indicated by the title of the subsection in which they occur. In later sections, we often omit the perturbative order designation from a quantity when its order is clear from the context.

A. Zeroth-Order Field

The zeroth-order field is simply flat space–time, and in null-spherical polar coordinates it is given by

$$\begin{aligned} &\Psi_A = 0, \\ &r\mu = U = -1, \quad \rho = -r^{-1}, \\ &\alpha = -\beta = -(2r)^{-1}\cot\theta, \\ &\xi^i = r^{-1}(1, i/\sin\theta), \end{aligned} \tag{3.1}$$

with all other TF variables equal to zero.

B. First-Order Field

The first-order field has been studied before in this formalism,[7-9] and we only sketch the essential points here. We first use the linearized form of the Bianchi identities, Eqs. (2.15) and (2.16), as field equations for the Ψ_A. These equations are

$$\dot{\Psi}_A + D\Psi_A + \frac{A+1}{r}\Psi_A - \frac{1}{r}\delta\Psi_{A+1} = 0, \tag{3.2}$$

$$D\Psi_{A+1} + \frac{4-A}{r}\Psi_{A+1} + \frac{1}{r}\bar{\delta}\Psi_A = 0, \tag{3.3}$$

where $A = 0, 1, 2, 3$, the dot denotes $\partial/\partial u$, and δ and $\bar{\delta}$ are angular differential operators.[10]

The assumption that $\Psi_0 = O(r^{-5})$[11] as $r \to \infty$, which insures us that our space is asymptotically flat at future null infinity, allows us the following solutions of Eq. (3.3):

$$\Psi_{A+1} = \frac{\Psi^\circ_{A+1}}{r^{4-A}} - \frac{1}{r^{4-A}}\int_\infty^r r'^{3-A}\bar{\delta}\Psi_A\, dr', \tag{3.4}$$

where the Ψ°_{A+1}, $A = 0, 1, 2, 3$, are arbitrary functions of u, θ, and ϕ. Substituting Eq. (3.4) into Eqs. (3.2), we find that

$$\begin{aligned} \dot{\Psi}^\circ_3 &= \delta\Psi^\circ_4, \\ \dot{\Psi}^\circ_2 &= \delta\Psi^\circ_3, \\ \dot{\Psi}^\circ_1 &= \delta\Psi^\circ_2, \end{aligned} \tag{3.5}$$

and

$$\Psi'_0 + D\Psi'_0 + \frac{\Psi'_0}{r} + \frac{1}{r^5}\int_\infty^r r'^3 \delta\bar{\delta}\Psi'_0\, dr' - \frac{\delta\Psi'^\circ_1}{r^5} = 0. \tag{3.6}$$

If we are given $\Psi'^\circ_4(u, \theta, \phi)$ and Ψ'°_3, Ψ'°_2 and Ψ'°_1 at a particular time u_0, solving Eq. (3.6) leads to a solution Ψ'_A of Eqs. (3.2) and (3.3).

To complete our discussion of the first-order calculation, it is necessary to find the rest of the TF variables. They are obtained by linearizing Eqs. (2.11) and (2.12) and integrating them. The results are

$$\rho = 0, \tag{3.7a}$$

$$\sigma = \frac{\sigma^\circ}{r^2} + \frac{1}{r^2}\int_\infty^r r'^2\Psi'_0\, dr', \tag{3.7b}$$

$$\alpha = -\overset{0}{\alpha}\int_\infty^r \bar{\sigma}\, dr', \tag{3.7c}$$

$$\beta = -\bar{\alpha} + \frac{1}{r}\int_\infty^r r'\Psi'_1\, dr', \tag{3.7d}$$

$$\tau = \bar{\alpha} + \beta, \tag{3.7e}$$

$$\gamma = -r\overset{0}{\alpha}\int_\infty^r \frac{1}{r'}(\bar{\tau} - \tau)\, dr' + \int_\infty^r \Psi'_2\, dr', \tag{3.7f}$$

$$\mu = \frac{1}{r}\int_\infty^r r'\Psi'_2\, dr', \tag{3.7g}$$

$$\lambda = -\frac{1}{r}\int_\infty^r \bar{\sigma}\, dr' + \frac{\lambda^\circ}{r}, \tag{3.7h}$$

$$\nu = -\int_\infty^r \frac{\bar{\tau}}{r'}\, dr' + \int_\infty^r \Psi'_3\, dr', \tag{3.7i}$$

$$U = -\int_\infty^r (\gamma + \bar{\gamma})\, dr', \tag{3.7j}$$

$$\omega = -\frac{1}{r}\int_\infty^r r'\tau\, dr' + \frac{\omega^\circ}{r}, \tag{3.7k}$$

$$X^i = r\overset{0}{\xi}{}^i\int_\infty^r \frac{\tau}{r'}\, dr' + r\overset{0}{\xi}{}^i\int_\infty^r \frac{\bar{\tau}}{r'}\, dr', \tag{3.7l}$$

$$\xi^i = \overset{0}{\xi}{}^i\int_\infty^r \sigma\, dr', \tag{3.7m}$$

where σ°, λ°, and ω° are functions independent of r arising from the integration of Eqs. (2.11) and (2.12). The remaining field equations, Eqs. (2.13) and (2.14), are then satisfied if

$$\lambda^\circ = -\dot{\bar{\sigma}}^\circ, \tag{3.8a}$$

$$\omega^\circ = -\bar{\delta}\sigma^\circ, \tag{3.8b}$$

$$\Psi'^\circ_2 - \bar{\Psi}'^\circ_2 = \bar{\delta}^2\sigma^\circ - \delta^2\bar{\sigma}^\circ, \tag{3.8c}$$

$$\Psi'^\circ_3 = \delta\dot{\bar{\sigma}}^\circ, \tag{3.8d}$$

$$\Psi'^\circ_4 = -\ddot{\bar{\sigma}}^\circ. \tag{3.8e}$$

A statement of the first-order initial-data problem, consistent with that of the full theory,[5] can now, with the aid of Eqs. (3.8), be given. If one gives $\Psi'^\circ_2(\theta, \phi) + \bar{\Psi}'^\circ_2(\theta, \phi)$, $\Psi'^\circ_1(\theta, \phi)$, $\Psi'_0(\theta, \phi, r)$ at a particular value of u, and $\sigma^\circ(u, \theta, \phi)$ for all u, one can obtain all of the first-order TF variables. Information about the retarded and advanced radiation is contained in $\sigma^\circ(u, \theta, \phi)$ and $\Psi'_0(r, \theta, \phi)$,[9] respectively, while these quantities combine with the rest of the initial data to fix the nonradiative (including the stationary) part of the solution. To avoid angular singularities, we must take $\sigma^\circ = \sum_{l=2}^\infty \sum_{m=-l}^l \sigma^\circ_{lm2} Y_{lm}$; then Eq. (3.8c) prevents imaginary ("magnetic") monopoles from appearing.[8]

C. Second-Order Field

We now give the equations needed to find the second-order corrections to first-order solutions of Eqs. (3.2), (3.3), and (3.7). The relevant equations are those obtained by collecting all second-order terms when the small-parameter perturbation expansions are substituted into Eqs. (2.11)–(2.16). The questions which we wish to investigate can be answered from the second-order Ψ'_A alone; thus, we shall limit ourselves to the second order of the Bianchi identities, Eqs. (2.15) and (2.16), obtaining

$$\Psi'_A + D\Psi'_A + \frac{A+1}{r}\Psi'_A - \frac{1}{r}\delta\Psi'_{A+1} = D_A \tag{3.9}$$

and

$$D\Psi'_{A+1} + \frac{4-A}{r}\Psi'_{A+1} + \frac{1}{r}\bar{\delta}\Psi'_A = R_{A+1}, \tag{3.10}$$

where $A = 0, 1, 2, 3$, and

$$\begin{aligned} -D_A = {} & \left(\overset{1}{\xi}{}^i\frac{\partial}{\partial x^i} + 2(1-A)\overset{1}{\alpha} \right. \\ & \left. + 3(A-2)\overset{1}{\tau} + \overset{1}{\omega}\frac{\partial}{\partial r}\right)\overset{1}{\Psi}'_{A+1} \\ & + \left(-\overset{1}{X}{}^i\frac{\partial}{\partial x^i} + 2(2-A)\overset{1}{\gamma}\right. \\ & \left. - \overset{1}{U}\frac{\partial}{\partial r} - (A+1)\overset{1}{\mu}\right)\overset{1}{\Psi}'_A \\ & + A\overset{1}{\nu}\overset{1}{\Psi}'_{A-1} - (A-3)\overset{1}{\sigma}\overset{1}{\Psi}'_{A+2}, \end{aligned} \tag{3.11}$$

$$\begin{aligned} R_{A+1} = {} & \left(\overset{1}{\bar{\omega}}\frac{\partial}{\partial r} + \overset{1}{\bar{\xi}}{}^i\frac{\partial}{\partial x^i} - 2(2-A)\overset{1}{\alpha}\right)\overset{1}{\Psi}'_A \\ & - A\overset{1}{\lambda}\overset{1}{\Psi}'_{A-1}. \end{aligned} \tag{3.12}$$

As was indicated in the Introduction, Eqs. (3.9) and (3.10) are simply inhomogeneous versions of Eqs. (3.2) and (3.3).

We can integrate Eqs. (3.10), introducing second-order constants of integration Ψ_A°, to obtain

$$\Psi_{A+1} = \frac{\Psi^\circ_{A+1}}{r^{4-A}} - \frac{1}{r^{4-A}}\int_\infty^r r'^{3-A}(\bar\delta\Psi_A + rR_{A+1})\,dr'. \tag{3.13}$$

This is, of course, analogous to our treatment of Eqs. (3.3), and substituting Eqs. (3.13) into Eqs. (3.9) yields a second-order version of Eqs. (3.5),[7] and Eq. (3.6), which are

$$\begin{aligned}\Psi_3^\circ &= \delta\Psi_4^\circ,\\ \Psi_2^\circ &= \delta\Psi_3^\circ - \overset{1}{\sigma}{}^\circ\overset{1}{\Psi}{}_4^\circ,\\ \Psi_1^\circ &= \delta\Psi_2^\circ - 2\overset{1}{\sigma}{}^\circ\overset{1}{\Psi}{}_3^\circ,\end{aligned} \tag{3.14}$$

$$\Psi_0 + D\Psi_0 + \frac{\Psi_0}{r} + \frac{1}{r^5}\int_\infty^r r'^3\delta\bar\delta\Psi_0\,dr' - \frac{\delta\Psi_1^\circ}{r^5} = \mathcal{D}_0. \tag{3.15}$$

These equations, along with $\Psi_4^\circ = -\ddot{\bar\sigma}^\circ$, make the initial data discussion of Sec. 3A adequate to this part as well. Since we base our interpretations on the Riemann tensor alone in second order, it is not necessary to consider the second-order corrections to Eqs. (3.7); and we have now assembled everything that we need in Secs. 4 and 5.

4. FIRST-ORDER SOLUTIONS AND SINGULARITIES

A. The Solutions

Those solutions of Eqs. (3.2) and (3.3) corresponding to axially symmetric gravitational radiation are known, and are given here.[7-9] The retarded radiation fields are given by

$$\Psi_A = K_{2-A}(l)\,{}_{2-A}Y_{l0}r^{l-2}d^{l-2+A}\left(\frac{a(u)}{r^{l+3-A}}\right), \tag{4.1}$$

while the advanced radiation fields are

$$\Psi_A = K_{A-2}(l)\,{}_{2-A}Y_{l0}r^{l-2}D^{l+2-A}\left(\frac{b(u-r)}{r^{l-1+A}}\right), \tag{4.2}$$

where

$$K_p(l) \equiv \left(\frac{(l+p)!}{(l-p)!}\right)^{\frac{1}{2}}, \quad d \equiv \frac{\partial}{\partial u} + \frac{\partial}{\partial r}, \quad D \equiv \frac{\partial}{\partial r},$$

and

$$l = 2, 3, \cdots,$$

and the ${}_sY_{l0}$ are generalized spherical harmonics.[12] The functions $a(u)$ and $b(u-r)$ are arbitrary functions of their respective arguments, which determine the profiles of the retarded and advanced waves, respectively. The profile functions for each different value of l are, of course, unrelated. If $d^{l+1}a(u)/du^{l+1} = 0$, a retarded solution is called nonradiative, since energy is not being carried to infinity,[5,13] and an analogous definition is used for advanced solutions.

When one knows the Ψ_A, one can then use Eqs. (3.7) and (3.8) to obtain the spin coefficients and tetrad vector components to first order. This calculation has been discussed elsewhere,[7] and in this paper we simply give the results: The retarded solutions are

$$\begin{aligned}
\rho &= 0,\\
\sigma &= -K_{-2}Y_l^2\left(\frac{\bar a_l}{r^2} + \sum_{n=0}^{l-2} C_{ln0}\frac{a_{l-2-n}}{(n+2)r^{n+4}}\right),\\
\alpha &= -\overset{0}{\alpha}K_{-2}Y_l^{-2}\left(\frac{a_l}{r} + \sum_{n=0}^{l-2} C_{ln0}\frac{\bar a_{l-2-n}}{(n+2)(n+3)r^{n+3}}\right),\\
\tau &= -K_{-1}Y_l^1\sum_{n=0}^{l-1} C_{ln1}\frac{a_{l-1-n}}{(n+2)r^{n+3}},\\
\beta &= \tau - \bar\alpha,\\
\gamma &= -\overset{0}{\alpha}K_{-1}\sum_{n=0}^{l-1} C_{ln1}\frac{Y_l^1 a_{l-1-n} - Y_l^{-1}\bar a_{l-1-n}}{(n+2)(n+3)r^{n+2}}\\
&\quad - Y_l^0\sum_{n=0}^{l} C_{ln2}\frac{a_{l-n}}{(n+2)r^{n+2}},\\
\mu &= -Y_l^0\sum_{n=0}^{l} C_{ln2}\frac{a_{l-n}}{(n+1)r^{n+2}},\\
\lambda &= -K_{-2}Y_l^{-2}\left(-\frac{a_{l+1}}{r} + \frac{a_l}{r^2}\right.\\
&\quad \left. + \sum_{n=0}^{l-2} C_{ln0}\frac{\bar a_{l-2-n}}{(n+2)(n+3)r^{n+4}}\right),\\
\nu &= -K_1Y_l^{-1}\sum_{n=0}^{l+1} C_{ln3}\frac{a_{l+1-n}}{(n+1)r^{n+1}}\\
&\quad + K_{-1}Y_l^{-1}\sum_{n=0}^{l-1} C_{ln1}\frac{\bar a_{l-1-n}}{(n+2)(n+3)r^{n+3}},\\
U &= -Y_l^0\sum_{n=0}^{l} C_{ln2}\frac{a_{l-n} + \bar a_{l-n}}{(n+1)(n+2)r^{n+1}},\\
\omega &= -K_{-1}Y_l^1\left(\frac{\bar a_l}{r} + \sum_{n=0}^{l-1} C_{ln1}\frac{a_{l-1-n}}{(n+1)(n+2)r^{n+2}}\right),\\
X^i &= \overset{0}{\xi}{}^iK_{-1}Y_l^1\sum_{n=0}^{l-1} C_{ln1}\frac{a_{l-1-n}}{(n+2)(n+3)r^{n+2}}\\
&\quad + \text{complex conjugate},\\
\xi^i &= \overset{0}{\xi}{}^iK_{-2}Y_l^2\left(\frac{\bar a_l}{r} + \sum_{n=0}^{l-2} C_{ln0}\frac{a_{l-2-n}}{(n+2)(n+3)r^{n+3}}\right),
\end{aligned} \tag{4.3a}$$

where

$$C_{lnA} \equiv (-1)^n(K_{2-A+n}(l))^2\frac{1}{n!}$$

and we have indicated derivatives of $a(u)$ by a subscript, as $a_l \equiv d^l a(u)/du^l$; the advanced solutions are

$$\begin{aligned}
\rho &= 0, \\
\sigma &= K_{-2} Y_l^2 r^{l-1} D^l [r^{-1} D(b/r^{l-1})], \\
\alpha &= -\overset{0}{\alpha} K_{-2} Y_l^{-2} r^l D^{l-1} [r^{-2} D(\bar{b}/r^{l-1})], \\
\tau &= K_{-1} Y_l^1 r^{l-1} D^{l-1} [r^{-1} D(b/r^l)], \\
\beta &= \tau - \bar{\alpha}, \\
\gamma &= \overset{0}{\alpha} K_{-1} r^l D^{l-2} [r^{-2} D((Y_l^1 b - Y_l^{-1} \bar{b})/r^l)] \\
&\quad + Y_l^0 r^{l-1} D^{l-2} [r^{-1} D(b/r^{l+1})], \\
\mu &= Y_l^0 r^{l-1} D^{l-1} (b/r^{l+2}), \qquad (4.3b) \\
\lambda &= -K_{-2} Y_l^{-2} r^{l-1} D^{l-1} [r^{-2} D(\bar{b}/r^{l-1})], \\
\nu &= K_1 Y_l^{-1} r^{l-1} D^{l-2} (b/r^{l+3}) \\
&\quad + K_{-1} Y_l^{-1} r^{l-1} D^{l-2} [r^{-2} D(\bar{b}/r^l)], \\
U &= -Y_l^0 r^l D^{l-2} ((b + \bar{b})/r^{l+3}), \\
\omega &= -K_{-1} Y_l^1 r^l D^{l-1} (b/r^{l+2}), \\
X^i &= \overset{0}{\xi^i} K_{-1} Y_l^1 r^l D^{l-2} [r^{-2} D(b/r^l)] + \text{c. c.}, \\
\xi^i &= \overset{0}{\xi^i} K_{-2} Y_l^2 r^l D^{l-1} [r^{-2} D(b/r^{l-1})].
\end{aligned}$$

Equations (4.3b) are easily obtained from Eqs. (3.7) and (4.2), when one uses the identity

$$D^p [r^{q+t} D^q (F/r^t)] \equiv r^{q+t-p} D^q (D^p F/r^{t-p})$$

which holds for any arbitrary function F and all nonnegative integers p, q, and t.

The above solutions, considering the advanced and retarded solutions separately, are of course singular along the worldline $r = 0$. In this paper, we are particularly interested in imploding–exploding waves, and the corresponding first-order solutions are obtained by adding together the ${}_{(1)}\Psi_A$ of Eq. (4.1) and those of Eq. (4.2), with the added condition that $b(x) = -a(x)$, where a and b are the functional forms appearing in Eqs. (4.1) and (4.2). This latter condition simply means that the same wave comes out that went in. Thus, the first-order Ψ_A with which we deal are

$$\Psi_A = {}_{2-A}Y_{l0} r^{l-2} \left[K_{2-A} \left(\frac{\partial}{\partial u} + \frac{\partial}{\partial r} \right)^{l-2+A} \left(\frac{a(u)}{r^{l+3-A}} \right) - K_{A-2} \left(\frac{\partial}{\partial r} \right)^{l+2-A} \left(\frac{a(u-r)}{r^{l-1+A}} \right) \right]. \qquad (4.4)$$

In the same way, one can easily construct from Eqs. (4.3) the retarded-minus-advanced first-order spin coefficients and tetrad vector components, corresponding to the Ψ_A of Eq. (4.4), but we do not display them explicitly.

The pure retarded Ψ_A given by Eqs. (4.1), and the pure advanced Ψ_A given by Eqs. (4.2), are obviously singular at $r = 0$. This relates to the fact that the pure retarded wave must have originated at $r = 0$, while the pure advanced wave must have been absorbed there. The Ψ_A given by Eq. (4.4), however, are for an imploding–exploding wave, and such a situation involves no emission or absorption; thus, it is of interest to determine the structure of these Ψ_A [Eq. (4.4)] at $r = 0$. In order to do this, it is sufficient that we assume that

$$a(u - r) = \sum_{n=0}^{l+2} \frac{a_n (-r)^n}{n!} + O(r^{l+3}), \qquad (4.5)$$

for small r. We actually make the stronger assumption

$$a(u - r) = \sum_{n=0}^{N} \frac{a_n (-r)^n}{n!} + O(r^{N+1}), \quad N \geq 2l + 1, \qquad (4.6)$$

for convenience, but that of Eq. (4.5) is sufficient. If one inserts Eq. (4.6) into Eq. (4.4), one obtains, after some manipulation,

$$\Psi_A = K_{A-2\,2-A} Y_{l0} r^{l-2} \times \sum_{n=2l+1}^{N} \frac{a_n (-1)^{n+1} (n - l + 1 - A)!}{n!\,(n - 2l - 1)!} r^{n-2l-1} + O(r^{N-l-2}). \qquad (4.7)$$

Thus, we see that, as far as the variable r is concerned, the Ψ_A are well behaved at $r = 0$. In fact, only for the case $l = 2$ are they nonvanishing there. For the $l = 2$ case, there remains the problem that θ and ϕ, and thus the ${}_{2-A}Y_{l0}$, are not well defined at $r = 0$. This difficulty is resolved later.

Using the results embodied in Eq. (4.7), it is possible to find all the first-order spin coefficients and tetrad vector components for small r directly from Eqs. (3.7) and (3.8). These calculations are quite complicated, and the results are not simple; therefore, we give only a sample calculation (the spin coefficient σ), and we give results only for the tetrad vector components, which are all we need later.

We begin by writing each of Eqs. (3.7) in terms of an integral from 0 to r, instead of an integral from ∞ to r. Thus, for example,

$$\sigma = \frac{\sigma^\circ}{r^2} + \frac{1}{r^2} \int_\infty^r r'^2 \Psi_0 \, dr'$$

is rewritten as

$$\sigma = \frac{1}{r^2} \left(\sigma^\circ + \int_\infty^0 r'^2 \Psi_0 \, dr' \right) + \int_0^r r'^2 \Psi_0 \, dr',$$

and in view of Eq. (4.7), we have

$$\sigma = \frac{1}{r^2}\left(\sigma^\circ + \int_\infty^0 r'^2 \Psi'_0\, dr'\right) + O(r^{l-1}). \quad (4.8)$$

The integral in the first term is, from Eq. (4.4),

$${}_2Y_{l0}K_2\int_\infty^0 r^l\left[\left(\frac{\partial}{\partial r} - \frac{\partial}{\partial u}\right)^{l-2}\left(\frac{a(u)}{r^{l+3}}\right) - \frac{(l-2)!}{(l+2)!}\left(\frac{\partial}{\partial r}\right)^{l+2}\left(\frac{a(u-r)}{r^{l-1}}\right)\right] dr, \quad (4.9)$$

which, after substituting Eq. (4.6), becomes

$${}_2Y_{l0}K_2\int_\infty^0 r^l\left[\sum_{n=0}^{l-2}\frac{(l-2)!}{(l-2-n)!\,n!}\left(\frac{\partial}{\partial r}\right)^{l-2-n}\frac{a_n(-1)^n}{r^{l+3}} - \frac{(l-2)!}{(l+2)!}\left(\frac{\partial}{\partial r}\right)^{l+2}\sum_{n=0}^{N}\frac{a_n r^n}{n!\,r^{l-1}}\right] dr. \quad (4.10)$$

If we use the previously mentioned identity with $p = t = 1$ on the second line of Eq. (4.10), then the integration can be performed, yielding, after many manipulations,

$$\int_\infty^0 r'^2\Psi'_0\, dr' = -{}_2Y_{l0}K_{-2}a_l(u). \quad (4.11)$$

Substituting this into Eq. (4.8), and evaluating σ° via Eq. (3.8c), gives

$$\sigma = -[(a_l + \bar{a}_l)Y_l^2K_{-2}]r^{-2} + O(r^{l-1}). \quad (4.12)$$

We now list the results of similar calculations on the tetrad vector components:

$$U = -(a_{l+1} + \bar{a}_{l+1})K_{-1}^2Y_l^0 + (a_{l+2} + \bar{a}_{l+2})K_{-2}^2Y_l^0 r + O(r^l), \quad (4.13a)$$

$$\omega = -(a_l + \bar{a}_l)K_{-1}Y_l^1 r^{-1} + a_{l+1}K_{-1}^3Y_l^1 + O(r^l), \quad (4.13b)$$

$$X^2 = (a_{l+1} + \bar{a}_{l+1})K_{-1}^3Y_l^1 r^{-1} - (a_{l+2} + \bar{a}_{l+2})2K_{-1}K_{-2}^2Y_l^1 + O(r^{l-1}), \quad (4.13c)$$

$$X^3/i\csc\theta = (\bar{a}_{l+1} - a_{l+1})K_{-1}^3Y_l^1 r^{-1} + (a_{l+2} - \bar{a}_{l+2})2K_{-1}K_{-2}^2Y_l^1 + O(r^{l-1}), \quad (4.13d)$$

$$\xi^2 = (a_l + \bar{a}_l)K_{-2}Y_l^2 r^{-2} - a_{l+1}2K_{-2}K_{-1}^2Y_l^2 r^{-1} + O(r^{l-1}), \quad (4.13e)$$

$$\xi^3/i\csc\theta = -(a_l + \bar{a}_l)K_{-2}Y_l^2 r^{-2} + a_{l+1}2K_{-2}K_{-1}^2Y_l^2 r^{-1} + O(r^{l-1}). \quad (4.13f)$$

We also want to know the behavior of the metric tensor for small r, and this can be calculated from Eqs. (4.13) and Eqs. (2.9) yielding

$$g^{11} = {}_0Y_{l0}[-2K_{-1}^2(a_{l+1} + \bar{a}_{l+1}) + 2K_{-2}^2(a_{l+2} + \bar{a}_{l+2})r] + O(r^l), \quad (4.14a)$$

$$g^{12} = {}_1Y_{l0}[2K_{-1}(a_l + \bar{a}_l)r^{-2} - 2K_{-1}K_{-2}^2(a_{l+2} + \bar{a}_{l+2})] + O(r^{l-1}), \quad (4.14b)$$

$$g^{13} = i\csc\theta 2K_{-1}K_{-2}^2(a_{l+2} - \bar{a}_{l+2})\,{}_1Y_{l0} + O(r^{l-1}), \quad (4.14c)$$

$$g^{22} = [-4K_{-2}(a_l + \bar{a}_l)r^{-3} + 4K_{-2}K_{-1}^2(a_{l+1} + \bar{a}_{l+1})r^{-2}]\,{}_2Y_{l0} + O(r^{l-2}), \quad (4.14d)$$

$$g^{23} = [i\csc\theta 4K_{-1}^2K_{-2}(a_{l+1} - \bar{a}_{l+1})r^{-2}]\,{}_2Y_{l0} + O(r^{l-2}), \quad (4.14e)$$

$$g^{33} = \csc^2\theta[4K_{-2}(a_l + \bar{a}_l)r^{-3} - 4K_{-2}K_{-1}^2(a_{l+1} + \bar{a}_{l+1})r^{-2}]\,{}_2Y_{l0} + O(r^{l-2}). \quad (4.14f)$$

B. The Nature of the Singularities

We have noted that the Ψ'_A are ambiguous at $r = 0$ due to the presence of the angular variables θ and ϕ, and we have found that some of the tetrad vector components and metric tensor components actually diverge at $r = 0$. We now show that these difficulties are a consequence of the choice of coordinates and of the choice of tetrad vectors made in the problem. This choice, which has proven highly useful in studying radiation fields for large r, is unsatisfactory for $r \sim 0$. We consider first the metric tensor components since they are independent of the choice of tetrad vectors, and enable us to consider the question of coordinate conditions by itself.

The zeroth-order metric tensor, which follows directly from Eqs. (3.1), gives the line element

$$ds^2 = -2\,du\,dr + 2\,du^2 - 2r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \quad (4.15)$$

for which $\det(g_{\mu\nu}) = 0$, at $r = 0$. One knows, of course, that the coordinate transformation

$$\begin{aligned} x^0 &= 2^{\frac{1}{2}}(-u + \tfrac{1}{2}r), \\ x^1 &= r\sin\theta\cos\phi/2^{\frac{1}{2}}, \\ x^2 &= r\sin\theta\sin\phi/2^{\frac{1}{2}}, \\ x^3 &= r\cos\theta/2^{\frac{1}{2}}, \end{aligned} \quad (4.16)$$

transforms ds^2 to the form

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \quad (4.17)$$

We now wish to find a first-order correction to Eqs. (4.16) so that these corrected transformations, when applied to the singular metric tensor ${}_{(0)}g^{\mu\nu} + {}_{(1)}g^{\mu\nu}$ given by Eqs. (4.15) and (4.14), result in a metric tensor well behaved for $r \sim 0$. It turns out to be convenient to assume the desired transformations to be

$$\begin{aligned} x^0 &= 2^{\frac{1}{2}}(-u + \tfrac{1}{2}r + \epsilon^0), \\ x^1 &= (r + \epsilon^1)\sin\theta\cos\phi/2^{\frac{1}{2}}, \\ x^2 &= (r + \epsilon^2)\sin\theta\sin\phi/2^{\frac{1}{2}}, \\ x^3 &= r\cos\theta/2^{\frac{1}{2}} + \epsilon^3/2^{\frac{1}{2}}. \end{aligned} \tag{4.18}$$

If we then do this coordinate transformation to ${}_{(0)}g^{\mu\nu} + {}_{(1)}g^{\mu\nu}$, and ask that the resulting ${}_{(0)}g^{\mu\nu'} + {}_{(1)}g^{\mu\nu'}$ be well behaved at $r = 0$, we obtain differential equations for ϵ^0, ϵ^1, ϵ^2, and ϵ^3. Three of these can be easily solved to yield

$$\begin{aligned} \epsilon^0 &= -\tfrac{1}{2}K_{-1}^2(a_l + \tilde{a}_l){}_0Y_{l0}, \\ \epsilon^1 = \epsilon^2 &= K_{-2}(a_l + \tilde{a}_l){}_2Y_{l0}, \end{aligned} \tag{4.19}$$

while the equation for ϵ^3 is

$$\epsilon^3_{,\theta} = (\cos\theta Y_l^1 K_{-1} + \sin\theta Y_l^2 K_{-2})(a_l + \tilde{a}_l). \tag{4.20}$$

This last equation cannot be solved by putting ϵ^3 equal to any simple combination of generalized spherical harmonics, but one can see that for each value of l it can be easily solved. We have no need of the actual solution, so we are content with Eq. (4.20). It has now been established that the first-order geometry for our problem is everywhere nonsingular.

We are particularly interested in the behavior of the Ψ_A at $r = 0$. They, of course, depend on the choice of tetrad vectors, and, in as much as our l^μ and n^μ vector fields diverge from the world-line $r = 0$, it is to be expected that a tetrad transformation is also needed.

In the zeroth order, the tetrad vectors are given by

$$\begin{aligned} l^\mu &= \delta^\mu_1, \\ \overset{0}{n}{}^\mu &= -\delta^\mu_0 - \delta^\mu_1, \\ \overset{0}{m}{}^\mu &= r^{-1}(\delta^\mu_2 + i\csc\theta\delta^\mu_3) \end{aligned} \tag{4.21}$$

in the u, r, θ, and ϕ coordinate system. We denote this tetrad by ${}_{(0)}Z^\mu_B = (l^\mu, {}_{(0)}n^\mu, {}_{(0)}m^\mu, {}_{(0)}\bar{m}^\mu)$. If we make the tetrad transformation

$$\overset{0}{V}{}^\mu_B = L_B{}^C \overset{0}{Z}{}^\mu_C, \tag{4.22}$$

where

$$L_B{}^C \equiv (2)^{-\frac{1}{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ \dfrac{x^1}{r} & -\dfrac{x^1}{r} & \dfrac{1}{w}\left(\dfrac{x^1x^3}{r} - ix^2\right) & \dfrac{1}{w}\left(\dfrac{x^1x^3}{r} - ix^2\right) \\ \dfrac{x^2}{r} & -\dfrac{x^2}{r} & \dfrac{1}{w}\left(\dfrac{x^2x^3}{r} - ix^1\right) & \dfrac{1}{w}\left(\dfrac{x^2x^3}{r} + ix^1\right) \\ \dfrac{x^3}{r} & -\dfrac{x^3}{r} & -\dfrac{w}{r} & -\dfrac{w}{r} \end{pmatrix}, \tag{4.23}$$

with $r = [(x^1)^2 + (x^2)^2 + (x^3)^2]^{\frac{1}{2}}$ and

$$w = [(x^1)^2 + (x^2)^2]^{\frac{1}{2}},$$

and then the coordinate transformation given by Eq. (4.16), we obtain

$$\overset{0}{V}{}^\mu_B = \delta^\mu_B. \tag{4.24}$$

The tensor indices in Eq. (4.24) refer to the coordinates (x^0, x^1, x^2, x^3) of Eq. (4.16).

Let us now define tetrad components of ${}_{(1)}C_{\mu\nu\rho\sigma}$ with respect to the new tetrad ${}_{(0)}V^\mu_B$:

$$C_{BCDE} \equiv \overset{1}{C}_{\mu\nu\rho\sigma}\overset{0}{V}{}^\mu_B\overset{0}{V}{}^\nu_C\overset{0}{V}{}^\rho_D\overset{0}{V}{}^\sigma_E. \tag{4.25}$$

A lengthy but straightforward calculation then gives us the C_{BCDE} in terms of the Ψ_A. The result is

$$\begin{aligned} -C_{BCDE} &= \Psi_0(L_{[B}{}^0L_{C]}{}^2L_{[D}{}^0L_{E]}{}^2) + \Psi_4(L_{[B}{}^1L_{C]}{}^3L_{[D}{}^1L_{E]}{}^3) \\ &\quad + \Psi_1(L_{[B}{}^0L_{C]}{}^1L_{[D}{}^0L_{E]}{}^2 + L_{[B}{}^1L_{C]}{}^3L_{[D}{}^1L_{E]}{}^2 \\ &\quad - L_{[B}{}^0L_{C]}{}^2L_{[D}{}^2L_{E]}{}^3 - L_{[B}{}^2L_{C]}{}^3L_{[D}{}^1L_{E]}{}^2) \\ &\quad + \Psi_3(-L_{[B}{}^0L_{C]}{}^1L_{[D}{}^1L_{E]}{}^3 - L_{[B}{}^1L_{C]}{}^3L_{[D}{}^0L_{E]}{}^1 \\ &\quad + L_{[B}{}^1L_{C]}{}^3L_{[D}{}^2L_{E]}{}^3 + L_{[B}{}^2L_{C]}{}^3L_{[D}{}^1L_{E]}{}^3) \\ &\quad + \Psi_2[-L_{[B}{}^0L_{C]}{}^2L_{[D}{}^1L_{E]}{}^3 - L_{[B}{}^1L_{C]}{}^3L_{[D}{}^0L_{E]}{}^2 \\ &\quad + (L_{[B}{}^0L_{C]}{}^1 - L_{[B}{}^2L_{C]}{}^3)(L_{[D}{}^0L_{E]}{}^1 - L_{[D}{}^2L_{E]}{}^3)] \\ &\quad + \text{complex conjugate.} \end{aligned} \tag{4.26}$$

The expressions given by Eq. (4.7) for the Ψ_A's, for $r \sim 0$, show that

$$\lim_{r\to 0}\Psi_A = 0, \quad l > 2. \tag{4.27}$$

Since the C_{BCDE} are just linear combinations of the Ψ_A, we conclude that

$$\lim_{r\to 0} C_{BCDE} = 0, \quad l > 2. \tag{4.28}$$

When $l = 2$, one obtains

$$\lim_{r\to 0} C_{0101} = \lim_{r\to 0} C_{0123} = \frac{2}{3(5\pi)^{\frac{1}{2}}} \frac{d^5}{du^5}(a + \bar{a}),$$

and that C_{0102}, C_{0103}, C_{0112}, and C_{0113} all vanish as $r \to 0$. These particular components are a complete and independent set for C_{BCDE}. Thus, we see that the tetrad components of the Weyl tensor are in fact well defined at $r = 0$ for arbitrary l. Since the tetrad has the form given by Eq. (4.24), we also see that the Weyl tensor itself is nonsingular.

The significance of the preceding calculations is twofold. They show, first of all, that the first-order imploding–exploding wave corresponding to the Ψ_A of Eqs. (4.4) leads to a geometry with no singularities despite the ambiguities of the Ψ_A at $r = 0$, for $l = 2$, and the explicit infinities at $r = 0$ in the metric tensor. The difficulties with the mathematical expressions are due to the choice of coordinate and tetrad conditions, and the transformations given by Eqs. (4.18) and Eqs. (4.23) remove the offending ambiguities and infinities and make the good behavior of the geometry manifest.

5. THE SECOND-ORDER SOLUTIONS AND SINGULARITIES

A. The Solution

We now find the second-order field in our problem of an imploding–exploding, first-order, quadrupole pulse of radiation. The nonsingular first-order field pictured in Fig. 1 is given by Eqs. (4.4) with $l = 2$, and by the sum of the retarded and advanced fields of Eqs. (4.3) with $l = 2$ and $a(x) = -b(x)$. The function $a(u)$ is nonvanishing only in the pulse region ($|u| < u_0$), separating regions A and B and hence b is nonvanishing only in the pulse region ($|v| < u_0$), separating regions B and C.

To find the second-order field, we need the driving terms D_A and R_{A+1} for this case. From Eqs. (3.11) and (3.12), we find them to be

$$\begin{aligned}
&6^{\frac{1}{2}}\left(\frac{(4-A)!}{A!}\right)^{\frac{1}{2}} R_{A+1} \\
&= Y_2^{-1}Y_2^{2-A}\left[r^2 D\left(\frac{\dot{\bar{b}}}{r^4}\right) + \frac{3\dot{\bar{a}}}{r^2} - \frac{4\bar{a}}{r^3} + \frac{\ddot{\bar{a}}}{r}\right] \\
&\quad \times \left[D^{5-A}\left(\frac{b}{r^{A+1}}\right) - \frac{(5-A)!}{A!}\, d^A\left(\frac{a}{r^{6-A}}\right)\right] \\
&\quad - Y_2^{-2}Y_2^{3-A}[A(5-A)]^{\frac{1}{2}} \\
&\quad \times \left\{\tfrac{1}{2}rD\left[\frac{1}{r^2} D\left(\frac{\bar{b}}{r}\right)\right] + \frac{\ddot{a}}{2r^2} + \frac{2\bar{a}}{r^4}\right\} \\
&\quad \times \left[D^{4-A}\left(\frac{b}{r^{A+1}}\right) + \frac{(4-A)!}{A!}\, d^A\left(\frac{a}{r^{5-A}}\right)\right] \\
&\quad + Y_2^{-2}Y_2^{3-A}\left(\frac{A}{5-A}\right)^{\frac{1}{2}} \\
&\quad \times \left\{\tfrac{1}{2}rD\left[\frac{1}{r^2} D\left(\frac{\bar{b}}{r}\right)\right] - \frac{\dddot{a}}{2r} + \frac{\ddot{a}}{2r^2} + \frac{2\bar{a}}{r^4}\right\} \\
&\quad \times \left[D^{5-A}\left(\frac{b}{r^A}\right) + \frac{(5-A)!}{(A-1)!}\, d^{A-1}\left(\frac{a}{r^{6-A}}\right)\right],
\end{aligned} \tag{5.1}$$

$$\begin{aligned}
&6^{\frac{1}{2}}\left(\frac{(4-A)!}{A!}\right)^{\frac{1}{2}} D_A \\
&= Y_2^2 Y_2^{-A}[(A+1)(A+2)(3-A)(4-A)]^{\frac{1}{2}} \\
&\quad \times \left\{rD\left[\frac{1}{r^2} D\left(\frac{\dot{b}}{r}\right)\right] + \frac{\ddot{\bar{a}}}{2r^2} + \frac{2a}{r^4}\right\} P_{A+1} \\
&\quad - \tfrac{1}{2}Y_2^1 Y_2^{1-A}[(A+1)(4-A)]^{\frac{1}{2}} \\
&\quad \times \left[\frac{1}{r} D\left(\frac{b}{r^2}\right) + \frac{\dot{a}}{r^3} - \frac{2a}{r^4}\right] P_A \\
&\quad + \tfrac{1}{2}Y_2^{-1}Y_2^{3-A}[A(5-A)]^{\frac{1}{2}}\left[\frac{1}{r} D\left(\frac{\dot{\bar{b}}}{r^2}\right) + \frac{\dot{\bar{a}}}{r^3} - \frac{2\bar{a}}{r^4}\right] P_A \\
&\quad + \tfrac{1}{2}\sqrt{3}\, Y_2^0 Y_2^{2-A}(2-A) \\
&\quad \times \left[D\left(\frac{b}{r^3}\right) - \frac{\ddot{a}}{2r^2} + \frac{2\dot{a}}{r^3} - \frac{3a}{r^4}\right] P_A \\
&\quad + 3Y_2^1 Y_2^{1-A}(A-2)[(A+1)(4-A)]^{\frac{1}{2}} \\
&\quad \times \left\{rD\left[\frac{1}{r} D\left(\frac{b}{r^2}\right)\right] - \frac{3\dot{a}}{r^3} + \frac{8a}{r^4}\right\} P_{A+1} \\
&\quad + \sqrt{3}\, Y_2^0 Y_2^{2-A}\left(\frac{b+\bar{b}}{r^3} + \frac{\ddot{a}+\ddot{\bar{a}}}{2r}\right. \\
&\quad \left. - \frac{\dot{a}+\dot{\bar{a}}}{r^2} + \frac{a+\bar{a}}{r^3}\right) DP_A \\
&\quad + Y_2^1 Y_2^{1-A}[(A+1)(4-A)]^{\frac{1}{2}} \\
&\quad \times \left[-r^2 D\left(\frac{\ddot{\bar{b}}}{r^4}\right) - \frac{\dddot{\bar{a}}}{r} + \frac{3\ddot{a}}{r^2} + \frac{4\dot{a}}{r^3}\right] DP_{A+1} \\
&\quad + Y_2^{-1}Y_2^{3-A}\left(\frac{A}{5-A}\right)^{\frac{1}{2}}\left[\frac{1}{2r} D\left(\frac{\dot{\bar{b}}}{r^2}\right) + \frac{3b}{r^4} + \frac{\dot{\bar{a}}}{2r^3}\right. \\
&\quad \left. - \frac{\bar{a}}{r^4} - \frac{\ddot{\bar{a}}}{2r} + \frac{3\ddot{\bar{a}}}{2r^2} - \frac{3\dot{a}}{r^3} + \frac{3a}{r^4}\right] P_{A-1} \\
&\quad - \sqrt{3}\, Y_2^0 Y_2^{2-A}\left[rD\left(\frac{\dot{b}}{r^4}\right) - \frac{\ddot{a}}{r^2} + \frac{3\dot{a}}{r^3} - \frac{4a}{r^4}\right] P_A \\
&\quad + \tfrac{1}{2}Y_2^2 Y_2^{-A}(3-A)[(A+1)(A+2)(3-A)(4-A)]^{\frac{1}{2}} \\
&\quad \times \left\{rD^2\left[\frac{1}{r} D\left(\frac{b}{r}\right)\right] - \frac{\ddot{\bar{a}}}{r^2} - \frac{12a}{r^4}\right\} P_{A+2},
\end{aligned} \tag{5.2}$$

where $A = 0, 1, 2, 3$, and

$$P_A \equiv D^{4-A}\left(\frac{b}{r^{A+1}}\right) + \frac{(4-A)!}{A!}\, d^A\left(\frac{a}{r^{5-A}}\right).$$

The products of spin-weighted functions occurring in Eqs. (5.1) and (5.2) are all special cases of the general product $Y_l^s Y_{l'}^{s'}$. The product $Y_l^s Y_{l'}^{s'}$ may be expanded in a finite series in the spin-weighted functions $Y_L^{s+s'}$ with the numerical coefficients of $Y_L^{s+s'}$ in the expansion given in terms of Wigner 3-j symbols by

$$(-1)^{s+s'}\frac{1}{2(\pi)^{\frac{1}{2}}}[(2l+1)(2l'+1)(2L+1)]^{\frac{1}{2}} \times \begin{pmatrix} l & l' & L \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l & l' & L \\ -s & -s' & s+s' \end{pmatrix}.$$

In the expansions used in Eqs. (5.1) and (5.2), only the terms $L = 0, 2, 4$ occur (the $Y_0^{s+s'}$ term occurs only when $s + s' = 0$). These expansions determine the angular dependence of the second-order field since the angular dependence of the equations determining the ${}_{(2)}\Psi_A$, Eqs. (3.13) and (3.15), is eliminated by taking the ${}_{(2)}\Psi_A$ to be a linear combination of Y_L^{2-A}, $L = 0, 2, 4$.

From R_1 and D_1, we obtain the quantity $\mathfrak{D}_0$ which drives Eq. (3.15):

$$\mathfrak{D}_0 = D_1 + \frac{1}{r^5}\int_\infty^r r'^4 \eth R_1\, dr'. \tag{5.3}$$

A method of solving Eq. (3.15) with $\mathfrak{D}_0$ known was given in Ref. 7, Sec. 3 and Appendix B. We apply the method to the case at hand and obtain ${}_{(2)}\Psi_0$ to be

$$\begin{aligned}{}^2\Psi_0 = {}& \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}}\, Y_2^2 r^{-2} D^2\left(\frac{Q(u)-Q(v)}{r}\right) \\ & - \frac{5^{\frac{1}{2}}}{7(2\pi)^{\frac{1}{2}}}\, Y_4^2\Bigg[r^{-2}D^2\left(\frac{\ddot{S}(u)-S''(v)}{4r} - \frac{7}{6}\frac{\dot{S}(u)+S'(v)}{r^2} + \frac{7}{3}\frac{S(u)-S(v)}{r^3}\right) \\ & + 28\frac{E(u)-E(v)}{r^7}\Bigg] + T_0,\end{aligned} \tag{5.4a}$$

where

$$\begin{aligned} Q(u) &\equiv -\tfrac{1}{4}\int_{-\infty}^{u} dx_2 \int_{-\infty}^{x_2} dx_1 \int_{-\infty}^{x_1} \bar{a}(x)\frac{d^6 a(x)}{dx^6}\, dx, \\ S(u) &\equiv 2\int_{-\infty}^{u} dx_1 \int_{-\infty}^{x_1} Q(x)\, dx, \\ E(u) &\equiv -\tfrac{15}{4}\int_{-\infty}^{u} a(x)\frac{d^2 a(x)}{dx^2}\, dx, \\ v &\equiv u - r. \end{aligned}$$

Note that

$$\begin{aligned} Q(v) &= -\tfrac{1}{4}\int_{-\infty}^{v} dx_2 \int_{-\infty}^{x_2} dx_1 \int_{-\infty}^{x_1} \bar{a}(x)\frac{d^6 a(x)}{dx^6}\, dx \\ &= -\tfrac{1}{4}\int_{-\infty}^{v} dx_2 \int_{-\infty}^{x_2} dx_1 \int_{-\infty}^{x_1} \bar{b}(x)\frac{d^6 b(x)}{dx^6}\, dx. \end{aligned}$$

The primes denote differentiation with respect to v, i.e., $Q'(v) \equiv dQ(v)/dv$.

Knowing ${}_{(2)}\Psi_0$ and R_{A+1}, we calculate the remaining ${}_{(2)}\Psi_A$'s from Eqs. (3.13) and obtain

$$\begin{aligned}{}^2\Psi_1 = {}& \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}}\, Y_2^1 \frac{1}{r^2} D\Bigg(-\frac{\dot{Q}(u)+Q'(v)}{r} + 2\frac{Q(u)-Q(v)}{r^2}\Bigg) \\ & - \frac{3}{4}\frac{5^{\frac{1}{2}}}{7(\pi)^{\frac{1}{2}}}\, Y_4^1\Bigg[\frac{1}{r^2}D\Bigg(-\frac{1}{9}\frac{\dddot{S}(u)+S'''(v)}{r} + \frac{\ddot{S}(u)-S''(v)}{r^2} - \frac{14}{3}\frac{\dot{S}(u)+S'(v)}{r^3} \\ & + \frac{28}{3}\frac{S(u)-S(v)}{r^4}\Bigg) - \frac{112}{3}\frac{E(u)-E(v)}{r^7}\Bigg] + T_1,\end{aligned} \tag{5.4b}$$

$$\begin{aligned}{}^2\Psi_2 = {}& \frac{1}{12(\pi)^{\frac{1}{2}}}\, Y_0^0 \frac{\ddot{Q}(u)+Q''(v)}{r^3} \\ & + \frac{5^{\frac{1}{2}}}{7(\pi)^{\frac{1}{2}}}\, Y_2^0\Bigg(\frac{\ddot{Q}(u)-Q''(v)}{6r^3} - \frac{\dot{Q}(u)+Q'(v)}{r^4} + 2\frac{Q(u)-Q(v)}{r^5}\Bigg) \\ & - \frac{15}{2}\frac{1}{7(\pi)^{\frac{1}{2}}}\, Y_4^0\Bigg(\frac{\ddot{S}(u)-S''''(v)}{180r^3} - \frac{\dddot{S}(u)+S'''(v)}{9r^4} \\ & + \frac{\ddot{S}(u)-S''(v)}{r^5} - \frac{14}{3}\frac{\dot{S}(u)+S'(v)}{r^6} \\ & + \frac{28}{3}\frac{S(u)-S(v)}{r^7} + \frac{28}{3}\frac{E(u)-E(v)}{r^7}\Bigg) + T_2,\end{aligned} \tag{5.4c}$$

$$\begin{aligned}{}^2\Psi_3 = {}& \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}}\, Y_2^{-1}\Bigg(-\frac{\ddot{Q}(u)}{r^3} + \frac{3\dot{Q}(u)}{r^4} - \frac{4Q(u)}{r^5} + \frac{Q'(v)}{r^4} + \frac{4Q(v)}{r^5}\Bigg) \\ & + \frac{5^{\frac{1}{2}}}{7(\pi)^{\frac{1}{2}}}\, Y_4^{-1}\Bigg[\frac{1}{4}\Bigg(\frac{\dddot{S}(u)}{3r^3} - \frac{10\ddot{S}(u)}{3r^4} + \frac{20\ddot{S}(u)}{r^5} - \frac{70\dot{S}(u)}{r^6} + \frac{112S(u)}{r^7} \\ & - \frac{S'''(v)}{6r^4} - \frac{3S''(v)}{r^5} - \frac{21S'(v)}{r^6} - \frac{56S(v)}{r^7}\Bigg) \\ & + 28\frac{E(u)-E(v)}{r^7}\Bigg] + T_3,\end{aligned} \tag{5.4d}$$

$$\overset{2}{\Psi}_4 = \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}} Y_2^{-2}\left(\frac{\ddot{Q}(u)}{r^3} - \frac{2\dot{Q}(u)}{r^4} + \frac{2Q(u)}{r^5} - \frac{2Q(v)}{r^5}\right)$$
$$- \frac{3}{4}\frac{5^{\frac{1}{2}}2^{\frac{1}{2}}}{7(\pi)^{\frac{1}{2}}} Y_4^{-2}\left(\frac{\dddot{S}(u)}{6r^3} - \frac{10\ddot{S}(u)}{9r^4} + \frac{5\dot{S}(u)}{r^5}\right.$$
$$- \frac{14\dot{S}(u)}{r^6} + \frac{56S(u)}{3r^7} - \frac{S''(v)}{3r^5} - \frac{14S'(v)}{3r^6}$$
$$\left. - \frac{56S(v)}{3r^7} + \frac{56}{3}\frac{E(u) - E(v)}{r^7}\right) + T_4. \qquad (5.4e)$$

The quantities T_A are transient terms vanishing in regions A, B, and C. Their exact forms are necessary for the calculations that follow and are given in Appendix A.

In Eqs. (5.4), we have incorporated the results of integrating Eqs. (3.14) with the initial data of our problem; namely,

$$\overset{2}{\Psi}_0|_{u=-\infty} = \overset{2}{\Psi}_1^{\circ}|_{u=-\infty} = (\overset{2}{\Psi}_2^{\circ} + \overset{\tilde{2}}{\Psi}_2^{\circ})|_{u=-\infty} = 0.$$

The results of the integration are

$$\overset{2}{\Psi}_2^{\circ} = \frac{1}{3(2\pi)^{\frac{1}{2}}128}\left(\ddot{a}\dddot{a} - \int_{-\infty}^{u} \dddot{a}\,\dddot{a}\right)$$
$$\times \left(Y_0^0 - \frac{2(5^{\frac{1}{2}})}{7} Y_2^0 + \frac{1}{7} Y_4^0\right), \qquad (5.5a)$$

$$\overset{2}{\Psi}_1^{\circ} = \frac{5^{\frac{1}{2}}}{21(6\pi)^{\frac{1}{2}}16}\left(3\int_{-\infty}^{u} \ddot{a}\dddot{a} - \int_{-\infty}^{u} dx_1 \int_{-\infty}^{x_1} \dddot{a}\,\ddot{a}\right)$$
$$\times (-3Y_2^1 + \tfrac{1}{2}(6)^{\frac{1}{2}} Y_4^1), \qquad (5.5b)$$

$$\overset{2}{\Psi}_3^{\circ} = \overset{2}{\Psi}_4^{\circ} = 0, \qquad (5.5c)$$

where we have taken $\sigma^{\circ}(u) = 0$.

The solution given by Eqs. (5.4) is unique; it is picked out from all possible solutions by the fact that it satisfies our second-order boundary conditions. There is no freedom of adding a formally linear advanced field because of the requirement that there be no radiation coming in from $\mathfrak{I}^-$, and there is no freedom of adding a formally linear retarded field because of the requirement that there be no singularities at $r = 0$. We now investigate the solution in regard to four properties: outgoing radiation, singularities, incoming radiation, and the nature of the solution in regions A, B, and C.

Outgoing radiation is characterized by the news function,

$$\dot{\sigma}^{\circ}_{\text{out}} \equiv \lim_{r\to\infty} r^2 \overset{\tilde{2}}{\Psi}_3, \quad u = \text{const.}$$

This definition[14] is essentially equivalent to that of Bondi.[13] The nonvanishing of $\dot{\sigma}^{\circ}_{\text{out}}$ means that there is radiation reaching $\mathfrak{I}^+$. It follows easily from Eq. (5.4d) that $\dot{\sigma}^{\circ}_{\text{out}} = 0$. This result depends on our choice of ${}_{(2)}\sigma^{\circ}(u) = 0$ in obtaining Eqs. (5.4) and (5.5). However, this is not an arbitrary choice: It is the only choice possible with our boundary conditions. A nonzero ${}_{(2)}\sigma^{\circ}(u)$ can only be obtained (except for ${}_{(2)}\sigma^{\circ} = \text{const}$) by adding a formally linear retarded field, but this would introduce singularities at $r = 0$ into the ${}_{(2)}\Psi_A$ which are nonsingular as given by Eqs. (5.4).

To prove that the ${}_{(2)}\Psi_A$ are nonsingular, we simply evaluate them near $r = 0$. We assume that $\partial^7 b/\partial r^7$ exists, then b has an expansion near $r = 0$ powers of r up to r^7, $\partial b/\partial r$ has an expansion through r^6, etc. Similarly, $Q(v)$ has an expansion up to r^4, and $S(v)$ and $E(v)$ have expansions up to r^6. Direct substitution of these expansions into Eq. (5.4) with T_A given by Eq. (A1) and use of the fact that $a(u) = -b(u)$ gives immediately that

$$\overset{2}{\Psi}_A = O(1), \quad r \to 0. \qquad (5.6)$$

The powers of r^{-1} that ostensibly occur in the ${}_{(2)}\Psi_A$ cancel identically to zero in a neighborhood of $r = 0$.

The solution given by Eqs. (5.4) has been chosen so that it has no radiation coming in from $\mathfrak{I}^-$. To see this, one must show that the news function for incoming radiation is zero. This is done by transforming to a tetrad system $(\tilde{l}^\mu, \tilde{n}^\mu, \tilde{m}^\mu, \tilde{\bar{m}}^\mu)$ in which the vector $\tilde{n}^\mu$ that points along the backward null surfaces satisfies the same conditions as the vector l^μ associated with the forward null surfaces. Then the incoming news function, defined in the same way at $\mathfrak{I}^-$ as $\dot{\sigma}^{\circ}_{\text{out}}$ is at $\mathfrak{I}^+$, must be given as

$$\overset{2}{\dot{\sigma}}^{\circ}_{\text{in}} = \lim_{r\to\infty} r^2 \overset{\tilde{2}}{\tilde{\Psi}}_1, \quad u - r = \text{const.} \qquad (5.7)$$

Note that, although $u - r = \text{const}$ does not label backward null surfaces correct to first order inside the pulse, the limit in Eq. (5.7) is the same as one taken along the proper first-order backward null surfaces because ${}_{(2)}\tilde{\Psi}_1$ is second order.

The calculation of $\dot{\sigma}^{\circ}_{\text{in}}$ is straightforward but somewhat long: We give it in Appendix B. The important result is that $\dot{\sigma}^{\circ}_{\text{in}} = 0$.

We now evaluate the solution in regions A, B, and C.

It follows directly from Eqs. (5.4) and the fact that

$$Q(u) = Q(v) = S(u) = S(v) = E(u) = E(v) = 0$$

in region C that the solution in region C is ${}_{(2)}\Psi_A = 0$. In region A, the Y_2^2 part of Ψ_0 reduces immediately

to $r^{-2}D[(Q(u)-Q(v))/r]$, where now, since $v > u_o$,

$$\begin{aligned} Q(v) &= Q(u_o) - \frac{1}{4}\int_{u_0}^{v} dx_2 \int_{-\infty}^{x_2} dx_1 \int_{-\infty}^{x_1} \bar{b}(x)\frac{d^6 b}{dx^6}\,dx \\ &= Q(u_o) - \frac{1}{4}\Bigg(\tfrac{1}{2}(v-u_o)^2 \int_{-u_o}^{u_o} \bar{b}\,\frac{d^6 b}{dx^6}\,dx \\ &\qquad + (v-u_o)\int_{-u_o}^{u_o} dx_1 \int_{-u_o}^{x_1} \bar{b}\,\frac{d^6 b}{dx^6}\,dx\Bigg) \\ &= Q(u_o) + (v-u_o)Q'(u_o) + \tfrac{1}{2}(v-u_o)^2 Q''(u_o). \end{aligned}$$

With this form of $Q(v)$, direct calculation shows that

$$r^{-2}D^2(Q(v)/r) = 2Q(u)/r^5.$$

Hence, $r^{-2}D[(Q(u)-Q(v))/r] = 0$. Similar calculations for the Y_4^2 part of ${}_{(2)}\Psi_0$ and for the rest of the ${}_{(2)}\Psi_A$ show that in region A the solution is ${}_{(2)}\Psi_A = 0$.

In region B, we have $Q(u) = S(u) = E(u) = 0$, and $Q(v)$, $S(v)$, and $E(v)$ are the same as they are in region A. We find then for the solution in region B

$$\overset{2}{\Psi}_A = Y_0^0 \frac{M}{r^3}\,\delta_A^2 + K_{2-A}(2)Y_2^{2-A}d^A\left(\frac{q(u)}{r^{5-A}}\right) + K_{2-A}(4)Y_4^{2-A}r^2 d^{2+A}\left(\frac{s(u)}{r^{7-A}}\right), \quad (5.8)$$

where

$$M \equiv \frac{7}{2(5^{\frac{1}{2}})}\,\ddot{q}, \quad (5.9a)$$

$$q(u) \equiv -\frac{5^{\frac{1}{2}}}{42(\pi)^{\frac{1}{2}}}\,[Q(u_o) + (u-u_o)Q'(u_o) + \tfrac{1}{2}(u-u_o)^2 Q''(u_o)], \quad (5.9b)$$

$$s(u) \equiv \frac{1}{84(\pi)^{\frac{1}{2}}}\left(\sum_{n=0}^{4}\frac{(u-u_o)^n}{n!}\,\frac{d^n S}{dv^n}\bigg|_{u_o} + E(u_o)\right). \quad (5.9c)$$

The solution in region B is thus, as expected, a formally linear field with 2^L-pole structure of $L = 0, 2, 4$; it is nonradiative since $\dot{M} = \dddot{q}(u) = d^5 s(u)/du^5 = 0$.

The nontransient part of the solution can be characterized rather simply even in the regions where the pulse is present. By comparing Eqs. (5.4) with Eqs. (4.1), it may be seen that the retarded parts of the $({}_{(2)}\Psi_A - T_A)$ are identical in form to nonradiative-motion fields with monopole, quadrupole, and 16-pole moments proportional to $\ddot{Q}(u)$, $Q(u)$, and $S(u) + E(u)$, respectively. The only difference is that $\ddot{Q}$ and $d^5(S+E)/du^5$ do not vanish in the region where the outgoing pulse is present. The exactly analogous property holds for the parts of the $({}_{(2)}\Psi_A - T_A)$ that depend on the advanced time v; this may be seen by comparison with Eqs. (4.2).

B. The Nonsingular Nature of the Field

The result that the ${}_{(2)}\Psi_A$ are finite at $r = 0$ can be anticipated from the fact that the driving terms D_A and R_{A+1} have only a r^{-1} infinity as $r \to 0$ in the case of retarded-minus-advanced fields. All other powers of $1/r$ cancel to zero when the expansion for $b(u-r)$ in the neighborhood of $r = 0$ is substituted into Eqs. (5.1) and (5.2). It is easy to see that with these driving terms the differential equations, Eqs. (3.9) and (3.10), may be satisfied near $r = 0$ by quantities which are $O(1)$. Hence, the driving terms do not contribute, near $r = 0$, any powers of $1/r$ to a solution and any singularities of a solution must satisfy the homogeneous equations. They are therefore removable by an appropriate formally linear solution.

From Eqs. (2.5), we see that the quantity ${}_{(2)}\Psi_0$ is given by

$$\overset{2}{\Psi}_0 = 2\overset{1}{C}_{\mu\nu\rho\sigma} l^\mu \overset{0}{m}{}^\nu l^\rho \overset{1}{m}{}^\sigma + \overset{2}{C}_{\mu\nu\rho\sigma} l^\mu \overset{0}{m}{}^\nu l^\rho \overset{0}{m}{}^\sigma, \quad (5.10)$$

and that ${}_{(2)}\Psi_1$ through ${}_{(2)}\Psi_4$ are given by similar expressions. In addition to the fact that the ${}_{(2)}\Psi_A$ are finite at $r = 0$, we know from Sec. 4 that ${}_{(1)}C_{\mu\nu\rho\sigma}$ is nonsingular. However, as discussed in Sec. 4, the zeroth-order and first-order tetrads are infinite at $r = 0$. Hence, consideration only of Eq. (5.10) and the similar equations leaves open the possibility that ${}_{(2)}C_{\mu\nu\rho\sigma}$ might be infinite even in the coordinate system in which ${}_{(0)}g_{\mu\nu} + {}_{(1)}g_{\mu\nu}$ is manifestly nonsingular.

We now show that, in fact, ${}_{(2)}C_{\mu\nu\rho\sigma}$ is nonsingular. The first step is to find a tetrad correct to first order which is manifestly nonsingular when expressed in terms of the coordinates (x^0, x^1, x^2, x^3) given by Eq. (4.18). It is a straightforward calculation to check that such a tetrad (in coordinates u, r, θ, ϕ) is given by

$$\overset{0}{V}{}_B^\mu + \overset{1}{V}{}_B^\mu = L_B{}^C(\overset{0}{Z}{}_C^\mu + \overset{1}{Z}{}_C^\mu), \quad (5.11)$$

where ${}_{(1)}Z_C^\mu$ is the first-order null tetrad $(0, {}_{(1)}n^\mu, {}_{(1)}m^\mu, {}_{(1)}\bar{m}^\mu)$. One would expect that Eq. (5.11) would have a term ${}_{(1)}L_B{}^C{}_{(0)}Z_C^\mu$ on the right-hand side, since the tetrad might need a first-order correction. However, it turns out that such a term is not necessary since, as it stands, the tetrad of Eq. (5.11) is nonsingular.

Solving Eq. (5.11) for ${}_{(0)}m^\mu$ and ${}_{(1)}m^\mu$ and substituting the result in Eq. (5.10) gives

$$\overset{2}{\Psi}_0 = (L_2{}^B)^{-1}(L_2{}^D)^{-1} \times (2\overset{1}{C}_{\mu\nu\rho\sigma} l^\mu \overset{0}{V}{}_B^\nu l^\rho \overset{1}{V}{}_D^\sigma + \overset{2}{C}_{\mu\nu\rho\sigma} l^\mu \overset{0}{V}{}_B^\nu l^\rho \overset{0}{V}{}_D^\sigma), \quad (5.12)$$

where $(L_B{}^C)^{-1}L_C{}^D = \delta_B^D$, and similar expressions for ${}_{(2)}\Psi_1$ through ${}_{(2)}\Psi_4$. In this form, the terms involving ${}_{(1)}C_{\mu\nu\rho\sigma}$ are nonsingular and then it follows, from Eq. (4.26) and the fact that the ${}_{(2)}\Psi_A$ are finite at $r = 0$, that ${}_{(2)}C_{\mu\nu\rho\sigma}$ is nonsingular in the coordinates of Eq. (4.18).

Hence, we have established that the second-order field of our problem is nonsingular. Since there is no *a priori* reason to expect that it is possible to obtain a nonsingular second-order field from the retarded-minus-advanced first-order field, the fact that it is possible is an expression of a property of the Einstein field equations.

6. CONCLUSION

With the results obtained in Sec. 5, we are now able to answer the questions raised in the introduction concerning the self-scattering of gravitational radiation. Specifically, does a first-order imploding–exploding pulse of gravitational radiation self-interact in such a way that second-order radiation is produced?

The second-order correction given in Sec. 5 was found to contain no retarded radiation field. This by itself does not mean that no second-order retarded radiation field was generated. It is possible that one was generated, but that it was cancelled by destructive interference with a second-order retarded radiation field already present. However, there seems to be only two possible ways for another second-order retarded radiation field to be present; either it was spontaneously generated along the line $r = 0$, or it began as an advanced second-order radiation field, and after focusing on the line $r = 0$, became a retarded radiation field. However, the possibility that second-order retarded radiation fields were generated along the line $r = 0$ is precluded by the nonsingularity of our second-order correction along $r = 0$ for all u. Likewise, our second-order correction is characterized by vanishing incoming radiation on $\mathfrak{I}^-$, and therefore, no radiation came in from past null infinity and became available for destructive interference at future null infinity. *We are forced to conclude that the nonexistence of retarded radiation in our second-order correction is a consequence of nongeneration of second-order radiation by the first-order pulse.* The distinction between forward, backward, and mixed scattering which was made in the introduction (and which seems to us a meaningful distinction in the context of a perturbative expansion) becomes a distinction of no importance; no scattering of any type occurs.

In order to emphasize the nontriviality of this conclusion, it should be pointed out that scattering *does* occur under different conditions. The problem of a first-order pulse of outgoing quadrupole radiation with a first-order mass at its focus was investigated,[7] *and in this case, back scattering occurs.* The absence of such scattering in the pure radiation problem discussed in this paper was a surprise to us.

The other question raised in the introduction concerned the nature of the second-order correction in regions where the background is Minkowski space–time (regions labeled A, B, or C on Fig. 2). Before and behind the first-order pulse (in regions A or C), the second-order correction vanishes for our solution. In region B, the second-order correction is a formally linear, nonradiative field characterized by a positive monopole moment $-M$, and nonradiative quadrupole and 16-pole moments. These moments are given by Eq. (5.9). The second-order mass is the Bondi mass[13] equivalent to the amount of energy which is carried into the focus by the first-order wave. These nonradiative effects may be understood as follows. Suppose an observer follows the world line indicated by the curved solid line from I^- to I^+ in Fig. 2. From I^- to the point where the world line encounters the incoming pulse, the observer observes a vanishing curvature tensor in his neighborhood. While passing through the incoming pulse, he observes a very complicated field with both first- and second-order parts. In region B, the observer experiences the relatively simple gravitational field characterized by a constant monopole moment and nonradiative quadrupole and 16-pole moments. They are due to the focusing of the ingoing quadrupole wave toward $r = 0$; however, the effects on the observer are indistinguishable from those which would be caused by monopole, quadrupole, and 16-pole moments associated with real singularities at $r = 0$. After the observer observes the outgoing pulse, represented by his world line intersecting the retarded crosshatching, he is again in a space–time with vanishing curvature tensor and no gravitational field.

In conclusion, we should note that the qualitative features described above seem to be quite general features of radiative interactions of the gravitational field, to second-order. Computer calculations[15] have shown that a *nonradiative* second-order correction arises from first-order fields which are any linear combination of 2^l-pole fields for $l = 2, 3, 4$. Most probably, this is true for all $l \geq 2$. On the other hand, it appears likely that if such calculations are continued to third order, *radiative* corrections will appear.

ACKNOWLEDGMENTS

The authors wish to thank Professor P. G. Bergmann, Professor J. N. Goldberg, and Dr. William

Kinnersley for many helpful discussions. One author (W. E. C.) wishes to thank the Centro de Investigacion y de Estudious Avanzados for its hospitality during a visit there.

APPENDIX A

The transient quantities T_A occurring in Eqs. (5.4) are given here. We then have the tetrad components of the Riemann tensor of the second-order field given completely and explicitly in all of space-time. The T_A are

$$
\begin{aligned}
T_0 = {} & \frac{5^{\frac{1}{2}}}{28(6\pi)^{\frac{1}{2}}} Y_2^2\Big[-2\frac{3\dot{a}\ddot{a}-\bar{a}\ddot{a}}{r^5}-\frac{360a\dot{a}}{r^7}+\frac{840aa}{r^8}+\frac{1}{2}D^4\Big(\frac{G}{r}\Big)-r\Big(\frac{b''}{r}\Big)'''+\frac{\bar{a}}{r^2}\Big(\frac{b'''}{r}\Big)''\Big] \\
& + \frac{5^{\frac{1}{2}}}{28(6\pi)^{\frac{1}{2}}} Y_4^2\Big\{\frac{3\dot{a}\ddot{a}-\bar{a}\ddot{a}}{r^5}-14\frac{4\dot{a}\dot{a}-\bar{a}\ddot{a}}{r^6}+8\frac{5a\dot{a}-\bar{a}\dot{a}-30a\dot{a}}{r^7} \\
& + \frac{560aa}{r^8}+\frac{\dot{a}}{2r}\Big(\frac{b''}{r}+\frac{6b'}{r^2}+\frac{6b}{r^3}\Big)'''+aD^4\Big(-\frac{b'}{2r^3}+\frac{4b}{3r^4}\Big)+\frac{\dot{\bar{a}}}{2r}\Big(\frac{b''}{r}+\frac{14b'}{3r^2}+\frac{28b}{3r^3}\Big)''' \\
& - \frac{\bar{a}}{r^2}\Big(\frac{b'''}{2r}+\frac{7b''}{3r^2}+\frac{14b'}{3r^3}\Big)''+\frac{1}{3}D^4\Big[\frac{1}{r}\Big(\frac{bb}{r^2}\Big)'\Big]-\frac{7}{6}r\Big(\frac{b''b}{r^4}\Big)'''-14\frac{(b''b)'}{r^5}\Big\}, \\
T_1 = {} & \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}} Y_2^1\Big\{\frac{\bar{a}a_4-2\dot{a}\ddot{a}}{4r^4}+\frac{3\dot{a}\ddot{a}-\bar{a}\ddot{a}}{r^5}-\frac{15a\ddot{a}}{r^6}-30\frac{a\dot{a}-2a\dot{a}}{r^7}+15\frac{a(2\bar{a}-7a)}{r^8}-\frac{1}{4}\Big(\frac{G}{r^2}\Big)''' \\
& - \frac{\dot{a}}{2r}\Big(\frac{b''}{r^2}\Big)''-\frac{\bar{a}}{2r^2}\Big(\frac{b'''}{r^2}\Big)'+\frac{30a\bar{b}}{r^8}-\frac{3\dot{\bar{a}}}{4r}\Big[\frac{1}{r}\Big(\frac{b}{r}\Big)''\Big]''+\frac{\bar{a}}{r^2}\Big[\frac{1}{r}\Big(\frac{b}{r}\Big)'''\Big]'-\frac{1}{4}\ddot{a}\Big[\frac{1}{r}\Big(\frac{b}{r}\Big)'\Big]'''-\frac{1}{4}\frac{\bar{b}}{r^2}D^5\Big(\frac{b}{r}\Big)\Big\} \\
& + \frac{5^{\frac{1}{2}}}{14(\pi)^{\frac{1}{2}}} Y_4^1\Big\{\frac{2\dot{a}\ddot{a}-\bar{a}a_4}{12r^4}-\frac{3}{2}\frac{3\dot{a}\ddot{a}-\bar{a}\ddot{a}}{r^5}+\frac{21}{2}\frac{4\dot{a}\dot{a}-\bar{a}\ddot{a}}{r^6}+\frac{5a\ddot{a}}{r^6}+2\frac{2\bar{a}\dot{a}-5a\dot{a}+60a\dot{a}}{r^7} \\
& - 10\frac{a(\bar{a}+21a)}{r^8}-\frac{3\dot{a}}{4r}\Big(\frac{b''}{r^2}+\frac{6b'}{r^3}+\frac{6b}{r^4}\Big)''-\frac{3}{2}a\Big(-\frac{b'}{2r^4}+\frac{4b}{3r^5}\Big)''' \\
& - \frac{\dot{\bar{a}}}{4r}\Big(\frac{2b''}{r^2}+\frac{12b'}{r^3}+\frac{26b}{r^4}\Big)''+\frac{\bar{a}}{2r^2}\Big(\frac{5b'''}{6r^2}+\frac{5b''}{r^3}+\frac{10b'}{r^4}-\frac{4b}{r^5}\Big)'-\frac{10a\bar{b}}{r^8} \\
& - \frac{1}{12}\ddot{a}\Big[\frac{1}{r}\Big(\frac{b}{r}\Big)'\Big]'''+\frac{\bar{b}}{12r^2}D^5\Big(\frac{b}{r}\Big)+\frac{7}{4}\Big[\frac{(b''b)''}{r^4}+\frac{12(b''b)'}{r^5}+\frac{60b''b}{r^6}\Big)\Big]-\frac{1}{2}\Big[\frac{1}{r^2}\Big(\frac{bb}{r^2}\Big)'\Big]'''\Big\}, \\
T_2 = {} & \frac{1}{12(\pi)^{\frac{1}{2}}} Y_0^0\Big\{\frac{\dddot{a}\ddot{a}}{4r^3}+3\frac{a\ddot{a}-3\dot{a}\dot{a}}{r^5}-6\frac{4\dot{a}\dot{a}-5a\ddot{a}}{r^6}+30\frac{\bar{a}\dot{a}+3\dot{a}a}{r^7}-\frac{120a\bar{a}}{r^8} \\
& + \frac{3\dot{a}}{r}\Big[\frac{1}{r}\Big(\frac{b}{r^2}\Big)''\Big]'-\frac{5\bar{a}}{r^3}\Big(\frac{b}{r^2}\Big)'''+\frac{3}{2}\ddot{a}\Big[\frac{1}{r}\Big(\frac{b}{r^2}\Big)'\Big]''+\frac{6\dot{a}}{r}\Big(\frac{\bar{b}}{r^5}\Big)' \\
& - \frac{30a}{r^3}\Big(\frac{\bar{b}}{r^4}\Big)'+\frac{1}{4}\ddot{a}\Big[\frac{1}{r}\Big(\frac{\bar{b}}{r}\Big)'\Big]''-\frac{1}{4}r^4\Big(\frac{\bar{b}}{r^5}\Big)'D^4\Big(\frac{b}{r^2}\Big)-\frac{1}{4}\frac{\bar{b}}{r^2}D^5\Big(\frac{b}{r}\Big)\Big\} \\
& + \frac{5^{\frac{1}{2}}}{14(\pi)^{\frac{1}{2}}} Y_2^0\Big\{\frac{\bar{a}a_5-\dot{a}a_4}{12r^3}+\frac{2\dot{a}\ddot{a}-\bar{a}a_4}{2r^4}-\frac{3\dot{a}\ddot{a}-\bar{a}\ddot{a}+3\dot{a}\ddot{a}+a\dddot{a}}{r^5}+4\frac{5a\ddot{a}-\dot{a}\dot{a}}{r^6} \\
& - 2\frac{15a\dot{a}-\bar{a}\dot{a}-15a\dot{\bar{a}}}{r^7}+2\frac{21aa-10a\bar{a}}{r^8}+\frac{1}{2}\Big(\frac{G}{r^3}\Big)''+\frac{10a}{r^9}(r^2\bar{b})' \\
& - \frac{2\dot{a}}{r^7}(r\bar{b})'+\frac{\ddot{a}}{12r^3}\Big[2r^3\Big(\frac{4b'}{r^3}+\frac{5b}{r^4}\Big)''-r^2\Big(\frac{b''}{r^2}+\frac{6b'}{r^3}+\frac{10b}{r^4}\Big)'\Big] \\
& + \frac{1}{12}\dddot{a}\Big[\frac{1}{r}\Big(\frac{b}{r}\Big)'\Big]''-\frac{\bar{a}}{3r^3}\Big[\Big(\frac{b}{r^2}\Big)'''-\frac{3b'''}{r^2}+\frac{6}{r}\Big(\frac{b}{r}\Big)'''\Big] \\
& + \frac{\dot{\bar{a}}}{2r}\Big[\frac{1}{r}\Big(\frac{b}{r^2}\Big)''-\frac{2b''}{r^3}+\frac{3}{r^2}\Big(\frac{b}{r}\Big)''\Big]'+\frac{1}{12}\Big[\frac{\bar{b}}{r}D^5\Big(\frac{b}{r^2}\Big)+r^3\Big(\frac{\bar{b}}{r^4}\Big)'D^4\Big(\frac{b}{r^2}\Big)\Big]\Big\} \\
& + \frac{1}{14(\pi)^{\frac{1}{2}}} Y_4^0\Big\{\frac{\dot{a}a_4-\bar{a}a_5}{24r^3}-\frac{5}{6}\frac{2\dot{a}\ddot{a}-\bar{a}a_4}{r^4}+\frac{45\dot{a}\ddot{a}-15\bar{a}\ddot{a}+17\dot{a}\ddot{a}+a\dddot{a}}{2r^5}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{35(4\dot{a}\dot{a} - \bar{a}\ddot{a}) + 45a\ddot{a} - 16\dot{a}\dot{\bar{a}}}{r^6} - 5\,\frac{7a\dot{\bar{a}} + \bar{a}\dot{a} + 60a\dot{a}}{r^7} + \frac{20(21aa + 4a\bar{a})}{r^8} \\
&+ \frac{\ddot{a}}{24}\left[\frac{1}{r}\left(\frac{b}{r}\right)'\right]'' - \frac{\ddot{a}}{24}\left[\frac{17}{r}\left(\frac{b}{r^2}\right)' + \frac{20}{r^2}\left(\frac{b}{r}\right)'\right]'' + \frac{15\dot{a}}{2r}\left(\frac{b''}{r^3} + \frac{6b'}{r^4} + \frac{6b}{r^5}\right) \\
&- 2\dot{a}\left[\frac{8\bar{b}}{r^7} - \frac{1}{2r^5}\left(\frac{\bar{b}}{r}\right)'\right] + 15a\left(-\frac{b'}{2r^5} + \frac{4b}{3r^6}\right)'' + 5a\left(-\frac{\bar{b}'}{r^7} + \frac{16\bar{b}}{r^8}\right) \\
&+ \frac{\dot{\bar{a}}}{r}\left(\frac{3b''}{r^3} + \frac{22b'}{r^4} + \frac{53b}{r^5}\right)' - 5\bar{a}\left(\frac{b'''}{3r^5} + \frac{2b''}{r^6} + \frac{b'}{r^7} - \frac{16b}{r^8}\right) \\
&- \frac{35}{2}\left[\frac{(b''b)'}{r^5} + \frac{10b''b}{r^6}\right] + 5\left[\frac{1}{r^3}\left(\frac{bb}{r^2}\right)'\right]'' + \frac{1}{12}D^4\left(\frac{b}{r^2}\right)\left[\frac{8\bar{b}}{r^2} - \frac{1}{2r^3}\left(\frac{\bar{b}}{r}\right)'\right] - \frac{\bar{b}}{24r^2}\,D^5\left(\frac{b}{r}\right)\Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
T_3 = \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}}\,Y_2^{-1}\Bigg\{&\frac{\dddot{a}a_4 + \bar{a}a_5}{4r^3} - \frac{3}{4}\,\frac{\dddot{a}\ddot{a} - \bar{a}a_4 + 2\dot{a}\ddot{a}}{r^4} + \frac{9\dddot{a}\ddot{a} - 2\bar{a}\ddot{a} + 6\dot{a}\ddot{a} + 10a\ddot{a}}{2r^5} - 3\,\frac{2\dddot{a}\dot{a} + 5a\ddot{a} + \bar{a}\ddot{a}}{r^6} \\
&+ \frac{18\dot{a}(a + \bar{a})}{r^7} - \frac{3a(7a + 10\bar{a})}{r^8} - \frac{3}{2}\left(\frac{G}{r^4}\right)' \\
&- \frac{3\dddot{a}}{2r^2}\left[-\left(\frac{b}{r^3}\right)'' + \frac{1}{r}\left(\frac{b}{r^2}\right)'' + \frac{3}{r^2}\left(\frac{b}{r}\right)'' + \frac{2b''}{r^3}\right] + \frac{1}{4}\,\ddot{a}\left[\frac{2}{r}\left(\frac{b}{r^2}\right)' + \frac{1}{r^2}\left(\frac{b}{r}\right)'\right]' \\
&- \frac{\ddot{a}}{4r^2}\left[\frac{6\bar{b}}{r^4} + \frac{6}{r^2}\left(\frac{\bar{b}}{r}\right)' + \frac{4}{r}\left(\frac{b}{r^2}\right)' + \frac{1}{r}\left(\frac{b}{r}\right)'' + 2r^2\left(\frac{4b'}{r^4} + \frac{5b}{r^5}\right)'\right] \\
&- \frac{30a}{r^6}\left(\frac{\bar{b}}{r}\right)' + 6\dot{a}\left(\frac{2\bar{b}'}{r^6} + \frac{3\bar{b}}{r^7}\right) - 6\bar{a}\left(\frac{2b'}{r^7} + \frac{5b}{r^8}\right) \\
&+ \frac{1}{4r^2}\left[-2\bar{b}\left(\frac{b}{r^3}\right)''' - 2r\left(\frac{\bar{b}}{r}\right)'\left(\frac{b}{r^2}\right)''' - 6r\left(\frac{\bar{b}}{r}\right)'\left(\frac{b}{r^3}\right)'' + \bar{b}D^4\left(\frac{b}{r^2}\right)\right]\Bigg\} \\
&+ \frac{5^{\frac{1}{2}}}{7(\pi)^{\frac{1}{2}}}\,Y_4^{-1}\Bigg\{\frac{32}{3}\,\frac{\dddot{a}a_4 - \bar{a}a_5}{r^3} + \frac{10\dddot{a}\ddot{a} - 5\bar{a}a_4 + 3\dddot{a}\ddot{a} + {}^{2}{}_{2}{}^{1}\ddot{a}\ddot{a}}{12r^4} + \frac{-12\dddot{a}\ddot{a} + 5\bar{a}\ddot{a} - 5a\ddot{a} - {}^{1}{}_{3}{}^{1}\dddot{a}\ddot{a}}{2r^5} \\
&+ 2\,\frac{11\dddot{a}\dot{a} - 5\bar{a}\ddot{a} + 10a\ddot{a}}{r^6} + \frac{25a\dddot{a} + 13\bar{a}\dot{a} + 60a\dot{a}}{r^7} - 10\,\frac{a(7a + 3\bar{a})}{r^8} \\
&- \frac{\ddot{a}}{24}\left[\frac{1}{r^2}\left(\frac{b}{r}\right)' + \frac{2}{r}\left(\frac{b}{r^2}\right)'\right]' - \frac{15}{2}\,\dot{a}\left(\frac{b''}{r^5} + \frac{6b'}{r^6} + \frac{6b}{r^7}\right) \\
&+ \frac{1}{24}\,\ddot{a}\left[\frac{17}{r^2}\left(\frac{b}{r^2}\right)' + \frac{20}{r^3}\left(\frac{b}{r}\right)' + \frac{14}{r}\left(\frac{b}{r^3}\right)'\right]' + \frac{1}{4}\,\ddot{a}\left[\frac{1}{r^4}\left(\frac{\bar{b}}{r}\right)' - \frac{6\bar{b}}{r^6}\right] \\
&+ 15a\left[\frac{b'}{2r^6} - \frac{4b}{3r^7}\right]' - \frac{1}{2}\,\dddot{a}\left(\frac{3b''}{r^5} + \frac{26b'}{r^6} + \frac{70b}{r^7}\right) + 2\dot{a}\left[\frac{13\bar{b}}{2r^7} - \frac{1}{r^5}\left(\frac{\bar{b}}{r}\right)'\right] \\
&- 5a\left[\frac{7\bar{b}}{r^8} - \frac{1}{r^6}\left(\frac{\bar{b}}{r}\right)'\right] - \frac{5\bar{a}}{r}\left(\frac{b}{r^6}\right)' + \frac{35b''b}{2r^6} - 5\left[\frac{1}{r^4}\left(\frac{bb}{r^2}\right)'\right]' \\
&- \frac{r^2}{12}\left(\frac{b}{r^3}\right)'''\left[\frac{6\bar{b}}{r^4} - \frac{1}{r^2}\left(\frac{\bar{b}}{r}\right)'\right] - \frac{\bar{b}}{24r^2}\,D^4\left(\frac{b}{r^2}\right)\Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
T_4 = \frac{5^{\frac{1}{2}}}{7(6\pi)^{\frac{1}{2}}}\,Y_2^{-2}\Bigg\{&\frac{\bar{a}a_5 - \dot{a}a_4 - 3\dddot{a}\ddot{a}}{4r^3} + \frac{4\dddot{a}\ddot{a} - \bar{a}a_4 - 3\dddot{a}\ddot{a} + 7\dot{a}\ddot{a}}{2r^4} - \frac{9\dddot{a}\dot{a} + \bar{a}\ddot{a} - 3\dot{a}\ddot{a} + 7a\ddot{a}}{r^5} \\
&+ 6\,\frac{4\dddot{a}\dot{a} + 2\bar{a}\ddot{a}}{r^6} - 6\,\frac{a\dot{a} + 7\bar{a}\dot{a} + 5a\dot{\bar{a}}}{r^7} + 12\,\frac{a(a + 10\bar{a})}{r^8} + \frac{3G}{r^5} \\
&+ \frac{3\ddot{a}}{2r}\left(\frac{3\bar{b}'}{r^4} + \frac{8\bar{b}}{r^5}\right) + \frac{3\ddot{a}b'}{r^5} + 30a\left(\frac{b'}{r^7} + \frac{2b}{r^8}\right) - 6\dot{a}\left(\frac{3b'}{r^6} + \frac{7b}{r^7}\right) \\
&+ 6\dot{\bar{a}}\left(\frac{b'}{r^6} - \frac{5b}{r^7}\right) + \frac{1}{2}\,\ddot{a}\left(\frac{4b}{r^5} - \frac{\bar{b}'}{r^4} - \frac{3\bar{b}}{r^5}\right) + 6\bar{a}\left(\frac{3b'}{r^7} + \frac{10b}{r^8}\right) + \frac{3\bar{b}}{r^2}\left(\frac{b}{r^4}\right)''
\end{aligned}
$$

$$+3b\left(\frac{\bar{b}'}{r^7}+\frac{\bar{b}}{r^8}\right)+\frac{3}{r^2}\left(\frac{b}{r^3}\right)'\left(\frac{\bar{b}}{r}\right)'-\frac{\bar{b}}{2r^2}\left(\frac{b}{r^3}\right)'''+\frac{3}{2r}\left(\frac{b}{r^3}\right)''\left(\frac{\bar{b}}{r}\right)'\Bigg\}$$
$$+\frac{(5)^{\frac{1}{2}}(2)^{\frac{1}{2}}}{7(\pi)^{\frac{1}{2}}}Y_4^{-2}\Bigg\{\frac{3\dot{a}\bar{a}_4-3\bar{a}a_5+5\ddot{a}\ddot{a}}{12r^3}+\frac{5\bar{a}a_4-6\dot{a}\ddot{a}-12\dot{a}\ddot{a}+27\dot{a}\dot{a}}{12r^4}$$
$$+\frac{9\dot{a}\dot{a}-20\bar{a}\ddot{a}/3+12a\ddot{a}-22\dot{a}\dot{a}}{4r^5}+\frac{6\bar{a}\dot{a}-12\dot{a}\dot{\bar{a}}+35a\dot{a}}{2r^6}+\frac{19\bar{a}\dot{a}-45a\dot{\bar{a}}+60a\dot{a}}{2r^7}$$
$$+\frac{10a(3a-2\bar{a})}{r^8}+\frac{1}{4}\dddot{a}\left(\frac{3b'}{r^4}+\frac{7b}{r^5}\right)-\frac{1}{4}\ddot{a}\left(\frac{70b'}{r^5}+\frac{31b}{r^6}\right)+\frac{1}{24}\dddot{a}\left(\frac{3\bar{b}'}{r^4}-\frac{5\bar{b}}{r^5}\right)$$
$$-45a\left(\frac{b'}{2r^7}-\frac{4b}{3r^8}\right)+\frac{45}{2}\dot{a}\left(\frac{b'}{r^6}+\frac{b}{r^7}\right)-\frac{9}{8}\ddot{a}\left(\frac{\bar{b}'}{r^5}-\frac{2\bar{b}}{r^6}\right)+\frac{3}{2}\dot{a}\left(\frac{3\bar{b}'}{r^6}-\frac{7\bar{b}}{r^7}\right)$$
$$-\frac{5}{2}a\left(\frac{\bar{b}'}{r^7}-\frac{8\bar{b}}{r^8}\right)+\frac{5}{2}\bar{a}r\left(\frac{b}{r^8}\right)'+\frac{3\dot{a}}{2r}\left(\frac{b}{r^5}\right)'+\frac{15}{r^5}\left(\frac{bb}{r^2}\right)'-\frac{1}{8}\left(\frac{b}{r^4}\right)''\left(\frac{3\bar{b}'}{r}-\frac{5\bar{b}}{r^2}\right)+\frac{1}{8}\frac{\bar{b}}{r^2}\left(\frac{b}{r^3}\right)'''\Bigg\},$$

where

$$G\equiv-2\dot{a}\left(\frac{b}{r}\right)'+2a\left(\frac{b}{r^2}\right)'+\left(\frac{bb}{r^2}\right)'.$$

APPENDIX B

We give here the calculation establishing that the field given by Eqs. (5.4) contains no incoming radiation.

The backward null surfaces, that is, the surfaces whose normal vector $V_{,\mu}$ satisfies $V_{,\mu}V_{,\nu}g^{\mu\nu}=0$ and $l^\mu V_{,\mu}\neq 0$, are given, correct to first order, by

$$V=\text{const},$$
$$V\equiv u-r+\tfrac{1}{2}Y_2^0\left(\frac{b+\bar{b}}{r^2}-\frac{\dot{a}+\dot{\bar{a}}}{r}+\frac{a+\bar{a}}{r^2}\right). \tag{B1}$$

The vector $\tilde{n}^\mu$, correct to first order, that bears the same relation to the surfaces given by Eq. (B1) that l^μ bears to the surfaces $u=\text{const}$ is given by

$$\tilde{n}^\mu=\overset{0}{n}{}^\mu+\overset{1}{n}{}^\mu+L\overset{0}{n}{}^\mu+P\overset{0}{m}{}^\mu+\bar{P}\overset{0}{\bar{m}}{}^\mu, \tag{B2}$$

where

$$L\equiv-\tfrac{1}{2}Y_2^0\left(-\frac{\dot{a}+\dot{\bar{a}}}{r^2}+2\frac{a+\bar{a}}{r^3}+\frac{b'+\bar{b}'}{r^2}+2\frac{b+\bar{b}}{r^3}\right),$$
$$P\equiv\frac{6^{\frac{1}{2}}}{6}Y_2^1\left(\frac{\ddot{a}}{2r}-\frac{3\dot{a}}{2r^2}+\frac{3a-\bar{a}}{2r^3}-\frac{b'}{2r^2}-\frac{\bar{b}-3b}{2r^3}\right). \tag{B3}$$

That $\tilde{n}^\mu$ has this relation to the backward null surfaces may be checked directly by transforming $\tilde{n}^\mu$, given by Eq. (B2), to the coordinate system (u',r',θ',ϕ'), which is constructed from the surfaces $V=\text{const}$ in the same manner that the coordinates u, r, θ, and ϕ are constructed from the surfaces $u=\text{const}$. This coordinate system is given by

$$u'=V, \tag{B4a}$$
$$r'=r-\tfrac{1}{2}Y_2^0\left(\frac{b'+\bar{b}'}{r}+\frac{\dot{a}+\dot{\bar{a}}}{r}\right), \tag{B4b}$$
$$\theta'=\theta+\frac{6^{\frac{1}{2}}}{6}Y_2^1\left(\frac{b'+\bar{b}'}{r^2}+\frac{\dot{a}+\dot{\bar{a}}}{r^2}\right), \tag{B4c}$$
$$\phi'=-\phi+\frac{6^{\frac{1}{2}}}{6}\frac{i}{\sin\theta}\times Y_2^1\left(\frac{b'-\bar{b}'}{r^2}+\frac{\dot{\bar{a}}-\dot{a}}{r^2}-2\frac{\bar{a}-a}{r^3}\right) \tag{B4d}$$

For example, it is easy to show that $\tilde{n}^{\mu'}=\delta^\mu_1$.

The remainder of the tetrad constructed around $\tilde{n}^\mu$ in the same manner as the original tetrad is constructed around l^μ is given, correct to first order, by

$$\tilde{l}^\mu=l^\mu+\tfrac{1}{2}Ll^\mu+N\overset{0}{m}{}^\mu+\bar{N}\overset{0}{\bar{m}}{}^\mu,$$
$$\tilde{m}^\mu=Pl^\mu+N\overset{0}{n}{}^\mu+\overset{0}{\bar{m}}{}^\mu+\overset{1}{\bar{m}}{}^\mu+M\overset{0}{\bar{m}}{}^\mu, \tag{B5}$$

where

$$N\equiv-\frac{6^{\frac{1}{2}}}{6}Y_2^1\left(-\frac{\dot{a}}{r^2}+\frac{a-3\bar{a}}{r^3}-\frac{\bar{b}''}{r}-\frac{3\bar{b}'}{r^2}+\frac{3\bar{b}-b}{r^3}\right),$$
$$M\equiv\frac{6^{\frac{1}{2}}}{6}Y_2^1\left(-\frac{\dot{a}-\dot{\bar{a}}}{r^2}+2\frac{a-\bar{a}}{r^3}+\frac{b'-\bar{b}'}{r^2}+2\frac{b-\bar{b}}{r^3}\right). \tag{B6}$$

It may be seen by direct calculation that the tetrad $\tilde{n}^{\mu'}$, $\tilde{l}^{\mu'}$, $\tilde{m}^{\mu'}$, and $\bar{\tilde{m}}^{\mu'}$ satisfies all the tetrad and coordinate conditions imposed in Sec. 2. Further

details concerning the coordinates and tetrad built around the backward null surfaces are contained in Ref. 9, Sec. 4.

From Eq. (2.5b), it is seen that ${}_{(2)}\tilde{\Psi}_1$ is given by

$$\overset{2}{\tilde{\Psi}}_1 = \overset{2}{\Psi}_1 + \overset{1}{\Psi}_1(\bar{M} - L) + \overset{1}{\Psi}_0 P + 3\bar{N}\overset{1}{\Psi}_1. \quad \text{(B7)}$$

We want to evaluate $r'^2{}_{(2)}\tilde{\Psi}_1$, as $r' \to \infty$ with $V =$ const, that is, at $\mathfrak{I}^-$. All terms involving $a(u)$ will not enter into this evaluation, since $a(u)$ vanishes at $\mathfrak{I}^-$. From Eqs. (B1) and (B4b), we see that $V = u - r + O(r^{-1})$ and $r' = r + O(r^{-1})$ for $r \to \infty$, $u - r =$ const. Hence, we may go to $\mathfrak{I}^-$ by taking the limit $r \to \infty$, $u - r =$ const. In this limit, we have from Eqs. (B3) and (B6) that

$$\bar{M} - L = O(r^{-2}) = P, \quad N = O(r^{-1}),$$

and from Eqs. (4.2) that

$$\overset{1}{\Psi}_0 = O(r^{-1}), \quad \overset{1}{\Psi}_1 = O(r^{-2}), \quad \overset{1}{\Psi}_2 = O(r^{-3}).$$

Hence, from Eq. (B7), we have

$$\overset{2}{\tilde{\Psi}}_1 = \overset{2}{\Psi}_1 + O(r^{-3}).$$

Therefore, we have

$$\overset{2}{\dot{\sigma}}{}^{\circ}_{\text{in}} \equiv \lim_{r\to\infty}(r^2\overset{2}{\tilde{\Psi}}_1), \quad u - r = \text{const},$$

$$= \lim_{r\to\infty}(r^2\overset{2}{\tilde{\Psi}}_1), \quad u - r = \text{const}.$$

By using ${}_{(2)}\Psi_1$, as given by Eq. (5.4b) and Appendix A, this latter limit is easily evaluated and found to be zero. Hence, the second-order news function for incoming radiation vanishes.

We have considered the tetrad $\tilde{n}^\mu$, $\tilde{l}^\mu$, $\tilde{m}^\mu$, and $\tilde{\bar{m}}^\mu$ only to first order: This is sufficient because second-order parts of the tetrad do not enter into ${}_{(2)}\tilde{\Psi}_A$.

APPENDIX C

The coordinate and tetrad conventions of this paper differ from those of preceding papers (Refs. 4, 5, 7–9) by the coordinate and tetrad transformations

$$\begin{aligned} l^\mu &= \tilde{l}^\mu(2)^{-\frac{1}{2}}, \quad n^\mu = \tilde{n}^\mu(2)^{\frac{1}{2}}, \\ r &= (2)^{\frac{1}{2}}r', \qquad u = -u'(2)^{-\frac{1}{2}}, \end{aligned} \quad \text{(C1)}$$

where the primed and tilded quantities are those used in the preceding papers. The consequent changes in various fundamental variables are collected here: We have

$$\begin{aligned} \Psi_0 &= \tfrac{1}{2}\tilde{\Psi}_0, \quad \Psi_1 = \tilde{\Psi}_1(2)^{-\frac{1}{2}}, \quad \Psi_2 = \tilde{\Psi}_2, \\ \Psi_3 &= 2^{\frac{1}{2}}\tilde{\Psi}_3, \quad \Psi_4 = 2\tilde{\Psi}_4, \end{aligned} \quad \text{(C2)}$$

and

$$\begin{aligned} \kappa &= \tfrac{1}{2}\tilde{\kappa}, & \pi &= \pi, & \epsilon &= \tilde{\epsilon}(2)^{-\frac{1}{2}}, \\ \rho &= \tilde{\rho}(2)^{-\frac{1}{2}}, & \lambda &= (2)^{\frac{1}{2}}\tilde{\lambda}, & \alpha &= \tilde{\alpha}, \\ \sigma &= \tilde{\sigma}(2)^{-\frac{1}{2}}, & \mu &= (2)^{\frac{1}{2}}\tilde{\mu}, & \beta &= \tilde{\beta}, \\ \nu &= 2\tilde{\nu}, & \gamma &= (2)^{\frac{1}{2}}\tilde{\gamma}, & \tau &= \tilde{\tau}, \end{aligned} \quad \text{(C3)}$$

along with

$$D = \tilde{D}(2)^{-\frac{1}{2}}, \quad \Delta = (2)^{\frac{1}{2}}\tilde{\Delta}, \quad \delta = \tilde{\delta}, \quad \text{(C4)}$$

and

$$\begin{aligned} U &= 2U', & X^i &= (2)^{\frac{1}{2}}X^{i'}, \\ \omega &= (2)^{\frac{1}{2}}\omega', & \xi^i &= \xi^{i'}. \end{aligned} \quad \text{(C5)}$$

As a consequence, the constants of integration change, and are related to the old ones by

$$\begin{aligned} &\sigma^\circ = (2)^{\frac{1}{2}}\tilde{\sigma}^{\circ\prime}, && \xi^{i\circ} = (2)^{\frac{1}{2}}\tilde{\xi}^{i\circ\prime}, && \alpha^\circ = (2)^{\frac{1}{2}}\tilde{\alpha}^{\circ\prime}, \\ &\beta^\circ = (2)^{\frac{1}{2}}\tilde{\beta}^{\circ\prime}, && \tau^\circ = (2)^{\frac{1}{2}}\tilde{\tau}^{\circ\prime}, && \omega^\circ = 2\tilde{\omega}^{\circ\prime}, \\ &\lambda^\circ = 2\tilde{\lambda}^{\circ\prime}, && u^\circ = 2\tilde{u}^{\circ\prime}, && \gamma^\circ = (2)^{\frac{1}{2}}\tilde{\gamma}^{\circ\prime}, \\ &\nu^\circ = 2\tilde{\nu}^{\circ\prime}, && U^\circ = 2\tilde{U}^{\circ\prime}, && \Psi_1^\circ = 2(2)^{\frac{1}{2}}\tilde{\Psi}_1^{\circ\prime}, \\ &\Psi_2^\circ = 2(2)^{\frac{1}{2}}\tilde{\Psi}_2^{\circ\prime}, && \Psi_3^\circ = 2(2)^{\frac{1}{2}}\tilde{\Psi}_3^{\circ\prime}, && \Psi_4^\circ = 2(2)^{\frac{1}{2}}\tilde{\Psi}_4^{\circ\prime}. \end{aligned} \quad \text{(C6)}$$

Changes in various equations appearing in the older papers and in this one can be determined by direct comparison.

* Supported in part by Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force; Office of Scientific Research; National Science Foundation; and National Research Council of Canada.

† This work was begun when one author (W. E. C.) was at Syracuse University and the other author (R. J. T.) was at the Centro de Investigacion y de Estudious Avanzados del Instituto Politechnico Nacional, Mexico.

[1] (a) Tensor indices denoted by Greek letters range and sum from 0 to 3. Tensor indices denoted by lower case Latin letters range and sum from 2 to 3. The letters *B*, *C*, *D*, and *E* denote tetrad indices and range and sum from 0 to 3. The letter *A* used as an index will range and sum from 0 to 4 unless otherwise specified. Ordinary partial differentiation is denoted in the usual way or by a comma, and covariant differentiation is denoted by a semicolon. (b) In the text, we use a left subscript in parentheses as an alternative notation for the perturbative order. For example, ${}_{(1)}h_{\mu\nu}$ is the same as $\overset{1}{h}_{\mu\nu}$.

[2] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Englewood Cliffs, N.J., 1942).

[3] R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

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

[5] E. Newman and T. Unti, J. Math. Phys. **3**, 891 (1962).

[6] For convenience, we have in this paper made changes in some coordinate and tetrad conventions from what has hitherto been used (i.e., in Refs. 4, 5, 7–9). This paper is self-contained, but comparisons with earlier papers must be made with care. The changes are discussed in Appendix C.

[7] W. E. Couch, R. J. Torrence, A. I. Janis, and E. T. Newman, J. Math. Phys. **9**, 484 (1968).

[8] A. Janis and E. Newman, J. Math. Phys. **6**, 902 (1965).

[9] R. Torrence and A. Janis, J. Math. Phys. **8**, 1355 (1967).

[10] E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).

[11] We define $f(x) = O(g(x))$ as $x \to \infty$ to mean that there exist Λ and ζ independent of x such that $f(x) \leq \Lambda g(x)$ for $x > \zeta$, and similarly $f(x) = O(g(x))$ as $x \to 0$ if $f(x) \leq \Lambda g(x)$ for $|x| < \zeta$.

[12] The quantities ${}_sY_{lm}$ are a complete set of spin-weighted functions of spin weight s obtained from the spherical harmonics,

$$Y_{lm} = {}_0Y_{lm}, \quad \text{by} \quad {}_sY_{lm} = K_{-s}(l)\eth^s Y_{lm}, \qquad 0 \leq s \leq l,$$
$$= (-1)^s K_s(l)\bar{\eth}^{-s} Y_{lm}, \quad -l \leq s \leq 0.$$

The operators $\eth$ and $\bar{\eth}$ are defined by

$$\eth\eta = -(\sin\theta)^s\left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)[(\sin\theta)^{-s}\eta]$$

and

$$\bar{\eth}\eta = -(\sin\theta)^{-s}\left(\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)[(\sin\theta)^{s}\eta],$$

where η has spin weight s. The most commonly used relations in calculations of this paper are $\eth_s Y_{lm} = K_{-s}(l)K_{s+1}(l)_{s+1}Y_{lm}$ and $\bar{\eth}_s Y_{lm} = -K_{1-s}(l)K_s(l)_{s-1}Y_{lm}$. The ${}_{(1)}\Psi_A$ and ${}_{(2)}\Psi_A$ have spin weights $2 - A$, ${}_{(1)}\sigma^\circ$ and ${}_{(2)}\sigma^\circ$ have spin weight 2. It is helpful to adopt the convention ${}_sY_{lm} = 0$ whenever $|s| > l$. We often use the abbreviation $Y_l^s \equiv {}_sY_{l0}$. For more details, see Ref. 7, Appendix A, and references given there.

[13] H. Bondi, M. Van der Burg, A. Metzner, Proc. Roy. Soc. (London) A **269**, 21 (1962).

[14] This definition differs only by angular factors (in each term of an expansion in ${}_sY_{l0}$) from that of Newman and Unti or of Bondi. This is immaterial because it turns out that we need only consider the vanishing or nonvanishing of the news.

[15] The authors are indebted to Dr. William Kinnersley for these calculations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Some Simple Lattice-Spin Systems

JONG-WON LIEH
Physics Department, Saint Francis College, Loretto, Pennsylvania 15940

(Received 18 December 1969)

A simple mathematical scheme is derived here. With this scheme, three problems of lattice-spin systems are solved exactly. The first one is the problem of solving, thermodynamically, a linear chain of the Lenz–Ising model in a zero magnetic field with nearest, as well as next-nearest, neighbor couplings. The problem turns out to be equivalent to the problem of a linear chain with only nearest-neighbor couplings but in a finite magnetic field. The second one is to solve an imperfect 1-dimensional Heisenberg–Dirac model, similar to the partially solved "Ising–Heisenberg" model of Lieb–Mattis–Schultz, in a zero magnetic field. The problem is solved completely in the sense that all the elementary excitations of this model are shown in terms of some pseudofermions and the spectra are given as

$$\epsilon_1(q) = [(A_q + B_q)^2 + E_q^2]^{\frac{1}{2}},$$
$$\epsilon_2(q) = [(A_q - B_q)^2 + E_q^2]^{\frac{1}{2}},$$

where A_q, B_q, and E_q are three different functions of sin (q), cos (q), and coupling strengths involved. The third one, the XY model, is used to study the contribution of "inhomogeneity" in the coupling strengths to the system, as compared to the anisotropicity contribution to it.

I. MATHEMATICAL PRELIMINARIES

We first derive here and put in order some of the mathematical formulas to be used in this calculation. The main guiding line in this study is to transform spin-pair operators to fermion-pair operators. Such process has already been used by Mattis *et al.*,[1] by the way of a so-called "drone operator" in order to simplify the stupendous solution of the Lenz–Ising model of Onsager and Kaufmann and to solve some other simple lattice-spin systems. However, we follow here a little different path by looking into the commutation relations of these two kinds of operators. Spin operators obey a mixture of anticommutability and commutability; that is to say, spin operators of the same site anticommute with one another, while they commute with spins of other sites. Such a mixture in commutation relation causes the main mathematical difficulty[2] in handling a lattice-spin system. On the other hand, fermion operators follow a simple rule, namely the anticommutation rule, and this fact makes a Fourier transformation to reduce a complex spin system possible.

Spin operators anticommute among themselves as

$$\{S_m^{(i)}, S_m^{(j)}\} = 2\delta_{ij}, \quad i, j = x, y, z,$$

because

$$\mathbf{S}_m \times \mathbf{S}_m = 2i\mathbf{S}_m$$

and

$$[S_m^{(i)}]^2 = 1,$$

where m is the index to show different spins. But spins of different lattice sites commute with one another. On the other hand, fermion operators follow the anticommutation rule

$$\{C_m, C_n\} = \{C_m^+, C_n^+\} = 0$$

and

$$\{C_m, C_n^+\} = \delta_{mn}.$$

[12] The quantities ${}_sY_{lm}$ are a complete set of spin-weighted functions of spin weight s obtained from the spherical harmonics,

$$Y_{lm} = {}_0Y_{lm}, \quad \text{by} \quad {}_sY_{lm} = K_{-s}(l)\eth^s Y_{lm}, \qquad 0 \leq s \leq l,$$
$$= (-1)^s K_s(l)\bar{\eth}^{-s} Y_{lm}, \quad -l \leq s \leq 0.$$

The operators $\eth$ and $\bar{\eth}$ are defined by

$$\eth\eta = -(\sin\theta)^s\left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)[(\sin\theta)^{-s}\eta]$$

and

$$\bar{\eth}\eta = -(\sin\theta)^{-s}\left(\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta}\frac{\partial}{\partial\phi}\right)[(\sin\theta)^{s}\eta],$$

where η has spin weight s. The most commonly used relations in calculations of this paper are $\eth_s Y_{lm} = K_{-s}(l)K_{s+1}(l)_{s+1}Y_{lm}$ and $\bar{\eth}_s Y_{lm} = -K_{1-s}(l)K_s(l)_{s-1}Y_{lm}$. The ${}_{(1)}\Psi_A$ and ${}_{(2)}\Psi_A$ have spin weights $2 - A$, ${}_{(1)}\sigma^\circ$ and ${}_{(2)}\sigma^\circ$ have spin weight 2. It is helpful to adopt the convention ${}_sY_{lm} = 0$ whenever $|s| > l$. We often use the abbreviation $Y_l^s \equiv {}_sY_{l0}$. For more details, see Ref. 7, Appendix A, and references given there.

[13] H. Bondi, M. Van der Burg, A. Metzner, Proc. Roy. Soc. (London) A **269**, 21 (1962).

[14] This definition differs only by angular factors (in each term of an expansion in ${}_sY_{l0}$) from that of Newman and Unti or of Bondi. This is immaterial because it turns out that we need only consider the vanishing or nonvanishing of the news.

[15] The authors are indebted to Dr. William Kinnersley for these calculations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Some Simple Lattice-Spin Systems

JONG-WON LIEH
Physics Department, Saint Francis College, Loretto, Pennsylvania 15940

(Received 18 December 1969)

A simple mathematical scheme is derived here. With this scheme, three problems of lattice-spin systems are solved exactly. The first one is the problem of solving, thermodynamically, a linear chain of the Lenz–Ising model in a zero magnetic field with nearest, as well as next-nearest, neighbor couplings. The problem turns out to be equivalent to the problem of a linear chain with only nearest-neighbor couplings but in a finite magnetic field. The second one is to solve an imperfect 1-dimensional Heisenberg–Dirac model, similar to the partially solved "Ising–Heisenberg" model of Lieb–Mattis–Schultz, in a zero magnetic field. The problem is solved completely in the sense that all the elementary excitations of this model are shown in terms of some pseudofermions and the spectra are given as

$$\epsilon_1(q) = [(A_q + B_q)^2 + E_q^2]^{\frac{1}{2}},$$
$$\epsilon_2(q) = [(A_q - B_q)^2 + E_q^2]^{\frac{1}{2}},$$

where A_q, B_q, and E_q are three different functions of sin (q), cos (q), and coupling strengths involved. The third one, the XY model, is used to study the contribution of "inhomogeneity" in the coupling strengths to the system, as compared to the anisotropicity contribution to it.

I. MATHEMATICAL PRELIMINARIES

We first derive here and put in order some of the mathematical formulas to be used in this calculation. The main guiding line in this study is to transform spin-pair operators to fermion-pair operators. Such process has already been used by Mattis *et al.*,[1] by the way of a so-called "drone operator" in order to simplify the stupendous solution of the Lenz–Ising model of Onsager and Kaufmann and to solve some other simple lattice-spin systems. However, we follow here a little different path by looking into the commutation relations of these two kinds of operators. Spin operators obey a mixture of anticommutability and commutability; that is to say, spin operators of the same site anticommute with one another, while they commute with spins of other sites. Such a mixture in commutation relation causes the main mathematical difficulty[2] in handling a lattice-spin system. On the other hand, fermion operators follow a simple rule, namely the anticommutation rule, and this fact makes a Fourier transformation to reduce a complex spin system possible.

Spin operators anticommute among themselves as

$$\{S_m^{(i)}, S_m^{(j)}\} = 2\delta_{ij}, \quad i, j = x, y, z,$$

because

$$\mathbf{S}_m \times \mathbf{S}_m = 2i\mathbf{S}_m$$

and

$$[S_m^{(i)}]^2 = 1,$$

where m is the index to show different spins. But spins of different lattice sites commute with one another. On the other hand, fermion operators follow the anticommutation rule

$$\{C_m, C_n\} = \{C_m^+, C_n^+\} = 0$$

and

$$\{C_m, C_n^+\} = \delta_{mn}.$$

As usual, one represents a set of spin operators by direct products of the Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

as

$$X_m \equiv S_m^x = I \otimes \cdots \otimes I \otimes X \otimes I \otimes \cdots \otimes I$$

and

$$Y_m \equiv S_m^y = I \otimes \cdots \otimes I \otimes Y \otimes I \otimes \cdots \otimes I,$$

where X and Y are in the mth position, and, therefore,

$$Z_m \equiv S_m^z = (-i)X_m Y_m = I \otimes \cdots \otimes I \otimes Z \otimes I \otimes \cdots \otimes I,$$

where Z is in the mth position.

Now let one define a set of new operators, tentatively to be called "fermion generating operators," as

$$C_m^x = Z \otimes \cdots \otimes Z \otimes X \otimes I \otimes \cdots \otimes I$$

and

$$C_m^y = Z \otimes \cdots \otimes Z \otimes Y \otimes I \otimes \cdots \otimes I,$$

where X and Y are in the mth position; then one sees easily that

$$\{C_m^x, C_y^x\} = \{C_m^y, C_n^y\} = 2\delta_{mn}$$

and

$$\{C_m^x, C_n^y\} = 0, \quad \text{for all } m \text{ and } n, \cdots. \tag{1}$$

These relations enable one to define a set of operators of the Fermi type by the following relations:

$$\begin{aligned} C_m &= \tfrac{1}{2}(C_m^x - iC_m^y), \\ C_m^+ &= \tfrac{1}{2}(C_m^x + iC_m^y) \end{aligned}$$

or

$$\begin{aligned} C_m^x &= C_m + C_m^+, \\ C_m^y &= i(C_m - C_m^+). \end{aligned} \tag{2}$$

Consequently, one obtains a set of one-to-one correspondence between pairs of spin operators of two adjacent lattice sites and pairs of such "fermion generating operators" as follows:

$$\begin{aligned} X_m X_{m+1} &= (-i)C_m^y C_{m+1}^x = (C_m C_{m+1} - C_m^+ C_{m+1}^+) \\ &\quad - (C_m^+ C_{m+1} - C_m C_{m+1}^+), \\ Y_m Y_{m+1} &= iC_m^x C_{m+1}^y = -(C_m C_{m+1} - C_m^+ C_{m+1}^+) \\ &\quad - (C_m^+ C_{m+1} - C_m C_{m+1}^+), \\ X_m Y_{m+1} &= (-i)C_m^y C_{m+1}^y = i[(C_m C_{m+1} + C_m^+ C_{m+1}^+) \\ &\quad - (C_m^+ C_{m+1} + C_m C_{m+1}^+)], \\ Y_m X_{m+1} &= iC_m^x C_{m+1}^x = i[(C_m C_{m+1} + C_m^+ C_{m+1}^+) \\ &\quad + (C_m^+ C_{m+1} + C_m C_{m+1}^+)], \end{aligned} \tag{3}$$

and

$$Z_m = (-i)C_m^x C_m^y = 2C_m^+ C_m - 1 \equiv (-i)C_m^z.$$

Spin pairs of two sites a distance apart cannot, in general, be shown in this fashion except with a few exceptions (see Sec. III). This is the reason why a 3-dimensional Lenz–Ising model and some other more complicated lattice-spin systems have not been solved with this method. Once this transformation is established, one will see how easily the stupendous mathematical involvement of the Onsager's solution of the 2-dimensional model can be resolved to an SU_2-type problem and solved in a one-page work.[3]

However, one may generalize the relations (3) a little further. If there are a set of matrices (A_m, B_m, C_m), where $m = 1, 2, \cdots, M$, such that

$$\begin{aligned} A_m B_m &= (-)B_m A_m = iC_m, \\ A_m^2 &= B_m^2 = C_m^2 = 1, \quad \text{for all } m, \end{aligned}$$

and matrices of different indices commute with one another, then a set of "fermion generating operators" again can be established as

$$\begin{aligned} C_m^x &= C_1 \otimes C_2 \otimes \cdots \otimes C_{m-1} \otimes A_m \otimes I \otimes \cdots \otimes I, \\ C_m^y &= C_1 \otimes C_2 \otimes \cdots \otimes C_{m-1} \otimes B_m \otimes I \otimes \cdots \otimes I \end{aligned} \tag{4}$$

and then

$$C_m^z = C_m^x C_m^y = (i)I \otimes \cdots \otimes I \otimes C_m \otimes I \otimes \cdots \otimes I,$$

where A_m, B_m, and C_m are in the mth position.

The spin operator can be rotated as follows:

$$\exp(\theta Z)(X)\exp(-\theta Z) = X\cosh(2\theta) + iY\sinh(2\theta)$$

and

$$\exp(\theta Z)(Y)\exp(-\theta Z) = Y\cosh(2\theta) - iX\sinh(2\theta) \quad \text{(cyclically)}.$$

A generalized version of this rotational relation is

$$\begin{aligned} &\exp(\theta S_1 \cdots S_{m-1} Z_m S_{m+1} \cdots S_m) \\ &\quad \times (X_m)\exp(-\theta S_1 \cdots S_{m-1} Z_m S_{m+1} \cdots S_M) \\ &= X_m \cosh(2\theta) \\ &\quad + iS_1 \cdots S_{m-1} Y_m S_{m+1} \cdots S_M \sinh(2\theta) \cdots, \end{aligned} \tag{5}$$

where S_m are components of spins. A special case of them, which is to be used many times in the following, is

$$\begin{aligned} &\exp[-i(\tfrac{1}{4}\pi)S_1 \cdots S_{m-1} Z_m S_{m+1} \cdots S_M] \\ &\quad \times (X_m)\exp[i(\tfrac{1}{4}\pi)S_1 \cdots S_{m-1} Z_m S_{m+1} \cdots S_M] \\ &\qquad = S_1 \cdots S_{m-1} Y_m S_{m+1} \cdots S_M, \end{aligned} \tag{6}$$

and so forth.

Equipped with these formulas, we proceed to solve, in Sec. II, the linear Lenz–Ising model with nearest, as well as next-nearest, neighbor couplings. In Sec. III, an imperfect 1-dimensional Heisenberg–Dirac model is solved exactly in the sense that all the elementary excitations are found and defined. In the last section, the XY model is studied with a simple anisotropicity and inhomogeneity introduced in the coupling strengths.

II. THE 1-DIMENSIONAL LENZ–ISING MODEL

If the model contains both the nearest-neighbor as well as the next-nearest-neighbor couplings, the Hamiltonian is

$$H_1 = -J_1 \sum_{m=1}^{M} X_m X_{m+1} - J_2 \sum_{m=1}^{M} X_m X_{m+2}.$$

Now the operator,

$$T = \exp\left[i\tfrac{1}{4}\pi X_1 Y_2\right] \exp\left(i\tfrac{1}{4}\pi X_2 Y_3\right) \cdots \exp\left(i\tfrac{1}{4}\pi X_M Y_{M+1}\right) \exp\left(i\tfrac{1}{4}\pi X_{M+1} Y_{M+2}\right) \cdots \quad (7)$$

is used on the Hamiltonian for a transformation. Note that factors in this operator do not always commute with one another. The Hamiltonian, if one recalls the relation (6) in Sec. I, becomes simply

$$H_1' = -J_1 \sum_{m=1}^{M} Z_{m+1} - J_2 \sum_{m=1}^{M} Z_{m+1} Z_{m+2}.$$

It is nothing but the Hamiltonian of a 1-dimensional model with only the nearest-neighbor couplings but in a finite magnetic field. The apparent coupling strength is that of the original next-nearest-neighbor couplings, and the strength of the apparent magnetic field is the coupling strength of the original nearest-neighbor couplings. The only discrepancy which can possibly occur in this correspondence is in that at the boundary a few terms are compromised. However, at the thermodynamical limit in which the total number of spins $M \to \infty$, this difference does not affect the result. The statistical mechanical solution of such Hamiltonian is, of course, well known. And the solution verifies the above transformation.

On the other hand, if one operates the operator (7) to transform the Hamiltonian

$$H_2 = -J \sum_{m=1}^{M} X_m X_{m+1},$$

one gets simply

$$H_2' = -J \sum_{m=1}^{M} Z_{m+1},$$

which is nothing but the one for many independent spins; that is to say, such Hamiltonian is mathematically "decouplable." Consequently, such "decouplability" accounts for why a linear chain Lenz–Ising model in a zero magnetic field behaves like many independent spins.

III. THE IMPERFECT 1-DIMENSIONAL HEISENBERG–DIRAC MODEL

This model takes a shape defined by the Hamiltonian

$$H = -J_1 \sum_{m=1}^{M} \left[X_{3m+1} X_{3m+2} + X_{3m+2} X_{3(m+1)}\right] - J_2 \sum_{m=1}^{M} Y_{3m+1} Y_{3(m+1)} - J_3 \sum_{m=1}^{M} \left[Z_{3m+1} Z_{3(m+1)} + Z_{3(m+1)} Z_{3(m+1)+1}\right]. \quad (8)$$

It can be visualized in Fig. 1, in which one strike on the lines indicates couplings between x components, two strikes on the lines indicate couplings between y components, and three strikes on the lines indicate interactions between z components. In Fig. 2, the "Ising–Heisenberg" model of Lieb–Schultz–Mattis[4] is drawn for a comparison. One sees that these two models are physically very similar. In their work, Mattis *et al.* partially solved this model; that is, it was done for the antiferromagnetic case and in a small subset defined by $M_z = 0$. In this study, the problem is solved completely in the sense that all excitations are found in terms of pseudofermions without making any distinction whether the model is ferromagnetic or antiferromagnetic. For a mathematical convenience, let an auxiliary structure of Fig. 3 be used for the calculation instead of the original one, Fig. 1. In the new figure, the sites indicated by solid circles are introduced for the manipulation purpose. The Hamiltonian (8) should take a shape in terms of the new lattice:

$$H' = -J_1 \sum_{m=1}^{M} \left[X_{4m+1} X_{4m+2} + X_{4m+2} X_{4m+3}\right] - J_2 \sum_{m=1}^{M} Y_{4m+1} Y_{4m+3} - J_3 \sum_{m=1}^{M} \left[Z_{4m+1} Z_{4m+3} + Z_{4m+3} Z_{4(m+1)+1}\right].$$

Now under a transformation by an operator like the operator (7) as introduced in Sec. II, but with more terms in the product to suit the present purpose, it

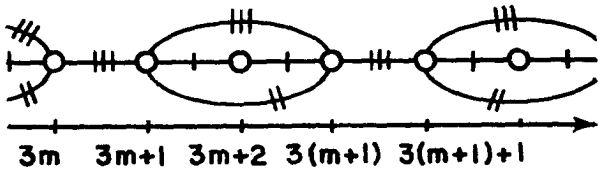


FIG. 1. The lattice structure and couplings of the model.

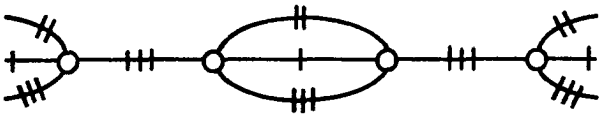


FIG. 2. The "Ising–Heisenberg" model of Lieb, Mattis, and Schultz.

becomes

$$H'' = -J_1 \sum_{m=1}^{M} [Z_{4m+2} + Z_{4m+3}] - J_2 \sum_{m=1}^{M} Y_{4m+1}X_{4m+2}X_{4m+3}Y_{4(m+1)} - J_3 \sum_{m=1}^{M} [Y_{4m+1}Y_{4m+2}Y_{4m+3}Y_{4(m+1)} + Y_{4m+3}Y_{4(m+1)}Y_{4(m+1)+1}Y_{4(m+1)+2}].$$

If one defines the identities

$$A_{2m-1} = Y_{4m+1}X_{4m+2}, \quad B_{2m-1} = Y_{4m+1}Y_{4m+2}, \quad C_{2m-1} = (-i)A_{2m-1}B_{2m-1} = Z_{4m+2}$$

and

$$A_{2m} = X_{4m+3}Y_{4(m+1)}, \quad B_{2m} = Y_{4m+3}Y_{4(m+1)}, \quad C_{2m} = (-i)A_{2m}B_{2m} = Z_{4m+3},$$

one finds that the set of operators (A_m, B_m, C_m) so defined anticommute among themselves, while they commute with operators of other sets with different indices. Squares of these operators are unity; that is to say, they are just like spin operators in every way. Hence, with the relation (4) in Sec. I, one can construct a set of "fermion-generating operators," C_m^x and C_m^y, out of these operators as

$$C_m^x = C_1 \otimes C_2 \otimes \cdots \otimes C_{m-1} \otimes A_m \otimes I \otimes \cdots \otimes I,$$
$$C_m^y = C_1 \otimes C_2 \otimes \cdots \otimes C_{m-1} \otimes B_m \otimes I \otimes \cdots \otimes I,$$

where A_m and B_m are in the mth position. Here one has to keep in mind the fact that A_m take different shapes for the odd and even sites. Finally, the Hamiltonian can be easily shown, in terms of the "fermion-generating operators," as

$$H'' = iJ_1 \sum_{m=1}^{2M} C_m^x C_m^y + iJ_2 \sum_{m=1}^{M} C_{2m-1}^y C_{2m}^x - iJ_3 \sum_{m=1}^{M} (C_{2m-1}^x C_{2m}^y + C_{2m}^x C_{2m+1}^y)$$

and, in terms of the fermion operators C_m and C_m^+,

$$H'' = -J_1 \sum_{m=1}^{2M} (2C_m^+ C_m - 1) - J_2 \sum_{m=1}^{M} (C_{2m-1}C_{2m} - C_{2m-1}^+ C_{2m}^+ - C_{2m-1}^+ C_{2m} + C_{2m-1}C_{2m}^+) + J_3 \sum_{m=1}^{2M} (C_m C_{m+1} - C_m^+ C_{m+1}^+ + C_m^+ C_{m+1} - C_m C_{m+1}^+).$$

From the translation symmetry of the lattice structure and the Hamiltonian, one defines the Fourier transforms in the following fashion:

$$C_q = M^{-\frac{1}{2}} \sum_{m=1}^{M} C_{2m-1} \exp[i(2m-1)],$$
$$d_q = M^{-\frac{1}{2}} \sum_{m=1}^{M} C_{2m} \exp[i(2m-1)];$$

and the Hamiltonian in terms of the Fourier transforms C_q and d_q is

$$H'' = -2J_1 \sum_q (C_q^+ C_q + d_q^+ d_q - 1) - J_2 \sum_q \exp(iq)(C_q d_{-q} - C_{-q}^+ d_q^+ - C_{-q}^+ d_{-q} + C_q d_q^+) + J_3 \sum_q \exp(iq) \times [(C_q d_{-q} - C_{-q}^+ d_q^+ + C_{-q}^+ d_{-q} - C_q d_q^+) + (d_q C_{-q} - d_{-q}^+ C_q^+ + d_{-q}^+ C_{-q} - d_q C_q^+)].$$

Now one sees that it is a sum of the terms which contain the operators of k wave vector, as well as $(-)k$ wave vector, the reflections of k at the origin $(k = 0)$ in k space. Hence it is profitable to get rid of $k < 0$ from the sum. Then

$$H'' = H_0 + \sum_{q>0} H_q,$$

where

$$H_0 = -2J_1(C_0^+ C_0 + d_0^+ d_0 - 1) - J_2(C_0 d_0 - C_0^+ d_0^+) + (J_2 + 2J_3)(C_0^+ d_0 - C_0 d_0^+)$$

and

$$\begin{aligned} H_q = {} & -2J_1(C_q^+ C_q + C_{-q}^+ C_{-q} + d_q^+ d_q + d_{-q}^+ d_{-q} - 2) \\ & + C_q d_{-q}[-J_2 \exp(iq) + J_3 \exp(iq) - J_3 \exp(-iq)] \\ & + C_{-q}^+ d_q^+[J_2 \exp(iq) - J_3 \exp(iq) + J_3 \exp(-iq)] \\ & + C_{-q} d_q[-J_2 \exp(-iq) + J_3 \exp(-iq) - J_3 \exp(iq)] \\ & + C_q^+ d_{-q}^+[J_2 \exp(-iq) - J_3 \exp(-iq) + J_3 \exp(iq)] \\ & + C_{-q}^+ d_{-q}[J_2 \exp(iq) + J_3 \exp(iq) + J_3 \exp(-iq)] \\ & + C_q d_q^+[-J_2 \exp(iq) - J_3 \exp(iq) - J_3 \exp(-iq)] \\ & + C_q^+ d_q[J_2 \exp(-iq) + J_3 \exp(-iq) + J_3 \exp(iq)] \\ & + C_{-q} d_{-q}^+[-J_2 \exp(-iq) - J_3 \exp(-iq) - J_3 \exp(iq)]. \end{aligned}$$

Since pair operators of the Fermi type of different kinds commute with one another, the Hamiltonian is reduced into many subspaces of 16 × 16, namely, reduced to an SU_4 problem, because the reduced Hamiltonian H_q is one of this kind.

To solve this SU_4 problem of

$$\begin{aligned}H_q = &-2J_1(C_q^+C_q + C_{-q}^+C_{-q} + d_q^+d_q + d_{-q}^+d_{-q} - 2)\\ &+ C_qd_{-q}[-J_2\exp(iq) + 2J_3 i\sin(q)]\\ &+ C_{-q}^+d_q^+[J_2\exp(iq) - 2J_3 i\sin(q)]\\ &+ d_qC_{-q}[J_2\exp(-iq) + 2J_3 i\sin(q)]\\ &+ d_{-q}^+C_q^+[-J_2\exp(-iq) - 2J_3 i\sin(q)]\\ &+ C_{-q}^+d_{-q}[J_2\exp(iq) + 2J_3\cos(q)]\\ &+ C_qd_q^+[-J_2\exp(iq) - 2J_3\cos(q)]\\ &+ d_{-q}^+C_{-q}[J_2\exp(-iq) + 2J_3\cos(q)]\\ &+ d_qC_q^+[-J_2\exp(-iq) - 2J_3\cos(q)],\end{aligned}$$

let one define two vectors as

$$\mathbf{x} = \begin{pmatrix} C_q \\ C_{-q}^+ \\ d_q \\ d_{-q}^+ \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} C_q^+ \\ C_{-q} \\ d_q^+ \\ d_{-q} \end{pmatrix},$$

in which the components are operators. Then the reduced Hamiltonian can be shown in a quadratic form by these two vectors as

$$H_q = \sum_{m,n=1}^{4} U_{mn}(q)x_m y_n,$$

where the matrix $U(q)$ is

$$\begin{pmatrix} 2J_1 & 0 & -(J_2e^{iq} + 2J_3\cos q) & (-J_2e^{iq} + 2J_3 i\sin q) \\ 0 & -2J_1 & (J_2e^{iq} - 2J_3 i\sin q) & (J_2e^{iq} + 2J_3\cos q) \\ -(J_2e^{-iq} + 2J_3\cos q) & (J_2e^{-iq} + 2J_3 i\sin q) & 2J_1 & 0 \\ -(J_2e^{-iq} + 2J_3 i\sin q) & (J_2e^{-iq} + 2J_3\cos q) & 0 & -2J_1 \end{pmatrix}.$$

It can be expressed in terms of the Dirac's matrices of quantum mechanics. For convenience, we go apart from the conventional notations but choose the following one for the matrices:

$$\gamma_{ij} = X_i \otimes X_j, \quad i,j = 0, 1, 2, 3,$$

where $X_1 = X$, $X_2 = Y$, $X_3 = Z$, and $X_0 = I$ (the 2 × 2 unit matrix). Consequently, the matrix $U(q)$ is shown as

$$\begin{aligned}U(q) = &\,2J_1\gamma_{03} - \gamma_{13}(J_2 + 2J_3)\cos(q)\\ &+ \gamma_{12}(J_2 - 2J_3)\sin(q) + \gamma_{23}J_2\sin(q)\\ &+ \gamma_{22}J_2\cos(q).\end{aligned}$$

So far, no distinction whatsoever was made on whether the model was a ferromagnetic or antiferromagnetic one. We try to proceed without doing this.

Now one can make a few similarity transformations on the matrix $U(q)$ similar to the principal axis transformation one may encounter in algebra. Such transformation is equivalent to a linear transformation of the vectors **x** and **y**, and the latter transformation is

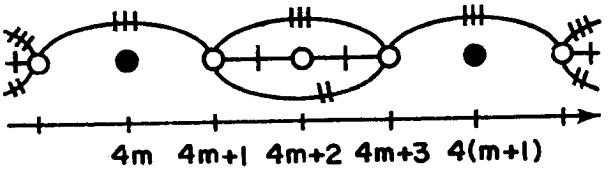


FIG. 3. The auxiliary lattice structure of the model.

nothing but a generalized version of the Bogoliubov–Valatin transformation (see also Appendix). Since the components of the vectors involved are operators of variables k, it is essential that the transformation be canonical.

Suppose that the linear transformation is

$$\mathbf{x} = T^{\mathrm{T}}\mathbf{x}' \quad \text{and} \quad \mathbf{y} = T^{-1}\mathbf{y}';$$

then

$$\begin{aligned}H_q &= \sum_{m,n} U_{mn}(q)\left(\sum_j T_{mj}^{\mathrm{T}}x_j'\right)\left(\sum_k T_{nk}^{-1}y_k'\right)\\ &= \sum_{j,k}\left(\sum_{m,n} T_{jm}U_{mn}(q)T_{nk}^{-1}\right)x_j'y_k',\end{aligned}$$

where T^{T} and T^{-1} are the transpose and inverse matrix of T, respectively.

Six similarity transformations in the following order are used on the matrix $U(q)$ and their corresponding transformations of the Bogoliubov–Valatin type are also presented:

(1) The transformation operator is

$$T_1 = \exp[i(\tfrac{1}{4}\pi)\gamma_{30}] = \cosh(\tfrac{1}{4}i\pi) + \gamma_{30}\sinh(\tfrac{1}{4}i\pi)$$

$$= 2^{-\frac{1}{2}}\begin{pmatrix} 1+i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 0 & 1-i \end{pmatrix},$$

and the corresponding linear transformation is

$$\mathbf{x} = T_1^{\mathrm{T}}\mathbf{X}' \rightarrow \begin{cases} C_q \rightarrow 2^{-\frac{1}{2}}(1+i)C_q \\ C_{-q}^+ \rightarrow 2^{-\frac{1}{2}}(1+i)C_{-q}^+ \\ d_q \rightarrow 2^{-\frac{1}{2}}(1-i)d_q \\ d_{-q}^+ \rightarrow 2^{-\frac{1}{2}}(1-i)d_{-q}^+ \end{cases}$$

and

$$\mathbf{y} = T_1^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow 2^{-\frac{1}{2}}(1-i)C_q^+ \\ C_{-q} \rightarrow 2^{-\frac{1}{2}}(1-i)C_{-q} \\ d_q^+ \rightarrow 2^{-\frac{1}{2}}(1+i)d_q^+ \\ d_{-q} \rightarrow 2^{-\frac{1}{2}}(1+i)d_{-q}; \end{cases}$$

it is a canonical transformation. The matrix $U(q)$, after this transformation by the relation (6) of Sec. I, can be easily shown to be

$$\begin{aligned} U_1(q) = {} & 2J_1\gamma_{03} + \gamma_{23}(J_2 + 2J_3)\cos(q) \\ & - \gamma_{22}(J_2 - 2J_3)\sin(q) + J_2\gamma_{13}\sin(q) \\ & + \gamma_{12}J_2\cos(q). \end{aligned}$$

(2) The transformation operator is

$$T_2 = \exp(-\theta\gamma_{30}) = \cosh(2\theta) - \gamma_{30}\sinh(2\theta)$$

$$= \begin{pmatrix} \exp(-\theta) & 0 & 0 & 0 \\ 0 & \exp(-\theta) & 0 & 0 \\ 0 & 0 & \exp(\theta) & 0 \\ 0 & 0 & 0 & \exp(\theta) \end{pmatrix},$$

where

$$\cosh(2\theta) = \frac{(J_2 + 2J_3)\cos(q)}{[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2(q)]^{\frac{1}{2}}},$$

$$\sinh(2\theta) = \frac{iJ_2\sin(q)}{[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2(q)]^{\frac{1}{2}}},$$

and, therefore, $\theta^*(q) = (-)\theta(q)$ and $\theta(-q) = (-)\theta(q)$. The corresponding linear transformation is

$$\mathbf{x} = T_2^{\mathrm{T}}\mathbf{x}' \rightarrow \begin{cases} C_q \rightarrow C_q \exp[-\theta(q)] \\ C_{-q}^+ \rightarrow C_{-q}^+ \exp[-\theta(q)] \\ d_q \rightarrow d_q \exp[\theta(q)] \\ d_{-q}^+ \rightarrow d_{-q}^+ \exp[\theta(q)] \end{cases}$$

and

$$\mathbf{y} = T_2^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow C_q^+ \exp[\theta(q)] \\ C_{-q} \rightarrow C_{-q} \exp[\theta(q)] \\ d_q^+ \rightarrow d_q^+ \exp[-\theta(q)] \\ d_{-q} \rightarrow d_{-q} \exp[-\theta(q)]; \end{cases}$$

it is a canonical transformation. The matrix $U_1(q)$ with the relation (5) becomes

$$\begin{aligned} U_2(q) = {} & 2J_1\gamma_{03} + (J_2 + 2J_3)\cos(q) \\ & \times (\gamma_{23}\cosh 2\theta + i\gamma_{13}\sinh 2\theta) \\ & + J_2\sin(q)(\gamma_{13}\cosh 2\theta - i\gamma_{23}\sinh 2\theta) \\ & - (J_2 - 2J_3)\sin(q)(\gamma_{22}\cosh 2\theta + i\gamma_{12}\sinh 2\theta) \\ & + J_2\cos(q)(\gamma_{12}\cosh 2\theta - i\gamma_{22}\sinh 2\theta) \\ = {} & 2J_1\gamma_{03} + \gamma_{23}[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2(q)]^{\frac{1}{2}} \\ & + \frac{\gamma_{12}\{J_2[J_2 + 2J_3\cos(2q)]\}}{[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2(q)]^{\frac{1}{2}}} \\ & + \frac{\gamma_{22}[2J_3^2\sin(2q)]}{[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2 q]^{\frac{1}{2}}}. \end{aligned}$$

(3) The transformation operator is

$$T_3 = \exp[(-\tfrac{1}{4}i\pi)\gamma_{03}]$$

$$= 2^{-\frac{1}{2}}\begin{pmatrix} 1-i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 0 & 1+i \end{pmatrix},$$

and its corresponding Bogoliubov–Valatin transformation is

$$\mathbf{x} = T_3^{\mathrm{T}}\mathbf{x}' \rightarrow \begin{cases} C_q \rightarrow 2^{-\frac{1}{2}}(1-i)C_q \\ C_{-q}^+ \rightarrow 2^{-\frac{1}{2}}(1+i)C_{-q}^+ \\ d_q \rightarrow 2^{-\frac{1}{2}}(1-i)d_q \\ d_{-q}^+ \rightarrow 2^{-\frac{1}{2}}(1+i)d_{-q}^+ \end{cases}$$

and

$$\mathbf{y} = T_3^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow 2^{-\frac{1}{2}}(1+i)C_q^+ \\ C_{-q} \rightarrow 2^{-\frac{1}{2}}(1-i)C_{-q} \\ d_q^+ \rightarrow 2^{-\frac{1}{2}}(1+i)d_q^+ \\ d_{-q} \rightarrow 2^{-\frac{1}{2}}(1-i)d_{-q}; \end{cases}$$

it is a canonical transformation. The matrix $U_2(q)$ after a transformation by the operator T_3 with the relation (6) becomes

$$\begin{aligned} U_3(q) = {} & 2J_1\gamma_{03} + \gamma_{23}[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2(q)]^{\frac{1}{2}} \\ & - \frac{\gamma_{11}\{J_2[J_2 + 2J_3\cos(2q)]\}}{[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2 q]^{\frac{1}{2}}} \\ & - \frac{\gamma_{21}[2J_3^2\sin(2q)]}{[J_2^2\sin^2(q) + (J_2 + 2J_3)^2\cos^2(q)]^{\frac{1}{2}}}. \end{aligned}$$

(4) The transformation operator is

$$T_4 = \exp\{[\tfrac{1}{4}\pi i]\gamma_{30}\}$$

$$= 2^{-\frac{1}{2}}\begin{pmatrix} 1+i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 0 & 1-i \end{pmatrix},$$

and its corresponding Bogoliubov–Valatin transformation is

$$\mathbf{x} = T_4^{\mathrm{T}}\mathbf{x}' \rightarrow \begin{cases} C_q \rightarrow 2^{-\frac{1}{2}}(1+i)C_q \\ C_{-q}^+ \rightarrow 2^{-\frac{1}{2}}(1+i)C_{-q}^+ \\ d_q \rightarrow 2^{-\frac{1}{2}}(1-i)d_q \\ d_{-q}^+ \rightarrow 2^{-\frac{1}{2}}(1-i)d_{-q}^+ \end{cases}$$

and

$$\mathbf{y} = T_4^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow 2^{-\frac{1}{2}}(1-i)C_q^+ \\ C_{-q} \rightarrow 2^{-\frac{1}{2}}(1-i)C_{-q} \\ d_q^+ \rightarrow 2^{-\frac{1}{2}}(1+i)d_q^+ \\ d_{-q} \rightarrow 2^{-\frac{1}{2}}(1+i)d_{-q}; \end{cases}$$

it is a canonical transformation. The matrix $U_3(q)$ after a transformation by the operator T_4 with the relation (6) becomes

$$U_4(q) = 2J_1\gamma_{03} + \gamma_{13}[J_2^2\sin^2(q) + (J_2+2J_3)^2\cos^2(q)]^{\frac{1}{2}} + \frac{\gamma_{21}\{J_2[J_2+2J_3\cos(2q)]\}}{[J_2^2\sin^2(q)+(J_2+2J_3)^2\cos^2 q]^{\frac{1}{2}}} - \frac{\gamma_{11}[2J_3^2\sin(2q)]}{[J_2^2\sin^2(q)+(J_2+2J_3)^2\cos^2(q)]^{\frac{1}{2}}}.$$

(5) The transformation operator is

$$T_5 = \exp[\alpha\gamma_{22}] = \begin{pmatrix} \cosh(\alpha) & 0 & 0 & -\sinh(\alpha) \\ 0 & \cosh(\alpha) & \sinh(\alpha) & 0 \\ 0 & \sinh(\alpha) & \cosh(\alpha) & 0 \\ -\sinh(\alpha) & 0 & 0 & \cosh(\alpha) \end{pmatrix},$$

where

$$\cosh(2\alpha) = 2J_1\left(\frac{[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]}{[4J_1^2(J_1^2\sin^2 q + (J_2+2J_3)^2\cos^2 q) + J_2^2(J_2+2J_3\cos 2q)^2]}\right)^{\frac{1}{2}},$$

$$\sinh(2\alpha) = \frac{iJ_2(J_2+2J_3\cos 2q)}{\{4J_1^2[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q] + J_2^2(J_2+2J_3\cos 2q)^2\}^{\frac{1}{2}}},$$

and, therefore,

$$\alpha^*(q) = (-)\alpha(q), \quad \alpha(-q) = \alpha(q).$$

The corresponding linear transformation is

$$\mathbf{x} = T_5^{\mathrm{T}}\mathbf{x}' \rightarrow \begin{cases} C_q \rightarrow C_q\cosh\alpha - d_{-q}^+\sinh\alpha \\ C_{-q}^+ \rightarrow C_{-q}^+\cosh\alpha + d_q\sinh\alpha \\ d_q \rightarrow d_q\cosh\alpha + C_{-q}^+\sinh\alpha \\ d_{-q}^+ \rightarrow d_{-q}^+\cosh\alpha - C_q\sinh\alpha \end{cases}$$

and

$$\mathbf{y} = T_5^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow C_q^+\cosh\alpha + d_{-q}\sinh\alpha \\ C_{-q} \rightarrow C_{-q}\cosh\alpha - d_q^+\sinh\alpha \\ d_q^+ \rightarrow d_q^+\cosh\alpha - C_{-q}\sinh\alpha \\ d_{-q} \rightarrow d_{-q}\cosh\alpha + C_q^+\sinh\alpha; \end{cases}$$

it is a canonical transformation. The matrix $U_4(q)$ after a transformation by the operator T_5 with the relation (5) becomes

$$\begin{aligned} U_5(q) &= 2J_1[\gamma_{03}\cosh(2\alpha) + i\gamma_{21}\sinh(2\alpha)] + [\gamma_{21}\cosh(2\alpha) - i\gamma_{03}\sinh(2\alpha)]\{J_2[J_2+2J_3\cos(2q)]\} \\ &\quad\times[J_2^2\sin^2(q) + (J_2+2J_3)^2\cos^2 q]^{-\frac{1}{2}} + \gamma_{13}[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]^{\frac{1}{2}} \\ &\quad - \gamma_{11}[2J_3^2\sin(2q)][J_2^2\sin^2(q) + (J_2+2J_3)^2\cos^2 q]^{-\frac{1}{2}}. \\ &= \gamma_{03}\left\{\frac{[4J_1^2(J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q) + J_2^2(J_2+2J_3\cos 2q)^2]}{[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]}\right\}^{\frac{1}{2}} \\ &\quad + \gamma_{13}[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]^{\frac{1}{2}} - \frac{\gamma_{11}[2J_3^2\sin 2q]}{[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]^{\frac{1}{2}}}. \end{aligned}$$

(6) The transformation operator is

$$T_6 = \exp[-(\tfrac{1}{4}i\pi)\gamma_{20}] = 2^{-\frac{1}{2}}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

and the corresponding linear transformation is

$$\mathbf{x} = T_6^{\mathrm{T}}\mathbf{x}' \to \begin{cases} C_q \to 2^{-\frac{1}{2}}(C_q + d_q) \\ C_{-q}^+ \to 2^{-\frac{1}{2}}(C_{-q}^+ + d_{-q}^+) \\ d_q \to 2^{-\frac{1}{2}}(d_q - C_q) \\ d_{-q}^+ \to 2^{-\frac{1}{2}}(d_{-q}^+ - C_{-q}^+) \end{cases} \quad \text{and} \quad \mathbf{y} = T_6^{-1}\mathbf{y}' \to \begin{cases} C_q^+ \to 2^{-\frac{1}{2}}(C_q^+ + d_q^+) \\ C_{-q} \to 2^{-\frac{1}{2}}(C_{-q} + d_{-q}) \\ d_q^+ \to 2^{-\frac{1}{2}}(d_q^+ - C_q^+) \\ d_{-q} \to 2^{-\frac{1}{2}}(d_{-q} - C_{-q}); \end{cases}$$

it is a canonical transformation. The matrix $U_5(q)$ with the relation (6) finally becomes

$$U_6(q) = A(q)\gamma_{03} + B(q)\gamma_{33} + E(q)\gamma_{31}$$
$$= \begin{pmatrix} A(q)+B(q) & E(q) & 0 & 0 \\ E(q) & -[A(q)+B(q)] & 0 & 0 \\ 0 & 0 & A(q)-B(q) & -E(q) \\ 0 & 0 & -E(q) & -[A(q)-B(q)] \end{pmatrix},$$

where

$$A(q) = \left(\frac{[4J_1^2(J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q) + J_2^2(J_2+2J_3\cos 2q)^2}{[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]}\right)^{\frac{1}{2}},$$
$$B(q) = [J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]^{\frac{1}{2}},$$

and

$$E(q) = \frac{-2J_3^2\sin 2q}{[J_2^2\sin^2 q + (J_2+2J_3)^2\cos^2 q]^{\frac{1}{2}}}.$$

Consequently, the reduced Hamiltonian is transformed to be

$$\begin{aligned} H'(q) = &-[A(q)+B(q)](C_q^+C_q + C_{-q}^+C_{-q} - 1) \\ &+ E(q)(C_qC_{-q} + C_{-q}^+C_q^+) \\ &- [A(q)-B(q)](d_q^+d_q + d_{-q}^+d_{-q} - 1) \\ &- E(q)(d_qd_{-q} + d_{-q}^+d_q^+) \end{aligned}$$

and is reduced to an SU_2 problem. With the aid of the Appendix,

$$\begin{aligned} H_q' \to H_q'' = &(C_q^+C_q + C_{-q}^+C_{-q} - 1) \\ &\times \{[A(q)+B(q)]^2 + E^2(q)\}^{\frac{1}{2}} \\ &+ (d_q^+d_q + d_{-q}^+d_{-q} - 1) \\ &\times \{[A(q)-B(q)]^2 + E^2(q)\}^{\frac{1}{2}}. \end{aligned}$$

For the term of $k = 0$, the reduced Hamiltonian can also be solved, but it is simpler, because it is already an SU_2 problem of C_0 and d_0. With the definition given in the Appendix,

$$H_0 = -2J_1\alpha^z - J_2\alpha^x + (J_2+2J_3)\beta^x$$

and, with its aid, the reduced Hamiltonian becomes

$$\begin{aligned} H_0 \to H_0'' &= \alpha^z(4J_1^2 + J_2^2)^{\frac{1}{2}} + \beta^z(J_2+2J_3) \\ &= (C_0^+C_0 - \tfrac{1}{2})[(J_2+2J_3) + (4J_1^2+J_2^2)^{\frac{1}{2}}] \\ &\quad + (d_0^+d_0 - \tfrac{1}{2})[-(J_2+2J_3) + (4J_1^2+J_2^2)^{\frac{1}{2}}]. \end{aligned}$$

This expression is nothing but that of the reduced Hamiltonian H_q'' when k vanishes, because

$$A(k=0) = (4J_1^2 + J_2^2)^{\frac{1}{2}},$$
$$B(k=0) = (J_2+2J_3),$$

and

$$E(k=0) = 0.$$

Consequently,

$$\begin{aligned} H_0'' = &(C_0^+C_0 - \tfrac{1}{2})[(A_0+B_0)^2 + E_0^2]^{\frac{1}{2}} \\ &+ (d_0^+d_0 - \tfrac{1}{2})[(A_0-B_0)^2 + E_0^2]^{\frac{1}{2}}. \end{aligned}$$

And, also, since

$$A_{-q} = A_q,$$
$$B_{-q} = B_q,$$

and

$$E_{-q} = E_q,$$

one gets

$$\begin{aligned} H_q'' = &(C_q^+C_q - \tfrac{1}{2})[(A_q+B_q)^2 + E_q^2]^{\frac{1}{2}} \\ &+ (C_{-q}^+C_{-q} - \tfrac{1}{2})[(A_{-q}+B_{-q})^2 + E_{-q}^2]^{\frac{1}{2}} \\ &+ (d_q^+d_q - \tfrac{1}{2})[(A_q-B_q)^2 + E_q^2]^{\frac{1}{2}} \\ &+ (d_{-q}^+d_{-q} - \tfrac{1}{2})[(A_{-q}-B_{-q})^2 + E_{-q}^2]^{\frac{1}{2}}. \end{aligned}$$

Finally, the total Hamiltonian in a diagonal form is

$$\begin{aligned} H'' \to H'' &= H_0'' + \sum_{q\neq 0} H_q'' \\ &= \sum_{(\text{all } q)} \{(2C_q^+C_q - 1)\tfrac{1}{2}[(A_q+B_q)^2 + E_q^2]^{\frac{1}{2}} \\ &\quad + (2d_q^+d_q - 1)\tfrac{1}{2}[(A_q-B_q)^2 + E_q^2]^{\frac{1}{2}}. \end{aligned}$$

Therefore, the spectra of this 1-dimensional model in terms of pseudofermions are in two branches, and they are

$$\epsilon_{1,2} = \tfrac{1}{2}[(A_q \pm B_q)^2 + E_q^2]^{\frac{1}{2}},$$

and the diagonalized Hamiltonian,

$$H'' = \sum_{(\text{all } q)} [(2C_q^+ C_q - 1)\epsilon_1(q) + (2d_q^+ d_q - 1)\epsilon_2(q)],$$

where

$$A_q = \left(\frac{[4J_1^2(J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q) + J_2^2(J_2 + 2J_3 \cos 2q)^2]}{[J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q]}\right)^{\frac{1}{2}},$$

$$B_q = [J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q]^{\frac{1}{2}},$$

and

$$E_q = \frac{-2J_3^2 \sin(2q)}{[J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q]^{\frac{1}{2}}}.$$

The result can be easily checked with a few special cases of this problem with known results:

(i) For $J_1 = J_2 = 0$, $A_q = 0$, $B_q = 2J_3 \cos(q)$, and $E_q = (-)2J_3 \sin(q)$. Consequently,

$$\epsilon_{1,2} = J_3,$$

and the Hamiltonian is

$$H = J_3 \sum_{(\text{all } q)} [(2C_q^+ C_q - 1) + (2d_q^+ d_q - 1)].$$

Since in this special case the problem is reduced to a linear chain Lenz–Ising model, the eigenvalues of the Hamiltonian are $\pm J_3$, which agree with the solution here.

(ii) For $J_2 = 0$, $A_q = 2J_1$, $B_q = 2J_3 \cos(q)$, and $E_q = -2J_3 \sin(q)$. Consequently,

$$\epsilon_{1,2} = \{[J_1 \pm J_3 \cos(q)]^2 + J_3^2 \sin^2 q\}^{\frac{1}{2}}.$$

This special case is an imperfect XY model and, in fact, shows the resemblance with the result obtained by Lieb, Schultz, and Mattis [4] and the result of Sec. VI, the special case (i).

(iii) For $J_2 = J_3 = 0$, $A_q = 2J_1$ and $B_q = E_q = 0$. Consequently,

$$\epsilon_{1,2} = J_1.$$

The result fits for this case of many independent clusters of spin pairs.

(iv) For $J_1 = 0$,

$$A_q = \frac{J_2(J_2 + 2J_3 \cos 2q)}{[J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q]^{\frac{1}{2}}},$$

$$B_q = [J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q]^{\frac{1}{2}},$$

and

$$E_q = \frac{-2J_3^2 \sin(2q)}{[J_2^2 \sin^2 q + (J_2 + 2J_3)^2 \cos^2 q]^{\frac{1}{2}}}.$$

The result agrees with the one of Sec. IV, the special case (iii).

(v) For $J_1 = J_3 = 0$, $A_q = J_2$, $B_q = J_2$, and $E_q = 0$,

$$\epsilon_1 = J_2 \quad \text{and} \quad \epsilon_2 = 0.$$

This is the solution for many independent spins.

(vi) For $J_3 = 0$, $A_q = [4J_1^2 + J_2^2]^{\frac{1}{2}}$, $B_q = J_2$, and $E_q = 0$. Consequently,

$$\epsilon_{1,2} = \tfrac{1}{2}[(4J_1^2 + J_2^2)^{\frac{1}{2}} \pm J_2].$$

As expected, there is no correlation between clusters of spins.

In the above calculation, spectra of the elementary excitations are in two branches, and both of them are in the form of

$$\epsilon(q) = [f(J_m, \sin q, \cos q)]^{\frac{1}{2}},$$

regardless of whether it is a ferromagnetic or antiferromagnetic model. In fact, all the exact spectra so far known for the 1-dimensional model, including the one for the XY model,[4] the one for the "Ising–Heisenberg" model,[4] and the numerical solution on the full Heisenberg–Dirac model of Cloiseaux and Pearson,[5] are of this shape. It is commonly assumed that there is a distinction between a ferromagnetism and an antiferromagnetism such that the spectrum of the former is of the form of

$$\lim_{q \to \text{small}} \epsilon(q) \propto q^2,$$

and that of the latter is of the form of

$$\lim_{q \to \text{small}} \epsilon(q) \propto q.$$

For a 1-dimensional model, such distinction may be an oversimplication of the issue, because only for an isotropic ferromagnetism in which

$$\begin{aligned} H_f &= -\sum_m (J_1 X_m X_{m+1} + J_2 Y_m Y_{m+1} + J_3 Z_m Z_{m+1}) \\ &= \sum [-J_3 S_m^z S_{m+1}^z - \tfrac{1}{4}(J_1 + J_2) \\ &\quad \times (S_m^{(+)} S_{m+1}^{(-)} + S_m^{(-)} S_{m+1}^{(+)}) \\ &\quad + \tfrac{1}{4}(-J_1 + J_2)(S_m^{(+)} S_{m+1}^{(+)} + S_m^{(-)} S_{m+1}^{(-)})] \\ &= -J \sum [S_m^z S_{m+1}^z + \tfrac{1}{2}(S_m^{(+)} S_{m+1}^{(-)} + S_m^{(-)} S_{m+1}^{(+)})], \end{aligned}$$

where $J_1 = J_2 = J_3 = J > 0$, is the ground state the state when all spins are parallel, and the first excited

state the state with one spin deviation. In the case, in which $J_1 \neq J_2$, a ferromagnetic model does not show such a simple solution. Rather,all existing evidence (see Refs. 4 and 5 and Secs. III and IV of this work) suggests that it may also have a spectrum of the shape of

$$[f(J_m, \sin q, \cos q)]^{\frac{1}{2}}.$$

The simple solution for an isotropic ferromagnetism may, thus, be a special case of this general type of solutions.

The ground-state energy of this model is

$$E_{gr.} = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\{[(A_q + B_q)^2 + E_q^2]^{\frac{1}{2}} + [(A_q - B_q)^2 + E_q^2]^{\frac{1}{2}}\}\, dq$$

when no pseudofermion is present.

IV. THE XY MODEL

The purpose of this section is to study the contribution of inhomogeneity in the coupling strength to the lattice-spin system. The anisotropicity in the coupling strength has been under attention[6] for a good many years. However, the inhomogeneity has eluded concerns in the study of lattice-spin systems. As a simple example, the XY model of Lieb–Schultz–Mattis is studied with two coupling strengths of different sizes coming into play alternatively. In order to compare its effect with the effect caused by anisotropicity, the latter is also admitted in the model. The Hamiltonian for this purpose is

$$H = -\sum_{m=1}^{M}(J_1 X_{2m-1}X_{2m} + J_2 X_{2m}X_{2m+1} + J_3 Y_{2m-1}Y_{2m} + J_4 Y_{2m}Y_{2m+1}),$$

namely, the coupling strength of the model is neither homogeneous nor isotropic. By the relation (3) of Sec. I, the Hamiltonian in terms of operators of the Fermi type can be shown to be

$$\begin{aligned} H = \sum[&(-J_1 + J_3)(C_{2m-1}C_{2m} - C_{2m-1}^{+}C_{2m}^{+}) \\ &+ (J_1 + J_3)(C_{2m-1}^{+}C_{2m} - C_{2m-1}C_{2m}^{+}) \\ &+ (-J_2 + J_4)(C_{2m}C_{2m+1} - C_{2m}^{+}C_{2m+1}^{+}) \\ &+ (J_2 + J_4)(C_{2m}^{+}C_{2m+1} - C_{2m}C_{2m+1}^{+})]. \end{aligned}$$

With the Fourier transforms defined as

$$C_q = (M)^{-\frac{1}{2}}\sum_{m=1}^{M} C_{2m-1}\exp[(2m-1)iq]$$

and

$$d_q = (M)^{-\frac{1}{2}}\sum_{m=1}^{M} C_{2m}\exp(2miq),$$

as it was done in Sec. III,

$$H = H_0 + \sum_{q>0} H_q,$$

where

$$\begin{aligned} H_0 = &(-J_1 + J_2 + J_3 - J_4)(C_0 d_0 + d_0^{+}C_0^{+}) \\ &+ (J_1 + J_2 + J_3 + J_4)(C_0^{+}d_0 + d_0^{+}C_0) \end{aligned}$$

and

$$\begin{aligned} H_q = &L_q(C_q d_{-q} - C_{-q}^{+}d_q^{+}) - L_{-q}(d_q C_{-q} - d_{-q}^{+}C_q^{+}) \\ &+ M_q(C_{-q}^{+}d_{-q} - C_q d_q^{+}) + M_{-q}(d_{-q}^{+}C_{-q} - d_q C_q^{+}) \end{aligned}$$

with

$$\begin{aligned} L_q = &(-J_1 + J_2 + J_3 - J_4)\cos(q) \\ &+ (-J_1 - J_2 + J_3 + J_4)i\sin(q), \end{aligned}$$

$$\begin{aligned} M_q = &(J_1 + J_2 + J_3 + J_4)\cos(q) \\ &+ (J_1 - J_2 + J_3 - J_4)i\sin(q). \end{aligned}$$

Now for H_0 it is simple to diagonalize these terms. With definitions given in the Appendix,

$$\begin{aligned} H_0 = &(-J_1 + J_2 + J_3 - J_4)\alpha_x \\ &+ (J_1 + J_2 + J_3 + J_4)\beta_x. \end{aligned}$$

Consequently, by the operator,

$$\exp[-(\tfrac{1}{4}\pi)i(\alpha_y - \beta_y)],$$

it is transformed as

$$\begin{aligned} H_0'' &= (J_1 - J_2 - J_3 + J_4)\alpha_z \\ &\quad + (J_1 + J_2 + J_3 + J_4)\beta_z \\ &= (J_1 - J_2 - J_3 + J_4)(C_0^{+}C_0 + d_0^{+}d_0 - 1) \\ &\quad + (J_1 + J_2 + J_3 + J_4)(C_0^{+}C_0 - d_0^{+}d_0) \\ &= (2C_0^{+}C_0 - 1)\tfrac{1}{2}[(J_1 - J_2 - J_3 + J_4) \\ &\quad + (J_1 + J_2 + J_3 + J_4)] \\ &\quad + (2d_0^{+}d_0 - 1)\tfrac{1}{2}[(J_1 - J_2 - J_3 + J_4) \\ &\quad - (J_1 + J_2 + J_3 + J_4)]. \end{aligned}$$

On the other hand, for H_q as it was done in Sec. III, the vectors

$$\mathbf{x} = \begin{pmatrix} C_q \\ C_{-q}^{+} \\ d_q \\ d_{-q}^{+} \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} C_q^{+} \\ C_{-q} \\ d_q^{+} \\ d_{-q} \end{pmatrix}$$

are introduced. With them,

$$H_q = \sum_{m,n=1}^{4} U_{mn}(q)X_m Y_n,$$

where the matrix

$$U(q) = \begin{pmatrix} 0 & 0 & -M_q & L_q \\ 0 & 0 & -L_q & M_q \\ -M_{-q} & -L_{-q} & 0 & 0 \\ L_{-q} & M_{-q} & 0 & 0 \end{pmatrix}$$
$$= -A_3\gamma_{13}\cos(q) + A_2\gamma_{12}\sin(q) + A_4\gamma_{23}\sin(q) + A_1\gamma_{22}\cos(q),$$

with

$$A_1 = (J_1 - J_2 - J_3 + J_4),$$
$$A_2 = (J_1 + J_2 - J_3 - J_4),$$
$$A_3 = (J_1 + J_2 + J_3 + J_4),$$

and

$$A_4 = (J_1 - J_2 + J_3 - J_4).$$

A four-step transformation, similar to the one done in Sec. III, will reduce the matrix $U(q)$ into two subspaces.

(1) The transformation operator is

$$T_1 = \exp(i\tfrac{1}{4}\pi\gamma_{30}) = \cosh(\tfrac{1}{4}i\pi) + \gamma_{30}\sin(\tfrac{1}{4}i\pi)$$
$$= 2^{-\frac{1}{2}}\begin{pmatrix} 1+i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 0 & 1-i \end{pmatrix},$$

and the corresponding linear transformation is

$$\mathbf{x} = T_1^{\mathrm{T}}\mathbf{x}' \rightarrow \begin{cases} C_q \rightarrow \frac{1}{2}^{\frac{1}{2}}(1+i)C_q \\ C_{-q}^+ \rightarrow \frac{1}{2}^{\frac{1}{2}}(1+i)C_{-q}^+ \\ d_q \rightarrow \frac{1}{2}^{\frac{1}{2}}(1-i)d_q \\ d_{-q}^+ \rightarrow \frac{1}{2}^{\frac{1}{2}}(1-i)d_{-q}^+ \end{cases}$$

and

$$\mathbf{y} = T_1^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow \frac{1}{2}^{\frac{1}{2}}(1-i)C_q^+ \\ C_{-q} \rightarrow \frac{1}{2}^{\frac{1}{2}}(1-i)C_{-q} \\ d_q^+ \rightarrow \frac{1}{2}^{\frac{1}{2}}(1+i)d_q^+ \\ d_{-q} \rightarrow \frac{1}{2}^{\frac{1}{2}}(1+i)d_{-q}; \end{cases}$$

it is a canonical transformation. The matrix $U(q)$, after this transformation by the relation (6) of Sec. I, can be easily shown to be

$$U_1(q) = A_3\gamma_{23}\cos(q) - A_2\gamma_{22}\sin(q) + A_4\gamma_{13}\sin(q) + A_1\gamma_{12}\cos(q).$$

(2) The transformation operator is

$$T_2 = \exp(-\theta\gamma_{30})$$
$$= \begin{pmatrix} \exp(-\theta) & 0 & 0 & 0 \\ 0 & \exp(-\theta) & 0 & 0 \\ 0 & 0 & \exp(\theta) & 0 \\ 0 & 0 & 0 & \exp(\theta) \end{pmatrix},$$

where

$$\cosh(2\theta) = \frac{A_3\cos(q)}{(A_4^2\sin^3 q + A_3^2\cos^2 q)^{\frac{1}{2}}},$$
$$\sinh(2\theta) = \frac{iA_4\sin(q)}{(A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}}}.$$

Therefore,

$$\theta^*(q) = -\theta(q); \quad \theta(-q) = -\theta(q).$$

The matrix $U_1(q)$ with the relation (5) becomes

$$U_2(q) = A_1\cos(q)[\gamma_{12}\cosh(2\theta) - i\gamma_{22}\sinh(2\theta)] - A_2\sin(q)[\gamma_{22}\cosh(2\theta) + i\gamma_{12}\sinh(2\theta)] + A_3\cos(q)[\gamma_{23}\cosh(2\theta) + i\gamma_{13}\sinh(2\theta)] + A_4\sin(q)[\gamma_{13}\cosh(2\theta) - i\gamma_{23}\sinh(2\theta)]$$
$$= \gamma_{23}(A_4^2\sin^2(q) + A_3^2\cos^2(q)^{\frac{1}{2}} + \frac{\gamma_{12}[A_1A_3\cos^2 q + A_2A_4\sin^2 q]}{(A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}}} + \frac{\gamma_{22}(A_1A_4 - A_2A_3)\sin(q)\cos(q)}{(A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}}}.$$

The corresponding linear transformation is

$$\mathbf{x} = T_2^{\mathrm{T}}\mathbf{x}' \rightarrow \begin{cases} C_q \rightarrow C_q\exp[-\theta(q)] \\ C_{-q}^+ \rightarrow C_{-q}^+\exp[-\theta(q)] \\ d_q \rightarrow d_q\exp[\theta(q)] \\ d_{-q}^+ \rightarrow d_{-q}^+\exp[\theta(q)] \end{cases}$$

and

$$\mathbf{y} = T_2^{-1}\mathbf{y}' \rightarrow \begin{cases} C_q^+ \rightarrow C_q^+\exp[\theta(q)] \\ C_{-q} \rightarrow C_{-q}\exp[\theta(q)] \\ d_q^+ \rightarrow d_q^+\exp[-\theta(q)] \\ d_{-q} \rightarrow d_{-q}\exp[-\theta(q)], \end{cases}$$

which is canonical.

(3) The transformation operator is

$$T_3 = \exp[-i\tfrac{1}{4}\pi\gamma_{11}]$$
$$= 2^{-\frac{1}{2}}\begin{pmatrix} 1 & 0 & 0 & -i \\ 0 & 1 & -i & 0 \\ 0 & -i & 1 & 0 \\ -i & 0 & 0 & 1 \end{pmatrix}$$

The matrix $U_2(q)$ with the relation (6) becomes

$$U_3(q) = A_q\gamma_{03} + B_q\gamma_{23} + E_q\gamma_{22},$$

where

$$A_q = \frac{(A_1A_3\cos^2 q + A_2A_4\sin^2 q)}{(A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}}},$$
$$B_q = (A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}},$$

and

$$E_q = \frac{(A_1A_4 - A_2A_3)\sin(q)\cos(q)}{(A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}}}.$$

The corresponding linear transformation is

$$\mathbf{x} = T_3^{\mathrm{T}}\mathbf{x}' \to \begin{cases} C_q \to 2^{-\frac{1}{2}}(C_q - id_{-q}^+) \\ C_{-q}^+ \to 2^{-\frac{1}{2}}(C_{-q}^+ - id_q) \\ d_q \to 2^{-\frac{1}{2}}(d_q - iC_{-q}^+) \\ d_{-q}^+ \to 2^{-\frac{1}{2}}(d_{-q}^+ - iC_q) \end{cases}$$

and

$$\mathbf{y} = T_3^{-1}\mathbf{y}' \to \begin{cases} C_q^+ \to 2^{-\frac{1}{2}}(C_q^+ + id_{-q}) \\ C_{-q} \to 2^{-\frac{1}{2}}(C_{-q} + id_q^+) \\ d_q^+ \to 2^{-\frac{1}{2}}(d_q^+ + iC_{-q}) \\ d_{-q} \to 2^{-\frac{1}{2}}(d_{-q} + iC_q^+), \end{cases}$$

which is canonical.

(4) The last transformation operator is

$$T_4 = \exp[-i\tfrac{1}{4}\pi\gamma_{10}] = 2^{-\frac{1}{2}}\begin{pmatrix} 1 & 0 & -i & 0 \\ 0 & 1 & 0 & -i \\ -i & 0 & 1 & 0 \\ 0 & -i & 0 & 1 \end{pmatrix}.$$

The matrix $U_3(q)$, with the relation (6), becomes

$$U_4(q) = A(q)\gamma_{03} + B(q)\gamma_{33} + E(q)\gamma_{32} = \begin{pmatrix} A_q + B_q & -iE_q & 0 & 0 \\ iE_q & -(A_q + B_q) & 0 & 0 \\ 0 & 0 & A_q - B_q & iE_q \\ 0 & 0 & -iE_q & -(A_q - B_q) \end{pmatrix}.$$

The corresponding linear transformation is

$$\mathbf{x} = T_4^{\mathrm{T}}\mathbf{x}' \to \begin{cases} C_q \to 2^{-\frac{1}{2}}(C_q - id_q) \\ C_{-q}^+ \to 2^{-\frac{1}{2}}(C_{-q}^+ - id_{-q}^+) \\ d_q \to 2^{-\frac{1}{2}}(d_q - iC_q) \\ d_{-q}^+ \to 2^{-\frac{1}{2}}(d_{-q}^+ - iC_{-q}^+) \end{cases}$$

and

$$\mathbf{y} = T_4^{-1}\mathbf{y}' \to \begin{cases} C_q^+ \to 2^{-\frac{1}{2}}(C_q^+ + id_q^+) \\ C_{-q} \to 2^{-\frac{1}{2}}(C_{-q} + id_{-q}) \\ d_q^+ \to 2^{-\frac{1}{2}}(d_q^+ + iC_q^+) \\ d_{-q} \to 2^{-\frac{1}{2}}(d_{-q} + iC_{-q}). \end{cases}$$

Consequently, the reduced Hamiltonian H_q becomes

$$\begin{aligned} H_q' = &-(A_q + B_q)(C_q^+C_q + C_{-q}^+C_{-q} - 1) \\ &- E_q i(C_qC_{-q} - C_{-q}^+C_q^+) \\ &- (A_q - B_q)(d_q^+d_q + d_{-q}^+d_{-q} - 1) \\ &+ E_q i(d_qd_{-q} - d_{-q}^+d_q^+). \end{aligned}$$

With the aid of the Appendix it can be easily diagonalized to become

$$\begin{aligned} H_q'' = &(C_q^+C_q + C_{-q}^+C_{-q} - 1)[(A_q + B_q)^2 + E_q^2]^{\frac{1}{2}} \\ &+ (d_q^+d_q + d_{-q}^+d_{-q} - 1)[(A_q - B_q)^2 + E_q^2]^{\frac{1}{2}}. \end{aligned}$$

Furthermore, because

$$A(q=0) = A_1 = (J_1 - J_2 - J_3 + J_4),$$
$$B(q=0) = A_3 = (J_1 + J_2 + J_3 + J_4),$$
$$E(q=0) = 0$$

and

$$A_{-q} = A_q,$$
$$B_{-q} = B_q,$$
$$E_{-q} = E_q,$$

the total Hamiltonian finally becomes

$$\begin{aligned} H'' = \sum_{(\text{all } q)} \{&(2C_q^+C_q - 1)\tfrac{1}{2}[(A_q + B_q)^2 + E_q^2]^{\frac{1}{2}} \\ &+ (2d_q^+d_q - 1)\tfrac{1}{2}[(A_q - B_q)^2 + E_q^2]^{\frac{1}{2}}\}, \end{aligned}$$

and the spectra of elementary excitations of this model in terms of pseudofermions are, therefore,

$$\epsilon_{1,2} = \tfrac{1}{2}[(A_q \pm B_q)^2 + E_q^2]^{\frac{1}{2}},$$

where

$$A_q = \frac{(A_1A_3\cos^2 q + A_2A_4\sin^2)}{(A_4^2\sin_q^2 + A_3^2\cos^2 q)^{\frac{1}{2}}},$$
$$B_q = (A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}},$$
$$E_q = \frac{(A_1A_4 - A_2A_3)\sin(q)\cos(q)}{(A_4^2\sin^2 q + A_3^2\cos^2 q)^{\frac{1}{2}}}$$

and

$$A_1 = (J_1 - J_2 - J_3 + J_4),$$
$$A_2 = (J_1 + J_2 - J_3 - J_4),$$
$$A_3 = (J_1 + J_2 + J_3 + J_4),$$
$$A_4 = (J_1 - J_2 + J_3 - J_4).$$

The result can be easily checked as follows:

(i) For $J_1 = J_2$, $J_3 = J_4$, $A_1 = A_4 = 0$, $A_2 = 2(J_1 - J_3)$, and $A_3 = 2(J_1 + J_3)$. Consequently,

$$A_q = 0,$$
$$B_q = 2(J_1 + J_3)\cos q,$$

and

$$E_q = -2(J_1 - J_3)\sin q.$$

Finally, the spectra for this case is

$$\epsilon_{1,2} = [(J_1 - J_3)^2\sin^2 q + (J_1 + J_3)^2\cos^2 q]^{\frac{1}{2}},$$

which agrees with the result obtained by Lieb, Schultz, and Mattis.[4]

(ii) For $J_1 = J_2 = J_3 = J_4 = J$, the spectra can be obtained from the above case, and it is

$$\epsilon_{1,2} = 2J \cos(q),$$

which also agrees with the result obtained by Lieb, Schultz, and Mattis.[4]

(iii) For $J_2 = 0$ and $J_3 = J_4$, $A_1 = A_4 = J_1$, $A_2 = J_1 - 2J_3$ and $A_3 = J_1 + 2J_3$. Consequently,

$$A_q = \frac{J_1(J_1 + 2J_3 \cos 2q)}{[J_1^2 \sin^2 q + (J_1 + 2J_3)\cos^2 q]^{\frac{1}{2}}},$$

$$B_q = [J_1^2 \sin^2 q + (J_1 + 2J_3)\cos^2 q]^{\frac{1}{2}},$$

and

$$E_q = \frac{2J_3^2 \sin(2q)}{[J_1^2 \sin^2 q + (J_1 + 2J_3)\cos^2 q]^{\frac{1}{2}}}.$$

Finally, the spectra are the same as the special case (iv) of Sec. III.

From the spectra so obtained, one sees that inhomogeneity in the coupling strengths affects the system in a similar way, as an anisotropicity does to it.

CONCLUSION

We present here three models of linear-lattice systems to investigate the following several points. A mathematical scheme is derived to transform the Hamiltonians into the forms in which spin operators are replaced by fermion operators. Thus, the problems become mathematically more tractable. The points discussed are as follows:

(1) Two more models, exactly solved here, show that the spectrum of elementary excitations of lattice-spin systems in a 1-dimensional case is in the form of

$$[f(J_m, \sin q, \cos q)]^{\frac{1}{2}},$$

as shown in a number of other exact calculations[1,4] on the similar models and a numerical calculation[5] on the full Heisenberg–Dirac model.

(2) The contribution of an extended force range is studied in a small way here, as its importance was suggested by Kac, Hemmer, and Uhlenbeck.[7]

(3) "Inhomogeneity" in the coupling strength which can be created by some impurities in the system affects the system similarly as an anisotropicity in it does to the system.

(4) Some models can be mathematically "decoupled" into simpler ones.

APPENDIX: THE BOGOLIUBOV–VALATIN TRANSFORMATION

We are concerned here with a 2-fermion system for the following six operators:

$$\alpha_x = C_1C_2 + C_2^+C_1^+, \qquad \beta_x = C_1^+C_2 + C_2^+C_1,$$
$$\alpha_y = i(C_1C_2 - C_2^+C_1^+), \qquad \beta_y = i(C_1^+C_2 - C_2^+C_1),$$
$$\alpha_z = C_1^+C_1 + C_2^+C_2 - 1, \text{ and } \beta_z = C_2^+C_2 - C_1^+C_1.$$

The size of the subspaces occupied by these operators in the Hilbert space is 4×4. The set of $(\alpha_x, \alpha_y, \alpha_z)$ occupies the Cooper-pair subspace of 2×2, while the set of $(\beta_x, \beta_y, \beta_z)$ occupies the single-particle subspace of the remaining 2×2 subspace. Consequently, one can make out two kinds of transformations, one for each of these subspaces.

(1) R transformation for the Cooper-pair subspace:

(a) R_1 transformation:

$$C_1 = d_1 \cos\theta + d_2^+ \sin\theta,$$
$$C_2 = d_2 \cos\theta - d_1^+ \sin\theta,$$
$$\alpha_x \to \alpha_x' \cos 2\theta + \alpha_z' \sin 2\theta,$$
$$\alpha_y \to \alpha_y',$$
$$\alpha_z \to \alpha_z' \cos 2\theta - \alpha_x' \sin 2\theta$$

and

$$\beta_x \to \beta_x',$$
$$\beta_y \to \beta_y',$$
$$\beta_z \to \beta_z'.$$

(b) R_2 transformation:

$$C_1 = d_1 \cos\theta + id_2^+ \sin\theta,$$
$$C_2 = d_2 \cos\theta - id_1^+ \sin\theta,$$
$$\alpha_x \to \alpha_x',$$
$$\alpha_y \to \alpha_y' \cos 2\theta - \alpha_z' \sin 2\theta,$$
$$\alpha_z \to \alpha_z' \cos 2\theta + \alpha_y' \sin 2\theta$$

and

$$\beta_x \to \beta_x',$$
$$\beta_y \to \beta_y',$$
$$\beta_z \to \beta_z'.$$

(2) S transformation for the single-particle subspace.

(a) S_1 transformation:

$$C_1 = d_1^+ \cos\theta + d_2 \sin\theta,$$
$$C_2 = d_2 \cos\theta - d_1^+ \sin\theta,$$
$$\alpha_x \to \beta_x',$$
$$\alpha_y \to \beta_y',$$
$$\alpha_z \to \beta_z'$$

and

$$\beta_x \to \alpha_x' \cos 2\theta + \alpha_z' \sin 2\theta,$$
$$\beta_y \to \alpha_y',$$
$$\beta_z \to \alpha_z' \cos 2\theta - \alpha_x' \sin 2\theta.$$

(b) S_2 transformation:

$$C_1 = d_1^+ \cos\theta - id_2 \sin\theta,$$
$$C_2 = d_2 \cos\theta - id_1^+ \sin\theta,$$
$$\alpha_x \to \beta_x',$$
$$\alpha_y \to \beta_y',$$
$$\alpha_z \to \beta_z'$$

and

$$\beta_x \to \alpha_x',$$
$$\beta_y \to \alpha_y' \cos 2\theta - \alpha_z' \sin 2\theta,$$
$$\beta_z \to \alpha_z' \cos 2\theta + \alpha_y' \sin 2\theta.$$

With these basic facts in mind, one immediately can utilize the following often used formula.

Under the transformation by the operator $\exp(\theta X)$,

$$(mZ + nY) \to Z[m \cosh(2\theta) - in \sinh(2\theta)] + Y[n \cosh(2\theta) + im \sinh(2\theta)] = Z(m^2 + n^2)^{\frac{1}{2}},$$

where m and n are constants and

$$\cosh(2\theta) = m/(m^2 + n^2)^{\frac{1}{2}},$$
$$\sinh(2\theta) = in/(m^2 + n^2)^{\frac{1}{2}}.$$

ACKNOWLEDGMENT

The author is grateful to his wife for taking interest in this work and helping to type the manuscript for him.

[1] D. C. Mattis, *Theory of Magnetism* (Harper and Row, New York, 1965); E. Lieb, T. Schultz, and D. M. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961).
[2] R. Whitlock and P. R. Zilsel, Phys. Rev. **131**, 2409 (1963).
[3] D. C. Mattis, Ref. 1.
[4] E. Lieb, T. Schultz, and D. M. Mattis, Ref. 1.
[5] J. Des Cloiseaux and J. J. Pearson, Phys. Rev. **128**, 2131 (1962).
[6] E. Lieb, T. Schultz, and D. M. Mattis, see Ref. 1. P. W. Anderson, Phys. Rev. **86**, 694 (1952); R. Kubo, Rev. Mod. Phys. **25**, 344 (1953); P. W. Kasteleijin, Physics **18**, 104 (1952); H. Takeda and T. Nakamura, J. Phys. Soc. (Japan) **11**, 919 (1956); W. Marshall, Proc. Roy. Soc. **A232**, 48 (1955); R. Orbach, Phys. Rev. **116**, 1089 (1959); T. W. Ruijgrok and S. Rodriguez, Phys. Rev. **119**, 596 (1959); H. L. Davis, Phys. Rev. **120**, 789 (1960); T. Oguchi, J. Phys. Chem. Solids **24**, 1049 (1963).
[7] M. Kac, *1966 Brandeis University Summer Institute in Theoretical Physics* (Gordon and Breach, New York, 1968), Vol. I, pp. 245.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Criteria of Accuracy of Approximate Wavefunctions

FRANK WEINHOLD*†
Department of Chemistry, University of California, Berkeley, California 94720

(Received 6 November 1969)

The accuracy with which a trial function ϕ approximates the true wavefunction ψ is quantitatively assessed by the overlap integral $S = \langle\phi \mid \psi\rangle$. Upper and lower bounds to S therefore furnish direct criteria of accuracy of the approximation ϕ and also of the associated physical properties. The available literature on overlap estimates is assembled and critically discussed from a unified point of view, based upon a method of determinantal inequalities. In particular, the relationships among the various approaches are pointed up, several results are extended or generalized, and some new results are obtained. Finally, the various formulas are illustrated by numerical applications to some simple soluble problems.

1. INTRODUCTION

According to the fundamental quantum-mechanical Schrödinger equation, a physical system represented by the Hamiltonian operator H has stationary states described by certain wavefunctions ψ which are solutions of the eigenvalue problem

$$H\psi = E\psi, \tag{1.1}$$

where E is the energy of the state. Other properties of the system are calculated from this (normalized) wavefunction as

$$\langle\psi| F |\psi\rangle \equiv \int \psi^* F \psi \, d\tau$$

for the appropriate Hermitian operator F.

In practice, however, it is almost never possible to actually solve Eq. (1.1) for systems of interest. Rather, one must seek some *approximate* wavefunction ϕ which, it is hoped, resembles the true solution ψ as closely as possible. For example, following the Ritz variational procedure (for the ground state), one might adjust ϕ in such a way as to minimize the energy integral $\langle\phi| H |\phi\rangle$. Other properties $\langle\psi| F |\psi\rangle$ are then approximated correspondingly as $\langle\phi| F |\phi\rangle$. However, the approximate nature of this procedure raises several theoretical and practical difficulties: How serious are the errors remaining in the approximation ϕ? How does the nature and extent of these errors effect the calculated values of various properties? What criteria may be employed to properly assess these effects?

For quantitative purposes, the natural criterion of accuracy for an approximate wavefunction ϕ is the approach toward unity of the *overlap integral*

$$S \equiv \langle\phi \mid \psi\rangle = \int \phi^* \psi \, d\tau, \tag{1.2}$$

and

$$\beta_x \to \alpha_x',$$
$$\beta_y \to \alpha_y' \cos 2\theta - \alpha_z' \sin 2\theta,$$
$$\beta_z \to \alpha_z' \cos 2\theta + \alpha_y' \sin 2\theta.$$

With these basic facts in mind, one immediately can utilize the following often used formula.

Under the transformation by the operator $\exp(\theta X)$,

$$(mZ + nY) \to Z[m \cosh(2\theta) - in \sinh(2\theta)] + Y[n \cosh(2\theta) + im \sinh(2\theta)] = Z(m^2 + n^2)^{\frac{1}{2}},$$

where m and n are constants and

$$\cosh(2\theta) = m/(m^2 + n^2)^{\frac{1}{2}},$$
$$\sinh(2\theta) = in/(m^2 + n^2)^{\frac{1}{2}}.$$

ACKNOWLEDGMENT

The author is grateful to his wife for taking interest in this work and helping to type the manuscript for him.

[1] D. C. Mattis, *Theory of Magnetism* (Harper and Row, New York, 1965); E. Lieb, T. Schultz, and D. M. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961).
[2] R. Whitlock and P. R. Zilsel, Phys. Rev. **131**, 2409 (1963).
[3] D. C. Mattis, Ref. 1.
[4] E. Lieb, T. Schultz, and D. M. Mattis, Ref. 1.
[5] J. Des Cloiseaux and J. J. Pearson, Phys. Rev. **128**, 2131 (1962).
[6] E. Lieb, T. Schultz, and D. M. Mattis, see Ref. 1. P. W. Anderson, Phys. Rev. **86**, 694 (1952); R. Kubo, Rev. Mod. Phys. **25**, 344 (1953); P. W. Kasteleijin, Physics **18**, 104 (1952); H. Takeda and T. Nakamura, J. Phys. Soc. (Japan) **11**, 919 (1956); W. Marshall, Proc. Roy. Soc. **A232**, 48 (1955); R. Orbach, Phys. Rev. **116**, 1089 (1959); T. W. Ruijgrok and S. Rodriguez, Phys. Rev. **119**, 596 (1959); H. L. Davis, Phys. Rev. **120**, 789 (1960); T. Oguchi, J. Phys. Chem. Solids **24**, 1049 (1963).
[7] M. Kac, *1966 Brandeis University Summer Institute in Theoretical Physics* (Gordon and Breach, New York, 1968), Vol. I, pp. 245.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Criteria of Accuracy of Approximate Wavefunctions

FRANK WEINHOLD*†
Department of Chemistry, University of California, Berkeley, California 94720

(Received 6 November 1969)

The accuracy with which a trial function ϕ approximates the true wavefunction ψ is quantitatively assessed by the overlap integral $S = \langle\phi \mid \psi\rangle$. Upper and lower bounds to S therefore furnish direct criteria of accuracy of the approximation ϕ and also of the associated physical properties. The available literature on overlap estimates is assembled and critically discussed from a unified point of view, based upon a method of determinantal inequalities. In particular, the relationships among the various approaches are pointed up, several results are extended or generalized, and some new results are obtained. Finally, the various formulas are illustrated by numerical applications to some simple soluble problems.

1. INTRODUCTION

According to the fundamental quantum-mechanical Schrödinger equation, a physical system represented by the Hamiltonian operator H has stationary states described by certain wavefunctions ψ which are solutions of the eigenvalue problem

$$H\psi = E\psi, \tag{1.1}$$

where E is the energy of the state. Other properties of the system are calculated from this (normalized) wavefunction as

$$\langle\psi| F |\psi\rangle \equiv \int \psi^* F \psi \, d\tau$$

for the appropriate Hermitian operator F.

In practice, however, it is almost never possible to actually solve Eq. (1.1) for systems of interest. Rather, one must seek some *approximate* wavefunction ϕ which, it is hoped, resembles the true solution ψ as closely as possible. For example, following the Ritz variational procedure (for the ground state), one might adjust ϕ in such a way as to minimize the energy integral $\langle\phi| H |\phi\rangle$. Other properties $\langle\psi| F |\psi\rangle$ are then approximated correspondingly as $\langle\phi| F |\phi\rangle$. However, the approximate nature of this procedure raises several theoretical and practical difficulties: How serious are the errors remaining in the approximation ϕ? How does the nature and extent of these errors effect the calculated values of various properties? What criteria may be employed to properly assess these effects?

For quantitative purposes, the natural criterion of accuracy for an approximate wavefunction ϕ is the approach toward unity of the *overlap integral*

$$S \equiv \langle\phi \mid \psi\rangle = \int \phi^* \psi \, d\tau, \tag{1.2}$$

which may be taken real and positive. As S approaches unity, the integral $\int |\phi - \psi|^2 \, d\tau$ tends to zero, and ϕ and ψ become equal in the quantum mechanical sense of mean convergence. Since S cannot usually be known exactly, the accuracy of ϕ is assessed instead by determining *upper and lower bounds*[1] to the true value of S.

As discussed below (Sec. 1A), the overlap S also enters the criteria of accuracy of specific properties, including the available formulas for error bounds to expectation values. Such formulas are found to depend in a very sensitive manner on the value employed for S, and this has lent new urgency to the problem of accurately estimating this important quantity.

Previous discussions of overlap estimates are actually fairly numerous, although widely scattered throughout the physical and mathematical literature. It appears that insufficient attention has been given to the problem of examining and comparing the various approaches, and some needless duplication of effort has resulted. We wish here to review the available literature,[2] assembling the known results and clarifying the relationships among them as far as possible from a unified point of view. New derivations are presented, and some new results are obtained in addition to extensions and generalizations of previous results. Some illustrative numerical applications are included, and various considerations for the practical applications are pointed out.

A. Accuracy of Physical Properties

Clearly, the overlap integral S of (1.2) measures the "fit" of ϕ to ψ only in some global sense, with all regions of configuration space weighted equally. But a particular property depends principally on regions of configuration space weighted most heavily by the corresponding operator F. Thus, if r is the usual electronic radial coordinate, it is known that the accuracy of $\langle\phi| r^{-2} |\phi\rangle$ depends mainly on the accuracy of ϕ in the region near the nucleus, whereas $\langle\phi| r^2 |\phi\rangle$ depends more sensitively on outer regions.

For any particular operator F, therefore, one might consider the *weighted* overlap integral

$$\langle\phi| F |\psi\rangle = \int \phi^* F \psi \, d\tau \tag{1.3}$$

in place of (1.2) as the natural criterion of accuracy. But for arbitrary Hermitian F, it has been shown[3] that this quantity has upper and lower bounds given by

$$\langle\phi| F |\psi\rangle \lessgtr S\langle F\rangle \pm \Delta F(1 - S^2)^{\frac{1}{2}}, \tag{1.4}$$

where ΔF, the "width" of F in the state ϕ, is defined by

$$(\Delta F)^2 \equiv \langle F^2\rangle - \langle F\rangle^2. \tag{1.5}$$

Here we have also introduced the abbreviation

$$\langle F\rangle \equiv \langle\phi| F |\phi\rangle,$$

and $\langle\phi| F |\psi\rangle$ is assumed real.

According to (1.4), the weighted overlap integral (1.3) is bounded by quantities which depend only on the width ΔF and the "global" overlap S, or on upper and lower bounds $S_u \geq S \geq S_l$. Suppose, for example, that $\langle F\rangle$ is a positive number. Then, for the lower bound of (1.4), one uses the lower bound S_l,

$$\langle\phi| F |\psi\rangle \geq S_l\langle F\rangle - \Delta F(1 - S_l^2)^{\frac{1}{2}}. \tag{1.6}$$

For the upper member of (1.4), it is always safe to use

$$\langle\phi| F |\psi\rangle \leq S_u\langle F\rangle + \Delta F(1 - S_l^2)^{\frac{1}{2}}, \tag{1.7}$$

but this result can often be improved. If, as is usually the case,

$$\frac{(\Delta F)^2}{\langle F\rangle^2} \geq \frac{1 - S_l^2}{S_l^2} \geq \frac{1 - S^2}{S^2}, \tag{1.8a}$$

then

$$\langle\phi| F |\psi\rangle \leq S_l\langle F\rangle + \Delta F(1 - S_l^2)^{\frac{1}{2}}, \tag{1.8b}$$

whereas, if

$$\frac{(\Delta F)^2}{\langle F\rangle^2} \leq \frac{1 - S_u^2}{S_u^2} \leq \frac{1 - S^2}{S^2}, \tag{1.9a}$$

then

$$\langle\phi| F |\psi\rangle \leq S_u\langle F\rangle + \Delta F(1 - S_u^2)^{\frac{1}{2}}. \tag{1.9b}$$

If neither (1.8a) nor (1.9a) can be assured, (1.7) must be used.

Therefore, according to (1.6)–(1.9) the problem of assessing the accuracy of a calculated property is again closely related to the problem of determining accurate upper and lower bounds to the overlap integral S. This conclusion persists when (1.4) is replaced by a somewhat sharper inequality involving matrix elements of FH and H^2,[4] or when one considers the available formulas for error bounds to expectation values.[5]

B. Method of Determinantal Inequalities

The principal method employed in the present analysis is based on the positivity of the Gram determinants.[6] A convenient geometrical language is possible when the wavefunctions are taken to represent vectors in Hilbert space.

Consider, therefore, a certain set of n vectors $|v_1\rangle$, $|v_2\rangle, \cdots, |v_n\rangle$ which form the columns of an operator A. Then the (Hermitian) matrix $A^\dagger A$, with elements

$$(A^\dagger A)_{ij} = \langle v_i \,|\, v_j\rangle, \tag{1.10}$$

is called the *overlap* (or *metric*) matrix of the system $|v_i\rangle$. Clearly, $A^\dagger A$ is positive semidefinite (i.e., has no

negative eigenvalues), $A^\dagger A \geq 0$, since, for *any* vector $|x\rangle$,

$$\langle x|\, A^\dagger A\, |x\rangle = \langle Ax \,|\, Ax\rangle = \|Ax\|^2 \geq 0$$

according to fundamental properties of the scalar product. Then also

$$G = \det |A^\dagger A| \geq 0, \tag{1.11}$$

since G is the product of all eigenvalues of $A^\dagger A$.

In (1.11), the determinant $G = G\{|v_i\rangle\}$ is called the *Gram determinant* (or "Gramian") of the vectors $|v_i\rangle$. Its value is just the squared volume of the parallelepiped generated by the $|v_i\rangle$, and approaches zero as the vectors become linearly dependent. Since the positive semidefiniteness of $A^\dagger A$ is both necessary and sufficient for the *existence* of a set of vectors $|v_i\rangle$ having the required norms and inner products (1.10), relations such as (1.11) are the *strongest possible* restrictions that can be placed on a particular element $\langle v_i \,|\, v_j\rangle$ if all other elements of G are known. This "best possible" feature of the determinantal inequalities is of considerable practical importance in the analysis.

A simple generalization may be mentioned. Let M be any Hermitian operator which is positive semidefinite, at least on the subspace spanned by the $|v_i\rangle$. That is, if P_n denotes the projector onto the subspace spanned by $|v_1\rangle, |v_2\rangle, \cdots, |v_n\rangle$, we assume that

$$\tilde{M} \equiv P_n M P_n \geq 0. \tag{1.12}$$

Then M (or, equivalently, $\tilde{M}$) may be adopted as the *metric* for a new inner product $[\ ,\ ]$,

$$[v_i, v_j] \equiv \langle v_i|\, M\, |v_j\rangle,$$

which may be used in place of (1.10). This is equivalent to considering a new set of vectors

$$|\tilde{v}_i\rangle \equiv \tilde{M}^{\frac{1}{2}} |v_i\rangle$$

in the original treatment (1.10)–(1.11), where the definition of the square-root operator $\tilde{M}^{\frac{1}{2}}$ is made possible by (1.12).

A simple example may be used to illustrate the general procedure. Consider normalized vectors ψ, ϕ, and χ, where $S_1 \equiv \langle \psi \,|\, \chi\rangle$ and $S_2 \equiv \langle \phi \,|\, \chi\rangle$ are known but $S \equiv \langle \phi \,|\, \psi\rangle$ is unknown. If these numbers are taken to be real, the positivity of the Gramian, $G = G(\psi, \phi, \chi)$,

$$G = \begin{vmatrix} 1 & S & S_1 \\ S & 1 & S_2 \\ S_1 & S_2 & 1 \end{vmatrix} \geq 0,$$

requires that

$$S^2 - 2SS_1S_2 + S_1^2 + S_2^2 - 1 \leq 0,$$

which shows that S must lie between the two roots of a simple quadratic equation. The final result is[7]

$$S \lessgtr S_1S_2 \pm [(1 - S_1^2)(1 - S_2^2)]^{\frac{1}{2}}, \tag{1.13}$$

which is the strongest possible bound for S in terms of the given S_1, S_2.

2. UPPER BOUNDS TO OVERLAP

A. Rayner's Upper Bound

As the first formula for upper bounds to the overlap S,[8] we derive by the method of determinantal inequalities an effective formula due originally to Rayner.[9] For the eigenvalue problem $H\psi_0 = E_0\psi_0$, we begin by obtaining a set of m normalized functions $\chi_1, \chi_2, \cdots, \chi_m$, which are chosen to satisfy

$$\langle \chi_i|\, (H - E_0)^2\, |\chi_j\rangle = 0, \quad i \neq j. \tag{2.1}$$

This might be done, e.g., by the Schmidt process or by diagonalizing an $m \times m$ representation of the operator $(H - E_0)^2$. Then, for the set of vectors ψ_0, ϕ, $(H - E_0)\chi_i$, the Gramian $G = G\{\psi_0, \phi, (H - E_0)\chi_i\}$ is found to be

$$G = \begin{vmatrix} 1 & S & 0 & 0 & 0 & \cdots \\ S & 1 & k_1 & k_2 & k_3 & \cdots \\ 0 & k_1^* & K_1 & 0 & 0 & \cdots \\ 0 & k_2^* & 0 & K_2 & 0 & \cdots \\ 0 & k_3^* & 0 & 0 & K_3 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix} \geq 0, \tag{2.2}$$

where

$$k_i \equiv \langle \phi|\, H - E_0\, |\chi_i\rangle,$$
$$K_i \equiv \langle \chi_i|\, (H - E_0)^2\, |\chi_i\rangle.$$

Using the Laplace expansion down the first two columns, the determinant (2.2) is easily evaluated as

$$G = (1 - S^2) \prod_{i=1}^{m} K_i + \sum_{j=1}^{m} (-1)^j k_j^* \times \left((-1)^{j+1} k_j \prod_{\substack{i=1 \\ (i \neq j)}}^{m} K_i \right) \geq 0.$$

This is immediately rewritten as

$$S^2 \leq 1 - \sum_{j=1}^{m} \frac{|k_j|^2}{K_j}$$

or

$$S^2 \leq 1 - \sum_{j=1}^{m} \frac{|\langle \phi|\, H - E_0\, |\chi_j\rangle|^2}{\langle \chi_j|\, (H - E_0)^2\, |\chi_j\rangle}, \tag{2.3}$$

which will be called *Rayner's inequality*. For the excited state k, the corresponding result is

$$|\langle\phi \mid \psi_k\rangle|^2 \leq 1 - \sum_{j=1}^{m} \frac{|\langle\phi|\, H - E_k\,|\chi_j\rangle|^2}{\langle\chi_j|\,(H - E_k)^2\,|\chi_j\rangle}. \quad (2.4)$$

The special case $m = 1$ of (2.3), namely,

$$S^2 \leq 1 - |\langle\phi|\, H - E_0\,|\chi\rangle|^2/\langle\chi|\,(H - E_0)^2\,|\chi\rangle, \quad (2.5)$$

was derived independently by Wang.[10] The special case $\chi = \phi$ of (2.5), which may also be written in the form

$$S^2 \leq (\Delta H)^2/\langle(H - E_0)^2\rangle, \quad (2.6)$$

had been previously found by Combet-Farnoux and Allard,[11] and has been subsequently rediscovered (also written as a lower bound to the eigenvalue) by several authors.[12–15] For future reference, we note also the more general form of (2.6):

$$|\langle\phi \mid \psi_k\rangle|^2 \leq (\Delta H)^2/\langle(H - E_k)^2\rangle. \quad (2.7)$$

B. Gordon's Upper Bounds

Recently, Gordon[15] has obtained a sequence of upper bounds to S based on the method of moments and the theory of generalized Gaussian quadrature. In this method, one requires the sequence of matrix elements $\langle H\rangle, \langle H^2\rangle, \cdots, \langle H^{2n}\rangle$ and the eigenvalue E_0, but the resulting bounds were only defined *implicitly* [except for the first member, (2.6)] in terms of the solutions of a certain set of equations. It was shown subsequently by Wang[10] that the same results could be obtained by choosing the function

$$\chi = \sum_n c_n H^n \phi$$

in formula (2.5), and determining the c_n to optimize the bound, but again the final result was not determined.

We present here an alternative derivation of Gordon's bounds which generalizes his result (by omitting the restrictions that n in the moments $\langle H^n\rangle$ be sequential and positive) and which displays the final result explicitly.

Consider the Gramian G of the vectors $S^{-1}\psi_0$, $H^i\phi$, $H^j\phi$, $H^k\phi, \cdots$, where $i, j, k, \cdots$ are arbitrary positive or negative integers[16]:

$$G = \begin{vmatrix} S^{-2} & E_0^i & E_0^j & E_0^k & \cdots \\ E_0^i & \langle H^{2i}\rangle & \langle H^{i+j}\rangle & \langle H^{i+k}\rangle & \cdots \\ E_0^j & \langle H^{i+j}\rangle & \langle H^{2j}\rangle & \langle H^{j+k}\rangle & \cdots \\ E_0^k & \langle H^{i+k}\rangle & \langle H^{j+k}\rangle & \langle H^{2k}\rangle & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix} \geq 0. \quad (2.8)$$

Now let D be the determinant obtained by deleting the first row and column of G,

$$D \equiv \begin{vmatrix} \langle H^{2i}\rangle & \langle H^{i+j}\rangle & \cdots \\ \langle H^{i+j}\rangle & \langle H^{2j}\rangle & \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix}, \quad (2.9)$$

and let $-\tilde{D}$ be the determinant obtained by setting the first element of G to zero,

$$\tilde{D} \equiv -\begin{vmatrix} 0 & E_0^i & E_0^j & \cdots \\ E_0^i & \langle H^{2i}\rangle & \langle H^{i+j}\rangle & \cdots \\ E_0^j & \langle H^{i+j}\rangle & \langle H^{2j}\rangle & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix}. \quad (2.10)$$

Then the bound (2.8) is immediately written explicitly as (note that $D, \tilde{D} \geq 0$)

$$S^2 \leq D/\tilde{D}. \quad (2.11)$$

Gordon's bounds are obtained as the special case $\{i, j, k \cdots\} = \{0, 1, 2 \cdots\}$ of (2.11). This identification is assured, since both Gordon's bounds and the inequality (2.8) are known to be the *best possible* bounds that can be placed on S for the amount of information required, so that they must also be identical. The result (2.11) is also readily generalized to any excited state ψ_k.

For atomic and molecular systems it is unfortunate that the moments $\langle H^n\rangle$ generally diverge for $n > 2$, because ϕ behaves improperly at the singularities of the Coulomb potential. However, for other (nonsingular) potentials, the upper bounds (2.11) are apparently very effective, and even the special case (2.6), which requires only $\langle H\rangle$ and $\langle H^2\rangle$, appears to give good results.[15]

It is well known that, when ϕ is sufficiently accurate, even the integral $\langle H^2\rangle$ may be quite difficult to compute.[17] In this respect, Rayner's bound (2.3) enjoys a considerable advantage, since the χ_j may be chosen to simplify the calculation of $\langle\chi_j|\, H^2\,|\chi_j\rangle$. Particularly, the use of Gaussian functions may be considered.[18]

C. Upper Bounds of Rayner and Delves

A principal disadvantage of the methods of Secs. 2A and 2B is the requirement that the energy level E_0 (or E_k) be known independently.[19] This difficulty is partially avoided in a method introduced by Rayner[20] and generalized subsequently by Delves.[21]

In the Rayner–Delves method, one introduces an estimate α which is at least *closer* to E_k, the eigenvalue of interest, than to any other eigenvalue. Then, for positive integer n, it is shown[21] that

$$|\langle\phi \mid \psi_k\rangle|^2 \leq \langle(H-\alpha)^{-2n+2}\rangle^n / \langle(H-\alpha)^{-2n}\rangle^{n-1}, \quad n = 1, 2, 3, \cdots, \tag{2.12}$$

and, further, that the bound sharpens monotonically with increasing n.

For the ground state $k = 0$, if α is also known to be a *lower* bound to E_0, one can follow the Rayner treatment to obtain the result

$$S^2 \leq \langle(H-\alpha)^{-n+1}\rangle^n / \langle(H-\alpha)^{-n}\rangle^{n-1}, \quad E_0 > \alpha, \tag{2.13}$$

which requires moments of lower absolute order than does (2.12). Various similar formulas are also considered by Rayner, but (2.13) is shown to be the most efficient.

The conspicuous difficulty of the Rayner–Delves bounds is the required knowledge of negative-moment integrals such as $\langle(H-\alpha)^{-n}\rangle$. These are obtained in the form $\langle f_n \mid f_0\rangle = \langle f_{n-m} \mid f_m\rangle$ from the solutions of an inhomogeneous set of equations

$$(H-\alpha)f_{n+1} = f_n, \quad n = 0, 1, 2, \cdots, \tag{2.14}$$

where $f_0 \equiv \phi$. But the system (2.14) cannot usually be solved, and an approximation method, such as a variation–perturbation approach,[22] must be introduced. Further, although only a minimum of information about the energy level E_k is required, the convergence of (2.12) and (2.13) will in fact depend strongly on how well α approximates the true eigenvalue. Finally, since the system (2.14), if soluble, will be solved sequentially, it would be preferable to use a formula in which all lower moments could also be employed as additional information.

To illustrate these formulas numerically, we consider the example (suggested by Rayner) of a particle in a 1-dimensional box of unit length, with $\phi = 1$ and $\alpha = 0$. For comparison, we consider also the formula

$$S^2 \leq (E_0-\alpha)^n\langle(H-\alpha)^{-n}\rangle, \quad n = 1, 2, 3, \cdots, \text{ for } \alpha < E_0, \quad = 2, 4, 6, \cdots, \text{ otherwise}, \tag{2.15}$$

derived from the Gramian of ψ_0 and $(H-\alpha)^{-n/2}\phi$, as well as the corresponding member of the generalized upper bound (2.11) which includes all moments up to the same order. Table I exhibits the upper bounds (2.13), (2.15), and (2.11) for various values of n. From this table, one observes the advantage of introducing E_0 explicitly, and of using all lower moments available.

TABLE I. Upper bounds to overlap for a particle in a 1-dimensional box, as calculated from the formulas of the text.

n	Eq. (2.13)	Eq. (2.15)	Eq. (2.11)
1	1.000000	0.822467	0.822467
2	0.833333	0.811742	0.811053
3	0.813841	0.810695	0.810582
4	0.811031	0.810583	0.810570
		exact = 0.810569	

D. Additional Upper Bounds

Recalling the discussion of Sec. 1B, we introduce a new metric $H - E_0$ (≥ 0) and consider the vectors ϕ, ψ_k. The Gramian is

$$G = \begin{vmatrix} \langle H\rangle - E_0 & (E_k - E_0)\langle\phi \mid \psi_k\rangle \\ (E_k - E_0)\langle\psi_k \mid \phi\rangle & E_k - E_0 \end{vmatrix} \geq 0,$$

which gives immediately an elementary upper bound for $|\langle\phi \mid \psi_k\rangle|^2$,

$$|\langle\phi \mid \psi_k\rangle|^2 \leq (\langle H\rangle - E_0)/(E_k - E_0), \tag{2.16}$$

which is sometimes useful (cf. Sec. 3E).

Finally, we wish to introduce an upper bound which does not require any knowledge of the energy levels E_k. Rather, it is necessary to know the true value (or an upper bound) of some *other* property $\langle\psi| F |\psi\rangle$ which corresponds to a positive semidefinite operator $F \geq 0$ and for which also $\langle\phi| F |\phi\rangle$ is too large. In such a case, one has formulas such as[3]

$$\langle\psi| F |\psi\rangle \geq [S\langle F^\nu\rangle - \Delta F^\nu(1-S^2)^{\frac{1}{2}}]^2 / \langle F^{2\nu-1}\rangle, \tag{2.17}$$

where F^ν denotes the νth power of F, and ν is arbitrary so long as the expression in brackets is nonnegative. Then, clearly, the value of S for which (2.17) becomes an *equality* is surely too large, so that an upper bound to S is obtained as a solution of the equation

$$S^2\langle F^{2\nu}\rangle - 2S\langle F^\nu\rangle\langle F^{2\nu-1}\rangle^{\frac{1}{2}}\langle\psi| F |\psi\rangle^{\frac{1}{2}} + \langle\psi| F |\psi\rangle\langle F^{2\nu-1}\rangle - (\Delta F^\nu)^2 = 0. \tag{2.18}$$

A similar procedure may be used for any of the other available lower bound formulas analogous to (2.17).[5b]

E. Numerical Applications and Discussion

A simple numerical example may be used to suggest the relative effectiveness, in a particular case, of the several formulas discussed in Secs. 2A–2D. For this purpose, we choose the simple soluble problem of atomic hydrogen,[23] so that the exact value of $S = |\langle\phi \mid \psi_0\rangle|$ is known in advance. For ϕ, we select the somewhat artificial function

$$\phi \equiv (32/13\pi)^{\frac{1}{2}}(1+r)e^{-2r}, \tag{2.19}$$

for which

$$S^2 = 0.864. \tag{2.20}$$

This is not a particularly good approximation to the true ground-state wavefunction $\psi_0 = \pi^{-\frac{1}{2}}e^{-r}$, but it is chosen in such a way that the integrals $\langle H^3\rangle$ and $\langle H^4\rangle$ do not diverge. This will permit an application of Gordon's bound (2.11), which, in the more general case, would not be useful.

For the simple upper bound (2.6) of Combet-Farnoux and Allard, which uses values of $\langle H^2\rangle$, $\langle H\rangle$, and E_0, the result is found to be

$$S^2 \leq 0.923, \tag{2.21}$$

and this is also the simplest case of Gordon's bound (2.11). If we include also $\langle H^3\rangle$ and $\langle H^4\rangle$, the next member of Gordon's sequence of bounds (2.11) is found to give

$$S^2 \leq 0.916, \tag{2.22}$$

which indicates that the convergence of Gordon's bounds towards the true value (2.20) may be rather slow when ϕ is not a good approximation to ψ_0.

Both bounds above could be regarded as special cases of the Rayner formula (2.5) for suitable trial functions χ. But,to further illustrate Rayner's formula, we consider here also the simple functional form

$$\chi \sim [1 + (\alpha - 1)r]e^{-\alpha r}, \tag{2.23}$$

where "$\sim$" denotes that the function is not yet normalized. Inserting (2.23) into (2.5) and choosing the parameter α to optimize the bound, we obtain

$$S^2 \leq 0.865 \quad (\alpha = 1.17), \tag{2.24}$$

and this result could be improved as much as desired by including additional variational functions χ_i in (2.3). Rayner's formula (2.3) certainly appears to be the most flexible and practical means available of systematically improving the upper bounds when a knowledge of E_0 is available.

Finally,[24] we consider the formula (2.18), which requires a knowledge of some property other than the energy. For this purpose, we choose $F = r^{-1}$, which is overestimated in the approximation (2.19),

$$\langle r^{-1}\rangle = \tfrac{19}{13}, \quad \langle\psi_0|\, r^{-1}\,|\psi_0\rangle = 1,$$

atomic units being used throughout. The result of (2.18) is found to be

$$S^2 \leq 0.928 \quad (\nu = 0.39),$$

which is not much worse than (2.21). In this special case, the true value of $\langle\psi_0|\, r^{-1}\,|\psi_0\rangle$ is known from the *virial theorem* if E_0 is known, so that no additional experimental information is required.

3. LOWER BOUNDS TO OVERLAP

An upper bound to the overlap S determines that the approximation ϕ is still defective to at least a certain degree. However, it is practically more useful (and more reassuring) to determine how *good,* rather than how bad, is the approximation ϕ. For this reason, the determination of lower bounds to S is usually of principal interest and importance.

A. The Eckart Criterion

In his important early paper on the theory and calculation of screening constants, Eckart[25] established a simple relation which, in many respects, is still the best simple criterion of accuracy of an approximate ground state wavefunction. To obtain this relation, we note that the operator[26] $H - E_1$ is *positive semidefinite* in the subspace of functions orthogonal to ψ_0, so that

$$Q(H - E_1)Q \geq 0, \quad \text{where} \quad Q \equiv 1 - |\psi_0\rangle\langle\psi_0|. \tag{3.1}$$

Then, immediately,

$$0 \leq \langle Q(H - E_1)Q\rangle = \langle H\rangle - E_1 - S^2(E_0 - E_1),$$

which gives Eckart's criterion[25]

$$S^2 \geq (E_1 - \langle H\rangle)/(E_1 - E_0). \tag{3.2}$$

When terms of order $(1 - S^2)^2$ can be neglected, an inexact form of this relation (introduced by Eckart himself) is obtained

$$S \gtrsim 1 - (\langle H\rangle - E_0)/[2(E_1 - E_0)], \tag{3.3}$$

and this result is sometimes quoted as Eckart's formula.[27] However, there appears to be no compelling reason for this additional approximation, and we refer only to the correct form (3.2) as "the Eckart criterion." Formula (3.2) has often been rediscovered in the literature.[28]

The Eckart criterion becomes exact in the limit as ϕ is a linear combination of only ψ_0 and ψ_1, as is apparent from the derivation. Furthermore, one confirms readily that the inequality (3.2) remains true if E_0 and E_1 are replaced by any *lower* bounds to their true values, so that a strictly theoretical result is still possible if E_0 and E_1 are not known experimentally.[29]

B. Weinberger's Formula

An important generalization and strengthening of the Eckart relation was obtained by Weinberger.[30] In this approach, one requires a knowledge of the $n + 2$ lowest eigenvalues $E_0, E_1, \cdots, E_{n+1}$ (or lower bounds) and an orthonormal set of approximations $\phi_0, \phi_1, \cdots, \phi_n$ to the eigenfunctions $\psi_0, \psi_1, \cdots, \psi_n$. It is assumed that the approximations ϕ_i are solutions

of a secular determinant

$$\langle\phi_i|\, H\,|\phi_j\rangle = J_i\delta_{ij}$$

and that the corresponding Rayleigh–Ritz estimates $J_i = \langle\phi_i|\, H\,|\phi_i\rangle$ are at least sufficiently accurate that

$$E_0 \le J_0 \le E_1 \le J_1 \le \cdots \le E_n \le J_n \le E_{n+1}. \quad (3.4)$$

Under these assumptions, Weinberger[30] was able to show that

$$S^2 \ge \frac{E_{n+1}-J_0}{E_{n+1}-E_0}\prod_{\nu=1}^{n}\frac{E_\nu - J_0}{E_\nu - E_0}\frac{J_\nu - E_0}{J_\nu - J_0}, \quad (3.5)$$

with $S^2 \equiv |\langle\phi_0 \mid \psi_0\rangle|^2$. The corresponding result for excited state k is

$$\langle\phi_k \mid \psi_k\rangle|^2 \ge \frac{E_{n+1}-J_k}{E_{n+1}-E_k}\prod_{\substack{\nu=0\\(\nu\neq k)}}^{n}\frac{E_\nu - J_k}{E_\nu - E_k}\frac{J_\nu - E_k}{J_\nu - J_k}. \quad (3.6)$$

Formula (3.5) may also be put into the form $S^2 \ge S_n^2$, where the S_n^2 satisfy the recursion

$$S_n^2 = \frac{E_{n+1}-J_0}{E_{n+1}-E_0}\frac{J_n - E_0}{J_n - J_0}S_{n-1}^2$$

and where $S_0^2 \equiv (E_1 - J_0)/(E_1 - E_0)$ is the Eckart overlap. Then, clearly,

$$S_0^2 \le S_1^2 \le \cdots \le S_n^2 \le S^2, \quad (3.7)$$

which shows that the Eckart formula can always be improved if additional eigenvalues E_i and Rayleigh–Ritz estimates J_i satisfying (3.4) are known.

Unfortunately, relation (3.4) becomes increasingly difficult to satisfy in atomic and molecular systems because of the Rydberg-like "bunching" of the true eigenvalues. Furthermore, even if an infinite set of bound states could be accommodated in (3.7), the final result will not, in general, be exact,

$$\lim_{n\to\infty} S_n^2 \neq S^2, \quad (3.8)$$

because of the omitted continuum contribution, which may be very significant. In fact, even if all the estimates J_i were exactly *equal* to the true eigenvalues E_i for all $i \ge 1$, formula (3.7) would only yield the limiting result

$$\lim_{n\to\infty} S_n^2 = \frac{E_\infty - \langle H\rangle}{E_\infty - E_0}, \quad (3.9)$$

where E_∞ is the first limit point of the bound spectrum. The right side of (3.9) is therefore the best result that could be achieved, even in principle, from Weinberger's formula, and even this number is often far short of the true value of S^2. In spite of these difficulties, Weinberger's formula often gives a worthwhile improvement over the simple Eckart criterion when the required additional information is available; see Jennings and Wilson[5a] for a numerical example.

For $k = 1$, formula (3.6) strongly improves the result

$$|\langle\phi_1 \mid \psi_1\rangle|^2 \ge \frac{E_2 - J_1}{E_2 - E_1} - \frac{E_2 - E_0}{E_2 - E_1}\frac{J_1 - E_1}{E_1 - J_0}$$

obtained by Shull and Löwdin.[31]

C. Analog of Rayner's Inequality; Wang's Inequality

By introducing the new metric $Q(H - E_1)Q$ of (3.1), it is possible to follow closely the discussion of Sec. 2A and obtain a *lower* bound to S which is the analog of Rayner's inequality (2.3).

We introduce normalized functions $\chi_1, \chi_2, \cdots, \chi_m$ which are now chosen to satisfy

$$\langle\chi_i|\,(H - E_0)^2(H - E_1)\,|\chi_j\rangle = 0, \quad i \neq j.$$

Then with metric $Q(H - E_1)Q$, the Gramian of the vectors ϕ, $(H - E_0)\chi_i$ is found to be

$$G = \begin{vmatrix} \langle H\rangle - E_1 + S^2(E_1 - E_0) & l_1 & l_2 & \cdots \\ l_1^* & L_1 & 0 & \cdots \\ l_2^* & 0 & L_2 & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix} \ge 0, \quad (3.10)$$

where

$$l_i \equiv \langle\phi|\,(H - E_0)(H - E_1)\,|\chi_i\rangle,$$
$$L_i \equiv \langle\chi_i|\,(H - E_0)^2(H - E_1)\,|\chi_i\rangle, \quad (3.11)$$

and where we have used the familiar relations

$$HQ - QH = 0, \quad Q^2 = Q, \quad (H - E_0)Q = H - E_0$$

in writing (3.11). The Gramian determinant (3.10) can again be readily evaluated, and one obtains

$$S^2 \ge \frac{E_1 - \langle H\rangle}{E_1 - E_0} + \frac{1}{E_1 - E_0}\sum_{i=1}^{m}\frac{|\langle\phi|\,(H - E_0)(H - E_1)\,|\chi_i\rangle|^2}{\langle\chi_i|\,(H - E_0)^2(H - E_1)\,|\chi_i\rangle}, \quad (3.12)$$

which is the lower-bound analog of (2.3). Notice that each term in the summation is positive, so that (3.12) always improves on the Eckart result (3.2).

The important special case $m = 1$ of (3.12) was first derived by Wang[10] as

$$S^2 \ge \frac{E_1 - \langle H\rangle}{E_1 - E_0} + \frac{1}{E_1 - E_0}\frac{|\langle\phi|\,(H - E_0)(H - E_1)\,|\chi\rangle|^2}{\langle\chi|\,(H - E_0)^2(H - E_1)\,|\chi\rangle}, \quad (3.13)$$

and will be called *Wang's inequality*. It has the significant practical advantage that χ may be *chosen to insure that the integral* $\langle\chi|\, H^3\,|\chi\rangle$ *does not diverge*, by a proper treatment of the singularities of the potential. Notice that (3.13) can become an equality when χ is the portion of ϕ which is orthogonal to both ψ_0 and ψ_1.

D. Gordon's Lower Bounds and Extensions; Kato's Bound

Gordon[15] has derived a sequence of lower bounds to S which complement the upper bounds of Sec. 2B, and which again require moments $\langle H^n\rangle$ of the Hamiltonian. Wang[10] has pointed out that the same results are obtained from formula (3.13) when χ is taken to be

$$\chi = \sum_n c_n (H - E_0)^n \phi$$

and the c_n are chosen to optimize the bound.

Here, we can again obtain the explicit (and more general) form of these bounds by considering the Gramian of the set of vectors ϕ, $(H - E_0)^i\phi$, $(H - E_0)^j\phi, \cdots$, in conjunction with the metric $Q(H - E_1)Q$ of (3.1). Following the procedure of Sec. 2B, the result is easily found to be

$$S^2 \geq \frac{E_1 - \langle H\rangle}{E_1 - E_0} + \frac{1}{E_1 - E_0}\frac{\tilde{d}}{d}, \tag{3.14}$$

where d and $\tilde{d}$ are the (nonnegative) determinants defined by

$$d \equiv \begin{vmatrix} h_{2i} & h_{i+j} & \cdots \\ h_{i+j} & h_{2j} & \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix},$$

$$-\tilde{d} \equiv \begin{vmatrix} 0 & h_i & h_j & \cdots \\ h_i & h_{2i} & h_{i+j} & \cdots \\ h_j & h_{i+j} & h_{2j} & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix} \tag{3.15}$$

and where

$$h_n \equiv \langle (H - E_0)^n (H - E_1)\rangle. \tag{3.16}$$

Gordon's lower bounds are obtained as the special case $\{i, j, \cdots\} = \{1, 2, \cdots\}$ of (3.14), but it is now also possible to consider *arbitrary* positive integer or noninteger values.

The practical difficulty of these bounds, as in those of Sec. 2B, lies in the evaluation of $\langle H^n\rangle$ for $n > 2$. But when these integrals exist, one can obtain a substantial improvement over the Eckart criterion, as Gordon's numerical examples show.

We remark finally that the bounds of this and the previous section can be easily extended in a natural manner to higher excited states k by adopting the metric

$$Q_k(H - E_{k-1})(H - E_{k+1})Q_k, \quad Q_k \equiv 1 - |\psi_k\rangle\langle\psi_k| \tag{3.17}$$

in place of (3.1) in the derivations. The corresponding formulas are again easily written down, and are not presented here.

Similarly, following Sec. 3A, one may consider the corresponding extension of the Eckart criterion,

$$\langle Q_k(H - E_{k-1})(H - E_{k+1})Q_k\rangle \geq 0,$$

which gives[32]

$$|\langle\phi \mid \psi_k\rangle|^2 \geq \frac{\langle (H - E_{k-1})(E_{k+1} - H)\rangle}{(E_k - E_{k-1})(E_{k+1} - E_k)}. \tag{3.18}$$

Formula (3.18) also results from an earlier formula of Kato[33]

$$|\langle\phi \mid \psi_k\rangle|^2 \geq \frac{\langle (H - \alpha)(\beta - H)\rangle}{(E_k - \alpha)(\beta - E_k)}, \tag{3.19}$$

if α and β are chosen to make (3.19) as strong as possible under Kato's assumption, i.e., $E_{k-1} \leq \alpha \leq E_k \leq \beta \leq E_{k+1}$. Kato[33] also noted that for the ground state $k = 0$, one could take $\alpha = E_{k-1} = -\infty$, whereupon (3.18) and (3.19) reduce back again to the Eckart criterion (3.2). Formula (3.18) was more recently found also by Delves.[21]

E. Lower Bounds of Bazley and Fox

Bazley and Fox[12] have suggested lower bounds to S which start from the relation

$$\langle H\rangle - E_{n+1} \geq \sum_{k=0}^{n} (E_k - E_{n+1})\,|\langle\phi \mid \psi_k\rangle|^2,$$

which follows upon introducing the expansion $\phi = \sum \langle\phi \mid \psi_k\rangle\psi_k$ into $\langle\phi|\, H\,|\phi\rangle = \langle H\rangle$ and replacing all E_k by E_{n+1} for $k > n$. This is solved for $S^2 = |\langle\phi \mid \psi_0\rangle|^2$ to give

$$S^2 \geq \frac{E_{n+1} - \langle H\rangle}{E_{n+1} - E_0} - \sum_{k=1}^{n} \frac{E_{n+1} - E_k}{E_{n+1} - E_0}\,|\langle\phi \mid \psi_k\rangle|^2, \tag{3.20}$$

which now requires some *upper* bounds to the overlap integrals $|\langle\phi \mid \psi_k\rangle|^2$. For this purpose, Bazley and

Fox[12] derived formula (2.7), which was then introduced into (3.20) to give their final result

$$S^2 \geq \frac{E_{n+1} - \langle H\rangle}{E_{n+1} - E_0} - \frac{(\Delta H)^2}{E_{n+1} - E_0} \sum_{k=1}^{n} \frac{E_{n+1} - E_k}{\langle (H - E_k)^2\rangle}. \tag{3.21}$$

But one might alternatively consider the upper bound (2.16) in place of (2.7). Indeed, it is readily verified that (2.16) is *superior* to (2.7) whenever

$$(E_k - \langle H\rangle)\langle (H - E_0)(H - E_k)\rangle \geq 0, \tag{3.22}$$

and, in turn, (3.22) is certainly satisfied at least for the first excited state $k = 1$, since $(H - E_0)(H - E_1)$ has no negative eigenvalues. Thus, (3.21) is certainly strengthened if (2.16) is used in the $k = 1$ term of (3.20):

$$\begin{aligned} S^2 &\geq \frac{E_{n+1} - \langle H\rangle}{E_{n+1} - E_0} - \frac{E_{n+1} - E_1}{E_{n+1} - E_0} \cdot \frac{\langle H\rangle - E_0}{E_1 - E_0} \\ &\quad - \sum_{k=2}^{n} \frac{E_{n+1} - E_k}{E_{n+1} - E_0} |\langle \phi \mid \psi_k\rangle|^2 \\ &= \frac{E_1 - \langle H\rangle}{E_1 - E_0} - \sum_{k=2}^{n} \frac{E_{n+1} - E_k}{E_{n+1} - E_0} |\langle \phi \mid \psi_k\rangle|^2. \end{aligned} \tag{3.23}$$

But formula (3.23) is clearly inferior to the Eckart relation (3.2), no matter what bounds are used for $|\langle \phi \mid \psi_k\rangle|^2$, $k \geq 2$. Therefore, also, the Bazley–Fox bound (3.21) is *always weaker than the Eckart criterion*, even though it requires considerable additional information ($\langle H^2\rangle$ and $E_2, \cdots, E_{n+1}$ in addition to E_0, E_1, and $\langle H\rangle$). This example indicates clearly the desirability of having bounds which are always the best possible for the amount of information employed.

F. Bounds of Swanson and Bohenblust, Wilkinson, Falk, Nitsche, Weinberger, and Davis and Kahan

In this section, we discuss various other bounds to overlap which have appeared principally in the mathematical literature. These are frequently obtained in connection with discussions of the effects of rounding errors, etc., in the numerical solution of eigenvalue problems, but arise also in a more general context.

By introducing the usual expansion $\phi = \sum \langle \phi \mid \psi_k\rangle \psi_k$ into $\langle (H - \lambda)^2\rangle$, replacing E_k by E_1 for $k > 1$, and *neglecting* the term containing E_0,

$$\begin{aligned} \langle (H - \lambda)^2\rangle &= \sum_k |\langle \phi \mid \psi_k\rangle|^2 (E_k - \lambda)^2 \\ &\geq S^2(E_0 - \lambda)^2 + (E_1 - \lambda)^2(1 - S^2) \\ &\geq (E_1 - \lambda)^2(1 - S^2), \end{aligned}$$

one obtains the inequality, valid when λ is at least closer to E_1 than to any higher eigenvalue,

$$S^2 \geq 1 - \langle (H - \lambda)^2\rangle/(E_1 - \lambda)^2. \tag{3.24}$$

This inequality was given by Swanson [34] and also by Wilkinson.[35] The special case $\lambda = \langle H\rangle$ of (3.24) was given by Wilkinson.[36] However, the correct choice of λ was not discussed by these authors.

The proper choice of λ is that value $\tilde{\lambda}$ which makes (3.24) as strong as possible, and, by differentiation, this is found to be

$$\tilde{\lambda} = (\langle H\rangle E_1 - \langle H^2\rangle)/(E_1 - \langle H\rangle).$$

Noting that $\langle (H - \tilde{\lambda})^2\rangle = (\tilde{\lambda} - \langle H\rangle)(\tilde{\lambda} - E_1)$, we obtain the proper form of (3.24),

$$S^2 \geq 1 - \frac{(\Delta H)^2}{\langle (H - E_1)^2\rangle} = \frac{(E_1 - \langle H\rangle)^2}{\langle (H - E_1)^2\rangle}, \tag{3.25}$$

which is the strongest result possible in this approach.

However, (3.25) is recognized to be just the *Eckart criterion* (3.2) with the special choice of *Temple's* lower bound[37,33]

$$E_0 \geq \langle H\rangle - (\Delta H)^2/(E_1 - \langle H\rangle) \tag{3.26}$$

for the eigenvalue E_0.[38] Clearly, therefore, (3.25) is always inferior to the Eckart criterion unless one has to use even a worse lower bound to E_0 than the Temple formula (3.26).

Formula (3.25) was also derived (under a slightly more general formulation) by Falk,[39] who also rederives formula (3.26) for eigenvalues.

Another bound requiring the same information was given by Nitsche,[40] who obtained

$$(2 - 2S)^{\frac{1}{2}} \leq \frac{\Delta H}{E_1 - \langle H\rangle}\left(1 + \frac{\Delta H}{E_1 - \langle H\rangle}\right), \tag{3.27}$$

as well as an alternative form essentially using *Weinstein's*[41] lower bound (also derived here by Nitsche) for E_1. However, we may show that (3.27) is an easy consequence of (3.25). For this purpose, denote

$$x \equiv \Delta H/(E_1 - \langle H\rangle)$$

so that (3.25) is written as $S^2 \geq (1 + x^2)^{-1}$. Then

$$\begin{aligned} (2 - 2S)^{\frac{1}{2}} &\leq [2 - 2(1 + x^2)^{-\frac{1}{2}}]^{\frac{1}{2}} \\ &\leq [2 - 2(1 - x^2/2)]^{\frac{1}{2}} \\ &= x \\ &\leq x(1 + x), \end{aligned}$$

which is Nitsche's inequality (3.27). Clearly, (3.25) gives a much stronger result using the same information.

A somewhat similar, but more complicated, formula was obtained in a report by Weinberger.[42] Adopting the notation

$$z \equiv \langle (H - E_0)^2\rangle^{\frac{1}{2}}/(E_1 - E_0), \tag{3.28}$$

we may write Weinberger's formula[42] in the form (for the ground state)

$$(2-2S)^{\frac{1}{2}} \leq \frac{z}{1-z} + \left(\frac{z^2}{(1-z)^2} + \frac{2}{1-z}[1-(1-z^4)^{\frac{1}{2}}]\right)^{\frac{1}{2}} \equiv \mathcal{E}_W \tag{3.29}$$

and we note the necessary assumption $z < 1$, which may be difficult to satisfy with simple trial functions ϕ. Developing the right side of (3.29) as a power series in z, we obtain

$$\mathcal{E}_W = 2z + 2z^2 + \tfrac{5}{2}z^3 + 2z^4 + \tfrac{15}{8}z^5 + \cdots. \tag{3.30}$$

Let us now compare this with the Eckart criterion (3.2), which requires less information. We first note the inequality

$$z^2 \geq (\langle H\rangle - E_0)/(E_1 - E_0), \tag{3.31}$$

which is an easy consequence of Temple's formula (3.26). The Eckart formula (3.2) requires

$$(2-2S)^{\frac{1}{2}} \leq \left[2 - 2\left(1 - \frac{\langle H\rangle - E_0}{E_1 - E_0}\right)^{\frac{1}{2}}\right]^{\frac{1}{2}} \equiv \mathcal{E}_E, \tag{3.32}$$

and in turn, noting (3.31),

$$\mathcal{E}_E \leq [2 - 2(1-z^2)^{\frac{1}{2}}]^{\frac{1}{2}} = z + \tfrac{1}{8}z^3 + \tfrac{7}{128}z^5 + \cdots. \tag{3.33}$$

Comparison of (3.30) and (3.33) shows that $\mathcal{E}_W \gtrsim 2\mathcal{E}_E$ even when z is small enough for the Weinberger formula (3.29) to be effective. Thus, the Eckart criterion again proves to be superior.

More recently, Davis and Kahan[43] have derived four theorems giving error bounds for entire subspaces of eigenvectors. However, if we restrict attention to the usual case of the single, nondegenerate eigenvector ψ_0, their results can be related to other approaches. Their "sin θ theorem" is equivalent to (3.24), where λ is an arbitrary estimate of the eigenvalue E_0, and their "tan θ theorem" is identical to (3.25). They also present a "sin 2θ theorem" which, in its strongest form, is

$$S^2 \geq \frac{1}{2}\left[1 + \left(1 - \frac{4(\Delta H)^2}{(E_1 - E_0)^2}\right)^{\frac{1}{2}}\right]. \tag{3.34}$$

But (3.34) is again *weaker* than the Eckart criterion (3.2), since we can write Temple's formula (3.26) in the form

$$(\Delta H)^2 \geq (\langle H\rangle - E_0)(E_1 - \langle H\rangle)$$

and substitute into (3.34) to obtain

$$\frac{1}{2}\left[1 + \left(1 - \frac{4(\Delta H)^2}{(E_1 - E_0)^2}\right)^{\frac{1}{2}}\right] \leq \frac{1}{2}\left[1 + \left(1 - \frac{4(\langle H\rangle - E_0)(E_1 - \langle H\rangle)}{(E_1 - E_0)^2}\right)^{\frac{1}{2}}\right] = \frac{E_1 - \langle H\rangle}{E_1 - E_0}.$$

Finally, their "tan 2θ theorem" can be written as

$$\frac{1}{2S^2 - 1}\left(\frac{1-S^2}{S^2}\right)^{\frac{1}{2}} \leq \frac{\Delta H}{|\gamma - \langle H\rangle|}, \tag{3.35}$$

where γ is the lowest eigenvalue of the operator

$$(1 - |\phi\rangle\langle\phi|)H(1 - |\phi\rangle\langle\phi|).$$

In practice, however, the eigenvalue γ is usually unknown, so that (3.35) is not presently a practical result.

G. Lower Bounds of Rayner and Delves

With the help of the quantities

$$a_n \equiv \langle (H-\alpha)^{-n}\rangle, \quad n = 0, 1, 2, \cdots, \tag{3.36}$$

with $\alpha < E_0$, one can also get lower bounds to S to complement the upper bounds of Sec. 2C. Rayner[20] gives the inequality

$$S^2 \geq \frac{1}{2}\frac{a_{n-1}^n}{a_n^{n-1}}\left\{1 + \left[1 - 4n\left(\frac{E_0-\alpha}{E_1-\alpha}\right)^{n-1}\frac{a_n^{n-1}}{a_{n-1}^n}\right]^{\frac{1}{2}}\right\}, \tag{3.37}$$

which is also discussed by Delves.[21] Combining (2.13) with a slightly weakened form of (3.37), we infer

$$\frac{a_{n-1}^n}{a_n^{n-1}} \geq S^2 \geq \frac{a_{n-1}^n}{a_n^{n-1}} - n\left(\frac{E_0-\alpha}{E_1-\alpha}\right)^{n-1}, \tag{3.38}$$

which shows the convergence of the upper and lower bounds as $\alpha \to E_0$. Rayner and Delves also derive

$$S^2 \geq [(E_0-\alpha)^n/(E_1-E_0)][(E_1-\alpha)a_n - a_{n-1}] \tag{3.39}$$

and various extensions to excited states are considered, as well as the use of upper and lower bounds to E_0 and E_1.

Table II presents some numerical applications of formulas (3.37) and (3.39) to the particle in a box considered in Sec. 2C, with $\phi = 1$ and $\alpha = 0$. Formula (3.37) is not so strong, but the convergence of (3.39) is again quite rapid in this simple example, the results comparing fairly well with the upper bounds of Table I.

TABLE II. Lower bounds to overlap for a particle in a 1-dimensional box, as calculated from the formulas of the text.

n	Eq. (3.37)	Eq. (3.39)
1	—	0.763289
2	—	0.808168
3	0.520890	0.810346
4	0.742789	0.810546
	exact = 0.810569	

However, the integrals a_n of (3.36) are seldom available, and formulas based on these integrals are not yet of practical utility in atomic and molecular problems.

Braun and Rebane[44] have recently derived a related expression, in which the variational functional J for arbitrary f,[45]

$$J \equiv \frac{\langle f|\,\phi\rangle^2}{\langle f|\,H-\alpha\,|f\rangle} \leq \langle\phi|\,(H-\alpha)^{-1}\,|\phi\rangle = a_1,$$

has been used in place of the integral a_1 in the $n=1$ case of (3.39),

$$S^2 \geq \frac{E_0-\alpha}{E_1-E_0}\,[(E_1-\alpha)J-1]$$
$$= \frac{E_0-\alpha}{E_1-E_0}\left[\langle f|\,\phi\rangle^2\left(\frac{E_1-\alpha}{E_f-\alpha}\right)-1\right], \quad (3.40)$$

with $E_f \equiv \langle f|\,H\,|f\rangle$, and with f taken normalized without loss of generality. Braun and Rebane note that (3.40) reverts back to the Eckart criterion (3.2), for $f=\phi$ and $\alpha \to -\infty$. Goscinski[46] has subsequently obtained a generalized form of (3.40), in which f is taken to be an optimal linear combination of linearly independent basis functions.

When the parameter α is chosen to *optimize* the bound, the Braun–Rebane result (3.40) is found[47] to be equivalent to the lower member of (1.13),

$$S \geq S_f\langle\phi\,|f\rangle - [(1-S_f^2)(1-\langle\phi\,|f\rangle^2)]^{\frac{1}{2}},$$

in the special case when $S_f \equiv \langle f|\,\psi_0\rangle$ (or a lower bound) has been evaluated from the Eckart criterion.

H. Numerical Applications and Conclusion

As in Sec. 2E, we wish to illustrate the lower-bound formulas numerically with an application to the simple representative example of the hydrogen atom. Again, ϕ is chosen to be the function (2.19) with

$$S^2 = 0.864.$$

In this case, the simple Eckart criterion (3.2) gives the rather poor lower bound (note that $E_0=-0.5$ and $E_1=-0.125$):

$$S^2 \geq 0.692, \quad (3.41)$$

which shows that the contribution to ϕ of $\psi_2, \psi_3, \cdots$ is still very substantial. On the other hand, the weakened relation (3.25), which requires $\langle H^2\rangle$ but not E_0, would lead to the much weaker result

$$S^2 \geq 0.297.$$

Similar results could also be cited for the other formulas discussed in Secs. 3E and 3F.

We cannot immediately apply Weinberger's formula (3.5), since we would need to have ϕ as one root of a secular determinant whose zeros interleave the lowest true eigenvalues according to (3.4). However, Weinberger's formula (3.5) could never surpass the hypothetical limit (3.9),

$$\lim_{n\to\infty} S_n^2 = 0.769,$$

which is still rather far below the true answer. This example illustrates the severe practical disadvantage of Weinberger's formula (3.5) in many applications of interest.

In applying Gordon's lower bound formulas (3.14), we can only compute the first member of the sequence, which requires $\langle H^2\rangle$ and $\langle H^3\rangle$, and which gives the result

$$S^2 \geq 0.775. \quad (3.42)$$

This certainly improves the Eckart result (3.41), but still lies farther from the true value than do either of the corresponding upper bounds (2.21), (2.22).

Finally, for Wang's inequality (3.13), we again choose for χ the simple function (2.23),

$$\chi \sim [1+(\alpha-1)r]e^{-\alpha r},$$

which will give finite matrix elements over H^3. Choosing the parameter α to optimize the lower bound (3.13), we obtain

$$S^2 \geq 0.857 \quad (\alpha = 1.03). \quad (3.43)$$

Still better results could be obtained by using the more general formula (3.12) in conjunction, for example, with the functions

$$\tilde{\chi}_1 \sim \chi_1,$$
$$\tilde{\chi}_2 \sim \chi_1 - \chi_2\,\langle\chi_1|\,(H-E_0)^2(H-E_1)\,|\chi_1\rangle/$$
$$\langle\chi_1|\,(H-E_0)^2(H-E_1)\,|\chi_2\rangle,$$

where χ_1 and χ_2 are arbitrary so long as the required integrals exist. Thus, it appears that Wang's inequality (3.13) and the more general formula (3.12) are the most flexible and practical means available of calculating accurate lower bounds to overlap when E_0 and E_1 are known.

If ϕ is so complex that the integrals $\langle\phi|\, H^2\,|\chi\rangle$ are still difficult to evaluate, one may adopt an alternative procedure: (a) choose some simpler function $\tilde{\phi}$ 'close' to ϕ so that $S_1 \equiv |\langle\tilde{\phi}\,|\,\phi\rangle|$ is near unity and $\langle\tilde{\phi}|\, H^2\,|\chi\rangle$ is easier to compute; (b) use formulas (3.12) or (3.13) to calculate a lower bound to $S_2 \equiv |\langle\tilde{\phi}\,|\,\psi_0\rangle|$; and (c) use formula (1.13) to obtain the final lower bound for $S = |\langle\phi\,|\,\psi_0\rangle|$.

Returning to our numerical example and combining (2.24) with (3.43), we conclude finally

$$S^2 = 0.861 \pm 0.004,$$

which would be a useful accuracy for many purposes.

ACKNOWLEDGMENTS

I wish to thank Professor T. Kato and, particularly, Professor W. M. Kahan for stimulating conversations, and Professor E. Bright Wilson, Jr., for a careful reading of the manuscript. The support of a Miller Research Fellowship is gratefully acknowledged.

* Miller Fellow.

† Present address: Department of Chemistry, Stanford University, Stanford, California 94305.

[1] G. G. Hall [J. Chem. Phys. **38**, 1104 (1963)] has also discussed the idea of a "most probable" value of the overlap, and M. H. Alexander [*ibid.* **51**, 5650 (1969)] has recently given semirigorous bounds to overlap based upon extrapolation procedures.

[2] Although results in the mathematical literature are frequently obtained under more general assumptions (larger invariant subspaces, more general classes of operators, etc.), attention here is concentrated on the special cases of principal physical interest. These refer to a single normalized nondegenerate eigenfunction ψ_k, usually the ground state $k = 0$, belonging to the bound (point) spectrum E_k of an Hermitian operator H.

[3] F. Weinhold, J. Phys. **A1**, 305 (1968).

[4] F. Weinhold, J. Phys. **A1**, 535 (1968).

[5] See, e.g., (a) P. Jennings and E. B. Wilson, Jr., J. Chem. Phys. **47**, 2130 (1967); (b) F. Weinhold, Phys. Rev. **183**, 142 (1969).

[6] For a more complete mathematical discussion, see, e.g., G. E. Shilov, *An Introduction to the Theory of Linear Spaces* (Prentice-Hall, Englewood Cliffs, N.J., 1961), Chap. 8.

[7] F. Weinhold, J. Chem. Phys. **46**, 2448 (1967).

[8] Hereafter, we generally reserve the symbol S for the ground state $S \equiv |\langle\phi\,|\,\psi_0\rangle|$.

[9] M. E. Rayner, Quart. J. Math. (Oxford) **13**, 61 (1962).

[10] P. S. C. Wang, Intern. J. Quantum Chem. **3**, 57 (1969). Relation (2.5) has also been obtained by M. G. Hegyi, M. Mezei, and T. Szondy, Theoret. Chim. Acta **15**, 273 (1969).

[11] F. Combet-Farnoux and G. Allard, Compt. Rend. **252**, 999 (1961).

[12] N. W. Bazley and D. W. Fox, J. Res. Natl. Bur. Std. (U.S.) **66B**, 1 (1962).

[13] H. Conroy, J. Chem. Phys. **47**, 930 (1967).

[14] J. B. Delos and S. M. Blinder, J. Chem. Phys. **47**, 2784 (1967).

[15] R. G. Gordon, J. Chem. Phys. **48**, 4984 (1968).

[16] If H is positive semidefinite, one may consider noninteger values as well.

[17] J. Goodisman and D. Secrest, J. Chem. Phys. **45**, 1515 (1966).

[18] See, e.g., I. T. Keaveny and R. E. Christofferson, J. Chem. Phys. **50**, 80 (1969).

[19] This requirement may often be somewhat relaxed upon closer study. For example, formula (2.6) remains true if E_0 is replaced by any *upper* bound to its true value.

[20] M. E. Rayner, Quart. J. Math. (Oxford) **13**, 137 (1962).

[21] L. M. Delves, J. Math. Phys. **5**, 1055 (1964).

[22] C.-Y. Hu, Phys. Rev. **167**, 112 (1968).

[23] See, e.g., L. Pauling and E. B. Wilson, Jr., *Introduction to Quantum Mechanics* (McGraw-Hill, New York, 1935).

[24] The Rayner–Delves bound (2.13) is not considered further, since the quantities it requires are generally unavailable in more-difficult problems.

[25] C. Eckart, Phys. Rev. **36**, 878 (1930).

[26] Of course, E_1 and other E_k may be taken to refer only to excited states having the same symmetry as the ground state.

[27] See, e.g., Ref. 23, p. 182; P.-O. Löwdin, Advan. Chem. Phys. **2**, 207 (1959), p. 264.

[28] See, e.g., R. A. Newing, Phil. Mag. **24**, 114 (1937); T. Kato, J. Phys. Soc. Japan **4**, 334 (1949); T. Kinoshita, Phys. Rev. **105**, 1490 (1957); H. F. Weinberger, J. Res. Natl. Bur. Std. (U.S.) **64B**, 217 (1960); S. Aranoff and J. K. Percus, Phys. Rev. **162**, 878 (1967); I. R. Afnan and Y. C. Tang, *ibid.* **175**, 1337 (1968).

[29] A similar relaxation is possible in many of the other formulas to be discussed. Usually a simple differentiation suffices to determine whether the formula remains valid when an upper or lower bound is employed.

[30] H. F. Weinberger, J. Res. Natl. Bur. Std. (U.S.) **64B**, 217 (1960); see also N. W. Bazley and D. W. Fox, Rev. Mod. Phys. **35**, 712 (1963).

[31] H. Shull and P.-O. Löwdin, Phys. Rev. **110**, 1466 (1958).

[32] Noting the restriction that the metric (3.17) must remain positive semidefinite, we may replace E_{k-1} by an *upper* bound and E_{k+1} by a *lower* bound in this formula.

[33] T. Kato, J. Phys. Soc. Japan **4**, 334 (1949).

[34] C. A. Swanson, Bull. Am. Math. Soc. **67**, 607 (1961). Swanson attributes the result to H. F. Bohenblust.

[35] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford U.P., London, 1965), p. 172.

[36] J. H. Wilkinson, Computer J. **4**, 230 (1961).

[37] G. Temple, Proc. Roy. Soc. (London) **A119**, 276 (1928).

[38] The possibility of introducing Temple's bound for E_0 into the Eckart criterion had previously been noted by Löwdin [Advan. Chem. Phys. **2**, 207 (1959)], but his resulting expression was based upon the inexact form (3.3) of Eckart's relation.

[39] S. Falk, Z. Angew. Math. Mech. **45**, 47 (1965).

[40] J. A. Nitsche, Z. Angew. Math. Mech. **39**, 322 (1959).

[41] D. H. Weinstein, Proc. Natl. Acad. Sci. (U.S.) **20**, 529 (1934).

[42] H. F. Weinberger, Technical Note BN-183, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland (1959).

[43] C. Davis and W. M. Kahan, Tech. Report No. 6, Computer Science Department, University of Toronto (1968); Bull. Amer. Math. Soc. **75**, 863 (1969); SIAM J. Num. Anal., to appear.

[44] P. A. Braun and T. K. Rebane, Intern. J. Quantum Chem. **3**, 1058 (1969).

[45] This form of J is obtained from the functional

$$J = 2\langle u\,|\,\phi\rangle - \langle u|\, H - \alpha\,|u\rangle$$

of Ref. 22 by taking $u = \mu f$, where f is normalized, and optimizing μ. Alternatively, it follows directly from the theory of inner projections on the positive definite operator $H - \alpha$. See, e.g., P. Lindner and P.-O. Löwdin, Intern. J. Quantum Chem. **S2**, 161 (1968).

[46] O. Goscinski, Intern. J. Quantum Chem. **3**, 1064 (1969).

[47] F. Weinhold and P. S. C. Wang, Intern. J. Quantum Chem. (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Foundations for Quantum Mechanics

ROBIN GILES
Department of Mathematics, Queen's University, Kingston, Ontario

(Received 5 September 1969)

A discussion is given of the structure of a physical theory and an "ideal form" for such a theory is proposed. The essential feature is that all concepts should be defined in operational terms. Quantum (and classical) mechanics is then formulated in this way (the formulation being, however, restricted to the kinematical theory). This requires the introduction of the concept of a *mixed test*, related to a pure test (or "question") just as a mixed state is related to a pure state. In the new formulation, the primitive concepts are not states and observables but certain operationally accessible mixed states and tests called *physical*. The notion of a *C*-system* is introduced; each such system is characterized by a certain C^*-algebra. The structure of a general C^*-system is then studied, all concepts being defined in terms of physical states and tests. It is shown first how pure states and tests can be so defined. The quantum analog of the phase space of classical mechanics is then constructed and on it is built a mathematical structure, called a *q-topology*, which is a quantum analog of the topology of classical phase space. Mathematically, a q-topology is related to a noncommutative C^*-algebra as an ordinary topology is related to a commutative C^*-algebra. Some properties of the q-topology of a C^*-system are given. An appendix contains some physically motivated examples illustrating the theory.

1. INTRODUCTION

Formulations of the foundations of quantum mechanics, although they may be adequately rigorous from a mathematical point of view, are frequently seriously lacking in the attention that is paid to the rules of interpretation by which the mathematical assertions of the theory are translated into experimentally meaningful statements concerning the physical world. In this paper, which represents a summary of work that will be published in detail elsewhere,[1] we attempt to remedy this defect. This results not only in a more satisfactory theory, but the interplay between the physical and mathematical aspects reveals features which are not present in the conventional formulations.

Exactly what should be required of a physical theory in respect of the rules of interpretation is itself not clear. This is discussed in Sec. 2. We arrive at the conclusion that important advantages accrue to a theory in which all concepts are defined in terms of a few *primitive concepts*, these being given precise interpretations in practical (or operational) terms. The requirement concerning the interpretation of the primitive concepts imposes a severe limitation on the choice of these concepts and thus on the acceptable formulations of the theory.

In Sec. 3, we apply these considerations to quantum mechanics. It is common, in formulations of this subject, to assume the physical interpretations of the concepts of classical mechanics to be already clear. While this attitude may be valuable from a didactic point of view, it is not logically satisfactory since, fundamentally, quantum mechanics is not a sophisticated variation of classical mechanics but should be regarded rather as the basic theory from which classical mechanics arises as a limiting case. For this reason, in choosing primitive concepts for quantum mechanics, we reject coordinates and momenta and the like and examine the concepts of *state* and *test* (i.e., *question* in the terminology of Mackey[2]). We discuss at some length the interpretation of these terms, assigning them operational meanings not dependent in any way on classical mechanics. (Indeed, our formulation of quantum mechanics includes classical mechanics as a special case.)

As a result of this discussion, we conclude that by no means can all states and tests (as the terms are normally understood) be given an effective operational interpretation; those that can we designate as *physical*. All physical states are *mixed states*; this is generally recognized. On analyzing physical tests, we find that they are not represented at all in conventional quantum mechanics. It is necessary to introduce a new concept, that of a *mixed test*. Mixed tests have a natural place in the theory, being related to ordinary (or *pure*) tests just as mixed states are related to pure states. Moreover, with the aid of mixed tests it is possible to form weighted means of tests in the same way as for states. Physical tests are normally mixed. The physical tests of a given quantum mechanical system form a distinguished set which plays a major role in our analysis—in a sense characterizing the particular system.

Since we are able to give an operational interpretation only of *physical* states and tests, it follows from the principles developed in Sec. 2 that *these* should be taken as the primitive concepts. Thus, we are faced with the program of reformulating both

classical and quantum mechanics in this way. In Sec. 4, we indicate how this may be done in the case of classical mechanics, considering in particular the way in which the concept of phase space can be constructed. On examining conventional quantum mechanics from the same point of view, we observe that, although it contains an analog of phase space, this analog carries nothing corresponding to a topological structure. However, comparison with the case of classical mechanics indicates a procedure for defining such a superstructure.

Sections 5–8 are concerned with the implementation of this program. In Sec. 5, the concept of a *C*-system* is introduced. This is a (classical or quantum) mechanical system whose physical tests are isomorphic to a dense subset of the region $\{a: a \in A, \mathbf{0} \leq a \leq \mathbf{1}\}$ in a C^*-algebra A. Some justification is given for the claim that all systems arising in practice are C^*-systems; further support for this proposal is given in the Appendix. The concept of a C*-system, introduced originally by Segal,[3] has been discussed frequently in the literature,[4–6] usually, however, in the context of quantum field theory and often with little physical justification.

In accordance with the policy of defining everything in observable terms, Secs. 6 and 7 are devoted to the construction, for an arbitrary C^*-system, of pure states and tests. In the cases of classical mechanics over a compact phase space and of quantum mechanics over a finite-dimensional Hilbert space, the construction yields exactly the usual pure states and tests—in the classical case, for instance, we get just the points and subsets of phase space. In other cases, there appear in addition certain "nonconventional" states and tests. In the classical case this is related to the occurrence of a certain compactification of phase space. (It turns out later that the situation in the quantum-mechanical case is analogous.)

In the case of a quantum-mechanical system, the pure tests may be identified with the projections of a certain W^*-algebra A_w, namely, the weak operator closure of the reduced atomic representation[7] of the C^*-algebra A. They form a lattice which is the quantum-mechanical analog of the lattice of all subsets of (the relevant compactification of) phase space and which we call a *q-space*.

In Sec. 8, we introduce an analog of a topological structure in this q-space by distinguishing certain elements as *open*, these distinguished elements being determined by the algebra A. The *q-topology* which is defined in this way turns out to have many of the properties of an ordinary topology. In particular, the q-space is *compact Hausdorff* and even *normal*, the definitions of these terms being natural extensions of the usual topological definitions. Moreover, the q-space equipped with its q-topology *determines* the C^*-algebra A: Indeed, the real part of A consists exactly of all "q-continuous" operators in A_w, the term *q-continuous* being here defined in terms of the spectral resolution of the operator in just the same way as continuity of a function is defined in terms of its inverse map.

In the Appendix we consider several interpretations of a particular quantum-mechanical system, thus illustrating the sort of physical arguments that can be used to identify the physical states and tests. These examples also serve to support the contention that it is C^*-systems which arise in practice and to clarify the nature and significance of the pure states and tests. Finally, they provide some concrete examples of q-spaces and q-topologies.

2. THE STRUCTURE OF A PHYSICAL THEORY

Although a physical theory normally has an underlying mathematical structure, it can be distinguished from a mathematical theory in that it makes assertions about the physical world which can be tested by experiment. Of course, a physical theory cannot be *proved* experimentally; however, it is usually claimed, as a necessary condition for an acceptable theory, that it should be amenable to experimental test: that is, that it should lay itself open to the possibility of *refutation* by experiment. This does not, of course, imply anything about the truth of the theory; it merely indicates a sense in which it is meaningful.

The requirement of being acceptable in this sense has an important consequence. To explain it, let us consider an example. Suppose a theory has been put forward which gives a particular value for the fine structure constant. An experimental determination, made with great accuracy, yields a different value and the experimenter accordingly claims to have refuted the theory. The theorist replies that this is not so—the experiment, as performed, is not a test of the theory since certain effects have not been taken into account. He describes these effects and shows how they should be allowed for, concluding that there is, in fact, no contradiction with his theory. The theory has thus been saved. Moreover, the theorist is still in a position to repel further attacks: it is only necessary for him to have enough ingenuity to think, on each occasion, of other effects that may be invoked to explain discrepancies between theory and experiment.

It is clear that this situation is unsatisfactory. Indeed, from a strict point of view, the theory appears to

be irrefutable: We have no sure way of refuting it. It must therefore be regarded as unacceptable. Notice the origin of this unacceptability: It is the fact that, in its original formulation, the circumstances under which the theory is applicable were not precisely stated. For it is just this feature which allows the theorist to save the theory from apparent refutation.

We conclude from this argument that, to be acceptable in the sense explained above, a physical theory must be accompanied by instructions explaining in detail the experimental circumstances under which it may be applied. Naturally, these *rules of interpretation* must be to some extent vague, if only because they must be expressed in common language, but they should be made as precise as possible, since the acceptability of the theory varies directly with the precision of the instructions.

We arrive at the same conclusion concerning the existence and precision of rules of interpretation if we consider practical applications of the theory. The considerable investment that may be involved in an engineering application, for instance, will be in jeopardy unless the conditions of applicability of the theory are clearly understood.

To minimize the possibility of a misinterpretation of the theory, the rules of interpretation should use only the simplest of terms, all concepts whose experimental significance can be further clarified being explained in terms of simpler concepts. Here, "simpler" should be understood to mean "more closely related to experience"—or, as we say, more *direct*—for we are concerned not with what the experimenter may think the terms "mean" but with what he should actually *do* in using the theory. An account of the rules of interpretation that fulfills these requirements is called *adequate*.

It is clear that an adequate account of the rules of interpretation of a physical theory is a lengthy affair. For instance, in the case of a theory of the hydrogen atom, it may well involve a complete description of how hydrogen is to be prepared, or at least recognized. In fact, a general understanding of the rules of interpretation of physical theories is exactly what is conveyed in the training of an experimental physicist. (Hence, the common attitude of regarding the details of the application of a physical theory as being a matter for the experimentalist.) An adequate account of the rules of interpretation of a given physical theory can thus be regarded as a specialist training manual in the application of that theory. Since potential users of this manual may have widely varying backgrounds, the manual should assume a minimum background knowledge—ideally none at all. Therefore, it should itself contain any general training in elementary physical concepts that is required for the elucidation of the more sophisticated concepts appearing in the particular theory.

Of course, it is quite unrealistic to demand that every formulation of a physical theory should be accompanied by an adequate account of the rules of interpretation. In fact, frequently, a formulation of a physical theory is accompanied by no indication of the rules of interpretation at all. There is no harm in this, provided it is realized that the theory, as thus presented, is incomplete; it becomes complete only when a specific set of rules of interpretation is adjoined. For instance, a theory of the fine structure constant would have to be completed by a careful description of at least one way in which this constant may be measured. (If several ways are described, then the completed theory is asserting that these distinct procedures yield the same result. If not, then it makes no such assertion.) The various rules of interpretation that can be adjoined in this way give rise to different *completions* of the original theory. One cannot test this theory itself; only the completions can be tested. Thus, only the completions can be acceptable, and only to them can the notions *true* or *false* be meaningfully applied. Consequently, although it is very inconvenient in practice to give explicit rules of interpretation for every physical theory, for the *precise* formulation of any theory it is essential to do just this.

The task of adjoining an adequate set of rules of interpretation to an incomplete physical theory is facilitated if the theory can first be formulated so that two conditions are satisfied. The first of these—which is also desirable for purely mathematical reasons—is that all the concepts in the theory should be defined in terms of a few *primitive* concepts. If this condition is satisfied, it is then *sufficient* to give interpretations to these primitive concepts alone. For then an interpretation is automatically determined for every other concept. (For instance, in the case of Euclidean geometry, an interpretation of *circle* is uniquely determined once interpretations for *point* and *distance* have been given.) Indeed, it is desirable that no other concepts should be given formal interpretations since this would, in any case, be redundant and could only produce confusion. The second condition is that the primitive concepts in the theory should be so chosen that their interpretations (which we shall call the *primitive physical concepts*) are as "direct" as possible, for only then is it possible to give an adequate account of these interpretations which is susceptible to the least possible extent to the vagaries of ordinary language.

We summarize these desirable features of a physical theory:

(1) *The theory should consist of a mathematical structure together with a set of rules of interpretation.*

(2) *The mathematical structure should be expressed in terms of axioms and primitive concepts.*

(3) *The rules of interpretation should give interpretations for all the primitive concepts and only these concepts.*

(4) *The physical concepts referred to in the rules of interpretation should be as direct as possible.*

We shall refer to a theory formulated in accordance with these requirements as an *ideal physical theory*.[8] Our main concern in this paper is to give an ideal formulation of quantum mechanics.

In order to apply an ideal physical theory, an experimenter need not be in full possession of his faculties. Since the theory refers only to certain primitive physical concepts, it is sufficient that he should be aware only of *these* aspects of his environment. We may even imagine an observer whose physical experience consists of these concepts alone; he is the simplest being to whom the theory makes sense. We call such an observer a *primitive observer* for the theory. Not only is a primitive observer able to understand and use the theory, but in principle he could develop it for himself, by extracting and formalizing systematic properties of his experience. We can often reach a deeper understanding of a physical theory by imagining its being developed in this way.

3. THE FUNDAMENTAL CONCEPTS OF QUANTUM MECHANICS

In the formulation of quantum mechanics, we can recognize two distinct stages. First, there is the so-called kinematical theory, usually expressed in terms of the concepts of state and observable. Secondly, there is the dynamical theory, in which the time development of quantum-mechanical systems is discussed.

A complete formulation of the dynamical theory involves giving an adequate account of the interpretation of the concept of time: i.e., we are faced with the problem of giving an explicit description of the experimental procedures involved in measuring a time interval. This is clearly a nontrivial matter. (Even more severe problems arise if we attempt to include other concepts such as space, particle, momentum, etc.) Of course, it is possible to ignore this difficulty—to pretend that the notion of time and the procedure to be used in measuring it are perfectly clear—and this is commonly done. But this entails a serious limitation on the acceptability of the theory: The experimental meaning of any assertion of the theory that involves the concept of time will be obscure. Because of these difficulties we restrict ourselves here to the kinematical theory, into which the notion of time does not enter.[2,9–11]

An immediate simplification is possible.[2,9–11] For the formulation of quantum mechanics, it is sufficient to consider only observables that can take just two values, which we may denote 0 and 1. Mackey calls such observables "questions"; we call them *tests*.

The restriction to tests is not so severe as might at first be supposed. For, in practice, an experiment set up with the object of measuring a continuous variable, say the coordinate X of a particle, measures not X but a function $f(X)$ which takes only finitely many values. For instance, if the experiment is carried out by allowing the particle to trigger one of a row of 100 counters, $f(X)$ may be taken to be the number of the counter that responds. Similarly, if the measurement is carried out by microscopic examination of tracks in a photographic plate, only a finite range of values of the coordinate X are covered and the experimenter records his results to only a finite number of decimal places. Thus most, if not all, experiments in physics, in fact, measure observables which have a spectrum consisting of a finite number of points, and our limitation is only to restrict this number to two.

We denote tests by $a, b, c, \cdots$ and states by $x, y, z, \cdots$. The experiment consisting of applying the test a to a system in the state x is denoted $[x, a]$ and is called an *elementary experiment*, and the value (0 or 1) which is obtained when this experiment is carried out is referred to as the *outcome*.

A typical assertion of the kinematical theory is now of the following form:

> The probability is p that, when the test a is applied to a system in the state x, the outcome will be 1. (1)

If we wish to give an adequate formulation of the theory, we must give detailed interpretations of the terms involved in each such assertion so that it acquires a precise experimental meaning—i.e., in such a way that it is refutable by experiment. A difficulty at once arises in connection with the interpretation of the term "probability." Strictly speaking, no finite number of executions of the experiment $[x, a]$ will serve to refute the assertion (1); this assertion refers rather to the relative frequency of positive outcomes in a long series of repetitions of the experiment $[x, a]$.[12] Clearly, this difficulty is common to every theory which asserts probabilities and is an essential limitation on the requirement that such a theory be refutable. Evidently, then, our original requirement (Sec. 2) that a physical

theory should be refutable is too strong. We therefore now relax this requirement on a physical theory and demand only that it should ascribe probabilities to the outcome of experiments which (a) are described so clearly and completely that there can be no doubt, on each occasion, as to whether or not the experiment has been carried out and what the outcome is and which (b) are,in fact, capable of being carried out.

Let us now apply these considerations to the assertion (1). Any procedure for carrying out the elementary experiment $[x, a]$ consists of two parts. In the first part, a system is prepared in the state x; in the second part, the test a is applied to this system. Applying the requirement (a), we conclude, in particular, that it should always be clear whether or not a certain procedure constitutes a preparation of the system in the state x.[14] Thus, we cannot say we know the state unless we know the corresponding class of procedures. Conversely, this class of procedures is all that we need know about x for the interpretation of every assertion of the form (1) involving x. Consequently, we now *define* a state of a system to be a (nonempty) collection of *methods of preparation* of the system, where by a method of preparation we mean a *document* giving detailed instructions for the preparation. In the same way, we define a test of a system to be a collection of *methods of analysis*, where by a method of analysis we mean a document giving detailed instructions for the execution of the test; this is, of course, a procedure which can be applied to the system no matter what its state may be and results in an outcome, either 0 or 1.

The execution of the experiment $[x, a]$ now involves arbitrarily selecting a document from the collection x and a document from the collection a. The pair of documents constitutes a complete set of instructions for the experiment $[x, a]$. These documents must be sufficiently explicit that it is perfectly clear in each particular case whether or not the instructions have been followed. Of course, we do not demand that the document *uniquely specifies* the actions that the experimenter is to perform: there will always be experimental details that are not prescribed. Thus, a document itself really corresponds to a *class* of experimental procedures. We say that a document A is a *refinement* of another document B if it is agreed that every experimental procedure which satisfies A also satisfies B. This relation of refinement clearly determines a partial ordering of the documents. It is clear that, without loss of generality, we may assume that every refinement of a document that belongs to a state x also belongs to that state (and the same for tests).

Two further points of clarification should be made. First, it must be granted that to follow the instructions on a pair of documents for an elementary experiment does not necessarily yield an "outcome." If, for instance, the laboratory is struck by a thunderbolt during the work, the experiment will be vitiated. However, the documents themselves should contain instructions for recognizing whether any such disturbance has occurred and give a signal "experiment void" if it has.

The second point concerns the *relative* nature of states and tests. Of course, different experimenters following the same instructions at different places or at different times do not produce states that are *objectively* alike. The notion of state is thus a relative one: "the system is in the state x" asserts really a relation of the system to the observer who has prepared it. In general, it is not in this state for any other observer, or indeed for the same observer operating at a later time. Accordingly, if a test is subsequently to be made on the system, it must be carried out by the same observer, and the time of commencement of the operations constituting the test must be related to the time during which the method of preparation was effected.[15]

These observations show that there must be some reference in the documents to the times at which the instructions should be carried out. The times referred to cannot be absolute times, since it is essential that every experiment can be repeated; they must be relative to the observer. To be more specific, we may envisage the following situation. Each observer is equipped with a reference frame and a clock. Between experiments he may move the reference frame and reset the clock as he wishes, but neither must be disturbed during the course of an experiment. To carry out the elementary experiment $[x, a]$, he first chooses documents from the collections x and a to form a set of instructions for the experiment and sets his clock to a time earlier than any referred to on these documents. Then he carries out the instructions.

We can avoid the possibility that the instructions given in the two documents might conflict with each other by supposing the instructions on all documents to be such that every method of preparation is completed before time zero, while every method of analysis starts after this time.

We now return to the interpretation of the term "probability" in the assertion (1). Let us take the view that all the applicable statements of a physical theory can be expressed in terms of assertions of this form. Then, from a pragmatic viewpoint, we may say:

The theory held by any physicist is the set of all assertions of the form (1) *to which he subscribes.*

What commitment is assumed by the theorist who makes the assertion (1)? If we employ the frequency

interpretation of probability, it seems he has placed himself under no obligation at all. For, as we have seen, his assertion can never be refuted. We therefore reject this interpretation. Instead, we understand his assertion as an offer of certain *odds* on the outcome of the experiment $[x, a]$. We can still never disprove the theory; but we can at least render bankrupt the theorist.

A given theory thus associates with each elementary experiment $[x, a]$ a probability (of the outcome 1) which we call the *expectation* of this experiment and denote by (x, a). Our interpretation of probability lies between the objective and purely subjective views. It is not objective, since the probabilities are not asserted to be properties of nature; they belong to a physical theory. Nor is it purely subjective, since in general the theory will be accepted by many persons.[13]

We do not attempt to discuss the process by which an observer comes to accept a particular theory (in the above sense). However, there is one matter concerning the simplification of such a theory which should be discussed. We have defined a state as a collection of methods of preparation. One may ask how are these collections formed: In what way does the observer decide that two methods of preparation may be assigned to the same state? In reply, we now describe how the collections may be *enlarged*. Since there is no problem in starting with collections of one member, this answers the question.

The enlargement is carried out by the identification of "equivalent" states. We define two states x and y to be *equivalent* if, whenever x appears in an assertion of the form (1), the validity[16] of this assertion is unaffected by replacing x by y and conversely. As far as all statements of the form (1) are concerned, equivalent states need not be distinguished. We may therefore identify all states in each equivalence class.

A similar discussion applies in the case of tests. Two tests are called *equivalent* provided they are interchangeable in every assertion of the form (1). Equivalent tests may be identified without ambiguity, if desired.

Of course, by definition, the equivalence of states and tests is relative to a given theory. An improvement in the theory may result in methods of preparation that had previously been considered equivalent becoming inequivalent. As an example, two methods of preparation that produce electrons of distinct polarization but that are otherwise similar are interpreted as realizing the same or different states according to whether or not there are available polarization-sensitive detectors which are recognized as such in the current theory. Similar remarks apply to methods of analysis.

We next consider the requirement (b). According to it, any state mentioned in an assertion of the form (1) must denote a method of preparation that *can actually be carried out*. We call such states *physical states*, leaving the general term "state" free for other purposes.

Let us determine the place of physical states in conventional quantum mechanics. In this theory,[9] there is associated with a physical system a Hilbert space H_c.[17] Any state x is represented by a positive Hermitian operator (density matrix) D_x of trace 1, and any test a by a projection A_a in such a way that always

$$(x, a) = \mathrm{Tr}\,(D_x A_a). \tag{2}$$

For $0 < \alpha < 1$, the weighted mean $z = \alpha x + (1 - \alpha)y$ of two states x and y is defined by $D_z = \alpha D_x + (1 - \alpha)D_y$ or, equivalently, by

$$(z, a) = \alpha(x, a) + (1 - \alpha)(y, a), \tag{3}$$

for every test a, and is again a state. The states thus form a convex set. The extreme points of this set are the *pure states*, represented by projections of rank 1; all other states are called *mixed*. Any pure state x determines, up to a phase factor, a unit vector $|\psi_x\rangle$ in H_c such that $D_x = |\psi_x\rangle\langle\psi_x|$, the projection on the subspace spanned by $|\psi_x\rangle$.

Now suppose x and y are physical states. Then so is $z = \frac{1}{2}(x + y)$, for the following constitutes a method of preparation for z: "Throw a die. If the number shown is odd, apply the method of preparation x; if it is even, apply y." In the same way, we can show that any weighted mean of two physical states is physical, so that the set S_p^o of all physical states is a convex set.[18]

Before proceeding, we deal with three criticisms that may be offered concerning the above description of the method of preparation z. First, if the elementary experiment $[z, a]$ is performed, we might be inclined to defer our declaration of the probability of the outcome until the die had been thrown or, equivalently, to make it dependent on the result of the throw. But this we cannot do: We are required to offer odds on the outcome of the experiment $[z, a]$ on the evidence of the documents z and a alone. It is clear that the only consistent thing to do, if we believe the die to be unbiased, is to take for (z, a) the value given by (3), with $\alpha = \frac{1}{2}$.

A second criticism might be expressed in the following terms. "The method of preparation proposed for z coincides, on every occasion, either with a preparation of x or with a preparation of y. It is therefore not an admissible method of preparation. Indeed, we can

give a more formal argument as follows. Suppose z is a state, and assume $x \neq y$. Every execution of a method of preparation, x' say, for x could be imagined as preceded by a throw of the die with an odd number being obtained. But it is then, at the same time, a method of preparation for the state z. Consequently, x' is, in effect, a "refinement" of a method of preparation for z and thus is itself a method of preparation for z. But this means $x = z$. Similarly $y = z$. Thus $x = y$, a contradiction."

In reply, we observe that no proper instructions have been given for the method of preparation x'. The only instructions which might have been offered would be: "Throw the die. If the number shown is odd, apply the method of preparation x; if it is even, consider the experiment void." However, because of the last phrase (which cannot simply be omitted) the execution of this method of preparation does not in general constitute an execution of z. Indeed, in the case of z, the experiment is never rendered void simply as a result of the number shown on the die.

A third, relatively minor, objection to the definition of z is the observation that it involves the introduction of a foreign element—namely, the reference to a die or some similar chance event. However, this need not be so. The operation of throwing the die may itself be replaced by the execution of an arbitrary elementary experiment whose expectation is $\frac{1}{2}$; indeed, an elementary experiment has exactly the properties—a well-defined expectation independent of other events—for which we employed the die. (In fact, with a slight elaboration we can use *any* elementary experiment whose expectation is neither 0 nor 1.) This concludes the discussion of possible objections to the definition of the weighted mean of two physical states.

We now consider two examples which suggest that the conventional theories contain nonphysical states and, in particular, that all pure states are nonphysical. The first is the example of the states of polarization of a light photon—we are dealing here with a quantum theory with a 2-dimensional Hilbert space. Suppose the light is travelling horizontally and we wish to prepare a pure state in which the polarization is vertical. To do this, we would pass the light beam through a suitable device (Nicol prism, polarizing filter, etc.). However, the resultant light is never perfectly plane polarized, so that such a device never produces a pure state; it seems that a pure state is not physical.

Note, by the way, that according to the present approach every photon in the beam is in the same (mixed) state, for they have all been prepared in the same way and the same predictions [i.e., assertions of the form (1)] would be made for any of them. A common attitude is to remark that the beam is indistinguishable from one obtained by superimposing (incoherently) beams, of suitable intensity and of vertical and horizontal polarization, and therefore to regard each photon as either vertically polarized or horizontally polarized. According to this attitude, different photons are in different "states." It is not clear what definition of "state" justifies this view, but it cannot be the one we have explained above.

As our second example, we take the case of classical mechanics. Here, according to the conventional view,[2,10,19] a test is represented by a Borel subset of phase space Ω and a state is a probability measure $d\mu$ on Ω. A pure state is therefore a "δ function" (a measure whose support is one point)—for instance, the state "at rest at the origin." It is clear that no actual procedure can realize such a state—there is always a slight uncertainty in position and momentum. (This, of course, has nothing to do with quantum theory—we are working with classical mechanics. It represents the necessary presence of "noise," e.g., thermal vibration, in the apparatus employed in the preparation of the state.) Thus, as in case of light polarization, every physical state is mixed. However, not every mixed state is physical—consider, for instance, a weighted mean of two pure states. Indeed, it is not unreasonable to suppose that every physical state is of the form $d\nu = f(\lambda)\, d\mu$, where $f(\lambda)$ is a smooth (i.e., infinitely differentiable) function on the phase space and $d\mu$ denotes Lebesgue measure on the manifold Ω.

We have seen that the physically significant assertions of a quantum theory must refer to *physical* states. Moreover, it is strongly suggested by the above examples that every physical state is mixed. It follows that mixed states *must* have a place in the theory. Here the situation is satisfactory: Mixed states have, in fact, a well-defined representation in conventional quantum mechanics. The situation is not so satisfactory in the case of tests, as we now see.

To illustrate this, we consider the same two examples as before. As a typical test in the case of light polarization consider that which asks the question: Is the photon polarized vertically? To realize this test, we require an arrangement which always responds to a photon if the plane of polarization is vertical and never responds if the plane of polarization is horizontal. Such an arrangement may be approximated by a device consisting of a detector placed behind a suitable filter. This device responds with a high probability $1 - \epsilon$ to a vertically polarized photon and with a low probability δ to a horizontally polarized

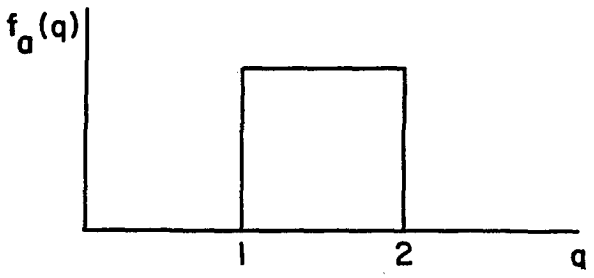


FIG. 1. Graph giving the probability of response as a function of the coordinate of the particle.

photon, where ϵ and δ are some numbers which may be small but which are not zero.

Certainly, on each occasion, the device either does or does not respond, so that it must realize some test a. If we compute the probability of response to a photon in a (not necessarily pure) state of polarization x, we obtain:

$$\text{Probability of response} = (x, a) = (1 - \epsilon)\,\mathrm{Tr}\,(D_x E_v) + \delta\,\mathrm{Tr}\,(D_x E_h),$$

where D_x is the density matrix representing the state and E_v and E_h are the projections corresponding to tests of horizontal and vertical polarization respectively, i.e., $E_v = |\psi_v\rangle\langle\psi_v|$, $E_h = |\psi_h\rangle\langle\psi_h|$, where $|\psi_v\rangle$ and $|\psi_h\rangle$ are vectors representing pure states of vertical and horizontal polarization, respectively. This gives

$$(x, a) = \mathrm{Tr}\,(D_x A_a), \tag{4}$$

where $A_a = (1 - \epsilon)E_v + \delta E_h$. Thus, Eq. (2) still holds for the test a, provided we represent this test not by a projection but by the Hermitian operator A_a.

By considering other devices of the above type, we arrive at the conclusion that, for any physical test a, (x, a) is always given by Eq. (2) for some Hermitian operator A_a whose eigenvalues may lie anywhere in the interval (0, 1). We describe a test a as a *mixed* test whenever A_a is not a projection, as a *pure* test if A_a is a projection. Only pure tests are discussed in conventional quantum mechanics. On the other hand, our example suggests that physical tests are normally mixed.

To throw more light on the situation, let us discuss our second example. As we have already observed, in conventional classical mechanics any test a is represented by a Borel set E_a in phase space Ω, and the probability of response when this test is applied to a system in a state x, represented by a measure dv_x on the phase space, is

$$(x, a) = \int_{E_a} dv_x = \int_{\Omega} f_a(\lambda)\, dv_x, \tag{5}$$

where $f_a(\lambda)$ is the characteristic function of E_a, defined by $f_a(\lambda) = 1$ for $\lambda \in E_a$ and $f_a(\lambda) = 0$ for $\lambda \notin E_a$. For example, consider a particle which moves on a line, and suppose that a is the test which answers the question: Does the coordinate q lie between 1 and 2? Then f_a can be regarded as a function of the coordinate q alone, and its graph (Fig. 1) gives the probability of response as a function of the coordinate of the particle.

To realize a, we need a kind of "spectrometer" that always responds when $1 < q < 2$ and never responds otherwise. It is clear that no actual device behaves in this way. Rather (Fig. 2), a real spectrometer is characterized by a probability of response which is at least a *continuous* function of q; for near the ends of its range the signal becomes dominated by noise, so that the probability of response drops continuously from 1 to 0. Indeed, it is reasonable to expect that this *response function* is a *smooth* function of q. For instance, in a simple model it is obtained by convolution of an ideal response function with a Gaussian probability distribution of errors.

Such a device certainly realizes a test, since on every occasion it either does or does not respond. What will be its probability of response to a (not necessarily pure) state x represented by a measure dv_x? The answer is surely that Ep. (5) still applies, where f_x is, however, no longer the characteristic function of a Borel set but rather the continuous response function of the device.

We call a test a in classical mechanics a *mixed* test if (x, a) is given by (5) with a function f_a which is not idempotent, in contrast with the *pure* tests of conventional classical mechanics for which f_a is idempotent. The above discussion strongly suggests that in classical mechanics, as in quantum mechanics, physical states are mixed.

The same argument as was used above in the case of states now applies: The physically significant assertions of classical or quantum mechanics must refer to physical tests, for no other tests can actually be realized. Thus, the concept of a mixed test must appear in the theory. It is a serious fault of conventional

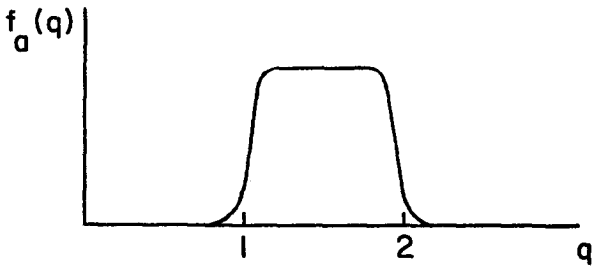


FIG. 2. A real spectrometer is characterized by a probability of response which is at least a continuous function of q.

quantum mechanics (and, for that matter, of classical mechanics) that this concept is not present.

The concept of a mixed test, which we have introduced, should not, of course, be confused with the conventional concept of an observable. Although both are represented in quantum mechanics by Hermitian operators (and in classical mechanics by functions on the phase space), the *interpretation* is quite different. Measurement of a test always gives either 0 or 1 as the outcome, whereas measurement of an observable gives, according to the conventional interpretation, a number in the spectrum of the corresponding operator. On the other hand, for a mixed test a, as for a conventional observable, the *expectation value* in the state x is given by $\mathrm{Tr}\,(D_x A_a)$ [or $\int f_a(\lambda)\,dv_x$ in classical mechanics].

In the same way as for states, we can form the *weighted mean* of two tests. This time, we describe the construction of an arbitrary weighted mean rather than the special case $\alpha = \frac{1}{2}$.

Let a and b be physical tests and let α be a number in the interval $(0, 1)$. We define the weighted mean $c = \alpha a + (1 - \alpha)b$ of a and b as a physical test by giving the following method of analysis for it. To apply the test c, first choose a number ρ *at random* in the interval $(0, 1)$. (This can be done, for instance, by determining, by throws of a coin, the successive digits in the binary expansion of ρ.) If $\rho < \alpha$, apply the test a; if $\rho > \alpha$, apply b. It follows from the construction that, for any state x,

$$(x, c) = \alpha(x, a) + (1 - \alpha)(x, b). \tag{6}$$

This equation justifies the notation $c = \alpha a + (1 - \alpha)b$.

The possibility of a direct construction of weighted means gives rise to another advantage of the concept of a mixed test. As we have seen, mixed tests are represented in the same way as observables. Now, in an axiomatic treatment of quantum mechanics, if a, b, and c are observables, it is customary to *define* $c = a + b$ to mean $(x, c) = (x, a) + (x, b)$ for every state x, but we have to *postulate* that, if a and b are observables, there is an observable c such that $c = a + b$. However, if a and b are physical tests, we can *construct* not quite their sum but their mean $c = \frac{1}{2}(a + b)$; no corresponding postulate is required.

The notion of a weighted mean provides a justification for the terms "mixed test" and "pure test." First consider classical mechanics. To any physical test a there corresponds a unique[20] function f_a and, since the expression (x, a) in Eq. (4) is a probability, the range of this function must certainly lie in the closed interval $[0, 1]$. Clearly, any such function determines (up to equivalence) not more than one physical test. Let us use the general term "test" to describe the idealized procedure represented by *any* such function (not necessarily even a Borel function), so that a physical test is a test which can actually be realized in practice. It is easy to show that the tests form a convex set of functions (of which the physical tests form a convex subset) and the extreme points are exactly the pure tests.

Similarly in quantum mechanics: on the basis of Eq. (3), we can set up a 1-to-1 correspondence between the physical tests and a certain class of Hermitian operators whose spectrum lies in the closed interval $[0, 1]$.[20] Let us designate as a *test* the idealized procedure represented by *any* Hermitian operator that has this property. The tests then form a convex set of Hermitian operators, of which the extreme points are the pure tests. Again, the physical tests form a convex subset.[22]

For every system, there are two special tests $\mathbf{0}$ and $\mathbf{1}$ defined by the equations $(x, \mathbf{0}) = 0$ and $(x, \mathbf{1}) = 1$ for every state x. They correspond to the always-false and the always-true propositions, respectively. These tests should always be regarded as physical tests: $\mathbf{0}$, for instance, is realized by the trivial method of analysis: "no action is required; the outcome is 0 in any case." Apart from these special cases, a physical analysis of each particular system is required to determine the class of physical tests. As we have seen, in the case of classical mechanics a reasonable necessary condition is that, for any physical test a, f_a should be a smooth function on the infinitely differentiable manifold Ω. If Ω is compact, it is reasonable to suppose that this condition is also sufficient. In the noncompact case, however, we must usually also impose some condition on the "behavior at infinity" of f_a. In all cases, it seems, a sufficient condition is that f_a should tend to a constant, but this is by no means necessary. On the other hand, to impose no condition at infinity is too weak for any realistic interpretation of "physical." In the case of light polarization, our discussion suggests that the physical tests are those represented by all Hermitian operators in the interior of the region $\{A: \mathbf{0} \leq A \leq \mathbf{1}\}$, together with $\mathbf{0}$ and $\mathbf{1}$.[23] It is reasonable to suppose that the same holds for any quantum mechanical system provided the Hilbert space H_c is finite dimensional. When H_c is infinite dimensional, the situation is much less simple, each case requiring its own analysis. We discuss this further in Sec. 6 and a few examples are treated in detail in the Appendix.

4. THE CONCEPT OF PHASE SPACE

One of our objects in the rest of this paper is the determination of the quantum analog of the concept

of phase space. To pave the way for this discussion, we treat now the case of classical mechanics.

Here we have a phase space Ω, which we assume to be an infinitely differentiable manifold. For simplicity, we shall at present assume Ω to be compact. We assume not only (as in Sec. 3) that every physical test, other than **0** and **1**, is represented uniquely by a smooth (i.e., infinitely differentiable) function on Ω to (0, 1), but also that every such function represents a physical test. We denote by T_p^o the class of all physical tests.[18] As explained in Sec. 3, we use the plain term "test" to denote any function on Ω to [0, 1]; for simplicity, we do not distinguish between a test and the function which represents it. Also, following Sec. 3, we assume that every physical state is represented by a probability measure of the form $f(\lambda)\, d\mu$, where $f(\lambda)$ is a smooth function on Ω and $d\mu$ denotes Lebesgue measure, and that every such measure represents a physical state. We denote by S_p^o the set of all physical states.[21]

Mathematically, there are three levels of detail at which we can view the phase space Ω. At the finest level, we regard Ω as it really is—as an infinitely differentiable manifold. At the second level, we choose to ignore the manifold structure on Ω. It then appears as a locally Euclidean topological space. Finally, at Level 3, we ignore the topology as well, and Ω then appears simply as a set.

From a physical point of view, each of these viewpoints may be regarded as arising from a certain attitude towards the tests. Suppose, for the present, that we are given the *set* Ω. Then the finest level corresponds to the possibility of distinguishing among the tests the subset T_p^o consisting of those which can be exactly realized in practice. For not only does the manifold structure on Ω determine this class, but the set of functions T_p^o uniquely determines, and indeed practically *is*, the manifold structure.

The second level is reached as follows. First observe that any continuous function on Ω to [0, 1], and only such a function, can be approximated uniformly on Ω by a physical test. Being so nearly realizable, these functions might be considered to be practically physical tests. Let us call them *uniform tests*. Of course, the set T^o of all uniform tests may arise from many different sets T_p^o.[24] Now we might take the view that, since every uniform test can be "practically" realized in the laboratory, there is no point in distinguishing the particular subset T_p^o from which T^o arose. If we decide to ignore this subset, we—in this very act—renounce the possibility of recovering the manifold structure on Ω, for T^o does not determine this structure. However, T^o does uniquely determine the topology[25–27]; and, of course, the topology on Ω determines T^o. So we have passed to Level 2.

The passage from Level 2 to Level 3 is accomplished in a similar manner. *Any* function on Ω to [0, 1] can be approximated in a rather crude way by uniform tests: namely, given any finite number of points in Ω, we can find a uniform test which differs arbitrarily little from the given function at all these points. So we can, with a stretch of the imagination, pretend that any such function represents a sort of test. Call these *weak tests*.[28] The set T_w^o of all weak tests is, of course, independent of the topology on Ω. So, if we choose to take the—admittedly crude—view that, since all weak tests can be approximated by uniform tests, all are equally good, and therefore decide to ignore the distinguished subset T^o, we lose in so doing the possibility of recovering, from the structure of the tests, the topology of Ω. In this way, we have passed to Level 3: Ω appears now as merely a set. Observe that, in passing from Level 2 to Level 3, one loses one very important physical property of the system—namely, the number of degrees of freedom. This property manifests itself through the dimension number of phase space, which is a topological property.

There are usually very few pure uniform tests: if Ω is connected, the only ones are **0** and **1**. On the other hand, there are plenty of pure weak tests. Indeed, they are in 1-to-1 correspondence with the subsets of Ω and so form a complete atomic Boolean algebra,[29] the atoms being the 1-point sets. The abundance of pure tests, combined with the mathematical simplicity and physical convenience of this concept, is the main justification for the introduction of weak tests.

There are actually two other levels of detail slightly finer than Level 3. The first places the emphasis on measurable functions. In 1936, Birkhoff and von Neumann[11] observed that any set of (Lebesgue) measure zero corresponded to a pure test a for which $(x, a) = 0$ for every physical state x. Accordingly, they proposed to ignore such sets and to represent (pure) tests by measurable sets modulo sets of measure zero. In terms of mixed tests, this corresponds to introducing *measurable tests* as represented by measurable functions on Ω to [0, 1] modulo functions of measure zero. There are plenty of pure measurable tests; they form a Boolean algebra which is complete but not atomic.[30]

Starting with the uniform tests, we may reach the measurable tests by following the procedure of Daniell integration theory.[31] (This requires the use of one physical state—any physical state defines a measure which is absolutely continuous with respect to Lebesgue measure.) If we then ignore the subclass

(consisting of the uniform tests) from which we started, we pass to *Level 3a*. Here, Ω takes the form of a *measurable space*.

More recently, it has become customary[2,10,19] to assign a basic role to the Borel sets. These arise in our treatment in the following way. Let us define a *Borel test* to be a Borel function from Ω to $[0, 1]$. A Borel test can be approximated by physical tests in a slightly stronger way than a weak test. Also, if a is a Borel test, (x, a) exists for every physical state x (and even, as we shall see later, for every "weak state" x), whereas this is by no means the case for an arbitrary weak test a. There are plenty of pure Borel tests—corresponding to the Borel subsets of Ω; they form a Boolean algebra which, although atomic, is not complete.

If, starting at Level 2, we construct the class T_B^o of all Borel tests and then ignore the subset T^o of T_B^o consisting of the uniform tests, we lose sight of the topology of Ω but are still able to distinguish its Borel structure; i.e., Ω appears as a Borel space. We call this level of detail *Level 3b*.

With these thoughts regarding the structure of classical mechanics in mind, let us now examine quantum mechanics. We have pointed out one serious fault in conventional quantum mechanics: It fails entirely to discuss mixed tests—all the tests mentioned in that theory are pure. A second major blemish is this: it fails to represent, in the underlying mathematical structure, the number of *degrees of freedom* of the system. Thus, whether we are discussing a harmonic oscillator, a rigid rotator, a particle moving on a closed curve, or on a line, or in two or three dimensions, or even a system of many particles, the Hilbert space H_c is always the same, and the observables—according to the conventional theory—form the set of all bounded Hermitian operators on H_c.

Regarding these two faults in the light of the above discussion of classical mechanics, it is clear what has happened: the system is being viewed at Level 3. This is shown both by the fact that the treatment can be given in terms of pure tests—only at Level 3 are there (in general) any nontrivial pure tests—and by the fact that no trace remains, in the general structure, of the number of degrees of freedom. This is exactly what would happen in classical mechanics if we ignored the topology of phase space and proceeded to Level 3.

A clear program now lies before us. To recover the lost detail and reveal the structure of a quantum-mechanical system at Levels 1 and 2, we must first make a careful physical analysis of the nature of the system, with the aim of recognizing the physical tests. If the conventional theory is correct (as far as it goes), then every mixed test should be represented by a Hermitian operator with spectrum lying in $[0, 1]$. Hopefully, our physical analysis will allow us to decide which of these operators represent physical tests. The class T_p^o of all physical tests is thus identified with a convex set of such operators; this set is the quantum mechanical analog of the set of smooth functions that corresponds to T_p^o in the classical case.

We now have two possibilities. On the one hand, we may investigate the structure of T_p^o; this should lead us to the quantum analog of a manifold. On the other hand, recognizing that a manifold is a much more sophisticated concept than its underlying topological space, we might try, in the first instance, to pass to Level 2 by identifying the class T^o of all uniform tests. This leaves us with the presumably easier task of investigating the structure of T^o in the hope of discovering the quantum analog of the concept of a topological space.

This same program also arises if we attempt to give a formulation of mechanics which is "ideal" in the sense of Sec. 2. To this end, we must first select suitable primitive concepts, which should be chosen so that the physical concepts to which they refer are as direct as possible. Now, although the (pure) states and tests of conventional quantum mechanics purport to represent physical concepts, we have seen above that they are really only limiting cases of physical states and tests. The latter concepts are much more direct, and we therefore choose *them* as primitive concepts for the new formulation. For the kinematic part of mechanics, these concepts, together with the expectations of the form (x, a), suffice.

The axiomatic theory which arises in this way is sketched in Secs. 5–8. It can be interpreted as representing the progress of a *primitive observer for mechanics*, who is able only to recognize physical states and tests and carry out elementary experiments. In this theory, all other concepts are defined in terms of physical states and tests and expectations. These definitions may be understood as describing the process of concept formation that goes on in the mind of our primitive observer.

So far, we have not distinguished classical and quantum mechanics, and indeed it is not clear, *a priori*, how our primitive observer is to tell whether he is dealing with a classical or a quantum system. The theory proceeds for some time without distinguishing the two cases. Even after the separation has occurred, it is natural, as far as possible, to use the same procedures in developing the two branches; in this way, quantum analogs for classical concepts appear in a natural way.

This approach leads also to a new view of classical

mechanics. Our primitive observer's first physical example of a manifold is phase space, and to him the primitive concept is not the points but the *algebra of smooth functions*. This ties in with modern differential geometry, where this algebra plays a fundamental role. Similarly, whereas to us the simplest concept connected with a topological space is a point and the concept of a continuous function is relatively sophisticated, to our primitive observer the order is reversed, for it is the continuous function which (for him) is closer to having a concrete physical significance.

In their conventional formulations, the mathematical structures of classical and quantum mechanics are very different. Hilbert space, for example, which is sometimes considered to be the quantum analog of phase space, is a very different structure from phase space; it is simpler in being always a linear space, but more complicated in being infinite dimensional. But, from the viewpoint of our primitive observer, the difference between classical mechanics and quantum mechanics is relatively small. For instance, in both cases it turns out that the uniform tests T^o fill the interval $\{a: \mathbf{0} \leq a \leq \mathbf{1}\}$ in the real part of a C^*-algebra A. In the classical case, A is commutative; in the quantum case, it is not.

The "inverted" view of various mathematical structures, which arises in this way, is valuable even in a purely mathematical respect in suggesting new developments and generalizations. Perhaps it is we, and not our primitive observer, who are looking at things upside down!

In the rest of this paper, we describe in more detail the mathematical structure of this formulation of quantum mechanics. Naturally, the axioms, and to some extent the definitions, are motivated by concrete examples of physical systems. To avoid interrupting the continuity of the argument, these examples are discussed separately in the Appendix, and it may be desirable that this should be read in parallel with Secs. 5–8.

5. C^*-SYSTEMS

In the last section, we introduced the concept of a primitive observer for mechanics. Let us imagine we are watching such an observer as he carries out experiments and develops a theory of the world around him. We assume he has available a variety of systems.[32] Each system can be subjected to any one of a set S_p^o of methods of preparation (states) and, when so prepared, can be tested by any of a set T_p^o of methods of analysis (tests). By induction based on experiment, our observer arrives at an expectation (x, a) that a positive outcome will be obtained when the method of analysis a is applied to a system prepared by the method of preparation x. Being only interested in the values of the expectations, he decides not to distinguish two states x and y, which have the same expectation for all tests, and similarly for tests. He tabulates the values of (x, a) as a function of x and a. His object is to discover any structure which may underlie these observations. Such a structure would (a) expedite the making of predictions and (b) serve as a basis for the assignment of expectations to elementary experiments which have never been carried out.

Of course, there are certain properties whose validity is evident without any reference to his experimental results. For instance, he can state with certainty:

Axiom 1: S_p^o and T_p^o are two sets whose elements are called *physical states* and *physical tests*, respectively. T_p^o contains distinguished elements $\mathbf{0}$ and $\mathbf{1}$. To every physical state x and physical test a, there corresponds a real number (x, a), such that

$$\text{(i)} \quad \begin{cases} (x, a) = (y, a) & \text{for all } a \text{ implies } \quad x = y, \\ (x, a) = (x, b) & \text{for all } x \text{ implies } \quad a = b; \end{cases}$$

and

$$\text{(ii)} \quad \begin{cases} 0 \leq (x, a) \leq 1, & \text{for all } x \text{ and } a; \\ 0 = (x, \mathbf{0}) \quad \text{and} \quad 1 = (x, \mathbf{1}), & \text{for all } x. \end{cases}$$

Next, our observer recognizes that he has always the possibility of forming weighted means of states and tests, as described in Sec. 3. In virtue of the "separation property" (i), each state may be regarded as a function on T_p^o and each test as a function on S_p^o. From the experimental meaning of the concept of weighted mean, our observer concludes that the linearity properties [Eqs. (3) and (6) of Sec. 3] must hold. So he can state with confidence:

Axiom 2: Regarded as a set of functions on T_p^o, S_p^o is a convex set. Similarly, regarded as a set of functions on S_p^o, T_p^o is a convex set.

Next, if a is any test, our primitive observer can always obtain another test a' by exchanging the labels 0 and 1 on the apparatus which realizes a (or rather, on the document which describes the method of analysis a). Since a' is the same method of analysis, with merely the reverse interpretation of the outcome, he knows that $(x, a') + (x, a) = 1$ for every state x. With the function interpretation of tests, this gives:

Axiom 3: If a is a physical test, so is $\mathbf{1} - a$.

All these assertions follow from the meanings of the terms employed, regardless of the experimental values of the expectations.

The next stage in his work is the construction of the uniform tests. In Sec. 4, we introduced these, in the case of classical mechanics, as those tests which could be approximated, uniformly on phase space Ω, by physical tests. Here, even if he is confronted with a classical system, our primitive observer has no access to Ω, so this definition is unavailable. However, in the example of classical mechanics described in Sec. 4, an equivalent condition is to demand uniform approximation on the physical states. Let us suppose, then, that our primitive observer adopts the following definition:

Definition 1: In the linear space of all real-valued bounded functions on S_p^o, let $\| \ \|$ denote the sup norm and let $\geq$ denote the pointwise ordering:

$$\|a\| = \sup \{|(x, a)| : x \in S_p^o\},$$

$$a \geq 0 \quad \text{means} \quad (x, a) \geq 0, \quad \text{for all } x \text{ in } S_p^o.$$

Let T^o denote the norm closure of T_p^o. The elements of T^o are called *uniform tests*. Let T_p and T be the linear spaces spanned by T_p^o and T^o, respectively.

It follows at once from Definition 1 that T is a partially ordered linear space with order unit $\mathbf{1}$, containing a distinguished convex subset T^o. Given $a \in T$, a *necessary* condition that $a \in T^o$ is clearly $\mathbf{0} \leq a \leq \mathbf{1}$. Although it does not follow from Axioms 1–3, it seems to be true in practical cases that this condition is also sufficient:

$$T^o = \{a : \mathbf{0} \leq a \leq \mathbf{1}\}.$$

This property would be recognized by our primitive observer as a *law of nature*. However, there is a much more profound observation which is also accessible to him: *T, as a partially ordered linear space, is isomorphic to the real part of a C^*-algebra.* To appreciate the evidence for this, consider first the case, introduced at the beginning of Sec. 4, of a classical-mechanical system with a compact phase space Ω. The space T, introduced in Definition 1 as a partially ordered linear space of real-valued functions on S_p^o, appears in this case as the partially ordered linear space of *all* continuous functions on Ω. Moreover, this representation of T preserves the order and the norm. Now, this latter space is, via the identity map, isomorphic to the real part Re $[C(\Omega)]$ of the C^*-algebra[33] $C(\Omega)$ of all complex-valued functions on Ω (with pointwise multiplication on Ω). Conversely, by the well-known representation theorem for commutative C^*-algebras, if T is isomorphic to the real part of any commutative C^*-algebra, then it may be represented as the linear space of all real-valued continuous functions on a compact Hausdorff space.[26]

If Ω is not compact, the nature of T depends on the conditions at infinity satisfied by the physical tests (see the end of Sec. 3). A study of a limited number of physical examples suggests that always $T = \text{Re } [C(X)]$, where X is some *compactification* of Ω. If there are no conditions at infinity, X is the Stone–Čech compactification[27] of Ω; if—the opposite extreme—every physical test tends to a constant at infinity, then X is the 1-point compactification of Ω; in general, intermediate compactifications arise.

Now let us consider the quantum-mechanical case. The simplest example is the case of light polarization discussed in Sec. 3. In this case, T_p is finite dimensional and so coincides with its completion T. Thus, $T = \text{Re } [L(H_c)]$, where H_c is the Hilbert space (in this case, 2-dimensional) associated with this system in the conventional theory; and, of course, $L(H_c)$ is certainly a norm-closed $*$-algebra of operators—i.e., a C^*-algebra. The case of quantum mechanics over an n-dimensional Hilbert space presents no new features. In the case of infinite dimension, however, T turns out in examples (see the Appendix) to be $T = \text{Re } (A)$ for some C^*-algebra A which is now a proper subalgebra of $L(H_c)$. (Indeed, in these examples, even T_p is the real part of a $*$-algebra of operators A_p. A_p is not norm closed; it may be regarded as the quantum analog of the algebra of smooth functions on phase space that appears in the case of classical mechanics. This algebra should be important in connection with the quantum generalization of a manifold.)

Let us make the following definition:

Definition 2: A *C^*-system* is a structure satisfying Axioms 1–3, such that

(i) $T^o = \{a : a \in T, \mathbf{0} \leq a \leq \mathbf{1}\}$, and

(ii) T is isomorphic, as a partially ordered linear space, to the real part Re (A) of a C^*-algebra A with $\mathbf{1}$ corresponding to the unit of A.

With this definition, our primitive observer can express the preceding observations by announcing the *law of nature*:

Every system is a C^*-system.

He will have less confidence in this law than in the preceding axioms, since it is based not only on definitions but on experience. However, its apparent validity certainly motivates a thorough investigation of C^*-systems. This program occupies us throughout the next three sections.

6. THE CONSTRUCTION OF PURE STATES

In this section and the next two, we continue the development of classical and quantum mechanics

from the point of view of a primitive observer. It is assumed, unless otherwise stated, that the system under study is a C^*-system.

The concept of a pure state plays an important role in both classical and quantum mechanics. This is partly because the specification of a pure state is simpler than that of a mixed state and any mixed state can be obtained as a superposition of pure states, and partly because of the behavior of pure states in connection with the operation of *uniting* systems: If a composite system is formed by the union of two subsystems, then a unique (pure) state of the whole is determined by specifying pure states for the two parts—the situation for mixed states is more complicated. Now, physical states are normally mixed, so our primitive observer has no immediate access to pure states; for him, such states must be *constructed.*

In the case of classical mechanics on a compact phase space Ω (see Sec. 4), the physical states are represented by certain smooth probability measures on Ω and the pure states should surely be represented by the points of Ω or, equivalently, by the point measures or δ functions on Ω. Any pure state or physical state thus determines a regular probability measure on Ω or, equivalently, a normalized positive linear functional on T. This suggests the following definition:

Definition 3: A *weak state* x is a linear functional $a \mapsto (x, a)$ on T to the reals such that

(i) x is *positive*: i.e., $a \geq 0$ implies $(x, a) \geq 0$;

(ii) x is *normalized*: i.e., $(x, \mathbf{1}) = 1$.

It follows easily from the definition that the weak states form, in the dual S_w of the Banach space T, a convex set S_w^o which is compact in the weak topology. The Krein–Milman theorem then guarantees that there are plenty of pure (weak) states: S_w^o is the weakly closed convex hull of the set S_w^p of all pure states.

The term "weak state" is justified by the following fact. Any weak state x (and, in particular, any pure state) may be "weakly approximated" by physical states: i.e., given any finite set of (uniform) tests $a_1, \cdots, a_n$ and any $\epsilon > 0$, there is a physical state y with $|(x, a_i) - (y, a_i)| < \epsilon$, for $1 \leq i \leq n$.[4] This follows from Definition 2 with the aid of the Hahn–Banach theorem.

These immediate conclusions from Definition 3 are valid for any system satisfying Axioms 1–3. Henceforth, however, we consider only C^*-systems.

We have introduced the term "weak state" so that "state" is available for other purposes, for instance in physical arguments. However, in the case of a C^*-system, the weak states correspond to the "states" of the C^*-algebra A, in the sense used by mathematicians,[34] i.e., to the normalized positive linear functionals on A. For the present, then, we refer to weak states simply as *states.*

Given any representation of a C^*-algebra A in a Hilbert space, any unit vector ψ determines a state x_ψ of A by $(x_\psi, a) = \langle \psi, a\psi \rangle$, and this *vector state* x is pure if and only if the representation is irreducible. Moreover, *every* pure state of A arises in this way and two irreducible representations give rise to the same pure states if and only if they are equivalent. Thus the *reduced atomic representation*[35–37] π of the C^*-algebra A of a C^*-system—which is defined as the direct sum of a maximal set of inequivalent irreducible representations of A—is characterized by the property that every pure state occurs in it exactly once as a vector state. The reduced atomic representation is always faithful; so we can, and sometimes shall, identify A with the concrete C^*-algebra $\pi(A)$.

The Hilbert space H in which $\pi(A)$ acts—which we call the *large* Hilbert space—is a direct sum of Hilbert spaces, one for each inequivalent irreducible representation. Borrowing the terminology of superselection rules, we call these *coherent subspaces* of H. Every pure state is associated with an *allowed* ray, i.e., a ray that lies on one coherent subspace. Moreover, every element of A, and in particular every uniform test, is represented by an allowed linear operator, i.e., one that commutes with the projection on every coherent subspace.

Let us define a state to be *discrete* if it is a countable linear combination of pure states. For any discrete state x, we can give an expression for (x, a) analogous to Eq. (4). First, for each pure state x, let $\tau(x)$ denote the projection on the ray in H that represents x. We clearly have

$$(x, a) = \mathrm{Tr}\,[\tau(x)\pi(a)]. \tag{7}$$

Now, the mapping τ extends by linearity and continuity to an order- and norm-preserving isomorphism, which we still denote by τ, of the norm-closed subspace S_{wt} of S_w, spanned by the pure states, onto the linear space A_{wt} of all allowed operators of trace class[38] with the trace norm. Moreover, (7) continues to hold for every x in S_{wt}. Consequently, it holds for every discrete state x, since the discrete states are exactly the states in S_{wt}.

For other states, this equation does not apply. The reason for this is easily seen by considering the case of classical mechanics over a compact phase space Ω. Here, $\pi(a)$ and $\tau(x)$ may be identified with

functions on Ω and the operator, trace, becomes summation (not integration) over all points in Ω; $\pi(a)$ is, of course, a continuous function. If x is a discrete state, $\tau(x)$ vanishes, except on a countable set of points in Ω. In fact, any state is represented by a measure on Ω; in the case of a discrete state, the measure is discrete, i.e., it is a countable linear combination of δ functions, and the function $\tau(x)$ gives the coefficients in this linear combination.

It is clear that Eq. (7) can only be expected to apply to discrete states. For general states, it should be replaced by some generalization of Eq. (5).

Not all the pure states of a C^*-system appear in the conventional theory. We see this in the discussion of particular quantum-mechanical systems in the Appendix. There we start, for practical reasons, with the conventional theory, in which the observables are the Hermitian operators on the conventional Hilbert space H_c. By physical considerations we distinguish, among these operators, a subset that corresponds to the physical tests, and this leads us to the C^*-algebra A. A is thus obtained as a concrete C^*-algebra of linear operators on H_c. We call this the *conventional* representation of A. Of course, it is faithful. Usually it is also irreducible; let us assume this to be the case. The conventional representation may then be taken to be a subrepresentation of the reduced atomic representation, so that H_c is one of the coherent subspaces in the large Hilbert space H. Now, a *conventional* pure state—i.e., a pure state in the sense of conventional quantum mechanics—is represented by a ray in H_c and is therefore a weak pure state. However, except in the trivial case when H_c is finite dimensional, there are other coherent subspaces also[39] and thus pure states other than conventional pure states. Indeed, given any uniform test a and any number α in the spectrum of a, there is[34] a pure state x such that $(x, a) = \alpha$; unless α is in the discrete spectrum, x is not a conventional pure state. We meet some concrete examples of such nonconventional pure states in the Appendix.

These nonconventional pure states are, of course, orthogonal (as rays in H) to every conventional pure state. One may wonder how it is that conventional quantum mechanics manages very well without such states. This is because any pure state can be weakly approximated (in the sense explained above) by conventional pure states; this follows from the fact that the conventional representation is faithful.[40]

It is important to remember that A is obtained above as a *concrete* C^*-algebra only because we have taken a short cut by appealing to the conventional theory. Faced with the task of determining A directly from experiment, we should have to adopt the same procedure as a primitive observer: to determine the values of (x, a) for many elementary experiments $[x, a]$ until the structure of S_p^o and T_p^o becomes clear and then to construct T and seek an A such that $T = \mathrm{Re}\,(A)$.[41] Thus, in principle, A is always obtained as an abstract C^*-algebra: the conventional representation has, *a priori*, no particular physical significance. In cases where this representation is the only faithful irreducible representation, it is, of course, determined by A, and this seems to happen with simple quantum-mechanical systems. But in more complicated cases, there could be several faithful irreducible representations,[42] and in classical mechanics there are none.

Nonconventional pure states arise also in the case of classical mechanics whenever the phase space Ω of the system is not compact. (This includes most practical cases.) In this case (see Sec. 5), $A = C(X)$, where X is some compactification of Ω. The pure states now correspond 1-to-1 to the points of X. There is thus always at least one nonconventional pure state. As in the case of quantum mechanics, any pure state can be weakly approximated by conventional pure states.

7. THE CONSTRUCTION OF PURE TESTS

It would be an understatement to say that pure tests play an important role in the foundations of quantum mechanics. As we have seen, they are the only tests considered, and through their interpretation as propositions they form the basis for the most popular approach to the subject.[2,10,11]

Nevertheless, the situation is the same as for states: Both in classical and quantum mechanics, pure tests have no actual realization in practical terms; to give them a physical meaning, they must be constructed out of physical states and tests. The construction (which may, as before, be regarded as a description of the process of concept formation in the mind of a primitive observer) is slightly more complicated than the corresponding procedure for states. We first describe this construction and then justify it.

Let us identify the C^*-algebra A of a C^*-system with its reduced atomic representation $\pi(A)$, which acts on the Hilbert space H (see Sec. 6). Thus, each uniform test a is identified with the Hermitian operator $\pi(a)$ which represents it. We also identify each discrete state x with the corresponding trace-class operator $\tau(x)$. These identifications are assumed until the end of Sec. 8.

It can be shown that the weak operator closure of A in $L(H)$ coincides[36] with the set A_w of all allowed operators. If H is expressed as a direct sum $H = \oplus \{H_i\}$ of coherent subspaces, then $A_w = \oplus \{L(H_i)\}$. The Hermitian operators in A_w form a partially

ordered linear space T_w; we set $T_w^o = \{a: a \in T_w, \mathbf{0} \leq a \leq \mathbf{1}\}$ and call the elements of T_w^o *weak tests*.

The extreme points of the convex set T_w^o are the pure (weak) tests. They coincide with the allowed projections in $L(H)$, i.e., with the projections in A_w. We denote the set of all pure tests by T_w^p. Since A_w is a von Neumann algebra, T_w^p is a complete orthocomplemented weakly modular lattice.[29] Obviously, it is also atomic.

To appreciate the significance of this construction, consider first the case of classical mechanics. In the general case, when the phase space Ω is not necessarily compact, $A = C(X)$ for a certain compactification X of Ω (see Sec. 5), and A_w is the algebra of all bounded complex-valued functions on X. The pure tests are thus in 1-to-1 correspondence with the subsets of X. Since $\Omega \subset X$ and the conventional pure tests (in the sense of Level 3—see Sec. 4) correspond to the subsets of Ω, every conventional pure test is a weak test. If Ω is compact, $X = \Omega$; in this case, the weak pure tests and the conventional pure tests coincide. In all other cases, there are nonconventional pure tests as well.

Now consider the case of quantum mechanics. Assume that A acts irreducibly in the conventional Hilbert space H_c so that we can embed H_c in H, as in Sec. 6. Then $A_w \supset L(H_c)$, so that every conventional pure test is a pure weak test. However, as in the case of the states, there are nonconventional pure tests also—unless $H_c = H$, which happens only when H_c is finite dimensional.[39]

We have yet to justify the term "weak test" in the general case. It is clear (see below) that we cannot expect to give a meaning to (x, a) for an arbitrary state x and test a. However, if x is a discrete state, then Eq. (7), which holds for every uniform test a, can be used to define (x, a) also for every weak test a. It is easy to show that, with this bilinear form, T_w can be identified with the dual of the Banach space S_{wt}.[38] (The complexification of S_{wt} then becomes the predual of A_w.) This fact has a number of important consequences.

First, the Krein–Milman theorem can be applied in the same way as in the case of states: The set T_w^o of all weak tests is convex and weakly[43] compact and is therefore the weakly closed hull of the set T_w^p of all pure tests. Thus, any weak test can be weakly approximated by a weighted mean of pure tests: In other words, given any weak test a, any finite set $x_1, \cdots, x_n$ of discrete states, and any number $\epsilon > 0$, there exists a weighted mean $b = \sum \alpha_i a_i$ of a finite number of pure tests $a_1, \cdots, a_N$ such that $|(x_i, a) - (x_i, b)| < \epsilon$, for $1 \leq i \leq n$.

Secondly, it is easy to show that T_w^o is the weak[43] closure of T_p^o so that every weak test, and in particular every pure test, can be weakly approximated (in the sense just explained) by physical tests. This justifies the application of the term "weak test" to the elements of T_w^o.

Note that the expectation (x, a) has not been defined for every weak state x and weak test a. In fact, even in the commutative case, where (x, a) is the integral of the function representing a with respect to the measure representing x, we could (unless x is discrete) take for a the pure test corresponding to a nonmeasurable set; (x, a) cannot then be defined in any useful way. In the commutative case, this difficulty is mitigated by the Daniell integration theory,[31] in which the integral, defined initially for the continuous functions, is extended to the bounded Baire functions. The analogous procedure in the quantum case is the following.[6,35]

Call a set of operators in $L(H)$ *σ-closed* if it contains, together with any *sequence* that converges in the weak operator topology, also the limit of the sequence. Let A_B be the *σ-closure* of A, i.e., the smallest σ-closed set containing A. A_B is called a Σ^*-algebra.[35] Let $T_B = \mathrm{Re}\,(A_B)$ and $T_B^o = \{a: a \in T_B, \mathbf{0} \leq a \leq \mathbf{1}\}$. Then, $T^o \subset T_B^o \subset T_w^o$. We call the elements of T_B^o *Baire tests*. Then, just as in the commutative case, the domain of the bilinear form $(\ ,\ )$ can be extended in a natural way from $S_w \times T$ to $S_w \times T_B$, so that the expectation (x, a) becomes defined for every weak state x and Baire test a. The reader is referred to Davies' paper[35] for details.

This quantum analog of the Daniell integral is of fundamental importance; it furnishes the construction corresponding to Level 3b of Sec. 4. However, our immediate concern is with the quantum analog of general topology. For this purpose, we return now to the discussion of the family of *all* pure tests.

The lattice of all projections in the von Neumann algebra A_w plays the same role in the mechanics of an arbitrary C^*-system as the lattice of all subsets of phase space does in conventional classical mechanics. It is convenient to have a terminology which emphasizes this analogy. We call these projections *q-sets*, and the minimal q-sets (the analog of points in phase space) will be called *q-points*. In the general case, as in the case of classical mechanics, the q-sets may be identified with the pure tests and the q-points with the pure states. Union and intersection of q-sets are to be taken in the lattice T_w^p of all q-sets. A *finite* q-set is a union of a finite number of q-points.

Formally, the q-sets are the projections in A_w. However, it is often convenient to visualize them as

the norm-closed allowed subspaces of H. Thus, given q-sets e and f, we say e *is contained in* f if $ef = e$, and e *is orthogonal to* f if $ef = 0$. Of course, in the case of classical mechanics, the q-sets are also in 1-to-1 correspondence with the subsets of a certain compactification of phase space, or, equivalently, with the characteristic functions of these subsets.

If a is a weak test, e a q-set, and α a number, we say $a = \alpha$ *on* e if the eigenspace of a of eigenvalue α contains e. Clearly, q-sets e and f are orthogonal if and only if there exists a weak test a with $a = 0$ on e and $a = 1$ on f. Thus *orthogonality*, rather than the property of having a zero intersection, is the analog for q-sets of *disjointness* of sets.

The q-sets of a C^*-system form a particular case of what we might call a *q-space*: namely, the lattice of projections of an "atomic" von Neumann algebra. Here, we call a von Neumann algebra *atomic* if its lattice of projections is atomic; this happens if and only if the algebra is of the form $\oplus\{L(H_i)\}$ for some set $\{H_i\}$ of Hilbert spaces. If its algebra is commutative, a q-space is isomorphic to the lattice of all subsets of a set; conversely, any such lattice is isomorphic to a q-space. In fact, the theory of q-spaces may be regarded as the quantum analog of the theory of sets.

8. THE QUANTUM ANALOG OF TOPOLOGY

We have now followed the route, outlined in Sec. 4, that, starting with any C^*-system, leads to the construction of pure states and tests. This shows us how a primitive observer would pass from Level 1 to Level 3. The pure tests form a q-space which we shall call "the q-space **1**"—naming it by giving its largest element, in the same way as a space is named in set theory and in topology. The q-space **1** is the quantum analog of phase space, devoid at this level of anything resembling a topology. As we saw in Sec. 4, in order to introduce a "topological" structure our observer must retrace his steps to Level 2: i.e., he must refer to the class T^o of uniform tests. To see how he might proceed, consider first the case of classical mechanics, assuming for simplicity that phase space Ω is compact. If Ω is metrizable, its topology is easily determined in terms of T^o: Define the support $\mathbf{e}(a)$ of a test a to be the complement in Ω of the set on which a vanishes[44]; then the open sets coincide with the supports of the uniform tests. If Ω is not metrizable, only a slight elaboration is required: Any open set e is now the support $e = \mathbf{e}(K)$ of some set K of uniform tests, where $\mathbf{e}(K)$ is defined to be the union of the supports of the elements of K.[27]

Now the same procedure can be adopted in the case of an arbitrary C^*-system. We define the support $\mathbf{e}(a)$ of a test a to be the smallest q-set such that $\mathbf{e}(a)a = a$ and the support $\mathbf{e}(K)$ of a set K of tests to be the sup of the supports of the elements of K. Finally, a q-set is called *open* if it has the form $\mathbf{e}(K)$ for some set K of uniform tests. In this way, a certain class of *open q-sets* is distinguished. We might call this structure on the q-space **1** a *q-topology* since it reduces, in the classical case, to a true topology.

Concerning the q-topology of the q-space of a C^*-system, the following can be proved.[45] We define a q-set e to be *closed* if $\mathbf{1} - e$ is open.

(1) Any union of open q-sets is open; any intersection of closed q-sets is closed. (But an intersection of two open q-sets is not necessarily open, and a union of two closed q-sets is not necessarily closed.)

(2) Any q-point is closed. So is any finite q-set.

(3) The q-space **1** is *Hausdorff*. (Given any two orthogonal q-points x and y, there are orthogonal open q-sets e and f with $e \geq x$ and $f \geq y$.)

(4) The q-space **1** is *compact*. (Define an *open covering* of the q-space **1** to be a set $\{e_i: i \in I\}$ of open q-sets with $\mathbf{1} = \bigvee\{e_i: i \in I\}$. Then any open covering has a finite subcovering.)

(5) The q-space is *normal*. (Given any two orthogonal closed q-sets e and f, there exist orthogonal open q-sets g and h, with $g \geq e$ and $h \geq f$.)

(6) It is possible to obtain a quantum analog of the Gel'fand representation theorem for *commutative* C^*-algebras. This theorem states that any commutative C^*-algebra A is isomorphic to $C(X)$ for a compact Hausdorff space X. Its relevance for classical mechanics is explained in Sec. 5, the real part of A being the space T and X being a compactification of phase space Ω. Now any measurable function f on X to the reals is determined by its inverse map f^{-1}, which assigns to each measurable set in the reals a Borel set in X, and f is continuous if and only if the inverse image of each open set is open. Remembering that a set in X is just a pure test (i.e., a projection in A_w), we see that f^{-1} is nothing but the spectral resolution of the multiplication operator given by the function f. We can now state the quantum analog of the Gel'fand representation theorem. Define an element a of T_w to be *q-continuous* if its spectral resolution assigns to every open set in the reals an *open* q-set. Then not only is every element of T q-continuous, but every q-continuous element of T_w belongs to T. [The first assertion is elementary; the second follows from (5).]

We conclude this section with some remarks on the form taken by the q-topology of a C^*-system in some special cases. Let us assume, as in Sec. 6, that A has

a faithful irreducible representation π_c on the Hilbert space H_τ.

In the finite-dimensional case, it is trivial to prove that every q-set is open. This is the *discrete* q-topology. In the infinite-dimensional case, the discrete q-topology cannot occur, since it is clearly not compact. The simplest case is that given in Model 4 of Sec. 9. Here there are two coherent subspaces H_c and H_∞, which are countably infinite and 1-dimensional, respectively. Denote the identities in $L(H_c)$ and $L(H_\infty)$ by $\mathbf{1}_c$ and $\mathbf{1}_\infty$, respectively, so that $\mathbf{1} = \mathbf{1}_c + \mathbf{1}_\infty$. The open q-sets are of two types: (a) any projection $e \leq \mathbf{1}_c$ and (b) any projection of the form $e + \mathbf{1}_\infty$ where e is cofinite (i.e., $\mathbf{1}_c - e$ is finite). It is clear that we have here the quantum analog of the 1-point compactification of a countable set!

This example and Examples 1, 2, and 5 in the Appendix (for which the open q-sets have not been determined explicitly) all exemplify the following situation.

Although the whole q-space $\mathbf{1}$ is compact, the q space $\mathbf{1}_c$, endowed with the "subspace q-topology," is in all these examples a discrete q-space. This follows from the fact that $\pi_c(A)$ contains all compact operators. Secondly, from the fact that π_c is faithful, it follows at once that any element of T which vanishes on $\mathbf{1}_c$ vanishes everywhere. This implies that the *closure* of $\mathbf{1}_c$ (defined as the smallest closed q-set containing $\mathbf{1}_c$) is the whole q-space $\mathbf{1}$, i.e., $\mathbf{1}_c$ is *dense* in $\mathbf{1}$. Thus, $\mathbf{1}$ can be described as *a compactification* of the discrete countable q-space $\mathbf{1}_c$.

This is analogous to the situation in classical mechanics. There, as we saw in Secs. 5 and 6, the space X composed of the pure states is always a compactification of the conventional phase space Ω. The analogy is not complete, however, since Ω is not usually discrete. In the example discussed in Sec. 9, for instance, it can be regarded as the direct product of the circle and the reals. In this respect, Model 3 of the Appendix is of interest. Here, the q-topology of $\mathbf{1}_c$ is not discrete; in fact, no finite q-set is open.

APPENDIX

To illustrate the contents of Secs. 4–8, we now discuss the case of a particle that moves on a closed curve. Although we describe this curve as a circle, using a coordinate which runs from 0 to 2π, and even speak of rigid rotations around this circle, I believe that the final structures which we set up will reflect only the structure of the curve as a manifold: i.e., they will be independent of any smooth change of coordinate (for instance, from θ to $\theta + \frac{1}{2}\sin\theta$).

If a primitive observer (Sec. 4) were faced with the task of studying such a system, his work would fall into two stages. First, the task of determining experimentally the structure of the sets T_p^o and S_p^o of physical tests and states. Secondly, the purely theoretical matter of developing the mathematical theory on this foundation, as in Secs. 5–8.

As explained in Sec. 4, in discussing the first stage, we make full use of the conventional theory. However, once the physical states and tests have been chosen, no further reference to it will be made, the argument thereafter proceeding from the physical tests and states alone.

Let Γ denote the circle $0 \leq \theta \leq 2\pi$ on which the particle moves. The Hilbert space of the conventional quantum-mechanical theory is $H_c = L_2(\Gamma)$. Assuming the measure in Γ to be $d\mu/2\pi$, where $d\mu$ is Lebesgue measure, the functions $\psi_n = e^{in\theta}$, $n = 0, \pm 1, \pm 2, \cdots$, form an orthonormal basis in H_c, and any bounded Hermitian operator a in H_c may be represented by its matrix elements $a_{mn} = \langle \psi_m, a\psi_n \rangle$. We denote the identity operator in $L(H_c)$ by $\mathbf{1}_c$ and define the *momentum* p of the particle along its curved path by $p\psi_n = n\psi_n$.

We assume (cf. Sec. 4) that the conventional theory is correct insofar as every physical test is represented by some bounded Hermitian operator in $L(H_c)$. Our problem is to decide, on physical grounds, which operators arise in this way. We give several answers (described below as Models 1–4) to this question. These answers correspond to distinct assumptions concerning the class of experimental procedures available to our primitive observer. Naturally, this class is not uniquely determined simply by the statement that we are concerned with a particle moving on a circle. There are, however, certain conditions on the physical tests which must hold in any reasonable model of this situation. This gives us an upper limit to the set T_p^o of physical tests. In Model 1, we assume this biggest possible T_p^o.

At the other extreme, T_p^o cannot be void. In this connection, one may ask the following questions. Is it possible, in principle, for a primitive observer to be able to recognize this particular system, in the sense that he can distinguish it from other systems? [We have seen (Sec. 4) that this is not possible in the conventional theory.] And if it is, what is a minimal set of physical tests that suffices for this purpose? These questions provide the motivation for Model 2. Models 3–5 can be regarded as variations of Model 2.

In principle, the choice of the set S_p^o of physical states would be expected to be as important as that of T_p^o. We assume, indeed, that each physical state is represented by a positive Hermitian operator of trace 1 (as in the conventional theory). However, it is

not necessary to discuss in detail which operators arise in this way. We need only make the rather weak assumption that there are sufficiently many to determine the usual operator ordering in T_p: i.e., if $a \in T_p$ and a is not a positive Hermitian operator, then there is a physical state x with $(x, a) < 0$. (See also Footnote 21.)

Our first condition on the class T_p^o of physical tests is the analog of the requirement in the case of classical mechanics (see Sec. 4) that every physical test a is represented by a smooth function on phase space. To translate this requirement into quantum mechanics, we first re-express it as follows. Let R_ϕ denote the operation of rotation around the circle through an angle ϕ; thus, $R_\phi(a)$ is the test obtained by applying the method of analysis a after having rotated all the apparatus involved rigidly around the circle through the angle ϕ. Then our requirement can be expressed by asserting that, for every state x, $(x, R_\phi(a))$ is a smooth function of ϕ. (Observe that it would not be sufficient to make this assertion only for every *physical* state x.) Now, in the quantum-mechanical case, we obtain easily

$$(R_\psi(a))_{mn} = e^{i(n-m)}a_{mn}.$$

This gives

$$a'_{mn} = i(n - m)a_{mn},$$

where

$$a' = \left[\frac{dR_\phi(a)}{d\phi}\right]_{\phi=0}$$

could be called the *derivative of a with respect to the coordinate* θ. The condition for smoothness of a may now be expressed in the requirement that the Nth derivative $a^{(N)}$ should be bounded for every positive integer N. It can be shown[1] that a necessary and sufficient condition for this is that, for every integer $N \geq 0$, there exists a number K_N such that

$$|(m - n)^N a_{mn}| < K_N,$$

for all m and n. When this condition is satisfied, we shall say *a is smooth in θ*. If a is also a Hermitian operator with spectrum in $[0, 1]$, we shall call it a *smooth test*.

Our condition can now be expressed as follows: In any model of our system the physical tests must all be smooth tests.

Model 1

As a first attempt, let us assume also that every smooth test is a physical test. Then T_p is the real part Re (A_p) of the *algebra* A_p of all operators that are smooth in θ. Let A be the C^*-algebra which is the closure of A_p in $L(H_c)$. Then the completion T of T_p may be identified with Re (A) so that the system is a C^*-system.

To determine the structure of A is a nontrivial matter. Certainly $A \neq L(H_c)$, for the operator b given by $b_{mn} = \delta_{m,2n}$ is not in A. Nevertheless, A is large; indeed, it is *nonseparable*. This follows from the fact that $f(p) \in A$ for every bounded complex-valued function f on the integers.

The last remark suggests at the same time that our original assumption about the extent of the class T_p^o of physical tests is unrealistic. Surely there are too many uniform tests; how, for example, could the test "p is prime" be *uniformly* approximated by physical tests?

In the next model we assume a much smaller class of physical tests.

Model 2

For our second model let us attempt to keep the class T_p^o of all physical tests as small as possible, while still retaining enough tests to (perhaps) characterize the particular system that concerns us. There seem to be three types of test that we should admit.

(i) First, there should be tests that can be regarded as approximating observations of position on the circle. Accordingly, we assume that every smooth test a, which is a function $a = f(\theta)$ of θ, is a physical test. (It is easy to see that the condition of smoothness implies that f is smooth in the usual sense.) Of course, by a function of θ, we really mean a function on the circle Γ; if θ is treated as a real variable, then the only functions to be admitted are those which are periodic with period 2π. Any such function corresponds to a *multiplication operator* in $L(H_c)$, i.e., $\phi = a\psi$ means $\phi(\theta) = f(\theta)\psi(\theta)$. For simplicity, we henceforth denote the multiplication operator that corresponds to the function $f(\ )$ on Γ by f.

(ii) Similarly, there should be tests which correspond to observations of the momentum p. We take for this purpose those tests (they are all smooth) whose operators are functions of p. However, warned by Model 1, we do not wish to call every test of the form $a = F(p)$ a physical test: Although the question of smoothness of the function F does not arise (since the spectrum of p is discrete), the "behavior at infinity" is important. However, it seems reasonable to assume that, if F is constant outside some finite interval, then $F(p)$ is (or at least can be uniformly approximated by) a physical test. (In any case, we find that these tests are included in the third type, below.) We might wish to recognize as physical tests also certain other functions of p. For instance, in Model 5 below, we assume the existence of tests which

distinguish the directon of rotation even for arbitrarily large $|p|$.

(iii) With only these two types of physical test, it would be impossible to recognize states that are coherent superpositions of states of different p-values. We must therefore admit some other physical tests. Consideration of the example of light polarization and other cases where the Hilbert space is finite dimensional suggests, as a reasonable assumption, that if a Hermitian operator a with spectrum in [0, 1] is such that a_{mn} vanishes for all but a finite number of values of m and n, then a is (uniformly approximated by) a physical test.

It is easy to deduce from (iii) that the uniform closure T of T_p contains all compact Hermitian operators in $L(H_c)$. Similarly, from (i) it follows that T contains the multiplication operator corresponding to every continuous function on Γ; these operators form the real part of a commutative subalgebra of $L(H_c)$ that may be identified with $C(\Gamma)$. The smallest subalgebra of $L(H_c)$ that can contain T is thus the direct sum $A = C(\Gamma) + K$, where K is the algebra of all compact operators in $L(H_c)$. Indeed, if we assume T_p^o to be the smallest convex set containing the three classes of physical tests mentioned above, we can show that $T = \mathrm{Re}\,(A)$. Since A is a C^*-algebra, this shows that we again have a C^*-system.

We now pass to Stage 2 of the development, in which we regard A as an abstract C^*-algebra. The irreducible representations of A are as follows.[46] First, there is the conventional representation π_c on the Hilbert space H_c. Secondly, for each point θ in Γ, there is a representation π_θ, on a 1-dimensional Hilbert space H_θ, given by $\pi_\theta(f + k) = f(\theta)\mathbf{1}_\theta$, where $\mathbf{1}_\theta$ is the identity operator in $L(H_\theta)$. [We have written an arbitrary element of A in the form $f + k$, where $f \in C(\Gamma)$ and $k \in K$.] The "large" Hilbert space is thus $H = H_c \oplus H_\Gamma$, where H_Γ is the direct sum $\oplus_{\theta\in\Gamma} H_\theta$; and the reduced atomic representation π on H is the direct sum $\pi = \pi_c \oplus \pi_\Gamma$, where $\pi_\Gamma = \oplus_{\theta\in\Gamma} \pi_\theta$: i.e., $\pi_\Gamma(f + k) = f_\Gamma$, where f_Γ denotes the multiplication operator in $L(H_\Gamma)$ that corresponds to the continuous function $f(\ )$ on Γ.

As well as the conventional pure states, there is one pure state, represented by the projection $\mathbf{1}_\theta$ on H_θ, corresponding to each point θ in Γ. This state represents an ideal method of preparation in which the particle is localized exactly at the point θ. As well as the conventional pure tests, represented by projections in $L(H_c)$, there is a pure test in $L(H_\Gamma)$ corresponding to each subset of Γ. The general pure test is a sum of pure tests of these two types.

The projection $\mathbf{1}_c$ on the coherent subspace H_c may be described as the greatest conventional pure test [though it must not be confused with the pure test $\mathbf{1}$, which is the identity in $L(H)$]. It corresponds to the proposition "*the momentum is finite.*" Of course, $(x, \mathbf{1}_c) = 1$ for every conventional pure state x; for any reasonable choice of S_p^o this also holds for every physical state. The nonconventional states and tests arise as limits of conventional ones; for instance, $\mathbf{1}_0$ is the weak limit (see Sec. 6) of the sequence $\{y_n\}$, where y_n is the conventional pure state $y_n = |\phi_n\rangle\langle\phi_n|$ given by the wavefunction $\phi_n(\theta) = (2n)^{-\frac{1}{2}}$ for $|\theta| < n^{-1}$ and $\phi_n(\theta) = 0$ for $|\theta| > n^{-1}$.

The significance of the nonconventional states can also be explained in terms of "q-topology" as in Sec. 8.

Model 3

A difficulty with Model 2 appears if we attempt to discuss the time development of the system. Physically, this is most easily seen in the Schrödinger picture. The pure state $\mathbf{1}_\theta$ must develop, with the passage of time, into a nonconventional pure state in which the particle is spread all around the circle—but there is no such state. Mathematically, it is more convenient to use the Heisenberg picture. Corresponding to any time interval t there should be an automorphism T_t of the physical tests: The test a goes into the test $T_t(a)$ whose method of analysis is *wait a time t and then apply the test a.* Extended by linearity, the mapping T_t becomes an automorphism of the complex linear space A_p spanned by T_p^o; thus T_p^o and A_p should both be invariant under the linear transformation T_t. Now the trouble with Model 2 is that, if we use, for instance, the free nonrelativistic Hamiltonian p^2, this condition is not satisfied. Indeed, from elementary quantum mechanics, we obtain $T_t(a) = e^{ip^2t}ae^{-ip^2t}$. If $a = e^{in\theta}$, where n is an integer, then a is in A_p, but a short calculation gives $T_t(a) = e^{in\theta}e^{i2ntp}e^{in^2t}$ so that $T_t(a)$ is not in A_p.

We can avoid this difficulty with Model 2 in several ways. The first, which we call *Model 3*, is obtained as follows. As in the case of Model 2, we assume (a) that the physical tests form a subset T_p^o of $L(H_c)$ and denote by A_p the complex linear space spanned by T_p^o, (b) that T_p^o contains all the tests that are represented by smooth functions of θ, and (c) that T_p^o is invariant under the time development corresponding to the nonrelativistic Hamiltonian p^2. It then follows that, for every integer n and every number α in the interval $0 \leq \alpha \leq 2\pi$ (other values of α are redundant), A_p contains the operator $e^{in\theta}e^{i\alpha p}$. It is mathematically natural, though physically unjustified, to remove the exceptional case by enlarging A_p to contain also the functions of p of the form $e^{i\alpha p}$ with $0 \leq \alpha \leq 2\pi$. If

we do this, then we find, in view of the Weyl relations

$$e^{i\alpha p}e^{in\theta} = e^{in\theta}e^{i\alpha p}e^{i\alpha n} \tag{A1}$$

(where n is any integer and $0 \leq \alpha \leq 2\pi$), that A_p is a $*$-algebra. Its norm-closure A is thus a C^*-algebra, the smallest C^*-algebra generated by the continuous functions of θ together with the almost-periodic functions of p. The C^*-algebra A is naturally associated with the problem of representations of the canonical commutation relations[47] [in the modified form (A1) applicable to a particle moving on a closed curve]. We do not discuss the representations[48] and other properties of A here, except to remark that A has been shown[49] to be not a type I C^*-algebra; indeed, even the identity representation of A does not contain the compact operators.

One physically unsatisfactory feature of Model 3 is that the only functions of p recognized as physical tests are the linear combinations of periodic functions of p—such simple functions as the characteristic function of the point $p = 0$ are excluded. The situation can be remedied by admitting as physical the tests described in Model 2 as *type (iii)*. Doing this, we again obtain a C^*-system, the relevant C^*-algebra being the direct sum of the algebra A and the compact operators.

Model 4

The assumption in Model 3, that tests represented by operators that are smooth functions of θ are physical tests, can be criticized as follows. For simplicity, the argument is given in terms of classical mechanics.

Consider a method of analysis which realizes some physical test a. When this method of analysis is applied, there will be, among other things, some random error, of order of magnitude ϵ say, in the time at which the operation is initiated. Now consider a state c in which the particle is moving so fast that it makes many revolutions in the time ϵ. Clearly, the expectation (x, a) will then be practically independent of the (initial) position of the particle. Thus, as a function on phase space, the response function of the test a must tend to a constant as $p \to \pm\infty$. In particular, no test whose response function is a nonconstant function of θ alone can be physical.

More generally, one can give for the operation of time displacement an argument analogous to that given at the beginning of this section for rotation around the circle. This leads to the conclusion that, for any physical test a, $T_t(a)$ should be an infinitely differentiable function of t. If we now *add* this condition to those already imposed on the physical tests in Model 2, we obtain a smaller class of physical tests which again determines a C^*-system.

This time, the C^*-algebra A is relatively simple: It consists of all operators which differ by a compact operator from a multiple of the identity. (This C^*-algebra is discussed briefly by Kadison.[33]) There are now only two inequivalent irreducible representations: the identity representation and the 1-dimensional representation π_∞ given by $\pi_\infty(\alpha\mathbf{1}_c + k) = \alpha\mathbf{1}_\infty$, where $\mathbf{1}_\infty$ is the identity operator in the 1-dimensional Hilbert space H_∞ in which π_∞ acts. There is thus only one nonconventional pure state, namely, $\mathbf{1}_\infty$. This represents the fact that, with his reduced collection of tests, our primitive observer can no longer distinguish between the nonconventional pure states $\mathbf{1}_\theta$ of Model 2.

Model 5

The difficulty raised at the beginning of the discussion of Model 3 is considerably abated if we assume instead the relativistic Hamiltonian $(p^2 + 1)^{\frac{1}{2}}$. Since there is now an upper limit to the speed of the particle, the difficulties connected with $T_t(a)$ no longer arise. The situation regarding the time development of the state $\mathbf{1}_\theta$ is also improved. If the particle is initially localized exactly at the point $\theta = 0$, then $|p|$ is very large. So, after a time t we expect to find the particle either at $\theta = t$ and moving with velocity $+1$ or at $\theta = -t$ and moving with velocity -1. This suggests that the original localized state was not pure: It should still be possible to distinguish the direction of rotation.

This is borne out by the mathematical analysis.[1] We enlarge slightly the class T_p^o of Model 2, admitting the possibility of distinguishing the direction of motion of a particle even in the limit of arbitrarily large $|p|$. This leads again to a C^*-system. The algebra A now consists of all operators in $L(H_c)$ of the form $a = e_+fe_+ + e_-ge_- + k$, where f and g are multiplication operators in $C(\Gamma)$, k is a compact operator, and $e_+ = \mathbf{1}_c - e_-$ is the projection on the subspace corresponding to $p \geq 0$.

The reduced atomic representation is now slightly more complicated.[46] There are *two* 1-dimensional irreducible representations, and thus two pure states, for each point on the circle—one corresponding to each direction of rotation.

ACKNOWLEDGMENTS

This work was carried out at the University of Glasgow, Scotland, Queen's University, Kingston, Canada, and for a short period at Tulane University, New Orleans. I should like to thank all those who, through illuminating discussions, have assisted with

this work. I mention, in particular, Dr. W. K. Burton, Dr. G. Wyllie, Dr. K. H. Hofmann, and Dr. H. Kummer. Dr. Kummer, especially, has been of constant help during the final stages of the work.

During the last three years, this work was supported by grants from the National Research Council of Canada.

[1] R. Giles, unpublished manuscript.

[2] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).

[3] I. E. Segal, Ann. Math. (2) **48**, 930 (1947).

[4] R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).

[5] R. V. Kadison, Topology **3**, Suppl. (2), 177 (1965).

[6] R. J. Plymen, Commun. Math. Phys. **8**, 132 (1968).

[7] R. V. Kadison and J. R. Ringrose, Commun. Math. Phys. **4**, 32 (1967).

[8] For a nontrivial example of such a theory see R. Giles, *Mathematical Foundations of Thermodynamics* (Pergamon, Oxford, 1964).

[9] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1955).

[10] J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).

[11] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, N.J., 1968), Vol. 1; G. Birkhoff and J. von Neumann, Ann. Math. (2) **37**, 823 (1936).

[12] Observe that this is true even if $p = 0$ or 1. Of course, we are employing the frequency interpretation of probability. With other interpretations of probability (see Ref. 13) the statement (1) is certainly no nearer to being experimentally refutable.

[13] For a concise survey of the various interpretations of probability, see I. J. Good, *The Estimation of Probabilities* (M.I.T. Press, Cambridge, Mass., 1965).

[14] In general, many procedures will be acceptable in this respect. Not only may they differ in irrelevant respects, they may even have apparently nothing in common and yet be acceptable by the physicist as preparations of the same state. This is discussed further below.

[15] Naturally, the operations constituting a method of analysis need not be performed "at an instant"; they may occupy any finite time. The same applies to a method of preparation.

[16] To say that an assertion is *valid* does not, of course, refer in any way to absolute "truth"; it means merely that it belongs to the current theory.

[17] "c" stands for "conventional."

[18] Our notation is as follows. T_p, T, and T_w denote certain linear spaces connected with the tests. Subsets of a linear space are denoted by adding a superscript. Thus T_p^o is (i.e., will eventually be exhibited as) a subset of T_p. A similar notation is used in the case of states.

[19] R. J. Plymen, Helv. Phys. Acta **41**, 69 (1968).

[20] Here, we have made a rather mild assumption (see Footnote 21) about the extent of the class of all physical states.

[21] The assumptions on S_p^o (and even those on T_p^o) are not critical for the following considerations. Technically speaking, all that is required is that S_p^o should be a "full family of states" (see Ref. 5) for the algebra $C(\Omega)$.

[22] In the general theory (see Sec. 7), it is convenient to admit a slightly larger class of tests (called *weak tests*) than that described in this and the previous paragraph. The two classes differ in the case of classical mechanics whenever phase space Ω is not compact and in an analogous situation in the quantum-mechanical case.

[23] The ordering is the usual one, in which $A \geq 0$ means the spectrum of A is nonnegative.

[24] This corresponds to the fact that many different manifold structures can be defined on a given topological space Ω. For example, take for Ω the closed interval $0 \leq \lambda \leq 1$. Two distinct manifold structures are given by (a) the set of all smooth functions of λ, (b) the set of all smooth functions of λ^2.

[25] R. V. Kadison, Ann. Math. **56**, 494 (1952).

[26] R. V. Kadison, Mem. Am. Math. Soc., No. 7 (1951).

[27] L. Gillman and M. Jerison, *Rings of Continuous Functions* (Van Nostrand, Princeton, N.J., 1960).

[28] Here the tests and the weak tests coincide. However, in the general theory (see Sec. 7) the term "weak test" is given a precise meaning, while "test" is not.

[29] A partially ordered set is a *lattice* if every finite subset has a sup and an inf. It is *complete* if this is also true for every infinite subset. It is *atomic* if every element exceeds an atom, an *atom* being a minimal nonzero element. The lattice of all subsets of a set is a complete atomic Boolean algebra, and every such Boolean algebra is of this form.

[30] G. Birkhoff, *Lattice Theory* (revised edition: Am. Math. Soc. Colloq. Publ. XXV, New York, 1948).

[31] See, for instance, E. Hewitt and K. Stromberg, *Real and Abstract Analysis* (Springer-Verlag, Berlin, 1965).

[32] The notion of a *system* is a very important one and deserves a thorough discussion, but we do not attempt this in the present paper.

[33] For the definition and properties of C^*-algebras see J. Dixmier, *Les C^*-algebras et leurs representations* (Gauthier-Villars, Paris, 1965), or, for a brief account, Haag and Kastler, Ref. 4, I. Kaplansky [in *Surveys in Applied Mathematics* (Wiley, New York, 1958), Vol. 4], or R. V. Kadison, "Lectures on Operator Algebras," in *Applications of Mathematics to Problems in Theoretical Physics*, F. Lurçat, Ed. (Gordon and Breach, New York, 1967).

[34] I. E. Segal, Bull. Am. Math. Soc. **53**, 73 (1947).

[35] E. B. Davies, Commun. Math. Phys. **8**, 147 (1968).

[36] J. Glimm and R. V. Kadison, Pacific J. Math. **10**, 547 (1960).

[37] References 6 and 7.

[38] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960).

[39] I. Kaplansky, Ref. 33.

[40] R. V. Kadison, Ann. Math. **66**, 304 (1957).

[41] In many cases, A is uniquely determined by T, up to an isomorphism or anti-isomorphism (see Ref. 5).

[42] But only if A is not postliminaire. (See Dixmier, Ref. 33).

[43] The topology referred to is the weak topology in T_w regarded as the dual of S_{wf}; it is the ultraweak operator topology in T_w regarded as a subspace of the von Neumann algebra A_w.

[44] Not the closure of this complement—a sense that is often used in the literature.

[45] The proofs will be given in a forthcoming paper (R. Giles and H. Kummer, "A Noncommutative Generalization of Topology," submitted to Can. J. Math.). After (1)–(4) had been established, a preprint arrived from C. A. Akemann [see Notices Am. Math. Soc. **16**, 500 (1969)], in which these results and others including "regularity" were obtained—in a slightly different form. Akemann's work was of assistance to H. Kummer, to whom I am indebted for the proofs of (5) and (6).

[46] I am indebted to L. A. Coburn and to T. L. Gardiner for discussions and correspondence.

[47] See, for instance, I. E. Segal, "Representations of the Canonical Commutation Relations," in F. Lurçat, Ref. 33. See also G. W. Mackey, Duke Math. J. **16**, 313 (1949).

[48] Two inequivalent irreducible representations are mentioned by Segal, Ref. 47.

[49] R. A. Derrig, Notices Am. Math. Soc. **16**, 676 (1969). See also C. A. Berger and L. A. Coburn, Bull. Am. Math. Soc. **74**, 1008 (1968); **75**, 468 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Finite Integral Equations for Green's Functions for $:\phi^4:$ Coupling. I*

RODNEY W. JOHNSON
Physics Department, Syracuse University, Syracuse, New York 13210

(Received 1 May 1969; Revised Manuscript Received 12 September 1969)

A fully renormalized set of integral equations is derived for the Green's functions of a theory with $:\phi^4:$ coupling.

The purpose of this paper is to propose a set of renormalized integral equations for determining the Green's functions (specifically, the τ functions) for a field theory with $:\phi^4:$ coupling.

The equations can be iterated to yield a solution in the form of a formal power series in the coupling constant, each term being free of ultraviolet divergences. The relation between this series and the usual renormalized perturbation series for the theory will be discussed in a future publication.

Finding an exact, nonperturbative solution to the equations would be one possible route to the construction of a nontrivial model of a field theory. The problem of obtaining rigorous solutions is, however, a formidable one and will not be discussed here.

We give a formal derivation of the equations, without pretentions to rigor, but sufficient to provide the heuristic motivation for considering this particular set of equations. The equations themselves, in any event, mention only quantities which are, at least in perturbation theory, well defined. To give a rigorous sense to some of the intermediate steps of the derivation would require interpreting renormalization constants and products of field operators in terms of a limiting process, as Zimmermann,[1,2] for example, does. The derivation is in part patterned after the treatment of the $:\phi^3:$ coupling given by Symanzik.[3]

We start with the field equation

$$(\Box + m^2)\phi = -\lambda Z_v Z^{-1}:\phi^3: + \delta m^2\phi, \tag{1}$$

where Z and Z_v are the wavefunction and vertex renormalization constants, λ and m are the renormalized coupling constant and mass, the renormalized field ϕ obeys the commutation relation

$$[\phi(\mathbf{x}, t), \dot{\phi}(\mathbf{y}, t)] = iZ^{-1}\delta^{(3)}(\mathbf{x} - \mathbf{y}), \tag{2}$$

and where $:\phi(x)^3:$ is formally defined by setting

$$\begin{aligned}:\phi(x_1)\phi(x_2)\phi(x_3):&\\ &= T(\phi(x_1)\phi(x_2)\phi(x_3)) - \Delta'_F(x_1 - x_2)\phi(x_3)\\ &= \Delta'_F(x_1 - x_3)\phi(x_2) - \Delta'_F(x_2 - x_3)\phi(x_1),\end{aligned} \tag{3}$$

where

$$\Delta'_F(x - y) = \langle 0|\, T(\phi(x)\phi(y))\,|0\rangle.$$

From (2) it follows that

$$\begin{aligned}&(\Box_{x_0} + m^2)\langle 0|\, T(\phi(x_0)\cdots\phi(x_n))\,|0\rangle\\ &= \langle 0|\, T((\Box_{x_0} + m^2)\phi(x_0)\phi(x_1)\cdots\phi(x_n))\,|0\rangle\\ &\quad - iZ^{-1}\sum_{j=1}^{n}\delta^{(4)}(x_0 - x_j)\\ &\quad \times \langle 0|\, T(\phi(x_1)\cdots\phi(x_{j-1})\phi(x_{j+1})\cdots\phi(x_n))\,|0\rangle.\end{aligned} \tag{4}$$

Writing $\tau(x_0, \cdots, x_n)$ for $\langle 0|\, T(\phi(x_0)\cdots\phi(x_n))\,|0\rangle$ and using the field equation (1), we get

$$\begin{aligned}&(\Box_{x_0} + m^2)\tau(x_0, \cdots, x_n)\\ &= -\lambda Z_v Z^{-1}\tau(x_0, x_0, x_0, x_1, \cdots, x_n)\\ &\quad + 3\lambda Z_v Z^{-1}\Delta'_F(0)\tau(x_0, x_1, \cdots, x_n)\\ &\quad + \delta m^2\tau(x_0, x_1, \cdots, x_n)\\ &\quad - iZ^{-1}\sum_{j=1}^{n}\delta^{(4)}(x_0 - x_j)\tau(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n)\end{aligned} \tag{5}$$

or, equivalently,

$$\begin{aligned}&(\Box_{x_0} + m^2)\tau(x_0, \cdots, x_n)\\ &= -\lambda Z_v\tau(x_0, x_0, x_0, x_1, \cdots, x_n)\\ &\quad + 3\lambda Z_v\Delta'_F(0)\tau(x_0, x_1, \cdots, x_n)\\ &\quad + [Z\delta m^2 - (Z - 1)(\Box_{x_0} + m^2)]\tau(x_0, x_1, \cdots, x_n)\\ &\quad - i\sum_{j=1}^{n}\delta^{(4)}(x_0 - x_j)\tau(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n).\end{aligned} \tag{6}$$

We introduce the truncated τ functions by

$$\tau(x_0, \cdots, x_n) = \sum_P \tau^{\mathrm{T}}(P_1)\cdots\tau^{\mathrm{T}}(P_\mu), \tag{7}$$

where the sum runs over all (unordered) partitions

$$P = \{P_1, \cdots, P_\mu\}$$

of the variables $x_0, \cdots, x_n$ into any number $1 \leq \mu \leq n$ of sets

$$P_i = \{x_{i1}, \cdots, x_{iv(i)}\},$$

where $x_{11}, \cdots, x_{1v(1)}, \cdots, x_{\mu 1}, \cdots, x_{\mu v(\mu)}$ is a renumbering of $x_0, \cdots, x_n$. In terms of the τ^{T}, we can

express (5) as

$$
\begin{aligned}
&(\Box_{x_0} + m^2)\tau^{\mathrm{T}}(x_0, \cdots, x_n) \\
&= -\lambda Z_v \tau^{\mathrm{T}}(x_0, x_0, x_0, x_1, \cdots, x_n) \\
&\quad - 3\lambda Z_v \sum_{Q} \tau^{\mathrm{T}}(x_0, x_0, Q_1)\tau^{\mathrm{T}}(x_0, Q_2) \\
&\quad - 6\lambda Z_v \sum_{P} \tau^{\mathrm{T}}(x_0, P_1)\tau^{\mathrm{T}}(x_0, P_2)\tau^{\mathrm{T}}(x_0, P_3) \\
&\quad + 3\lambda Z_v \Delta'_F(0)\tau^{\mathrm{T}}(x_0, x_1, \cdots, x_n) \\
&\quad + [Z\delta m^2 - (Z-1)(\Box_{x_0} + m^2)]\tau^{\mathrm{T}}(x_0, x_1, \cdots, x_n) \\
&\quad - \begin{cases} i\delta^{(4)}(x_0 - x_1), & n = 1, \\ 0, & n \neq 1, \end{cases}
\end{aligned} \tag{8}
$$

where the first sum runs over *ordered* partitions

$$Q = (Q_1, Q_2)$$

of $x_1, \cdots, x_n$ into two unordered sets, while the second runs over partitions

$$P = \{P_1, P_2, P_3\}$$

of $x_1, \cdots, x_n$ into three sets. If Q_1 is empty, we get a term which cancels with the fourth term on the right-hand side of (8). If Q_2 or P_1, P_2, or P_3 is empty, the corresponding term fails to contribute, since τ^{T} with a single argument vanishes.

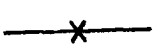
It will be convenient to use a graphical notation. We write

(9)

for Δ_F and

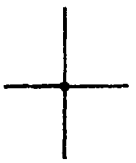
(10)

for Δ'_F, use a cross



(11)

to denote $iZ\delta m^2 - i(Z-1)(\Box + m^2)$, and let a vertex

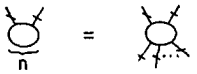


(12)

stand for $-6iZ_v\lambda$. An oval with n legs

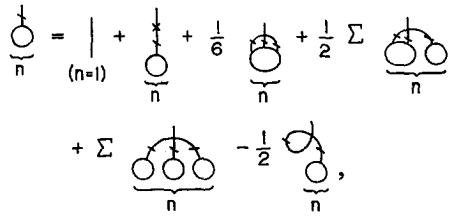
(13)

stands for τ^{T} with n arguments, and a brace is an abbreviation for a set of n legs; thus,



(14)

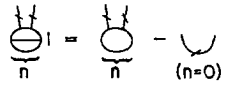
where the oval on the right has n legs below. The figures may, like Feynman diagrams, be interpreted in either momentum or configuration space. With these conventions we obtain the following from (8):



(15)

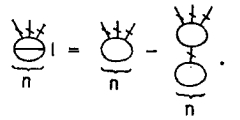
where the first sum is over ordered partitions of the n lower legs into two sets and the second sum is over partitions into three sets.

We define 1-particle irreducible parts by



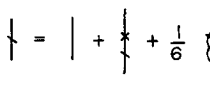
(16)

and



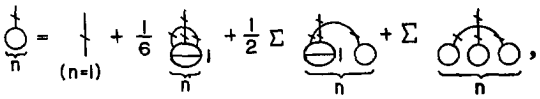
(17)

If one wishes to think of the τ^{T} functions as sums of connected Feynman diagrams, then (16) and (17) single out those diagrams which cannot be disconnected by one cut or less in the indicated channel. If $n = 1$, Eq. (15) is



(18)

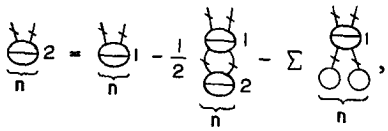
By use of (16)–(18), Eq. (15) becomes



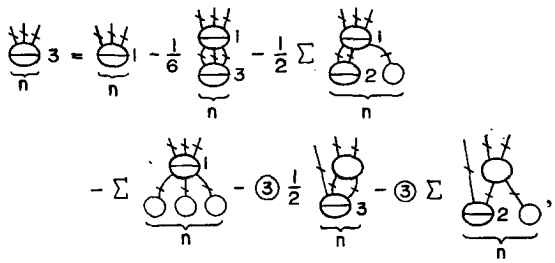
(19)

which reduces to an identity when $n = 1$, but is equivalent to (15) when supplemented by (16)–(18).

For 2- and 3-particle irreducible functions, we write

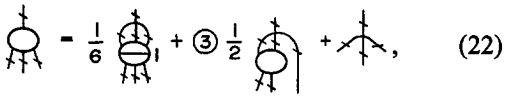


(20)



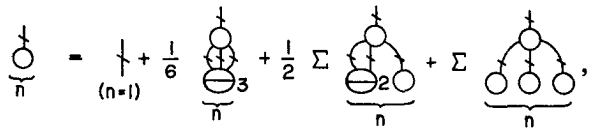
(21)

where ③ indicates three terms obtained from the one shown by permuting the three upper legs. If $n = 3$, Eq. (19) is



(22)

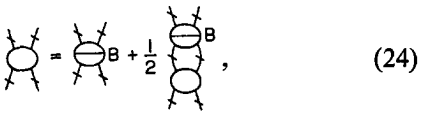
where ③ indicates three terms obtained by permuting the lower legs and we have used (16) with $n = 2$. By use of (20)–(22), Eq. (19) becomes



(23)

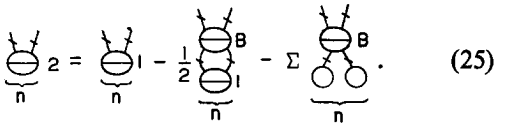
which reduces to an identity when $n < 3$, but is equivalent to (15) when supplemented by equations for $n = 1$ and $n = 3$ and by auxiliary definitions.

Although the Bethe–Salpeter kernel B is (at least in perturbation theory) logarithmically divergent, we temporarily introduce it formally with



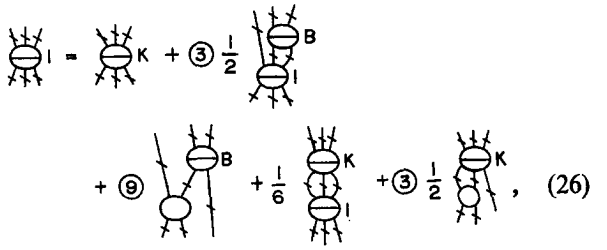
(24)

after which we can express the solution to (20) as



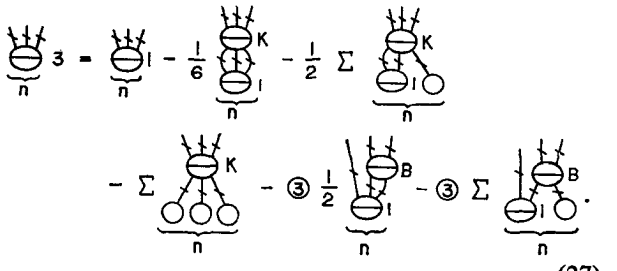
(25)

Further, we introduce the kernel K by



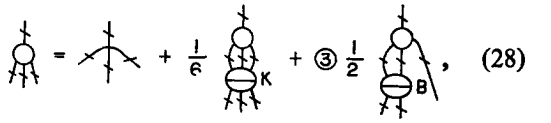
(26)

where the first ③ refers to permutations of upper legs, the second, of lower, and the ⑨, of both. We can then write the solution of (21) as



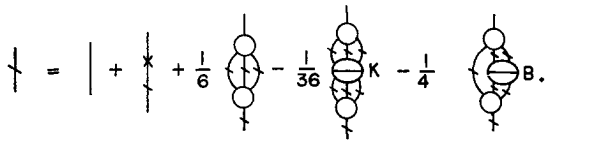
(27)

By the use of (24) and (26) and the assumption of symmetry of B and K between their upper and lower arguments, (22) becomes



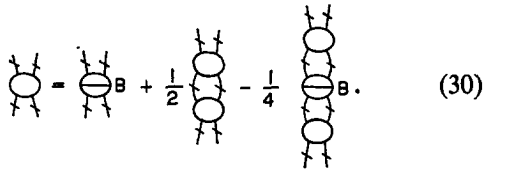
(28)

and then (18) becomes



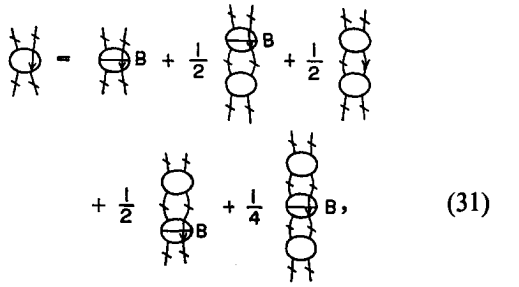
(29)

We now eliminate B in favor of its momentum-space derivative, which (if perturbation theory is a reliable guide) may be expected to exist even though B itself does not. From (24) we get



(30)

Differentiation then gives

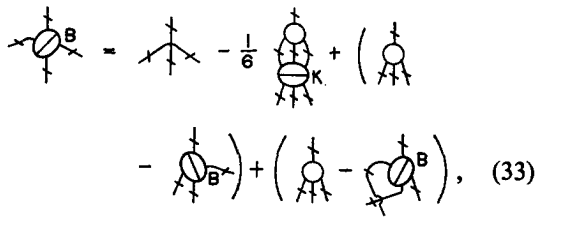


(31)

where an arrow connecting two legs of an oval denotes partial differentiation of the corresponding quantity with respect to the momentum flowing between the two indicated legs and

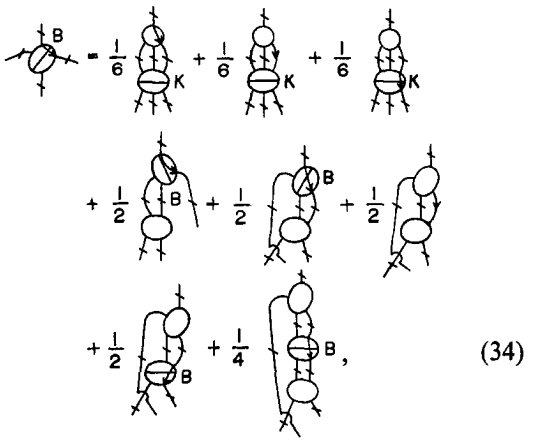
(32)

is the momentum-space derivative of Δ'_F. From (28), by the use of (24), we get



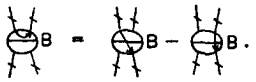
(33)

whereupon differentiation gives



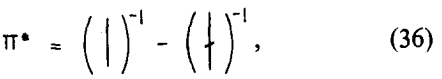
(34)

where we may set



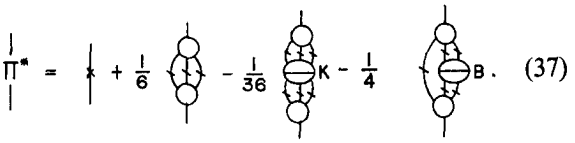
(35)

Defining the proper self-energy part by



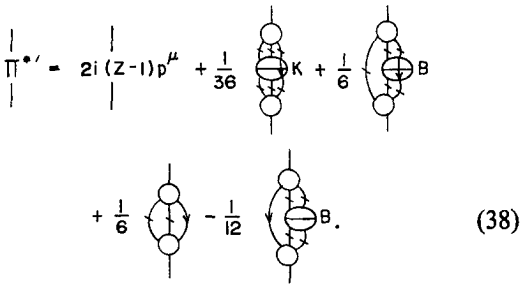
(36)

we find from (29) that



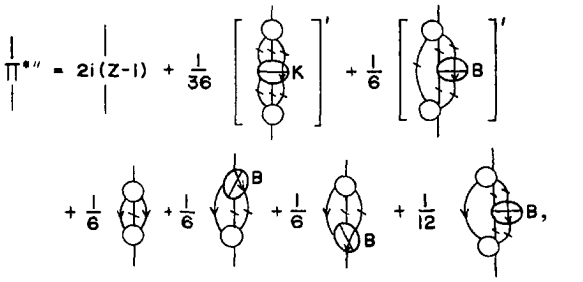
(37)

Differentiating once, we get



(38)

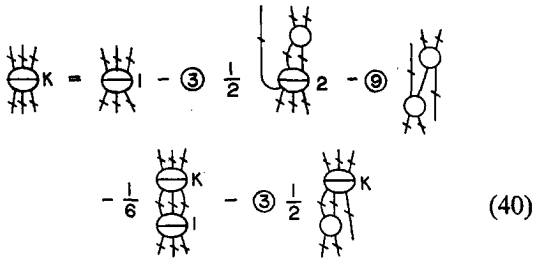
Using (30) to rewrite the last two terms and differentiating again, we find, with the help of (31),



(39)

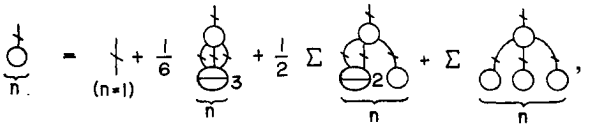
where [ ]′ indicates the momentum-space derivative of the enclosed quantity. In (39) the overlapping divergences have been untangled, and only one more differentiation is required to remove the over-all divergence.

It is still necessary to get rid of the undifferentiated B in (26), the equation for K. We transform the second and third terms on the right in (26) by using (17) with $n = 3$, (16) with $n = 4$, (25) with $n = 4$, (24), and (20) with $n = 4$, and we obtain

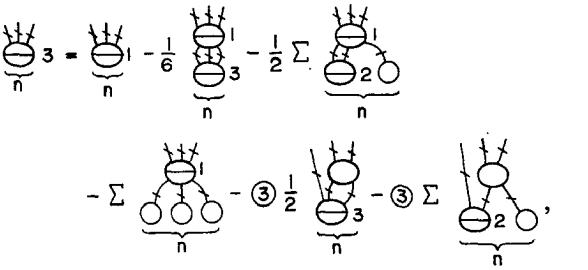


(40)

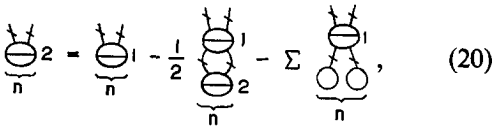
We collect here the final proposed set of divergence-free equations:



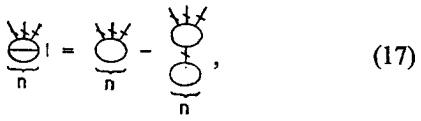
(23)



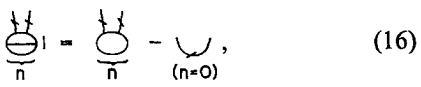
(21)



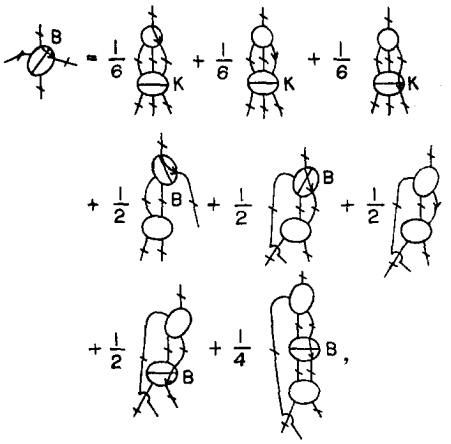
(20)



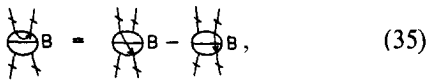
(17)



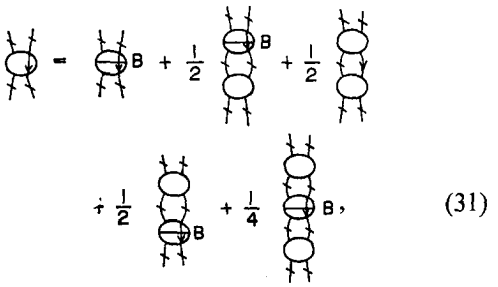
(16)



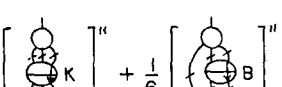
(34)



(35)

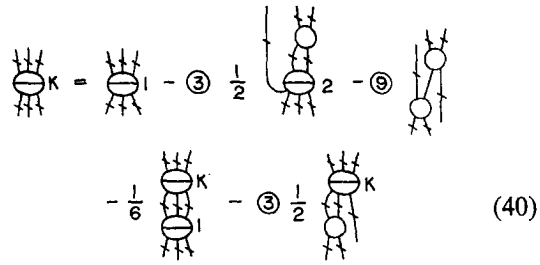


(31)



(36′)

(39′)



(40)

These must be supplemented by conditions which allow the vertex function and Π^* to be determined by (31) and (39′). We require that the vertex function (that is, τ^T with four arguments and Δ_F' amputated for each argument) reduce in momentum space to $(-6i\lambda)[\delta^{(4)}(p_1+p_2+p_3+p_4)/(2\pi)^4]$ at the point ($p_1=p_2=p_3=p_4=0$, for instance) chosen for the definition of the coupling constant. We further require that $\Pi^*/(p^2-m^2)$ vanish at the point $p^2=m^2$.

Some of the equations derived here are contained in the work of Taylor,[4] who has also investigated the possibility of finding exact solutions to similar sets of equations. Wu[5] has described a prescription for differentiating Feynman graphs for $:\phi^4:$ coupling which gives results comparable to the perturbation expansion obtained by iterating the present set of equations. The full set of equations has been obtained independently by Symanzik.[6] A related set was studied in the author's thesis.[7]

ACKNOWLEDGMENTS

Although I do not know the exact date of Dr. K. Symanzik's derivation of these equations, it is probable that it somewhat preceded my own. In preparing this paper, I have had the benefit of a number of remarks which he was kind enough to make when I showed him a preliminary version and which have in particular made it possible to reduce by one the list of equations required. I extend my warmest thanks to Dr. Symanzik for offering those remarks. I am grateful to Dr. S. D. Drell, Dr. J. D. Bjorken, and Dr. W. A. Strauss for advice and guidance in the preparation of Ref. 7. I thank Dr. A. S. Wightman for helpful discussions.

* Work supported in part by the U.S. Air Force (Contract 2456013) and in part by the National Science Foundation.

[1] W. Zimmermann, Commun. Math. Phys. **6**, 161 (1967), and earlier references therein.
[2] W. Zimmermann, Commun. Math. Phys. **10**, 325 (1968).
[3] K. Symanzik, in *Lectures on High Energy Physics*, B. Jakšić, Ed. (Gordon and Breach, New York, 1966).
[4] J. G. Taylor, Nuovo Cimento Suppl. **1**, 859 (1963).
[5] T. T. Wu, Phys. Rev. **125**, 1436 (1962).
[6] K. Symanzik, private communication.
[7] R. Johnson, thesis, Stanford University, 1967.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Photoelectron Shot Noise

B. PICINBONO AND C. BENDJABALLAH
Institut d'Electronique Fondamentale, Faculté des Sciences, 91 Orsay, France*
AND
J. POUGET
Institut d'Etudes Nucléaires, Algiers, Algeria

(Received 26 June 1969; Revised Manuscript Received 8 October 1969)

The instants of time emission of photoelectrons generated by a detector immersed in an optical field constitute a point compound Poisson process. A complete definition of such a process is introduced to calculate some average values of the distribution. The shot noise due to this point process is also considered and we study the difference between the "deterministic" and the "random" shot noises. They are completely defined by the set of their characteristic functions. We consider also the asymptotic properties of the shot noise and we show that for large mean density of the point process the fluctuations are not described by a Gaussian, but by a Gaussian compound random function. Thus the central limit theorem is not strictly valid. An experimental setup to obtain these fluctuations is described and some statistical properties of the asymptotic shot noise are presented.

1. INTRODUCTION

Several papers have been recently devoted to the study of statistical properties of optical fields. Particularly, the time instants $\{t_i\}$ at which a photon is absorbed by a detector in an electromagnetic field or a photoelectron is emitted constitute a point stochastic process, whose many properties are now clarified. This process can be defined by using either the classical representation of fields[1–3] or the quantum one.[4,5] The relations between the classical and quantum descriptions are now well established.[6–8] We also have many experimental results which confirm the theoretical calculations, particularly on photon counting[9–13] and photon coincidences.[14–17] One of the main results of those studies is that the time instants $\{t_i\}$ define not a simple Poisson process, but a compound Poisson process, for which there are some correlation properties which appear, for example, in bunching effects.

In order to describe a point process, it is convenient to introduce the random function $N(t)$, number of random points $\{t_i\}$ between an arbitrary origin of time and the time instant t. For Poisson processes, $N(t)$ has independent increments $dN(\theta)$, and is almost surely a discontinuous function.[18] The statistical properties of the random increments $dN(\theta)$, or number of points $\{t_i\}$ between θ and $\theta + d\theta$, can be used to specify the probability distribution of the process $N(t)$. We will study this point in the case of a compound Poisson process. Particularly, we will compute expectation values, such as $E[dN(\theta_1) \cdots dN(\theta_n)]$, and their Fourier transforms, which are used in many applications, and particularly for the photoelectron shot noise.

Usually shot noise appears when a point process passes through a linear time invariant system with impulse function (Rt). Therefore, the shot noise is described by the random function

$$X(t) = \int_{-\infty}^{+\infty} R(t - \theta)\, dN(\theta). \tag{1.1}$$

Nevertheless, for optical detectors this expression is not always convenient. In fact, the output to one photoelectron at the time instant θ is generally not a deterministic function $R(t - \theta)$, but a random impulsion $R(t - \theta, \omega)$. This is particularly true in the case of the photomultiplier (PM): For very low light intensity it is well known that at random times we observe pulses which are also random in shape. This fact is due to the amplification by secondary emission which is essentially a random process of amplification. Thus it is necessary to introduce the concepts of "deterministic" shot noise, described by Eq. (1.1), and "random" shot noise in which R is a function of a point ω in some probability space. In the two cases, we will define completely the statistical properties of these noises by the set of the characteristic functions of the random variable $X(t_1) \cdots X(t_n)$ for arbitrary t_i and n.

With the complete definition of the shot noise $X(t)$, we can now explore its asymptotic properties. This problem is well known and very important in classical theory of shot noise, i.e., in the case where the $dN(\theta)$ are the increments of a stationary pure Poisson process with density ρ, and $R(t)$ a deterministic function. The result is that, for large ρ, $X(t)$ becomes a Gaussian random function, which is a particular case of the central limit theorem. In practice this result is very significant because ρ is often large, and therefore the shot noise Gaussian. Therefore it is important to

explore the same problem for photoelectrons, i.e., for a Poisson compound process.

At first we will show that, as for a pure Poisson process, the *fluctuations* of the deterministic or random shot noises become very small compared to the mean value. This means that for high light intensity the instantaneous output of the detector is a good estimation of this intensity. Nevertheless, if we study these fluctuations we obtain that, for large mean density of the point process, they are no longer described by a Gaussian stochastic process, but by a new process defined by its characteristic function and called compound Gaussian stochastic process. Therefore, the central limit theorem is no longer valid for photoelectrons. But the question arises if it is possible to separate the fluctuations from the mean value of the shot noise, because for large ρ the fluctuations are very small compared to the mean value. For this purpose, we describe an experimental method which allows this observation by using two optical detectors in the same optical field, as in the Hanbury Brown and Twiss experiment, and an adapted signal processing of the outputs of the detectors.

To explain this experiment, some properties of Gaussian compound random functions are reviewed. Particularly, we calculate some probability distributions of the asymptotic shot noise fluctuations for optical fields, obtained by superposition of thermal light and ideal laser light.

2. DEFINITION AND SOME PROPERTIES OF POISSON COMPOUND POINT PROCESS

Let us call $\{t_i\}$ the time instants at which photoelectrons are emitted by a detector in an optical field. The instants $\{t_i\}$ constitute a random point process.[19] The first derivation of some properties of such process by using a quantum theory of electromagnetic measurement was given by Kelley and Kleiner.[5] We will present more general results, but without discussing the microscopy physical origin of the process.

This point process is represented by the random function $N(t) = \int_0^t dN(\theta)$, where $dN(\theta)$ are the random increments of $N(t)$. The probability distribution of $N(t)$ can be defined if, for every n and every set of $\theta_1, \theta_2, \cdots, \theta_n$, we know the probability distribution of the n-dimensional random variable $dN(\theta_1), \cdots, dN(\theta_n)$.

We characterize a Poisson compound point process by the following properties:

(i) $P_r\,[dN(\theta) > 1] = o(d\theta)$, which means that the random variable $dN(\theta)$ has only two values, 0 or 1. As a consequence we have $dN^k(\theta) = dN(\theta)$.

(ii) For different θ_i,

$$P_r\{[dN(\theta_1) = 1] \cdot [dN(\theta_2) = 1] \cdot \cdots \cdot [dN(\theta_n) = 1]\} = E[\rho(\theta_1)\rho(\theta_2) \cdots \rho(\theta_n)]\, d\theta_1\, d\theta_2 \cdots d\theta_n, \quad (2.1)$$

where $\rho(\theta)$ is a stationary, nonnegative random function. Evidently, if $\rho(\theta)$ is a nonrandom constant, Eq. (2.1) defines a stationary pure Poisson process; if now $\rho(\theta)$ is a nonrandom function of time, we have a nonstationary pure Poisson process.[20] We obtain a compound (or *a posteriori*) process, because, for a given trial of the random function $\rho(\theta)$, the *a posteriori* distribution of $dN(\theta)$ is given by

$$P_r\{[dN(\theta_1) = 1] \cdot \cdots \cdot [dN(\theta_n) = 1] \mid \rho\} = \rho(\theta_1) \cdots \rho(\theta_n)\, d\theta_1 \cdots d\theta_n, \quad (2.2)$$

which is the definition of a Poisson nonstationary process. Obviously *a priori* distribution given by Eq. (2.1) is obtained by averaging *a posteriori* distribution given by Eq. (2.2).

In the case of optical fields, it is well established that $\rho(\theta)$ is proportional to the instantaneous light intensity.[1,5] For the following discussion, we will write

$$\rho(\theta) = \alpha F(\theta), \quad (2.3)$$

where α is a nonrandom parameter and $F(\theta)$ a nonnegative random function. By varying α, we describe the variations of the mean light intensity which can be obtained by various means and for the asymptotic problem we will study the case where $\alpha \to \infty$, which appears for fields with very large intensities.

For many problems, as for example shot noise, it is important to have expression of the moments of $dN(\theta)$, and we will now establish some properties of such moments.

A. First-Order Moment

The random variable $dN(\theta)$ has only two values, 0 and 1. Therefore,

$$E[dN(\theta)] = P_r\,[dN(\theta) = 1] = \alpha E[F(\theta)]\, d\theta. \quad (2.4)$$

If $F(t)$ is stationary, we can introduce the mean density ρ of the process and write

$$E[dN(\theta)] = \rho\, d\theta, \quad (2.5)$$

where ρ is evidently $E[\rho(\theta)] = \alpha E[F]$.

B. Second-Order Moment

The notation of the second-order moment

$$E[dN(\theta_1)\, dN(\theta_2)]$$

has the meaning of a distribution on the space $\mathbf{R}^2(\theta_1 \otimes \theta_2)$. This distribution can be decomposed in two parts. First, *if* $\theta_1 \neq \theta_2$, we obtain, by using Eq. (2.1)

and the same method as for the first-order moment,

$$E[dN(\theta_1)\,dN(\theta_2)]_1 = \alpha^2 E[F(\theta_1)F(\theta_2)]\,d\theta_1\,d\theta_2. \quad (2.6)$$

Moreover, *if* $\theta_1 = \theta_2$, we have

$$E[dN^2(\theta_1)]_2 = E[dN(\theta_1)] = \alpha E[F(\theta_1)]\,d\theta_1, \quad (2.7)$$

which is proportional to $d\theta_1$ and therefore is a distribution on the curve $\theta_1 = \theta_2$ of the space $\mathbf{R}^2$. Therefore, we can write the complete expression of the moment

$$E[dN(\theta_1)\,dN(\theta_2)] = \{\alpha^2 E[F(\theta_1)F(\theta_2)] + \alpha E[F(\theta_1)]\delta(\theta_1 - \theta_2)\}\,d\theta_1\,d\theta_2, \quad (2.8)$$

where δ is the Dirac distribution.

With this notation,

$$\int\int_{\Delta\theta_1 \otimes \Delta\theta_2} E[dN(\theta_1)\,dN(\theta_2)],$$

where $\Delta\theta_1$ and $\Delta\theta_2$ are arbitrary finite time intervals, is finite.

Now, if $F(\theta)$ is a wide sense stationary random function, this moment depends only on $\theta_1 - \theta_2$, and is

$$E[dN(\theta_1)\,dN(\theta_2)] = \{\alpha^2 \Gamma_F(\theta_1 - \theta_2) + \rho\delta(\theta_1 - \theta_2)\}\,d\theta_1\,d\theta_2, \quad (2.9)$$

where $\Gamma_F(\tau)$ is the correlation function of $F(t)$. In this equation, the first term describes the Hanbury Brown, and Twiss effect, or bunching effect of photoelectrons, and the second is the contribution of a stationary Poisson process with density ρ.

For the following, it is convenient to introduce the random function $\tilde{F}(\theta)$ defined by

$$\tilde{F}(\theta) = F(\theta) - E[F], \quad (2.10)$$

which is clearly 0, if $F(\theta)$ is nonrandom (Poisson process). Thus, using $\alpha E[F] = \rho$, we can introduce the second-order density $M(\theta_1, \theta_2)$, defined by

$$E[dN(\theta_1)\,dN(\theta_2)] = M(\theta_1, \theta_2)\,d\theta_1\,d\theta_2, \quad (2.11)$$

and from Eqs. (2.9) and (2.10) we obtain

$$M(\theta_1, \theta_2) = \rho^2 + \alpha^2 \tilde{\Gamma}_F(\theta_1 - \theta_2) + \rho\delta(\theta_1 - \theta_2), \quad (2.12)$$

where $\tilde{\Gamma}_F(\tau)$ is the correlation function of $\tilde{F}(t)$.

The Fourier transform of $M(\theta_1, \theta_2)$ plays an important role in the following discussion. As we suppose that $F(t)$ is stationary, this Fourier transform can be written

$$\gamma(\nu_1, \nu_2) = \rho^2\delta(\nu_1)\delta(\nu_2) + \delta(\nu_1 + \nu_2)[\alpha^2\tilde{\gamma}_F(\nu_1) + \rho], \quad (2.13)$$

where $\tilde{\gamma}_F(\nu)$ is the spectrum of $\tilde{F}(t)$, or the Fourier transform of $\tilde{\Gamma}_F(\tau)$. Obviously, the term $\delta(\nu_1)\delta(\nu_2)$ means that $dN(t)$ has a nonzero mean value, and $\delta(\nu_1 + \nu_2)$ that it is stationary.

C. Third-Order Moment

By a simple extension of the preceding calculations, we obtain

$$E[dN(\theta_1)\,dN(\theta_2)\,dN(\theta_3)] = \Big\{\alpha^3 E[F(\theta_1)F(\theta_2)F(\theta_3)] + \alpha^2 \sum_{[i,j]} \delta(\theta_i - \theta_j)E[F(\theta_i)F(\theta_k)] + \rho\delta(\theta_1 - \theta_2)\delta(\theta_1 - \theta_3)\Big\}\,d\theta_1\,d\theta_2\,d\theta_3. \quad (2.14)$$

In the sum $\sum_{[i,j]}$, we have three terms corresponding to the three combinations (θ_1, θ_2), (θ_1, θ_3), and (θ_2, θ_3).

D. Fourth-Order Moment

For a general discussion, it is necessary to write explicitly this moment which has the structure of higher-order moments, but is considerably simpler to write down. By extension of previous calculations, this moment can be written

$$E[dN(\theta_1)\cdots dN(\theta_4)] = M(\theta_1, \cdots, \theta_4)\,d\theta_1\cdots d\theta_4, \quad (2.15)$$

where

$$\begin{aligned} M(\theta_1, \cdots, \theta_4) &= \alpha^4 E[F(\theta_1)\cdots F(\theta_4)] \\ &+ \alpha^3 \sum_{\substack{[i,j] \\ \binom{4}{2}}} \delta(\theta_i - \theta_j)E[F(\theta_i)F(\theta_k)F(\theta_l)] \\ &+ \alpha^2 \sum_{\substack{[i,j,k] \\ \binom{4}{3}}} \delta(\theta_i - \theta_j)\delta(\theta_i - \theta_k)E[F(\theta_i)F(\theta_l)] \\ &+ \alpha^2 \sum_{\substack{i,j,k,l \\ 3!!}} \delta(\theta_i - \theta_j)\delta(\theta_k - \theta_l)E[F(\theta_i)F(\theta_k)] \\ &+ \rho\delta(\theta_1 - \theta_2)\delta(\theta_1 - \theta_3)\delta(\theta_1 - \theta_4). \end{aligned} \quad (2.16)$$

For each sum we have indicated the number of terms. The term $\sum_{[i,j,k,l]}$ is particularly interesting. Indeed, it is a sum of $3!! = 3$ terms, which are obtained by considering in all the possible permutations of the θ_i, only the Gaussian ones,[21] i.e., $(\theta_1, \theta_2)(\theta_3, \theta_4)$, $(\theta_1, \theta_3)(\theta_2, \theta_4)$, and $(\theta_1, \theta_4)(\theta_2, \theta_3)$. This point is important for the following discussion.

As for the second-order moment, it is necessary to write Eq. (2.16) with the function $\tilde{F}(\theta)$. After some

calculations, we obtain

$$
\begin{aligned}
&M(\theta_1, \cdots, \theta_4) \\
&= \rho^4 + \rho^2\alpha^2 \sum_{\substack{[i,j] \\ \binom{4}{2}}} E[\tilde{F}(\theta_i)\tilde{F}(\theta_j)] \\
&\quad + \rho\alpha^3 \sum_{\substack{[i] \\ \binom{4}{1}}} E[\tilde{F}(\theta_j)\tilde{F}(\theta_k)\tilde{F}(\theta_l)] \\
&\quad + \alpha^4 E[\tilde{F}(\theta_1)\tilde{F}(\theta_2)\tilde{F}(\theta_3)\tilde{F}(\theta_4)] \\
&\quad + \sum_{\substack{[i,j] \\ \binom{4}{2}}} \delta(\theta_i - \theta_j) \Bigg\{ \rho^3 + \rho\alpha^2 \sum_{\substack{[i,k,l] \\ \binom{3}{2}}} E[\tilde{F}(\theta_i)\tilde{F}(\theta_k)] \\
&\quad + \alpha^3 E[\tilde{F}(\theta_i)\tilde{F}(\theta_k)\tilde{F}(\theta_l)] \Bigg\} \\
&\quad + \sum_{\substack{[i,j,k] \\ \binom{4}{3}}} \delta(\theta_i - \theta_j)\delta(\theta_i - \theta_k)\{\rho^2 + \alpha^2 E[\tilde{F}(\theta_i)\tilde{F}(\theta_l)]\} \\
&\quad + \sum_{\substack{[i,j,k,l] \\ 3!!}} \delta(\theta_i - \theta_j)\delta(\theta_k - \theta_l)\{\rho^2 + \alpha^2 E[\tilde{F}(\theta_i)\tilde{F}(\theta_k)]\} \\
&\quad + \rho\delta(\theta_1 - \theta_2)\delta(\theta_1 - \theta_3)\delta(\theta_1 - \theta_4). \qquad (2.17)
\end{aligned}
$$

We will not write explicitly the Fourier transform of the function $M(\theta_1, \cdots, \theta_4)$, which is obviously not very simple. For our discussion it is only necessary to extract from this Fourier transform the part which is distributed on the "Gaussian manifolds" of the space $\mathbf{R}^4$, $(\nu_1 \otimes \nu_2 \otimes \nu_3 \otimes \nu_4)$.[22] These manifolds are defined by the equations

$$\nu_i + \nu_j = 0, \quad \nu_k + \nu_l = 0 \qquad (2.18)$$

and in $\mathbf{R}^4$ there are evidently $3!! = 3$ different Gaussian manifolds corresponding to the 3 Gaussian permutations of (i, j, k, l).

Let us, for instance, consider the Gaussian manifold defined by

$$\nu_1 + \nu_2 = 0, \quad \nu_3 + \nu_4 = 0. \qquad (2.19)$$

The distribution $g(\nu_1, \nu_3)$ on this manifold is the coefficient of $\delta(\nu_1 + \nu_2)\delta(\nu_3 + \nu_4)$ in the Fourier transform of $M(\theta_1, \cdots, \theta_4)$. By inspection of all the terms of Eq. (2.17), we find

$$
\begin{aligned}
g(\nu_1, \nu_3) &= \alpha^4 h(\nu_1, \nu_3) \\
&\quad + \rho\alpha^2[\tilde{\gamma}_F(\nu_1) + \tilde{\gamma}_F(\nu_3)] + \rho^2, \qquad (2.20)
\end{aligned}
$$

where $h(\nu_1, \nu_3)$ is the contribution on this manifold of the Fourier transform of $E[\tilde{F}(\theta_1) \cdots \tilde{F}(\theta_4)]$. This function $h(\nu_1, \nu_3)$ is called "Gaussian density" [21] if it has the particular structure

$$h(\nu_1, \nu_3) = \tilde{\gamma}_F(\nu_1)\tilde{\gamma}_F(\nu_3), \qquad (2.21)$$

i.e., a product of power spectra of $F(t)$. If Eq. (2.21) holds, we obtain

$$g(\nu_1, \nu_3) = [\rho + \alpha\tilde{\gamma}_F(\nu_1)][\rho + \alpha\tilde{\gamma}_F(\nu_3)], \qquad (2.22)$$

and, from Eq. (2.13) and the definition (2.21), we see that $g(\nu_1, \nu_3)$ is also a "Gaussian density." Evidently, if $h(\nu_1, \nu_3)$ is not a "Gaussian density," i.e., if Eq. (2.21) does not hold, $g(\nu_1, \nu_3)$ cannot be a "Gaussian density." The same results can be easily found for the two other Gaussian manifolds of the space $\nu_1 \otimes \cdots \otimes \nu_4$, and by longer calculations extended for the higher-order moments of $dN(\theta)$.[23]

As an example, we can study the form of the function $h(\nu_i, \nu_j)$ for the thermal light, i.e., when the optical field is a Gaussian quasimonochromatic field. In this case, the function $F(\theta)$ [defined by Eq. (2.3)] can be written as

$$F(\theta) = Z(\theta)Z^*(\theta), \qquad (2.23)$$

where $Z(\theta)$ is the analytic signal of a zero mean, Gaussian, and quasimonochromatic real rf. Therefore, we have

$$
\begin{aligned}
&E[F(\theta_1) \cdots F(\theta_4)] \\
&\quad = E[Z(t_1) \cdots Z(t_4)Z^*(t_1) \cdots Z^*(t_4)], \qquad (2.24)
\end{aligned}
$$

which can be expressed only with

$$\Gamma_Z(\tau) = E[Z(t)Z^*(t - \tau)],$$

by classical expressions for the Gaussian case. From this expression we can prove that the distribution of the Fourier transform on the Gaussian manifold $(\nu_1 + \nu_2 = 0, \nu_3 + \nu_4 = 0)$ is given by Eq. (2.21) in which $\tilde{\gamma}_F(\nu)$ is the Fourier transform of $\tilde{\Gamma}_F(\tau)$, which for a Gaussian field is equal to $|\Gamma_Z(\tau)|^2$. Therefore, for the non-Gaussian random function $F(t)$, we have a "Gaussian density" on the Gaussian manifolds, and this result is true for higher-order moments.

The results of this section have a direct application to the deterministic shot noise described by Eq. (1.1). For instance, the general moment of $X(t)$ can be written

$$
\begin{aligned}
E[X(t_1) \cdots X(t_n)] &= \int_{-\infty}^{+\infty} \cdots \int^{n} R(t_1 - \theta_1) \cdots R(t_n - \theta_n) \\
&\quad \times E[dN(\theta_1) \cdots dN(\theta_n)]. \qquad (2.25)
\end{aligned}
$$

Particularly by using Eq. (2.5), we obtain

$$E[X(t)] = \rho G(0), \qquad (2.26)$$

where $G(\nu)$ is the transfer function, Fourier transform of $R(t)$. Similarly, we obtain from Eq. (2.10) the power spectrum of $X(t)$ by

$$
\begin{aligned}
\gamma_X(\nu) &= G^2(0)\rho^2\delta(\nu) + \rho\,|G(\nu)|^2 \\
&\quad + \alpha^2\tilde{\gamma}_F(\nu)\,|G(\nu)|^2. \qquad (2.27)
\end{aligned}
$$

In this expression the first term is due to the nonzero mean value of $X(t)$, the second describes the shot noise of a Poisson process with density ρ, and the third, owing to the fluctuations of $F(t)$, describes the Hanbury Brown and Twiss effect. A similar expression, obtained by a different method, was already presented by Mandel.[24]

Moreover, the structure of the higher-order moments of $dN(\theta)$ can be used for the determination of some asymptotic properties of $X(t)$.[21] For a given mean density ρ of the point process, let us consider the limit of $X(t)$ when the time constant T of the linear system becomes very large ($\rho T \gg 1$). The asymptotic properties of $X(t)$, after such very narrow band filtering, depend only on the structure of the Fourier transform of the higher-order moments of $dN(\theta)$ in the space $\nu_1 \otimes \nu_2 \otimes \cdots \otimes \nu_n$. If we have a "Gaussian density" on the "Gaussian manifolds" of this space, the asymptotic shot noise is described by a Gaussian stochastic process. That is the case for a stationary Poisson process,[21] and for a compound Poisson process we have seen that the result depends on the properties of the random function $F(t)$ which describes the light intensity. Nevertheless, we have shown that for thermal light the result is still true and therefore the asymptotic shot noise is Gaussian.

3. PHOTOELECTRON RANDOM SHOT NOISE

In this section, we study the statistical properties of the "random" shot noise which, as we have seen, describes more precisely the output of a detector of such a photomultiplier than the deterministic one defined by Eq. (1.1). Evidently, the "deterministic" shot noise is a particular case of "random" shot noise, and therefore we will obtain general expression valid in the two cases.[25]

The random function describing this noise can be written

$$X(t) = \int_{-\infty}^{+\infty} dX(t;\theta), \tag{3.1}$$

in which the $dX(t;\theta)$ are random increments depending on the fixed parameter t. The increments $dX(t;\theta)$ are connected with the $dN(\theta)$, defined previously by the fact that $dX(t;\theta) = 0$ if $dN(\theta) = 0$ and $dX(t;\theta) = R_\theta(t-\theta;\omega)$ if $dN(\theta) = 1$, where $R_\theta(t;\omega)$ are a sequence of nonstationary random functions defined on the same probability space and depending on the parameter θ. They are assumed independent and with the same probability distribution. This assumption means that the random impulses due to different photoelectrons are independent, which can be considered in a first time as a good physical approximation. In the *a posteriori* distribution, i.e., for a given $F(t)$, the $dN(\theta)$ are independent (Poisson distribution) and our assumption on the $R_\theta(t;\omega)$ functions means that the increments $dX(t;\theta)$ are also independent. This allows the calculation of the *a posteriori* characteristic function. For that, let us consider n arbitrary-time instants $t_1 \cdots t_n$. The *a posteriori* characteristic function is

$$\varphi[\{u_i\} \mid F] = E\left\{\exp\left[i\sum_1^n{}_i\, u_i X(t_i)\right] \,\middle|\, F\right\}. \tag{3.2}$$

We can write

$$\begin{aligned} u_1 X(t_1) + \cdots + u_n X(t_n) &= \int_{-\infty}^{+\infty} [u_1\, dX(t_1;\theta) + \cdots + u_n\, dX(t_n;\theta)] \\ &= \int_{-\infty}^{+\infty} dX'(t_1, \cdots, t_n;\theta), \end{aligned} \tag{3.3}$$

in which the increments dX' are still independent. This allows us to write the second characteristic function by the expression

$$\begin{aligned} \Psi[\{u_i\} \mid F] &= \log \varphi[\{u_i\} \mid F] \\ &= \int_{-\infty}^{+\infty} \log E[e^{i dX'(\theta)}]. \end{aligned} \tag{3.4}$$

From the structure of $dX(t;\theta)$, we deduce that

$$E[e^{i dX'(\theta)}] = 1 + [A(\{u_i\};\{t_i\};\theta) - 1]\alpha F(\theta)\, d\theta, \tag{3.5}$$

where

$$A(\{u_i\};\{t_i\};\theta) = E\exp\left[i\sum_1^n{}_i\, u_i R_\theta(t_i - \theta)\right]. \tag{3.6}$$

By using Eq. (3.4) we obtain finally

$$\varphi[\{u_i\} \mid F] = \exp\alpha\int_{-\infty}^{+\infty} [A(\{u_i\};\{t_i\};\theta) - 1]F(\theta)\, d\theta, \tag{3.7}$$

and the *a priori* characteristic function is obtained by taking the ensemble average on F, which gives

$$\varphi[\{u_i\}] = E\exp\left\{\alpha\int_{-\infty}^{+\infty} [A(\{u_i\};\{t_i\};\theta) - 1]F(\theta)\, d\theta\right\}. \tag{3.8}$$

As previously noticed, this expression is still valid in the case of the nondeterministic shot noise and the only difference is that we do not have to take the ensemble average in Eq. (3.6) because R is not random. For this case, a similar result was already presented by Hellstrom.[26]

By a limited expansion of the characteristic function we can obtain the moments of $X(t)$. Thus, in the stationary case, the mean value of $X(t)$ is

$$E[X(t)] = \rho\int_{-\infty}^{+\infty} E[R(t-\theta)]\, d\theta = \rho h(0), \tag{3.9}$$

in which $h(\nu)$ is the Fourier transform of

$$H(t) = E[R(t)]. \tag{3.10}$$

The result given by the Eq. (3.9) has the same form as the mean value of the deterministic shot noise given by a stationary Poisson process with the mean density ρ and a linear filter with impulse function $H(t)$, mean value of the random function $R(t)$. Thus, $H(t)$ can be considered as an equivalent impulse function.

The correlation function of $X(t)$ is given by an expansion limited to the second order and we obtain

$$\begin{aligned} &E[X(t_1)X(t_2)] \\ &= \rho^2 h^2(0) + \rho \int_{-\infty}^{+\infty} E[R(t_1 - \theta)R(t_2 - \theta)]\, d\theta \\ &\quad + \alpha^2 \iint_{-\infty}^{+\infty} H(t_1 - \theta_1)H(t_2 - \theta_2)\tilde{\Gamma}_F(\theta_1 - \theta_2)\, d\theta_1\, d\theta_2. \end{aligned} \tag{3.11}$$

By Fourier transformation we obtain the power spectrum of $X(t)$ which can be written

$$\gamma_X(\nu) = \rho^2 h^2(0)\delta(\nu) + \rho\, |g(\nu)|^2 + \alpha^2 \tilde{\gamma}_F(\nu)\, |h(\nu)|^2, \tag{3.12}$$

where $g(\nu)$ is the Fourier transform of

$$\int_{-\infty}^{+\infty} E[R(t)R(t - \tau)]\, dt.$$

It is interesting to compare this expression with Eq. (2.27), which is the power spectrum of the deterministic shot noise. For the random shot noise, $R(t)$, and, therefore, its Fourier transform $G(\nu)$, are random, and we have

$$|g(\nu)|^2 = E[|G(\nu)|^2]. \tag{3.13}$$

Moreover, from Eq. (3.10) we obtain

$$|h(\nu)|^2 = |E[G(\nu)]|^2, \tag{3.14}$$

and, therefore, $|g(\nu)|^2$ and $|h(\nu)|^2$ are in general quite different. In particular the last term can disappear completely if $E[G(\nu)] = 0 = E[R(t)]$. Evidently, that is not in general the case, if the randomness of $R(t)$ only due to the fluctuations of secondary emission in a photomultiplier.

Now we will consider some asymptotic properties of the shot noise $X(t)$. In the previous section, we have seen that the deterministic shot noise becomes Gaussian for very large time constant of the linear system and with some conditions on the light intensity. Now, we will study the asymptotic problem appearing if α, i.e., the mean light intensity, or the mean density of points [see Eq. (2.4)], becomes very large.

For this discussion let us introduce the two random functions $Y_\alpha(t)$ and $Z_\alpha(t)$ defined from $X(t)$ by

$$Y_\alpha(t) = \tilde{X}(t) = X(t) - E[X(t)], \tag{3.15}$$

$$Z_\alpha(t) = X(t) - \alpha F_R(t), \tag{3.16}$$

where $F_R(t)$ is

$$\begin{aligned} F_R(t) &= \int_{-\infty}^{+\infty} H(t - \theta)F(\theta)\, d\theta \\ &= \int_{-\infty}^{+\infty} E[R(t - \theta)]F(\theta)\, d\theta. \end{aligned} \tag{3.17}$$

At first, let us suppose that $F(t)$ is a *nonrandom function*, i.e., that the point process is a *pure nonstationary Poisson process*. In this case, we obtain from Eq. (3.7) $E[X(t)] = \alpha F_R(t)$, and, therefore,

$$Y_\alpha(t) \equiv Z_\alpha(t). \tag{3.18}$$

Moreover, the standard deviation of $Y_\alpha(t)$ is

$$E[Y_\alpha^2(t)] = \alpha \int E[R^2(t - \theta)]F(\theta)\, d\theta. \tag{3.19}$$

If we assume that $F(\theta)$ is bounded and $R(t)$ square integrable, the integral in Eq. (3.19) is finite and we obtain, therefore,

$$\lim_{\alpha \to \infty} \text{q.m.} \frac{Y_\alpha(t)}{\alpha} = 0, \tag{3.20}$$

where the quadratic mean limit is understood. This means that

$$\lim_{\alpha \to \infty} \text{q.m.} \frac{X(t)}{\alpha} = F_R(t). \tag{3.21}$$

This well-known result means that for very large α the *fluctuations* of the shot noise are suppressed. Evidently as convergence in the quadratic mean gives convergence in distribution, the characteristic function of $X(t)/\alpha$ converges to the characteristic function of $F_R(t)$.

But for the deterministic shot noise there is also a well-known result concerning the *fluctuations* of $Y_\alpha(t)$.[27] For large α, $Y_\alpha(t)/\alpha^{\frac{1}{2}}$ converges in distribution to a Gaussian random function defined by the covariance

$$\Gamma(t_1, t_2) = \int_{-\infty}^{+\infty} R(t_1 - \theta)R(t_2 - \theta)F(\theta)\, d\theta. \tag{3.22}$$

By using Eq. (3.7) it is possible to show that this particular form of the central limit theorem is still valid for the random shot noise and the covariance of the limit process is now

$$\int_{-\infty}^{+\infty} E[R(t_1 - \theta)R(t_2 - \theta)]F(\theta)\, d\theta.$$

Now let us consider the same problem for a stationary *Poisson compound process*, i.e., for the case where $F(t)$ is a *stationary random* function. In this

case $Y_\alpha(t)$ and $Z_\alpha(t)$ are different and Eq. (3.18) must be replaced by

$$Y_\alpha(t) = Z_\alpha(t) + \alpha\tilde{F}_R(t), \tag{3.23}$$

where

$$\tilde{F}_R(t) = F_R(t) - E[F_R(t)].$$

It is easy to show that Eq. (3.19) and Eq. (3.20) are no longer valid, because of the term in α^2 in Eq. (2.8). Therefore, we will calculate $E[Z_\alpha^2(t)]$ and by using Eqs. (3.1), (3.11), and (3.17) we obtain

$$\begin{aligned} E[Z_\alpha^2(t)] &= E\{[X(t) - \alpha F_R(t)]^2\} \\ &= E[X^2(t)] + \alpha^2 E[F_R^2(t)] - 2\alpha E[X(t)F_R(t)] \\ &= \alpha E[F]\int_{-\infty}^{+\infty} E[R^2(t-\theta)]\,d\theta. \end{aligned} \tag{3.24}$$

From this equation we deduce that

$$\lim_{\alpha\to\infty} \text{q.m.}\ \frac{Z_\alpha(t)}{\alpha} = 0. \tag{3.25}$$

Thus we obtain from Eq. (3.16)

$$\lim_{\alpha\to\infty} \text{q.m.}\ \frac{X(t)}{\alpha} = F_R(t), \tag{3.26}$$

but this equation is quite different from Eq. (3.21), because here $F_R(t)$ is a random function. Evidently the interpretation is the same and as previously the fluctuations of the shot noise are suppressed. Likewise the characteristic function of $X(t)/\alpha$ converges to the characteristic function of $F_R(t)$. Moreover, we obtain from Eq. (3.23) that

$$\lim_{\alpha\to\infty} \text{q.m.}\ \frac{Y_\alpha(t)}{\alpha} = \tilde{F}_R(t), \tag{3.27}$$

which is to compare with Eq. (3.20), because for a nonrandom $F(t)$, $\tilde{F}_R(t) = 0$.

Now let us study as previously the fluctuations of the shot noise. Evidently they are no longer described by $Y_\alpha(t)$, because we see from Eq. (3.27) that $Y_\alpha/\alpha^{\frac{1}{2}}$ is not finite when $\alpha \to \infty$. But we see from Eq. (3.24) that $Z_\alpha/\alpha^{\frac{1}{2}}$ remains finite for $\alpha \to \infty$, and therefore it is interesting to study the stochastic limit of this random function. The characteristic function $\varphi_\alpha[\{u_i\}]$ of $Z_\alpha/\alpha^{\frac{1}{2}}$ is obtained directly from Eq. (3.8) and after simple calculations we obtain that

$$\lim_{\alpha\to\infty} \varphi_\alpha[\{u_i\}] = \varphi_\infty[\{u_i\}], \tag{3.28}$$

where

$$\varphi_\infty[\{u_i\}] = E\exp\left[-\tfrac{1}{2}\sum_{ij} u_i u_j \times \int_{-\infty}^{+\infty} E[R(t_i-\theta)R(t_j-\theta)F(\theta)\,d\theta\right]. \tag{3.29}$$

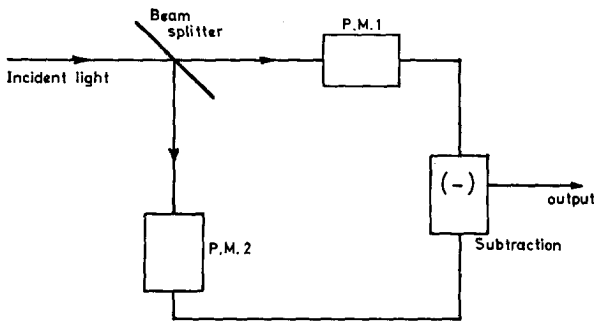


FIG. 1. Experimental method to obtain the asymptotic fluctuations of the random shot noise. The two photomultipliers receive the same light and have the same statistical properties. The two signals are subtracted and analyzed.

If $F(t)$ is nonrandom, we do not have to take the first expectation value in this equation and therefore, as we have seen previously, $\varphi_\infty[\{u_i\}]$ is the characteristic function of a Gaussian random function. If $F(t)$ is random, which is generally the case in optical problems, $\varphi_\infty[\{u_i\}]$ is the characteristic function of a *compound Gaussian random function.* Therefore, the asymptotic fluctuations of the shot noise are no longer Gaussian, and the central-limit theorem is not strictly valid. The same kind of situation appears in the problem of the limit of a random number of independent random variables.[28]

At this point of the discussion it is important to explore if it is physically possible to obtain the function $Z_\alpha(t)$. Indeed to define $Z_\alpha(t)$ we must know the random function $F_R(t)$ which appears in Eq. (3.16). This function defined by Eq. (3.17) is obtained by filtering $F(t)$ in a linear system with impulse function $H(t)$. But $F(t)$ is not exactly known in a real photodetector, because it is the instantaneous mean density of the compound Poisson process which is not measurable experimentally by studying only the point process. Evidently $F(t)$ is proportional to the light intensity which can be measured in some cases by another detector in the same field.

However, without a measurement of $F(t)$ it is possible to observe the random function $Z_\alpha(t)$, i.e., the fluctuations of the shot noise, by using a symmetrical system described in Fig. 1. As in the Hanbury Brown and Twiss experiment, we take two identical photodetectors (i.e., photomultipliers) in the same coherence area of an optical field. We assume that the two detectors are independent but identical, and therefore the two random shot noises have the same statistical properties and distribution, but correlated by the same light intensity $F(t)$ in the *a priori* distribution. Thus after subtraction we obtain the signal

$$S(t) = \int_{-\infty}^{+\infty} dS(t;\theta) \tag{3.30}$$

with

$$dS(t;\theta) = dX_1(t;\theta) - dX_2(t;\theta), \quad (3.31)$$

where the increments dX_1 and dX_2 have the same meaning as in Eq. (3.1). Now we can use the same procedure to calculate the characteristic function $\varphi_S[\{u_i\}]$ of $S(t_1)\cdots S(t_n)$. The only difference appears in Eq. (3.6), which is now, with the same notations,

$$A_S(\{u_i\};\{t_i\};\theta) = E\left[\cos\sum_1^n u_i R_\theta(t_i-\theta)\right]. \quad (3.32)$$

From this equation and Eq. (3.8) we can calculate the characteristic function of $S(t)/\alpha^{\frac{1}{2}}$ which for large α converge to $\varphi_\infty[\{u_i\}]$ defined by Eq. (3.29). Thus $Z_\alpha(t)/\alpha^{\frac{1}{2}}$ and $S(t)/\alpha^{\frac{1}{2}}$ have the same limit in distribution for $\alpha \to \infty$, which describes the "fluctuations" of the asymptotic shot noise.

Now we will explore more carefully the properties of these asymptotic fluctuations.

4. PROPERTIES OF THE ASYMPTOTIC FLUCTUATIONS OF THE SHOT NOISE-COMPOUND GAUSSIAN PROCESSES

The random function $M(t)$ which describes the asymptotic fluctuations of the shot noise is completely defined by the characteristic functions of $M(t_1),\cdots,M(t_n)$ given by Eq. (3.29). This function can be written

$$\varphi[u_1,\cdots,u_n] = E\exp\left(-\tfrac{1}{2}\sum_{i,j} u_i u_j \Gamma_{ij}\right), \quad (4.1)$$

where Γ_{ij} is a random variable defined by

$$\Gamma_{ij} = \int_{-\infty}^{+\infty} F(\theta)E[R(t_i-\theta)R(t_j-\theta)]\,d\theta. \quad (4.2)$$

The random function $M(t)$ is called a compound Gaussian process because if $F(\theta)$ is nonrandom, $M(t)$ is a pure zero mean Gaussian random function. From the probability distribution of the random function $F(\theta)$ which describes the fluctuations of the light intensity, we can in principle obtain the probability distribution of the n^2 random variable Γ_{ij}, of which evidently only $n(n-1)$ are different. This distribution is defined by a characteristic function

$$\varphi_\Gamma[\{v_{ij}\}] = E\left(\exp i\sum_{ij}^n v_{ij}\Gamma_{ij}\right). \quad (4.3)$$

With this function, Eq. (4.1) can be written

$$\varphi[u_1,\cdots,u_n] = \varphi_\Gamma\left[\left\{\frac{i}{2}u_i u_j\right\}\right]. \quad (4.4)$$

This equation defines completely $M(t)$, and we will at first consider the 1-dimensional case. Thus the characteristic function of $M(t)$ is

$$\varphi(u) = \varphi_\Gamma\left[\frac{i}{2}u^2\right], \quad (4.5)$$

where $\varphi_\Gamma(u)$ is the characteristic function of the random variable

$$\Gamma = \int_{-\infty}^{+\infty} E[R^2(t-\theta)]F(\theta)\,d\theta. \quad (4.6)$$

The probability distribution of $M(t)$ is obtained by taking the Fourier transform of $\varphi(u)$, and we obtain

$$p(x) = (2\pi)^{-\frac{1}{2}}\int_0^\infty \frac{1}{y^{\frac{1}{2}}} e^{-x^2/2y} p_\Gamma(y)\,dy, \quad (4.7)$$

where $p_\Gamma(y)$ is the probability distribution of Γ. We have evidently $p_\Gamma(y) = 0$ for $y < 0$, because Γ is a positive random variable. We have the same kind of equation for the probability distribution of the photon counting, in which the kernel is a Poisson kernel instead of a Gaussian one, and therefore we can apply the same kinds of methods. Particularly, it is in general very difficult to calculate the probability distribution of the random variable Γ, integral of a random function. Thus we will suppose that the correlation time of $F(t)$ is much greater than the average time constant of the random filter $R(t)$. In this case the integration can be omitted, and Γ has the same statistical distribution as the light intensity F.

To perform the calculations we will consider some examples of optical fields.

At first, let us suppose that the optical field has a constant light intensity (which is for example the case of the ideal amplitude stabilized laser light with only phase fluctuations). Thus $\Gamma = a$ and $\varphi_\Gamma(u) = e^{iau}$. Therefore $\varphi(u) = \exp -\frac{1}{2}au^2$, which is quite obvious, because if $F(\theta)$ is constant, the compound Gaussian process becomes a pure Gaussian process.

If now the optical field is created by a thermal source (natural light, or pseudothermal light), it is well known that the probability distribution of the light intensity is $a^{-1}e^{-x/a}$ and the characteristic function $1/(1-iau)$. Thus from Eq. (4.5) we obtain

$$\varphi(u) = \frac{1}{1+\frac{1}{2}au^2} \quad (4.8)$$

and the probability distribution is

$$p(x) = \frac{1}{(2a)^{\frac{1}{2}}} e^{-(2/a)^{\frac{1}{2}}|x|}. \quad (4.9)$$

Finally we will consider the superposition of the two previous fields. In this case the probability

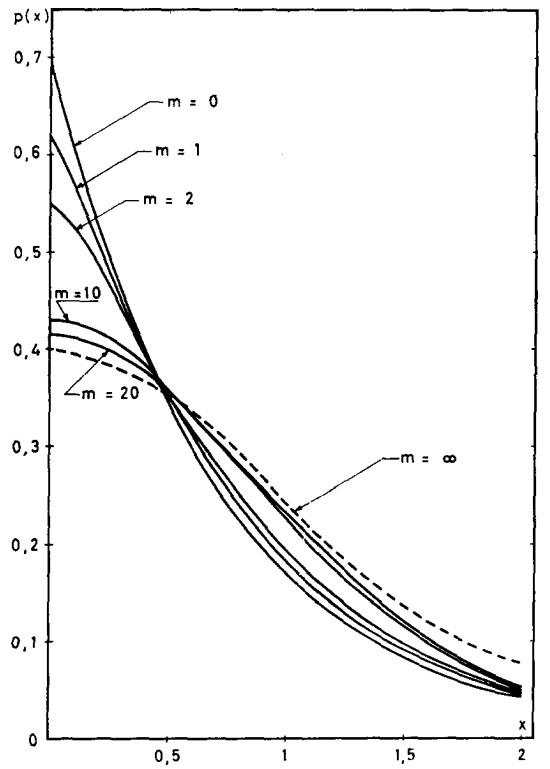


FIG. 2. Probability distribution functions $p(x)$ of the asymptotic shot noise for mixed light fields with some values of m, ratio of laser intensity to mean thermal intensity ($m = 0, 1, 2, 10, 20, \infty$).

distribution of the light intensity is[29]

$$p_\Gamma(y) = (m+1)\exp -[(m+1)y + m] \times I_0[2(m(m+1)y)^{\frac{1}{2}}], \quad (4.10)$$

where $m = I_l/I_t$, ratio of the mean light intensities of the laser and the thermal fields, and y is a reduced variable, ratio of the instantaneous intensity and the total mean intensity $I_l + I_t$. The characteristic function of this distribution is obtained by Fourier transformation

$$\varphi_\Gamma(u) = \frac{m+1}{m+1-iu}\exp\left(-\frac{imu}{m+1-iu}\right), \quad (4.11)$$

and therefore the characteristic function of the asymptotic shot noise is

$$\varphi(u) = \frac{m+1}{m+1+u^2/2}\exp\left(\frac{1}{2}\frac{mu^2}{m+1+u^2/2}\right). \quad (4.12)$$

We have no simple explicit form of the Fourier transform of $\varphi(u)$, and thus we can also use for computation Eq. (4.7), where $p_\Gamma(y)$ is given by Eq. (4.10). Computation of $p(x)$ for many values of m have been realized and the results appear in Fig. 2. We have only considered the positive values of x, because $p(x)$ is an even function. Evidently, for the limit cases $m = 0$ and $m \to \infty$, we find Eq. (4.9) and the Gaussian distribution.

Now let us consider two time instants t_1 and t_2 and calculate the probability distribution of the 2-dimensional random variable $M(t_1)$, $M(t_2)$. In this case, Eq. (4.4) can be written

$$\varphi[u_1, u_2] = E\exp[-\tfrac{1}{2}(\Gamma_{11}u_1^2 + 2\Gamma_{12}u_1u_2 + \Gamma_{22}u_2^2)], \quad (4.13)$$

where Γ_{ij} is still given by Eq. (4.2). As previously, we will suppose that $F(\theta)$ is stationary and that

$$T \ll t_c, \quad (4.14)$$

where T is the average time constant of the random linear filter and t_c the correlation time of $F(\theta)$. As a consequence of this inequality, we have

$$\Gamma_{\substack{11\\22}} = \lambda F_{\substack{1\\2}}, \quad (4.15)$$

where λ is $\int_{-\infty}^{+\infty} E[R^2(t)]\,dt$ and $F_{\substack{1\\2}} = F\left(t_{\substack{1\\2}}\right)$. Moreover, we can introduce $\tau = |t_1 - t_2|$ and write

$$\Gamma_{12} = \Gamma(\tau) = F\int_{-\infty}^{+\infty} E[R(t)R(t-\tau)]\,dt = \lambda F\tilde{\Gamma}_R(\tau), \quad (4.16)$$

where $\tilde{\Gamma}_R(\tau)$ is a normalized correlation function ($\tilde{\Gamma}_R(0) = 1$).

Different cases depending on the parameter τ can now be considered:

(i) $T \ll t_c \ll \tau$.

Here we have evidently $\tilde{\Gamma}_R(\tau) = 0$, and $F(t_1)$ and $F(t_2)$ are practically independent. Therefore, $\varphi[u_1, u_2]$ can be written

$$\varphi[u_1, u_2] = \varphi(u_1)\varphi(u_2) = \varphi_F\left(\frac{i}{2}\lambda u_1^2\right)\varphi_F\left(\frac{i}{2}\lambda u_2^2\right), \quad (4.17)$$

where $\varphi_F(v)$ is the characteristic function of F. Equation (4.17) shows that the variables $M(t_1)$ and $M(t_2)$ are independent.

(ii) $\tau \simeq t_c$.

We have always $\tilde{\Gamma}_R(\tau) = 0$, but F_1 and F_2 are now correlated and have the characteristic function $\varphi_F(v_1, v_2)$. From Eq. (4.13) we have

$$\varphi[u_1, u_2] = \varphi_F\left[\frac{i}{2}\lambda u_1^2, \frac{i}{2}\lambda u_2^2\right]. \quad (4.18)$$

(iii) $T \ll \tau \ll t_c$.

We have still $\tilde{\Gamma}_R(\tau) = 0$, but $F_1 \simeq F_2$, and thus

$$\varphi[u_1, u_2] = \varphi_F\left[\frac{i}{2}\lambda(u_1^2 + u_2^2)\right]. \quad (4.19)$$

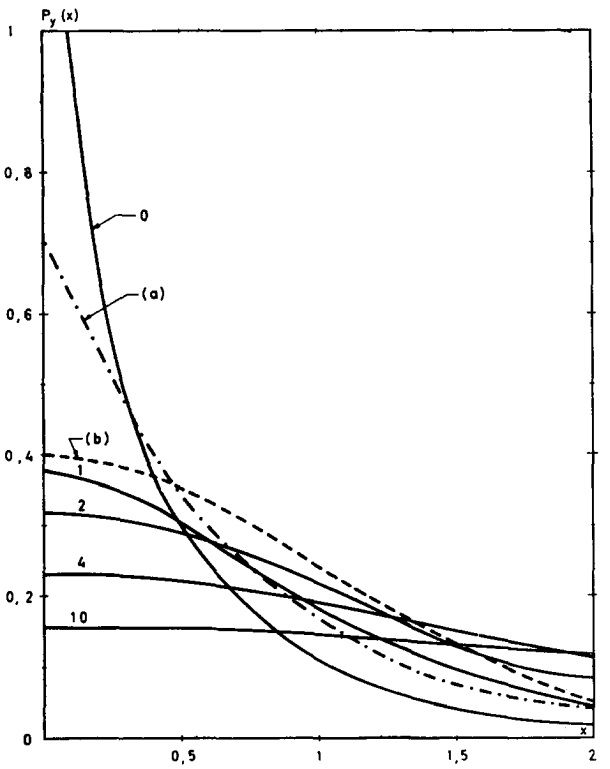


FIG. 3. *A posteriori* probability-distribution function $p_y(x)$ for $m = 0$ (thermal field) and some values of y ($y = 0, 1, 2, 4, 10$). The curves (*a*) and (*b*) are the *a priori* distribution functions for $m = 0$ (thermal field) and $m = \infty$ (ideal laser field).

(iv) $\tau \simeq T$.

Now $\tilde{\Gamma}_F(\tau) \neq 0$, $F_1 = F_2$, and we have

$$\varphi[u_1, u_2] = \varphi_F\left[\frac{i}{2}\lambda(u_1^2 + u_2^2 + 2\tilde{\Gamma}(\tau)u_1 \cdot u_2)\right]. \quad (4.20)$$

We must consider carefully the case (iii) for which $\varphi(u_1, u_2)$ is practically independent on τ. The structure of $\varphi(u_1, u_2)$ given by Eq. (4.19) is characteristic of a 2-dimensional spherically invariant random variable.[30-32] The particular property of $M(t_1)$ and $M(t_2)$, in this case, is that they are uncorrelated but nonindependent random variables. This appears clearly on Eq. (4.19). Therefore an interesting parameter describing the statistical dependence is the *a posteriori* probability distribution defined by

$$P_{x_2}(x_1) = \frac{\rho(x_1, x_2)}{p(x_2)}, \quad (4.21)$$

where $\rho(x_1, x_2)$ is the Fourier transform of $\varphi(u_1, u_2)$ and $p(x_2)$ the *a priori* distribution of $M(t)$ studied previously.

We will calculate this distribution for a thermal optical field. In this case $\varphi_F(u) = 1/(1 - iu)$, and by supposing $\lambda = 1$, we have a new random variable (X, Y) instead of (M_1, M_2) whose characteristic function is

$$\varphi(u, v) = [1 + (1/2)(u^2 + v^2)]^{-1} \quad (4.22)$$

and the Fourier transforms is

$$\rho(x, y) = \frac{1}{\pi} K_0\{[2(x^2 + y^2)]^{\frac{1}{2}}\}, \quad (4.23)$$

where K_0 is the zero-order modified Bessel function of the second kind.

By using Eq. (4.9) we find the *a posteriori* probability distribution

$$p_y(x) = \frac{\sqrt{2}}{\pi} e^{\sqrt{2}|y|} K_0\{[2(x^2 + y^2)]^{\frac{1}{2}}\}. \quad (4.24)$$

This probability distribution is represented on the Fig. 3 for some values of y. For large y, $p_y(x)$ becomes approximately a Gaussian probability density, which can be seen by the asymptotic form of the K_0 function.

All these probability distributions show the difference between compound and true Gaussian random functions which describe the fluctuations of the asymptotic shot noise of thermoelectrons and photoelectrons. Compound Gaussian stochastic processes have also many other interesting properties which will be studied in another paper.

* Laboratoire associé au Centre National de la Recherche Scientifique.

[1] L. Mandel, *Progress in Optics*, E. Wolf, Ed. (North-Holland Publishing Co., Amsterdam, 1963), Vol. II, p. 181.
[2] L. Mandel, E. C. G. Sudarshan, and E. Wolf, Proc. Phys. Soc. (London) **84**, 435 (1964).
[3] L. Mandel and E. Wolf, Phys. Rev. **149**, 1033 (1966).
[4] R. J. Glauber, in *Quantum Optics and Electronics*, C. de Witt, A. Blandin, and C. Cohen-Tannoudji, Eds. (Gordon & Breach, New York, 1965).
[5] P. L. Kelley and W. H. Kleiner, Phys. Rev. **136**, A316 (1964).
[6] F. Rocca, J. Phys. (Paris) **28**, 113 (1967).
[7] J. R. Klauder, J. McKenna, and D. G. Currie, J. Math. Phys. **6**, 733 (1965).
[8] C. L. Mehta and E. C. G. Sudarshan, Phys. Rev. **138**, B274 (1965).
[9] C. Freed and H. A. Haus, *Physics of Quantum Electronics*, P. L. Kelley, B. Lax, and P. E. Tannewald, Eds. (McGraw-Hill, New York, 1965), p. 715.
[10] F. A. Johnson, T. P. McLean, and E. R. Pike, *Physics of Quantum Electronics*, P. L. Kelley, B. Lax, and P. E. Tannewald, Eds. (McGraw-Hill, New York, 1965), p. 706.
[11] F. T. Arecchi, A. Berne, and P. Burlamacchi, Phys. Rev. Letters **16**, 32 (1966).
[12] W. Martienssen and E. Spiller, Phys. Rev. Letters **16**, 531 (1966).
[13] A. W. Smith and J. A. Armstrong, Phys. Rev. Letters **16**, 1169 (1966).
[14] J. A. Armstrong and A. W. Smith, Phy. Rev. **140**, A155 (1965).
[15] B. L. Morgan and L. Mandel, Phys. Rev. Letters **16**, 1012 (1966).
[16] F. T. Arecchi, E. Gatti, and A. Sona, Phys. Letters **20**, 27 (1966).

[17] D. B. Scarl, Phys. Rev. Letters **17**, 663 (1966).
[18] A. Blanc-Lapierre and R. Fortet, *Théorie des Fonctions aléatoires* (Masson, Paris, 1953), p. 130.
[19] A. Blanc-Lapierre, *Modèles statistiques pour l'étude des Phénomènes de fluctuations* (Masson, Paris, 1963), p. 39.
[20] See Ref. 19, p. 41.
[21] B. Picinbono, Compt. Rend. (Paris) **250**, 1174 (1960); B. Picinbono, Ann. Telecomm. (Paris) **16**, 12 (1961).
[22] This problem is presented with details in Refs. 21.
[23] B. Picinbono and J. Pouget, Compt. Rend. (Paris) **268**, 1481 (1969).
[24] L. Mandel, "Communication to the Symposium on Modern Optics (New York, N.Y., March 1967)" (Polytechnic Press of the Polytechnic Institute of Brooklyn, Brooklyn, N.Y., 1967), p. 143.
[25] The "random" shot noise due to a stationary Poisson process is also considered in Ref. 19, p. 49.
[26] C. W. Hellstrom, J. Opt. Soc. Am. **157**, 353 (1967).
[27] See Ref. 18, p. 161.
[28] H. Robbins, Bull. Am. Math. Soc. **54**, 1151 (1948).
[29] G. Lachs, Phys. Rev. **138B**, 1012 (1965).
[30] A. M. Vershik, Theoret. Prob. Appl. **9**, 353 (1964).
[31] I. F. Blake and J. B. Thomas, IEEE Trans. Inform. Theory **14**, 12 (1968).
[32] B. Picinbono, IEEE Trans. Inform. Theory **15**, 611 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Local Gauge Fields in General Relativity*

RICHARD P. TREAT
University of California, Riverside, California, and *West Virginia University, Morgantown, West Virginia 26506*†

(Received 2 October 1969)

The carrier space for internal degrees of freedom is assumed to be the 6-dimensional space spanned by generators $L_{i\kappa}{}^{\lambda}$ of local Lorentz transformations defined at each event of a Riemann space. Gauge potentials are introduced in this internal space which are like the Yang–Mills potentials except that they are related to parameters of the event connection $\Gamma_{\mu\kappa}^{\lambda}$ by a map which involves the $L_{i\kappa}{}^{\lambda}$ as connecting quantities. The map is similar to the one which relates the Dirac spin space and spin connection to the Riemannian geometry of event space. An algebraic uniqueness condition is shown which is necessary and sufficient for the map to be one-to-one in a neighborhood of a solution $\Gamma_{\mu\kappa}^{\lambda}$ of the mapping equations: The metric field is undetermined by a uniform scale transformation. If the gauge potentials have an internal holonomy group which is a subgroup of the Lorentz group and satisfy the free Yang–Mills equations, and if the uniqueness condition is satisfied, then the $\Gamma_{\mu\kappa}^{\lambda}$ committed by the inverse map automatically satisfy the free Einstein field equations, indicating that only the gravitational field is included in this mathematical framework. In an attempt to include nongravitational properties of matter, we have then considered internal holonomy groups larger than the Lorentz group. The part of the gauge potentials causing the enlargement is shown to form a current which may be interpreted as a source of the gravitational field.

I. INTRODUCTION

Some of the fields that appear in physical theories, such as the Dirac spinor field, do not have a direct geometric meaning, as they appear overlayed on the event space without influencing the event geometry in any way. This kind of field is in contrast to the gravitational field which has a direct geometrical consequence and, in fact, determines the Riemannian geometry of physical space. For this reason, the theory of gravitation is said to be a geometric theory. The electromagnetic potentials and gauge potentials (Yang–Mills potential[1] and their generalizations[2]) are akin to the gravitational field as they have a similar direct geometrical role. The electromagnetic potentials provide the definition of equivalence of Dirac spinors, and the gauge potentials define the equivalence of particle multiplet fields at neighboring events.

The mathematical definition of gauge potentials and gauge fields was given sometime prior to their use in physics in König,[3] who introduced "parameters of a linear connection" which provide the definition of equivalence of internal vectors at neighboring points of a geometric manifold. This extension of the notion of equivalence to vector spaces which are independent of the underlying geometric manifold was viewed as a generalization of the parallel displacement of the tangent space of the event manifold. König's definition of parameters of a linear connection for internal vectors is independent of the dimension and structure of the vector space and includes the electromagnetic potentials, spin connection,[4] and Yang–Mills potentials and their generalizations as special cases. These parameters of a linear connection are considered in the mathematical literature of modern differential geometry as a special case of fiber bundles with a linear connection.[5]

Parameters of a linear connection for vectors other than event vectors were first considered in physical

[17] D. B. Scarl, Phys. Rev. Letters **17**, 663 (1966).
[18] A. Blanc-Lapierre and R. Fortet, *Théorie des Fonctions aléatoires* (Masson, Paris, 1953), p. 130.
[19] A. Blanc-Lapierre, *Modèles statistiques pour l'étude des Phénomènes de fluctuations* (Masson, Paris, 1963), p. 39.
[20] See Ref. 19, p. 41.
[21] B. Picinbono, Compt. Rend. (Paris) **250**, 1174 (1960); B. Picinbono, Ann. Telecomm. (Paris) **16**, 12 (1961).
[22] This problem is presented with details in Refs. 21.
[23] B. Picinbono and J. Pouget, Compt. Rend. (Paris) **268**, 1481 (1969).
[24] L. Mandel, "Communication to the Symposium on Modern Optics (New York, N.Y., March 1967)" (Polytechnic Press of the Polytechnic Institute of Brooklyn, Brooklyn, N.Y., 1967), p. 143.
[25] The "random" shot noise due to a stationary Poisson process is also considered in Ref. 19, p. 49.
[26] C. W. Hellstrom, J. Opt. Soc. Am. **157**, 353 (1967).
[27] See Ref. 18, p. 161.
[28] H. Robbins, Bull. Am. Math. Soc. **54**, 1151 (1948).
[29] G. Lachs, Phys. Rev. **138B**, 1012 (1965).
[30] A. M. Vershik, Theoret. Prob. Appl. **9**, 353 (1964).
[31] I. F. Blake and J. B. Thomas, IEEE Trans. Inform. Theory **14**, 12 (1968).
[32] B. Picinbono, IEEE Trans. Inform. Theory **15**, 611 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Local Gauge Fields in General Relativity*

RICHARD P. TREAT
University of California, Riverside, California, and *West Virginia University, Morgantown, West Virginia 26506*†

(Received 2 October 1969)

The carrier space for internal degrees of freedom is assumed to be the 6-dimensional space spanned by generators $L_{i\kappa}{}^{\lambda}$ of local Lorentz transformations defined at each event of a Riemann space. Gauge potentials are introduced in this internal space which are like the Yang–Mills potentials except that they are related to parameters of the event connection $\Gamma_{\mu\kappa}^{\lambda}$ by a map which involves the $L_{i\kappa}{}^{\lambda}$ as connecting quantities. The map is similar to the one which relates the Dirac spin space and spin connection to the Riemannian geometry of event space. An algebraic uniqueness condition is shown which is necessary and sufficient for the map to be one-to-one in a neighborhood of a solution $\Gamma_{\mu\kappa}^{\lambda}$ of the mapping equations: The metric field is undetermined by a uniform scale transformation. If the gauge potentials have an internal holonomy group which is a subgroup of the Lorentz group and satisfy the free Yang–Mills equations, and if the uniqueness condition is satisfied, then the $\Gamma_{\mu\kappa}^{\lambda}$ committed by the inverse map automatically satisfy the free Einstein field equations, indicating that only the gravitational field is included in this mathematical framework. In an attempt to include nongravitational properties of matter, we have then considered internal holonomy groups larger than the Lorentz group. The part of the gauge potentials causing the enlargement is shown to form a current which may be interpreted as a source of the gravitational field.

I. INTRODUCTION

Some of the fields that appear in physical theories, such as the Dirac spinor field, do not have a direct geometric meaning, as they appear overlayed on the event space without influencing the event geometry in any way. This kind of field is in contrast to the gravitational field which has a direct geometrical consequence and, in fact, determines the Riemannian geometry of physical space. For this reason, the theory of gravitation is said to be a geometric theory. The electromagnetic potentials and gauge potentials (Yang–Mills potential[1] and their generalizations[2]) are akin to the gravitational field as they have a similar direct geometrical role. The electromagnetic potentials provide the definition of equivalence of Dirac spinors, and the gauge potentials define the equivalence of particle multiplet fields at neighboring events.

The mathematical definition of gauge potentials and gauge fields was given sometime prior to their use in physics in König,[3] who introduced "parameters of a linear connection" which provide the definition of equivalence of internal vectors at neighboring points of a geometric manifold. This extension of the notion of equivalence to vector spaces which are independent of the underlying geometric manifold was viewed as a generalization of the parallel displacement of the tangent space of the event manifold. König's definition of parameters of a linear connection for internal vectors is independent of the dimension and structure of the vector space and includes the electromagnetic potentials, spin connection,[4] and Yang–Mills potentials and their generalizations as special cases. These parameters of a linear connection are considered in the mathematical literature of modern differential geometry as a special case of fiber bundles with a linear connection.[5]

Parameters of a linear connection for vectors other than event vectors were first considered in physical

theories in investigations of the spinor calculus in a Riemann space[4] and in Einstein and Mayer's version of the Klein–Kalutza theory.[6] Later the Yang–Mills potentials were introduced in relation to isospin multiplet fields by a physical argument which may be stated as follows: At a given event, the bases of a 3-dimensional linear vector space which is endowed with a metric tensor are indistinguishable to the extent of transformations which leave the metric tensor invariant. If by the choice of a basis at one event a preferred basis at another event were somehow selected, this would be in conflict with the concept of a local-field theory and, hence, the choice of bases should be arbitrary up to local orthogonal base transformations. When local base transformations are allowed, it then becomes necessary to introduce gauge potentials to obtain an invariant definition of the differential of an isovector, or what is the same, a definition of the equivalence of isovectors at neighboring events. Of course, the demand for consistency under local base transformations does not compel one to introduce nontrivial gauge potentials. For example, in the Minkowskian event space, parameters of a connection arise through the use of curvilinear coordinates, but they are of no basic interest. Likewise, integrable gauge potentials could be employed in order to obtain a consistent transformation theory under local gauge transformations, but the gauge potentials and their concomittants would not have any physical significance. Physically significant gauge potentials may be introduced by simply identifying the field that provides the definition of equivalence of internal vectors as the field which forms the substrata of physical quantities, which is the point of view of Loos[7] and is the one which we take.

One may consider the gauge field coupled to particle fields, as Yang and Mills have done, or consider the properties of the free field, as one does when investigating the electromagnetic field free of its sources. The free field, which we study here, satisfies nonlinear field equations, which presents great difficulties in the quantization of the field and also in the investigation at the classical level. However, these free-field equations hold promise of admitting permanently localized regular solutions which could possibly avoid the point particle concept and obtain a field description of extended particles. This is the role of the field in physical theory envisioned by Weyl[8]: "The theory of fields has to explain why the field is granular in structure and why these energy knots preserve themselves permanently from energy and momentum in their passage to and fro," and "It is not the field that requires matter as its carrier, but matter is, on the contrary, an offspring of the field."

Although the geometric nature of local gauge theory was recognized in the beginning,[2] the use of already existing concepts and methods of differential geometry are usually not emphasized. Rather it seems that the fundamental geometric concept of equivalence of vectors is usually avoided even though the gauge potentials necessarily provide such a definition by virtue of their transformation law. Here, we follow Loos[7,9] and make free use of equivalence and related concepts and existing methods of differential geometry. The use of integrability conditions, which lead to the Einstein field equation in the theory of gravitation and the conservation of charge in electrodynamics, is also a powerful tool in the investigation of local gauge fields. The internal holonomy group of a gauge field[10] has been a useful group theoretical concept in this investigation, while the gauge group has not been of as much importance.

Here we are concerned with gauge potentials and fields defined on a Riemannian event space and acting in the 6-dimensional internal space formed by the generators of local Lorentz transformations.[11] This particular space is considered as a possible carrier space for internal degrees of freedom primarily because of the especially simple mathematical relationship of the space and the gauge potentials to the event geometry, a relation which is similar to the one that exists between the Dirac spin space, spin connection, and Riemannian event geometry.[4] If the Riemannian geometry is given, a map exists such that the event metric field determines uniquely the gauge potentials acting in the internal space formed by the local generators of the event holonomy group,[11,12] which in physically interesting cases is the homogeneous Lorentz group. In this work, we consider the problem of the inverse map, where we take the gauge potentials as given fundamental fields and study the extent to which the gauge potentials then determine an underlying Riemannian event geometry, a problem similar to one which has been posed and studied for the spin connection.[13] The inverse map is studied for gauge potentials not satisfying field equations and also for gauge potentials satisfying free Yang–Mills equations. The case where the internal holonomy group of the gauge field is the Lorentz group or a subgroup thereof is considered first. Finally, we study a more general class of gauge fields which can have an internal holonomy group larger than the Lorentz group.

The investigation is carried out at the classical level. In the case of the electromagnetic field, properties of

the classical field can be determined which also hold for the quantized field, e.g., the range of interaction. But it is not clear what conclusions with respect to the classical Yang–Mills field carry over to a quantized theory. Nevertheless, the difficulties are such that an investigation on the classical level appears to be useful and perhaps essential to further development and an eventual quantization of the theory.

II. THE INTERNAL SPACE AND THE GEOMETRY OF GAUGE FIELDS

A Riemannian geometry of event space is established by the parameters $\Gamma^{\lambda}_{\mu\kappa}$, $\mu, \kappa, \cdots = 0, 1, 2, 3$, of a symmetric linear connection and a symmetric metric tensor $g_{\kappa\lambda}$ which satisfies

$$\nabla_{\mu} g_{\kappa\lambda} = 0, \tag{1}$$

where ∇_{μ} denotes covariant differentiation with respect to $\Gamma^{\lambda}_{\mu\kappa}$.

It is possible to associate in a natural way to each point in the event space a 6-dimensional space spanned by the generators of local Lorentz transformations.[12] This space is considered here as a possible carrier space for the internal degrees of freedom of elementary particles.[11] In this section, we review how this space arises and how the gauge fields acting in the space are related to the Riemannian event geometry. In the context of the discussion, we introduce the geometrical concepts associated with the gauge fields which are relevant to this investigation.

In a Riemann space, the generators of local Lorentz transformations $L_{i\kappa}{}^{\lambda}(x^{\mu})$, $i, j, \cdots = 1, 2, \cdots, 6$, are introduced as the 6-linearly independent solutions of the algebraic equation

$$L_{i\kappa}{}^{\rho} g_{\rho\lambda} + L_{i\lambda}{}^{\rho} g_{\rho\kappa} = 0. \tag{2}$$

Coordinate transformations generated by $L_{i\kappa}{}^{\lambda}$ satisfying (2) leave the metric tensor $g_{\kappa\lambda}$ invariant and, hence, the $L_{i\kappa}{}^{\lambda}$ generate local Lorentz transformations. On account of (2) the $L_{i\kappa}{}^{\lambda}$ satisfy the commutation rules

$$L_{i\kappa}{}^{\rho} L_{j\rho}{}^{\lambda} - L_{j\kappa}{}^{\rho} L_{i\rho}{}^{\lambda} = c_{ij}{}^{k} L_{k\kappa}{}^{\lambda}, \tag{3}$$

where $c_{ij}{}^{k}$ are the structure constants of the Lorentz group satisfying the usual structural equations. In a Minkowski space and in a coordinate system, where $g_{\kappa\lambda}$ is constant, the $L_{i\kappa}{}^{\lambda}$ may be taken as constant matrices, but in a Riemann space the solutions of (2) are functions of the coordinates. The generators are determined by (2) up to algebra base transformations

$$L_{i'\kappa}{}^{\lambda} = A^{i}_{i'} L_{i\kappa}{}^{\lambda} \tag{4}$$

and similarity transformations with transformations belonging to the Lorentz group. Now, to each event of the Riemann space, we associate the 6-dimensional linear vector space of the Lie algebra of the local Lorentz group, which is spanned by the six linearly independent local generators $L_{i\kappa}{}^{\lambda}$.

Covariant differentiation of (2) with respect to event indices only and use of the covariant constancy of the metric tensor yield

$$({}^{e}\nabla_{\mu} L_{i\kappa}{}^{\rho}) g_{\rho\lambda} + ({}^{e}\nabla_{\mu} L_{i\lambda}{}^{\rho}) g_{\rho\kappa} = 0, \tag{5}$$

where ${}^{e}\nabla_{\mu}$ indicates the covariant differentiation is restricted to event indices. Equation (5) is satisfied if ${}^{e}\nabla_{\mu} L_{i\kappa}{}^{\lambda}$ is a linear combination of the $L_{i\kappa}{}^{\lambda}$, say

$${}^{e}\nabla_{\mu} L_{i\kappa}{}^{\lambda} = \Gamma_{\mu i}{}^{j} L_{j\kappa}{}^{\lambda}. \tag{6}$$

Under the change of algebra base (4), the parameters $\Gamma_{\mu i}{}^{j}$ introduced in (6) are specified to transform as

$$\Gamma_{\mu i'}{}^{j'} = A^{i}_{i'} \Gamma_{\mu i}{}^{j} A^{j'}_{j} - A^{i}_{i'} \partial_{\mu} A^{j'}_{i}, \tag{7}$$

where $A^{i}_{i'} A^{i'}_{j} = \delta^{i}_{j}$. By (6) and (7), we have extended covariant differentiation to the algebra indices $i, j, \cdots$. We may then write (6) in the form

$$\nabla_{\mu} L_{i\kappa}{}^{\lambda} = 0, \tag{8}$$

where now ∇_{μ} denotes covariant differentiation with respect to both event and algebra indices. Similarly to the geometric meaning of the event connection, the parameters $\Gamma_{\mu i}{}^{j}$ provide a definition of equivalence for quantities belonging to the 6-dimensional algebra space. That is, if for the vector $\eta^{i}(x^{\kappa})$ in the Lorentz algebra space at the event x^{κ},

$$dx^{\mu} \nabla_{\mu} \eta^{i}(x^{\kappa}) = 0, \tag{9}$$

then the vectors $\eta^{i}(x^{\kappa})$ and $\eta^{i}(x^{\kappa} + dx^{\kappa})$ are said to be equivalent. On account of the transformation law (7), the definition of equivalence and the covariant constancy of $L_{i\kappa}{}^{\lambda}$ are invariant properties under local algebra base transformations.

From the parameters $\Gamma_{\mu i}{}^{j}$ we construct, in perfect analogy to the Riemann tensor, the geometric quantity

$$\phi_{\mu\nu i}{}^{j} = \partial_{\mu} \Gamma_{\nu i}{}^{j} - \partial_{\nu} \Gamma_{\mu i}{}^{j} - \Gamma_{\mu i}{}^{k} \Gamma_{\nu k}{}^{j} + \Gamma_{\nu i}{}^{k} \Gamma_{\mu k}{}^{j}, \tag{10}$$

which, under the base transformation $A^{i}_{i'}$, transforms as

$$\phi_{\mu\nu i'}{}^{j'} = A^{i}_{i'} \phi_{\mu\nu i}{}^{j} A^{j'}_{j}. \tag{11}$$

In view of their geometric meaning, the parameters of connection $\Gamma_{\mu i}{}^{j}$ and the field $\phi_{\mu\nu i}{}^{j}$ can be regarded as generalized Yang–Mills potentials and Yang–Mills field, and we adopt the terminology of gauge potentials for $\Gamma_{\mu i}{}^{j}$ and gauge field for $\phi_{\mu\nu i}{}^{j}$, keeping in

mind that these fields are related to the Riemannian event geometry, a property which is not shared by the Yang–Mills field.

Covariant differentiation of (8) and alternation of the covariant derivatives yield the integrability condition

$$R_{\mu\nu\kappa}{}^{\rho}L_{i\rho}{}^{\lambda} - L_{i\kappa}{}^{\rho}R_{\mu\nu\rho}{}^{\lambda} = -\phi_{\mu\nu i}{}^{j}L_{j\kappa}{}^{\lambda}, \tag{12}$$

which the gauge field and the Riemann tensor must satisfy. Since a Riemann space is a metric space, the Riemann tensor is anti-self-adjoint in its last two indices and, hence, may be represented as a linear combination of $L_{i\kappa}{}^{\lambda}$:

$$R_{\mu\nu\kappa}{}^{\lambda} = B_{\mu\nu}{}^{k}L_{k\kappa}{}^{\lambda}. \tag{13}$$

Inserting this expression in (12) yields

$$B_{\mu\nu}{}^{k}(L_{k\kappa}{}^{\rho}L_{i\rho}{}^{\lambda} - L_{i\kappa}{}^{\rho}L_{k\rho}{}^{\lambda}) = -\phi_{\mu\nu i}{}^{j}L_{j\kappa}{}^{\lambda}. \tag{14}$$

With the commutation relations (3) and the linear independence of the $L_{i\kappa}{}^{\lambda}$, (14) shows that

$$\phi_{\mu\nu i}{}^{j} = -B_{\mu\nu}{}^{k}c_{ki}{}^{j}. \tag{15}$$

From (15), we see that the gauge field can be expressed as a linear combination of the structure constants and, hence, lies in the Lie algebra of local Lorentz transformations. According to (13) and (15), the Riemann tensor and the gauge fields are related to each other. In fact, the Lie algebra of the *event* holonomy group is spanned by the Riemann tensor and its covariant derivatives, and the gauge field and its covariant derivatives span a representation of the Lie algebra of the *internal* holonomy group which is just the adjoint representation of the Lie algebra of the event holonomy group. A full discussion of holonomy groups is given in Ref. 10. In a Riemann space the event holonomy group is contained in the Lorentz group, but may be a subgroup thereof. However, in physical applications where matter is present, the event holonomy group should be the full Lorentz group, since in such a simple situation as the spherically-symmetric gravitational field of an isolated particle, the event holonomy group has six parameters. Hence, we suppose that the event holonomy group is the Lorentz group and that the internal space is the 6-dimensional space spanned by the generators of local Lorentz transformations.

The symmetric tensor g_{ij} is defined by

$$g_{ij}(x^{\kappa}) = c_{ik}{}^{l}(x^{\kappa})c_{jl}{}^{k}(x^{\kappa}), \tag{16}$$

and may be used as a metric tensor in the internal space. Since the generators $L_{i\kappa}{}^{\lambda}$ are not constant then, according to (4), the structure constants need not be numerically constant as is indicated in (16). However, since $L_{i\kappa}{}^{\lambda}$ are covariant constant, we have

$$\nabla_{\mu}c_{ij}{}^{k} = 0 \tag{17}$$

with respect to the gauge potentials $\Gamma_{\mu i}{}^{j}$. From (16) and (17) it follows that

$$\nabla_{\mu}g_{ij} = 0. \tag{18}$$

The tensor g_{ij} is sometimes called the Cartan metric. Owing to the semisimplicity of the Lorentz group, g_{ij} is nonsingular and because of noncompactness it is indefinite. Since g_{ij} is nonsingular, it may be used to raise and lower indices and to define the norm of internal vectors. With g_{ij} as metric tensor, the internal space is a metric space, but with an indefinite metric with signature $(---+++)$ which may be verified by direct computation using the structure constants.

If the parameters of the Riemann connection are given, we may solve (8) for $\Gamma_{\mu i}{}^{j}$ in terms of $L_{i\kappa}{}^{\lambda}$ and $\Gamma_{\kappa\lambda}{}^{\rho}$. Multiplying (6) by $L_{k\lambda}{}^{\nu}$ and contraction of the event indices yield

$$L_{k\lambda}{}^{\kappa}\,{}^{e}\nabla_{\mu}L_{i\kappa}{}^{\lambda} = \Gamma_{\mu i}{}^{j}L_{k\lambda}{}^{\kappa}L_{j\kappa}{}^{\lambda}. \tag{19}$$

Since

$$L_{k\kappa}{}^{\lambda}L_{j\lambda}{}^{\kappa} = \lambda g_{kj}, \tag{20}$$

where λ is a nonvanishing real number, we can write (19) in the form

$$\Gamma_{\mu i}{}^{j} = \lambda^{-1}g^{kj}L_{k\lambda}{}^{\kappa}\,{}^{e}\nabla_{\mu}L_{i\kappa}{}^{\lambda}. \tag{21}$$

This shows that the gauge potentials (and, consequently, the gauge field) are determined uniquely by the Riemannian event connection.

There is a close similarity between the construction described here which gives the gauge potentials and the Lie algebra space of the local Lorentz group as an internal space and the construction by which the Dirac spin space and spin connection arise. Both the Dirac operators and anti-self-adjoint generators $L_{i\kappa}{}^{\lambda}$ are algebraically related to the event metric tensor. The spin connection, which defines equivalence of spinors at neighboring events, is defined through the requirement that the Dirac operators be covariant constant when covariant differentiation is extended to spinor indices; the gauge potentials, which provide the definition of equivalence at neighboring events of vectors in the algebra space of the Lorentz group, arise from the requirement that the generators of the local Lorentz group be covariant constant when covariant differentiation is extended to algebra indices.

III. GAUGE POTENTIALS AS FUNDAMENTAL FIELDS

If a gauge field is considered as a fundamental field with which the properties of matter are to be

described, it becomes of interest to investigate the extent to which a gauge field can determine the underlying Riemannian geometry. The main subject of this paper is the investigation of this "inverse problem." That is, we suppose a gauge field is *a priori* given and investigate the extent to which a Riemannian geometry is committed, when the relations between gauge potentials and the Riemannian geometry discussed in Sec. II are invoked. In particular, the uniqueness of a Riemannian geometry associated with a gauge field is considered.

The gauge field considered in this section could be restricted by some dynamical statement, for instance, by a field equation of the Yang–Mills type, but we shall not do so at this point, since the problem of uniqueness (and also of existence) can be investigated without specifying the dynamics of the field. However, we impose the algebraic requirement that the field and all of its covariant derivatives are contained in the Lie algebra of the Lorentz group. The holonomy group is then the Lorentz group or a subgroup thereof. We see that this insures that the gravitational field can be accounted for. In the internal space it is possible to have real gauge potentials which produce an internal holonomy group as large as the real linear group in six dimensions. Later, we consider an "enlargement" of the internal holonomy group beyond the Lorentz group in an attempt to introduce nongravitational properties of matter.

The starting point for the investigation of the inverse problem is a 4-dimensional geometric manifold as the event space with a 6-dimensional linear vector space at each event. The geometric structure of the event space is not completely specified initially, but we endow the event space with the Lorentz group structure constant field $c_{ij}{}^{k}$ satisfying the usual structural relations. As before, we use the Cartan metric for the internal metric tensor, and with g_{ij} the internal space becomes a metric space. We note that introducing the field $c_{ij}{}^{k}$ provides more geometrical structure than would be obtained by having just an internal metric at each event. This is evident because the field $c_{ij}{}^{k}$ is invariant under a 6-dimensional representation of the Lorentz group, while the internal metric field selects at each event a much larger group, the 15-dimensional group of local orthogonal transformations.

From this starting point, our motivation for introducing the gauge potentials is simply to provide the definition of equivalence of internal vectors at neighboring events, or, what is the same, the definition of the covariant derivative of internal quantities. The gauge potentials are restricted by the requirement that the structure constants are covariant constant as expressed by (17) and hence, as before, the Cartan metric is covariant constant. The first integrability condition for (18) is[14]

$$\phi_{\mu\nu(ik)} = 0, \tag{22}$$

which shows that gauge-field components belonging to gauge potentials satisfying (18) are anti-self-adjoint with respect to the internal metric.

A stronger integrability condition is obtained by covariant differentiation and alternation of (17). We obtain

$$\phi_{\mu\nu i}{}^{m}c_{mj}{}^{k} - \phi_{\mu\nu j}{}^{m}c_{im}{}^{k} + \phi_{\mu\nu m}{}^{k}c_{ij}{}^{m} = 0. \tag{23}$$

Multiplication of (23) by the structure constants, transvection over internal indices, and use of the definition of the Cartan metric leads to

$$\phi_{\mu\nu i}{}^{k} = -\phi_{\mu\nu j}{}^{l}c^{m}{}_{l}{}^{j}c_{mi}{}^{k}. \tag{24}$$

It follows that the gauge field is a linear combination of the structure constants

$$\phi_{\mu\nu i}{}^{m} = -B_{\mu\nu}{}^{k}c_{ki}{}^{m}, \tag{25}$$

where

$$B_{\mu\nu}{}^{k} = \phi_{\mu\nu}{}^{jl}c_{jl}{}^{k}. \tag{26}$$

Equation (25) shows that gauge-field components belonging to gauge potentials satisfying (17) are contained in the adjoint representation of the generators of local Lorentz transformations, guaranteeing that the internal holonomy group of the gauge potentials are a subgroup of the Lorentz group.

By a suitable choice of internal base we may make $c_{ij}{}^{k}$ numerically constant; this places no restriction on the structure constant field, but merely fixes the internal base up to the adjoint representation of a local Lorentz transformation. In such an internal base, (17) acquires the form

$$-\Gamma_{\kappa i}{}^{l}c_{lj}{}^{k} - \Gamma_{\kappa j}{}^{l}c_{il}{}^{k} + \Gamma_{\kappa l}{}^{k}c_{ij}{}^{l} = 0. \tag{27}$$

Following the same method as was used to show that a solution of (23) is a linear combination of the structure constants, we find

$$\Gamma_{\kappa i}{}^{m} = -b_{\kappa}^{k}c_{ki}{}^{m}, \tag{28}$$

with

$$b_{\kappa}^{k} = \Gamma_{\kappa j}{}^{l}c_{l}{}^{kj}. \tag{29}$$

It is now clear how to prescribe the gauge potentials in order to obtain an internal holonomy group which is a subgroup of the Lorentz group. In the special base where $c_{ij}{}^{k}$ are constants, one takes gauge potentials as linear combinations of the structure constants, as in (28), with arbitrary coefficient fields b_{κ}^{k}. By this prescription, we include all possible gauge

potentials whose internal holonomy group is contained in the Lorentz group.

Proceeding with the statement of the inverse problem, we note that the generators of local Lorentz transformations, related to the given structure-constant field by (3), are now considered as unknown fields. The map linking the internal space with the event space is determined, as before, by (8) and provides a differential equation which the $L_{i\kappa}{}^{\lambda}$ must satisfy. Writing out (8), we have

$$\partial_\mu L_{i\kappa}{}^{\lambda} - \Gamma^{\rho}_{\mu\kappa} L_{i\rho}{}^{\lambda} + \Gamma^{\lambda}_{\mu\rho} L_{i\kappa}{}^{\rho} - \Gamma_{\mu i}{}^{j} L_{j\kappa}{}^{\lambda} = 0, \quad (30)$$

where the gauge potentials $\Gamma_{\mu i}{}^{j}$ are now the given fields and we seek the event connection $\Gamma_{\mu\kappa}{}^{\lambda}$ as a solution of (30). The generators $L_{i\kappa}{}^{\lambda}$ cannot be chosen freely and appear as part of the solution. This can be seen as follows: Suppose the generator fields $L_{i\kappa}{}^{\lambda}$ satisfying (3) are prescribed. Then, from Eq. (2), we may determine the metric field that makes these generators anti-self-adjoint. Suppose there are two solutions of (2), say $h_{\kappa\lambda}$ and $g_{\kappa\lambda}$. Then it follows from (2) that

$$g^{\mu\lambda} L_{i\mu}{}^{\nu} g_{\nu\kappa} + L_{i\kappa}{}^{\lambda} = 0,$$
$$h^{\mu\lambda} L_{i\mu}{}^{\nu} h_{\nu\kappa} + L_{i\kappa}{}^{\lambda} = 0.$$

Subtracting these two equations, we obtain

$$h_{\mu\sigma} g^{\mu\lambda} L_{i\lambda}{}^{\nu} - L_{i\sigma}{}^{\lambda} h_{\mu\lambda} g^{\mu\nu} = 0. \quad (31)$$

Since the fundamental representation of the generators of Lorentz transformations form an irreducible set, Schur's lemma[15] applies and the solution of (31) has the form

$$h_{\mu\sigma} g^{\mu\lambda} = \mu \delta^{\lambda}_{\sigma}$$

or

$$h_{\mu\sigma} = \mu g_{\mu\sigma},$$

where μ is an arbitrary real number. Hence, the specification of the generators $L_{i\kappa}{}^{\lambda}$ determines the event metric field up to an arbitrary numerical factor.[16] Hence, if the $L_{i\kappa}{}^{\lambda}$ are chosen, the metric field calculated with the aid of (2) and the metric tensor related to the solution $\Gamma^{\lambda}_{\mu\kappa}$ of (30) through the Christoffel formula would not necessarily be in agreement.

In order to relate to the theory of general relativity, we shall consider solutions $\Gamma^{\lambda}_{\mu\kappa}$, $L_{i\kappa}{}^{\lambda}$ of (30) which give a Riemannian geometry. That is, we only admit solutions which satisfy (1) and

$$\Gamma_{[\kappa\lambda]}{}^{\mu} = 0. \quad (32)$$

The system of Eqs. (1)–(3), (30), and (32) determine the inverse map of gauge potentials restricted by (28) to the symmetric and metric linear connection of event space.

IV. UNIQUENESS OF SOLUTIONS OF THE INVERSE MAP

The uniqueness of the solutions is studied by assuming the existence of a solution and then investigating whether the governing equations admit solutions infinitesimally different from the assumed solution. It is clear from the outset that this method only exhibits continuous families of solutions and cannot reveal discretely different solutions. Let us assume that $\Gamma^{\lambda}_{\mu\kappa}$, $L_{i\kappa}{}^{\lambda}$, and $g_{\mu\kappa}$ satisfy the governing system of equations and then consider the fields

$$\Gamma'^{\lambda}_{\mu\kappa} = \Gamma^{\lambda}_{\mu\kappa} + \delta\Gamma^{\lambda}_{\mu\kappa},$$
$$L'_{i\kappa}{}^{\lambda} = L_{i\kappa}{}^{\lambda} + \delta L_{i\kappa}{}^{\lambda}, \quad (33)$$

infinitesimally different from the assumed solutions. In order that the primed objects also form a solution for fixed $\Gamma_{\mu i}{}^{j}$ and $c_{ij}{}^{k}$, the infinitesimal variations $\delta L_{i\kappa}{}^{\lambda}$ and $\delta\Gamma^{\lambda}_{\mu\kappa}$ must satisfy

$$[\delta L_i, L_j] + [L_i, \delta L_j] = c_{ij}{}^{k} \delta L_k, \quad (34)$$

$$\delta L_{i\kappa}{}^{\sigma} g_{\sigma\lambda} + L_{i\kappa}{}^{\sigma} \delta g_{\sigma\lambda} + \delta L_{i\lambda}{}^{\sigma} g_{\sigma\kappa} + L_{i\lambda}{}^{\sigma} \delta g_{\sigma\kappa} = 0, \quad (35)$$

$$\partial_\mu \delta L_i - [{}^{e}\Gamma_\mu, \delta L_i] - [\delta\, {}^{e}\Gamma_\mu, L_i] = \Gamma_{\mu i}{}^{j} \delta L_j, \quad (36)$$

$$\nabla_\mu \delta g_{\kappa\lambda} - \delta\Gamma^{\sigma}_{\mu\kappa} g_{\sigma\lambda} - g_{\kappa\sigma} \delta\Gamma^{\sigma}_{\mu\lambda} = 0, \quad (37)$$

$$\delta\Gamma_{[\mu\kappa]}{}^{\lambda} = 0, \quad (38)$$

where these equations are obtained by retaining only terms of first degree in the variations in Eqs. (1)–(3), (30), and (32). We have introduced matrix notation with L_i, for $L_{i\kappa}{}^{\lambda}$, and ${}^{e}\Gamma_\mu$, for $\Gamma^{\lambda}_{\mu\kappa}$.

Equation (34) is satisfied only if δL_i can be expressed in the form

$$\delta L_i = [K, L_i], \quad (39)$$

for an arbitrary infinitesimal operator K.

Use of (39) in Eq. (36) for fixed $\Gamma_{\mu i}{}^{j}$ yields

$$\nabla_\mu [K, L_i] = [\delta\, {}^{e}\Gamma_\mu, L_i], \quad (40)$$

where ∇_μ means covariant differentiation with respect to all indices. Since L_i is covariant constant, (40) may be written in the form

$$[\nabla_\mu K - \delta\, {}^{e}\Gamma_\mu, L_i] = 0. \quad (41)$$

The L_i form an irreducible set of matrices and, hence, by Schur's lemma,

$$\delta\Gamma^{\lambda}_{\mu\kappa} = \nabla_\mu K_\kappa{}^{\lambda} - a_\mu \delta^{\lambda}_{\kappa}, \quad (42)$$

where a_μ is an arbitrary infinitesimal vector field. Thus, if $\Gamma^{\lambda}_{\mu\kappa}$ is a solution of (36), then $\Gamma^{\lambda}_{\mu\kappa} + \delta\Gamma^{\lambda}_{\mu\kappa}$, with $\delta\Gamma^{\lambda}_{\mu\kappa}$ given by (42), is also a solution. According to (39), (35) will be satisfied only if

$$\delta g_{\kappa\lambda} = K_\kappa{}^{\sigma} g_{\sigma\lambda} + K_\lambda{}^{\sigma} g_{\sigma\kappa}. \quad (43)$$

Using the variations $\delta g_{\mu\nu}$ and $\delta\Gamma^{\kappa}_{\mu\nu}$ in (37) yields

$$a_\mu g_{\sigma\kappa} = 0.$$

Hence, the multiple of the identity in (42) is excluded and, consequently, the allowable variation in the parameters of the event connection is given by

$$\delta\Gamma^{\lambda}_{\mu\kappa} = \nabla_\mu K_\kappa{}^\lambda, \tag{44}$$

where $K_\kappa{}^\lambda$ is arbitrary.

To this point, we have shown that the system of equations excluding (38) which determines the mapping of the gauge potentials to the parameters of the event connection admits, for fixed $c_{ij}{}^k$ and $\Gamma_{\mu i}{}^j$, a continuous family of solutions $\Gamma^{\lambda}_{\mu\kappa} + \delta\Gamma^{\lambda}_{\mu\kappa}$ generated from a given solution by an operator $K_\kappa{}^\lambda$, which is arbitrary except for whatever smoothness restrictions one might wish to impose. The form of $\delta\Gamma^{\lambda}_{\mu\kappa}$ displayed in Eq. (44) shows that the family of connections $\Gamma^{\lambda}_{\mu\kappa} + \delta\Gamma^{\lambda}_{\mu\kappa}$ is not the result of a coordinate transformation and must be interpreted as geometrically different connections. Hence, we see that without invoking the torsion freeness of the event space, the map from the gauge potentials to the event connection is highly nonunique. If the event space is assumed to be torsionless, then from (38) it follows that

$$\nabla_{[\mu} K_{\kappa]}{}^\lambda = 0. \tag{45}$$

Equation (45) is a very strong restriction on $K_\kappa{}^\lambda$ and only in very special cases will nontrivial solutions exist.

Covariant differentiation of (45) and alternation of covariant indices yields

$$R_{[\nu\mu\,|\sigma|}{}^\lambda K_{\kappa]}{}^\sigma = 0. \tag{46}$$

Higher integrability conditions are satisfied identically, so we need consider only (46). Equation (46) is the same algebraic equation encountered by Loos[13] in the study of the mapping of parameters of the spin connection onto the parameters of a Riemannian event connection. It may be written as the following system of linear algebraic equations:

$$\begin{pmatrix} 0 & R^{\lambda}_{23\alpha} & R^{\lambda}_{31\alpha} & R^{\lambda}_{12\alpha} \\ R^{\lambda}_{23\alpha} & 0 & R^{\lambda}_{30\alpha} & R^{\lambda}_{02\alpha} \\ R^{\lambda}_{13\alpha} & R^{\lambda}_{30\alpha} & 0 & R^{\lambda}_{01\alpha} \\ R^{\lambda}_{12\alpha} & R^{\lambda}_{20\alpha} & R^{\lambda}_{01\alpha} & 0 \end{pmatrix} \begin{pmatrix} K^{\alpha}_0 \\ K^{\alpha}_1 \\ K^{\alpha}_2 \\ K^{\alpha}_3 \end{pmatrix} = 0. \tag{47}$$

From the form of the matrix of coefficients M, we see that the rank is not greater than 15; so there is at least one solution, namely

$$K^{\lambda}_{\kappa} = a\delta^{\lambda}_{\kappa}. \tag{48}$$

It is evident from (39) that for $K_\kappa{}^\lambda$ given by (48) the generators are invariant. Inserting (48) in the differential equation (45) yields

$$\delta^{\lambda}_{\kappa}\partial_\mu a - \delta^{\lambda}_{\mu}\partial_\kappa a = 0.$$

This equation implies that a is a constant. Then, according to (44), $\delta\Gamma^{\lambda}_{\mu\kappa}$ vanishes. The effect of the variation generated by (48) with constant a is to multiply the metric tensor by a constant factor and thereby represents a nonuniqueness in the inverse map which leaves the scale of the metric field undetermined. Combining this result with the results of Sec. II we have: *If the rank of M has the maximum value 15, the map between gauge potentials $\Gamma_{\mu i}{}^j$ and the parameters of the event connection $\Gamma^{\lambda}_{\mu\kappa}$ is one-to-one in a neighborhood of $\Gamma^{\lambda}_{\mu\kappa}$, and the metric field is determined up to a uniform scale transformation.* Our method, which is by infinitesimal variation of an assumed solution, does not rule out the possibility of discretely different Riemannian geometries corresponding to the same gauge potentials.

Loos[13] has shown that the identities satisfied by the Riemann tensor do not force the rank of M to be less than 15 and, hence, cases where this does occur should be considered exceptional. An example of such an exceptional case is the Schwarzschild metric.[17] For the Schwarzschild metric, use of nonvanishing components of the Riemann tensor in (47) shows by inspection that the rank of M is 12 or less and that generators of the form

$$K_\kappa{}^\lambda = k_\kappa \delta^{\lambda}_{\kappa}, \tag{49}$$

with no sum on κ and k_κ, an arbitrary vector field, are a solution of (46). In order for solutions, in addition to (49) to exist, it is necessary for the subdeterminant obtained by deleting the four rows and columns of zeros in M to vanish in a region of event spaces. This does not occur for components of the Riemann tensor computed from the Schwarzschild metric. This example suggests that the mapping between gauge potentials and parameters of the event connection fails to be locally one-to-one only for fields of high symmetry. For less symmetry, for example in a scattering situation, or if particle internal structure would be represented, we would expect the map to be one-to-one.

We have seen that a necessary condition for the existence of solutions of the inverse map which differ by more than a uniform scale transformation of the metric field is that the rank of the matrix M in (47) be less than 15. A condition which is sufficient for the existence of a nonuniqueness in the inverse mapping is demonstrated as follows. Let the generator

$K_\kappa{}^\lambda$ be of the form

$$K_\kappa{}^\lambda = \nabla_\kappa v^\lambda. \tag{50}$$

Substitution of (50) in (45) yields

$$\nabla_{[\mu}\nabla_{\kappa]}v^\lambda = 0. \tag{51}$$

But

$$\nabla_{[\mu}\nabla_{\kappa]}v^\lambda = \tfrac{1}{2}R_{\mu\kappa\nu}{}^\lambda v^\nu. \tag{52}$$

Hence, if the Riemann tensor has eigenvectors with zero eigenvalues, there are solutions of (45) of the form (50). These solutions represent a nonuniqueness in the mapping from the gauge potentials to the Riemannian geometry. If the Riemann tensor has a zero eigenvalue, the rank of M is less than 15, in agreement with the above integrability conditions.

It is perhaps of interest to examine the consequence of deleting the metric condition (1) from the mapping equations. According to (2) and (8) we must have

$$L_{i\kappa}{}^\lambda\nabla_\nu g_{\lambda\mu} + L_{i\mu}{}^\lambda\nabla_\nu g_{\lambda\kappa} = 0. \tag{53}$$

A metric field subject to the condition of the Weyl geometry, namely,

$$\nabla_\nu g_{\lambda\mu} = Q_\nu g_{\lambda\mu}, \tag{54}$$

satisfies (53) and is not excluded by (2).

The integrability condition (12) of (8) admits a curvature tensor of the form

$$R_{\mu\nu\kappa}{}^\sigma = {}^*B_{\mu\nu}{}^i L_{i\kappa}{}^\sigma + \nabla_{[\mu}Q_{\nu]}\delta_\kappa^\sigma, \tag{55}$$

which is compatible with (54). The uniqueness proof then goes through exactly as before, showing that we obtain the same uniqueness statements for the mapping between gauge potentials and a Weyl geometry as obtained for the Riemannian geometry.

V. YANG–MILLS AND EINSTEIN FREE-FIELD EQUATIONS

The gauge potentials considered thus far in the inverse problem have not been assumed to satisfy field equations; they were restricted only by algebraic conditions which assured that the holonomy group of the gauge potentials would be a subgroup of the Lorentz group. Let us now go on to investigate gauge fields which are subject to field equations, while maintaining the link between the gauge field and the event geometry which is established by requiring the covariant constancy of the generators of local Lorentz transformations. Only free gauge fields are considered, as we do not couple the gauge fields to particle current densities. For free fields the possible Lagrangians are scalar densities constructed from the gauge potentials. We consider the simplest Lagrangian, namely,

$$\tilde{\mathfrak{L}} = \tfrac{1}{4}\tilde{g}g^{\kappa\mu}g^{\lambda\nu}\phi_{\kappa\lambda i}{}^j\phi_{\mu\nu j}{}^i, \tag{56}$$

where

$$\tilde{g} = |\det g_{\kappa\lambda}|^{\frac{1}{2}}. \tag{57}$$

Under a coordinate transformation $A^\lambda_{\lambda'}$, $\tilde{g}$ transforms according to

$$\tilde{g}' = |\det A^\lambda_{\lambda'}|^{\frac{1}{2}}\,\tilde{g}, \tag{58}$$

and thus provides the density character to the Lagrangian. The gauge field equation is obtained from the postulate of stationary action under arbitrary variation of the gauge potentials for a fixed metric field. For variations which vanish outside a finite region of event space, we obtain the field equation

$$\partial_\kappa\tilde{g}\phi^{\kappa\lambda}{}_i{}^j - \tilde{g}(\Gamma_{\kappa i}{}^k\phi^{\kappa\lambda}{}_k{}^j - \phi^{\kappa\lambda}{}_i{}^k\Gamma_{\kappa k}{}^j) = 0. \tag{59}$$

However, the left-hand side of (59) is not a quantity unless the event connection is symmetric. Making the independent assumption that the event connection is symmetric, we obtain a field equation which transforms like a geometric quantity[18] without further *a priori* assumptions about the metric properties of the event space.

Originally, the Yang–Mills equations were defined on a Minkowskian event space, but here we are interested in the case where the event space is not flat. The event metric tensor appears in (59) through $\tilde{g}$ and the contravariant indices of $\phi^{\kappa\lambda}{}_i{}^j$ and, hence, the metric tensor and gauge potentials are interdependent even without having the map between gauge potentials and parameters of the event connection.

To establish the inverse map, we proceed as before by requiring that the generators $L_{i\kappa}{}^\lambda$ be covariant constant as expressed by (8). For a Riemann space the integrability conditions for (8) imply that (13) and (15) are satisfied. With (13) and (15), using the covariant constancy of $L_{ik}{}^\lambda$, we see from the field equation (59) that

$$\nabla_\nu R_\mu{}^\nu{}_\kappa{}^\lambda = 0, \tag{60}$$

which for a Riemann space implies

$$\nabla_\lambda R_{\mu\nu\kappa}{}^\lambda = 0. \tag{61}$$

Transvection of the Bianchi identity satisfied by the Riemann tensor and use of (61) give, for the Ricci tensor,

$$\nabla_{[\nu}R_{\mu]\kappa} = 0. \tag{62}$$

Equation (61) implies the weaker condition

$$\nabla_\lambda R_\nu{}^\lambda = 0, \tag{63}$$

while contraction of (62) yields

$$\nabla_\nu R = 0, \tag{64}$$

where R is the scalar curvature. So we see that when the inverse map between the gauge potentials and the parameters of the event connection is invoked, the field equations for the gauge potentials imply that Eqs. (60)–(64) hold for the Riemann tensor, the Ricci tensor, and the scalar curvature, respectively. Hence, the gauge-field equation together with the inverse map imply that the event space is a space of constant curvature. But, a Riemann space of constant scalar curvature need not be an Einstein space.[12] However, if we require that the inverse map determines the event geometry uniquely except for a constant scale transformation of the metric field, then the only solutions of (62) are

$$R_{\mu\nu} = bg_{\mu\nu},$$

where b is an arbitrary scalar factor.[19] Hence, parameters of the event connection, which are in one-to-one correspondence to gauge potentials which satisfy the free Yang–Mills equations, are automatically solutions of the empty-space Einstein equations with arbitrary cosmological constant. But there are solutions of the Einstein equation for empty space which do not satisfy the subsidiary requirement that the inverse map have a unique solution, the Schwarzschild metric providing such an example. So the Yang–Mills type field equation together with the subsidiary requirement of uniqueness admits a smaller set of solutions than the Einstein field equation.

VI. ENLARGEMENT OF THE HOLONOMY GROUP

If one insists that solutions of the inverse map be unique, then it appears that, to go beyond the gravitational theory in this mathematical framework, we must consider gauge potentials of a more general kind. At the same time, we wish to maintain a correspondence to the Riemannian geometry so that gravitational phenomena can be included. To accomplish this, we allow gauge potentials having a semisimple holonomy group which is larger than the Lorentz group. Then the map in the present form can no longer be maintained. That is, parameters of a Riemannian event connection such that the generators of local Lorentz transformation are covariant constant do not exist if $\phi_{\mu\nu i}{}^{j}$ and their covariant derivatives are not contained in the Lie algebra $\mathfrak{L}$ of the Lorentz group.

Let us again consider the Lagrangian density (56) and the field equation (59) for a symmetric event connection. We supplement the action principle with the modified mapping equation

$$^{e}\nabla_{\mu} L_{i\kappa}{}^{\lambda} = {}^{*}\Gamma_{\mu i}{}^{j} L_{j\kappa}{}^{\lambda}, \tag{65}$$

where ${}^{*}\Gamma_{\mu i}{}^{j}$ is that part of the gauge potential which belongs to $\mathfrak{L}$ in a canonical base system, i.e., in an internal base and an algebra base such that the total gauge potentials are contained in the Lie algebra of the internal holonomy group.[20] This insures that in any internal base the holonomy group of ${}^{*}\Gamma_{\mu i}{}^{j}$ is a subgroup of the Lorentz group. In the canonical base system

$$\Gamma_{\mu i}{}^{j} = {}^{*}\Gamma_{\mu i}{}^{j} + Q_{\mu i}{}^{j}, \tag{66}$$

where $Q_{\mu i}{}^{j}$ is not contained in $\mathfrak{L}$. When the gauge potentials are expressed in the form (66), the gauge field decomposes as

$$\phi_{\mu\nu} = \phi^{*}_{\mu\nu} + \psi_{\mu\nu}, \tag{67}$$

where

$$\phi^{*}_{\mu\nu} = \partial_{\mu} {}^{*}\Gamma_{\nu} - \partial_{\nu} {}^{*}\Gamma_{\mu} - [{}^{*}\Gamma_{\mu}, {}^{*}\Gamma_{\nu}], \tag{68}$$

$$\psi_{\mu\nu} = \partial_{\mu} Q_{\nu} - \partial_{\nu} Q_{\mu} - [Q_{\mu}, Q_{\nu}] - [{}^{*}\Gamma_{\mu}, Q_{\nu}] - [Q_{\mu}, {}^{*}\Gamma_{\nu}]. \tag{69}$$

The gauge field $\phi^{*}_{\mu\nu}$ is contained in $\mathfrak{L}$ and $\psi_{\mu\nu}$ is not. The map between gauge potentials and the event geometry is now established through

$$\partial_{\mu} L_{i\kappa}{}^{\lambda} - {}^{*}\Gamma^{\rho}_{\mu\kappa} L_{i\rho}{}^{\lambda} + {}^{*}\Gamma^{\lambda}_{\mu\rho} L_{i\kappa}{}^{\rho} - {}^{*}\Gamma_{\mu i}{}^{j} L_{j\kappa}{}^{\lambda} = 0. \tag{70}$$

The relation between the Riemann tensor and the part of the gauge field which belongs to $\mathfrak{L}$ is shown by

$$R_{\mu\nu\kappa}{}^{\lambda} = {}^{*}B_{\mu\nu}{}^{i} L_{i\kappa}{}^{\lambda}, \tag{71}$$

$${}^{*}\phi_{\mu\nu i}{}^{j} = -{}^{*}B_{\mu\nu}{}^{k} c_{ki}{}^{j}. \tag{72}$$

Invoking a unified action principle for the total gauge field and using the same Lagrangian density (56) as before, we see that the field equation is again (59). Decomposing the gauge potentials and field into Lorentzian and non-Lorentzian parts, which is understood to be done in a canonical base, we obtain from (59) the field equation

$${}^{*}\nabla_{\mu} {}^{*}\phi^{\mu\lambda} = J^{\lambda}, \tag{73}$$

where

$$J^{\lambda} = -\nabla_{\mu}\psi^{\mu\lambda} - [Q_{\mu}, {}^{*}\phi^{\mu\lambda}] \tag{74}$$

and ${}^{*}\nabla_{\mu}$ and ∇_{μ} are covariant derivatives with respect to event indices and gauge covariant with respect to ${}^{*}\Gamma_{\mu i}{}^{j}$ and the total gauge potential $\Gamma_{\mu i}{}^{j}$. Equation (73) implies that

$$\nabla_{\nu} R_{\kappa\lambda\mu}{}^{\nu} = J_{\mu\kappa\lambda} \tag{75}$$

holds rather than (61), where

$$J_{\mu\kappa\lambda} = j^{i}_{\mu} L_{i\kappa\lambda}, \tag{76}$$

$$j_{\mu i} = c_{ik}{}^{j} J_{\mu j}{}^{k}. \tag{77}$$

Using Bianchi's identity and (75), we obtain, after

contraction,

$$\nabla_\lambda R = -2J^\kappa{}_{\kappa\lambda}. \tag{78}$$

Equation (78) shows that the event space is not empty and, in this sense, the enlargement of the internal holonomy group results in a source of the gravitational field. If $J_{\mu i}{}^j$ vanishes, then $J_{\mu\kappa}{}^\lambda$ also vanishes, the internal holonomy group is the Lorentz group or a subgroup thereof and the parameters of the event connection satisfy the Einstein field equation for an empty space.

It is natural to identify the Lorentzian part of the gauge potentials with the gravitational field and the non-Lorentzian part with matter which acts as the source of the gravitational field. Regular localized solutions of (73), if they exist, could perhaps be interpreted as a classical field description of an extended particle with internal structure.

If a free-gauge field has an internal holonomy group which is a direct product of two semisimple groups, then the field can be decomposed into two noninteracting fields.[10] That is, the gauge field can be decomposed into parts that satisfy uncoupled field equations. If the gravitational part of the internal holonomy group is the adjoint representation of the full Lorentz group, then the only possible enlargement that gives an enlarged group which is a direct product of semisimple groups is obtained by adding a multiple of the identity to the gravitational part of the gauge potentials, i.e.,

$$\Gamma_\mu = {}^*\Gamma_\mu + Q_\mu I. \tag{79}$$

We then obtain the gauge field

$$\phi_{\mu\nu} = \phi^*_{\mu\nu} + (\partial_\mu Q_\nu - \partial_\nu Q_\mu)I. \tag{80}$$

Inserting (79) and (80) in (59) yields

$$\partial_\mu \tilde{g}\, {}^*\phi^{\mu\nu} - \tilde{g}[{}^*\Gamma_\mu, \phi^{\mu\nu}] + \partial_\mu \tilde{g}(\partial^\mu Q^\nu - \partial^\nu Q^\mu)I = 0.$$

The terms involving ${}^*\phi^{\mu\nu}$ and Q^ν belong to disjoint algebras, so we must have separately

$$\partial_\mu \tilde{g}\, {}^*\phi^{\mu\nu} - \tilde{g}[{}^*\Gamma_\mu, {}^*\phi^{\mu\nu}] = 0, \tag{81}$$

$$\partial_\mu \tilde{g}(\partial^\mu Q^\nu - \partial^\nu Q^\mu) = 0. \tag{82}$$

We have shown that if the mapping to the Riemannian geometry is unique up to uniform scale transformation, the field equation (81) implies the free-gravitational field equation

$$R_{\mu\nu} = b g_{\mu\nu}. \tag{83}$$

But a free-gravitational field is in conflict with the presence of another field which is the carrier of energy and momentum. Hence, in the case of a unique mapping we must regard the enlargement of the internal holonomy group by the direct product of a semisimple group with the adjoint representation of the full Lorentz group as physically untenable.

VII. DISCUSSION

The central ideas behind this investigation are that Yang–Mills-type gauge potentials should be fundamental in a field theory of elementary particles and that internal and space–time degrees of freedom should be somehow linked together. In a Riemannian event space, the Pauli spin space, the Dirac spin space, and the space spanned by generators of local Lorentz transformations (Lie algebra $\mathfrak{L}$), all provide mathematical frameworks where these ideas can be studied. The parameters of linear connection for vectors in the respective spaces all transform under local base transformations in the same way as the Yang–Mills potentials in isospin space and can be regarded as gauge potentials. Here the investigation has been confined to gauge potentials acting in $\mathfrak{L}$. This linear vector space and the gauge potentials acting in it arise in much the same way as the Dirac spin space and spin connection arise from a Riemannian event space. However, this mathematical structure and its application in physical theory have not been specifically investigated as they have in the case of the Dirac spin space.

If gauge potentials in the algebra space are considered fundamental, then it is natural to inquire into the extent to which the gauge potentials commit the underlying Riemannian geometry. The present investigation has been concerned mainly with this question. The generators of local Lorentz transformations $L_{i\kappa}{}^\lambda(x^\mu)$ provide the connecting link between internal and space–time degrees of freedom. When the $L_{i\kappa}{}^\lambda(x^\mu)$ are required to be covariant constant with respect to event and algebra indices, a very close correspondence is established between the gauge potentials and the parameters of the event connection. In fact, we have shown that the correspondence is almost always one-to-one in a neighborhood of parameters of the event connection $\Gamma^\lambda_{\mu\kappa}$ which satisfy the mapping equations, with the metric field remaining undetermined only by a uniform scale transformation. An algebraic condition is displayed which if satisfied insures such a locally one-to-one map. These results hold independently of whether the gauge potentials satisfy field equations. However, if the gauge potentials are subject to an algebraic restriction which insures that the internal holonomy group of the gauge potentials is the Lorentz group or a subgroup thereof, if the gauge potentials satisfy field equations of the free Yang–Mills type, and if the map is locally

one-to-one, then it follows that the parameters of the event connection necessarily satisfy the Einstein equations for empty space with arbitrary cosmological constant.[18]

This result suggested that gauge potentials of this class have only to do with the gravitational field. To introduce nongravitational fields, one could introduce particle fields which would appear as a source in the Yang–Mills field equation, but which would not have any direct geometric meaning. In this case, one can also establish a locally one-to-one map between gauge potentials and parameters of the event connection. Then the Yang–Mills equations no longer imply an empty Einstein space. However, it seemed appropriate to the spirit of this investigation to try to introduce nongravitational fields without introducing particle fields of a nongeometrical nature. It is shown that it is possible to do this by considering a more general class of gauge potential. When gauge potentials with internal holonomy groups not contained in the Lorentz group are allowed, the gauge potentials separate into a part contained in $\mathfrak{L}$ and a part not in $\mathfrak{L}$. Invoking the free Yang–Mills equations for the total gauge potentials, the field equation splits into a Lorentzian part and a non-Lorentzian part with the split being invariant if local gauge transformations are restricted to elements of the Lorentz group. The non-Lorentzian part contributes to a current which acts as a source of the Lorentzian part of the gauge potentials, which we identify as the gravitational field. When a modified map between gauge potentials and parameters of a Riemannian event connection is invoked, the Yang–Mills equations with "source" imply that the Riemann curvature tensor satisfies the Yang–Mills type of equation

$$\nabla_\mu R_\nu{}^\mu{}_\kappa{}^\lambda = J_{\nu\kappa}{}^\lambda,$$

with a tensor source $J_{\nu\kappa}{}^\lambda$. A nonvanishing current implies an event space of variable scalar curvature and, hence, that the Einstein field equations for empty space are not satisfied. If one acknowledges only the event geometry, this is merely a theory with a metric field plus a nongeometrical tensor field. However, there is a geometric unification if one acknowledges the gauge potentials acting in the internal space and their relation to the Riemann geometry. Moreover, the Yang–Mills equation for gauge potentials with an enlarged internal holonomy group is not equivalent to considering only the event space with a metric field plus a tensor field $J_{\nu\kappa}{}^\lambda$. The theory including the gauge potentials contains additional fields and differential relations which, however, have not been considered here.

ACKNOWLEDGMENT

It is a pleasure to thank Dr. H. G. Loos for many useful and interesting discussions and for much encouragement.

* Based on part of the author's Ph.D. thesis, University of California, Riverside.

† Present address: West Virginia University, Morgantown, West Virginia.

[1] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[2] R. Utiyama, Phys. Rev. **101**, 1597 (1956).

[3] R. König, J. Deut. Math. **28**, 213 (1920).

[4] H. Tetrode, Z. Physik **49**, 858 (1928); **50**, 336 (1928); E. Wigner, *ibid.* **53**, 592 (1928); H. Weyl, *ibid.* **56**, 330 (1929); V. Fock, *ibid.* **57**, 261 (1929); Schrödinger, Sitzber. Preuss. Akad. Wiss. Kl. Physik-Math. XI, 105 (1932).

[5] We cite here a few papers where further references may be found; J. A. Schouten, Proc. Kon. Akad. Amsterdam **27**, 407 (1924); A. W. Tucker, Ann. Math. **32**, 451 (1931); K. Yano, Proc. Japan Acad. **21**, 156, 164 (1945).

[6] A. Einstein and W. Mayer, Akad. Wiss. Berlin Kl. Physik-Math., **22**, 541 (1931).

[7] H. G. Loos, Nucl. Phys. **72**, 677 (1965).

[8] H. Weyl, *Space–Time–Matter* (Dover, New York, 1950), p. 203.

[9] H. G. Loos, Ann. Phys. (N.Y.) **36**, 486 (1966).

[10] H. G. Loos, J. Math. Phys. **8**, 2114 (1967).

[11] R. P. Treat, Phys. Rev. Letters **12**, 407 (1964).

[12] J. A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954).

[13] H. G. Loos, Ann. Phys. (N.Y.) **25**, 91 (1963).

[14] Round brackets around indices denote mixing: $T_{(ij)} = (1/2!) \times (T_{ij} + T_{ji})$. Square brackets denote alternation: $T_{[ij]} = (1/2!) \times (T_{ij} - T_{ji})$. When indices are to be excluded from mixing or alternation we use vertical lines, e.g., $T_{[i\,|j|\,k]} = (1/2!)(T_{ijk} - T_{kji})$ and this notation is used for both event and internal indices. We follow generally the notation of Ref. 12.

[15] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946), p. 83.

[16] From this result it follows that, if the Riemann tensor forms an irreducible set of matrices, the Riemann tensor determines the metric field algebraically up to an arbitrary numerical factor.

[17] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U.P., New York, 1960), p. 141.

[18] By definition, a geometric quantity is an object which, under a coordinate or base transformation, has new components which are linear homogeneous in the old components and homogeneous algebraic in the elements of the transformation matrix.

[19] H. G. Loos and R. P. Treat, Phys. Letters **26A**, 91 (1967).

[20] Such a base always exists; see Theorems 8 and 9, Ref. 10.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Anholonomic Bases in the Study of the Spin Connection and Gauge Potentials in a Riemann Space*

RICHARD P. TREAT
University of California at Riverside, and *West Virginia University, Morgantown, West Virginia 26506*†

(Received 29 December 1969)

If parameters of a spin connection have a holonomy group which is the Lorentz group or a subgroup thereof and if a certain algebraic restriction involving the Riemann tensor, say $\mathcal{R}$, is satisfied, then there is a one-to-one relation between the spin connection ${}^{s}\Gamma_{\mu a}{}^{b}$ and parameters of a Riemannian event connection $\Gamma_{\mu\kappa}{}^{\lambda}$, in a neighborhood of an assumed solution of the equations which establish the correspondence [see H. G. Loos, Ann. Phys. **25**, 91 (1963)]. We have, by the use of anholonomic bases in the Riemann space, simplified the equations which determine the map and have shown that if the algebraic condition $\mathcal{R}$ is satisfied, a one-to-one correspondence holds also in the large. The same method, involving anholonomic bases, has also been used to study the map between gauge potentials acting in the internal space formed by generators of local Lorentz transformations and parameters of connection of a Riemannian event space. An analogous theorem is proved which asserts that if $\mathcal{R}$ is satisfied, then the correspondence between the gauge potential and the parameters of the event connection is one-to-one.

I. INTRODUCTION

One may introduce parameters of a linear connection (the spin connection) for objects represented in the spin space, but only in a manner which is compatible with the quadratic decomposition of the metric field by the Dirac operators. One way in which this may be done is to require that the Dirac operators be covariant constant with respect to both spinor and event indices.[1] In this theory, the Dirac operators play the role of "connecting quantities" (not to be confused with parameters of a linear connection) by virtue of their having both spinor and event indices. Their existence and the requirement of covariant constancy forces certain relationships between the spin space and the tangent space to the event space. Similar structures also exist for the Pauli spin space and the space formed by the generators of local Lorentz transformations.[2,3] The Pauli operators $\sigma_{\kappa}(x^{\mu})$ "connect" the Pauli spin space and the Riemannian event space and the Pauli spin connection is introduced and its relationship to the Riemannian geometry established by requiring that the Pauli operators be covariant constant. The six linearly-independent local generators $L_{i\kappa}{}^{\lambda}(x^{\mu})$, $i, j, \cdots = 1, 2, \cdots, 6$, span the internal space formed by generators of local Lorentz transformations, which we will denote by $L(g_{\kappa\lambda})$, and provide the connecting quantities in this case. The parameters of linear connection (gauge potentials) acting in $L(g_{\kappa\lambda})$ are introduced and a map to the parameters of the event connection is established by requiring the covariant constancy of the $L_{i\kappa}{}^{\lambda}(x^{\mu})$.

If the spin connection or gauge potentials are to be considered as fundamental fields, it is natural to inquire into the uniqueness of the event geometry which they determine when the above maps are invoked. Loos[4] has posed and studied this so-called inverse problem, where he has considered the spin connection as a fundamental field and then investigated the extent to which a given, fixed spin connection determines the local Dirac operator field $\gamma_{\kappa}(x^{\mu})$ and the associated Riemannian event geometry. An investigation by Peres[5] of the Pauli and Dirac spin connections in a curved space has bearing on this problem, but from a different point of view. The analogous problem, where gauge potentials acting in $L(g_{\kappa\lambda})$ are assumed given, has also been studied.[6] In Ref. 4, algebraic conditions involving the Riemann tensor are shown, which, if satisfied, insure that the correspondence between parameters of the spin connection and event connection is one-to-one in a neighborhood of an assumed solution of the mapping equations (locally one-to-one). The Riemannian metric remains undetermined only by a uniform scale transformation. In Ref. 6, the uniqueness of the correspondence between the gauge potentials and parameters of the event connection was studied and the same algebraic condition was shown also to guarantee a locally one-to-one map. In both cases, the question of whether different Riemannian geometries, not infinitesimally close or not related by a continuous transformation, could correspond to the same spin connection or to the same gauge potentials was left open.

In Sec. II we recall the elements of the theory of a spin connection in a Riemannian event space and summarize results of prior work which are relevant here. At the same time we introduce our notation and mention some points which may not be entirely familiar. In Sec. III we go on to study, for the case of

the spin connection, the same inverse problem considered in Ref. 4, but by a different method, which employs anholonomic bases in the Riemann event space. In Sec. IV, the method using anholonomic bases is applied to the problem involving the gauge potentials in $L(g_{\kappa\lambda})$, where we state the problem, indicate the method of solution briefly, and cite the results.

II. THE SPIN CONNECTION IN A RIEMANN SPACE

We consider a 4-dimensional event space with time-space coordinates x^κ, $\kappa, \lambda \cdots = 0, 1, 2, 3$, which is endowed with parameters of a symmetric linear connection $\Gamma_{\mu\kappa}{}^{\lambda}$ and a symmetric, covariant-constant metric field $g_{\kappa\lambda}$. The "Dirac operators" are defined at each event by the quadratic decomposition[7]

$$\gamma_{(\mu}\gamma_{\nu)} = g_{\mu\nu}(x^\kappa)I, \tag{1}$$

where I is the identity. The solutions $\gamma_\mu(x^\kappa)$ of Eq. (1) are represented irreducibly at each event in local 4-dimensional Dirac spin spaces. According to Eq. (1), in a curved space the γ_μ are necessarily functions of the coordinates.

It is useful in this study to introduce the spin metric explicitly. The spin metric is Hermitian, a property which in index notation is expressed by

$$g_{a\bar{b}} = g_{\bar{b}a}. \tag{2}$$

We will use Latin indices $a, b, \cdots = 1, 2, 3, 4$ to denote components of objects represented in the spin space. Under a change of local spin base, $a \to a'$, we have the transformation matrix $S = S_a^{a'}$ and its inverse $S^{-1} = S_{a'}^{a}$. Unbarred contra- and covariant indices transform according to S and S^{-1}, respectively, and barred indices transform by the complex-conjugate matrices. We are using bars above indices to specify their transformation rule where dots are often used.[8] We require that the Dirac operators satisfying Eq. (1) be self-adjoint with respect to the spin metric, i.e.,

$$\gamma_{\mu a}{}^{b} g_{b\bar{c}} - \gamma_{\mu\bar{c}}{}^{\bar{b}} g_{\bar{b}a} = 0, \tag{3}$$

where $\gamma_{\mu\bar{a}}{}^{\bar{b}}$ is the complex conjugate of $\gamma_{\mu a}{}^{b}$. On account of the irreducibility of the γ_μ, Eq. (3), for a given fixed Dirac operator field, determines the spin metric $g_{a}{}^{\bar{b}}$ up to a multiple of a real number. We construct from the Dirac operators the 16 linearly independent operators

$$iI, \quad i\gamma_\mu, \quad \gamma_{[\mu}\gamma_{\nu]}, \quad \gamma_\mu\gamma_5, \quad i\gamma_5. \tag{4}$$

In Eq. (4),

$$\gamma_5 = (1/4!)\,|\det g_{\sigma\rho}|^{\frac{1}{2}}\, \tilde{\epsilon}^{\mu\nu\kappa\lambda}\gamma_\mu\gamma_\nu\gamma_\kappa\gamma_\lambda, \tag{5}$$

where $\tilde{\epsilon}^{\mu\nu\kappa\lambda}$ is the totally antisymmetric tensor Δ-density of weight 1. Real linear combinations of the 16 elements (4) form the Lie algebra $L(g_{a\bar{b}})$ of the unitary group $U(g_{a\bar{b}})$ of transformations which leave $g_{a\bar{b}}$ invariant.

The definition of equivalence of spinors at neighboring events is given by parameters of a linear connection, the spin connection $\Gamma_{\mu a}{}^{b}$. By virtue of the inhomogeneous transformation law

$$\Gamma'_\mu = S^{-1}\Gamma_\mu S - S^{-1}\partial_\mu S, \tag{6}$$

this definition of equivalence, or, what is the same, spin covariant differentiation, is invariant under local spin base transformation, where we are using the matrix notation $\Gamma_\mu = \Gamma_{\mu a}{}^{b}$, $\Gamma'_\mu = \Gamma_{\mu a'}{}^{b'}$.

We insist that the norm $\psi^a\psi_a$ of arbitrary equivalent spinors be invariant. This condition is satisfied if and only if

$$\nabla_\mu g_{a\bar{b}} \equiv \partial_\mu g_{a\bar{b}} - \Gamma_{\mu a}{}^{c} g_{c\bar{b}} - \Gamma_{\mu\bar{b}}{}^{\bar{c}} g_{\bar{c}a} = 0. \tag{7}$$

If the spin curvature

$$\Phi_{\mu\nu} = \partial_\mu\Gamma_\nu - \partial_\nu\Gamma_\mu - [\Gamma_\mu, \Gamma_\nu] \tag{8}$$

and its covariant derivatives $\nabla_\kappa\Phi_{\mu\nu}, \cdots, \nabla_\omega \cdots \nabla_\kappa\Phi_{\mu\nu}$ belong to $L(g_{a\bar{b}})$, and if, consequently, the spin holonomy group[9,10] is $U(g_{a\bar{b}})$ or a subgroup of $U(g_{a\bar{b}})$, then (7) is satisfied.

It is sometimes convenient to use special bases in spin space. In particular it is convenient and possible to choose a base at each event such that the spin connection belongs to the Lie algebra of the holonomy group, which at this point of the discussion is $U(g_{a\bar{b}})$ or a subgroup thereof. In such bases the spin metric is numerically constant, as can be deduced from (7). We call these normal bases.

In a normal basis, $\Gamma_{\mu a}{}^{b}$ belongs to $L(g_{a\bar{b}})$ and is therefore anti-self-adjoint with respect to $g_{a\bar{b}}$, i.e.,

$$\Gamma_{\mu a}{}^{c} g_{c\bar{b}} + \Gamma_{\mu\bar{b}}{}^{\bar{c}} g_{\bar{c}a} \overset{*}{=} 0, \tag{9}$$

where the $\overset{*}{=}$ is used to indicate that the equality is asserted only in normal bases. Now (9) is satisfied only if the trace part of Γ_μ calculated in a normal basis is purely imaginary.[11] Then in any base, we have

$$\Gamma_\mu = {}^{s}\Gamma_\mu + ia_\mu - \partial_\mu a, \tag{10}$$

where ${}^{s}\Gamma_\mu$ is the traceless part of the spin connection, a_μ is a real vector field, and a is the trace of the generator of a transformation relating an arbitrary spin basis to one where the spin metric is numerically constant.

The algebraic relation between the Dirac operators and the event metric field places certain restrictions on the traceless part of the spin connection and from now on when we speak of the spin connection, it will be

understood that we mean the traceless part. For a metric space, the most general relation defining the spin connection which is compatible with (1) is

$$\nabla_\omega \gamma_\kappa \equiv \partial_\omega \gamma_\kappa - \Gamma_{\omega\kappa}{}^\rho \gamma_\rho - [{}^s\Gamma_\omega, \gamma_\kappa] = iM_{\omega\kappa}, \quad (11)$$

where $M_{\omega\kappa}$ is a tensor and either vanishes or belongs to $L(g_{ab})$. In the latter case it must satisfy

$$M_{\omega(\kappa}\gamma_{\mu)} + \gamma_{(\mu}M_{|\omega|\kappa)} = 0. \quad (12)$$

The most general solution of (12) is of the form

$$M_{\omega\kappa} = [M_\omega, \gamma_\kappa], \quad (13)$$

where M_ω belongs to $L(g_{a\bar{b}})$. Here we will be concerned with the only case that has been studied, namely, $M_{\omega\kappa} = 0$, or

$$\nabla_\omega \gamma_\kappa = 0. \quad (14)$$

As a consequence of (14) and the definition (5),

$$\nabla_\omega \gamma_5 = 0. \quad (15)$$

Equations (14) and (1) define a relation between the metric field of a Riemann space and the spin connection. The first integrability condition for (14) is

$$R_{\omega\mu\nu}{}^\rho \gamma_\rho + [\Phi_{\omega\mu}, \gamma_\nu] = 0. \quad (16)$$

On account of the irreducibility of the Dirac operators, the traceless solution of (16) is unique and is given by

$$\Phi_{\omega\mu} = -\tfrac{1}{4}R_{\omega\mu}{}^{\kappa\lambda}\gamma_{[\kappa}\gamma_{\lambda]}. \quad (17)$$

It follows that in a Minkowski space the spin connection is integrable and hence of no physical interest. Moreover, we see from (17) that $\Phi_{\mu\nu}$ and its covariant derivatives are contained in the Lie algebra of the Lorentz group and hence the spin holonomy group is the Lorentz group or a subgroup thereof.

When we assume that the Riemann geometry is given, we may solve (14) for ${}^s\Gamma_\mu$. Choosing a normal basis, it follows from (16) that ${}^s\Gamma_\mu$ must be a linear combination of the bivectors $\gamma_{[\kappa}\gamma_{\lambda]}$. In a normal basis let

$${}^s\Gamma_\mu \overset{*}{=} a_\mu{}^{\kappa\lambda}\gamma_{[\kappa}\gamma_{\lambda]}. \quad (18)$$

Then, with (18) and (14),

$$\partial_\mu \gamma_\nu - \Gamma_{\mu\nu}{}^\rho \gamma_\rho + 4a_{\mu\nu}{}^\rho \gamma_\rho \overset{*}{=} 0. \quad (19)$$

Multiplying (19) by γ^σ and contraction of spinor indices yield

$$a_{\mu\nu}{}^\sigma \overset{*}{=} -\tfrac{1}{4}\operatorname{Tr}\gamma^\sigma {}^e\nabla_\mu \gamma_\nu, \quad (20)$$

where ${}^e\nabla_\mu$ means covariant with respect to the event indices only. With (18) and (20) we have an explicit solution for the spin connection in terms of the Riemannian geometry. This solution is simpler than those given by Fletcher[12] and Loos [13];however, their solutions hold in arbitrary local bases, while the solution given here holds only in the special bases, as indicated in (18) and (20).

Summarizing at this point, we have the well-known results that, for a given Riemannian geometry and a spin connection related to the Riemannian geometry by the covariant constancy of the Dirac operators, the spin connection exists and the traceless part is determined uniquely by the Riemannian metric field and is given explicitly by Eqs. (18) and (20). The spin connection is nonintegrable if and only if the parameters of the event connection are nonintegrable.

Let us turn now to the so-called inverse problem where the spin connection is assumed given and we seek the Dirac operators γ_μ and the parameters of the linear connection for a Riemann space as solutions of the mapping equations stated above. The physical interest in this problem arises when one wishes to consider the spin connection as a fundamental field in elementary particle theory. Since this field carries energy and momentum, it should give rise to an underlying Riemann geometry. The question which has been asked is whether this geometry is unique. Loos[4] has solved this problem by showing an algebraic condition for which the parameters of the event connection are determined uniquely by the spin connection in a neighborhood of an assumed solution ${}^0\Gamma_{\mu\kappa}{}^\lambda$ of the mapping equations. The metric field remains undetermined by a uniform scale transformation, say β, and the Dirac operators are hence undetermined by the factor $\beta^{\frac{1}{2}}$. In addition the γ_μ are undetermined by a transformation $\exp(i\alpha\gamma_5)$ with constant, real α. With the other known results summarized above, this uniqueness condition guarantees that the map between parameters of the event connection and the spin connection is locally one-to-one.

The invariance of the theory under the transformation $\exp(i\alpha\gamma_5)$ can be shown as follows.[14] Suppose ${}^0\Gamma_{\mu\kappa}{}^\lambda$ and ${}^0g_{\mu\kappa}$ are a solution of (1) and (14) for fixed ${}^s\Gamma_\mu$. Then, ${}^0\gamma_\mu + \delta\gamma_\mu$, for an infinitesimal variation $\delta\gamma_\mu$, will satisfy (1) and (14) if and only if

$$\delta\gamma_\mu = [L, \gamma_\mu] \quad (21)$$

and the variation $\delta\gamma_\mu$ is covariant constant. In view of (14), this condition is satisfied if

$$[\nabla_\nu L, \gamma_\mu] = 0. \quad (22)$$

By Schur's lemma, the general solution of (22) is

$$\nabla_\mu L = \beta_\mu I, \quad (23)$$

where β_μ is an arbitrary vector field. The first integrability condition for (23) is

$$-\tfrac{1}{2}[\Phi_{\mu\nu}, L] = \partial_{[\mu}\beta_{\nu]} I, \tag{24}$$

which shows that $\partial_{[\mu}\beta_{\nu]}$ vanishes and that the general solution is

$$L = \beta I + i\alpha\gamma_5,$$

where α and β are arbitrary scalars. Returning to (23) and recalling that γ_5 is covariant constant, we see that $\beta_\mu = \partial_\mu\beta$ and that α is a constant. We require also that $g_{a\bar{b}}$ is invariant, which forces α to be real. Hence the theory is invariant, for arbitrary fixed spin basis, under the object transformations of spinors generated by $i\alpha\gamma_5$, which is the only part of L of interest. This invariance arises because the Lie algebra of the spin holonomy group is not represented irreducibly. If the spin holonomy group were, for example, the full unitary group of $g_{a\bar{b}}$, then this invariance would not exist. But then, as we have seen, the map between the spin and event connections would not exist either, if we still insisted on a Riemannian event space.

III. THE USE OF ANHOLONOMIC BASES IN THE EVENT SPACE

The inverse problem appears to be nonlinear in the following sense. Invoking the Christoffel formula we may write (14) in terms of the Riemann metric and the Dirac operators. Then $g_{\kappa\lambda}$ can be eliminated from the equation with the aid of the quadratic decomposition of γ_μ. This gives what appears to be a nonlinear partial differential equation in γ_μ, which we have not studied further in this form.

It turns out that the apparent nonlinearity in the inverse problem can be avoided by using anholonomic bases in the Riemann space. We shall see that the problem of solving the system of equations which couple the local Dirac operators, the event tensor, and the parameters of the event connection can be reduced to solving a linear partial differential equation in the transformation coefficients which relate holonomic to anholonomic bases.

Let ${}^\kappa e_\lambda$ denote a system of reference vectors of the coordinate system x^λ. The ${}^\kappa e_\lambda$ satisfy

$$\partial_{[\mu}{}^\kappa e_{\lambda]} = 0, \tag{25}$$

and are called a holonomic basis. We may employ a new basis at each event which is obtained from the holonomic basis by the transformation

$${}^A e_\lambda = A_\kappa^A{}^\kappa e_\lambda. \tag{26}$$

In the sequel, we need the matrix A_A^λ as well; it is defined by

$$A_A^\lambda A_\kappa^A = \delta_\kappa^\lambda. \tag{27}$$

The new basis ${}^A e_\lambda$ is holonomic if and only if

$$\partial_{[\mu}{}^A e_{\lambda]} = 0. \tag{28}$$

If the transformation A_λ^A does not satisfy (28), the system of reference formed by the fields ${}^A e_\lambda$ is called an anholonomic basis. In the following we shall use Greek letters to indicate holonomic bases and capital Latin letters to denote anholonomic bases.

Suppose we choose bases at each event such that the event metric field is constant. It is clear that unless the event space is a Minkowski space we will have to use anholonomic bases to achieve this. Then

$$\gamma_{(A}\gamma_{B)} = g_{AB} \tag{29}$$

and covariant constancy of the Dirac operators reads

$$\nabla_\mu\gamma_A \equiv \partial_\mu\gamma_A - \Gamma_{\mu A}{}^B\gamma_B - [{}^s\Gamma_\mu, \gamma_A] = 0. \tag{30}$$

Further, we may choose a spin basis such that the γ_A are numerically constant. Then we have

$$-\Gamma_{\mu A}{}^B\gamma_B - [{}^s\Gamma_\mu, \gamma_A] \overset{*}{=} 0, \tag{31}$$

the solution of which is

$$\Gamma_{\mu A}{}^B \overset{*}{=} -\mathrm{Tr}\,\gamma^B[{}^s\Gamma_\mu, \gamma_A]. \tag{32}$$

We do not yet have a solution of the inverse problem since we do not know the parameters of the event connection in the holonomic coordinate system in which the spin connection is specified.

Under the transformation A_λ^A, the parameters of the connection transform according to

$$\Gamma_{AB}{}^C = A_A^\mu A_B^\kappa \Gamma_{\mu\kappa}{}^\lambda A_\lambda^C - A_A^\mu A_B^\kappa \partial_\mu A_\kappa^C. \tag{33}$$

Alternation over AB in (33) yields

$$\Gamma_{[AB]}{}^C = A_A^\mu A_B^\kappa S_{\mu\kappa}{}^\lambda A_\lambda^C - A_A^\mu A_B^\kappa \partial_{[\mu} A_{\kappa]}^C, \tag{34}$$

where $S_{\mu\kappa}{}^\lambda$ is the torsion tensor. In accord with general relativity we take a vanishing torsion. Then transvection of (34) with $A_\nu^A A_\sigma^B$ yields

$$\partial_{[\nu} A_{\sigma]}^C + \Gamma_{[\nu|A|}{}^C A_{\sigma]}^A = 0. \tag{35}$$

The coefficients $\Gamma_{\nu A}^C$ in (35) are determined by (32) when the gauge potentials are specified. The transformation matrices are thus determined by a linear equation and if one had a solution A_σ^C, the parameters of the event connection would be computed from the transformation law

$$\Gamma_{\mu\kappa}{}^\lambda = A_\kappa^A \Gamma_{\mu A}{}^B A_B^\lambda - A_\mu^A \partial_\kappa A_A^\lambda.$$

The event connection calculated in this way is symmetric and metric. The uniqueness of solutions $\Gamma_{\mu\kappa}{}^\lambda$ is thereby established by the number of solutions A_κ^A of (35).

Because of the symmetry of the event connection, we may write (35) in the form

$$\nabla_{[\nu} A^{A}_{\sigma]} = 0. \tag{36}$$

The first integrability condition for (36) is

$$R_{[\mu\nu|A|}{}^{B} A^{A}_{\sigma]} = 0, \tag{37}$$

where

$$R_{\mu\nu A}{}^{B} = \partial_\mu \Gamma_{\nu A}{}^{B} - \partial_\nu \Gamma_{\mu A}{}^{B} - \Gamma_{\mu A}{}^{C} \Gamma_{\nu C}{}^{B} + \Gamma_{\nu A}{}^{C} \Gamma_{\mu C}{}^{B}. \tag{38}$$

We may write (37) as an array of linear algebraic equations, with a 16 × 16 matrix of coefficients, say N:

$$\begin{bmatrix} 0 & R_{23A}{}^{B} & R_{31A}{}^{B} & R_{12A}{}^{B} \\ R_{23A}{}^{B} & 0 & R_{30A}{}^{B} & R_{02A}{}^{B} \\ R_{13A}{}^{B} & R_{30A}{}^{B} & 0 & R_{01A}{}^{B} \\ R_{12A}{}^{B} & R_{20A}{}^{B} & R_{01A}{}^{B} & 0 \end{bmatrix} \begin{bmatrix} A^{A}_{0} \\ A^{A}_{1} \\ A^{A}_{2} \\ A^{A}_{3} \end{bmatrix} = 0. \tag{39}$$

By inspection one sees that the rank of N is at most fifteen, so there is at least one solution.

With the aid of the integrability condition (39), we may simplify the differential Eq. (36). The general solution of (36) can be written in the form

$$A^{A}_{\sigma} = \varphi(x^{\mu})\, {}^{0}A^{A}_{\sigma}, \tag{40}$$

where it is necessary that ${}^{0}A^{A}_{\sigma}$ satisfy the integrability condition (39). If we restrict the determinant of ${}^{0}A^{A}_{\sigma}$ to be a fixed value, then ${}^{0}A^{A}_{\sigma}$ is determined uniquely if the rank of N is fifteen. For definiteness we take $\det {}^{0}A^{A}_{\sigma} = 1$. Substitution of (40) in (36) yields

$${}^{0}A^{A}_{[\sigma} \nabla_{\nu]} \varphi + \nabla_{[\nu}\, {}^{0}A^{A}_{\sigma]} \varphi = 0. \tag{41}$$

Covariant differentiation of (41) and alternation over holonomic indices yields

$$(\nabla_{[\omega}\, {}^{0}A^{A}_{\sigma}) \nabla_{\nu]} \varphi + {}^{0}A^{A}_{[\sigma} \nabla_{\omega} \nabla_{\nu]} \varphi + (\nabla_{[\omega} \nabla_{\nu}\, {}^{0}A^{A}_{\sigma]}) \varphi + \nabla_{[\nu}\, {}^{0}A^{A}_{\sigma} \nabla_{\omega]} \varphi = 0.$$

The first and last terms cancel and the second term vanishes because ∇_ω and ∇_ν acting on a scalar commute. The remaining term may be expressed in the form

$$(\nabla_{[\omega} \nabla_{\nu}\, {}^{0}A^{A}_{\sigma]}) \varphi = \tfrac{1}{2} R_{[\omega\nu|B|}{}^{A0} A^{B}_{\sigma]} \varphi,$$

which, by virtue of the definition of ${}^{0}A^{A}_{\sigma}$, also vanishes, and we see that this integrability condition for (41) is satisfied identically. Whether this condition is sufficient for existence of a scalar function φ has not been investigated. What is needed here is a theorem for the curl Eq. (41) [or (36)], of the type proved by Veblen and Thomas[15] for gradient equations.

We have seen that the general solution for the transformation matrix may be written in the form shown by (40). If the condition $\det {}^{0}A^{A}_{\sigma} = 1$ is imposed, then ${}^{0}A^{A}_{\sigma}$ is uniquely determined by (40) if the rank of N is maximum. The uniqueness is thus established by the number of different solutions φ of (41). Suppose there are two solutions φ and φ' of (41) with the same ${}^{0}A^{A}_{\sigma}$. Then it follows by subtracting the two equations for φ and φ' that

$${}^{0}A^{A}_{[\sigma} \nabla_{\nu]} \ln \frac{\varphi'}{\varphi} = 0. \tag{42}$$

Transvection of (42) with ${}^{0}A_{A}{}^{\sigma}$ yields

$$3\nabla_\nu \ln (\varphi'/\varphi). \tag{43}$$

Equation (43) implies

$$\varphi'(x^{\kappa}) = a\varphi(x^{\kappa}),$$

where a is an arbitrary real constant. If the rank of N is maximum, there is only one solution ${}^{0}A^{A}_{\sigma}$, with $\det {}^{0}A^{A}_{\sigma} = 1$, of the integrability condition (39). Then according to (40), the transformation matrix is determined uniquely up to the constant factor a. This nonuniqueness represents a uniform scale transformation of the metric field. The parameters of the event connection are unaffected by this scale transformation. Thus we have shown that if the rank of the matrix N is maximum, the correspondence between the spin connection and parameters of the event connection is one-to-one. Moreover, Loos[4] has shown that the identities satisfied by the Riemann curvature do not force the rank of N to be less than 15, so that cases where this does occur[6] should be considered exceptional, and in this sense we may say that the correspondence between the spin and event connection is almost always one-to-one.

IV. THE INVERSE PROBLEM FOR GAUGE POTENTIALS ACTING IN THE ALGEBRA SPACE OF LOCAL LORENTZ TRANSFORMATIONS[16]

The generators of local Lorentz transformations $L_{i\kappa}{}^{\lambda}$, $i, j \cdots = 1, \cdots, 6$, have both event and algebra indices and thereby play the role of connecting quantities. The local generators are anti-self-adjoint with respect to the event metric,

$$L_{i\kappa}{}^{\lambda} g_{\lambda\mu} + L_{i\mu}{}^{\lambda} g_{\lambda\kappa} = 0, \tag{44}$$

and hence satisfy the commutation relations

$$[L_i, L_j] = C_{ij}{}^{k} L_k, \tag{45}$$

where $C_{ij}{}^{k}$ are the structure constants of the group of local Lorentz transformations. The gauge potentials are introduced by insisting that

$$\nabla_\mu L_i = \partial_\mu L_i - [{}^{c}\Gamma_\mu, L_i] - \Gamma_{\mu i}{}^{j} L_j = 0. \tag{46}$$

Equations (45) and (46), together with the requirement that the event space be a Riemann space, establishes the map between the parameters of a symmetric and metric event connection and the gauge potentials.

Let us now consider the system of Eqs. (44)–(46), defining the inverse mapping in a particular anholonomic basis. We fix the algebra base up to constant Lorentz transformations by the requirement that $C_{ik}^{\ j}$ are numerically constant. In this base, the generators $L_{i\kappa}^{\ \lambda}$ at different points x^κ and y^κ satisfy

$$\begin{aligned}[L_i(x^\kappa), L_j(x^\kappa)] &= C_{ij}^{\ k}L_k(x^\kappa),\\ [L_i(y^\kappa), L_j(y^\kappa)] &= C_{ij}^{\ k}L_k(y^\kappa),\end{aligned} \tag{47}$$

with the same structure constants in both equations. The generators at x^κ and y^κ are therefore related by a similarity transformation. If we are not restricted by holonomity of basis, we may then obtain generators independent of the coordinates by a transformation

$$L_{iA}^{\ \ B}(y^\mu) = A_A^\kappa(y^\mu)L_{i\kappa}^{\ \ \lambda}(y^\mu)A_\lambda^B(y^\mu) = \delta_A^\kappa L_{i\kappa}^{\ \ \lambda}(x^\mu)\delta_\lambda^A.$$

In such an anholonomic basis, we have

$$\partial_\mu L_{iA}^{\ \ B} \overset{*}{=} 0. \tag{48}$$

In addition we have, for the special basis in $L(g_{\kappa\lambda})$,

$$\Gamma_{\mu j}^{\ \ \kappa} \overset{*}{=} b_\mu^i C_{ij}^{\ \ \kappa}, \tag{49}$$

where (49) guarantees that the holonomy group of the gauge potentials is the Lorentz group or a subgroup thereof, a condition which is necessary in order for solutions of the inverse map to exist.[6] Further, we have

$$L_{iA}^{\ \ C}g_{CB} + L_{iB}^{\ \ C}g_{CA} = 0. \tag{50}$$

With (48), the mapping Eq. (46) is now algebraic and reads

$$-\Gamma_{\mu A}^{\ \ C}L_{iC}^{\ \ B} + L_{iA}^{\ \ C}\Gamma_{\mu C}^{\ \ B} \overset{*}{=} \Gamma_{\mu i}^{\ \ j}L_{jA}^{\ \ B}, \tag{51}$$

where $\Gamma_{\mu i}^{\ \ j}$ is given by (49). Substitution of

$$\Gamma_{\mu A}^{\ \ B} \overset{*}{=} -b_\mu^i L_{iA}^{\ \ B} \tag{52}$$

into (51) and use of the commutation relations (45) shows that (52) is a solution of (51). We may add to this solution a matrix which commutes with the $L_{iA}^{\ \ B}$. Since the $L_{iA}^{\ \ B}$ form an irreducible set, the general solution is

$$\Gamma_{\mu A}^{\ \ B} \overset{*}{=} -b_\mu^i L_{iA}^{\ \ B} + \alpha_\mu \delta_A^B. \tag{53}$$

The event space is a metric space, so we have

$$\nabla_\mu g_{AB} = 0. \tag{54}$$

Using (53) in (54) yields

$$\partial_\mu g_{AB} - b_\mu^i(L_{iA}^{\ \ C}g_{CB} + L_{iB}^{\ \ C}g_{CA}) - \alpha_\mu(\delta_A^C g_{CB} + \delta_B^C g_{CA}) = 0.$$

The second term vanishes, since the $L_{iA}^{\ \ C}$ are anti-self-adjoint with respect to g_{AB} and, hence, we have

$$\partial_\mu g_{AB} - 2\alpha_\mu g_{AB} = 0. \tag{55}$$

However, the integrability condition for (54) is

$$\partial_{[\mu}\alpha_{\nu]} = 0. \tag{56}$$

It follows that the part of the connection in (53) proportional to the identity can be transformed away by a transformation that leaves $L_{iA}^{\ \ B}$ constant, so we may as well take $\alpha = 0$, and the desired solution of (51) is (52).

As in the case of the spin connection, a solution of the mapping equations can be constructed from (52), once one knows the transformation matrix A_σ^A. The A_σ^A is found as a solution of the same differential Eq. (36) as before. The uniqueness of the $\Gamma_{\mu\kappa}^{\ \ \lambda}$ then is established, as before, by the uniqueness of the solutions of the linear equation. The question of uniqueness has therefore been settled by the arguments already given in Sec. III. So we have also reached the conclusion here that the correspondence between gauge potentials is one-to-one providing the rank of the matrix N is maximum.

One difference in the two problems is that the $L_{i\kappa}^{\ \ \lambda}$, which appear as solutions of the mapping equations, are unique for given fixed gauge potentials, structure constants, and parameters of the event connection. While the Dirac operators, which played the analogous role as connecting quantities, are arbitrary by a uniform scale transformation $\beta^{\frac{1}{2}}$ and similarity transformation with transformation matrix $\exp(i\alpha\gamma_5)$. The latter differences arises from the fact that Lorentz holonomy group is represented reducibly in the spin space, while the Lie algebra space $L(g_{\kappa\lambda})$ affords an irreducible representation of the Lorentz holonomy group.

ACKNOWLEDGMENT

I am happy to acknowledge that discussions with H. G. Loos were very helpful in carrying out this work.

* Based on part of the author's Ph.D. thesis, University of California, Riverside.

† Present Address: West Virginia University, Morgantown, W. Va.

[1] H. Tetrode, Z. Physik **49**, 858 (1928); **50**, 336 (1928); E. Wigner, *ibid.* **53**, 592 (1928); H. Weyl, *ibid.* **56**, 330 (1929); V. Fock, *ibid.* **57**, 261 (1929); Schrödinger, Sitzber, Preuss. Akad. Wiss. Physik-math. Kl. **XI.**, 105 (1932).

[2] J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954).

[3] R. P. Treat, Phys. Rev. Letters **12**, 407 (1964).

[4] H. G. Loos, Ann. Phys. (N.Y.) **25**, 91 (1963).

[5] A. Peres, Nuovo Cimento Suppl. **24**, 2, 390 (1962).

[6] R. P. Treat, "Local Gauge Fields in General Relativity," J. Math. Phys. **11**, 2176 (1970).

[7] We will use the kernel index tensor notation and follow generally the notation of Ref. 2, where round brackets denote mixing, e.g.,

$$T_{(\kappa\lambda)} = \frac{1}{2!}(T_{\kappa\lambda} + T_{\lambda\kappa}).$$

Square brackets denote alternation:

$$T_{[\kappa\lambda]} = \frac{1}{2!}(T_{\kappa\lambda} - T_{\lambda\kappa}).$$

When indices are to be excluded from mixing or alternation, it is indicated by vertical lines enclosing the index as in

$$T_{[\kappa|\lambda|\mu]} = \frac{1}{2!}(T_{\kappa\lambda\mu} - T_{\mu\lambda\kappa}).$$

The notation for mixing and alternation is used for any kind of index.

[8] B. L. van der Waerden, Nachr. Akad. Wiss, Göttingen, Math. Phys. Kl., 100 (1929).

[9] The holonomy group of the spin connection is defined in the same way as for the Yang–Mills potentials. Much of the theory of internal holonomy groups of Yang–Mills fields found in Ref. 10 also pertains to the holonomy group of the spin connection.

[10] H. G. Loos, J. Math. Phys. **8**, 2114 (1967).

[11] H. G. Loos ["Spin Connection in General Relativity," 1963 (unpublished)] has shown that if the spin metric is covariant constant, it follows that the trace of the spin connection is necessarily pure imaginary. Our way of dealing with the metric structure of spin space has been given in this work.

[12] J. G. Fletcher, Nuovo Cimento **8**, 451 (1958).

[13] H. G. Loos, Nuovo Cimento **30**, 901 (1963).

[14] This has been proved in Ref. 1. We give a different proof here.

[15] O. Veblen and J. Y. Thomas, Ann. Math. **27**, 278 (1926).

[16] A more detailed analysis can be found in R. P. Treat, Ph.D. thesis, University of California, Riverside, 1967.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Solution of a Three-Body Scattering Problem in One Dimension*

C. MARCHIORO
Istituto Matematico dell'Università di Roma, Istituto di Fisica dell'Università di Roma, Rome, Italy

(Received 27 October 1969)

The scattering problem of three equal particles interacting via repulsive inverse-cube forces is solved in one dimension both in the classical and the quantal cases. The quantal S matrix is similar to that produced by repulsive δ-function interaction of infinite strength.

1. INTRODUCTION

The 1-dimensional problem of three equal particles interacting via repulsive inverse-cube forces has been recently studied and the whole set of eigenfunctions given.[1] However, the scattering of these three particles has not yet been considered; its study is the aim of the present paper.

We find that the scattering amplitude of this process is quite simple and very similar, in spite of the long-range forces involved, to the amplitude produced by an infinitely repulsive δ-function interaction.[2,3] Specifically, the S matrix has nonvanishing elements only between states differing in the interchange of the momenta of the two external particles.

In Sec. 2, the Hamiltonian of the system is stated; in Sec. 3, the classical problem is briefly studied and the (classical) scattering process discussed; in Sec. 4 the quantum scattering amplitude is derived.

2. HAMILTONIAN OF THE SYSTEM

Let us consider three equal particles moving on a straight line and interacting via repulsive inverse-cube forces. The Hamiltonian of this system is

$$H = (2m)^{-1}(p_1^2 + p_2^2 + p_3^2) + g[(X_1 - X_2)^{-2} + (X_1 - X_3)^{-2} + (X_2 - X_3)^{-2}], \tag{2.1}$$

where X_1, X_2, X_3 are the coordinates of the three particles, p_1, p_2, p_3 their conjugate momenta, m the mass of each particle, and g a positive (coupling) constant.

In order to exhibit the separability of this Hamiltonian[4,5] and to write it in a more manageable form, it is convenient to separate the center-of-mass coordinate from the internal ones, setting

$$\begin{aligned} R &= \tfrac{1}{3}(X_1 + X_2 + X_3), \\ \xi &= 6^{-\frac{1}{2}}(X_1 + X_2 - 2X_3), \\ \eta &= 2^{-\frac{1}{2}}(X_1 - X_2). \end{aligned} \tag{2.2}$$

We introduce next the 2-dimensional spherical coordinates r and φ:

$$\begin{aligned} \xi &= r\cos\varphi, \quad 0 \le r < \infty, \\ \eta &= r\sin\varphi, \quad 0 \le \varphi < 2\pi. \end{aligned} \tag{2.3}$$

In these variables, the Hamiltonian in the center-of-mass frame reads[1]

$$H = \frac{1}{2m}\left(p_r^2 + \frac{p_\varphi^2}{r^2}\right) + \frac{9g}{2}\frac{1}{r^2\sin^2(3\varphi)}. \tag{2.4}$$

This Hamiltonian is formally identical to that of a particle of mass m, moving in a plane and subject to the (noncentral) potential

$$\tfrac{9}{2}g(r\sin 3\varphi)^{-2}.$$

Square brackets denote alternation:

$$T_{[\kappa\lambda]} = \frac{1}{2!}(T_{\kappa\lambda} - T_{\lambda\kappa}).$$

When indices are to be excluded from mixing or alternation, it is indicated by vertical lines enclosing the index as in

$$T_{[\kappa|\lambda|\mu]} = \frac{1}{2!}(T_{\kappa\lambda\mu} - T_{\mu\lambda\kappa}).$$

The notation for mixing and alternation is used for any kind of index.

[8] B. L. van der Waerden, Nachr. Akad. Wiss, Göttingen, Math. Phys. Kl., 100 (1929).

[9] The holonomy group of the spin connection is defined in the same way as for the Yang–Mills potentials. Much of the theory of internal holonomy groups of Yang–Mills fields found in Ref. 10 also pertains to the holonomy group of the spin connection.

[10] H. G. Loos, J. Math. Phys. **8**, 2114 (1967).

[11] H. G. Loos ["Spin Connection in General Relativity," 1963 (unpublished)] has shown that if the spin metric is covariant constant, it follows that the trace of the spin connection is necessarily pure imaginary. Our way of dealing with the metric structure of spin space has been given in this work.

[12] J. G. Fletcher, Nuovo Cimento **8**, 451 (1958).

[13] H. G. Loos, Nuovo Cimento **30**, 901 (1963).

[14] This has been proved in Ref. 1. We give a different proof here.

[15] O. Veblen and J. Y. Thomas, Ann. Math. **27**, 278 (1926).

[16] A more detailed analysis can be found in R. P. Treat, Ph.D. thesis, University of California, Riverside, 1967.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Solution of a Three-Body Scattering Problem in One Dimension*

C. MARCHIORO
Istituto Matematico dell'Università di Roma, Istituto di Fisica dell'Università di Roma, Rome, Italy

(Received 27 October 1969)

The scattering problem of three equal particles interacting via repulsive inverse-cube forces is solved in one dimension both in the classical and the quantal cases. The quantal S matrix is similar to that produced by repulsive δ-function interaction of infinite strength.

1. INTRODUCTION

The 1-dimensional problem of three equal particles interacting via repulsive inverse-cube forces has been recently studied and the whole set of eigenfunctions given.[1] However, the scattering of these three particles has not yet been considered; its study is the aim of the present paper.

We find that the scattering amplitude of this process is quite simple and very similar, in spite of the long-range forces involved, to the amplitude produced by an infinitely repulsive δ-function interaction.[2,3] Specifically, the S matrix has nonvanishing elements only between states differing in the interchange of the momenta of the two external particles.

In Sec. 2, the Hamiltonian of the system is stated; in Sec. 3, the classical problem is briefly studied and the (classical) scattering process discussed; in Sec. 4 the quantum scattering amplitude is derived.

2. HAMILTONIAN OF THE SYSTEM

Let us consider three equal particles moving on a straight line and interacting via repulsive inverse-cube forces. The Hamiltonian of this system is

$$H = (2m)^{-1}(p_1^2 + p_2^2 + p_3^2) + g[(X_1 - X_2)^{-2} + (X_1 - X_3)^{-2} + (X_2 - X_3)^{-2}], \tag{2.1}$$

where X_1, X_2, X_3 are the coordinates of the three particles, p_1, p_2, p_3 their conjugate momenta, m the mass of each particle, and g a positive (coupling) constant.

In order to exhibit the separability of this Hamiltonian[4,5] and to write it in a more manageable form, it is convenient to separate the center-of-mass coordinate from the internal ones, setting

$$\begin{aligned} R &= \tfrac{1}{3}(X_1 + X_2 + X_3), \\ \xi &= 6^{-\frac{1}{2}}(X_1 + X_2 - 2X_3), \\ \eta &= 2^{-\frac{1}{2}}(X_1 - X_2). \end{aligned} \tag{2.2}$$

We introduce next the 2-dimensional spherical coordinates r and φ:

$$\begin{aligned} \xi &= r\cos\varphi, \quad 0 \le r < \infty, \\ \eta &= r\sin\varphi, \quad 0 \le \varphi < 2\pi. \end{aligned} \tag{2.3}$$

In these variables, the Hamiltonian in the center-of-mass frame reads[1]

$$H = \frac{1}{2m}\left(p_r^2 + \frac{p_\varphi^2}{r^2}\right) + \frac{9g}{2}\frac{1}{r^2\sin^2(3\varphi)}. \tag{2.4}$$

This Hamiltonian is formally identical to that of a particle of mass m, moving in a plane and subject to the (noncentral) potential

$$\tfrac{9}{2}g(r\sin 3\varphi)^{-2}.$$

The (ξ, η) plane may be separated in six sectors, according to the different ordering of the three particles:

$$\begin{aligned} &0 < \varphi < \tfrac{1}{3}\pi, && X_1 > X_2 > X_3, \\ &\tfrac{1}{3}\pi < \varphi < \tfrac{2}{3}\pi, && X_1 > X_3 > X_2, \\ &\tfrac{2}{3}\pi < \varphi < \pi, && X_3 > X_1 > X_2, \\ &\pi < \varphi < \tfrac{4}{3}\pi, && X_3 > X_2 > X_1, \\ &\tfrac{4}{3}\pi < \varphi < \tfrac{5}{3}\pi, && X_2 > X_3 > X_1, \\ &\tfrac{5}{3}\pi < \varphi < 2\pi, && X_2 > X_1 > X_3. \end{aligned} \tag{2.5}$$

This can be easily seen using the relations

$$\begin{aligned} X_1 - X_2 &= 2^{\frac{1}{2}} r \sin \varphi, \\ X_2 - X_3 &= 2^{\frac{1}{2}} r \sin (\varphi + \tfrac{2}{3}\pi), \\ X_3 - X_1 &= 2^{\frac{1}{2}} r \sin (\varphi + \tfrac{4}{3}\pi), \end{aligned} \tag{2.6}$$

which follow from Eqs. (2.2) and (2.3).

The singular nature of the repulsive interaction forbids any exchange in the ordering of the particles, so problems in different sectors are not correlated (apart from a possible symmetry requirement in the quantum-mechanical case).

3. CLASSICAL CASE

Before investigating the more interesting quantum-mechanical scattering problem, let us briefly study the classical case, which displays in a very intuitive way the most interesting features of the process. To do this, we derive the exact solution describing the motion of the classical system. We start from the two integrals of motion [yielded by the separability of the Hamiltonian (2.4)]:

$$(2m)^{-1} p_r^2 + B^2 r^{-2} = E, \tag{3.1}$$

$$(2m)^{-1} p_\varphi^2 + \tfrac{9}{2} g \sin^{-2}(3\varphi) = B^2. \tag{3.2}$$

Here E is the total energy of the system and B is an "angular" constant of motion. Recalling that

$$p_r = m \frac{dr}{dt}, \quad p_\varphi = m r^2 \frac{d\varphi}{dt}, \tag{3.3}$$

it is straightforward to first solve Eq. (3.1) and subsequently Eq. (3.2), using the knowledge of $r(t)$. In this manner one finds

$$r(t) = [(2E/m)(t - t_0)^2 + B^2/E]^{\frac{1}{2}}, \tag{3.4}$$

$$\begin{aligned} \cos 3\varphi = (1 - 9g/2B^2)^{\frac{1}{2}} &\times \sin \{\arcsin [(1 - 9g/2B^2)^{-\frac{1}{2}} \cos 3\varphi_0] \\ &- 3 \arctan [(2/m)^{\frac{1}{2}} (E/B)(t - t_0)]\}. \end{aligned} \tag{3.5}$$

These equations constitute an explicit solution of the equations of motion.

We turn now to discuss the process of scattering. The initial state of the system is completely specified by the asymptotic expressions

$$\begin{aligned} X_1 &\xrightarrow[t \to -\infty]{} v_1 t + a_1, \\ X_2 &\xrightarrow[t \to -\infty]{} v_2 t + a_2, \\ X_3 &\xrightarrow[t \to -\infty]{} v_3 t + a_3. \end{aligned} \tag{3.6}$$

The scattering problem consists in determining the asymptotic motion of the particles after the collision, which will, of course, again be of the form

$$\begin{aligned} X_1 &\xrightarrow[t \to +\infty]{} v_1' t + a_1', \\ X_2 &\xrightarrow[t \to +\infty]{} v_2' t + a_2', \\ X_3 &\xrightarrow[t \to +\infty]{} v_3' t + a_3'. \end{aligned} \tag{3.7}$$

We now show that v_1, a_1, v_1', a_1', etc., are related by the following relations:

$$\begin{aligned} v_1' = v_3, \quad a_1' = a_3, \\ v_2' = v_2, \quad a_2' = a_2, \\ v_3' = v_1, \quad a_3' = a_1. \end{aligned} \tag{3.8}$$

This means that particles 1 and 3 interchange their asymptotic behaviors, while particle 2 passes "undisturbed" through the quite complicated process of scattering! To prove Eq. (3.8), we confine ourselves to the case $X_1 > X_2 > X_3$; that is, $\frac{1}{3}\pi > \varphi > 0$, the generalization to other situations being trivial. From Eq. (3.5), it is possible to verify by explicit calculation that

$$\begin{aligned} &\varphi \xrightarrow[t \to -\infty]{} \alpha + \beta t^{-1}, \\ &\varphi \xrightarrow[t \to +\infty]{} \alpha' + \beta' t^{-1}, \\ &\text{with} \quad \alpha' = \tfrac{1}{3}\pi - \alpha, \quad \beta' = -\beta, \end{aligned} \tag{3.9}$$

and, from Eq. (3.4),

$$\begin{aligned} r &\xrightarrow[t \to -\infty]{} -(2E/m)^{\frac{1}{2}}(t - t_0), \\ r &\xrightarrow[t \to +\infty]{} (2E/m)^{\frac{1}{2}}(t - t_0). \end{aligned} \tag{3.10}$$

Inserting these equations into the Eqs. (2.6), which give the explicit dependence of X_1, X_2, and X_3 upon r and φ, yields Eqs. (3.6), (3.7), and (3.8). This ends our proof.

It is interesting to compare the process under consideration with the scattering of three particles interacting via infinitely repulsive δ-function interactions. In the latter case, every collision merely interchanges

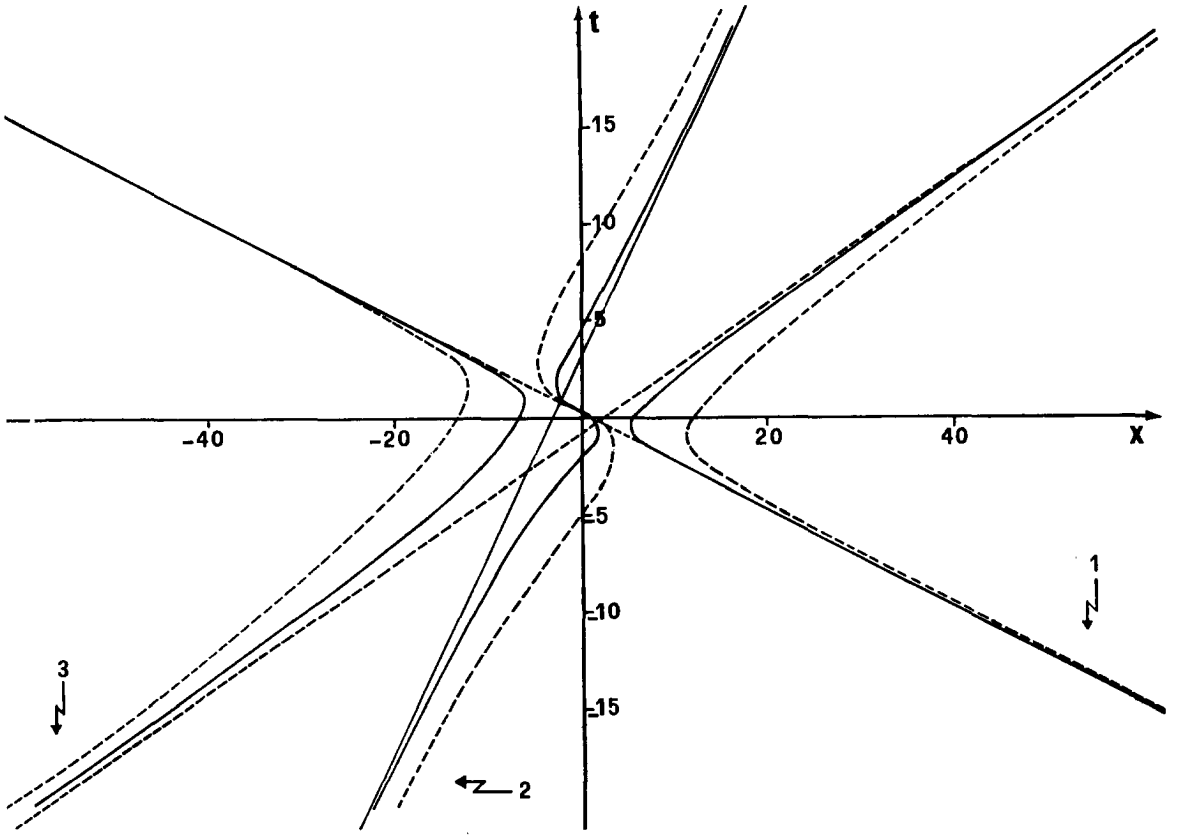


FIG. 1. Comparison between two motions with different g, but with the same asymptotic behavior. The corresponding infinitely repulsive δ interaction solution coincides with the asymptotic straight lines.

the free motion of the particles involved. Thus, the solutions of the motion correspond to the asymptotes of the solutions of the case discussed above. In Fig. 1 an example is drawn; the two curves correspond to different values of g and both tend asymptotically to the straight lines representing the solution of the motion of a system with infinite δ interactions. The solid curve and the dashed one correspond to $g = 100$, $m = 1$, and $g = 500$, $m = 1$ (in arbitrary units), respectively; it is clear that, for smaller g (that is, for weaker repulsive interactions), the solutions of the motion approach the straight lines in a larger zone.

In the next section, we see how the property (3.8) is reflected in the quantum scattering amplitude. In fact, the symmetry between initial and final states in the classical cases suggests that in the quantal case "incoming" and "outgoing" wavefunctions be related by a strict symmetry requirement, which, in the sense discussed in Ref. 2, forbids any "diffraction" during the scattering.

4. QUANTUM SCATTERING AMPLITUDE

We study the process of scattering in the framework of the time-independent scattering theory; we consider three particles obeying Boltzmann statistics and we confine ourselves to the sector $\frac{1}{3}\pi > \varphi > 0$, corresponding to $X_1 > X_2 > X_3$. (The extension to the other cases is trivial.)

We prove that to an "incoming" plane wavefunction

$$\text{const} \times \exp i(K_1X_1 + K_2X_2 + K_3X_3), \qquad X_1 \gg X_2 \gg X_3, \quad K_1 < K_2 < K_3, \tag{4.1}$$

there corresponds an "outgoing" wavefunction of the form

$$\text{const} \times \exp i(K_3X_1 + K_2X_2 + K_1X_3), \qquad X_1 \gg X_2 \gg X_3, \quad K_1 < K_2 < K_3. \tag{4.2}$$

The above statement implies that the scattering wavefunction must have the following asymptotic behavior:

$$\psi \xrightarrow[|X_i - X_j| \to \infty]{} \text{const}\,[\exp i(K_1X_1 + K_2X_2 + K_3X_3) + \gamma \exp i(K_3X_1 + K_2X_2 + K_1X_3)], \qquad X_1 \gg X_2 \gg X_3, \quad K_1 < K_2 < K_3, \tag{4.3}$$

where γ is a complex number of modulus 1.

We work in the center-of-mass frame, characterized by $K_1 + K_2 + K_3 = 0$, and we associate with K_1, K_2, K_3 a vector in a 2-dimensional space

$$K_\xi = 6^{-\frac{1}{2}}(K_1 + K_2 - 2K_3),$$
$$K_\eta = 2^{-\frac{1}{2}}(K_1 - K_2).$$

Then in this plane we introduce a system of spherical coordinates, in analogy with formulas (2.3), setting

$$K_\xi = K\cos\varphi_{in}, \quad 0 \le K < \infty,$$
$$K_\eta = K\sin\varphi_{in}, \quad 0 \le \varphi < 2\pi.$$

Then Eq. (4.3) in the coordinates r and φ reads[6]

$$\psi \xrightarrow[\substack{r\to\infty \\ 0<\varphi<\frac{1}{3}\pi}]{} \text{const}\,\{\exp[-iKr\cos(\pi + \varphi_{in} - \varphi) + \gamma\exp[iKr\cos(\tfrac{4}{3}\pi - \varphi_{in} - \varphi)]\}. \quad (4.4)$$

To prove Eq. (4.4), we develop the most general wavefunction in terms of the eigenfunctions ψ_{Kl} of the system as given in Ref. 1:

$$\psi_{Kl}(r, \varphi) = J_{3l+3a+\frac{3}{2}}(Kr)(\sin 3\varphi)^{a+\frac{1}{2}}C_l^{a+\frac{1}{2}}(\cos 3\varphi), \quad 0 \le \varphi \le \tfrac{1}{3}\pi, \quad (4.5)$$
$$a = \tfrac{1}{2}(1 + 2g)^{\frac{1}{2}}, \quad l = 0, 1, 2, \cdots,$$

in units so that $2m\hbar^{-2} = 1$. Here K^2 is the total energy (in the c.m. frame), C_l^λ is a Gegenbauer polynomial, and J_ν a Bessel function. Hence, we have

$$\psi(r, \varphi) = \sum_{l=0}^{\infty} b_l J_{3l+3a+\frac{3}{2}}(Kr)(\sin 3\varphi)^{a+\frac{1}{2}}C_l^{a+\frac{1}{2}}(\cos 3\varphi) \quad (4.6)$$

and, asymptotically,

$$\psi(r, \varphi) \xrightarrow[r\to\infty]{} \sum_{l=0}^{\infty} i(2\pi r)^{-\frac{1}{2}} \times b_l(\exp\{-i[Kr - \tfrac{1}{2}\pi(3l + 3a + 1)]\} - \exp\{i[Kr - \tfrac{1}{2}\pi(3l + 3a + 1)]\}) \times (\sin 3\varphi)^{a+\frac{1}{2}}C_l^{a+\frac{1}{2}}(\cos 3\varphi). \quad (4.7)$$

From the property of the Gegenbauer polynomials

$$C_l^\lambda(-X) = (-1)^l C_l^\lambda(X),$$

we have

$$C_l^{a+\frac{1}{2}}(\cos 3(\tfrac{1}{3}\pi - \varphi)) = \exp(il\pi)C_l^{a+\frac{1}{2}}(\cos 3\varphi). \quad (4.8)$$

Thus,

$$\psi(r, \varphi) \xrightarrow[r\to\infty]{} \sum_{l=0}^{\infty} i(2\pi r)^{-\frac{1}{2}}b_l \exp[i\tfrac{1}{2}\pi(3l + 3a + 1)] \times \{\exp(-iKr)(\sin 3\varphi)^{a+\frac{1}{2}}C_l^{a+\frac{1}{2}}(\cos 3\varphi) + \gamma\exp(iKr)[\sin 3(\tfrac{1}{3}\pi - \varphi)]^{a+\frac{1}{2}} \times C_l^{a+\frac{1}{2}}(\cos 3(\tfrac{1}{3}\pi - \varphi))\}, \quad (4.9)$$

with γ independent of l; specifically,

$$\gamma = \exp(-3ia\pi). \quad (4.10)$$

So, as a particular case of Eq. (4.9), Eq. (4.4) is proved.

Note that Eq. (4.10) implies that the S matrix is independent of the energy of the scattering particles and that it depends on the strength g of the interaction only through a phase factor.

Of course, the simplicity of the S matrix does not imply the simplicity of the whole collision process; in fact, the motion of a wave packet through the collision zone is quite complicated (as it was already suggested by the classical case).

In conclusion, we have proved that the S matrix of this 3-body problem has nonvanishing elements only between states that differ only in the interchange of the momenta of particles 1 and 3, exactly as in the infinitely repulsive δ interaction case. It is appealing to conjecture that a similar feature also holds, in analogy with the δ case,[2,3] in the scattering of N particles. However, in the 3-body problem we have made use of the complete set of eigenfunctions, while for the N-body problem only the ground-state function is known[7]; therefore, this problem is yet open.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor F. Calogero for many stimulating and helpful discussions. I wish to thank also Professor C. Cattaneo for some conversations on the classical case.

* Research supported by C.N.R. (Italian National Council of Research).

[1] F. Calogero, J. Math. Phys. **10**, 2191 (1969).
[2] J. B. McGuire, J. Math. Phys. **5**, 622 (1964).
[3] F. A. Berezin and V. N. Sushko, Zh. Eksp. Teor. Fiz. **48**, 1293 (1963) [Sov. Phys. JETP **21**, 865 (1965)]; E. Brezin and I. Zinn-Justine, Compt. Rend. **B263**, 670 (1966); C. N. Yang, Phys. Rev. Letters **19**, 1312 (1967); C. N. Yang, Phys. Rev. **168**, 1920 (1968).
[4] H. R. Post, Proc. Phys. Soc. **A69**, 936 (1956).
[5] J. Hurley, J. Math. Phys. **8**, 813 (1967).
[6] It is easily seen that interchange of K_1 with K_3 changes φ_{in} into $\frac{4}{3}\pi - \varphi_{in}$.
[7] F. Calogero, J. Math. Phys. **10**, 2197 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Expansions in Spherical Harmonics. V. Solution of Inverse Legendre Transforms

LL. G. CHAMBERS
Department of Applied Mathematics, University College of North Wales, Bangor, North Wales
AND
R. A. SACK
Department of Mathematics, University of Salford, Salford 5, Lancashire, England

(Received 10 November 1969)

Explicit formulas are derived for inverse Legendre transforms, i.e., for the solutions of the integral equation

$$f(r) = \int_b^1 G_L(ur) P_L(u)\, du$$

for arbitrary integer L and general functions $f(r)$. It is shown that the functions $f(r)$, defined by 2-sided transforms ($b = -1$), follow uniquely from the 1-sided ones ($b = 0$), but an arbitrary function of parity opposite to L (supplementary function) may be added to G_L in the 2-sided case. For $L \geq 2$, an arbitrary polynomial of the same parity as L, but of degree lower than L (complementary function), may be added to G_L. The complementary and supplementary functions do not affect the values of the integrals for the radial dependence in the expansion of a function of a vector sum in spherical harmonics. A power term in $f(r)$ leads to a power term in G_L, except for those powers which occur in the complementary function for which G_L involves logarithms. Inverse transforms are also obtained for a restricted number of negative powers and some recurrence relations are derived.

1. INTRODUCTION

It is well known that a function depending on a vector $\mathbf{r}_{AB} = (r_{AB}, \theta_{AB}, \phi_{AB})$, which in turn is the sum of s component vectors $\mathbf{r}_j = (r_j, \theta_j, \phi_j)$, can be expanded in spherical harmonics of the polar angles of the $\mathbf{r}_j$:

$$\begin{aligned} V(\mathbf{r}_{AB}) &= V(\mathbf{r}_1 + \mathbf{r}_2 + \cdots + \mathbf{r}_s) \\ &= \sum_{\mathbf{l},\mathbf{m}} \prod_{j=1}^{s} P_{l_j}^{m_j}(\cos\theta_j) \\ &\quad \times \exp(im_j\theta_j) R(V; \mathbf{l}, \mathbf{m}; \{r_j\}). \end{aligned} \quad (1.1)$$

In particular, if V is a function of the radius multiplied by a single harmonic of the angles

$$V = f(r_{AB}) P_L^M(\cos\theta_{AB}) \exp(iM\phi_{AB}), \quad (1.2)$$

then, as one of the authors has shown in a series of papers[1] (referred to as I–IV, respectively), the radial part can be split into two factors:

$$\begin{aligned} &R(f; L, M; \mathbf{l}, \mathbf{m}; \{r_j\}) \\ &\quad = K''(L, \mathbf{l}, M, \mathbf{m}) R''(f; L, \mathbf{l}; \{r_j\}). \end{aligned} \quad (1.3)$$

Of these factors, only the first, which is essentially an integral of a product of spherical harmonics over the unit sphere, involves the azimuthal quantum numbers m_j. The second factor R'', which alone involves the nature of the function f and the magnitudes of the vectors r_j, can be represented, for a wide class of functions f, either as a series expansion (Papers I–III), as a 1-dimensional Hankel integral[2,3] or as an s-dimensional integral over a unit hypercube (Paper IV):

$$\begin{aligned} &R''(f; L, \mathbf{l}, \{r_j\}) \\ &\quad = \int_{-1}^{1} \cdots \int_{-1}^{1} G_L(\textstyle\sum r_j u_j) \prod [P_{l_j}(u_j)\, du_j]. \end{aligned} \quad (1.4)$$

The kernel $G_L(w)$ is independent of the l_j and even of the number s of component vectors; hence, its relation to f and L is most easily found by choosing $s = 1$. In this case, the expansion (1.1) reduces to a single term with $l = L$, so that (1.4) becomes

$$f(r) = \int_{-1}^{1} G_L(ru) P_L(u)\, du. \quad (1.5)$$

Negative values of r are physically meaningless, and f can be chosen arbitrarily for negative values of its argument; similarly, the kernel G can be arbitrarily modified, provided the resulting integral is not altered for $r \geq 0$. In IV, it was assumed that the kernel vanishes for negative argument, so that the contribution to the integral (1.4) is confined to one-half of the hypercube and (1.5) can be written as

$$f(r) = \int_{0}^{1} g_L(ru) P_L(u)\, du. \quad (1.6)$$

The integrals (1.5) and (1.6) can be considered as 2-sided (2-s) and 1-sided (1-s) Legendre transforms, respectively, and the problem consists in finding the inverse Legendre transforms (ILT) for a given function f and rotational quantum number L:

$$G(w) = {}^2\mathfrak{L}_L^{-1}[f(r)], \quad g(w) = {}^1\mathfrak{L}_L^{-1}[f(r)]. \quad (1.7)$$

In IV, explicit formulas were derived for ${}^1\mathfrak{L}_L^{-1}(r^N)$, $N > L - 2$, and for ${}^1\mathfrak{L}_L^{-1}[f(r)]$ for arbitrary f and $L = 0$ or 1.

The purpose of the present paper is twofold. First, it will be shown that:

(a) solutions of the 1-s ILT (1.6) are not unique for $L > 1$;

(b) solutions of the 2-s ILT (1.5) exist only if $f(r)$ has the same parity as L; if this consistency condition is satisfied, an arbitrary function of the wrong parity can be added to $G_L(w)$ without affecting (1.5);

(c) the kernel $G_L(w)$ for the s-dimensional Legendre transform (1.4) exhibits all the arbitrariness of the 1-dimensional ILT for all physically relevant combinations of the l_j.

Secondly, explicit solutions of ILT's are derived for all values of L and all functions $f(r)$ possessing a sufficient number of derivatives. Recurrence relations are deduced, and the transforms for some nonanalytic functions are discussed in detail.

2. NONUNIQUENESS OF THE SOLUTIONS; COMPLEMENTARY AND SUPPLEMENTARY FUNCTIONS

The Legendre polynomials $P_L(u)$ are either even or odd functions depending on the parity of L. In consequence, they are orthogonal over the interval $(-1, 1)$ to any function of the opposite parity. Denoting an arbitrary even or odd function by ψ_e or ψ_o, respectively, we have

$$\begin{aligned} &\int_{-1}^{1} \psi_o(ru)P_L(u)\,du = 0, \quad L \text{ even}, \\ &\int_{-1}^{1} \psi_e(ru)P_L(u)\,du = 0, \quad L \text{ odd}. \end{aligned} \tag{2.1}$$

Furthermore, a Legendre polynomial $P_L(u)$ is orthogonal, over the same interval, to any polynomial $Q_n(u)$ of degree $n < L$:

$$\int_{-1}^{1} Q_n(u)P_L(u)\,du = 0, \quad n < L. \tag{2.2}$$

For powers u^n of parity opposite to L, Eq. (2.2) can be considered a consequence of (2.1), but for powers of the same parity the contribution to the integral for positive and negative values of u are the same, so that for even- or odd-powered polynomials $Q_{e,n}(u)$ and $Q_{o,n}(u)$, respectively, we obtain

$$\begin{aligned} &\int_{0}^{1} Q_{e,n}(u)P_L(u)\,du = 0, \quad L \text{ even}, \ n < L, \\ &\int_{0}^{1} Q_{o,n}(u)P_L(u)\,du = 0, \quad L \text{ odd}, \ n < L. \end{aligned} \tag{2.3}$$

If we call the set of polynomials satisfying (2.3) the *complementary function* and the set of functions satisfying (2.1) the *supplementary function*, we see that:

(a) the 1-s ILT $g_L(w)$ is indeterminate to within an additive complementary function;

(b) the 2-s ILT $G_L(w)$ is indeterminate to within additive complementary and supplementary functions.

For $L = 0$ and $L = 1$ no complementary function exists and, hence, the solutions $g_0(w)$ and $g_1(w)$ of (1.6) are unique; these are the cases explicitly solved in IV. The occurrence of the arbitrary complementary and supplementary functions would be a serious matter if they affected the value of the s-dimensional integral (1.4). Fortunately, they do not influence the value of R'' for all those combinations of $\{l_j\}$ for which the factor K'' of (1.3) differs from zero. Two necessary conditions for $K''(L, \mathbf{l}; M, \mathbf{m}) \neq 0$ are that:

(i) $L + \sum l_j$ is even;

(ii) none of the indices L and l_j exceeds the sum of all the others.

According to (i), the product $\prod P_{l_j}(u_j)$ has the parity of L on simultaneous sign reversal of all the u_j; hence, the supplementary function will contribute nothing to the integral over the hypercube. According to (ii), we have $\sum l_j > L$; since any power $(\sum r_j u_j)^n$ can be expanded in products of powers $\prod u_j^{n_j}$, $\sum n_j = n$, by means of the multinomial theorem, (2.2) implies that at least one $n_j < l_j$ in each term and the integral over the corresponding u_j vanish in view of the orthogonality. Integrals (1.4), for sets of $\{l_j\}$ for which K'' vanishes, are of no physical significance.

The converse problem concerns the nature of the 2-s ILT $G_L(ru)$ if $f(r)$ itself has the properties of the supplementary function, i.e., the opposite parity to L. If $G(ru)$ is split into an even and an odd component, the integral over the component of the same parity is again of the same parity, whereas the component of the opposite parity contributes nothing. Hence, the only functions $f(r)$ for which 2-s ILT's exist are those of the same parity as L. In fact, apart from the ineffective supplementary function, we can always put

$$G(w) = (-1)^L G(-w) = \tfrac{1}{2}g(w), \quad w > 0, \tag{2.4}$$

except when generalized functions occur at the origin. It is of interest to note that Ruedenberg,[2] in discussing the equivalence of the Hankel and Legendre transform representations of the radial functions, uses a 2-s ILT of the form (2.4) rather than the one-sided form of Paper IV. Formally, every 1-s ILT can be written as a two-sided one by introducing the Heaviside unit step function as a factor

$$H(w) = 0, \quad w < 0, \quad H(w) = 1, \quad w > 0, \tag{2.5}$$

by means of which (1.6) can be written as

$$f(r) = \int_{-1}^{1} g_L(ru) P_L(u) H(ru)\, du, \qquad (2.6)$$

which was the notation employed in IV. Since the one-sided Legendre transform defined by (1.6) or (2.6) is of the same generality as (1.5) without the additional complication of the supplementary function, the remainder of the paper will be concerned with 1-s ILT's wherever possible.

3. PARTICULAR SOLUTIONS OF THE ILT FOR POWERS AND POWER SERIES

We assume $f(r)$ to be a power r^N; it is clear that a particular solution for $g_L(w)$ is also proportional to w^N. We have

$$\int_0^1 (ru)^N P_L(u)\, du = r^N \frac{\pi^{\frac{1}{2}} \Gamma(1+N) 2^{-1-N}}{\Gamma(1 + \frac{1}{2}N - \frac{1}{2}L)\Gamma(\frac{3}{2} + \frac{1}{2}N + \frac{1}{2}L)} \qquad (3.1)$$

(cf. B3.12.23).[4] For even L, the integral converges for $N > -1$ and, for odd L, it converges for $N > -2$ [the pole in the numerator of (3.1) is cancelled by a similar pole in the denominator]. We thus obtain

$${}^1\mathfrak{L}_L^{-1}(r^N) = C_{NL} w^N, \qquad (3.2)$$

where C_{NL} is the reciprocal of the fraction in (3.1) and can be simplified by means of the duplication formula (B1.3.15) for the gamma function to

$$C_{NL} = \frac{(1+N)(3+N)\cdots(L+N+1)}{(2+N-L)(4+N-L)\cdots N}, \quad L \text{ even}, N > -1, N \neq 0, 2, \cdots, L-2, \qquad (3.3)$$

$$C_{NL} = \frac{(2+N)(4+N)\cdots(L+N+1)}{(2+N-L)(4+N-L)\cdots(N-1)}, \quad L \text{ odd}, N > -2, N \neq 1, 3, \cdots, L-2. \qquad (3.4)$$

For $L = 0$ and $L = 1$ the products in the denominator become empty; the value of such an empty product is always to be taken equal to unity. The results (3.3) and (3.4) were already derived in (IV.45), but the fact that a vanishing denominator leads to a pole for C_{NL} and that the equations resume their validity beyond these singularities was overlooked. These poles occur when r^N has the form of a complementary function (2.3) and solutions can be found by a limiting procedure $N = n + \epsilon \to n$. Before proceeding to the limit, a certain multiple of the complementary function is subtracted out and the limiting value is found by L'Hôpital's rule:

$$\begin{aligned} {}^1\mathfrak{L}_L^{-1}(r^N) &= \lim_{\epsilon \to 0} C_{n+\epsilon, L}(w^{n+\epsilon} - w^n) \\ &= D_{nL} \lim (w^{n+\epsilon} - w^n)/\epsilon \\ &= D_{nL} w^n \log w. \end{aligned} \qquad (3.5)$$

Here D_{nL} is the limit of the ratio of all the factors in (3.3) or (3.4) which do not vanish:

$$D_{nL} = \frac{(1+n)(3+n)\cdots(L+n+1)}{(2+n-L)\cdots(-4)(-2)\cdot 2\cdot 4\cdots n}, \quad L \text{ even}, \qquad (3.6)$$

$$D_{nL} = \frac{(2+n)(4+n)\cdots(L+n+1)}{(2+n-L)\cdots(-4)(-2)\cdot 2\cdot 4\cdots(n-1)}, \quad L \text{ odd}.$$

By means of the double factorial function

$$(2k)!! = 2\cdot 4\cdots(2k), \quad (2k+1)!! = 1\cdot 3\cdots(2k+1), \quad 0!! = 1!! = 1, \qquad (3.7)$$

Eqs. (3.6) can be written in a form independent of the parity of L

$$D_{nL} = (-1)^{(L-n)/2+1} \frac{(L+n+1)!!}{(L-n-2)!!\, n!}, \quad L - n = 2, 4, \cdots, L \text{ or } L-1. \qquad (3.8)$$

The integral in (3.1) diverges and, hence, the formulas (3.2)–(3.4) break down irremedially for $N \leq -1$ and L even or for $N \leq -2$ and L odd; however, solutions for the bilateral transform can be obtained for a limited number of negative powers in terms of generalized functions. We have

$$\int_{-1}^{1} \delta^{(m)}(ru) P_L(u)\, du = \frac{(-1)^m}{r^{m+1}} \frac{d^m P_L(0)}{du^m}. \qquad (3.9)$$

The derivative of the Legendre function in (3.9) is easily evaluated by using Rodrigues' formula (B3.6.17), leading to the result

$$\frac{d^m P_L(u)}{du^m} = \frac{(-1)^{\frac{1}{2}(L-m)}(L+m)!}{2^L [\frac{1}{2}(L+m)]!\, [\frac{1}{2}(L-m)]!}, \quad m = L, L-2, \cdots, 1 \text{ or } 0, \quad = 0, \text{ for other values of } m. \qquad (3.10)$$

Substitution of (3.10) in (3.9) yields the 2-s ILT

$${}^2\mathfrak{L}_L^{-1}\left(\frac{1}{r^{m+1}}\right) = \frac{(-1)^{(L+m)/2} 2^L [\frac{1}{2}(L-m)]!\, [\frac{1}{2}(L+m)]!}{(L+m)!} \delta^{(m)}(w), \qquad (3.11)$$

valid for

$$m = L, \quad L-2, \cdots, 1 \text{ or } 0. \tag{3.12}$$

For $m = L = 0$ and $m = L = 1$, this result was already given in IV [Eqs. (48) and (50)]. It is interesting to note that the same generalized function may occur as the ILT of very differently structured functions $f(r)$; thus, according to (IV.52), $\delta''(w)$ is proportional to the ILT for the 3-dimensional δ function when $L = 0$, whereas (3.11) shows that it describes the inverse transform for r^{-3} and positive even $L > 0$.

If a function $f(r)$ can be represented as a power series in r, its ILT can be found as the sum of the ILT's of the individual terms.

4. INVERSE LEGENDRE TRANSFORMS FOR ARBITRARY FUNCTIONS

If the Legendre functions $P_L(u)$ are expressed by means of Rodrigues' formula (B3.6.17), (1.6) becomes

$$f(r) = \int_0^1 g(ur)\,\frac{1}{2^L L!}\left(\frac{d}{du}\right)^L (u^2-1)^L\,du. \tag{4.1}$$

This formula can be repeatedly integrated by parts, provided that $g(w)$ has derivatives in $[0, r]$ up to the appropriate order; if we further assume that the integrated terms contribute nothing (this restriction will be discussed below), we obtain

$$\begin{aligned} f(r) &= \frac{(-r)^m}{2^L L!}\int_0^1 g^{(m)}(ur)\left(\frac{d}{du}\right)^{L-m}(u^2-1)^L\,du \\ &= \frac{(-1)^{L+m}}{2^L r^{L+1}}\int_0^r g^{(m)}(w)\left(\frac{d}{dw}\right)^{L-m}\frac{(r^2-w^2)^L}{L!}\,dw. \end{aligned} \tag{4.2}$$

Now the expression

$$F(x) = \int_\alpha^x \phi(\xi)\,\frac{(x-\xi)^L}{L!}\,d\xi \tag{4.3}$$

yields the $(L+1)$th iterate of the integral of $\phi(x)$ (cf. p. 221 of Ref. 5):

$$\left.\frac{d^l F(x)}{dx^l}\right|_\alpha = 0, \quad l = 0, 1, \cdots, L,$$
$$\frac{d^{L+1}F(x)}{dx^{L+1}} = \phi(x). \tag{4.4}$$

The integral (4.2) can be written in the form (4.3) when $m = L$, provided that r^2 and w^2 are taken as the independent variables:

$$\begin{aligned} (2r)^{L+1} f(r) &= \int_0^{r^2} \frac{g^{(L)}(w)}{w}\,\frac{(r^2-w^2)^L}{L!}\,d(w^2), \\ d(w^2) &= 2w\,dw, \end{aligned} \tag{4.5}$$

with the intermediate solution

$$\begin{aligned} \frac{g^{(L)}(r)}{r} &= \left(\frac{d}{dr^2}\right)^{L+1}[(2r)^{L+1}f(r)] \\ &= \left(\frac{1}{r}\frac{d}{dr}\right)^{L+1}[r^{L+1}f(r)], \end{aligned} \tag{4.6}$$

$$\left(\frac{d}{dr}\right)^L g_L(r) = \frac{d}{dr}\left(\frac{1}{r}\frac{d}{dr}\right)^L [r^{L+1}f(r)]. \tag{4.7}$$

In the special case $L = 0$, this yields the unique solution

$$g_0(r) = \frac{d}{dr}[rf(r)], \tag{4.8}$$

already derived by elementary means in IV; for $L > 0$ the last differentiation in (4.7) can be cancelled on both sides, leading to

$$\left(\frac{d}{dr}\right)^{L-1} g_L(r) = \left(\frac{1}{r}\frac{d}{dr}\right)^L [r^{L+1}f(r)]. \tag{4.9}$$

The same formula could have been derived by applying (4.3) and (4.4) to (4.2) with $m = L-1$; this derivation shows that no arbitrary constant enters (4.9). For $L = 1$ the ILT is obtained explicitly from (4.9):

$$g_1(r) = \frac{1}{r}\frac{d}{dr}[r^2 f(r)], \tag{4.10}$$

again in agreement with IV. For $L \geq 2$, (4.9) must be integrated, the explicit solution becoming

$$g_L(w) = \int_0^w \left(\frac{1}{r}\frac{d}{dr}\right)^L [r^{L+1}f(r)]\,\frac{(w-r)^{L-2}}{(L-2)!}\,dr \tag{4.11}$$

in view of (4.3) and (4.4).

However, comparison with (3.3) and (3.4) shows that it must be possible to simplify (4.11) since some of the differentiations under the integral sign are cancelled by subsequent integrations. These redundant operations can be eliminated with the use of the identities

$$\left(\frac{d}{dr}\right)^{2\lambda} = \left(\frac{d}{dr}\frac{1}{r}\right)^\lambda r^{2\lambda}\left(\frac{1}{r}\frac{d}{dr}\right)^\lambda, \tag{4.12a}$$

$$\left(\frac{d}{dr}\right)^{2\lambda+1} = \frac{1}{r}\left(\frac{d}{dr}\frac{1}{r}\right)^\lambda r^{2\lambda+2}\left(\frac{1}{r}\frac{d}{dr}\right)^{\lambda+1}, \tag{4.12b}$$

which can be verified by applying them to any power of r. Rodrigues' formula (B3.6.17) for the Legendre

polynomials becomes, in view of (4.12),

$$P_{2\lambda}(u) = \frac{1}{4^\lambda(2\lambda)!}\left(\frac{d}{du}\frac{1}{u}\right)^\lambda u^{2\lambda}\left(\frac{1}{u}\frac{d}{du}\right)^\lambda (u^2-1)^{2\lambda}$$
$$= \frac{1}{2^\lambda\lambda!}\left(\frac{d}{du}\frac{1}{u}\right)^\lambda [u^2(u^2-1)]^\lambda, \quad (4.13a)$$

$$P_{2\lambda+1}(u) = \frac{1}{2^\lambda\lambda!}\frac{1}{u}\left(\frac{d}{du}\frac{1}{u}\right)^\lambda [u^{2\lambda+2}(u^2-1)^\lambda]. \quad (4.13b)$$

Substitution of (4.13a) in (1.6) and integration by parts yield, in analogy to (4.2), for even $L = 2\lambda$,

$$f(r) = \frac{(-r^2)^\lambda}{2^\lambda\lambda!}\int_0^1\left[\left(\frac{1}{w}\frac{d}{dw}\right)^\lambda g(w)\right]u^{2\lambda}(u^2-1)^\lambda\,du$$
$$= \frac{1}{2^\lambda r^{2\lambda+1}\lambda!}\int_0^r\left[\left(\frac{1}{w}\frac{d}{dw}\right)^\lambda g(w)\right]w^{2\lambda}(r^2-w^2)^\lambda\,dw \quad (4.14)$$

or

$$r^{2\lambda-1}\left(\frac{1}{r}\frac{d}{dr}\right)^\lambda g(r) = \left(\frac{1}{r}\frac{d}{dr}\right)^{\lambda+1}[f(r)r^{2\lambda+1}], \quad (4.15)$$

with the solution

$$g_{2\lambda}(w) = \int_\epsilon^w\left[\frac{1}{r^{2\lambda-2}}\left(\frac{1}{r}\frac{d}{dr}\right)^{\lambda+1}[r^{2\lambda+1}f(r)]\right]\frac{(w^2-r^2)^{\lambda-1}}{2^{\lambda-1}(\lambda-1)!}\,dr. \quad (4.16)$$

Similarly, substitution of (4.13b) in (1.6) yields, for $L = 2\lambda + 1$,

$$f(r) = \frac{1}{2^\lambda r^{2\lambda+2}\lambda!}\int_0^r\left[\left(\frac{1}{w}\frac{d}{dw}\right)^\lambda\frac{g(w)}{w}\right]w^{2\lambda+2}(r^2-w^2)^\lambda\,dw, \quad (4.17)$$

$$\left(\frac{1}{r}\frac{d}{dr}\right)^\lambda\frac{g(r)}{r} = \frac{1}{r^{2\lambda+1}}\left(\frac{1}{r}\frac{d}{dr}\right)^{\lambda+1}[r^{2\lambda+2}f(r)], \quad (4.18)$$

$$g_{2\lambda+1}(w) = w\int_\epsilon^w\frac{1}{r^{2\lambda}}\left[\left(\frac{1}{r}\frac{d}{dr}\right)^{\lambda+1}[r^{2\lambda+2}f(r)]\right]\frac{(w^2-r^2)^{\lambda-1}}{2^{\lambda-1}(\lambda-1)!}\,dr. \quad (4.19)$$

The explicit solutions (4.16) and (4.19) for the ILT are not applicable for $\lambda = 0$, i.e., $L = 0$ and $L = 1$, and must be replaced by (4.8) and (4.10), respectively. The lower limits in the integrals in (4.16) and (4.19) have been left arbitrary as they affect $g(w)$ only through the addition of a complementary function; in (4.11) this was not permissible.

The crucial equation (4.6) can also be derived by an alternative, recursive, method which clearly shows its range of validity and that of its solution (4.11) at every stage. It is convenient to treat both one-sided and bilateral transforms simultaneously by leaving the lower limit of integration undetermined. If we write g_L for G_L in (1.5) and make use of the recurrence relation (B10.10.13)

$$\frac{d}{du}[uP_L(u) - P_{L-1}(u)] = (L+1)P_L(u), \quad (4.20)$$

we find the relation

$$\left(r\frac{d}{dr} + L + 1\right)f(r)$$
$$= \int_\alpha^1 P_L(u)[(L+1)g_L(ru) + rug_L'(ru)]\,du$$
$$= \int_\alpha^1\frac{d}{du}[u\cdot P_L(u)g_L(ru)]\,du - \int_\alpha^1 g_L(ru)P_{L-1}'(u)\,du$$
$$= [g(ru)[u\cdot P_L(u) - P_{L-1}(u)]]_\alpha^1$$
$$+ r\int_\alpha^1 g_L'(ru)P_{L-1}(u)\,du. \quad (4.21)$$

The integrated part vanishes for $\alpha = -1$ and for $\alpha = 0$ and L even; for $\alpha = 0$ and L odd it vanishes only if $g(0) = 0$. Assuming one of these conditions to hold, we obtain the recurrence relation

$$\left(\frac{L+1}{r} + \frac{d}{dr}\right)f(r) = \frac{1}{r^{L+1}}\frac{d}{dr}(r^{L+1}f)$$
$$= \int_\alpha^1 g_L'(ru)P_{L-1}(u)\,du \quad (4.22)$$

and, on repeated application of the same process,

$$\mathfrak{L}_{L-m}^{-1}\left[\frac{1}{r^{L-m}}\left(\frac{1}{r}\frac{d}{dr}\right)^m(r^{L+1}f)\right] = \left(\frac{d}{dw}\right)^m g_L(w), \quad (4.23)$$

provided that the integrated part vanishes at each stage. On putting $m = L$ and making use of the solutions (4.8) for $L = 0$, which is known from IV, Eq. (4.6) is obtained, of which (4.11) is the solution. For the bilateral ILT, the derivation is valid provided that $f(r)$ has derivatives to order $L + 1$ [and $g(w)$ to order L] everywhere in $[-w, +w]$. For the one-sided ILT, the derivatives to the same order must exist in $[0, w]$ and, in addition, all derivatives of parity opposite to L must vanish at $r = 0$ (This condition is automatically satisfied for the 2-s ILT in view of the parity conditions considered in Sec. 2). By contrast, the formulas (4.16) and (4.19) are valid provided that $f(w)H(w)$ possesses derivatives up to order $\lambda + 1$ in $(0, w]$; moreover,

even if the derivatives exist only in the generalized sense, the solution remains meaningful, provided that the integration range $[\epsilon, w]$ includes all the discontinuities, especially those occurring at $w = 0$. If $f(r)$ or any of the required derivatives have branch points at $r = 0$, the expressions (4.16) and (4.19) (using $\epsilon > 0$) are not necessarily solutions of (1.6) since the integral may diverge at its lower limit $w = 0$. However, they may still represent valid solutions when substituted in the multidimensional integrals (1.4). Fractional powers and generalized functions occurring in the radial part of expansions involving 2 or 3 constituent vectors have recently been discussed by Kay, Todd, and Silverstone.[6]

The recurrence relation (4.22) has a partner in which the index of P_L is stepped up; with the use of (B10.10.14)

$$\frac{d}{du}[uP_L(u) - P_{L+1}(u)] = -LP_L(u), \quad (4.24)$$

we obtain, under the same conditions as (4.22),

$$\left(\frac{d}{dr} - \frac{L}{r}\right) f(r) = r^L \frac{d}{dr}\frac{f(r)}{r^L} = \int g'_L(ru)P_{L+1}(u)\, du. \quad (4.25)$$

If (4.22) and (4.25) are combined, the relation

$$\frac{1}{r}\frac{d^2}{dr^2}[rf(r)] - \frac{L(L+1)}{r^2} f(r) = \int g''_L(ru)P_L(u)\, du \quad (4.26)$$

results, which was already obtained in IV. Applications of the formulas derived will be given in later papers.

[1] R. A. Sack, J. Math. Phys. **5**, 245, 252, 260 (1964); **8**, 1774 (1967).
[2] K. Ruedenberg, Theoret. Chim. Acta **7**, 359 (1967).
[3] H. S. Silverstone, J. Chem. Phys. **47**, 537 (1967).
[4] Bateman Manuscript Project, *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill, New York, 1953). Formulas in this work are directly referenced by the prefix B.
[5] R. Courant, *Differential and Integral Calculus*, (English Translation by E. J. McShane) (Blackie, London, 1936), Vol. II.
[6] K. G. Kay, H. D. Todd, and H. S. Silverstone, J. Chem. Phys. **51**, 2359, 2363 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Tensor Harmonics in Canonical Form for Gravitational Radiation and Other Applications*†

FRANK J. ZERILLI‡
Palmer Physical Laboratory, Princeton, New Jersey 08540

(Received 14 October 1969; Revised Manuscript Received 27 January 1970)

An analysis is made of the relation between the tensor harmonics given by Regge and Wheeler in 1957 and those given by Jon Mathews in 1962. This makes it possible to use the Regge–Wheeler harmonics, which are given in terms of derivatives of scalar spherical harmonics, for calculations while using Mathews' form of the harmonics [linear combinations of the elements of the product basis formed from a basis for scalar functions on the 2-sphere and a basis for symmetric tensors such that the product basis is split into sets which transform under the irreducible representations of $SO(3)$] to elucidate the properties of tensor harmonics. Thus, a convenient orthonormal set of harmonics is given which is useful in studying, for example, gravitational radiation.

1. INTRODUCTION

We discuss tensor harmonics as elements of bases for subspaces of the function space of tensor fields on a manifold (in our case, Euclidean 3-space E_3 or pseudo-Euclidean 4-space V_4). These bases are bases for the finite-dimensional irreducible representations of a group of transformations on the manifold [in our case, the group is $SO(3)$, the rotation group]. Di Sessa[1] has given a compact account of the relation between tensor harmonics and Lie-group representations. The relationship between the harmonics discussed by Mathews[2] and by Regge and Wheeler[3] is elucidated. We note that Lifshitz[4] and Lifshitz and Khalatnikov[5] have given harmonic scalars, vectors, and tensors on the 3-sphere [where the appropriate group would be $SO(4)$], but their work has not yet been considered from the group-theoretical approach of di Sessa. The tensor harmonics to be discussed are useful in the consideration of problems involving gravitational radiation considered as a perturbation on a spherically symmetric background geometry.

2. TENSOR HARMONICS

We proceed to construct rank-2 tensor harmonics as follows. We start with a basis for tensors of rank 2 by taking the tensor product of vectors of the basis $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$; that is, our basis is $\{\mathbf{e}_i \otimes \mathbf{e}_j;\ i, j = 1, 3\}$. For simplicity, we denote the tensor product by juxtaposition, $\mathbf{e}_i \otimes \mathbf{e}_j \equiv \mathbf{e}_i\mathbf{e}_j$. Now the transformation of second-rank tensors under rotations gives a representation of the rotation group, but this representation, as is well known,[6] is reducible to the sum of three irreducible representations $\mathfrak{D}^{(0)}$, $\mathfrak{D}^{(1)}$, and $\mathfrak{D}^{(2)}$. The particular linear combinations of the tensor components which transform under these irreducible representations are, respectively, the trace, the antisymmetric part, and the trace-free symmetric part of the tensor.

These three bases are not the ones which refer to the usual matrices of $\mathfrak{D}^{(0)}$, $\mathfrak{D}^{(1)}$, and $\mathfrak{D}^{(2)}$. To obtain the proper bases, we start with the spherical tensor basis for vectors $\{\mathbf{e}_{+1}, \mathbf{e}_0, \mathbf{e}_{-1}\}$, form the tensor products $\mathbf{e}_\mu \otimes \mathbf{e}_\nu$, and use the Wigner coefficients to form the appropriate spherical tensor bases for rank-2 tensors:

$$
\begin{aligned}
&\mathbf{t}_0^{(0)} = 3^{-\frac{1}{2}}(\mathbf{e}_{-}\mathbf{e}_{+1} - \mathbf{e}_0\mathbf{e}_0 + \mathbf{e}_{+1}\mathbf{e}_{-1}), \quad \text{for } \mathfrak{D}^{(0)},\\
&\mathbf{t}_{+1}^{(1)} = 2^{-\frac{1}{2}}(\mathbf{e}_0\mathbf{e}_{+1} - \mathbf{e}_{+1}\mathbf{e}_0),\\
&\mathbf{t}_0^{(1)} = 2^{-\frac{1}{2}}(\mathbf{e}_{-1}\mathbf{e}_{+1} - \mathbf{e}_{+1}\mathbf{e}_{-1}), \quad \text{for } \mathfrak{D}^{(1)},\\
&\mathbf{t}_{-1}^{(1)} = 2^{-\frac{1}{2}}(-\mathbf{e}_0\mathbf{e}_{-1} + \mathbf{e}_{-1}\mathbf{e}_0),\\
&\mathbf{t}_{+2}^{(2)} = \mathbf{e}_{+1}\mathbf{e}_{+1},\\
&\mathbf{t}_{+1}^{(2)} = 2^{-\frac{1}{2}}(\mathbf{e}_0\mathbf{e}_{+1} + \mathbf{e}_{+1}\mathbf{e}_0),\\
&\mathbf{t}_0^{(2)} = 6^{-\frac{1}{2}}(\mathbf{e}_{+1}\mathbf{e}_{-1} + 2\mathbf{e}_0\mathbf{e}_0 + \mathbf{e}_{-1}\mathbf{e}_{+1}), \quad \text{for } \mathfrak{D}^{(2)},\\
&\mathbf{t}_{-1}^{(2)} = 2^{-\frac{1}{2}}(\mathbf{e}_{-1}\mathbf{e}_0 + \mathbf{e}_0\mathbf{e}_{-1}),\\
&\mathbf{t}_{-2}^{(2)} = \mathbf{e}_{-1}\mathbf{e}_{-1}.
\end{aligned}
$$

We now take the products $\mathbf{t}_\mu^{(\lambda)} Y_{LM}$, where $\lambda = 0, 1, 2$. For $\lambda = 0$, $\mathbf{t}_0^{(0)} Y_{JM}$ transforms under $\mathfrak{D}^{(0)} \otimes \mathfrak{D}^{(J)}$, which is trivially equivalent to $\mathfrak{D}^{(J)}$. These harmonics are denoted

$$\mathbf{T}_{JM}^{(0)}(\theta, \phi) = \mathbf{t}_0^{(0)} Y_{JM}(\theta, \phi). \tag{1}$$

For $\lambda = 1$, we have a sum of three irreducible representations J, $J + 1$, and $J - 1$, and the appropriate harmonics are

$$\mathbf{T}_{JLM}^{(1)} = \sum_\mu (L, M - \mu, 1, \mu \mid L\, 1\, JM) Y_{L, M-\mu} \mathbf{t}_\mu^{(1)}, \tag{2}$$

where $L = J, J + 1$, and $J - 1$. For $\lambda = 2$, we have a sum of five irreducible representations J, $J \pm 1$, and $J \pm 2$. The harmonics are

$$\mathbf{T}_{JLM}^{(2)} = \sum_\mu (L, M - \mu, 2, \mu \mid L\, 2\, JM) Y_{L, M-\mu} \mathbf{t}_\mu^{(2)}. \tag{3}$$

The $\mathbf{T}_{JLM}^{(2)}$ were given by Mathews.[2] We remark that the tensors divide into two parity types: $\mathbf{T}_{J, J\pm1, M}^{(1)}$ and

$\mathbf{T}^{(2)}_{J,J\pm1,M}$ with parity $(-1)^{J+1}$, and $\mathbf{T}^{(0)}_{JM}$, $\mathbf{T}^{(1)}_{JJM}$, $\mathbf{T}^{(2)}_{JJM}$, and $\mathbf{T}^{(2)}_{J,J\pm2,M}$ with parity $(-1)^{J}$. Borrowing terminology from electrodynamics and following Mathews, we call the tensor harmonics with parity $(-1)^{J+1}$ "electric" and those with parity $(-1)^{J}$ "magnetic."

We define the dot product of two tensors by

$$(\mathbf{a}\otimes\mathbf{b}):(\mathbf{c}\otimes\mathbf{d})\equiv(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}).$$

Thus, if

$$\mathbf{T}=T^{ij}\mathbf{e}_i\otimes\mathbf{e}_j \quad\text{and}\quad \mathbf{S}=S_{kl}\mathbf{e}^k\otimes\mathbf{e}^l,$$

then

$$\mathbf{T}:\mathbf{S}=T^{ij}S_{kl}(\mathbf{e}_i\cdot\mathbf{e}^k)(\mathbf{e}_j\cdot\mathbf{e}^l)=T^{ij}S_{kl}\delta^k_i\delta^l_j=T^{ij}S_{ij}.$$

Now, $Y_{JM}\mathbf{e}_i\mathbf{e}_j$ is orthonormal in the inner product

$$(\mathbf{T},\mathbf{S})=\iint\mathbf{T}^*:\mathbf{S}\,d\Omega.$$

Thus, our tensor harmonics are also orthonormal, since they are obtained by a unitary transformation from the set $Y_{JM}\mathbf{e}_i\otimes\mathbf{e}_j$. The general covariant tensor function on the 2-sphere can be expanded in terms of $\mathbf{T}^{(0)}_{JM}$, $\mathbf{T}^{(1)}_{JLM}$, and $\mathbf{T}^{(2)}_{JLM}$ which give the trace, antisymmetric part, and trace-free symmetric part, respectively, of the tensor.

We can find linear combinations of the tensor harmonics of a particular degree J which are expressible in somewhat simpler terms involving spherical harmonics and their derivatives. Thus, we will show that the quantities

$$\nabla_i\nabla_jY_{JM},\quad L_i\nabla_jY_{JM},\quad L_iL_jY_{JM},$$

and

$$\tilde{x}_i\nabla_jJ_{JM},\quad \tilde{x}_jL_jY_{JM},\quad \tilde{x}_i\tilde{x}_jY_{JM} \tag{4}$$

are independent linear combinations of the harmonics of degree J. We do this straightforwardly by writing $\mathbf{L}$, $\boldsymbol{\nabla}$, and $\mathbf{x}$ in spherical tensor components:

$$\begin{aligned}
L_{\pm1}&=\mp2^{-\frac{1}{2}}(L_x\pm iL_y),\\
L_0&=L_z,\\
\nabla_{\pm1}&=\mp\frac{1}{2^{\frac{1}{2}}}\left(\frac{\partial}{\partial x}\pm\frac{i\partial}{\partial y}\right),\\
\nabla_0&=\frac{\partial}{\partial z},\\
\tilde{x}_{\pm1}&=\mp2^{-\frac{1}{2}}(x\pm iy),\\
\tilde{x}_0&=z.
\end{aligned}$$

The spherical tensor components of a vector $\mathbf{V}$ are defined by $V_\mu=\mathbf{e}_\mu\cdot\mathbf{V}$. Then, since $\mathbf{e}^*_\mu\cdot\mathbf{e}_\nu=\delta_{\mu\nu}$ and $(-1)^\mu\mathbf{e}_{-\mu}=\mathbf{e}^*_\mu$, we have the representation

$$\mathbf{V}=\sum_\mu(-1)^\mu\mathbf{e}_{-\mu}V_\mu.$$

The action of each of these operators on Y_{JM} is easily determined. We have the well-known formulas[7]

$$L_\mu Y_{JM}=(-1)^{J+M+\mu+1}[J(J+1)(2J+1)]^{\frac{1}{2}}\times\begin{pmatrix}J&1&J\\M+\mu&-\mu&-M\end{pmatrix}Y_{J,M+\mu} \tag{5a}$$

and

$$\nabla_\mu Y_{JM}=\frac{(-1)^{J+M+\mu}}{r}\times\left[J(J+1)^{\frac{1}{2}}\begin{pmatrix}J&1&J+1\\M&\mu&-M-\mu\end{pmatrix}Y_{J+1,M+\mu}+(J+1)J^{\frac{1}{2}}\begin{pmatrix}J&1&J-1\\M&\mu&-M-\mu\end{pmatrix}Y_{J-1,M+\mu}\right]. \tag{5b}$$

Now, noting that $\tilde{x}_\mu=r(\frac{4}{3}\pi)^{\frac{1}{2}}Y_{1\mu}$ and using the formula for the product of spherical harmonics found in Edmonds,[8] we obtain

$$\tilde{x}_\mu Y_{JM}=r(-1)^{J+M+\mu+1}\times\left[(J+1)^{\frac{1}{2}}\begin{pmatrix}J&1&J+1\\M&\mu&-M-\mu\end{pmatrix}Y_{J+1,M+\mu}-J^{\frac{1}{2}}\begin{pmatrix}J&1&J-1\\M&\mu&-M-\mu\end{pmatrix}Y_{J-1,M+\mu}\right]. \tag{5c}$$

We define the spherical tensor components of a second-rank tensor. Note that $\mathbf{t}^{(\lambda)*}_\mu:\mathbf{t}^{(\sigma)}_\nu=\delta_{\lambda\sigma}\delta_{\mu\nu}$. Thus, if $\mathbf{T}$ is any second-rank covariant tensor, let

$$T^{(\lambda)}_\mu\equiv\mathbf{t}^{(\lambda)}_\mu:\mathbf{T}.$$

Also, note that $\mathbf{t}^{(\lambda)*}_\mu=(-1)^{\mu+\lambda}\mathbf{t}^{(\lambda)}_{-\mu}$. Thus, $\mathbf{T}$ has the representation

$$\mathbf{T}=\sum_{\lambda,\mu}Y^{(\lambda)}_\mu\mathbf{t}^{(\lambda)*}_\mu=\sum_{\lambda,\mu}(-1)^{\mu+\lambda}T^{(\lambda)}_\mu\mathbf{t}^{(\lambda)}_{-\mu}.$$

Further, if $\mathbf{T}=\mathbf{U}\otimes\mathbf{V}$, then $T^{(\lambda)}_\mu=(\mathbf{UV})^{(\lambda)}_\mu$, where the $(\mathbf{UV})^{(\lambda)}_\mu$ are the spherical tensor components of $\mathbf{U}\otimes\mathbf{V}$ obtained by the usual angular-momentum coupling rules:

$$\begin{aligned}
(\mathbf{UV})^{(2)}_{+2}&=U_{+1}V_{+1},\\
(\mathbf{UV})^{(2)}_{+1}&=2^{-\frac{1}{2}}(U_0V_{+1}+U_{+1}V_0),\\
(\mathbf{UV})^{(2)}_{0}&=6^{-\frac{1}{2}}(U_{-1}V_{+1}+2U_0V_0+U_{+1}V_{-1}),\\
(\mathbf{UV})^{(2)}_{-1}&=2^{-\frac{1}{2}}(U_{-1}V_0+U_0V_{-1}),\\
(\mathbf{UV})^{(2)}_{-2}&=U_{-1}V_{-1},\\
(\mathbf{UV})^{(1)}_{+1}&=2^{-\frac{1}{2}}(U_0V_{+1}-U_{+1}V_0),\\
(\mathbf{UV})^{(1)}_{0}&=2^{-\frac{1}{2}}(U_{-1}V_{+1}-U_{+1}V_{-1}),\\
(\mathbf{UV})^{(1)}_{-1}&=2^{-\frac{1}{2}}(-U_0V_{-1}+U_{-1}V_0),\\
(\mathbf{UV})^{(0)}_{0}&=3^{-\frac{1}{2}}(U_{-1}V_{+1}-U_0V_0+U_{+1}V_{-1}).
\end{aligned}$$

Let $\mathbf{P}$ and $\mathbf{Q}$ denote two of the vector operators $\mathbf{L}$, ∇, and $\mathbf{e}_r$. Then, the tensor operator $\mathbf{P} \otimes \mathbf{Q}$ may be expressed in terms of its spherical tensor components

$$(\mathbf{PQ})^{(2)}_{+2} = P_{+1}Q_{+1},$$
$$(\mathbf{PQ})^{(2)}_{+1} = 2^{-\frac{1}{2}}(P_0Q_{+1} + P_{+1}Q_0),$$

etc., being careful to order the operators properly.

Utilizing the action of the operators ∇_μ, L_μ, and $\tilde{x}_\mu$ on Y_{JM} as given by Eqs. (5) together with the expressions for the second-rank tensor operators in spherical tensor components, we may now write down the relation between the tensor harmonics (3) and the expressions (4).

We now restrict our attention to symmetric tensors with trace. There are clearly six such tensor harmonics: the five trace-free harmonics $\mathbf{T}^{(0)}_{JLM}$, where $L = J$, $J \pm 1$, $J \pm 2$, and the trace $\mathbf{T}^{(2)}_{JM}$. We summarize the results of operating on Y_{JM} with $\mathbf{L}$, ∇, and $\mathbf{e}_r$ in six possible ways as follows (let the subscript s denote the "symmetric part with trace"):

$$[\nabla\nabla Y_{JM}]_s = \frac{1}{r^2}J(J+2)\left(\frac{(J+1)(J+2)}{(2J+1)(2J+3)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J+2,M} + \frac{1}{r^2}J(J+1)\left(\frac{2J(J+1)}{3(2J-1)(2J+3)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{JJM} + \frac{1}{r^2}(J+1)(J-1) \times [J(J-1)(2J-1)(2J+1)]^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J-2,M} + \left(\frac{1}{r^2}\right)3^{-\frac{1}{2}}J(J+1)\mathbf{T}^{(0)}_{JM}, \tag{6a}$$

$$[\mathbf{LL}Y_{JM}]_s = [\tfrac{1}{6}J(J+1)(2J-1)(2J+3)]^{\frac{1}{2}}\mathbf{T}^{(2)}_{JJM} - 3^{-\frac{1}{2}}J(J+1)\mathbf{T}^{(0)}_{JM}, \tag{6b}$$

$$[\mathbf{e}_r\mathbf{e}_rY_{JM}]_s = \left(\frac{(J+1)(J+2)}{(2J+1)(2J+3)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J+2,M} + \left(\frac{2J(J+1)}{3(2J-1)(2J+3)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{JJM} + \left(\frac{J(J-1)}{(2J-1)(2J+1)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J-2,M} - 3^{-\frac{1}{2}}\mathbf{T}^{(0)}_{JM}, \tag{6c}$$

$$[\mathbf{e}_r\nabla Y_{JM}]_s = \frac{1}{r}[J(J+1)]^{\frac{1}{2}} \times \left[-\left(\frac{J(J+2)}{(2J+1)(2J+3)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J+2,M} - 3[6(2J-1)(2J+3)]^{-\frac{1}{2}}\mathbf{T}^{(2)}_{JJM} + \left(\frac{(J+1)(J-1)}{(2J-1)(2J+1)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J-2,M}\right], \tag{6d}$$

$$[\mathbf{L}\nabla Y_{JM}]_s = -\frac{1}{r}J\left(\frac{J(J+1)(J+2)}{2(2J+1)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J+1,M} - \frac{1}{r}(J+1)\left(\frac{(J-1)J(J+1)}{2(2J+1)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J-1,M}, \tag{6e}$$

$$[\mathbf{e}_r\mathbf{L}Y_{JM}]_s = [\tfrac{1}{2}J(J+1)]^{\frac{1}{2}}\left[\left(\frac{(J+2)}{(2J+1)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J+1,M} - \left(\frac{(J-1)}{(2J+1)}\right)^{\frac{1}{2}}\mathbf{T}^{(2)}_{J,J-1,M}\right]. \tag{6f}$$

Thus, the symmetric parts of the four tensors $\mathbf{LL}Y_{JM}$, $\nabla\nabla Y_{JM}$, $\mathbf{e}_r\mathbf{e}_rY_{JM}$, and $\mathbf{e}_r\nabla Y_{JM}$, which have parity $(-1)^J$ ("magnetic"), and the two tensors $\mathbf{L}\nabla Y_{JM}$ and $\mathbf{e}_r\mathbf{L}Y_{JM}$, which have parity $(-1)^{J+1}$ ("electric"), will suffice to represent an arbitrary symmetric tensor on the 2-sphere.

We now write explicit expressions for $[\nabla\nabla Y_{JM}]_s$, $[\mathbf{LL}Y_{JM}]_s$, and $[\mathbf{L}\nabla Y_{JM}]_s$ in spherical coordinates. In each operator, we use covariant derivatives in the spherical coordinate system. Thus, the (ij) component of $\nabla\nabla Y_{JM}$ is

$$Y_{JM;i;j} = \nabla_j\nabla_iY_{JM},$$

the (ij) component of $\mathbf{L}\nabla Y_{JM}$ is

$$L_i\nabla_jY_{JM} = -ig_{ik}g^{-\frac{1}{2}}\epsilon^{kln}\tilde{x}_l\nabla_nY_{JM} = -ir\epsilon^k_iY_{JM;j;k},$$

where

$$\epsilon^2_3 = \sin\theta \quad \text{and} \quad \epsilon^3_2 = -1/(\sin\theta).$$

Finally, the (ij) component of $\mathbf{LL}Y_{JM}$ is

$$-r\epsilon^k_i\nabla_k r\epsilon^l_j\nabla_lY_{JM}.$$

In the following formulas, components determined by the symmetry of the tensor are denoted by $*$:

$$[\nabla\nabla Y_{JM}]_s = \begin{pmatrix} 0 & -\frac{1}{r}\frac{\partial Y_{JM}}{\partial\theta} & -\frac{1}{r}\frac{\partial Y_{JM}}{\partial\phi} \\ * & \frac{\partial^2Y_{JM}}{\partial\theta^2} & \tfrac{1}{2}X_{JM} \\ * & * & -\left(\frac{\partial^2Y_{JM}}{\partial\theta^2} + J(J+1)Y_{JM}\right)\sin^2\theta \end{pmatrix}, \tag{7a}$$

$$[\mathbf{L}\nabla Y_{JM}]_{\mathrm{s}} = -\tfrac{1}{2}ir\begin{pmatrix} 0 & \dfrac{1}{r}\sin\theta\dfrac{\partial Y_{JM}}{\partial\phi} & -\dfrac{\sin\theta}{r}\dfrac{\partial Y_{JM}}{\partial\theta} \\ * & -\dfrac{1}{\sin\theta}X_{JM} & (\sin\theta)W_{JM} \\ * & * & (\sin\theta)X_{JM} \end{pmatrix}, \tag{7b}$$

$$[\mathbf{L}\mathbf{L}Y_{JM}]_{\mathrm{s}} = -\tfrac{1}{2}r^2\begin{pmatrix} 0 & \dfrac{1}{r}\dfrac{\partial Y_{JM}}{\partial\theta} & \dfrac{1}{r}\dfrac{\partial Y_{JM}}{\partial\theta} \\ * & -2\left(\dfrac{\partial^2 Y_{JM}}{\partial\theta^2} + J(J+1)Y_{JM}\right) & -X_{JM} \\ * & * & 2\sin^2\theta\dfrac{\partial^2 Y_{JM}}{\partial\theta^2} \end{pmatrix}, \tag{7c}$$

$$[\mathbf{e}_r\mathbf{e}_r Y_{JM}]_{\mathrm{s}} = \begin{pmatrix} Y_{JM} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{7d}$$

$$[\mathbf{e}_r\nabla Y_{JM}]_{\mathrm{s}} = \tfrac{1}{2}\begin{pmatrix} 0 & \dfrac{\partial Y_{JM}}{\partial\theta} & \dfrac{\partial Y_{JM}}{\partial\phi} \\ * & 0 & 0 \\ * & 0 & 0 \end{pmatrix}, \tag{7e}$$

$$[\mathbf{e}_r\mathbf{L}Y_{JM}]_{\mathrm{s}} = \tfrac{1}{2}ir\begin{pmatrix} 0 & \dfrac{1}{\sin\theta}\dfrac{\partial Y_{JM}}{\partial\phi} & -(\sin\theta)\left(\dfrac{\partial Y_{JM}}{\partial\theta}\right) \\ * & 0 & 0 \\ * & 0 & 0 \end{pmatrix}, \tag{7f}$$

where

$$X_{JM} = 2\frac{\partial}{\partial\phi}\left(\frac{\partial}{\partial\theta} - \cot\theta\right)Y_{JM}$$

and

$$W_{JM} = \left(\frac{\partial^2}{\partial\theta^2} - (\cot\theta)\frac{\partial}{\partial\theta} - \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right)Y_{JM}.$$

It is now clear that we can separate these into three tensors, which are orthogonal to the 2-sphere, and three tensors, which are tangent to the 2-sphere. The three orthogonal to the 2-sphere (one electric and two magnetic) are

$$\mathbf{a}_{JM} = [\mathbf{e}_r\mathbf{e}_r Y_{JM}]_{\mathrm{s}}, \tag{8a}$$

$$\mathbf{b}_{JM} = 2^{\frac{1}{2}}r[J(J+1)]^{-\frac{1}{2}}[\mathbf{e}_r\nabla Y_{JM}]_{\mathrm{s}}, \tag{8b}$$

$$\mathbf{c}_{JM} = 2^{\frac{1}{2}}[J(J+1)]^{-\frac{1}{2}}[\mathbf{e}_r\mathbf{L}Y_{JM}]_{\mathrm{s}}. \tag{8c}$$

Those tangent to the 2-sphere (one electric and two magnetic) are

$$\mathbf{d}_{JM} = 2^{\frac{1}{2}}r[J(J+1)(J-1)(J+2)]^{-\frac{1}{2}} \times \{[\mathbf{L}\nabla Y_{JM}]_{\mathrm{s}} + (1/r)[\mathbf{e}_r\mathbf{L}Y_{JM}]_{\mathrm{s}}\}, \tag{9a}$$

$$\mathbf{e}_{JM} = r^2\{[\nabla\nabla Y_{JM}]_{\mathrm{s}} + (2/r)[\mathbf{e}_r\nabla Y_{JM}]_{\mathrm{s}}\}, \tag{9b}$$

$$\mathbf{h}_{JM} = [\mathbf{L}\mathbf{L}Y_{JM}]_{\mathrm{s}} + r[\mathbf{e}_r\nabla Y_{JM}]_{\mathrm{s}}. \tag{9c}$$

By looking at the expressions for these quantities in terms of $\mathbf{T}^{(\lambda)}_{JLM}$ and also the explicit representations (8) and (9), we see that all the inner products between the different quantities in (8) and (9) vanish except for the inner product of $\mathbf{e}_{JM}$ and $\mathbf{h}_{JM}$. However, the sum and difference of these give two orthogonal harmonics

$$\mathbf{f}_{JM} = [2J(J+1)(J-1)(J+2)]^{-\frac{1}{2}}(\mathbf{e}_{JM} + \mathbf{h}_{JM}), \tag{10a}$$

$$\mathbf{g}_{JM} = -[2^{\frac{1}{2}}J(J+1)]^{-1}(\mathbf{e}_{JM} - \mathbf{h}_{JM}). \tag{10b}$$

Note that

$$\mathbf{g}_{JM} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sin^2\theta \end{pmatrix} Y_{JM},$$

$$\mathbf{f}_{JM} = r^2[2J(J+1)(J-1)(J+2)]^{-\frac{1}{2}} \times \begin{pmatrix} 0 & 0 & 0 \\ 0 & W_{JM} & X_{JM} \\ 0 & * & -(\sin^2\theta)W_{JM} \end{pmatrix},$$

$$\mathbf{d}_{JM} = -ir^2[2J(J+1)(J-1)(J+2)]^{-\frac{1}{2}} \times \begin{pmatrix} 0 & 0 & \\ 0 & -(1/\sin\theta)X_{JM} & (\sin\theta)W_{JM} \\ 0 & * & (\sin\theta)X_{JM} \end{pmatrix}.$$

Thus, $\{\mathbf{a}_{JM}, \mathbf{b}_{JM}, \mathbf{c}_{JM}, \mathbf{d}_{JM}, \mathbf{g}_{JM}, \mathbf{f}_{JM}\}$ is an orthonormal set of tensor harmonics, distinguished by parity and by whether they are tangent or orthogonal to the 2-sphere. Note that, of the harmonics tangent to the 2-sphere, the electric harmonic $\mathbf{d}_{JM}$ and the magnetic harmonic $\mathbf{f}_{JM}$ are trace-free. These are, in fact, the transverse traceless electric and magnetic tensor harmonics $\mathbf{T}^{(e)}_{JM}$ and $\mathbf{T}^{(m)}_{JM}$ given by Mathews,[2] as can be seen by writing them in terms of the $\mathbf{T}^{(\lambda)}_{JLM}$ harmonics

$$\mathbf{d}_{JM} = -\mathbf{T}^{(e)}_{JM},$$
$$\mathbf{f}_{JM} = \mathbf{T}^{(m)}_{JM}.$$

Note that $\mathbf{b}_{JM}$ and $\mathbf{c}_{JM}$ are zero for $J = 0$, while $\mathbf{d}_{JM} = -\mathbf{T}^{(e)}_{JM}$ and $\mathbf{f}_{JM} = \mathbf{T}^{(m)}_{JM}$ are zero for $J = 0$ and $J = 1$. For gravitational radiation in the radiation field, $\mathbf{d}_{JM}$ and $\mathbf{f}_{JM}$ are the significant harmonics.

These harmonics can be identified with the decomposition given by Regge and Wheeler.[3] However, their set consisted of the harmonics $\mathbf{a}_{JM}$, $\mathbf{b}_{JM}$, $\mathbf{c}_{JM}$, $\mathbf{d}_{JM}$, $\mathbf{e}_{JM}$, and $\mathbf{g}_{JM}$. This is not an orthogonal set, since $(\mathbf{e}_{JM}, \mathbf{g}_{JM}) \neq 0$.

Up to now, we have considered tensors on a 2-sphere embedded in a 3-dimensional Euclidean space. Let us extend these tensor harmonics to tensors in a 4-dimensional pseudo-Euclidean space. We have three harmonics lying in the 2-spheres and, in 3-space, we have three additional harmonics orthogonal to the 2-spheres: $[\mathbf{e}_r\mathbf{e}_r Y_{JM}]_s$, $[\mathbf{e}_r\nabla Y_{JM}]_s$, and $[\mathbf{e}_r\mathbf{L}Y_{JM}]_s$. Let $\mathbf{e}_t$ denote the covector which is identified with the unit tangent vector along the time direction by means of the pseudo-Euclidean metric. Split the tensor space at each point of V_4 into the direct sum of the tensor space tangent to the Euclidean subspace E_3 plus the space spanned by $\{\mathbf{e}_t \otimes \mathbf{e}_t\} \oplus \{\mathbf{e}_t \odot \mathbf{e}_i;\ i = 1, 3\}$. $\odot$ means symmetric tensor product. Clearly, $\{\mathbf{e}_t \otimes \mathbf{e}_t\}$ is invariant under $SO(3)$, while $\{\mathbf{e}_t \odot \mathbf{e}_i;\ = 1, 3\}$ transforms like a vector basis. Thus, we write

$$\mathbf{T} = \sum_{i,j}^{3} T_{ij}\mathbf{e}_i \otimes \mathbf{e}_j + \sum_{i=1}^{3} T_{0i}\mathbf{e}_t \odot \mathbf{e}_i + T_{00}\mathbf{e}_t \otimes \mathbf{e}_t.$$

Thus, a tensor field is the direct sum of three terms: The first term can be expanded in terms of the previously described harmonics, the second term can be expanded in $\mathbf{e}_t \otimes \mathbf{Y}_{JLM}$, which is a basis for $\mathfrak{D}^{(0)} \otimes \mathfrak{D}^{(J)} \equiv \mathfrak{D}^{(J)}$, and the third term can be expanded in $\mathbf{e}_t \otimes Y_{JM}\mathbf{e}_t = \mathbf{e}_t\mathbf{e}_t Y_{JM}$, which is a basis for $\mathfrak{D}^{(0)} \otimes \mathfrak{D}^{(0)} \otimes \mathfrak{D}^{(J)} \equiv \mathfrak{D}^{(J)}$.

Let μ, ν range from 0 to 3; let $\mathbf{e}^\mu$ be the basis of covectors in Minkowski space. (Thus, in natural coordinates which we have been using, $\mathbf{e}^0 = \mathbf{e}_t$, $\mathbf{e}^1 = \mathbf{e}_x$, $\mathbf{e}^2 = \mathbf{e}_y$, and $\mathbf{e}^3 = \mathbf{e}_z$.) Then $\{\mathbf{e}^\mu \otimes \mathbf{e}^\nu\}$ is a basis for covariant tensors and, if

$$\mathbf{T} = T_{\mu\nu}\mathbf{e}^\mu \otimes \mathbf{e}^\nu, \quad \mathbf{S} = S_{\mu\nu}\mathbf{e}^\mu \otimes \mathbf{e}^\nu,$$

then

$$\mathbf{T}:\mathbf{S} = \eta^{\mu\lambda}\eta^{\nu\kappa}T_{\mu\nu}S_{\lambda\kappa},$$

where $\eta^{\mu\nu} = \{-1, +1, +1, +1\}$. As before, we define

$$(\mathbf{T}, \mathbf{S}) = \iint \mathbf{T}^*:\mathbf{S}\, d\Omega.$$

The four additional harmonics are

$$\begin{aligned} \mathbf{a}^{(0)}_{JM} &\equiv [\mathbf{e}_t\mathbf{e}_t Y_{JM}]_s, \\ \mathbf{a}^{(1)}_{JM} &\equiv 2^{\frac{1}{2}}i[\mathbf{e}_t\mathbf{e}_r Y_{JM}]_s, \\ \mathbf{b}^{(0)}_{JM} &\equiv 2^{\frac{1}{2}}ir[J(J+1)]^{-\frac{1}{2}}[\mathbf{e}_t\nabla Y_{JM}]_s, \\ \mathbf{c}^{(0)}_{JM} &\equiv 2^{\frac{1}{2}}i[J(J+1)]^{-\frac{1}{2}}[\mathbf{e}_t\mathbf{L}Y_{JM}]_s. \end{aligned} \tag{11}$$

Any symmetric covariant tensor can be expanded in terms of the harmonics (8), (9a), (10), and (11):

$$\begin{aligned} \mathbf{T} = \sum_{JM} [&A^{(0)}_{JM}\mathbf{a}^{(0)}_{JM} + A^{(1)}_{JM}\mathbf{a}^{(1)}_{JM} + A_{JM}\mathbf{a}_{JM} \\ &+ B^{(0)}_{JM}\mathbf{b}^{(0)}_{JM} + B_{JM}\mathbf{b}_{JM} \\ &+ Q^{(0)}_{JM}\mathbf{c}^{(0)}_{JM} + Q_{JM}\mathbf{c}_{JM} \\ &+ G_{JM}\mathbf{g}_{JM} + D_{JM}\mathbf{d}_{JM} + F_{JM}\mathbf{h}_{JM}], \end{aligned}$$

where the coefficients $A_{JM}, B_{JM}, \cdots$ are functions of r and t. In the Regge–Wheeler set of harmonics, $\mathbf{e}_{JM}$ is used in place of $\mathbf{f}_{JM}$.

3. CONCLUSION

This formalism is useful in the solution of the equations for linear perturbations from a spherically

symmetric background geometry. In particular, the orthogonal set of tensor harmonics enables one to solve such perturbation equations in the case where there is a source term. The author treats the problem of a particle falling in a Schwarzschild background geometry in his Ph.D. thesis.[9] The static and stationary parts of the gravitational field produced by the falling particle are given by the $J = 0$ and $J = 1$ harmonics, while the radiation is described by the $J \geq 2$ harmonics. Other possibilities suggest themselves: the interaction of electromagnetic fields with the gravitational field and the perturbations produced by continuous distributions of matter (dust) on a spherically symmetric background.

ACKNOWLEDGMENT

I am indebted to Professor John A. Wheeler for stimulating me to carry out this study.

* This material appears in the author's Ph.D. thesis, Princeton University, 1969.

† Work supported in part by NSF Grant GP 7769 and an NSF Graduate Fellowship.

‡ Present Address: Physics Department, University of North Carolina, Chapel Hill, N.C. 27514.

[1] A. di Sessa, A.B. thesis, Princeton University, 1969.
[2] J. Mathews, J. Soc. Ind. Appl. Math. **10**, 768 (1962).
[3] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[4] E. Lifshitz, J. Phys. (USSR) **10**, 116 (1946).
[5] E. Lifshitz and I. Khalatnikov, Advan. Phys. **12**, 185 (1963).
[6] E. P. Wigner, *Group Theory* (Academic, New York, 1959), p. 168.
[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957), p. 84.
[8] See Ref. 7, p. 63.
[9] F. Zerilli, Ph.D. thesis, Princeton University, 1969.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Separable Schrödinger Equations for Two Interacting Particles in External Fields*

R. SEILER AND R. H. PRATT
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 7 July 1969; Revised Manuscript Received 14 October 1969)

We derive the most general form of the Schrödinger equation for two interacting particles in an external field which can be separated by a linear transformation of the coordinates. In particular, the transformation to center-of-mass and relative coordinates is shown to separate only those Hamiltonians with external potentials of the 3-dimensional harmonic oscillator type. The most general allowed form contains in addition specified cubic terms in the potentials.

1. INTRODUCTION

Consider the Hamiltonian

$$H(x_1, x_2) = -(2m_1)^{-1}\Delta_1 - (2m_2)^{-1}\Delta_2 + V_1(x_1) + V_2(x_2) + V_3(x_1 - x_2) \quad (1)$$

of two interacting nonrelativistic massive particles in an external field. The question arises whether $H(x_1, x_2)$ is equivalent to the sum of two commuting 1-particle Hamiltonians

$$\tilde{H}(\tilde{x}_1, \tilde{x}_2) = \tilde{H}_1(\tilde{x}_1) + \tilde{H}_2(\tilde{x}_2) = UHU^{-1}, \quad (2)$$

the equivalence transformation being induced by a linear coordinate transformation. (Instead of linear transformations we could as well consider arbitrary transformations which leave the kinetic energy diagonal.) H is called separable $H \in S$ if such a decomposition is possible. Talmi[1] noted that for equal-mass particles and central potentials $V(r)$ H separates under a transformation to center of mass and relative coordinates only if $V_1(r) = V_2(r) = \frac{1}{2}Kr^2 + \text{const}$. He and others[1–4] have used this to discuss nuclear problems outside closed shells; Kestner and Sinanoğlu[5] have used the same separability in a discussion of 2-electron atoms. It is the purpose of this note to find all separable Hamiltonians of the form (1) where the potentials V are elements of the space of the distributions $\mathfrak{D}'$.[6]

Our principal interest lies in the classification of the potentials $V(x)$ of $H \in S$ and we do not care about the detailed form of $\tilde{H}_1(\tilde{x}_1)$ and $\tilde{H}_2(\tilde{x}_2)$. If there is a linear coordinate transformation T at all separating H, then there is a class equally acceptable which we get by multiplying T with all possible rotations and dilatations of the coordinates $\tilde{x}_1$ and $\tilde{x}_2$ separately. It is therefore enough to consider for H the special form

$$\tilde{H}(\tilde{y}_1, \tilde{y}_2) = -(2\mu)^{-1}(\Delta_1 + \Delta_2) + \tilde{U}_1(\tilde{y}_1) + \tilde{U}_2(\tilde{y}_2), \quad \mu = m_1m_2/(m_1 + m_2). \quad (2')$$

symmetric background geometry. In particular, the orthogonal set of tensor harmonics enables one to solve such perturbation equations in the case where there is a source term. The author treats the problem of a particle falling in a Schwarzschild background geometry in his Ph.D. thesis.[9] The static and stationary parts of the gravitational field produced by the falling particle are given by the $J = 0$ and $J = 1$ harmonics, while the radiation is described by the $J \geq 2$ harmonics. Other possibilities suggest themselves: the interaction of electromagnetic fields with the gravitational field and the perturbations produced by continuous distributions of matter (dust) on a spherically symmetric background.

ACKNOWLEDGMENT

I am indebted to Professor John A. Wheeler for stimulating me to carry out this study.

* This material appears in the author's Ph.D. thesis, Princeton University, 1969.

† Work supported in part by NSF Grant GP 7769 and an NSF Graduate Fellowship.

‡ Present Address: Physics Department, University of North Carolina, Chapel Hill, N.C. 27514.

[1] A. di Sessa, A.B. thesis, Princeton University, 1969.
[2] J. Mathews, J. Soc. Ind. Appl. Math. **10**, 768 (1962).
[3] T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[4] E. Lifshitz, J. Phys. (USSR) **10**, 116 (1946).
[5] E. Lifshitz and I. Khalatnikov, Advan. Phys. **12**, 185 (1963).
[6] E. P. Wigner, *Group Theory* (Academic, New York, 1959), p. 168.
[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957), p. 84.
[8] See Ref. 7, p. 63.
[9] F. Zerilli, Ph.D. thesis, Princeton University, 1969.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Separable Schrödinger Equations for Two Interacting Particles in External Fields*

R. SEILER AND R. H. PRATT
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 7 July 1969; Revised Manuscript Received 14 October 1969)

We derive the most general form of the Schrödinger equation for two interacting particles in an external field which can be separated by a linear transformation of the coordinates. In particular, the transformation to center-of-mass and relative coordinates is shown to separate only those Hamiltonians with external potentials of the 3-dimensional harmonic oscillator type. The most general allowed form contains in addition specified cubic terms in the potentials.

1. INTRODUCTION

Consider the Hamiltonian

$$H(x_1, x_2) = -(2m_1)^{-1}\Delta_1 - (2m_2)^{-1}\Delta_2 + V_1(x_1) + V_2(x_2) + V_3(x_1 - x_2) \quad (1)$$

of two interacting nonrelativistic massive particles in an external field. The question arises whether $H(x_1, x_2)$ is equivalent to the sum of two commuting 1-particle Hamiltonians

$$\tilde{H}(\tilde{x}_1, \tilde{x}_2) = \tilde{H}_1(\tilde{x}_1) + \tilde{H}_2(\tilde{x}_2) = UHU^{-1}, \quad (2)$$

the equivalence transformation being induced by a linear coordinate transformation. (Instead of linear transformations we could as well consider arbitrary transformations which leave the kinetic energy diagonal.) H is called separable $H \in S$ if such a decomposition is possible. Talmi[1] noted that for equal-mass particles and central potentials $V(r)$ H separates under a transformation to center of mass and relative coordinates only if $V_1(r) = V_2(r) = \frac{1}{2}Kr^2 + \text{const}$. He and others[1–4] have used this to discuss nuclear problems outside closed shells; Kestner and Sinanoğlu[5] have used the same separability in a discussion of 2-electron atoms. It is the purpose of this note to find all separable Hamiltonians of the form (1) where the potentials V are elements of the space of the distributions $\mathfrak{D}'$.[6]

Our principal interest lies in the classification of the potentials $V(x)$ of $H \in S$ and we do not care about the detailed form of $\tilde{H}_1(\tilde{x}_1)$ and $\tilde{H}_2(\tilde{x}_2)$. If there is a linear coordinate transformation T at all separating H, then there is a class equally acceptable which we get by multiplying T with all possible rotations and dilatations of the coordinates $\tilde{x}_1$ and $\tilde{x}_2$ separately. It is therefore enough to consider for H the special form

$$\tilde{H}(\tilde{y}_1, \tilde{y}_2) = -(2\mu)^{-1}(\Delta_1 + \Delta_2) + \tilde{U}_1(\tilde{y}_1) + \tilde{U}_2(\tilde{y}_2), \quad \mu = m_1m_2/(m_1 + m_2). \quad (2')$$

The Hamiltonian (1) is of the form, up to a trivial dilatation,

$$H = -(2\mu)^{-1}(\Delta_1 + \Delta_2) + U_1(y_1) + U_2(y_2) + U_3(\alpha y_1 - \beta y_2),$$
$$\alpha = \left(\frac{m_2}{m_1+m_2}\right)^{\frac{1}{2}}, \quad \beta = \left(\frac{m_1}{m_1+m_2}\right)^{\frac{1}{2}}, \qquad (1')$$
$$y_1 = \frac{1}{\alpha}x_1, \quad y_2 = \frac{1}{\beta}x_2.$$

The problem reduces therefore to finding all Hamiltonians H of the form (1') which can be transformed into $\tilde{H}$ as defined by (2'). If we consider the kinetic parts of H and $\tilde{H}$ so defined, we realize that the set of all possible linear coordinate transformations now reduces to the rotations $O(6)$ in y_1 and y_2.[7]

It is convenient to begin with the physical situation of separability in a transformation to center of mass and relative coordinates. In this case

$$V_3 = V_3(\tilde{x}_2), \qquad (3)$$

and there is no restriction on its form; V_1 and V_2 must be quadratic functions whose leading coefficients are related. We then consider the case of more general transformations and find that cubic functions are now also allowed for V_1 and V_2 at the expense of restricting V_3 to be a related cubic function.[8]

2. THE CENTER-OF-MASS CASE

In this section we use the transformation to the center of mass and the relative coordinate of the particles to identify a large number of potentials. The characteristic property of this simple transformation is that it places no restrictions on the form of V_3:

Theorem 1: The Hamiltonian (1) can be separated by the coordinate transformation

$$T_{\text{c.m.}}: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} \tilde{x}_1 = M^{-1}(m_1x_1 + m_2x_2) \\ \tilde{x}_2 = x_1 - x_2 \end{pmatrix}, \quad M = m_1 + m_2, \qquad (4)$$

if the potentials V_1 and V_2 are of the form

$$V_1 = \alpha^{-2}a_{ik}x_1^ix_1^k + b_lx^l + c,$$
$$V_2 = \beta^{-2}a_{ik}x_2^ix_2^k + b_l'x_2^l + c'. \qquad (5)$$

Conversely, if V_1 and V_2 are elements of[6] $\mathfrak{D}'$ and H can be separated by $T_{\text{c.m.}}$, then V_1 and V_2 are of the form (5) with no restrictions on the constants.

Proof:

(i) To prove the theorem,it is simpler to use instead of (1) and (2) the corresponding forms (1') and (2'). Expression (4) will change to

$$T_{\text{c.m.}}: \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \to \begin{pmatrix} \tilde{y}_1 = \beta y_1 + \alpha y_2 \\ \tilde{y}_2 = \alpha y_1 - \beta y_2 \end{pmatrix},$$

which is obviously a rotation in the coordinates y_1 and y_2.

(ii) The first part of the proof can be checked by simple calculation. We therefore concentrate on the second part and assume

$$U_1(y_1) + U_2(y_2) + U_3(\alpha y_1 - \beta y_2) = \tilde{U}_1(\beta y_1 + \alpha y_2) + \tilde{U}_2(\alpha y_1 - \beta y_2). \qquad (6)$$

Since U_3 is of the same form as U_2, it is enough to discuss the functional equation

$$U_1(y_1) + U_2(y_2) = \tilde{U}_1(\beta y_1 + \alpha y_2) + \tilde{U}_2(\alpha y_1 - \beta y_2). \qquad (6')$$

(iii) We show that U_i and $\tilde{U}_i$ are smooth functions. Smear the equation (6') with a test function $\phi \in \mathfrak{D}$ in the variable y_1:

$$\int dy_1\phi(y_1)U_1(y_1) + U_2(y_2)\int dy_1\phi(y_1) = \int dy_1\phi(y_1)\tilde{U}_1(\beta y_1 + \alpha y_2) + \int dy_2\phi(y_1)\tilde{U}_2(\alpha y_1 - \beta y_2).$$

On the left-hand side we have a linear form in $U_2(y_2)$ and on the right-hand side two integrals of the convolution type. The assertion follows from the fact that a convolution of $\phi \in \mathfrak{D}$ with $f \in \mathfrak{D}'$ is smooth.[9] A similar argument works for U_1, $\tilde{U}_1$, and $\tilde{U}_2$.

(iv) Since the differentiability of the U's has been established, we can apply the operator

$$D_{ik} = \frac{\partial}{\partial y_1^i}\frac{\partial}{\partial y_2^k}, \quad i,k = 1,2,3, \qquad (7)$$

on (6'). Since the left-hand side will vanish for all i and k, one obtains

$$\frac{\partial^2\tilde{U}_1}{\partial\tilde{y}_1^i\partial\tilde{y}_1^k} = \frac{\partial^2\tilde{U}_2}{\partial\tilde{y}_2^i\partial\tilde{y}_2^k}$$

Since U_1 and U_2 are functions of independent variables, the last expression has to be a constant $2a_{ik}$ and therefore

$$\tilde{U}_1 = a_{ik}\tilde{y}_1^i\tilde{y}_1^k + d_i\tilde{y}_1^i + e,$$
$$\tilde{U}_2 = a_{ik}\tilde{y}_2^i\tilde{y}_2^k + f_iy_2^i + g,$$

from which (5) follows.

With a rotation the quadratic terms take the form $\frac{1}{2}(k_1x^2 + k_2y^2 + k_3z^2)$; i.e., three 1-dimensional oscillators of different spring constants. The separability

of this type of potential has been exploited by Krieger[10] in work on 2-particle states in a deformed nucleus. If the masses $m_1 \neq m_2$, it should be noted that (5) requires that the particles be in potentials which differ as their masses. This is not likely to be realized in situations of physical interest. However, it is possible that the difference from this situation is small compared to all the potentials of the problem and can consequently be treated with perturbation methods. This approach has been followed by Ananthanaryanan[11] in discussing a Λ particle and a nucleon in a hypernucleus.

3. THE GENERAL CASE

In the previous section we considered the Hamiltonians which could be separated by the special coordinate transformation $T_{\text{c.m.}}$(4). Now we are looking for all separable Hamiltonians $H \in S$. We have seen it is sufficient to consider the potential part of H and restrict the transformations to the rotations $O(6)$ in y_1, y_2.

To simplify the discussion we introduce the following equivalence relation [not the usual equivalence relation of Eq. (2)] between Hamiltonians:

$$H \sim H' \quad \text{iff} \quad V_1(x_1) + V_2(x_2) + V_3(x_1 - x_2) = V_1'(R_1x_1) + V_2'(R_2x_2) + V_3'(R_1x_1 - R_2x_2), \quad R_1, R_2 \in O(3). \quad (8)$$

In each class of separable Hamiltonians (with respect to the equivalence relation $\sim$), there is a representative of the form specified below.

Theorem 2: Let the 6-dimensional vector space $W = \{(x_1, x_2)\}$ be decomposed in three 2-dimensional subspaces

$$W = W_1 \oplus W_2 \oplus W_3, \quad W_i = \{(x_1^{l(i)}, x_2^{k(i)})\}, \quad i = 1, 2, 3. \quad (9)$$

The Hamiltonian $H(1)$ can be separated by the coordinate transformation T reducing, according to (9), to

$$T = T_1 \oplus T_2 \oplus T_3,$$

$$T_i : \begin{pmatrix} x_1^{l(i)} \\ x_2^{k(i)} \end{pmatrix} \to \begin{pmatrix} (\lambda_i/\alpha)x_1^{l(i)} - (\mu_i/\beta)x_2^{k(i)} \\ (\mu_i/\alpha)x_1^{l(i)} + (\lambda_i/\beta)x_2^{k(i)} \end{pmatrix}, \quad \lambda_i^2 + \mu_i^2 = 1,$$

if the potentials are of the form (with an arbitrary linear and a constant term and parameters χ_i, λ_i, μ_i, a_i, and b_i)

$$V_1(x_1) = \sum_i [\chi_i K(\alpha, \beta, \lambda_i, \mu_i)(\alpha^{-1}x_1^{l(i)})^3 + L(\alpha, \beta, \lambda_i, \mu_i, a_i, b_i)(\alpha^{-1}x_1^{l(i)})^2],$$

$$V_2(x_2) = \sum_i [\chi_i K(\beta, -\alpha, \lambda_i, \mu_i)(\beta^{-1}x_2^{k(i)})^3 + L(\beta, -\alpha, \lambda_i, \mu_i, b_i, a_i)(\beta^{-1}x_2^{k(i)})^2],$$

$$V_3(x_1 - x_2) = \sum_i \frac{\lambda_i\mu_i}{\alpha\beta}[\chi_i(x_1^{l(i)} - x_2^{k(i)})^3 + (a_i - b_i)(x_1^{l(i)} - x_2^{k(i)})^2], \quad (10)$$

$$K(\alpha, \beta, \lambda, \mu) = \beta^{-1}[\alpha\beta(\lambda^2 - \mu^2) + \lambda\mu(\beta^2 - \alpha^2)],$$

$$L(\alpha, \beta, \lambda, \mu, a, b) = a\lambda^2 + b\mu^2 - \alpha\beta^{-1}(a - b)\lambda\mu.$$

Conversely, if the Hamiltonian H' is separable and the potentials V_1', V_2', and V_3' are elements from[6] $\mathfrak{D}$ then H' is equivalent to H ($H' \sim H$) with potentials having the form (10).

Proof:

(i) Again, the first part of the theorem is a straightforward algebraic calculation, and we consider the second part. Suppose

$$U_1'(y_1') + U_2'(y_2') + U_3'(\alpha y_1' - \beta y_2') = \tilde{U}_1(\tilde{y}_1) + \tilde{U}_2(\tilde{y}_2), \quad \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} = R\begin{pmatrix} y_1' \\ y_2' \end{pmatrix}, \quad R \in O(6). \quad (11)$$

Lemma: For a given rotation $R \in O(6)$ there exist $S \in O(3) \oplus O(3)$ and $Q \in O(3) \oplus O(3)$ such that $N(R) = SRQ$ is a direct sum of three 2-dimensional rotations T_i.

In an appendix we will reformulate the lemma in a more abstract and more transparent manner and give a proof. Now we consider again the Eq. (11) but using $N(R)$ instead of R. With our definition of equivalence, there are equivalent potentials U and $\tilde{U}$ such that

$$U_1(y_1) + U_2(y_2) + U_3(\alpha y_1 - \beta y_2) = \tilde{U}_1(\tilde{y}_1) + \tilde{U}_2(\tilde{y}_2), \quad (11')$$

where

$$\tilde{y} = N(R)y.$$

Since N is the direct sum of three 2-dimensional rotations, it is enough to consider the problem just in one of the three 2-dimensional subspaces $W_{i,k}$ [we do not indicate the dependence on (i, k)]:

$$U_1(y_1) + U_2(y_2) + U_3(\alpha y_1 - \beta y_2) = \tilde{U}_1(\lambda y_1 - \mu y_2) + \tilde{U}_2(\mu y_1 + \lambda y_2), \quad \lambda^2 + \mu^2 = 1. \quad (11'')$$

(ii) By a similar argument,as in the proof of Theorem 1,the U's are smooth functions in y_1, y_2. (We assume that the present situation is different from the one covered by Theorem 1.)

(iii) We apply the operator

$$\left(\beta\frac{\partial}{\partial y_1}+\alpha\frac{\partial}{\partial y_2}\right)\frac{\partial}{\partial y_1}\frac{\partial}{\partial y_2}$$

on Eq. (11″). The left-hand side vanishes and therefore (assume that $\lambda\mu \neq 0$)

$$\begin{aligned}(\beta\lambda-\alpha\mu)U_1'''&=(\alpha\lambda+\mu\beta)U_2'''\\&=6(\beta\lambda-\alpha\mu)(\alpha\lambda+\beta\mu)\chi,\end{aligned}$$

with the solutions (up to linear and constant terms)

$$\begin{aligned}\tilde{U}_1(\tilde{y}_1)&=\chi(\alpha\lambda+\beta\mu)\tilde{y}_1^3+a\tilde{y}_1^2,\\\tilde{U}_2(\tilde{y}_2)&=\chi(\beta\lambda-\alpha\mu)\tilde{y}_2^3+b\tilde{y}_2^2.\end{aligned}$$

In terms of the U's this gives (still for one pair of variables, say $y_1^i y_2^k$)

$$\begin{aligned}U_1(y_1)&=\chi K(\alpha,\beta,\lambda,\mu)y_1^3\\&\quad+L(\alpha,\beta,\lambda,\mu,a,b)y_1^2,\\U_2(y_2)&=\chi K(\beta,-\alpha,\lambda,\mu)y_2^3\\&\quad+L(\beta,-\alpha,\lambda,\mu,b,a)y_2^2,\\U_3(\alpha y_1-\beta y_2)&=(\lambda\mu/\alpha\beta)[\chi(\alpha y_1-\beta y_2)^3\\&\quad+(a-b)(\alpha y_1-\beta y_2)^2].\end{aligned}\tag{12}$$

From this, Eq. (10) follows.

4. CONCLUSIONS

We have shown that the Schrödinger equation for two interacting particles (1) can be separated by the transformation to the center-of-mass and relative coordinates only for external harmonic oscillator potentials $V_1(x_1)$, $V_2(x_2)$. The quadratic terms of V_1 and V_2 have to be the same up to a factor (ratio of the masses). There is no restriction on the interaction $V_3(x_1-x_2)$. For central symmetric potentials(spring constants independent of space direction), this is a result of Talmi.

We have further shown that the most general Hamiltonian which allows for a separation by a linear coordinate transformation is of the type (1), (10) up to arbitrary linear transformations in x_1 and x_2, separately. The external as well as the interaction potentials contain specified cubic terms in addition to the quadratic terms; no other terms are allowed.

APPENDIX

We reformulate the lemma used in Sec. 3 and provide a proof. The following notations and facts will be used:

(1) Let W be a 6-dimensional linear vector space with the basis $e_1, e_2, e_3, f_1, f_2, f_3$, and let $O(6)$ be the rotation group acting in W. There is a subgroup $O(3)\oplus O(3)$ of $O(6)$ acting in $W_1=\langle e_1,e_2,e_3\rangle$ and $W_2=\langle f_1,f_2,f_3\rangle$, separately. We will also use the decomposition of W into the 2-dimensional subspaces $W_{ik}=\langle e_i f_k\rangle$.

(2) Two elements of $O(6)$ are called equivalent $A\approx B$ iff there exist $R_1, R_2\in O(3)\oplus O(3)$ such that $A=R_1BR_2$. A is contained in the equivalence class $\hat{A}$.

(3) Consider the matrix function $M(A)=AgA^T$, $g=1\oplus(-1)\in O(3)\oplus O(3)$. Then the equivalence of A and B implies the similarity of $M(A)$ and $M(B)$;

$$A=R_1BR_2\Rightarrow M(A)=R_1M(B)R_1^T.$$

(4) $M(A)^2=1$.

The lemma then reads as follows:

Lemma: Each equivalence class $\hat{A}$ in $O(6)$ contains a characteristic rotation $N(\hat{A})$ which reduces according to a decomposition of W into three 2-dimensional subspaces W_{ik}.

Remarks:

(1) The decomposition of W into 2-dimensional subspaces is not unique.

(2) The introduction of the matrix function $M(A)$ will reduce the problem between equivalent matrices to similar matrices and "pick the part of A out which is alien to $O(3)\oplus O(3)$."

Proof:

(a) Consider $A\in O(6)$. Then by (3) and (4), $M(A)\in O(6)$ and $M(A)^2=1$. This is enough to show explicitly that there exists $R_1\in O(3)\oplus O(3)$ such that

$$M(A)=R_1M(N)R_1^T,$$

where N is a characteristic rotation of $\hat{A}$.

(b) By definition

$$\begin{aligned}AgA^T&=R_1NgN^TR_1^T,\\A&=R_1NR_2,\quad R_2=gN^TR_1^TAg.\end{aligned}$$

Now it can be easily verified that

$$R_2gR_2^T=g.$$

But since R_2 is also a rotation it has to be in $O(3)\oplus O(3)$.

* Supported in part by the U.S. Atomic Energy Commission under Contract No. AT(30-1)-3829.

[1] I. Talmi, Helv. Phys. Acta **25**, 185 (1952).
[2] T. Terasawa, Progr. Theoret. Phys. (Kyoto) **23**, 87 (1960).
[3] J. F. Dawson, I. Talmi, and J. D. Walecka, Ann. Phys. (N.Y.) **18**, 339 (1962).
[4] B. P. Nigam, Phys. Rev. **133**, B1381 (1964).
[5] N. R. Kestner and O. Sinanoğlu, Phys. Rev. **128**, 2687 (1962).
[6] $\mathfrak{D}'$ denotes the set of distributions over the infinitely differentiable functions with compact support. See L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1950).
[7] Or, more generally, $O(2n)$. Since the dimension $n=3$ of the configuration space does not enter into the argument, we believe all results hold for arbitrary n. However, we have not constructed a proof of the lemma in the appendix for arbitrary n.
[8] The mathematics can be generalized to the case of more functions V, allowing corresponding higher-order polynomial solutions.
[9] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1950).
[10] S. J. Krieger, Phys. Rev. Letters **22**, 97 (1969).
[11] K. Ananthanaryanan, Phys. Rev. **163**, 985 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Spatially Homogeneous Anisotropic Cosmological Models Containing Relativistic Fluid and Magnetic Field*

J. PETER VAJK AND PETER G. ELTGROTH
Lawrence Radiation Laboratory, University of California, Livermore, California 94550

(Received 30 June 1969; Revised Manuscript Received 18 February 1970)

We review spatially homogeneous, axially symmetric universes containing either an ideal fluid (with a γ-law equation of state) or a uniform magnetic field parallel to the symmetry axis or both. In many cases, the field equations may be solved by the technique (described in detail) of replacing the cosmic time by a suitably chosen timelike parameter. We systematically derive all known exact solutions for such universes.

I. INTRODUCTION

In recent years, experimental studies of the isotropy of the cosmic microwave radiation[1] and speculations about the amount of helium formed in the primordial fireball,[2] about cosmic magnetic fields,[3–6] and about the effects of neutrino viscosity in the primordial fireball[7,8] have all stimulated theoretical interest in anisotropic cosmological models,[7–15] and a large number of exact solutions have been obtained.

In this paper, we review three classes of spatially homogeneous, axially symmetric space–times which contain an ideal fluid obeying a γ-law equation of state and in which there is a uniform magnetic field parallel to the axis of symmetry. We systematically derive all known exact solutions of these classes, showing explicitly how the coupled Einstein field equations may be linearized and separated, using the same notation throughout.

We consider space–times having the metric forms

$$ds^2 = d\tau^2 - f^2(\tau)\,d\chi^2 - S^2(\tau)\begin{cases} d\Omega^2 & \text{(1a)} \\ dl^2 & \text{(1b)} \\ d\Sigma^2, & \text{(1c)} \end{cases}$$

where

$$d\Omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2, \quad \text{(1'a)}$$

$$dl^2 = d\eta^2 + d\zeta^2, \quad \text{(1'b)}$$

$$d\Sigma^2 = d\theta^2 + \sinh^2\theta\,d\phi^2, \quad \text{(1'c)}$$

where θ, ϕ, η, and ζ are dimensionless ("angular") coordinates. These metric forms were first investigated in detail by Thorne.[4,5] In each of these space–times, families of timelike world lines along which all three space coordinates are fixed form geodesic congruences orthogonal to the spacelike hypersurfaces along which τ = const. Throughout this discussion, the magnetic field is assumed to be directed along the χ axis.

The symmetries of each of these space–times may be described by four spacelike Killing vector fields. (These are the generators of the respective isometry groups acting on these space–times.) In terms of the coordinates used in (1), the Killing vector fields are, respectively,

$$\xi_{a1} = \frac{\partial}{\partial\phi}, \quad \xi_{a2} = \sin\phi\frac{\partial}{\partial\theta} + \cot\theta\cos\phi\frac{\partial}{\partial\phi},$$

$$\xi_{a3} = \cos\phi\frac{\partial}{\partial\theta} - \cot\theta\sin\phi\frac{\partial}{\partial\phi}, \quad \xi_{a4} = \frac{\partial}{\partial\chi}, \quad \text{(2a)}$$

$$\xi_{b1} = \eta\frac{\partial}{\partial\zeta} - \zeta\frac{\partial}{\partial\eta}, \quad \xi_{b2} = \frac{\partial}{\partial\eta},$$

$$\xi_{b3} = \frac{\partial}{\partial\zeta}, \quad \xi_{b4} = \frac{\partial}{\partial\chi}, \quad \text{(2b)}$$

$$\xi_{c1} = -\cos\phi\frac{\partial}{\partial\theta} + (\coth\theta\sin\phi - 1)\frac{\partial}{\partial\phi},$$

$$\xi_{c2} = \sin\phi\frac{\partial}{\partial\theta} + \coth\theta\cos\phi\frac{\partial}{\partial\phi},$$

$$\xi_{c3} = \cos\phi\frac{\partial}{\partial\theta} - (\coth\theta\sin\phi + 1)\frac{\partial}{\partial\phi},$$

$$\xi_{c4} = \frac{\partial}{\partial\chi}. \quad \text{(2c)}$$

The corresponding Lie algebras are specified by the following commutation relations, respectively:

$$[\xi_{a1},\xi_{a2}] = \xi_{a3}, \quad [\xi_{a2},\xi_{a3}] = \xi_{a1}, \quad [\xi_{a3},\xi_{a1}] = \xi_{a2},$$
$$[\xi_{a4},\xi_{a1}] = 0, \quad [\xi_{a4},\xi_{a2}] = 0, \quad [\xi_{a4},\xi_{a3}] = 0, \quad \text{(3a)}$$

$$[\xi_{b1},\xi_{b2}] = \xi_{b3}, \quad [\xi_{b2},\xi_{b3}] = 0, \quad [\xi_{b3},\xi_{b1}] = -\xi_{b2},$$
$$[\xi_{b4},\xi_{b1}] = 0, \quad [\xi_{b4},\xi_{b2}] = 0, \quad [\xi_{b4},\xi_{b3}] = 0, \quad \text{(3b)}$$

$$[\xi_{c1},\xi_{c2}] = \xi_{c1}, \quad [\xi_{c2},\xi_{c3}] = \xi_{c3}, \quad [\xi_{c3},\xi_{c1}] = 2\xi_{c2},$$
$$[\xi_{c4},\xi_{c1}] = 0, \quad [\xi_{c4},\xi_{c2}] = 0, \quad [\xi_{c4},\xi_{c3}] = 0. \quad \text{(3c)}$$

In each case, the Killing vector field ξ_4 is the generator of spatial translations parallel to the axis of

symmetry (the χ direction). The Killing vector fields ξ_1, ξ_2, and ξ_3 are generators of rotations about the χ axis and/or spatial translations normal to the χ axis, and are tangent to the 2-dimensional subspaces on which $\tau = \text{const}$ and $\chi = \text{const}$. These subspaces are surfaces of constant curvature with line elements proportional to (1'a), (1'b), or (1'c), which are the line elements, respectively, of the sphere (positive curvature), the plane (zero curvature), and the pseudosphere (negative curvature).

The components of the Riemann tensor relative to the orthonormal tetrads of 1-forms

$$\omega_a^0 = d\tau, \qquad \omega_a^1 = f(\tau)\, d\chi,$$
$$\omega_a^2 = S(\tau)\, d\theta, \quad \omega_a^3 = S(\tau) \sin\theta\, d\phi, \tag{4a}$$
$$\omega_b^0 = d\tau, \qquad \omega_b^1 = f(\tau)\, d\chi,$$
$$\omega_b^2 = S(\tau)\, d\eta, \quad \omega_b^3 = S(\tau)\, d\zeta, \tag{4b}$$
$$\omega_c^0 = d\tau, \qquad \omega_c^1 = f(\tau)\, d\chi,$$
$$\omega_c^2 = S(\tau)\, d\theta, \quad \omega_c^3 = S(\tau) \sinh\theta\, d\phi \tag{4c}$$

may be computed for the corresponding metric forms (1) by Misner's prescription.[16] The Einstein tensor is found to be diagonal; consequently, if the stress–energy tensor is to represent an ideal fluid and a magnetic field parallel to the χ axis, the fluid must be comoving with the spatial coordinates (χ, θ, ϕ) or (χ, η, ζ). The Einstein field equations without cosmological constant are then

$$G_{00} = \frac{\dot S}{S}\left(2\frac{f'}{f} + \frac{\dot S}{S}\right) - \frac{\epsilon}{S^2} = \kappa(\mu + \tfrac{1}{2}H^2), \tag{5}$$

$$G_{11} = -2\frac{\ddot S}{S} - \left(\frac{\dot S}{S}\right)^2 + \frac{\epsilon}{S^2} = \kappa(p - \tfrac{1}{2}H^2), \tag{6}$$

$$G_{22} = G_{33} = -\frac{f''}{f} - \frac{f'\dot S}{f S} - \frac{\ddot S}{S} = \kappa(p + \tfrac{1}{2}H^2), \tag{7}$$

$$G_{\alpha\beta} = 0, \quad \text{for} \quad \alpha \neq \beta, \tag{8}$$

where $\kappa = 8\pi G/c^4$ is the relativistic gravitational constant, μ is the proper density of the ideal fluid, p is the proper stress (pressure) of the fluid, H is the magnetic field, and $\epsilon = -1$, 0, or $+1$, according to whether (1a), (1b), or (1c) is chosen. Derivatives with respect to τ are denoted by a dot.

The Maxwell equations for a magnetic field parallel to the χ axis may also be expressed in terms of the orthonormal tetrads[17] (4); these equations are

$$\dot H + 2H\dot S/S = 0, \quad f^{-1}H' = 0, \tag{9a}$$

where the prime denotes differentiation with respect to χ, and may be integrated straightforwardly to give

$$H = H_0/S^2, \tag{9b}$$

where H_0 is a constant of integration having the dimensions of magnetic flux. We recognize (9b) as a statement of the Gauss theorem: The magnetic flux threading any closed spacelike loop is invariant if that loop is comoving with the spatial coordinates.

The explicit forms of the conservation laws $T^{\mu\nu}{}_{;\nu} = 0$ may be obtained by the usual tensorial methods. The resulting equations are

$$\dot\mu + (\mu + p)(f'/f + 2\dot S/S) = 0, \tag{10}$$
$$f^{-1}p' = 0, \tag{11}$$

while the remaining two equations are trivially satisfied because of the symmetries of the metric forms (1). Equation (11) could also have been obtained by inspection from the field equations (5)–(7) together with the integral (9b) of the Maxwell equations; from the form of these equations, it is clear that μ and p are functions only of τ.

If the equation of state has the form

$$p = (\gamma - 1)\mu, \quad \mu + p = \gamma\mu, \quad 1 \leq \gamma \leq 2 \tag{12}$$

(the γ-law equation of state),[18,19] then the conservation law (10) may be integrated to give

$$\mu = \mu_0/(fS^2)^\gamma, \quad p = (\gamma - 1)\mu_0/(fS^2)^\gamma, \tag{13}$$

where μ_0 is a constant of integration. Inserting (9b) and (13) into the Einstein field equations (5)–(7), we obtain

$$G_{00} = \frac{\dot S}{S}\left(2\frac{f'}{f} + \frac{\dot S}{S}\right) - \frac{\epsilon}{S^2} = \frac{\kappa\mu_0}{(fS^2)^\gamma} + \frac{q^2}{S^4}, \tag{14}$$

$$G_{11} = -2\frac{\ddot S}{S} - \left(\frac{\dot S}{S}\right)^2 + \frac{\epsilon}{S^2} = \frac{\kappa(\gamma - 1)\mu_0}{(fS^2)^\gamma} - \frac{q^2}{S^4}, \tag{15}$$

$$G_{22} = -\frac{f''}{f} - \frac{f'\dot S}{f S} - \frac{\ddot S}{S} = \frac{\kappa(\gamma - 1)\mu_0}{(fS^2)^\gamma} + \frac{q^2}{S^4}, \tag{16}$$

where $q^2 \equiv \frac{1}{2}\kappa H_0^2$. This is the system of equations we study in this paper. Although there are three field equations for the two functions $f(\tau)$ and $S(\tau)$, it follows from the Bianchi identities that only two of these are independent.

II. DYNAMIC MAGNETIC UNIVERSES

We first consider universes containing only a magnetic field. Setting $\mu_0 = 0$, we see that (15) then contains only S:

$$2\ddot S + S^{-1}(\dot S)^2 = (q^2/S^3) + (\epsilon/S). \tag{17}$$

Since the independent variable τ does not appear explicitly, we may regard $\dot S$ and $\ddot S$ as functions of S

rather than τ. If we define a new dependent variable g by

$$g(S) \equiv \dot{S}(\tau) = \dot{S}(S(\tau)) = \dot{S}(S), \tag{18a}$$

then

$$\ddot{S}(S) = g\frac{dg}{dS} \tag{18b}$$

and (17) becomes a Bernoulli equation for g which has the solution

$$g = \frac{dS}{d\tau} = \pm\left(\epsilon + \frac{C}{S} - \frac{q^2}{S^2}\right)^{\frac{1}{2}}, \tag{19}$$

where C is a constant of integration.[20]

A second integration now gives S as an implicit function of τ for each value of ϵ:

$$\tfrac{1}{2}[D^2 - (2S - C)^2]^{\frac{1}{2}} + \tfrac{1}{2}C\cos^{-1}((2S - C)/D) = \tau - \tau_0, \quad \epsilon = -1, \tag{20a}$$

$$(2/3C^2)(CS + 2q^2)(CS - q^2)^{\frac{1}{2}} = \tau - \tau_0, \quad \epsilon = 0, \tag{20b}$$

$$\tfrac{1}{2}[(2S + C)^2 - D^2]^{\frac{1}{2}} - \tfrac{1}{2}C\cosh^{-1}((2S + C)/D) = \tau - \tau_0, \quad \epsilon = +1, \tag{20c}$$

where

$$D^2 \equiv C^2 + 4q^2\epsilon \tag{21}$$

and τ_0 is a (physically trivial) constant of integration.

In principle, we could now insert these results in either (14) or (16) and integrate the resulting linear differential equation for f; in practice, it is more convenient to express the solutions (20) in parametric form and then to rewrite (14) or (16) in terms of the parameter. The solutions (20a) and (20c) suggest the "natural" parameters

$$\psi \equiv \cos^{-1}((2S - C)/D), \quad \epsilon = -1, \tag{22a}$$

$$\psi \equiv \cosh^{-1}((2S + C)/D), \quad \epsilon = +1, \tag{22c}$$

respectively, so that

$$S(\psi) = \tfrac{1}{2}C + \tfrac{1}{2}D\cos\psi, \quad \tau - \tau_0 = \tfrac{1}{2}C\psi + \tfrac{1}{2}D\sin\psi, \quad \epsilon = -1, \tag{23a}$$

$$S(\psi) = \tfrac{1}{2}D\cosh\psi - \tfrac{1}{2}C, \quad \tau - \tau_0 = \tfrac{1}{2}D\sinh\psi - \tfrac{1}{2}C\psi, \quad \epsilon = +1. \tag{23c}$$

[Note that (20a) and (23a) describe a cycloid in the (S, τ) plane.]

The solution (20b), however, does not suggest any "natural" parameter. But we note from (23) that the parameters (22) satisfy, in each case, the relation

$$d\psi = d\tau/S. \tag{24}$$

(This relation between the parameter ψ and the cosmic time τ is analogous to that between the timelike conformal coordinate sometimes used in discussions of Robertson–Walker cosmological models[21] and the cosmic time in those models.) If ψ is used as the timelike coordinate instead of τ, the metric forms (1) become

$$ds^2 = S^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)\begin{cases} d\Omega^2 & \text{(25a)} \\ dl^2 & \text{(25b)} \\ d\Sigma^2. & \text{(25c)} \end{cases}$$

To transform independent variables in any of our differential equations from τ to ψ, we must replace the operators $d/d\tau$ and $d^2/d\tau^2$ by

$$\frac{d}{d\tau} \to \frac{1}{S}\frac{d}{d\psi}, \quad \frac{d^2}{d\tau^2} \to \frac{1}{S^2}\left(\frac{d^2}{d\psi^2} - \frac{S'}{S}\frac{d}{d\psi}\right), \tag{26}$$

where the prime denotes differentiation with respect to ψ. Upon transforming (19) and integrating it anew, we obtain

$$S(\psi) = \tfrac{1}{4}C\psi^2 + q^2/C, \quad \tau - \tau_0 = \tfrac{1}{12}C\psi^3 + (q^2/C)\psi, \quad \epsilon = 0, \tag{23b}$$

as the parametrized form of (20b). In this case,

$$\psi \equiv (2/C)(CS - q^2)^{\frac{1}{2}}. \tag{22b}$$

The field equation (14) now becomes

$$f' + \tfrac{1}{2}\{[(S')^2 - \epsilon S^2 - q^2]/SS'\}f = 0, \tag{27}$$

and we readily obtain

$$f(\psi) = E\sin\psi/(C + D\cos\psi), \quad \epsilon = -1, \tag{28a}$$

$$f(\psi) = E\psi/(C^2\psi^2 + 4q^2), \quad \epsilon = 0, \tag{28b}$$

$$f(\psi) = E\sinh\psi/(D\cosh\psi - C), \quad \epsilon = +1, \tag{28c}$$

where E is a constant of integration.

The metric (23a), (25a), and (28a) describes the intermediate region $r_1 < r < r_2$ of the Reissner–Nordstrøm solution[22]

$$ds^2 = \Phi(r)\,dt^2 - [\Phi(r)]^{-1}dr^2 - r^2\,d\Omega^2, \tag{29}$$

where

$$\Phi(r) = 1 - \frac{2m}{r} + \frac{q^2}{r^2}, \tag{29'}$$

as extended by Graves and Brill,[23] where

$$r_1 = m - (m^2 - q^2)^{\frac{1}{2}} = \tfrac{1}{2}(C - D), \quad r_2 = m + (m^2 - q^2)^{\frac{1}{2}} = \tfrac{1}{2}(C + D). \tag{30}$$

The apparent singularities in (28a) for $\psi = n\pi$ correspond to the fact that the hypersurfaces $r =$ const become null as $r \to r_1$ or $r \to r_2$.

The metric (23b), (25b), (28b) is the axially symmetric purely magnetic special case (in a different notation) of the general Bianchi type I solution found by Rosen for a universe containing only

electromagnetic fields[24] parallel to one of the coordinate axes. Jacobs[13,14] also gives Rosen's solution in a different form.

Bianchi type I space-times have the metric

$$ds^2 = dt^2 - A^2(t)\,dx^2 - B^2(t)\,dy^2 - W^2(t)\,dz^2,$$

and have been examined most systematically by Jacobs.[13,14] For the special case $A(t) \equiv B(t)$, this metric reduces to the form (1b), i.e., that of a spatially homogeneous axially symmetric quasi-Euclidean space-time.

This solution was also obtained by Brill[25] (in a different notation), who has also discussed its analytic extension beyond the apparent (coordinate) singularity at $\psi = 0$, $S = S_{\text{crit}} \equiv q^2/C$. If S is permitted to decrease below S_{crit}, then we see from (22b) and (28b) that both ψ and $f(\psi)$ become pure imaginary. Inspection of (25b) shows that the coordinates ψ and χ then interchange roles, but the metric remains regular. The space-time region where $S < S_{\text{crit}}$ is an electromagnetic generalization of Taub space; the region $S > S_{\text{crit}}$ is a generalization of outer NUT space; and the null hypersurface $S = S_{\text{crit}}$ is the analog of the Misner boundary.[25,16]

III. STATIC MAGNETIC UNIVERSES

For $\epsilon \neq 0$, Eq. (17) has another solution for which

$$S^2 = -\epsilon q^2, \quad \dot{S} = \ddot{S} = 0, \quad \epsilon \neq 0, \tag{31}$$

so that (16) becomes

$$\ddot{f}/f = -1/q^2. \tag{32}$$

We thus obtain the metric forms

$$ds^2 = d\tau^2 - E^2\cos^2(\tau/q)\,d\chi^2 - q^2\,d\Omega^2, \quad \epsilon = -1, \tag{33a}$$

$$ds^2 = E^2\cos^2(\chi/q)\,d\tau^2 - d\chi^2 - q^2\,d\Sigma^2, \quad \epsilon = +1, \tag{33c}$$

where E is a constant of integration. [The coordinates have been renamed to obtain (33c), and an over-all minus sign, which has no physical significance, has been dropped for the sake of convention.] These are the Bertotti solutions,[26] which, despite the apparent singularities in (33), describe static nonsingular electromagnetic universes. The 2-dimensional subspaces $\theta = \text{const}$ and $\phi = \text{const}$ are surfaces of constant curvature, negative for (33a), positive for (33c).

IV. UNIVERSES CONTAINING DUST AND A MAGNETIC FIELD

Can the system of equations (14)–(16) still be solved in closed form if $\mu_0 \neq 0$? We note that, if $\gamma = 1$ (i.e., if the matter is pressureless dust), then (15) has the same form (17) as in the pure magnetic case. The solutions (20) or (23) again result, and the same timelike parameters (22) may be defined. The field equation (14) in terms of those parameters becomes

$$f' + \frac{1}{2}\left(\frac{(S')^2 - \epsilon S^2 - q^2}{SS'}\right) f = \frac{\kappa\mu_0}{2}\frac{S}{S'}, \tag{34}$$

which has the solutions

$$f(\psi) = \frac{E\sin\psi}{C + D\cos\psi} + \frac{\kappa\mu_0}{2D}\frac{D^2\psi\sin\psi + (C^2 + D^2)\cos\psi + 2CD}{C + D\cos\psi}, \quad \epsilon = -1, \tag{35a}$$

$$f(\psi) = \frac{E\psi}{C^2\psi^2 + 4q^2} + \frac{\kappa\mu_0}{6}\frac{C^4\psi^4 + 24q^2C^2\psi^2 - 48q^4}{C^2\psi^2 + 4q^2}, \quad \epsilon = 0, \tag{35b}$$

$$f(\psi) = \frac{E\sinh\psi}{D\cosh\psi - C} + \frac{\kappa\mu_0}{2D}\frac{D^2\psi\sinh\psi - (C^2 + D^2)\cosh\psi + 2CD}{D\cosh\psi - C}, \quad \epsilon = +1, \tag{35c}$$

where E is a constant of integration. These expressions together with (23) and (25) define the solutions first obtained by Doroshkevich[11] in a different form. Shikin[27] has also found these solutions, in a notation very similar to ours, as has Thorne[6] in yet another form. If μ_0 vanishes so that these universes contain only a magnetic field, these solutions reduce to those found in Sec. II above; the cases in which the magnetic fields vanish but μ_0 is finite are discussed in Secs. VI and VII below.

The dynamic equation (17) for S again has constant solutions for $\epsilon \neq 0$, but these are incompatible (for finite f) with (14), so that there exist no static universes containing dust and a magnetic field analogous to the Bertotti solutions (33).

V. UNIVERSES CONTAINING AN IDEAL FLUID AND A MAGNETIC FIELD

We now seek solutions of the system (14)–(16) for the case $\gamma \neq 1$, i.e., universes containing ideal fluids with pressure. The equations for f and S are now coupled by source terms of the form $(fS^2)^{-\gamma}$ so that we cannot, as in previous cases, first solve for $S(\tau)$ and then integrate the remaining equations to obtain

$f(\tau)$. But the technique of replacing the cosmic time τ by a suitably chosen timelike parameter has been so useful in solving the simpler cases that we are motivated to seek a modification of that technique.

A. General Solution

We first rearrange (15) to obtain

$$f^\gamma = -\kappa(\gamma - 1)\mu_0/S^{2\gamma}G, \quad \gamma \neq 1, \tag{36}$$

where

$$G \equiv 2\frac{\ddot{S}}{S} + \left(\frac{\dot{S}}{S}\right)^2 - \frac{\epsilon}{S^2} - \frac{q^2}{S^4} \tag{37}$$

so that

$$\frac{\dot{f}}{f} = -2\frac{\dot{S}}{S} - \frac{1}{\gamma}\frac{\dot{G}}{G}. \tag{38}$$

Substitution of (36) and (38) into (14) yields

$$\frac{2}{\gamma}\frac{\dot{S}}{S}\frac{\dot{G}}{G} = \frac{1}{\gamma - 1}\left[2\frac{\ddot{S}}{S} + (4 - 3\gamma)\left(\frac{\dot{S}}{S}\right)^2 - \gamma\left(\frac{\epsilon}{S^2} + \frac{q^2}{S^4}\right)\right]. \tag{39}$$

This equation and the relation (36) are completely equivalent to the field equations (14) and (15).

Next we replace the independent variable τ by a new parameter σ, but we leave unspecified until later the form of the relation between τ and σ. The operators $d/d\tau$ and $d^2/d\tau^2$ are replaced by

$$\frac{d}{d\tau} \to \frac{1}{\tau'}\frac{d}{d\sigma}, \quad \frac{d^2}{d\tau^2} \to \frac{1}{(\tau')^2}\left(\frac{d^2}{d\sigma^2} - \frac{\tau''}{\tau'}\frac{d}{d\sigma}\right), \tag{40}$$

where the prime denotes differentiation with respect to σ. The field equation (39) then becomes

$$\frac{G'}{G} = \frac{\gamma}{2(\gamma - 1)}\left[2\frac{S''}{S'} - 2\frac{\tau''}{\tau'} + (4 - 3\gamma)\frac{S'}{S} - \gamma\frac{(\tau')^2}{SS'}\left(\epsilon + \frac{q^2}{S^2}\right)\right], \tag{41}$$

which could be integrated if the last term were sufficiently simple. But since σ has not yet been specified, we may now choose σ so that τ' and S' satisfy

$$\frac{(\tau')^2}{SS'}\left(\epsilon + \frac{q^2}{S^2}\right) = S'\frac{dy(S)/dS}{y(S)}, \tag{42}$$

where y is some as yet unknown function of S. [Note that, if $\tau(\sigma)$ and $S(\sigma)$ were known, $y(S)$ would be determined uniquely except for a multiplicative constant.] With this choice of σ, we immediately obtain from (41) the result

$$G = g_0\left[\left(\frac{S'}{\tau'}\right)^2\frac{S^{4-3\gamma}}{y^\gamma}\right]^{\gamma/2(\gamma-1)}, \tag{43}$$

where g_0 is a constant of integration.

Upon making the replacements (40) in the original definition (37) of G, equating that result with the above integral (43), and using (42) to eliminate all reference to τ, we obtain a differential equation for $y(S)$:

$$\frac{d}{dS}\left(\frac{dy/dS}{\epsilon + q^2/S^2}\right) = -g_0\frac{S^{(4-3\gamma)/2}}{y^{(\gamma+2)/2}}\left[\frac{dy/dS}{\epsilon + q^2/S^2}\right]^{(3\gamma-4)/2(\gamma-1)} \tag{44}$$

Once this equation has been integrated to give $y = y(S)$, we may obtain $\tau(S)$ from (42) which may be written as

$$\left(\frac{d\tau}{dS}\right)^2 = \frac{S\,dy(S)/dS}{y(S)(\epsilon + q^2/S^2)}. \tag{45}$$

Next, we may invert $y = y(S)$ to find $S = S(y)$ and thus $\tau = \tau(y)$. Finally, we may calculate $f = f(y)$ from (36), thus obtaining the solution of the cosmological Eqs. (14), (15), and (16) in parametric form with y as parameter.

Jacobs (Ref. 28, Appendix D) obtains a closed-form solution for Bianchi type I cosmological models containing an ideal fluid with $\gamma = \frac{4}{3}$ and a magnetic field, using a different analysis.

Note that this solution is valid *only* if $\gamma \neq 1$ [see (15) and (36)]: the Doroshkevich solutions (23), (25), and (35) for $\gamma = 1$ cannot be obtained by the technique described here. Nor is this solution valid if $\epsilon = 0$ *and* $q = 0$ [see (42), (44), and (45)]: this case is discussed separately below. We now turn to several special cases of (44) which can be integrated analytically.

B. Quasi-Euclidean Case with Magnetic Field

The simplest case to consider is that of $\epsilon = 0$. The space–time metric then has the form (1b), and the intrinsic geometry of the spacelike hypersurfaces $\tau = \text{const}$ is Euclidean. The differential equation (44) for $y(S)$ becomes

$$\frac{d}{dS}\left(S^2\frac{dy}{dS}\right) = -g_0 q^{(2-\gamma)/(\gamma-1)}S^{(3\gamma-4)(3-\gamma)/2(\gamma-1)} \times y^{-(\gamma+2)/2}\left(\frac{dy}{dS}\right)^{(3\gamma-4)/2(\gamma-1)} \tag{46}$$

Provided that q does not vanish, a *particular* solution is

$$y = y_0 S^p, \quad p = (3\gamma - 4)(2 - \gamma)/\gamma^2, \tag{47}$$

with $\gamma \neq \frac{4}{3}$ or 2. We then obtain from (45) the result

$$\frac{d\tau}{dS} = \frac{p^{\frac{1}{2}}}{q}S, \tag{48}$$

so that we must have $p > 0$ or

$$\tfrac{4}{3} < \gamma < 2. \tag{47'}$$

Integrating (48) and solving for S, we find that

$$S(\tau) = [2\gamma q(\tau - \tau_0)]^{\frac{1}{2}}/[(3\gamma - 4)(2 - \gamma)]^{\frac{1}{2}}. \tag{49}$$

Finally, (36) gives

$$f(\tau) = \left(\frac{2\kappa\mu_0\gamma^2}{4-\gamma}\right)^{1/\gamma} \frac{[(3\gamma - 4)(2 - \gamma)]^{\frac{1}{2}}}{2\gamma q}(\tau - \tau_0)^{(2-\gamma)/\gamma}. \tag{50}$$

[The solution (49)–(50) is not the most general solution for $\frac{4}{3} < \gamma < 2$, as we have been unable to obtain the general solution of (46).] This exact solution was first given by Jacobs.[14]

Note that the fluid pressure diminishes the effectiveness of the Faraday pressure in resisting gravitational collapse transverse to the magnetic field [see (6)] so that $S(\tau)$ can vanish for this case in contrast to the Doroshkevich solution (23b), (25b), (35b) for pressureless matter and a magnetic field.

VI. QUASI-EUCLIDEAN SPACE-TIMES CONTAINING AN IDEAL FLUID

A. $\gamma = 1$

For $\epsilon = 0$, the spacelike hypersurfaces $\tau = \text{const}$ are Euclidean spaces. The exact solution for an axially symmetric quasi-Euclidean space-time containing only dust ($\gamma = 1$) can be obtained at once from the Doroshkevich solution (20b), (23b), (35b) by setting $q = 0$. This gives

$$S(\tau) = (\tfrac{3}{2})^{\frac{2}{3}}C^{\frac{1}{3}}(\tau - \tau_0)^{\frac{2}{3}}, \tag{51}$$

$$f(\tau) = (\tfrac{1}{12})^{\frac{1}{3}}EC^{-\frac{5}{3}}(\tau - \tau_0)^{-\frac{1}{3}} + (\tfrac{2}{3})^{\frac{1}{3}}\kappa\mu_0 C^{\frac{4}{3}}(\tau - \tau_0)^{\frac{2}{3}}, \tag{52}$$

or

$$S(\psi) = \tfrac{1}{4}C\psi^2, \quad \tau - \tau_0 = \tfrac{1}{12}C\psi^3, \tag{53b}$$

$$f(\psi) = E/C^2\psi + \tfrac{1}{6}\kappa\mu_0 C^2\psi^2. \tag{54b}$$

This is the axially symmetric special case of the general Bianchi type I solution found by Robinson[29] and by Heckmann and Schücking[30] for a universe containing only dust. That general solution has also been given in a different form by Jacobs.[13] If the constant of integration E vanishes, the solution is isotropic for all times, and is just the standard, flat Friedmann universe.[21] Kompaneets and Chernov[31] first obtained the axially symmetric solution (51)–(52) in a different notation.

B. $\gamma \neq 1$

As we noted earlier, the analysis leading to the differential equations (44) and (45) is not valid for $\epsilon = 0$ *and* $q = 0$, i.e., for a quasi-Euclidean space-time containing an ideal fluid but no magnetic field. In this case, (37) reduces to

$$G \equiv 2\frac{\ddot{S}}{S} + \left(\frac{\dot{S}}{S}\right)^2 = \frac{1}{S^2\dot{S}}\frac{d}{d\tau}(S\dot{S}^2), \tag{55}$$

and (41) may be integrated at once, *without* forcing a restriction such as (42) on the choice of the parameter σ. We thus obtain

$$G(\sigma) = g_0(S'/\tau')^{\gamma/(\gamma-1)}S^{\gamma(4-3\gamma)/2(\gamma-1)}, \tag{56}$$

where g_0 is a constant of integration. Upon transforming variables from τ to σ in (55) and equating that result with (56), we obtain

$$\frac{d}{d\sigma}\left[S\left(\frac{S'}{\tau'}\right)^2\right] = g_0(S')^{(2\gamma-1)/(\gamma-1)}(\tau')^{-\gamma/(\gamma-1)}S^{(3\gamma-2)(2-\gamma)/2(\gamma-1)}. \tag{57}$$

Suppose σ were chosen such that

$$(S')^{(2\gamma-1)/(\gamma-1)}(\tau')^{-\gamma/(\gamma-1)}S^{(3\gamma-2)(2-\gamma)/2(\gamma-1)} = \frac{dF(\sigma)}{d\sigma}, \tag{58}$$

where $F(\sigma)$ is some function of σ. Then (57) could be integrated to give

$$S(S'/\tau')^2 = g_0F(\sigma). \tag{59}$$

Using this relation to eliminate τ' from (58), we obtain

$$S^{(4-3\gamma)/2}S' = (g_0)^{-\gamma/2(\gamma-1)}F^{-\gamma/2(\gamma-1)}F', \tag{60}$$

which may be integrated at once to give

$$[F(\sigma)]^{(\gamma-2)/2(\gamma-1)} = -(g_0)^{\gamma/2(\gamma-1)}[A_1 + S^{3(2-\gamma)/2}/3(\gamma - 1)], \quad 1 < \gamma < 2, \tag{61a}$$

$$F(\sigma) = A_2S^{g_0}, \quad \gamma = 2, \tag{61b}$$

where A_1 and A_2 are arbitrary constants of integration.

Now (59) gives $d\tau/dS$ in terms of S and $F(\sigma)$; using (61), we obtain

$$\frac{d\tau}{dS} = (-g_0)^{(\gamma-1)/(2-\gamma)} \times S^{\frac{1}{2}}[A_1 + S^{3(2-\gamma)/2}/3(\gamma - 1)]^{(\gamma-1)/(2-\gamma)}, \quad 1 < \gamma < 2, \tag{62a}$$

$$\frac{d\tau}{dS} = (g_0A_2)^{-\frac{1}{2}}S^{(1-g_0)/2}, \quad \gamma = 2. \tag{62b}$$

[Upon integrating (62), we would obtain $S(\tau)$ in

implicit form.] Finally, we find from (36)

$$f(\tau) = [\kappa(\gamma-1)\mu_0]^{1/\gamma}(-g_0)^{(\gamma-1)/(2-\gamma)} \times S^{-\frac{1}{2}}[A_1 + S^{3(2-\gamma)/2}/3(\gamma-1)]^{1/(2-\gamma)}, \quad 1<\gamma<2, \quad (63a)$$

$$f(\tau) = (-\kappa\mu_0/g_0^2A_2)^{\frac{1}{2}}S^{-(g_0+1)/2}, \quad \gamma = 2. \quad (63b)$$

Note that (62) and (63) require both g_0 and A_2 to be negative; there is no restriction on A_1. We now examine the cases $1<\gamma<2$ and $\gamma=2$ separately.

Case 1. $1<\gamma<2$

For arbitrary nonzero values of A_1, (62a) cannot be integrated in closed form except for certain rational values of γ. However, if we define a new timelike parameter ψ by the relation

$$\tfrac{1}{2}|A_1|\cosh\psi \equiv S^{3(2-\gamma)/2}/3(\gamma-1) + \tfrac{1}{2}A_1, \quad A_1 \neq 0, \quad (64)$$

so that

$$S(\psi) = [\tfrac{3}{2}(\gamma-1)A_1(\cosh\psi - 1)]^{2/3(2-\gamma)}, \quad A_1>0, \quad (65a)$$

$$S(\psi) = [\tfrac{3}{2}(\gamma-1)|A_1|(\cosh\psi + 1)]^{2/3(2-\gamma)}, \quad A_1<0, \quad (65b)$$

then the metric (1b) may be written in the form

$$ds^2 = L^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)\,dl^2, \quad \epsilon = 0, \quad (66b)$$

where $L(\psi)\,d\psi = d\tau$. Using (64) in (62a) and (63a), we find, respectively,

$$L(\psi) = (-g_0)^{(\gamma-1)/(2-\gamma)}[3(\gamma-1)]^{1/(2-\gamma)}|\tfrac{1}{2}A_1|^{\gamma/(2-\gamma)} \times [2/3(2-\gamma)](\sinh\psi)^{\gamma/(2-\gamma)} \quad (67)$$

and

$$f(\psi) = (\kappa\mu_0)^{1/\gamma}(-g_0)^{(\gamma-1)/(2-\gamma)}(\gamma-1)^{2(3-2\gamma)/3\gamma(2-\gamma)} \times (\tfrac{1}{2}A_1)^{2/3(2-\gamma)}(\cosh\psi+1)^{1/(2-\gamma)} \times [3(\cosh\psi-1)]^{-1/3(2-\gamma)}, \quad A_1>0, \quad (68a)$$

$$f(\psi) = (\kappa\mu_0)^{1/\gamma}(-g_0)^{(\gamma-1)/(2-\gamma)}(\gamma-1)^{2(3-2\gamma)/3\gamma(2-\gamma)} \times |\tfrac{1}{2}A_1|^{2/3(2-\gamma)}(\cosh\psi-1)^{1/(2-\gamma)} \times [3(\cosh\psi+1)]^{-1/3(2-\gamma)}, \quad A_1<0. \quad (68b)$$

We are thus able to obtain a closed-form expression for the metric.

For $A_1>0$, we see from (65a) and (68a) that the universe expands without bound from an initial "cigar" singularity in which $f\to\infty$, $S\to 0$, and $fS^2\to 0$. In late stages of the evolution ($\psi\to\infty$), the expansion becomes isotropic ($\dot{f}/f \to \dot{S}/S$).

For $A_1<0$, we see from (65b) and (68b) that the universe expands without bound from an initial "pancake" singularity in which $f\to 0$, $S\to S_0 = [3(\gamma-1)|A_1|]^{2/3(2-\gamma)}$, and $fS^2\to 0$. In late stages of the evolution, the expansion becomes isotropic.

For certain rational values of γ, Eq. (62a) may be integrated in closed form. If we use the parameter ψ defined by (64), (62a) may be written in the form $d\tau = L(\psi)\,d\psi$, where $L(\psi)$ is given by (67). For integral values of $\gamma/(2-\gamma)$, we obtain

$$\tau - \tau_0 = (-g_0)^{n/2}n^{2/(n+2)}[\tfrac{1}{3}(n+2)]^{n/(n+2)}|\tfrac{1}{2}A_1|^{n+1} \times \int(\sinh\psi)^{n+1}\,d\theta, \quad (69)$$

(which may be expressed in closed form for $n = 1, 2, 3, \cdots$) and

$$\gamma = 2(n+1)/(n+2) = \tfrac{4}{3}, \tfrac{3}{2}, \tfrac{8}{5}, \tfrac{5}{3}, \tfrac{12}{7}, \tfrac{7}{4}, \cdots. \quad (70)$$

Upon eliminating ψ from (65), (68), and (69), we would obtain $S(\tau)$ and $f(\tau)$ in closed form.

For $A_1 = 0$, we may integrate (62a) in closed form to obtain

$$S(\tau) = [3(\gamma-1)/(-g_0)]^{2(\gamma-1)/3\gamma(2-\gamma)}[\tfrac{3}{2}\gamma(\tau-\tau_0)]^{2/3\gamma}, \quad (71)$$

$$f(\tau) = [\kappa(\gamma-1)\mu_0]^{1/\gamma}(-g_0)^{(3\gamma-2)(\gamma-1)/3\gamma(2-\gamma)} \times [3(\gamma-1)]^{-(\gamma+2)/3\gamma(\gamma-2)}[\tfrac{3}{2}\gamma(\tau-\tau_0)]^{2/3\gamma}. \quad (72)$$

The universe expands without bound from an initial "point" singularity, and the expansion is isotropic for all times. This solution is just the spatially flat Robertson–Walker solution[21] in a different form.

The axially symmetric solution (65)–(69), inclusive, for the case $\gamma = \frac{4}{3}$ was first obtained by Doroshkevich[11] with a different parametrization. The asymptotic form of that solution had been given earlier by Kompaneets and Chernov.[31]

Case 2. $\gamma = 2$

Integration of (62b) yields

$$S(\tau) = (g_0A_2)^{1/(3-g_0)}[\tfrac{1}{2}(3-g_0)(\tau-\tau_0)]^{2/(3-g_0)}, \quad (73)$$

so that

$$f(\tau) = [\kappa\mu_0/(-g_0)]^{\frac{1}{2}}(g_0A_2)^{-2/(3-g_0)} \times [\tfrac{1}{2}(3-g_0)(\tau-\tau_0)]^{-(1+g_0)/(3-g_0)}. \quad (74)$$

We have already noted that g_0 and A_2 must both be negative. The qualitative features of the expansion depend upon g_0, but in all cases the expansion continues without bound. For $-1<g_0<0$, the universe expands from an initial "cigar" singularity; for $g_0=-1$, the universe expands purely transversely from an initial "barrel" singularity with $f = (\kappa\mu_0/|A_2|)^{\frac{1}{2}} = \text{const}$. For $g_0<-1$, the initial singularity is "point"-like, with $f\to 0$ and $S\to 0$, and $\dot{f}/f < \dot{S}/S$ or $\dot{f}/f > \dot{S}/S$ for all times, accordingly as $-3<g_0<-1$ or

$g_0 < -3$, respectively. The case $g_0 = -3$ is just the spatially flat Robertson–Walker solution[21] for $\gamma = 2$ in a different form, since the expansion is isotropic for all times.

Exact solutions for general Bianchi type I universes containing only an ideal fluid with the rational values (70) of γ or with $\gamma = 2$ were first found by Jacobs[13,28] by a similar analysis. Jacobs also discusses in detail the nature of the singularity in each solution. The solutions we have given above for the rational values (70) of γ and for $\gamma = 2$ are the axially symmetric special cases of Jacobs' solutions. The solutions (65)–(69) and (71)–(72), however, hold for *all* γ in the range $1 < \gamma < 2$.

VII. NON-EUCLIDEAN SPACE-TIMES CONTAINING AN IDEAL FLUID

A. $\gamma = 1$

The exact solutions for axially symmetric non-Euclidean space–times ($\epsilon = -1$ and $\epsilon = +1$) containing only dust ($\gamma = 1$) can be obtained at once from the Doroshkevich solutions [(20a), (23a), (35a) and (20c), (23c), (35c), respectively] by setting $q = 0$, so that $D = C$. We then obtain for $\epsilon = -1$ the results

$$S(\psi) = \tfrac{1}{2}C(1 + \cos\psi), \quad \tau - \tau_0 = \tfrac{1}{2}C(\psi + \sin\psi), \tag{53a}$$

$$f(\psi) = \tfrac{1}{2}\kappa\mu_0[\psi \sin\psi + 2(1 + \cos\psi) + E^* \sin\psi]/(1 + \cos\psi), \tag{54a}$$

where E^* is a constant of integration, while for $\epsilon = +1$ we obtain

$$S(\psi) = \tfrac{1}{2}C(\cosh\psi - 1), \quad \tau - \tau_0 = \tfrac{1}{2}C(\sinh\psi - \psi), \tag{53c}$$

$$f(\psi) = \tfrac{1}{2}\kappa\mu_0[\psi \sinh\psi - 2(\cosh\psi - 1) + E^* \sinh\psi]/(\cosh\psi - 1). \tag{54c}$$

Note that (53a) is a cycloid in the (S, τ) plane.

The solution for $\epsilon = +1$ was first obtained by Kompaneets and Chernov,[31] with S and f given as implicit functions of τ. Kantowski and Sachs[9] independently obtained exact solutions for both $\epsilon = -1$ and $\epsilon = +1$, using a different parametrization, while Thorne[4] independently obtained the exact solution for $\epsilon = -1$, using the same parametrization as Kantowski and Sachs.

B. $\gamma \neq 1$

We return now to the differential equations (44) and (45) of Sec. V for the cases $\epsilon = \pm 1$ with $q = 0$, i.e., non-Euclidean space–times containing only an ideal fluid. Solutions may be obtained in closed form for $\gamma = \frac{4}{3}$ and for $\gamma = 2$.

1. $\gamma = \frac{4}{3}$

This value of γ corresponds to a collision-dominated gas of photons or ultrarelativistic free particles. Equation (44) then becomes

$$\frac{d^2y}{dS^2} \equiv \frac{1}{2}\frac{d}{dy}\left[\left(\frac{dy}{dS}\right)^2\right] = -\epsilon g_0 y^{-\frac{5}{3}}, \tag{75}$$

which may be integrated at once to give

$$\left(\frac{dy}{dS}\right)^2 = \frac{3\epsilon g_0}{y^{\frac{2}{3}}} + \alpha, \tag{76}$$

where α is a constant of integration. To simplify further analysis, we replace y by a new variable ψ defined by

$$y^{\frac{2}{3}} = \sinh\psi. \tag{77}$$

In terms of the new variable ψ, (76) now becomes

$$\frac{dS}{d\psi} = \frac{3}{2}\frac{\sinh\psi\cosh\psi}{(\alpha\sinh\psi + 3\epsilon g_0)^{\frac{1}{2}}}. \tag{78}$$

Depending on the value of α, we obtain two different results upon integrating (78):

$$S(\psi) = S_0 + \tfrac{1}{4}(3/\epsilon g_0)^{\frac{1}{2}}\sinh^2\psi, \quad \alpha = 0, \tag{79a}$$

$$S(\psi) = S_0 + \alpha^{-2}(\alpha\sinh\psi - 6\epsilon g_0)(\alpha\sinh\psi + 3\epsilon g_0)^{\frac{1}{2}}, \quad \alpha \neq 0. \tag{79b}$$

With the aid of (76), the differential equation (45) for $d\tau/dS$ may be written in the form

$$\frac{d\tau}{d\psi} \equiv L(\psi) = \left(\tfrac{3}{2}\epsilon\coth\psi S(\psi)\frac{dS}{d\psi}\right)^{\frac{1}{2}}. \tag{80}$$

If we use ψ instead of τ as the timelike coordinate, then the metric forms (1a) and (1c) may be expressed as

$$ds^2 = L^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)\,d\Omega^2, \quad \epsilon = -1, \tag{66a}$$

$$ds^2 = L^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)\,d\Sigma^2, \quad \epsilon = +1. \tag{66c}$$

Thus, the metric can be expressed in closed form, even if (80) cannot be integrated in closed form.

Inserting the solutions (79) in (80), we find

$$L(\psi) = \{\tfrac{3}{4}\epsilon(3/\epsilon g_0)^{\frac{1}{2}}\cosh^2\psi[S_0 + \tfrac{1}{4}(3/\epsilon g_0)^{\frac{1}{2}}\sinh^2\psi]\}^{\frac{1}{2}}, \quad \alpha = 0, \tag{81a}$$

$$L(\psi) = \{\tfrac{9}{4}\epsilon\cosh^2\psi[S_0(\alpha\sinh\psi + 3\epsilon g_0)^{-\frac{1}{2}} + \alpha^{-\frac{1}{2}}(\alpha\sinh\psi - 6\epsilon g_0)]\}^{\frac{1}{2}}, \quad \alpha \neq 0. \tag{81b}$$

Finally, we obtain from (36) and (43)

$$f(\psi) = (-\kappa\mu_0\epsilon)^{\frac{3}{4}}[S_0 + \tfrac{1}{4}(3/\epsilon g_0)^{\frac{1}{2}} \sinh^2 \psi]^{-\frac{1}{2}}, \quad \alpha = 0, \tag{82a}$$

$$f(\psi) = [(-\kappa\mu_0/3g_0)(\alpha \sinh \psi + 3\epsilon g_0)]^{\frac{3}{4}} \times [S_0 + \alpha^{-2}(\alpha \sinh \psi - 6\epsilon g_0) \times (\alpha \sinh \psi + 3\epsilon g_0)^{\frac{1}{2}}]^{-\frac{1}{2}}, \quad \alpha \neq 0. \tag{82b}$$

Recalling that the proper density of the fluid is $\mu = \mu_0(fS^2)^{-\gamma}$, we find that the solution (79a), (81a), (82a) is physically unacceptable, since it leads to negative densities if the metric coefficients L^2, f^2, and S^2 appearing in (66) are positive. With suitable restrictions on the constants of integration α, g_0, and S_0, acceptable solutions of the form (79b), (81b), (82b) exist for both $\epsilon = -1$ and $\epsilon = +1$. For $\epsilon = -1$, ψ is confined to a finite range, so that the universe begins and ends in a singularity.

These solutions for $\gamma = \frac{4}{3}$ with $\epsilon = \pm 1$ were first obtained by Kompaneets and Chernov[31] with f and S given as implicit functions of τ. Kantowski[10] later found these solutions using a parametrization different from ours.

2. $\gamma = 2$

Equation (44) now becomes

$$\frac{d^2y}{dS^2} = \frac{-g_0}{y^2 S}\frac{dy}{dS}, \tag{83}$$

which may be transformed to

$$\frac{d^2y}{dz^2} + \left(\frac{g_0}{y^2} - 1\right)\frac{dy}{dz} = 0, \tag{84}$$

where

$$z = \log S. \tag{85}$$

Upon applying the technique used in Sec. II to solve (17), we find the general solution of (83) is

$$S = S_0 \exp\left(\int \frac{y\,dy}{y^2 + \alpha y + g_0}\right), \tag{86}$$

where α and S_0 are constants of integration. Substituting (86) in (45), we find

$$\frac{d\tau}{dy} \equiv L(y) = \frac{\epsilon^{\frac{1}{2}} S(y)}{(y^2 + \alpha y + g_0)^{\frac{1}{2}}}, \tag{87}$$

while (36) and (43) give

$$f(y) = \left(\frac{-\kappa\mu_0}{\epsilon g_0}\right)^{\frac{1}{2}} \frac{(y^2 + \alpha y + g_0)^{\frac{1}{2}}}{S(y)}. \tag{88}$$

For $\epsilon = -1$, acceptable solutions result for $g_0 < 0$ and $y^2 + \alpha y + g_0 < 0$, in which case we obtain, from (86),

$$S = S_0 |y^2 + \alpha y + g_0|^{\frac{1}{2}} \times \exp\left[\frac{\alpha}{(\alpha^2 - 4g_0)^{\frac{1}{2}}} \tanh^{-1}\left(\frac{2y + \alpha}{(\alpha^2 - 4g_0)^{\frac{1}{2}}}\right)\right], \quad \epsilon = -1, \tag{86'a}$$

and the metric has the form (66a), with y (instead of ψ) as the timelike coordinate. The special case $\alpha = 0$ reduces to a solution found by Thorne[4] in which the evolution is purely transverse, with L and f constant, and $S(\tau)$ is a semicircle in the (S, τ) plane.

For $\epsilon = +1$, acceptable solutions result for $g_0 < 0$ and $(y^2 + \alpha y + g_0) > 0$, so that (86) yields

$$S = S_0(y^2 + \alpha y + g_0)^{\frac{1}{2}} \times \exp\left[\frac{-\alpha}{(\alpha^2 - 4g_0)^{\frac{1}{2}}} \coth^{-1}\left(\frac{2y + \alpha}{(\alpha^2 - 4g_0)^{\frac{1}{2}}}\right)\right], \quad \epsilon = +1, \tag{86'c}$$

with the metric given by (66c), with y as timelike coordinate.

Kantowski[10] first obtained these solutions for $\gamma = 2$ using a different parametrization.

VIII. SUMMARY

We have systematically derived a number of exact solutions of the Einstein–Maxwell equations for spatially homogeneous axially symmetric space-times containing an ideal fluid or a uniform magnetic field parallel to the axis of symmetry or both. The space-times we have considered have the metric forms

$$ds^2 = d\tau^2 - f^2(\tau)\,d\chi^2 - S^2(\tau)(d\theta^2 + \sin^2\theta\,d\phi^2), \quad \epsilon = -1, \tag{1a}$$

$$ds^2 = d\tau^2 - f^2(\tau)\,d\chi^2 - S^2(\tau)(d\eta^2 + d\zeta^2), \quad \epsilon = 0, \tag{1b}$$

$$ds^2 = d\tau^2 - f^2(\tau)\,d\chi^2 - S^2(\tau)(d\theta^2 + \sinh^2\theta\,d\phi^2), \quad \epsilon = +1. \tag{1c}$$

In many cases, the field equations did not readily permit exact solution for $f(\tau)$ and $S(\tau)$; in such cases, it frequently proved useful and convenient to replace the timelike coordinate τ by a timelike parameter ψ (or σ, or y) such that the metric forms (1) became

$$ds^2 = L^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)(d\theta^2 + \sin^2\theta\,d\phi^2), \quad \epsilon = -1, \tag{66'a}$$

$$ds^2 = L^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)(d\eta^2 + d\zeta^2), \quad \epsilon = 0, \tag{66'b}$$

$$ds^2 = L^2(\psi)\,d\psi^2 - f^2(\psi)\,d\chi^2 - S^2(\psi)(d\theta^2 + \sinh^2\theta\,d\phi^2), \quad \epsilon = +1, \tag{66'c}$$

and such that closed-form expressions could be obtained for $L(\psi)$, even if $\tau - \tau_0 = \int L(\psi)\, d\psi$ could not be reduced to closed form. In some cases, this parametrization permitted the coupled Einstein equations to be separated and solved, if the parameter was suitably chosen. (See Secs. V–VII above.) It is hoped that this technique will be useful to other workers.

For convenience, a key to the solutions is presented in the three tables below: Table I, universes containing only magnetic fields; Table II, universes containing both magnetic fields and an ideal fluid; Table III, universes containing only an ideal fluid.

TABLE I. Magnetic universes.

	Metric forms (1)			Metric forms (66)					Pertinent references
ϵ	ds^2	$f(\tau)$	$S(\tau)$	ds^2	$L(\psi)$	$f(\psi)$	$S(\psi)$	$\tau(\psi)$	(in chronological order)
−1	(1′a)	—	(20a)	(25a)	$L = S$	(28a)	(23a)	(23a)	22, 23
0	(1′b)	—	(20b)	(25b)	$L = S$	(28b)	(23b)	(23b)	24, 25, 14, 28
+1	(1′c)	—	(20c)	(25c)	$L = S$	(28c)	(23c)	(23c)	
−1	(33a)		(31)						26
+1	a	a	(31)						26

[a] In this case, the metric cannot be written in the form (1′c), but is given by (33c).

TABLE II. Universes containing a magnetic field and an ideal fluid.

	Metric forms (1)				Metric forms (66)					Pertinent references
γ	ϵ	ds^2	$f(\tau)$	$S(\tau)$	ds^2	$L(\psi)$	$f(\psi)$	$S(\psi)$	$\tau(\psi)$	(in chronological order)
1	−1				(25a)	$L = S$	(35a)	(23a)	(23a)	11, 27, 6
1	0				(25b)	$L = S$	(35b)	(23b)	(23b)	11, 27, 6
1	+1				(25c)	$L = S$	(35c)	(23c)	(23c)	11, 27, 6
$\frac{4}{3} < \gamma < 2$	0	(1′b)	(50)[a]	(49)[a]						14

[a] This is not the most general solution possible for this case; see Eq. (46) and subsequent discussion.

TABLE III. Universes containing only an ideal fluid.

	Metric forms (1)				Metric forms (66)					Pertinent references
γ	ϵ	ds^2	$f(\tau)$	$S(\tau)$	ds^2	$L(\psi)$	$f(\psi)$	$S(\psi)$	$\tau(\psi)$	(in chronological order)
1	−1				(25a)	$L = S$	(54a)	(53a)	(53a)	11, 4, 27, 9, 6
1	0	(1′b)	(52)	(51)	(25b)	$L = S$	(54b)	(53b)	(53b)	29, 30, 31, 11, 27, 6
1	+1				(25c)	$L = S$	(54c)	(53c)	(53c)	31, 11, 27, 9, 6
$\frac{4}{3}$	−1				(66a)	(81b)	(82b)	(79b)	—	31, 10
$\frac{4}{3}$	+1				(66c)	(81b)	(82b)	(79b)	—	31, 10
$1 < \gamma < 2$[a]	0				(66b)	(67)	(68a)	(65a)	(69)[d]	11, 13, 28[e]
$1 < \gamma < 2$[b]	0	(1′b)	(71)	(72)						11, 13, 28[e]
$1 < \gamma < 2$[c]	0				(66b)	(67)	(68b)	(65b)	(69)[d]	11, 13, 28[e]
2	−1				(66a)	(87)	(88)	(86′a)	—	10
2	0	(1′b)	(74)	(73)						13, 28
2	+1				(66c)	(87)	(88)	(86′c)	—	10

[a,b,c] These three cases correspond to different values of a constant of integration.
[d] $\tau(\psi)$ may be expressed in closed form only for certain rational values of γ, namely those given by (70).
[e] Reference 11 gives solutions only for $\gamma = \frac{4}{3}$; Refs. 13 and 28 give closed-form solutions only for γ given by (70). The closed-form solutions given in this paper apply for *all* γ in the range $1 < \gamma < 2$.

In all cases, the magnetic field is given by Eq. (9b). The solutions for models containing a magnetic field are expressed in terms of a constant of integration q which is related to H_0 by

$$q^2 \equiv \tfrac{1}{2}\kappa H_0^2.$$

For models containing an ideal fluid, the behavior of the fluid is given by Eq. (13).

ACKNOWLEDGMENT

We wish to thank Dr. K. C. Jacobs for calling a number of pertinent references to our attention.

* This work was performed under the auspices of the U.S. Atomic Energy Commission.

[1] R. B. Partridge and D. F. Wilkinson, Phys. Rev. Letters **18**, 557 (1967); E. K. Conklin and R. N. Bracewell, *ibid.*, 614 (1967); A. A. Penzias and R. W. Wilson, Science **156**, 1100 (1967).

[2] S. W. Hawking and R. J. Tayler, Nature **209**, 1278 (1966); R. V. Wagoner, Science **155**, 1369 (1967), and references given therein.

[3] Ya. B. Zel'dovich, Zh. Eksp. Teor. Fiz. **48**, 986 (1965) [Sov. Phys. JETP **21**, 656 (1965)].

[4] K. S. Thorne, thesis, Princeton University, Princeton, N.J., 1965.

[5] K. S. Thorne, Bull. Am. Phys. Soc. **11**, 340 (1966).

[6] K. S. Thorne, Astrophys. J. **148**, 51 (1967).

[7] A. G. Doroshkevich, Ya. B. Zel'dovich, and I. D. Novikov, Zh. Eksp. Teor. Fiz. Pis'ma **5**, 119 (1967) [Sov. Phys. JETP Letters **5**, 96 (1967)].

[8] C. W. Misner, Phys. Rev. Letters **19**, 533 (1967); Astrophys. J. **151**, 431 (1968).

[9] R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).

[10] R. Kantowski, thesis, University of Texas, Austin, Texas, 1966.

[11] A. G. Doroshkevich, Astrofizika **1**, 255 (1965).

[12] J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).

[13] K. C. Jacobs, Astrophys. J. **153**, 661 (1968).

[14] K. C. Jacobs, Astrophys. J. **155**, 379 (1969).

[15] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).

[16] C. W. Misner, J. Math. Phys. **4**, 924 (1963).

[17] H. Flanders, *Differential Forms* (Academic, New York, 1963).

[18] B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse* (University of Chicago Press, Chicago, 1965), pp. 101–104.

[19] Ya. B. Zel'dovich {Zh. Eksp. Teor. Fiz. **41**, 1609 (1961) [Sov. Phys. JETP **14**, 1143 (1962)]; **43**, 1561 (1962) [**16**, 1102 (1963)]} was one of the first to seriously popularize the γ-law equation of state (12). The limits $1 \leq \gamma \leq 2$ result from the requirements that the stresses in the fluid be pressures rather than tensions and that the speed of sound in the fluid not exceed the speed of light *in vacuo*.

[20] This technique for treating nonlinear differential equations in which the independent variable does not appear explicitly is given by E. Kamke, *Differentialgleichungen: Lösungsmethoden und Lösungen* (Chelsea, New York, 1959), Vol. I, p. 554, Sec. 6.54. We have presented the method in detail since it is frequently useful in obtaining exact solutions of the Einstein field equations, but does not appear to be universally known.

[21] J. P. Vajk, J. Math. Phys. **10**, 1145 (1969). See also Ref. 18, pp. 126–27.

[22] H. Reissner, Ann. Physik **50**, 106 (1916); G. Nordstrøm, Verhandl. Koninkl. Ned. Akad. Wetenschap, Afdel. Natuurk., Sec. I, **26**, 1201 (1918).

[23] J. C. Graves and D. R. Brill, Phys. Rev. **120**, 1507 (1960).

[24] G. Rosen, J. Math. Phys. **3**, 313 (1962); Phys. Rev. **136**, B297 (1964).

[25] D. R. Brill, Phys. Rev. **133**, B845 (1964).

[26] B. Bertotti, Phys. Rev. **116**, 1331 (1959). The solution (33a) was also discovered independently by I. Robinson, Bull. Acad. Polon. Sci., Ser. Math., **7**, 351 (1959), and, in a different form, by B. Hoffmann, in *Recent Developments in General Relativity* (Pergamon, New York, 1962), pp. 279–82.

[27] I. S. Shikin, Dokl. Akad. Nauk SSSR **171**, 73 (1966) [Sov. Phys. Dokl. **11**, 944 (1967)].

[28] K. C. Jacobs, thesis, California Institute of Technology, Pasadena, Calif., 1969.

[29] B. B. Robinson, Proc. Natl. Acad. Sci. **47**, 1852 (1961).

[30] O. Heckmann and E. Schücking, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (Wiley, New York, 1962), Chap. XI.

[31] A. S. Kompaneets and A. S. Chernov, Zh. Eksp. Teor. Fiz. **47**, 1939 (1964) [Sov. Phys. JETP **20**, 1303 (1965)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Wave Equations for Unstable Particles

MAYER HUMI
Department of Mathematics, University of Toronto, Toronto, Canada

(Received 27 January 1970)

A model is presented for the kinematical description of unstable particles in terms of equations which are invariant under broken $IO(4, 1)$. This model agrees with the current models of unstable particles, both those treating lifetimes and mass spectrum. It is shown that the concepts of generalized eigenfunction expansions and complex Fourier transforms enable us to deduce rigorously some of the guesses and assumptions made in these models. Finally, we deal with multiplets of unstable particles and discuss some of the predictions of the theory.

I. INTRODUCTION

The group-theoretical description of unstable particles has drawn the attention of many physicists in recent years. Two distinct approaches to this problem have evolved each of which treats different kinematical aspects of the problem.

In the first approach[1] attempts were concentrated on the description of unstable particles by means of nonunitary representations of the Poincaré group, i.e., emphasizing the problem of the mean lifetime. On the other hand, the second approach[2] concentrated on the continuous mass spectrum which was described by the concept of a generalized free field using multiplicity-free unitary representations of the Poincaré group.

It is a common statement in the literature that although these two approaches are inequivalent from the group-theoretical point of view, since they use different types of representations, they are still complementary in view of the uncertainty relations (which are introduced as external conditions).

However, in this paper we shall develop a model which treats both aspects of the problem (lifetime and mass spectrum) simultaneously and resolves both the group-theoretical problem stated above and the external use of the uncertainty relations. The main new features of this treatment are the construction of Lie-algebra-invariant equations for unstable particles and the use of generalized eigenfunctions[3] and complex Fourier transforms.[1] Admittedly, most of the

* This work was performed under the auspices of the U.S. Atomic Energy Commission.

[1] R. B. Partridge and D. F. Wilkinson, Phys. Rev. Letters **18**, 557 (1967); E. K. Conklin and R. N. Bracewell, *ibid.*, 614 (1967); A. A. Penzias and R. W. Wilson, Science **156**, 1100 (1967).

[2] S. W. Hawking and R. J. Tayler, Nature **209**, 1278 (1966); R. V. Wagoner, Science **155**, 1369 (1967), and references given therein.

[3] Ya. B. Zel'dovich, Zh. Eksp. Teor. Fiz. **48**, 986 (1965) [Sov. Phys. JETP **21**, 656 (1965)].

[4] K. S. Thorne, thesis, Princeton University, Princeton, N.J., 1965.

[5] K. S. Thorne, Bull. Am. Phys. Soc. **11**, 340 (1966).

[6] K. S. Thorne, Astrophys. J. **148**, 51 (1967).

[7] A. G. Doroshkevich, Ya. B. Zel'dovich, and I. D. Novikov, Zh. Eksp. Teor. Fiz. Pis'ma **5**, 119 (1967) [Sov. Phys. JETP Letters **5**, 96 (1967)].

[8] C. W. Misner, Phys. Rev. Letters **19**, 533 (1967); Astrophys. J. **151**, 431 (1968).

[9] R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).

[10] R. Kantowski, thesis, University of Texas, Austin, Texas, 1966.

[11] A. G. Doroshkevich, Astrofizika **1**, 255 (1965).

[12] J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).

[13] K. C. Jacobs, Astrophys. J. **153**, 661 (1968).

[14] K. C. Jacobs, Astrophys. J. **155**, 379 (1969).

[15] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).

[16] C. W. Misner, J. Math. Phys. **4**, 924 (1963).

[17] H. Flanders, *Differential Forms* (Academic, New York, 1963).

[18] B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse* (University of Chicago Press, Chicago, 1965), pp. 101–104.

[19] Ya. B. Zel'dovich {Zh. Eksp. Teor. Fiz. **41**, 1609 (1961) [Sov. Phys. JETP **14**, 1143 (1962)]; **43**, 1561 (1962) [**16**, 1102 (1963)]} was one of the first to seriously popularize the γ-law equation of state (12). The limits $1 \leq \gamma \leq 2$ result from the requirements that the stresses in the fluid be pressures rather than tensions and that the speed of sound in the fluid not exceed the speed of light *in vacuo*.

[20] This technique for treating nonlinear differential equations in which the independent variable does not appear explicitly is given by E. Kamke, *Differentialgleichungen: Lösungsmethoden und Lösungen* (Chelsea, New York, 1959), Vol. I, p. 554, Sec. 6.54. We have presented the method in detail since it is frequently useful in obtaining exact solutions of the Einstein field equations, but does not appear to be universally known.

[21] J. P. Vajk, J. Math. Phys. **10**, 1145 (1969). See also Ref. 18, pp. 126–27.

[22] H. Reissner, Ann. Physik **50**, 106 (1916); G. Nordstrøm, Verhandl. Koninkl. Ned. Akad. Wetenschap, Afdel. Natuurk., Sec. I, **26**, 1201 (1918).

[23] J. C. Graves and D. R. Brill, Phys. Rev. **120**, 1507 (1960).

[24] G. Rosen, J. Math. Phys. **3**, 313 (1962); Phys. Rev. **136**, B297 (1964).

[25] D. R. Brill, Phys. Rev. **133**, B845 (1964).

[26] B. Bertotti, Phys. Rev. **116**, 1331 (1959). The solution (33a) was also discovered independently by I. Robinson, Bull. Acad. Polon. Sci., Ser. Math., **7**, 351 (1959), and, in a different form, by B. Hoffmann, in *Recent Developments in General Relativity* (Pergamon, New York, 1962), pp. 279–82.

[27] I. S. Shikin, Dokl. Akad. Nauk SSSR **171**, 73 (1966) [Sov. Phys. Dokl. **11**, 944 (1967)].

[28] K. C. Jacobs, thesis, California Institute of Technology, Pasadena, Calif., 1969.

[29] B. B. Robinson, Proc. Natl. Acad. Sci. **47**, 1852 (1961).

[30] O. Heckmann and E. Schücking, *Gravitation: An Introduction to Current Research*, L. Witten, Ed. (Wiley, New York, 1962), Chap. XI.

[31] A. S. Kompaneets and A. S. Chernov, Zh. Eksp. Teor. Fiz. **47**, 1939 (1964) [Sov. Phys. JETP **20**, 1303 (1965)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Wave Equations for Unstable Particles

MAYER HUMI
Department of Mathematics, University of Toronto, Toronto, Canada

(Received 27 January 1970)

A model is presented for the kinematical description of unstable particles in terms of equations which are invariant under broken $IO(4, 1)$. This model agrees with the current models of unstable particles, both those treating lifetimes and mass spectrum. It is shown that the concepts of generalized eigenfunction expansions and complex Fourier transforms enable us to deduce rigorously some of the guesses and assumptions made in these models. Finally, we deal with multiplets of unstable particles and discuss some of the predictions of the theory.

I. INTRODUCTION

The group-theoretical description of unstable particles has drawn the attention of many physicists in recent years. Two distinct approaches to this problem have evolved each of which treats different kinematical aspects of the problem.

In the first approach[1] attempts were concentrated on the description of unstable particles by means of nonunitary representations of the Poincaré group, i.e., emphasizing the problem of the mean lifetime. On the other hand, the second approach[2] concentrated on the continuous mass spectrum which was described by the concept of a generalized free field using multiplicity-free unitary representations of the Poincaré group.

It is a common statement in the literature that although these two approaches are inequivalent from the group-theoretical point of view, since they use different types of representations, they are still complementary in view of the uncertainty relations (which are introduced as external conditions).

However, in this paper we shall develop a model which treats both aspects of the problem (lifetime and mass spectrum) simultaneously and resolves both the group-theoretical problem stated above and the external use of the uncertainty relations. The main new features of this treatment are the construction of Lie-algebra-invariant equations for unstable particles and the use of generalized eigenfunctions[3] and complex Fourier transforms.[1] Admittedly, most of the

predictions that follow from this model are known, *guessed*, or *stated as assumptions* in other treatments, but it seems to us that their consistent rigorous derivation from one single and simple equation deserves presentation.

II. EQUATIONS DESCRIBING UNSTABLE PARTICLES

In this section we derive equations for unstable particles by generalizing the 5-dimensional approach to the construction of free-stable-particle equations.[4] The main point in this method is the use of $IO(4, 1)$ [inhomogeneous $O(4, 1)$ group] as a symmetry group of the desired equations which is broken in the last step of the construction to yield $IO(3, 1)$ invariance only.

Explicitly, the method utilizes for the description of free stable particles those representations of $IO(4, 1)$ which satisfy $p_a p^a = 0$, while on the covariant fields the following conditions are imposed:

$$p_a p^a \psi(p) = 0, \quad \Omega_{ab}(p)\psi(p) = 0, \tag{1}$$

where

$$\Omega_{ab}(p) = \tfrac{1}{2}\epsilon_{abcde}P^c W^{de},$$
$$W_{ab}(p) = \tfrac{1}{2}\epsilon_{abcde}P^c M^{de}, \quad a, b, \cdots = 0, \cdots, 4. \tag{2}$$

Here, M^{de} and P^c are the generators of $O(4, 1)$ and translations, respectively. The desired free stable equations then result upon restriction of $IO(4, 1)$ to the $IO(3, 1)$ subgroups which leave p_4 invariant. Equation (1) now takes the form

$$p_\mu p^\mu \psi(p) = m^2 \psi(p),$$
$$\Omega_{ab}(|p_4| = m)(p) = 0. \tag{3}$$

Now, it can be easily seen that, if we put $m_2 \leq p_4 \leq m_1$, instead of $p_4 = m$, then the derived equations

$$p_a p^a \,(m_2 \leq p_4 \leq m_1)\, \psi(p) = 0,$$
$$\Omega_{ab}\,(m_2 \leq p_4 \leq m_1)\, \psi(p) = 0 \tag{4}$$

will be Lorentz invariant (as they remain invariant under transformations which leave p_4 unchanged) and describe a particle with a continuous mass. Nevertheless, since we require an explicit expression of the particle instability, we must refer to nonunitary representations of $IO(4, 1)$ and their decomposition with respect to $IO(3, 1)$.

The nonunitary representations of $IO(4, 1)$ which we need in our treatment are characterized by complex momenta which satisfy[1]

$$\pi_a \pi^a = 0, \quad \pi_a = p_a + iq_a, \quad q_a = -(\Gamma/2b)p_a, \tag{5}$$

where Γ and b are some constants.

The little group for a momentum vector in these representations will be the same as for the unitary representations of $IO(4, 1)$ characterized by $p_a p^a = 0$; namely,[5] $E(3)$ (Euclidean group in 3-dimensions) and the specific representations of this group which we use in the following are characterized by finite real spin (as for stable particles). Following the same arguments as in Ref. 4, one can prove that the nonunitary representations of $IO(3, 1)$ which appear in the reduction of the above-mentioned representations are[6]

$$[M, s] \quad M = (1 - i\Gamma/2b)m, \tag{6}$$

each value $-\infty < m < \infty$ appears once, $(0, s), \cdots, (0, -s)$.

Therefore, to derive the equations we are looking for, we must take the x representation of Eq. (1), apply the complex Fourier transform[1] to p_a, q_a which satisfy (5), and then restrict $p_4 = \mathrm{Re}\,\pi_4$ in the appropriate limits. Then, we find the equations which an unstable particle with a continuous mass spectrum must satisfy to be

$$\pi_a \pi^a \,(m_2 \leq p_4 \leq m_1)\, \psi(\pi) = 0,$$
$$\Omega_{ab}\,(\pi;\ m_2 \leq p_4 \leq m_1)\, \psi(\pi) = 0. \tag{7}$$

For spin-$\frac{1}{2}$ particles these equations reduce after some algebra to

$$[\pi_\mu \pi^\mu + \pi_4 \pi^4 \theta(m_1 m_2)]\psi(\pi) = 0, \tag{8a}$$
$$[\gamma^\mu \pi_\mu + \gamma^5 \pi_4 \theta(m_1 m_2)]\psi(\pi) = 0, \tag{8b}$$

where

$$\theta(x_1, x_2) = 1, \quad x_1 \leq x \leq x_2,$$
$$= 0, \quad \text{otherwise.}$$

Only (8b) is needed, as every solution of it satisfies the first. Here, we make the following remarks:

(1) In the limit when $\pi_4\theta(m_1 m_2)$ is replaced by $M\delta(p_4 - m)$, Eqs. (7) reduce to those given in Ref. 1.

(2) In our equations we use a multiplicity-free representation of $IO(3, 1)$ [as is evident from Eq. (6)]. Therefore, our description conforms to the assumptions made by Lurçat[2] in this respect. However, let us note that in our model this is a *result* rather than an *assumption*.

(3) It is important to realize that these equations possess a higher symmetry in some limit. Thus, the extra parameter which describes the continuous mass is not just a dummy variable as in other treatments[2] but appears as one of the generators of a broken higher symmetry. We shall see the significance of this fact in Sec. IV.

III. SECOND QUANTIZATION

State functions of free stable particles belong to the space S of fast decreasing functions. On the other hand, unstable-particle state functions belong to the space D of functions with compact support. This basic difference enables us to quantize the scalar free unstable field by using the functionals $e^{\pm i\pi\cdot x}$ for the Fourier analysis of $\psi(x, \Gamma)$. The fact that these functions form a complete orthonormal basis was proved

by Gel'fand[3] (as they are generalized eigenfunctions of a self-adjoint operator in the appropriate rigged Hilbert space).

Therefore, we may write the following generalized eigenfunction expansion:

$$\psi(x,\Gamma)=\frac{1}{(2\pi)^2}\int_c \frac{d^4\pi^\mu}{2w_\pi}[A(\pi)e^{-i\pi x}+B^+(\pi)e^{i\pi x}],\qquad \mu=1\cdots 4, \quad (9)$$

where the range of π_4 is restricted by $m_2 \leq \operatorname{Re}\pi_4 \leq m_1$, $w_\pi=(\pi_\mu\pi^\mu)^{\frac{1}{2}}$, $\pi\cdot x=\pi_0x_0-\pi_\mu x^\mu$ and $A(\pi)$, $B(\pi)$, etc., are defined by

$$A(\pi)=\frac{i}{(2\pi)^2}\int d^4x_\mu e^{i\pi x}\overleftrightarrow{\partial}_0\psi(x). \quad (10)$$

Now, if we postulate as usual the following commutation relations (CR's):

$$[\partial_0\psi^*(x),\psi(y)]=\delta(x-y),$$
$$[\partial_0\psi(x),\psi^*(x)]=\delta(x-y), \quad (11)$$

we can then deduce the CR

$$[A(\pi),A^+(\pi')]=2w_\pi\delta(\pi_\mu-\pi'_\mu),\quad \mu=1\cdots 4. \quad (12)$$

We notice here that w_π contains exactly the right phase factor so as to make the rhs of (12) real. This *deduction* appears in Ref. 1 as an *assumption*. However, following this result, the introduction and the calculation of the two point functions $\Delta_5(x_a-y_a)$ is parallel to that given in Ref. 1 (except for the necessary modifications for the continuous spectrum).

IV. APPLICATIONS

A. Multiplets of Unstable Particles

Starting with the $O(4,1)$ invariant Majorana equation

$$(\Gamma_a p^a-k_0)\psi(\pi)=0, \quad (13)$$

we can construct the appropriate Majorana equation describing unstable particles with continuous mass spectra by the same method described in Sec. II. The resulting equation is

$$[\Gamma_\mu\pi^\mu+\theta(m_1,m_2)\Gamma_4\pi^4-k_0]\psi(\pi)=0. \quad (14)$$

We solve this equation by the same method suggested in Ref. 7. We find the mass bands of this equation to be (timelike solutions only)

$$m_1^2+k_0^2/n^2\leq m^2\leq m_2^2+k_0^2/n^2. \quad (15)$$

However, we emphasize that, in contrast to the formal phenomenological way[7] in which the group $IO(4,1)$ is introduced, in general, and which affords no explanation for its breaking effect, Eq. (14) suggests, naturally, the introduction of $IO(4,1)$ group as a limiting exact symmetry for unstable elementary particles and an explanation of its breaking effect.

B. Currents Nonconservation ($s=\frac{1}{2}$)

The generalized eigenfunctions of Eq. (8b) are

$$\psi(\pi)=u(\pi)e^{i\pi x}, \quad (16)$$

where $u(\pi)$ is a 4-vector. The corresponding generalized currents will be defined as

$$J_\mu(\pi)=\bar\psi(\pi)\gamma_\mu\psi(\pi), \quad (17)$$

which are not conserved. In fact,

$$\partial_\mu J^\mu=i\theta(m_1m_2)\partial_5 J_5, \quad (18)$$

where

$$J_5=\bar\psi(\pi)\gamma_5\psi(\pi).$$

Thus, rather than just stating that the current J_μ is not conserved, Eq. (18) enables us to trace current nonconservation to the symmetry breaking of $IO(4,1)$. It is clear, however, that this nonconservation is due to the fact that we do not consider the formation of new particles out of the decaying particles. Thus, if we consider, e.g., the reaction $c\to a+b$ (a and b stable) total current conservation will imply that

$$\partial_\mu J_a^\mu+\partial_\mu J_b^\mu-\partial_\mu J_c^\mu=0. \quad (19)$$

To sum up, we suggested in this paper a new model [and definition by Eq. (7)] of unstable particles which is a natural extension of the stable one. This model treats lifetime and mass spectrum of unstable particles on equal footing. Moreover, it enabled us to rigorously deduce some of the arbitrary assumptions made in other unstable particle models and suggest an interpretation of the $IO(4,1)$ symmetry (and its breaking effect) used in elementary particle physics.

ACKNOWLEDGMENTS

The author is deeply indebted to Professor W. T. Sharp, Professor A. Grossman, and Professor D. Masson whose remarks clarified many points in this paper. I would also like to thank Professor H. S. M. Coxeter and Professor W. T. Sharp for the kind hospitality extended to me at the University of Toronto.

[1] T. Kawai and M. Gato, Nuovo Cimento **60B**, 21 (1969); D. Zwanzinger, Phys. Rev. **131**, 2818 (1963).
[2] F. Lurçat, Phys. Rev. **173**, 1461 (1968); A. L. Licht, Ann. Phys. **34**, 161 (1965); J. Lukierski, Bull. Acad. Sci. Polon. **16**, 343 (1968).
[3] J. P. Antoine, J. Math. Phys. **10**, 53 (1969).
[4] J. A. de Vos and J. Hilgeroord, Nucl. Phys. **B1**, 494 (1967).
[5] S. Sannikov, Yad. Fiz. **4**, 587 (1966) [Sov. J. Nucl. Phys. **4**, 416 (1967)].
[6] In Ref. 4 the authors employed representations with $m>0$ only; hence, each representation of $IO(3,1)$ appeared twice. However, the correct complete classification of the Poincaré group representations as given by Wigner [Ann. Math. **40**, 149 (1939)] shows that both $m>0$ and $m<0$ should be considered and hence the multiplicity-free decomposition.
[7] C. Fronsdal, Phys. Rev. **156**, 1665 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Another Note on Two Binomial-Coefficient Identities of Rosenbaum*

H. M. SRIVASTAVA
Department of Mathematics, University of Victoria, Victoria, British Columbia, Canada

(Received 6 February 1970)

This paper aims at providing a direct and unified approach to two binomial-coefficient identities proved recently by Rosenbaum [J. Math. Phys. **8**, 1973 (1967)], who made use of rather long and involved considerations of commutation relations. The analysis presented here verifies the existence of the wider ranges of validity of these identities, as was subsequently pointed out by Gould [J. Math. Phys. **10**, 49 (1969)].

For all real or complex x and integers $n \geq 0$, we let

$$\binom{x}{n} = \frac{x(x-1)\cdots(x-n+1)}{n!}, \quad \text{with} \quad \binom{x}{0} = 1, \tag{1}$$

denote a binomial coefficient, and put

$$A = \sum_{n=0}^{\alpha} (-1)^n \binom{n+\epsilon-1}{n} \binom{\epsilon}{\alpha-n} \tag{2}$$

and

$$B = \sum_{n=0}^{\alpha} (-1)^n \binom{n+\epsilon-1}{\alpha-1} \binom{\alpha}{n}. \tag{3}$$

In a recent paper, Rosenbaum[1] proved that $A = 0$ for integers α and ϵ such that $\epsilon \geq \alpha \geq 2$ and $B = 0$ for integers α and ϵ such that $\alpha \geq 1$, $\epsilon \geq 2$, and $\epsilon > \alpha$. Subsequently, Gould[2] made use of certain facts about binomial coefficients and finite differences to show that both A and B are, of course, zero for all integers $\alpha \geq 1$ and for all real or complex values of ϵ. It may be of interest to physicists as well as mathematical analysts to observe how systematically these identities could be viewed together in terms of hypergeometric functions.

We start from definition of the factorial function (Ref. 3, p. 19)

$$(a)_n = a(a+1)\cdots(a+n-1), \quad n \geq 1, \quad \text{with} \quad (a)_0 = 1, \tag{4}$$

whence it follows at once that

$$(-m)_n = 0 \tag{5}$$

for all integers m and n such that $0 \leq m < n$.

Now we put

$$S^{\nu,\lambda,m}_{p,q,r} = \sum_{n=0}^{m} (-1)^n \binom{\nu+n}{mp+nq-r} \binom{\lambda}{m-n}, \tag{6}$$

where p, q, and r are nonnegative integers such that $mp \geq r$. Then, comparing (6) with (2) and (3), we get

$$S^{\epsilon-1,\epsilon,\alpha}_{0,1,0} \equiv A \tag{7}$$

and

$$S^{\epsilon-1,\alpha,\alpha}_{1,0,1} \equiv B. \tag{8}$$

Making use of the definitions (1) and (4), we readily see that

$$S^{\nu,\lambda,m}_{p,q,r} = \binom{\lambda}{m} \binom{\nu}{mp-r} \times \sum_{n=0}^{m} \frac{(-m)_n(\nu+1)_n}{(\lambda-m+1)_n(mp-r+1)_{nq}(\nu-mp+r+1)_{n-nq}}. \tag{9}$$

For integers $q \geq 2$, the finite series on the right-hand side of (9) can be expressed as a hypergeometric ${}_{q+2}F_{q+1}$ function with argument $q^{-q}(1-q)^{q-1}$. In the special cases $q = 0$ and $q = 1$, which would serve our purpose here, if we particularize other parameters of (9) in accordance with (7) and (8), the resulting hypergeometric series can be summed by using the Gauss theorem in the form (Ref. 3, p. 23)

$${}_2F_1[-n, b; c; 1] = (c-b)_n/(c)_n, \tag{10}$$

and we finally have

$$\alpha A = \epsilon B = (1-\alpha)_\alpha/(\alpha-1)! = 0, \tag{11}$$

by virtue of (5), for all integers $\alpha \geq 1$.

Therefore, for all integers $\alpha \geq 1$, it follows that $A = 0$, for all real or complex values of ϵ, and $B = 0$, for all real or complex values of $\epsilon \neq 0$. In case of the identity $B = 0$, the additional restriction $\epsilon \neq 0$, which was not stated in Gould's paper, can be waived fairly easily from the definition (3). Indeed, when $\epsilon = 0$, (3) would reduce to

$$B = \sum_{n=0}^{\alpha} (-1)^n \binom{n-1}{\alpha-1} \binom{\alpha}{n} = (-1)^{\alpha-1} + (-1)^\alpha = 0, \tag{12}$$

leading us to the desired fact that $B = 0$ for any integer $\alpha \geq 1$ if $\epsilon = 0$.

* Supported in part by the National Research Council of Canada under Grant A7353.

[1] D. M. Rosenbaum, J. Math. Phys. **8**, 1973 (1967).
[2] H. W. Gould, J. Math. Phys. **10**, 49 (1969).
[3] I. N. Sneddon, *Special Functions of Mathematical Physics and Chemistry* (Oliver and Boyd, London, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Isotropy of Space

VITTORIO GORINI*
Istituto di Scienze Fisiche dell'Università, Milano, Italy, and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Italy

AND

ANTONIO ZECCA
Istituto di Scienze Fisiche dell'Università, Milano, Italy

(Received 13 February 1970)

It is well known that it is possible to deduce the Lorentz transformations solely on the basis of the principles of relativity, space–time homogeneity, and isotropy of space, without making any use of the principles of reciprocity and of the invariance of the velocity of light. Here we put forward a rigorous and simple formulation of the isotropy principle, which appears to be more general than those usually presented and which, together with the results of a preceding paper, allows us to carry out such a deduction in the physical case of a 4-dimensional space–time in a straightforward and considerably simplified way.

1. INTRODUCTION AND STATEMENT OF PURPOSE

It has been known since the pioneer paper by Frank and Rothe[1] that the principle (a) of invariance of the velocity of light is not necessary to derive the Lorentz transformations. Instead, it can be seen that these transformations can be obtained from a set of general principles in which no reference is made to a universal constant whatsoever. In this connection, it appears that the most primitive and fundamental principles are the principles of (i) relativity, (ii) space–time homogeneity, and (iii) isotropy of space. Indeed, it has been proved by Lalan in a remarkable paper which appeared more than thirty years ago[2] that (i)–(iii) alone are sufficient, together with an argument which is substantially equivalent to the assumption of existence of an arrow of time, to rule out all possible kinematics with the exception of the Lorentz and the Galilean ones. Lalan's treatment, which exploits the principle of relativity essentially through its group content and which is based on the study of the structure of the 1-dimensional Lie subgroups of $GL(2, R)$ with the choice of a velocity parameter as a coordinate over the group manifolds, can be considerably simplified by adding to the principles (i)–(iii) the so-called principle (b) of reciprocity which states that the velocity of an inertial frame of reference S with respect to another inertial frame of reference S' is the opposite of the velocity of S' with respect to S.[3] Recently, it has been shown[4] that the above reciprocity relation can be deduced in a very simple way from (i) and (iii). Thus, it is in fact not necessary to assume it as a postulate with the purpose of providing a straightforward and elementary way of deduction of the coordinate transformations between two inertial frames without the requirement of invariance of a given velocity.

Without the use of (a) the common feature of all derivations of the Lorentz transformations, which, to our knowledge, have appeared in the literature,[5] is that they are applied to a 2-dimensional space–time or (which is almost equivalent) they start from a velocity transformation along an axis between frames with parallel axes. This transformation is written in the form

$$\begin{aligned} x_1' &= a(v)x_1 + b(v)x_0, \\ x_2' &= e(v)x_2, \\ x_3' &= e(v)x_3, \\ x_0' &= c(v)x_1 + d(v)x_0, \quad x_0 = t, \end{aligned} \tag{1}$$

where $v = -b(v)/a(v)$ denotes the velocity of the primed frame with respect to the unprimed one. Here the conditions that the corresponding space axes are equally oriented and that the times flow in the same direction are (compare I)

$$\begin{aligned} &\frac{\partial x_1'}{\partial x_1} = a(v) > 0, \\ &\frac{\partial x_2'}{\partial x_2} = \frac{\partial x_3'}{\partial x_3} = e(v) > 0, \\ &\frac{\partial x_0'}{\partial x_0} = d(v) > 0, \\ &\frac{\partial x_1}{\partial x_1'} = \frac{d(v)}{a(v)[d(v) + vc(v)]} > 0, \\ &\frac{\partial x_0}{\partial x_0'} = \frac{1}{d(v) + vc(v)} > 0. \end{aligned} \tag{1'}$$

Of course, this procedure implies no loss of generality because, in view of the homogeneity of space–time and of the isotropy of space, it is always possible to

let the space–time origins coincide by means of a translation and, at the same time, it is possible to go over to parallel axes, while orienting the x_1 axes of both frames along the direction of the relative velocity, by means of appropriate space rotations. The justification for writing the transformations in form (1) once the above operations have been suitably performed is ascribed to the isotropy of space which is supposed to tell us that x_1' and x_0' do not depend on x_2 and x_3 and that x_2' and x_3' are proportional, with the same coefficient, to the corresponding unprimed coordinates x_2 and x_3. In this context, however, appeals to the isotropy principle are usually rather vague and unsatisfactory and there lacks a formal and general statement of the principle which would unavoidably lead to transformations (1) for parallel axes and relative velocity directed along the x_1 axis. When the Lorentz transformations are derived with the use of (a), this omission is not very important and can be forgiven. Indeed, in this case, once the linearity has been granted as a consequence of space–time homogeneity or, alternatively, of the principle of inertia, the whole Poincaré group P can be immediately deduced, as is well known, by applying (a) to the set of transformations[6]

$$x_\mu' = L_{\mu\nu}x_\nu + a_\mu, \quad \mu, \nu = 1, 2, 3, 0, \tag{2}$$

to get

$$g_{\mu\nu}L_{\mu\rho}L_{\nu\sigma} = g_{\rho\sigma},$$
$$\{g_{\rho\sigma}\} = \text{diag}\,(-1, -1, -1, c^2), \tag{3}$$

and the velocity transformations along an axis are easily obtained as special cases of Eqs. (2) plus (3).[7]

On the other hand, if one wants to avoid using (a), it does not seem to be easy to arrive at the structure of P by the use of (i)–(iii) alone, without passing through the particular transformation (1). Therefore, it would be advisable that in an axiomatic deduction of P from (i)–(iii), the isotropy principle be given a rigorous and general formulation which should both conform to intuition and allow an unambiguous and straightforward derivation of formulas (1) as the basis for the subsequent analysis. It is the purpose of the present paper to put forward and exploit such a formulation in a convenient way, thus filling the gap which was left open in I, where transformations (1) [with $e(v) = 1$] were taken for granted and the basic principles directly applied to the 2-dimensional case. A successful attempt in this direction has already been made by Lalan[8] who requires the set of transformations among frames with parallel axes to be invariant under space rotations. However, we believe this statement to require more than the isotropy principle actually implies, and we shall see that the same results can be obtained on a basis which is less restrictive and sticks closer to the actual content of the principle itself.

2. FORMULATION OF THE ISOTROPY PRINCIPLE

Let S and S' be two inertial frames using orthogonal space axes, and let

$$x_\mu' = f_\mu(x_1, x_2, x_3, x_0) \tag{4}$$

be the transformation equations which express the coordinates of an event as seen by S' in terms of the coordinates of the same event as seen by S.

We denote by e_1, e_2, e_3 (respectively, e_1', e_2', e_3') the space axes of S (respectively of S').

The assumption that space–time is homogeneous implies the linearity of the f_μ. The proof of this assertion, which was given in I for the 2-dimensional case, holds equally well here. Hence, we write (4) as

$$x_\mu' = L_{\mu\nu}x_\nu, \tag{5}$$

modulo a space–time translation, and, by the invertibility condition,

$$\det\{L_{\mu\nu}\} \neq 0. \tag{6}$$

We suppose that S and S' are not at rest with respect to each other.

Denote by $\mathbf{v} = \{v_1, v_2, v_3\}$ the velocity of S' with respect to S as given by its components along the space axes of S and by $\mathbf{w} = \{w_1', w_2', w_3'\}$ the velocity of S with respect to S' as given by its components along the space axes of S'. We have[9]

$$L_{ik}v_k + L_{i0} = 0 \tag{7}$$

and

$$w_i' = L_{i0}/L_{00}. \tag{8}$$

Now let S' perform a space rotation which directs the axis e_1' parallel to $\mathbf{w}$. This gives, from (8),

$$L_{20} = L_{30} = 0, \tag{9}$$

and we set

$$L_{10}/L_{00} = w_1' = w \neq 0. \tag{10}$$

Then let S perform a corresponding rotation which directs the axis e_1 parallel to $\mathbf{v}$. Setting[10]

$$v_1 = v \neq 0, \tag{11}$$

we have, from (7),

$$L_{i1}v + L_{i0} = 0; \tag{12}$$

whence, by (9),

$$L_{21} = L_{31} = 0 \tag{13}$$

and

$$v = -L_{10}/L_{11}. \tag{14}$$

Thus, the matrix $L = \{L_{\mu\nu}\}$, which characterizes a velocity transformation along the first space axis, has the following structure:

$$L = \begin{pmatrix} L_{11} & L_{12} & L_{13} & L_{10} \\ 0 & L_{22} & L_{23} & 0 \\ 0 & L_{32} & L_{33} & 0 \\ L_{01} & L_{02} & L_{03} & L_{00} \end{pmatrix}, \tag{15}$$

on which we impose the conditions [compare (1′)]

$$L_{11} > 0 \tag{15′}$$

and

$$L_{00}/[L_{11}(L_{00} + vL_{01})] > 0,$$

which express the requirement that the axes e_1 and e_1' be equally oriented (we can always reduce ourselves to this situation by possibly carrying out a rotation of π of one of the two frames about any axis perpendicular to its x_1 axis).

Once the form (15) has been attained, all the transformations and the only ones, which can be performed on the frames S and S'—without changing their respective states of motion—with the condition that (9), (13), and (15′) should continue to hold and by keeping unaltered the sign of v, are arbitrary and mutually independent rotations of both S and S' about their respective x_1 axes. In other words, if we denote by $K(v)$ the set of matrices which correspond to all possible velocity transformations along the first space axis for a given value v of the velocity of S' with respect to S and which satisfy (15′), and by $L(v)$ we denote an arbitrary fixed element of $K(v)$, we have that every matrix $M \in K(v)$ can be written at least in one way in the form

$$M = R(\alpha)L(v)R(\beta) = M_{(\alpha,\beta)}, \tag{16}$$

where

$$R(\vartheta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\vartheta & \sin\vartheta & 0 \\ 0 & -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{17}$$

and there is no restriction on the range of variation of the angles α and β as M varies in $K(v)$.

Lemma: The relation among the elements of $K(v)$, $N \sim N'$ iff $\exists$ angles α, β, and β' such that

$$N = M_{(\beta,\alpha)}, \quad N' = M_{(\beta',\alpha)}, \tag{18}$$

is an equivalent relation.

Proof: Reflexivity and symmetry are trivial. To prove transitivity, suppose $N \sim N'$ and $N' \sim N''$. Then, by hypothesis, there exist angles α, β, and β' such that

$$N = R(\beta)LR(\alpha), \quad N' = R(\beta')LR(\alpha),$$

and angles γ, δ', and δ'' such that

$$N' = R(\delta')LR(\gamma), \quad N'' = R(\delta'')LR(\gamma).$$

There follows

$$N'' = R(\delta'' - \delta')R(\delta')LR(\gamma) = R(\delta'' - \delta')R(\beta')LR(\alpha)$$
$$= R(\delta'' - \delta' + \beta')LR(\alpha),$$

which proves $N \sim N''$.

We denote the general equivalence class by

$$K_\alpha = \{N \mid \exists\, \beta \text{ such that } N = M_{(\beta,\alpha)}\}. \tag{19}$$

We state the principle of isotropy of space in regard to the set $K(v)$ as follows: *there exists only one equivalence class, namely the quotient set $K(v)/\sim$ has a single element.* Formally,

$$K_\alpha = K_{\alpha'}, \quad \forall\, \alpha \text{ and } \alpha'. \tag{20}$$

Indeed, as it appears from (16), K_α and $K_{\alpha'}$ represent the two sets of velocity transformations along the first space axis for which the unprimed frame S is (counterclockwise) rotated about the axis e_1 by an angle $-\alpha$ and, respectively, by an angle $-\alpha'$, starting from a given initial configuration. Now, by the isotropy of space expressed as the equivalence of all directions perpendicular to the direction of the relative motion, we must always get equal sets of transformations no matter how α and α' are chosen.

3. USE OF THE ISOTROPY PRINCIPLE

Explicitly, condition (20) states that for every α, α', and β there exists a β' such that

$$R(\beta')L(v)R(\alpha') = R(\beta)L(v)R(\alpha). \tag{21}$$

It is easily seen that this condition is equivalent to the following one: For every angle α there exists an angle β such that

$$R(\beta)L(v) = L(v)R(\alpha). \tag{22}$$

Equation (22) is a particular case of the standard equation $AX = XB$ in the theory of matrices, of which the general solution is known.[11] However, in order to keep our treatment as elementary as possible, we do not refer to the general theorem of Ref. 11 but, instead, work out the solution explicitly.

Inserting (15) and (17) into (22) and introducing the 2×2 matrices

$$P(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$$

and

$$A(v) = \begin{pmatrix} L_{22} & L_{23} \\ L_{32} & L_{33} \end{pmatrix},$$

we get the conditions

$$L_{r2} = L_{r2}\cos\alpha - L_{r3}\sin\alpha,$$
$$L_{r3} = L_{r2}\sin\alpha + L_{r3}\cos\alpha, \quad r = 1, 0, \tag{23}$$

and

$$P(\beta)A(v) = A(v)P(\alpha). \tag{24}$$

Introducing the matrix

$$W = \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}, \tag{25}$$

we have

$$WP(\alpha)W^{-1} = \tfrac{1}{2}\begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix} = Q(\alpha). \tag{26}$$

Further, we set

$$WAW^{-1} = B = \begin{pmatrix} B_{22} & B_{23} \\ B_{32} & B_{33} \end{pmatrix}. \tag{27}$$

Then, writing (24) as

$$Q(\beta)B = BQ(\alpha), \tag{28}$$

we get the equations

$$B_{22}e^{i\beta} = B_{22}e^{i\alpha}, \tag{29a}$$
$$B_{23}e^{i\beta} = B_{23}e^{-i\alpha}, \tag{29b}$$
$$B_{32}e^{-i\beta} = B_{32}e^{i\alpha}, \tag{29c}$$
$$B_{33}e^{-i\beta} = B_{33}e^{-i\alpha}. \tag{29d}$$

Condition (6) implies $B \neq 0$, and suppose first that $B_{22} \neq 0$. This implies

$$\beta = \alpha, \mod 2\pi, \tag{30}$$

and, by the arbitrariness of α,

$$B_{23} = B_{32} = 0, \tag{31}$$

while, by (6), B_{22} and B_{33} are nonzero complex numbers which are subject only to the condition that A be a real matrix.

On the other hand, if $B_{23} \neq 0$, Eq. (29b) implies

$$\beta = -\alpha, \mod 2\pi. \tag{32}$$

Hence,

$$B_{22} = B_{33} = 0, \tag{33}$$

while B_{23} and B_{32} are nonzero complex numbers which are again subject to the condition that A be real.

Corresponding to (30), we have

$$A = W^{-1}BW = W^{-1}\begin{pmatrix} B_{22} & 0 \\ 0 & B_{33} \end{pmatrix} W$$
$$= \tfrac{1}{2}\begin{pmatrix} B_{22} + B_{33} & i(B_{33} - B_{22}) \\ i(B_{22} - B_{33}) & B_{22} + B_{33} \end{pmatrix} \tag{34}$$

and the reality condition implies that

$$\operatorname{Im} B_{22} = -\operatorname{Im} B_{33} = F$$

and

$$\operatorname{Re} B_{22} = \operatorname{Re} B_{33} = E,$$

whence

$$A = \begin{pmatrix} E & F \\ -F & E \end{pmatrix}, \tag{35}$$

where, by (6), $(E^2 + F^2)^{\frac{1}{2}} = K \neq 0$. Setting $\cos\vartheta = E/K$, $\sin\vartheta = F/K$, we can write A as

$$A = \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}\begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix}, \quad K > 0. \tag{36}$$

Similarly, corresponding to (32), we obtain

$$A = \begin{pmatrix} G & H \\ H & -G \end{pmatrix}, \quad (G^2 + H^2)^{\frac{1}{2}} = K \neq 0. \tag{37}$$

Setting $\cos\vartheta = G/K$, $\sin\vartheta = H/K$, we can write A as

$$A = \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{pmatrix}, \quad K > 0. \tag{38}$$

Finally, by the arbitrariness of α, Eqs. (23) imply that

$$L_{12} = L_{13} = L_{02} = L_{03} = 0. \tag{39}$$

The preceding results can be summed up as follows.

Proposition: Equation (22) together with the condition $\det L(v) \neq 0$ admits of solutions only if $\beta = \alpha$ or $\beta = -\alpha$.

(1) If $\beta = \alpha$, the general solution is

$$L(v) = \begin{pmatrix} a(v) & 0 & 0 & b(v) \\ 0 & e(v) & 0 & 0 \\ 0 & 0 & e(v) & 0 \\ c(v) & 0 & 0 & d(v) \end{pmatrix} R(\vartheta(v))$$
$$= L_0(v)R(\vartheta(v)), \tag{40}$$

where the angle ϑ is an arbitrary function of v and, to the extent of the implications of (22) and provided (6) is satisfied, the coefficients a, b, c, d, and e are arbitrary as well (e can be chosen to be positive).

(2) If $\beta = -\alpha$, the general solution is

$$L(v) = L_0(v)\,\Sigma\,R(\vartheta(v)), \tag{41}$$

where

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{42}$$

Thus, if we impose conditions (1') on $L_0(v)$, we see that, according to whether the determinant of the transformations is positive or negative, by choosing $\vartheta(v) = 0$ we can select the transformation corresponding to the matrix $L_0(v)$ or the one corresponding to the matrix $L_0(v)\ \Sigma$ as a standard velocity transformation along an axis. The first transformation refers to space frames which are of the same type (both left-handed or both right-handed) and to space axes which are parallel and equally oriented. On the other hand, the second one refers to space frames of opposite types (one left-handed and the other one right-handed) having parallel and equally oriented x_1 and x_2 axes and oppositely oriented x_3 axes. The corresponding equations (30) and (32) express the obvious fact that an α rotation of S' about the axis e_1' appears to S as an α or as a $-\alpha$ rotation according to whether S and S' are of the same type or not.

Thus, we have come to the conclusion that the isotropy principle, as expressed by (20), uniquely leads to formulas (1) for a velocity transformation along an axis between frames with parallel axes.

Further, provided the two frames use equal length and time standards, the coefficients of (1) are uniquely determined functions of v which, by the principle of relativity, are independent of either frame.

To show that $e(v)$ has to be taken equal to 1, we must make use of the results of I. Let the two frames S and S' be connected by transformation (1) and suppose they both perform a rotation of π about their respective x_3 axes. Formally,

$$S \to \bar{S} = TS \quad \text{and} \quad S' \to \bar{S}' = TS', \tag{43}$$

where

$$T = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{44}$$

The matrix which connects $\bar{S}$ to $\bar{S}'$ is

$$\bar{L}_0(v) = TL_0(v)T = \begin{pmatrix} a(v) & 0 & 0 & -b(v) \\ 0 & e(v) & 0 & 0 \\ 0 & 0 & e(v) & 0 \\ -c(v) & 0 & 0 & d(v) \end{pmatrix}, \tag{45}$$

which is again of the form (1) and satisfies (1') if (1) does. Hence, we can write

$$\bar{L}_0(v) = L_0(\bar{v}) = \begin{pmatrix} a(\bar{v}) & 0 & 0 & b(\bar{v}) \\ 0 & e(\bar{v}) & 0 & 0 \\ 0 & 0 & e(\bar{v}) & 0 \\ c(\bar{v}) & 0 & 0 & d(\bar{v}) \end{pmatrix}, \tag{46}$$

where, by (45), $\bar{v} = -v$. Therefore, by comparison [see I, formulas (18)],

$$a(-v) = a(v), \tag{47a}$$

$$b(-v) = -b(v), \tag{47b}$$

$$c(-v) = -c(v), \tag{47c}$$

$$d(-v) = d(v), \tag{47d}$$

$$e(-v) = e(v). \tag{47e}$$

Writing

$$w = b(v)/d(v) = \varphi(v), \tag{48}$$

we have

$$L_0(\varphi(v)) = L_0^{-1}(v); \tag{49}$$

hence

$$e(\varphi(v)) = e^{-1}(v). \tag{50}$$

By using (47) and the principle of relativity [stating $\varphi(\varphi(v)) = v$], we can prove that

$$\varphi(v) = -v \tag{51}$$

exactly in the same way as in I.

Then, combining (50), (51), and (47e), we get

$$e^2(v) = 1, \tag{52}$$

whence, since $e(v)$ is positive,

$$e(v) = 1. \tag{53}$$

At this point, the whole treatment can be freely restricted to the 2-dimensional case, and we refer to I for the derivation of the Lorentz transformations.

4. CONCLUSION

The combined results of I and of the present paper provide a straightforward, rigorous, and most elementary deduction of the Lorentz transformations (hence of the Poincaré group) which does not make use of the reciprocity principle and of the existence of an invariant velocity, being based only on the principle of relativity together with the customary assumptions in regard to the structure of the space–time manifold in special relativity; namely, space–time homogeneity and isotropy of space.

* Address from March 15, 1970: Institut für Theoretische Physik (1) der Universität Marburg, 355 Marburg/Lahn, Renthof 7, Federal Republic of Germany.

[1] P. Frank and H. Rothe, Ann. Physik **34**, 825 (1911).

[2] V. Lalan, Bull. Soc. Math. France **65**, 83 (1937).

[3] For a derivation of the Lorentz transformations from the principles (i)–(iii) and (b) see, for instance, Ya. P. Terletskii, *Paradoxes in the Theory of Relativity* (Plenum, New York, 1968), Chap. II, Sec. 7.

[4] V. Berzi and V. Gorini, J. Math. Phys. **10**, 1518 (1969), hereafter referred to as I.

[5] A list of references can be found in I.

[6] The standard summation convention is adopted.

[7] For instance, V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon, London, 1959), Sec. 10.

[8] V. Lalan, Ref. 2.

[9] Latin indices run from 1–3.

[10] It is easy to control that $v \neq 0$ iff $w \neq 0$.

[11] F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1960), Vol. I, Chap. VIII, Theorem 1.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 7 JULY 1970

Classification of Finite-Dimensional Irreducible Representations of Connected Complex Semisimple Lie Groups*

A. K. BOSE
Département de Physique, Université de Montréal, Montréal, Canada
AND
J. PATERA
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Canada

(Received 1 December 1969)

Using the results of Malcev and Dynkin, the irreducible representations of connected components of complex semisimple Lie groups are classified. The simple Lie algebras A_1, B_n, C_n, G_2, F_4, E_7, and E_8 have no outer automorphism and, for the Lie algebra D_{2k}, the automorphism corresponding to the contragredient transformation is inner. Hence, all the irreducible representations of all simple Lie algebras except A_n $(n > 1)$, D_{2k+1}, and E_6 are self-contragredient. For A_n $(n > 1)$, D_{2k+1}, and E_6, an irreducible representation is self-contragredient, provided its highest weight possesses the symmetry to reduce the outer automorphism of the algebra into an inner one for the representation. All the self-contragredient irreducible representations are classified into orthogonal and symplectic types by reducing the problem to the case of the angular-momentum algebra.

I. INTRODUCTION

A unitary irreducible representation (UIR) of a compact simple Lie group belongs to one of the three Wigner[1] classes, complex, real, or pseudo-real. Recently Mehta[2] and Mehta and Srivastava[3] classified the finite-dimensional UIR's of (compact) simple Lie groups. In this article, we present the classification of the finite-dimensional irreducible representations (IR's) of the connected complex semisimple Lie groups.

Here we are, in fact, bringing together results of Malcev[4] and Dynkin.[5,6] In contrast to the method used in Refs. 2 and 3, which requires separate and laborious investigation of each simple Lie group, the present method is simple and does not require case considerations. It is based on the knowledge of the outer automorphisms of corresponding Lie algebras and on a certain property of the principal 3-dimensional subalgebra. Indeed, all IR's of a semisimple Lie algebra which has no nontrivial outer automorphism are trivially self-contragredient (real and pseudo-real in Wigner's classification). Also, all the IR's of the Lie algebra of type D_{2k} are automatically self-contragredient. For the other simple Lie algebras, an IR is self-contragredient provided its highest weight possesses the symmetry to reduce the outer automorphism of the algebra to an inner one for the representation. A self-contragredient IR of a semisimple Lie algebra induces a representation of its principal 3-dimensional subalgebra which is, in general, reducible but contains one irreducible component which always carries the property of orthogonality or symplecticity (reality or pseudo-reality in Wigner's classification) of the inducing representation.

In Sec. 2, all self-contragredient IR's of a semisimple Lie algebra are found and, in Sec. 3, they are classified into orthogonal and symplectic types.

Because of the 1:1 correspondence between the connected part of a semisimple complex Lie group and its Lie algebra,[7] one can go freely from a group to its corresponding algebra whenever this is convenient.

II. AUTOMORPHISMS AND SELF-CONTRAGREDIENT REPRESENTATIONS

Let $\varphi(g)$ be a linear representation of a semisimple Lie group g and $\bar{\varphi} \equiv [\varphi(g)]^{t-1}$ the contragredient representation. Here, t denotes the transpose (in a conveniently chosen basis). The representations of the corresponding Lie algebra G are then related as follows:

$$\bar{\varphi}(G) = -[\varphi(G)]^t. \tag{1}$$

A representation $\varphi(g)$ is called self-contragredient if there exists an inner automorphism such that $\bar{\varphi}(g) = S^{-1}\varphi(g)S$, $S \in \varphi(g)$.

An automorphism of a Lie algebra is an automorphism of the vector space of the algebra which preserves commutators. Let G be a semisimple Lie algebra of rank n, and let $\pi = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ be a system of its simple roots relative to a fixed Cartan subalgebra H. Let e_{α_i} denote the root vector corresponding to α_i. The matrix (A_{ij}), $A_{ij} = 2(\alpha_i, \alpha_j)/(\alpha_i, \alpha_i)$, is the Cartan matrix of G determined by π. A mapping $\alpha_i \to \alpha_{i'}$ of π onto itself preserving Cartan's scalar products (i.e., $A_{ij} = A_{i'j'}$) is an automorphism of G defined by $f(\alpha_i) = \alpha_{i'}$, $f(e_{\alpha_i}) = e_{\alpha_{i'}}$. Let P denote the group of all such automorphisms. Clearly, the

isometric mapping f is an automorphism of the Dynin diagram[8] of G. Only the identity mapping $\alpha_i \to \alpha_i$, $i = 1, 2, \cdots, n$, acts as the identity in H and is an inner automorphism of G. Thus, one can write uniquely each automorphism

$$q = q_0 f, \tag{2}$$

where q_0 is an inner automorphism and f is an automorphism of the Dynkin diagram.

Let $\varphi(G)$ be an IR of G. We can choose a basis in the representation space such that $\varphi(h)$, $h \in H$, are diagonal matrices. From (1), it follows that

$$\bar{\varphi}(h) = -\varphi(h), \quad h \in H. \tag{3}$$

Let us define θ by $\theta e_{\pm\alpha_i} = e_{\mp\alpha_i}$. Then the commutation relations yield $\theta h_i = -h_i$, $i = 1, 2, \cdots, n$, and $\theta\alpha = -\alpha$, $\alpha \in \pi$. Clearly, θ is the automorphism which transforms a representation of a semisimple Lie algebra into its contragredient partner. By (2), one has $\theta = q_0 f_\theta$, where f_θ denotes the corresponding automorphism of the Dynkin diagram. Since $\varphi(q_0 G)$ is equivalent to $\varphi(G)$, one has $\bar{\varphi}(G) = \varphi(\theta G) \simeq \varphi(f_\theta G)$. *Hence, $\varphi(G)$ is self-contragredient if and only if the diagram automorphism f_θ is the identity.*

Clearly, it suffices to consider only simple Lie algebras. By looking at the Dynkin diagrams of the simple Lie algebras (Fig. 1), one sees that the group P is the identity for A_1, B_n, C_n, G_2, F_4, E_7, and E_8 (and none other). Hence, all the IR's of these Lie algebras are necessarily self-contragredient.

We now determine f_θ for the remaining Lie algebras, namely A_n $(n > 1)$, D_n (≥ 4), and E_6. Let $\Lambda(\Lambda_1, \Lambda_2, \cdots, \Lambda_n)$ be the highest weight of an IR $\varphi(G)$, where

$$\Lambda_i = 2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i), \quad \alpha_i \in \pi, \tag{4}$$

are nonnegative integers relative to the basis of simple roots as numbered in Fig. 1. Let $\Delta\varphi$ denote the system of weights of φ (see Sec. 3). Hence, by (3), $\Delta\bar{\varphi} = -\Delta\varphi$. Since the highest weight of an IR determines its system of weights, it follows that φ is self-contragredient if the lowest weight $\Lambda_l = -\Lambda$ and conversely.

Using the general algorithm[8] for finding $\Delta\varphi$, one computes the lowest weight and finds that $\Lambda_l = -\Lambda$ for any IR of D_{2k}, i.e., all IR's of D_{2k} are self-contragredient. Similarly, one finds that, for A_n $(n > 1)$, D_{2k+1} $(k > 1)$, and E_6, there are IR's which are not self-contragredient, i.e., f_θ is not the identity. For Lie algebras of types A and D one can easily determine explicitly their automorphisms and thus find f_θ directly.[5]

By referring again to Fig. 1, one sees that each of the diagrams A_n $(n > 1)$, D_{2k+1} $(k > 1)$, and E_6 has only one nontrivial automorphism which must then be necessarily f_θ. Hence, one has for A_n $(n > 1)$:

$$f_\theta(\alpha_i) = \alpha_{n+1-i}, \tag{5a}$$

for D_{2k+1} $(k > 1)$,

$$f_\theta(\alpha_i) = \alpha_i, \; i \leq 2k - 1,$$
$$f_\theta(\alpha_{2k}) = \alpha_{2k+1}, f_\theta(\alpha_{2k+1}) = \alpha_{2k}, \tag{5b}$$

and, for E_6,

$$f_\theta(\alpha_i) = \alpha_{6-i}, \; i \leq 5, f_\theta(\alpha_6) = \alpha_6. \tag{5c}$$

The highest weight $\bar{\Lambda}$ of $\bar{\varphi}$ is obtained from Λ of φ by applying the permutation f_θ appropriate to the algebra on its system of roots. By (4), one then has

$$\bar{\Lambda}_i = f_\theta(\Lambda_i) = 2(\Lambda, \alpha_{i'})/(\alpha_{i'}, \alpha_{i'}), \quad \alpha_{i'} = f_\theta(\alpha_i).$$

Hence, those and only those IR's of A_n $(n > 1)$, D_{2k+1} $(k > 1)$, and E_6 whose highest weights remain invariant under the appropriate permutation f_θ of (5) are self-contragredient. All such representations are given in Fig. 2.

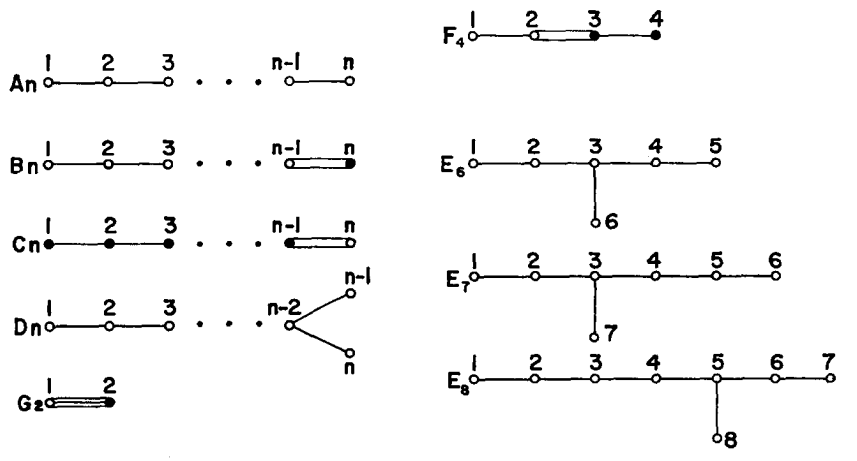


FIG. 1. Numbering of simple roots of simple Lie algebras (black dots represent shorter roots).

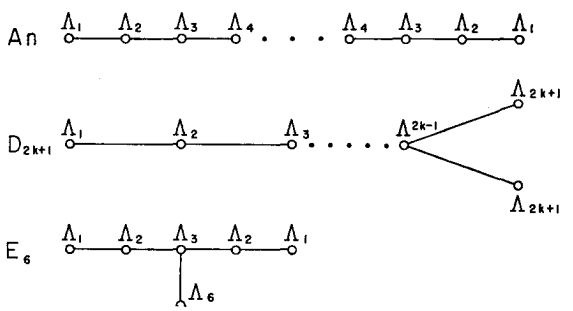


FIG. 2. Self-contragredient representations of Lie algebras of types A_n $(n > 1)$, D_{2k+1} $(k > 1)$, and E_6 $[\Lambda_i = 2(\Lambda, \alpha_i)/(\alpha_i, \alpha_i)]$.

III. ORTHOGONAL AND SYMPLECTIC REPRESENTATIONS AND PRINCIPAL SUBALGEBRA

Let φ be a linear representation of G. Let G be a semisimple subalgebra of G and $\tilde{\varphi}$ the representation of $\tilde{G}$ induced by φ, i.e., $\varphi(x) = \tilde{\varphi}(x)$, $x \in \tilde{G}$. Let us choose the basis of their Cartan subalgebras such that $\tilde{H} \subset H$. Let ξ be a weight vector of φ with the weight M. Then it is also a weight vector of $\tilde{\varphi}$, with the weight, say, $\tilde{M}$. This implies that $\varphi(h)\xi = (M, h)\xi$, $h \in H$, and $\tilde{\varphi}(h)\xi = (\tilde{M}, h)\xi$, $h \in \tilde{H}$. That is $(M, h) = (\tilde{M}, h)$, $h \in \tilde{H}$. Hence, $M - \tilde{M}$ is orthogonal to $\tilde{H}$, i.e., $\tilde{M}$ is obtained from M by an orthogonal projection onto $\tilde{H}$.

We now construct a particular subalgebra $\tilde{G}$ of G. We choose root vectors for each simple root of G in such a way that $(e_\alpha, e_{-\alpha}) = 1$. Let

$$\beta = \sum_{\alpha \in \pi} p_\alpha \alpha, \quad e_\beta = \sum_{\alpha \in \pi} u_\alpha e_\alpha, \quad e_{-\beta} = \sum_{\alpha \in \pi} u_{-\alpha} e_{-\alpha}. \tag{6}$$

We choose p_α so that, for every $\alpha \in \pi$,

$$(\beta, \alpha) = (\beta, \beta), \tag{7}$$

and let $u_\alpha u_{-\alpha} = p_\alpha$. Then it is easy to verify the commutation relations

$$[\beta, e_\beta] = (\beta, \beta)e_\beta, \quad [\beta, e_{-\beta}] = -(\beta, \beta)e_{-\beta},$$
$$[e_\beta, e_{-\beta}] = (e_\beta, e_{-\beta})\beta.$$

Hence β, e_β, and $e_{-\beta}$ form a canonical basis of a 3-dimensional simple subalgebra $\tilde{G}$, called the principal subalgebra of G.

Equations (6) and (7) define an orthogonal projection of the n-dimensional root space of G onto the 1-dimensional space generated by β. From (7), it is evident that any simple root is projected onto β, i.e., $\alpha_i \to \beta$, $\alpha_i \in \pi$.

Let $\varphi(G)$ be an IR of G with the highest weight Λ. Then every weight M of $\varphi(G)$ is of the form $\Lambda - \sum k_i \alpha_i$, where k_i are nonnegative integers. We decompose the system of weights into classes, called levels, defined by the integer $k = \sum k_i$, which denotes the number of simple roots that are subtracted from Λ. Thus, the kth level consists of weights of the form $\Lambda(k) = \Lambda - \alpha_{i_1} - \alpha_{i_2} - \cdots - \alpha_{i_k}$, $\alpha_{i_j} \in \pi$. One can then characterize Λ as the zeroth-level weight. Let r be the largest possible value of k. Then the total number of levels is $r + 1$.

A weight $\Lambda(k)$ of the kth level of $\varphi(G)$ is projected onto a weight, say $\tilde{\Lambda}(k)$, of the induced representation $\tilde{\varphi}(\tilde{G})$. By (7), one has then $\tilde{\Lambda}(k) = \tilde{\Lambda} - k\beta$. It follows that no weight but Λ is projected onto the zeroth level of the system of weights of $\tilde{\varphi}(\tilde{G})$. This means that $\tilde{\Lambda}$ is a highest weight of $\tilde{\varphi}(\tilde{G})$. If $\tilde{\varphi}(\tilde{G})$ has other highest weights, they must necessarily be projected from levels with $k \neq 0$ of $\varphi(G)$. Hence, we conclude that the induced representation $\tilde{\varphi}(\tilde{G})$, which is in general reducible, contains *one and only one irreducible component*, say $\tilde{\varphi}_1(\tilde{G})$, with the highest weight $\tilde{\Lambda}$ and of dimensionality $r + 1$.

Let us now determine r. It is clear that $r = 2(\tilde{\Lambda}, \beta)/(\beta, \beta)$ and, since $(\Lambda - \tilde{\Lambda}, \beta) = 0$, we get $r = 2(\Lambda, \beta)/(\beta, \beta)$. Let us introduce the coefficients r_α by the relation

$$\frac{\beta}{(\beta, \beta)} = \sum_{\alpha \in \pi} r_\alpha \frac{\alpha}{(\alpha, \alpha)}. \tag{8}$$

Multiplying the above equation by 2Λ, we obtain

$$r = \sum_{\alpha \in \pi} r_\alpha \Lambda_\alpha, \quad \Lambda_\alpha = \frac{2(\Lambda, \alpha)}{(\alpha, \alpha)}. \tag{9}$$

Multiplying (8) by $\gamma \in \pi$ and using (7), we get the system of linear equations

$$1 = \sum_{\alpha \in \pi} r_\alpha \frac{(\gamma, \alpha)}{(\alpha, \alpha)}, \quad \gamma \in \pi, \tag{10}$$

which determines r_α. For the simple algebras, the value of the coefficients r_α are given[6] in Fig. 3. If the algebra is semisimple, the system of its simple roots decomposes into several orthogonal subsystems. Therefore, the system (10) decomposes into several subsystems so that the value of r_α depends only on the subsystem to which α belongs.

Since $\tilde{\varphi}_1(\tilde{G})$ is an IR of a simple Lie algebra of rank one, it follows[1] that its odd-(even-)dimensional representations are orthogonal (symplectic). A self-contragredient IR of a semisimple Lie group always preserves a nondegenerate bilinear form and is orthogonal (symplectic) if the bilinear form is symmetric (skew-symmetric). By Malcev's theorem[9] on reducible representations, we conclude that a self-contragredient IR $\varphi(G)$ of a semisimple Lie algebra is orthogonal (symplectic) if r is even (odd). *Thus, Eq.* (9)

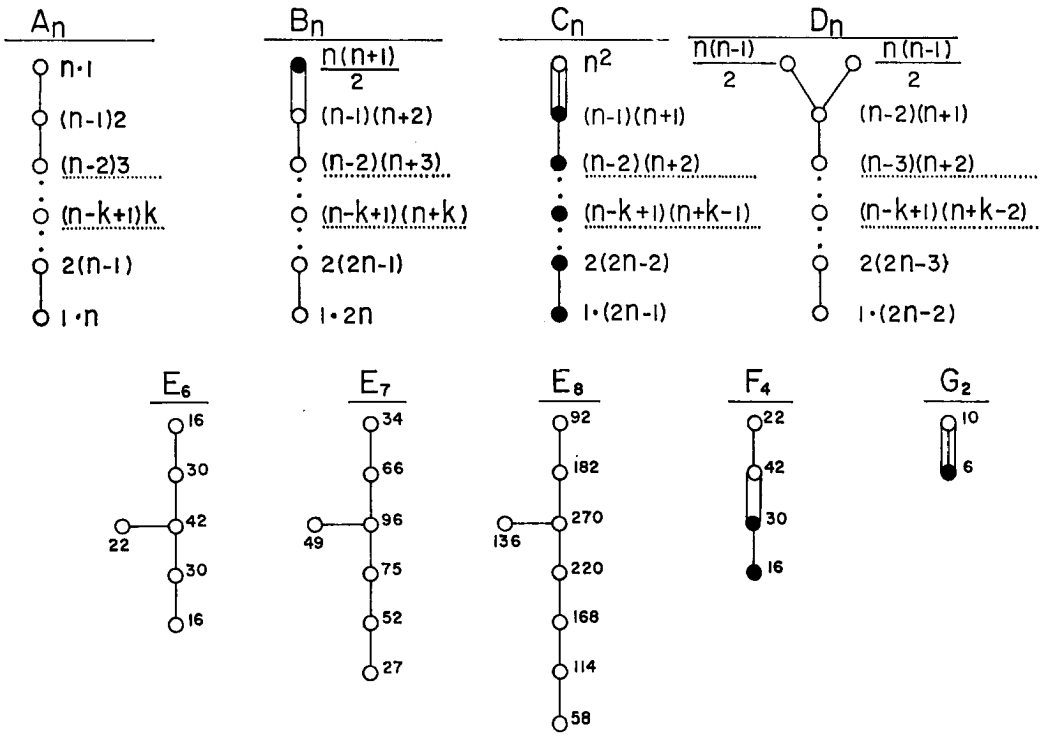


FIG. 3. Values of r_α for simple Lie algebras.

and Fig. 3 classify all the self-contragredient IR's of all semisimple Lie algebras.

The values of r_α in Fig. 3 contain more information than is required. For our purpose, it is sufficient to use $r_\alpha \pmod 2$. Using (9) and the symmetry of the coordinates Λ_i (Fig. 2), we then get the following explicit results. Actually, it involves determining $r_{\frac{1}{2}(n+1)}$ for A_n (n = odd), r_n for B_n, and $r_n(\Lambda_{n-1} + \Lambda_n)$ for D_n.

(i) The self-contragredient IR's of the Lie algebras A_{2k}, A_{4k+3}, B_{4k}, B_{4k+3}, D_{2k+1}, D_{4k}, G_2, F_4, E_6, and E_8 are always orthogonal. One may notice that there are no semisimple Lie algebras all of whose self-contragredient representations are symplectic.

(ii) A self-contragredient IR of A_{4k+1} is orthogonal (symplectic) if Λ_{2k+1} is even (odd).

(iii) An IR of B_n ($n = 4k + 1, 4k + 2$) is orthogonal (symplectic) if Λ_n is even (odd).

(iv) An IR of C_n is orthogonal (symplectic) if $\Lambda_1 + \Lambda_3 + \Lambda_5 + \cdots$ is even (odd).

(v) An IR of D_{4k+2} is orthogonal (symplectic) if $\Lambda_{4k+1} + \Lambda_{4k+2}$ is even (odd).

(vi) An IR of E_7 is orthogonal (symplectic) if $\Lambda_4 + \Lambda_6 + \Lambda_7$ is even (odd).

* Supported in part by the National Research Council of Canada.

[1] E. G. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[2] M. L. Mehta, J. Math. Phys. **7**, 1824 (1966).
[3] M. L. Mehta and P. K. Srivastava, J. Math. Phys. **7**, 1833 (1966).
[4] A. I. Malcev, Am. Math. Soc. Transl. Ser. 1, **9**, 172 (1962).
[5] E. B. Dynkin, Dokl. Akad. Nauk SSSR (N.S.) **76**, 629 (1951).
[6] E. B. Dynkin, Dokl. Akad. Nauk SSSR (N.S.) **71**, 221 (1950).
[7] L. S. Pontryagin, *Topological Groups* (Gordon & Breach, New York, 1966).
[8] E. B. Dynkin, Am. Math. Transl. Ser. 2 (Suppl.) **6**, 245 (1957).
[9] See Ref. 8, p. 361.