JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 4, NUMBER 7 JULY 1963

More Integral Representations for Scattering Amplitudes with Complex Singularities*

C. FRONSDAL, R. E. NORTON, AND K. T. MAHANTHAPPA
University of California, Los Angeles, California
(Received 14 December 1962)

Previous work is generalized in order to achieve a better understanding of the role of complex singularities in connection with integral representations. The most general conditions under which the box-diagram contribution to a scattering amplitude satisfies a representation with real integration contours are derived. Explicit representations are derived in several special cases. It is frequently found possible to obtain representations that are essentially of the Mandelstam type, although the more general Bergman–Oka–Weil representation must be invoked in general. One example of a three-dimensional representation is given, which exhibits the analytic structure in one of the external masses in addition to the kinematical variables. The significance of the "physical sheet" is discussed.

I. INTRODUCTION

ACCORDING to quantum field theory, the elements of the S matrix may be extended analytically to complex values of the kinematical variables.[1] Much effort has been directed to the study of their analytic structure, in the hope that the analytic properties may be combined with unitarity to provide a framework for calculation. With the notable exception of the work of Mandelstam,[2] most results obtained so far concern only the location of singularities,[3] and the evaluation of discontinuities across branch cuts in perturbation theory.[4] Before this information can be usefully exploited, the means must be found to generate the S matrix from the discontinuities. For simple scattering problems this is provided by the Mandelstam representation. For other amplitudes, including all production processes, the essential link is still missing.

The chief difficulty that besets the problem of finding suitable integral representations for production amplitudes is the occurrence of complex singularities.[3] The Mandelstam representation, or more generally multiple Cauchy representations, could, in principle, be applied in this new situation. But in addition to the fact that the integrations would extend over a complex domain, this domain would depend upon the values of the dynamical variables. As the discontinuities are known only in terms of the amplitudes themselves, it seems reasonable to conclude that the resulting integral equations would be impossible to handle.

This difficulty is also present in simple perturbation-theory contributions to ordinary scattering, which is why we feel that some insight can be gained by studying integral representations for a scattering amplitude that has complex singularities. In a previous report, we showed that a representation with real contours of integration, albeit with a rather complicated kernel, could be found even if complex

* Work supported in part by the National Science Foundation.

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

[2] S. Mandelstam, Phys. Rev. **112**, 1344 (1958); *ibid.* **115**, 1741 (1959).

[3] References of particular interest to us are A Chi-Tai-Wu, University of Maryland, Physics Department, Tech. Rept. No. 186 (July, 1960); and J. Tarski, J. Mod. Phys. **1**, 429 (1960).

[4] R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960).

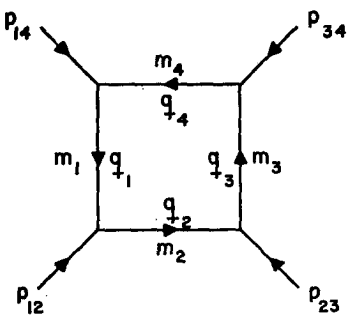

FIG. 1. The "box diagram." Variables s and t are defined by $s = (p_{12} + p_{23})^2$, $t = (p_{12} + p_{14})^2$.

singularities were present.[5] The representation, applicable to a very simple case only, was of the Bergman–Oka–Weil type, with weight functions that could be calculated by Cutkosky's rules.[4] In this paper the program is extended to representations of more general validity.

We study the contribution to the scattering amplitude arising from the diagram of Fig. 1. The external masses M_i are first restricted to the range of values where the Mandelstam representation is valid, and the function so defined is then continued analytically through the upper half planes of the M_i. This procedure[6] is correct for this particular diagram if s and t are taken as kinematical variables, for then Feynman's prescription of adding negative imaginary parts to the internal masses is equivalent to adding positive imaginary parts to s, t, and the M_i.

In all anomalous cases treated, several different integral representations are obtained. These represent different analytic functions that coincide when both variables are in the upper half complex planes, but differ elsewhere. The difference lies in the definition of the "physical sheet," i.e., the Riemann sheet in which the representation yields the amplitude without further analytic continuation. With the proper choice of the kinematical variables s and t, there are no singularities in the product of the six upper half planes of s, t, and M_i, and the amplitude may be extended uniquely into this domain. To continue further, it is necessary to decide how to go around the various branch points that are encountered, notably those on the real axes. A different choice gives a different "physical sheet" and a different integral representation.

We have reached, perhaps provisionally, the following point of view. Extending an amplitude into the complex domain serves two purposes. To connect the amplitudes for different physical processes, it is sufficient to continue the variables into the upper half planes. The sole purpose of further continuation is to relate the amplitude to its discontinuities, and the only obvious criterion for defining the lower half planes is that the integral representation which expresses this relationship assume the simplest possible form. In simple dispersion theory, the cuts may be selected on the consideration that the discontinuities be easily interpretable; in the dispersion relations for forward scattering, the discontinuity is related to the total cross section. At present it is not clear whether similar considerations are relevant for the more general processes.

There are two important corollaries. Generally, an amplitude is obtained as a sum of different terms with different analytic properties, and hence with different branch cuts. It frequently happens that the best choice of the branch cuts differs for the several terms, and that the union of the integration domains extend over the entire real axis. In such cases the complete amplitude can be extended to the lower half of the physical sheet only by way of another sheet. We have no prejudice against this situation, since we are not interested in analytic properties *per se* but only in integral representations. Second, the different diagrams that contribute to an amplitude may require different choices of kinematical variables to ensure that the product of the upper half planes be free of singularities. Therefore it is not convenient, and probably not fruitful, to study the extension of the whole amplitude into the complex domain. Instead, guided by perturbation theory, one should break the amplitude into several parts, each with the simplest possible analytic structure. This situation is already familiar; the Mandelstam representation is a sum of three terms, and the physical boundary value is defined differently in each term.

A simple rule of thumb for choosing the best "physical sheet" is to attempt to keep the complex singularities out of it. This has proved possible in a variety of special cases. As demonstrated explicitly in the examples in Sec. III, it is often possible to obtain representations of essentially the Mandelstam type, even for those cases which are normally characterized by complex singularities.

In Sec. II the amplitude of Fig. 1 is discussed in generality for all values of the external masses. The purpose of the discussion is to delineate those special cases where representations can be found which contain only integrations over real contours. In Sec. III explicit representations are given. One example of a three-dimensional integral formula is

[5] C. Fronsdal, K. T. Mahanthappa, and R. E. Norton, Phys. Rev. **127**, 1847 (1962). See also J. N. Islam, J. Math. Phys. **3**, 1098 (1962).

[6] S. Mandelstam, Phys. Rev. Letters **4**, 84 (1960).

also demonstrated, which displays the analytic structure of the amplitude in one of the external masses, in addition to s and t. This representation may be of particular interest in connection with production amplitudes.

The mathematical techniques are similar to those in the previous paper. The Bergman–Oka–Weil (BOW) representation is frequently employed, and it is perhaps useful to summarize its essential features.[7] Such a discussion will also serve to introduce some definitions and notation.

Consider a function $F(x, y)$ of two complex variables, analytic in the product of the two complex planes except for cuts labeled by indices i, j, etc. The BOW formula is

$$F(x, y) = \sum_{i<j} (2\pi i)^{-2} \int dx'\, dy' F(x', y')(q_x^i q_y^j - q_x^j q_y^i),$$

where the q_x^i and q_y^i are any functions of the primed and unprimed variables which satisfy the conditions

(i) $$(x' - x)q_x^i + (y' - y)q_y^i = 1,$$

(ii) when x' and y' are on the ith cut, the q_x^i and q_y^i are analytic in x and y in the domain of analyticity of $F(x, y)$. The integration extends over all four sides of the intersection of all pairs of cuts. This intersection is termed the distinguished boundary and the reality of this distinguished boundary is necessary and sufficient for the existence of a representation involving only real contours of integration.

If the q_x^i and q_y^i are specialized to satisfy the following additional condition,

(iii) the q_x^i and q_y^i are one-valued functions of x' and y' near the domain of integration,

then the representation simplifies to

$$F(x, y) = \sum_{i<j} (\pi i)^{-2} \int dx'\, dy' \rho^{ij}(x', y')(q_x^i q_y^j - q_x^j q_y^i),$$

where

$$\rho^{ij}(x, y) = (\tfrac{1}{4})(F_{++}^{ij} - F_{+-}^{ij} - F_{-+}^{ij} + F_{--}^{ij}).$$

Here $F_{\pm\pm}^{ij}$ are the four different boundary values of $F(x, y)$ on the intersection of the ith and the jth cuts.

II. REAL DISTINGUISHED BOUNDARIES

The Feynman diagram of Fig. 1 makes a contribution to the scattering amplitude which, in the normal case, can be expressed in the form[2,4,8]

$$A(s, t) = g_1 g_2 g_3 g_4 (2^7 m_1 m_2 m_3 m_4)^{-1} F(x, y), \qquad (1)$$

$$F(x, y) = (\pi i)^{-2} \times \int_1^\infty dx' \int_{y_{+'}}^\infty dy' \rho(x', y')/(x' - x)(y' - y). \qquad (2)$$

Here s and t are the usual invariant energy and momentum-transfer variables (see Fig. 1), and

$$x = [s - (m_1 - m_3)^2]/4m_1 m_3, \qquad (3)$$

$$y = [t - (m_2 - m_4)^2]/4m_2 m_4. \qquad (4)$$

The spectral function $\rho(x, y)$ is

$$\rho(x, y) = [\Delta(x, y)]^{-\frac{1}{2}}, \qquad (5)$$

$$\Delta(x, y) = (4m_1 m_2 m_3 m_4)^{-2} |(q_i q_j)| = (\tfrac{1}{16}) |y_{ij}|. \qquad (6)$$

The variables y_{ij} are the same as those that were used by Tarski[3]:

$$y_{ij} = -(p_{ij}^2 - m_i^2 - m_j^2)/2m_i m_j = (q_i q_j)/m_i m_j. \qquad (7)$$

In terms of these variables,

$$x = \tfrac{1}{2}(1 - y_{13}), \qquad y = \tfrac{1}{2}(1 - y_{24}). \qquad (8)$$

In order to save writing, and to emphasize the distinction between variables and mass parameters, we introduce

$$k_i = -y_{i,i+1}, \qquad (5 \equiv 1). \qquad (9)$$

Then

$$\Delta(x, y) = xy[(x - 1)(y - 1) - (\tfrac{1}{2})(k_2 k_4 + k_1 k_3)] + (\tfrac{1}{4})x(k_1 - k_4)(k_3 - k_2) + (\tfrac{1}{4})y(k_1 - k_2)(k_3 - k_4) + (\tfrac{1}{16})(k_1 k_3 - k_2 k_4)^2 - (\tfrac{1}{16})(k_1 - k_4 + k_3 - k_2)^2. \qquad (10)$$

The lower limit of the y' integration is one of the zero's of $\Delta(x, y)$. As was shown by Tarski,[3] these zero's are given by

$$2x(x - 1)y_{\pm} = x(x - 1) + (\tfrac{1}{2})x(k_1 k_3 + k_2 k_4) + (\tfrac{1}{4})(k_1 - k_2)(k_4 - k_3) \pm [(x - L_{12}^+)(x - L_{12}^-)(x - L_{43}^+)(x - L_{43}^-)]^{\frac{1}{2}}, \qquad (11)$$

$$L_{ij}^{\pm} = (\tfrac{1}{2})\{1 - k_i k_j \pm [(1 - k_i^2)(1 - k_j^2)]^{\frac{1}{2}}\}. \qquad (12)$$

It is sometimes convenient to define

$$k_i \equiv -\cos\theta_i, \qquad (13)$$

$$L_{ij}^{\pm} = (\tfrac{1}{2})[1 - \cos(\theta_i \pm \theta_j)]. \qquad (14)$$

If the order of integration in (2) is reversed, then the lower limits are $y' = 1$ and $x' = x_+'$, with

$$2y(y - 1)x_{\pm} = y(y - 1) + (\tfrac{1}{2})y(k_1 k_3 + k_2 k_4) + (\tfrac{1}{4})(k_1 - k_4)(k_2 - k_3) \pm [(y - L_{14}^+)(y - L_{14}^-)(y - L_{32}^+)(y - L_{32}^-)]^{\frac{1}{2}}. \qquad (15)$$

[7] See references cited by present authors in reference 5.
[8] T. W. B. Kibble, Phys. Rev. 117, 1159 (1960).

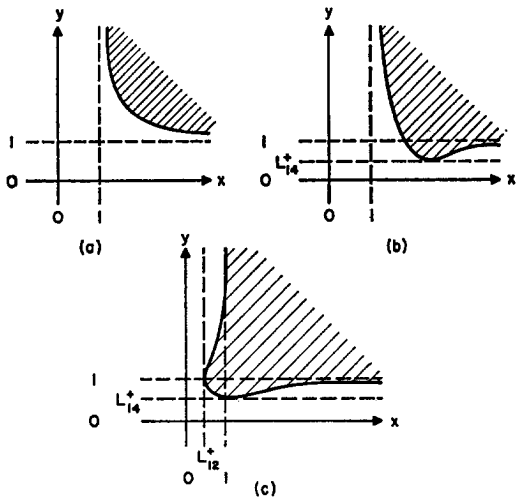

FIG. 2. Integration domains for the Mandelstam representation in (a) the normal case, (b) when $k_1 + k_4$ is positive, and (c) if both $k_1 + k_4$ and $k_1 + k_2$ are positive.

A sufficient condition for the validity of the Mandelstam representation (2) is that all the k_i be negative. A vertex is externally stable if the appropriate k_i is less than $+1$. We now proceed to discuss the manner in which the representation (2) breaks down when the mass parameters are increased.

Without loss of generality, the following conditions can be imposed:

$$k_1 \geq k_i, \qquad k_4 \geq k_2. \tag{16}$$

It is convenient to distinguish several special cases.

A. $1 \geq k_4, k_3, k_2 > -1$

The spectral function, for fixed x, has branch points at $y = y_\pm$, and vice versa. The inner integration starts at $x' = x_+'$ or at $y' = y_+'$, depending on the order of integration. In the normal case, which includes the case $k_i < 0$, all these branch points are real: also $x_+' > x_-'$ when $y' \geq 1$ and vice versa. Trouble develops when k_1 is increased to the point where x_-' or y_-' concide for $y' \geq 1$, or $x' \geq 1$, respectively. The first time this happens is when $L_{14}^+ = 1$; i.e., $k_1 = -k_4$, or $\theta_1 + \theta_4 = \pi$. At this point, the x_-' coincide when $y' = 1$. This

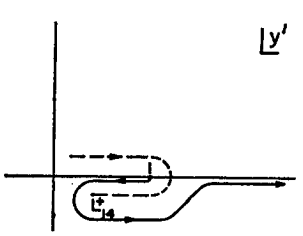

FIG. 3. The path of the point $y' = L_{14}^+$ as $k_1 + k_4$ is increased above zero, and the deformation of the y' contour.

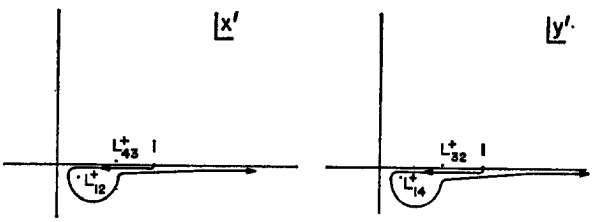

FIG. 4. The contours of the Mandelstam representation become complex if $\theta_1 + \theta_2 + \theta_3 + \theta_4 > 2\pi$.

is a rather trivial problem that disappears when the x' integration is carried out last. In that case, the only thing that happens is that y_-' ceases to be a nonincreasing function of x', and the integration region takes the shape shown in Fig. 2(b). Nevertheless, it is instructive to obtain the same result with the opposite order of integration.

Let k_1 have a small positive imaginary part and let it move over the value $-k_4 < 1$. Then the point L_{14}^+ moves to the right on top of the real axis, cuts the real axis just above 1, and moves back on the lower side. Because x_-' must not coincide when y' is in the integration region (remember that we integrate over y' last), it is necessary to deform the y' contour as shown in Fig. 3. Because the point $y' = L_{14}^+$ is a branch point of the function $x_+'(y')$, we find that the region of integration is $x' > x_+'$ when y' lies on the lower part of the contour and $x' > x_-'$ when y' lies on the upper part. Since the sense of the y' integration is opposite on the two parts of the sling, the net effect is that, when $L_{14}^+ < y' < 1$, x' is integrated from x_+' to x_-'. Thus the result that the integration region is as shown in Fig. 2(b) is confirmed.

Next, let the k_i be increased further, until $k_1 + k_2 > 0$, while $k_2 + k_3$ remains negative. Then the integration region of Fig. 2(c) is obtained. The x' and y' integrations now start at $x' = L_{12}^+$, $y' = y_+'$, or at $y' = L_{14}^+$, $x' = x_+'$. We have tacitly assumed that these limits are real. This is true if $k_1 \leq 1$, $L_{14}^+ \geq L_{32}^+$ and $L_{12}^+ \geq L_{43}^+$. For the present, take $k_1 \leq 1$. If the k_i are increased to the point that L_{32}^+ exceeds L_{14}^+, then the y' integration includes the stretch from L_{14}^+ to L_{32}^+, and here x_+' is complex. Similarly, if L_{43}^+ exceeds L_{12}^+. The condition for either of these disasters to happen is[9]

$$\theta_1 + \theta_2 + \theta_3 + \theta_4 > 2\pi. \tag{17}$$

In that case, the distinguished boundary generally contains the complex points $L_{14}^+ < y < L_{32}^+$, x on some cut starting at x_+ and ending at L_{12}^+, and the points $L_{12}^+ < x < L_{43}^+$, y on some cut starting at y_+

[9] See S. Mandelstam, Phys. Rev. 115, 1741 (1959).

and ending at $y = L_{14}^{+}$, say. This is not the only possible choice of cuts, but the most symmetrical one. The contours are shown in Fig. 4.

We next determine all the special cases when (17) is satisfied, $k_1 < 1$ and the distinguished boundary is real. Let the y' integration be performed last. Then it is clear from (15) that x_+' becomes real for all $y' \geq L_{14}^{+}$ if L_{32}^{+} is moved around 1 and down to coincide with L_{14}^{+}. That is, a real distinguished boundary is obtained if $\theta_1 + \theta_4 = \theta_2 + \theta_3$. Exchanging the order of integration, one obtains, in a similar way, another exceptional case, namely $\theta_1 + \theta_2 = \theta_3 + \theta_4$. It is interesting to see how this last result can be obtained when the y' integration is performed last. In that case, if L_{14}^{-} and L_{32}^{-} coincide, x_+' is real for all y' to the *left* of L_{14}^{+}. Therefore, a real distinguished boundary is obtained if $L_{14}^{-} = L_{32}^{-}$, i.e., $\theta_1 - \theta_4 = |\theta_2 - \theta_3|$, by deforming the y' contour till it runs along the real axis from L_{14}^{+} to $-\infty$ and from 1 to $+\infty$. Summarizing the result for $k_1 \leq 1$, we conclude that if the inequality (17) is not satisfied, then a Mandelstam representation holds, and if (17) *is* satisfied, it is possible to obtain an integral representation with real contours of integration if, and only if, any one of the following equalities is satisfied:

$$\theta_i + \theta_j = \theta_k + \theta_l, \quad i, j, k, l, \text{ all different.} \tag{18}$$

Next, consider the case $k_1 \geq 1$. If k_1 is given a positive imaginary part and increased beyond 1, then L_{14}^{+} and L_{12}^{+} move into the lower half plane (unless k_4 equals $+1$) from the positions shown in Fig. 4. The distinguished boundary is now irrevocably complex, unless either k_4 or k_3 equals $+1$. The case $k_3 = +1$ is special because then $L_{43}^{+} = L_{43}^{-}$, and the x' contour can be deformed till it goes from 1 to $-\infty$. (Fig. 5).

B. $1 \geq k_4 \geq k_2 > -1, k_3 \leq -1$

Now L_{32}^{*} and L_{43}^{*} are complex except in the limits $k_3 = -1$ or $k_4 = +1$. Therefore L_{14}^{+} never encounters L_{32}^{+} on its wanderings up and down the real axis, and x_+' is real. Trouble develops when k_1 reaches 1, for then L_{14}^{+} and L_{12}^{+} become complex, unless $k_4 = +1$. Because L_{32}^{*} are complex conjugates and L_{43}^{*} are complex conjugates, the way is not open for deforming the contour till it lies on the real axis, except in the special case $L_{43}^{+} = L_{43}^{-}$ or $L_{32}^{+} = L_{32}^{-}$, that is, when $k_3 = -1$. Thus the only cases of a real distinguished boundary for $k_1 > 1$ are

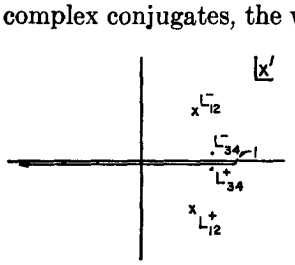

FIG. 5. Possible deformation of the contours when $k_1 > 1$ and $k_3 = +1$, to obtain a real distinguished boundary.

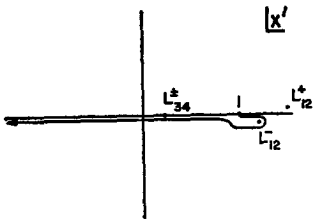

FIG. 6. Possible deformation of the contours in the case $k_1 > 1$, $k_3 = -1$, $k_4^2 < 1$, $k_2 < -1$.

$$k_3 = -1, \quad \text{or} \quad k_4 = +1. \tag{19}$$

C. $1 \geq k_4 > -1, k_3 > -1, k_2 \leq -1$

Now L_{12}^{*} and L_{23}^{*} are complex as long as $k_1 \leq 1$. The integration goes from $y' = L_{14}^{+}$ and $x' = x_+' > 1$, and the Mandelstam representation holds up to $k_1 = 1$. When k_1 reaches 1, L_{12}^{*} meet on the real axis above 1 from opposite sides. As k_1 is increased further, L_{12}^{*} move apart on the real axis, and y_+' is complex when x' lies between them. The only exceptions occur when $k_2 = -1$, or when $k_4 = +1$. Thus the only cases of a real distinguished boundary for $k_1 > 1$ are

$$k_2 = -1, \quad \text{or} \quad k_4 = +1. \tag{20}$$

D. $1 \geq k_4 > -1; k_3, k_2 \leq -1$

In this case too, the Mandelstam representation is valid for $k_1 \leq 1$. When k_1 reaches 1, L_{12}^{*} converge on the real axis above 1. Again, the distinguished boundary is real if $k_2 = -1$. Fig. 6 shows how it is possible to deform the contour to obtain a real distinguished boundary if $k_3 = -1$, as well. Hence, in this case, either (19) or (20) must be satisfied if $k_1 > 1$.

E. $1 \geq k_3 > -1; k_2, k_4 \leq -1$

Again the Mandelstam representation is valid if $k_1 \leq 1$. The condition for a real distinguished boundary for $k_1 > 1$ is that k_4 be equal to -1, or that $k_3 = +1$.

F. $k_2, k_3, k_4 \leq -1$

In this case, either k_2 or k_3 or k_4 must equal -1 in order that the distinguished boundary be real for $k_1 > 1$.

Summary. This completes the discussion of all cases where at most one vertex is externally unstable. The complete result is:

If k_2, k_3 and k_4 are not larger than 1, then the

distinguished boundary is real in the following cases:

(a) $k_i^2 \leq 1$, $\theta_1 + \theta_2 + \theta_3 + \theta_4 \leq 2\pi$,
(b) $k_i \leq 1$, and at least one $k_j \leq -1$,
} "normal case,"

(c) $k_i^2 \leq 1$, and for at least one permutation (i, j, k, l) of (1234), $\theta_i + \theta_j = \theta_k + \theta_l$.

(d) $k_1 > 1$; $k_2, k_3, k_4 \leq 1$, and at least one $k_i^2 = 1$.

The next simplest case, of two externally unstable vertices, is not treated here. In addition to the fact that this has rather few applications, there are two good reasons for stopping at this point. One is that the analysis in the case of two externally unstable vertices is exceedingly long, and the other reason is that some of the special cases that are of the greatest interest are particularly simple. These will be treated in the next section.

III. SIMPLE REPRESENTATIONS

In this section we select for special study those cases where $x_\pm$ and $y_\pm$ are rational functions of y and x, respectively. This restriction greatly simplifies the algebra and guarantees that the distinguished boundary is real; fortunately it is satisfied rather frequently in the applications.

Equations (11) and (15) show that $x_\pm$ and $y_\pm$ are rational if and only if one of the following conditions is met:

A1 $k_2 = k_4 = -1$,
2 $k_2 = -1$, $k_4 = +1$,
3 $k_2 = k_4 = +1$,
B1 $k_2 = k_4$, $k_1 = k_3$,
2 $k_2 = -k_4$, $k_1 = -k_3$.

Since A1 and B1 have the most numerous practical applications, representations are worked out in detail for these cases below.

A. Case A1: $k_2 = k_4 = -1$

1. One unstable vertex. This is the case with the most interesting applications. All the true box diagrams, with each line representing a single particle, could easily be evaluated by perturbation theory, since the "vertex" functions are simply renormalized coupling constants. However, if one side (or two opposite sides) of the box is a two-particle state, then two (or all four) of the vertices are exact two-particle scattering amplitudes. It has been pointed out that in this case the singularities of the amplitude are determined by the lower limits of the effective masses of the two-particle states that are represented by the internal lines.[10] In many applications the vertices are elastic amplitudes, and then the appropriate k_i equals -1 at the lower limit of the effective mass. For this reason the case $k_2 = k_4 = -1$ has many applications to decay amplitudes and production amplitudes.

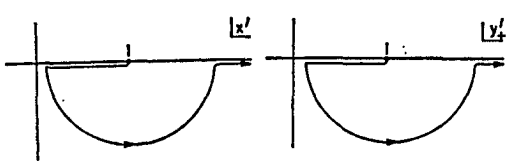

FIG. 7. Possible choice of the x' contour when $k_2 = k_4 = -1$, $k_1 > 1$, and the associated locus of the end point of the y' contour.

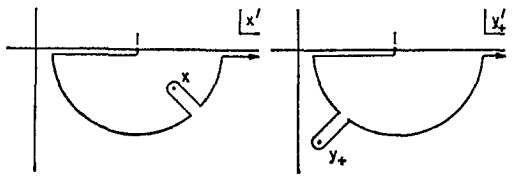

FIG. 8. Deformation of the x' contour of Fig. 7, necessary to calculate $F(x, y)$ by (2) if x has the value shown, and the corresponding locus of y_+'.

The expressions (10), (11), and (15) now simplify to

$$\Delta(x, y) = [(x - 1)(y - 1) - a][xy - b], \quad (21)$$

$$y_+ = 1 + a/(x - 1), \qquad y_- = b/x, \quad (22)$$

$$x_+ = 1 + a/(y - 1), \qquad x_- = b/y, \quad (23)$$

where

$$a = \tfrac{1}{4}(1 - k_1)(1 - k_3), \quad b = \tfrac{1}{4}(1 + k_1)(1 + k_3). \quad (24)$$

It follows from the general discussion that the Mandelstam representation is valid if k_1 and k_3 are less than 1; that is, if the external particles are stable at vertices 1 and 3. We have

$$L_{12}^+ = L_{12}^- = L_{14}^+ = L_{14}^- = \tfrac{1}{2}(1 + k_1),$$

$$L_{23}^+ = L_{23}^- = L_{34}^+ + L_{34}^- = \tfrac{1}{2}(1 + k_3),$$

both of which are less than 1 and are out of the integration region $x' \geq 1$, $y' \geq y_+ \geq 1$.

As discussed in the introduction, we are particularly interested in an analytic continuation which is valid in the product of the upper half planes. The remainder of the "physical sheet" is defined by analytic continuation in x and y. From a point in the upper half y-plane, one may pass to the lower half, either to the right of $y = 1$, or sufficiently far to the left. One possibility is to maintain the

[10] G. Barton and C. Kacser, Nuovo Cimento **21**, 988 (1961).

conventional physical cut to the right of $y = 1$ and define the lower half plane by continuation across the negative real axis. Although this gives a very unpleasant representation, we discuss this choice first and return to the other alternative below.

Rather than keeping the x' contour fixed, we may deform part of both contours near 1, so that neither is changed far to the right. A convenient choice is shown in Fig. 7. With this contour $F(x, y)$ can be evaluated so long as (x, y) lies outside the semicircles. To find $F(x, y)$ inside the semicircle in the x plane, say, the x' contour must be deformed as illustrated in Fig. 8. This produces a spike on the locus of the points y_+' and hence a complex branch point at $y = y_+$. The branch cut is conveniently made a straight line from $y = y_+$ to $y = 1$. (The other branch point, y_-, is out of the way in the upper half plane.)

When $k_1 > 1 > k_3$, the only cuts are

$$(1) \quad x > 1, \qquad (2) \quad y > 1,$$

$$(3) \quad (x-1)(y-1) = ra, \qquad 0 \leq r \leq 1.$$

It is important to remember that (3) is relevant only when both x and y have negative imaginary parts. The following set of q_x^i, q_y^i satisfies all the requirements[11]:

$$q_x^1 = (x' - x)^{-1}, \quad q_y^2 = (y' - y)^{-1}, \quad q_x^2 = q_y^1 = 0,$$

$$q_x^3 = \frac{1}{2N} \frac{(y'-1)^{\frac{1}{2}} + (y-1)^{\frac{1}{2}}}{(x'-1)^{\frac{1}{2}} + (x-1)^{\frac{1}{2}}}, \tag{25}$$

$$q_y^3 = \frac{1}{2N} \frac{(x'-1)^{\frac{1}{2}} + (x-1)^{\frac{1}{2}}}{(y'-1)^{\frac{1}{2}} + (y-1)^{\frac{1}{2}}},$$

$$N = (x'-1)^{\frac{1}{2}}(y'-1)^{\frac{1}{2}} - (x-1)^{\frac{1}{2}}(y-1)^{\frac{1}{2}}.$$

We define the square roots of unprimed variables by Im $(\ )^{\frac{1}{2}} \geq 0$; then these square roots are analytic in the domain of analyticity of $F(x, y)$. The square roots of primed variables are defined in the same way, except that the cuts are pushed up into the upper half planes; then q_x^i and q_y^i are one-valued functions of the integration variables near the distinguished boundary. The function q_x^3 occurs only in the integral over the intersection of cuts (2) and (3); here $x' < 1$, and $(x'-1)^{\frac{1}{2}}$ is positive imaginary; thus the denominator $(x'-1)^{\frac{1}{2}} + (x-1)^{\frac{1}{2}}$ does not vanish. In the region of integration the first term in N takes the values $-i(-ra)^{\frac{1}{2}}$, $0 \leq r \leq 1$, so N can vanish only when x and y are both in the lower half planes.

[11] Yu. A. Simonow [Zh. Experim. i Teor. Fiz. **43**, 2263 (1962), English translation to be published in Soviet Phys.—JETP]. has pointed out that the analyticity requirement for the q_x^i, q_y^i can be relaxed under certain special circumstances. We are grateful to Dr. Simonow for sending us a preprint of his paper.

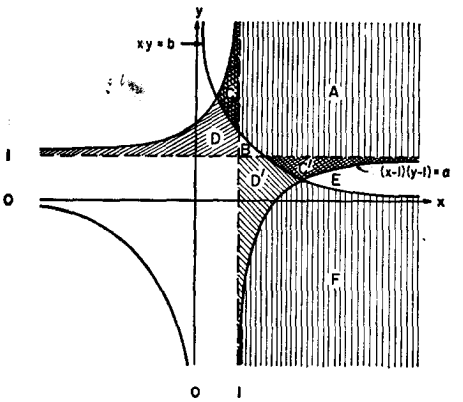

FIG. 9. The integration domains in Eqs. (27) and (28) are shown.

The distinguished boundary is

$$(1,2) \quad x > 1, \quad y > 1,$$

$$(1,3) \quad x > 1, \quad y_+ < y < 1,$$

$$(2,3) \quad y > 1, \quad x_+ < x < 1.$$

We already know that $\rho^{12} = \rho(x, y)$ is positive real for sufficiently large x and y. To find the spectral functions for smaller x or y we need only continue x and/or y past the zero of the second factor of $\Delta(x, y)$, [See Eq. (21)] and always below this zero. Hence the spectral functions get a factor $+i$ when x or y is decreased past this zero. Thus,

$$\rho^{12} = \rho(x, y) \quad \text{if} \quad xy > b,$$

$$\rho^{12} = i\rho'(x, y) \quad \text{if} \quad xy < b,$$

where

$$\rho'(x, y) = [-\Delta(x, y)]^{-\frac{1}{2}}. \tag{26}$$

The result is

$$F(x, y) = (\pi i)^{-2}\Big\{\iint_A dx'\, dy' \frac{\rho(x', y')}{(x'-x)(y'-y)} + \iint_B dx'\, dy' \frac{i\rho'(x', y')}{(x'-x)(y'-y)} + \Big[\iint_C dx'\, dy' \frac{\rho(x', y')}{y'-y} q_x^3 + \iint_D dx'\, dy' \frac{i\rho'(x', y')}{y'-y} q_x^3 + (x \leftrightarrow y)\Big]\Big\}. \tag{27}$$

The domains of integration are shown in Fig. 9. The domain "B" is empty if $b < 1$.

The potential usefulness of the above representation is seriously impeded by the complexity of the kernels.[11] Consider, therefore, the second alternative,

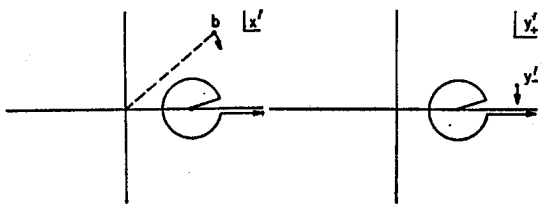

FIG. 10. With these contours, the physical cut is maintained. When b descends on the real axis and the line connecting b to the origin crosses the x' contour, then y_-' goes into the lower half plane for those values of x' that lie above this line.

of reaching the lower half of the y plane through the physical cut. Then the only cuts are $x > 1$ and $y < 1$. If we include both possibilities Im $k_1 \lesssim 0$,[12] the spectral function is $\pm(-i\rho')$ if $y < y_+$, $xy > b$; and ρ if $y < y_+$, $xy < b$. Thus the BOW representation is

$$F(x, y) = (\pi i)^{-2}\left\{\iint_F dx'\, dy' \frac{\rho(x', y')}{(x' - x)(y' - y)} \pm \iint_E dx'\, dy' \frac{-i\rho'(x', y')}{(x' - x)(y' - y)}\right\}; \quad (28)$$

the domains E and F are shown in Fig. 9. The sense of the integrations is from smaller to larger x' and y'.

There is another way of arriving at the representation (28). Consider again the Mandelstam representation (2) with ρ given by (5) and Δ and y_+ by (21) and (22). Introduce a new variable z' by

$$z' = (y' - 1)/a.$$

This has the effect of making the limits of integration in (2) independent of k_1 and k_3:

$$F(x, y) = (\pi i)^{-2} \int_1^\infty dx' \int_{(x'-1)^{-1}}^\infty dz' \times \frac{a\rho(x', z')}{(x' - x)(az' + 1 - y)}, \quad (29)$$

$$\rho(x, z) = \{[(x - 1)z - 1]a[x(az + 1) - b]\}^{-\frac{1}{2}}. \quad (30)$$

If y is in the upper half plane, then the denominator in (29) has no zero in the lower half a plane. As a is continued through the lower half plane and reaches the negative real axis, the phase of the factor $a[x(az + 1) - b]$ becomes either $-i\pi$ or $-2i\pi$. Hence $\rho(x', z')$ either becomes $-\rho(x', z')$ or $+i\rho'(x', z')$. As y' is reintroduced as an integration variable, the change of the sense of integration changes the overall sign, and (28) is obtained.

[12] If Im$k_1 < 0$, the anomalous cut lies in the product of the upper half planes, rather than the lower half planes. In this case it is the upper half of the y plane that is reached through the physical cut.

2. Triple representations. When $k_1 > 1$ it is often a kinematical variable, and it may be useful to exhibit the analytic structure of the amplitude as a function of three complex variables. The generalization of (27) involves very complex kernels, so we discuss the second alternative only. For $-1 \leq k_3 \leq 1$ and complex k_1, the cuts are

$$(1)\quad x \geq 1, \qquad (2)\quad k_1 \geq 1,$$
$$(3)\quad y - 1 = r(1 - k_1), \qquad 0 \leq r \leq \infty.$$

When k_1 varies in the complex plane, cut from $+1$ to $+\infty$, the cut (3) never catches on the branch point y_-, since y_+ and y_- lie in opposite half planes. A possible set of q_x^i, q_y^i, q_k^i is

$$\begin{aligned} &q_x^1 = (x' - x)^{-1}, \qquad q_y^1 = q_k^1 = 0, \\ &q_x^2 = q_y^2 = 0, \qquad q_k^2 = (k_1' - k_1)^{-1}, \\ &q_x^3 = 0, \quad q_y^3 = (1 - k_1)/N, \quad q_k^3 = (y - 1)/N, \\ &N = (1 - k_1)(y' - 1) - (1 - k_1')(y - 1). \end{aligned} \quad (31)$$

Equation (28) shows that the discontinuity function is $-i\rho'$ where ρ' is real, and zero elsewhere. Thus, for all complex values of k_1,

$$F(x, y, k_1) = \pi^{-3} \int_1^\infty dk_1' \int_{L_{1+}}^\infty dx' \int_{y_+'}^{y_-'} dy' \times \frac{\rho'(x', y', k_1')}{(x' - x)(k_1' - k_1)} \frac{1 - k_1}{N}. \quad (32)$$

This result can be obtained without the help of the BOW formula. The function $\rho(x', z', k_1)$, for x' and z' fixed, has two branch points in the k_1 plane, at $k_1 = 1$ and at

$$k_1 = K \equiv \frac{(1 - k_3)x'z' + 4x' - 1 - k_3}{(1 - k_3)x'z' + 1 + k_3} \geq 1.$$

The function $\rho(x', z', k_1)$ is two-valued on the cut $1 \leq k_1 \leq K$, with boundary values $\pm i\rho'(x', z', k_1)$, and satisfies the dispersion relation

$$\rho(x', z', k_1) = \pi^{-1} \int_1^K dk_1' \rho'(x', z', k_1')/(k_1' - k_1).$$

Inserting this into (29) we get

$$F(x, y, k_1) = \pi^{-3} \int_1^\infty dx' \int_{(x'-1)^{-1}}^\infty dz' \times \int_1^K dk_1' \frac{-a\rho'(x', z', k_1')}{(x' - x)(k_1' - k_1)(az' + 1 - y)},$$

which turns into (32) as the order of integrations is reversed.

3. Two unstable vertices. Let k_1 and k_3 both move into the upper half plane, while keeping the x'

integration contour fixed for large x' and the y' contour fixed for large y' (starting from the normal case, k_1, $k_3 < 1$). If the x' integration is carried out last, it starts at $x' = 1$, follows the straight line $x' - 1 = ra$, $0 \leq r$, then follows a circle to the real axis above 1 in the counterclockwise sense, and goes along the real axis to infinity. The locus of y_+' is the same in the opposite direction. This is shown in Fig. 7 for the particular case that a is about to cross the negative real axis. We would like to let the y' integration follow a straight line from y_+ to 1, and whence to infinity along the real axis. We can show (Appendix) that the branch point of $\rho(x', y')$ at $x'y' = b$ never crosses the straight line from y_+' to 1 when both k_1 and k_3 have positive imaginary parts. However, $x'y' = b$ can occur for $1 < y' < \infty$ if x' lies on the straight line connecting b with the origin, and this can happen if a lies in the upper half plane.

As k_1 and k_3 approach the real axis above 1, the x' integration and the locus of y_+' are complete circles around 1 plus the real axes to the right of the circles, and the line connecting b to the origin descends on the real x' axis from above, with $b > 1$; see Fig. 10. We now have to decide what contour to take for the y' integration.

If we try to keep the y' contour unchanged, i.e., from y_+' to 1 and from 1 to ∞ along the real axis, then the point $y_-' = b/x'$ must not be allowed to cross the real axis above 1. To accomplish this, however, the radius of the semicircle in the upper half x' plane must be shrunk to zero with the result that the semicircle in the lower half y' plane becomes infinitely large. Thus the lower half y' plane becomes inaccessible, and we conclude that the y' contour must be altered.

A possible procedure is to retain the contours of Fig. 10, letting the line from b to the origin cross the x' contour. In this case, y_-' crosses into the lower half y' plane and the y' integration contour must be depressed so as to remain below y_-'. We will not write out explicitly the BOW representation which results, but concentrate on the simpler representation described below.

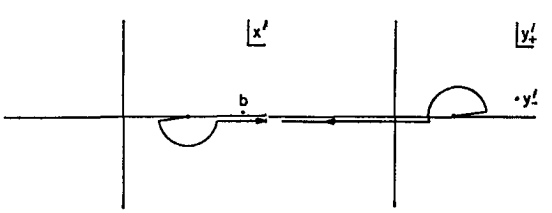

FIG. 11. With these contours, the lower half of the physical y-plane has been redefined.

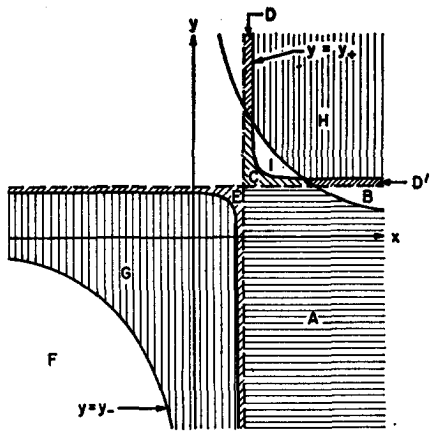

FIG. 12. Integration domains in Eqs. (33), (34), and (36).

Rather than simply depressing the y' contour to avoid y_-', a better choice is to integrate y' from y_+' to 1 and then from 1 to $-\infty$. This procedure will leave the lower half y' plane accessible as long as the x' contour is simultaneously deformed so that y_+' goes to $-\infty$ rather than $+\infty$ as x' approaches 1. A symmetrical set of contours is shown in Fig. 11. The cuts are:

(1) $x > 1$,

(2) $y < 1$,

(3) $(x - 1)(y - 1) = ra, \quad 0 \leq r \leq 1,$

the last being relevant if Im $x < 0$ and Im $y > 0$ only. The kernel functions may again be taken to be given by (25), except that now $(x - 1)^{\frac{1}{2}}$[or $(y - 1)^{\frac{1}{2}}$] should be defined with the cut to the right (or left) of 1. The cut of $(x' - 1)^{\frac{1}{2}}$[or $(y' - 1)^{\frac{1}{2}}$] should lie in the upper (or lower) half plane, so that the definitions of $(x - 1)^{\frac{1}{2}}$ and $(x' - 1)^{\frac{1}{2}}$[or $(y - 1)^{\frac{1}{2}}$ and $(y' - 1)^{\frac{1}{2}}$] agree in the lower (or upper) half planes. The distinguished boundary is

(1,2) $x > 1, \quad y < 1,$

(1,3) $x > 1, \quad 1 \leq y \leq y_+,$

(2,3) $y < 1, \quad x_+ \leq x \leq 1.$

The spectral functions are easily determined and the result is

$$F(x, y) = (\pi i)^{-2}\left\{\iint_A dx'\, dy' \frac{\rho(x', y')}{(x' - x)(y' - y)}\right.$$

$$+ \iint_B dx'\, dy' \frac{-i\rho'(x', y')}{(x' - x)(y' - y)}$$

$$\left. + \iint_C dx'\, dy' \frac{\rho(x', y')}{x' - x}\, q_v\right.$$

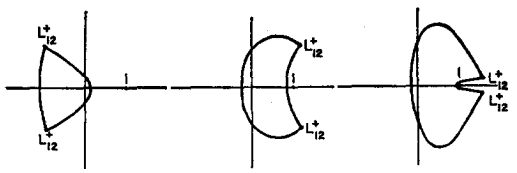

FIG. 13. If k_1 and k_2 are restricted to the upper half planes, and L_{12}^+ have one of the pair of complex values indicated, then L_{12}^- is restricted to the inside of closed domains of this approximate shape.

$$+\iint_{D+D'} dx'\,dy'\,\frac{-i\rho'(x',y')}{x'-x}\,q_v + \iint_E dx'\,dy'\,\frac{\rho(x',y')}{y'-y}\,q_x\Big\}. \tag{33}$$

The integration regions are shown in Fig. 12.

An even simpler representation is obtained if the x' contour is also deformed through the lower half plane so that it ranges from 1 to $-\infty$ along the real axis. This yields a Mandelstam-type representation:

$$F(x,y) = (\pi i)^{-2}\Big\{\iint_F dx'\,dy'\,\frac{\rho(x',y')}{(x'-x)(y'-y)} + \iint_G dx'\,dy'\,\frac{-i\rho'(x',y')}{(x'-x)(y'-y)}\Big\}, \tag{34}$$

where the integration regions are again given by Fig. 12. This result may be obtained without using the BOW representation, and without deforming the contours. Introduce new variables

$$u' = 2(y'-1)/(1-k_1), \quad v' = 2(x'-1)/(1-k_3).$$

This has the effect of making the limits of integration in (2) independent of k_1 and k_3. For $k_1 < 1$, $k_3 < 1$ we now get

$$F(x,y) = (\pi i)^{-2}\int_0^\infty dv' \int_{1/v'}^\infty du' \times \frac{a\rho(v',u')}{[\frac{1}{2}(1-k_3)v'+1-x][\frac{1}{2}(1-k_1)u'+1-y]},$$

$$\rho(v,u) = ((uv-1)a\{[\tfrac{1}{2}(1-k_3)v+1] \times [\tfrac{1}{2}(1-k_1)u+1]-b\})^{-\frac{1}{2}}.$$

If x and y are in the upper half planes, then the denominators have no singularities in the upper half k_1 or k_3 planes. As k_1 and k_3 are moved through the product of the upper half planes, we can prove that the curly bracket never crosses the positive real axis. Therefore its phase is unique. For very large u and v, the phase is the same as that of a, namely -2π, and ρ is real and positive. For smaller u and v, where $xy - b < 0$, the curly bracket has phase $-\pi$ and ρ has the phase $-\frac{1}{2}(-2\pi-\pi) = \frac{3}{2}\pi$. Now we reintroduce x' and y' and verify (34).

B. Case B1: $k_2 = k_4$, $k_1 = k_3$

This case has applications to $\pi - D$ scattering, $\pi - \rho$ scattering, etc.

Equations (21)–(23) hold in this case too, but with

$$a = \tfrac{1}{4}(k_1+k_2)^2, \qquad b = \tfrac{1}{4}(k_1-k_2)^2,$$

$$L_{12}^{\pm} = L_{34}^{\pm} = L_{14}^{\pm} = L_{23}^{\pm} = \tfrac{1}{2}\{1 - k_1k_2 \pm [(1-k_1^2)(1-k_2^2)]^{\frac{1}{2}}\}. \tag{35}$$

Without loss of generality we may take

$$k_1 \geq k_2.$$

The "normal" region is given by $k_1 \leq 0$; here $y_-' \leq y_+'$ for x' on the integration contour.

Starting in the "normal" region, k_1 and k_2 should move into the upper half planes, and then back to the real axes in "anomalous" regions. One of the main difficulties is that of making sure that the result is independent of the paths through which the k_i are moved to the final values. Let any set of paths be given, containing finite points only. Then a finite real positive number x_0 exist, such that, for all $x' > x_0$, $|y_-'| < \frac{1}{2}$ and $|y_+' - 1| < \frac{1}{2}$, throughout the continuation in k_1, k_2. Thus, for $x' > x_0$, the y' contour may run from y_+' to 1 and from 1 to $+\infty$ without interference from y_-'.

Next we establish the relationship between the x' contour and the singular points $L_{ij}^{\pm}$. If we use the representation (13), (14), defining Im $\theta_i > 0$, it is easy to see that, as k_1, k_2 vary in the upper half planes, $L_{12}^+(L_{12}^-)$ varies all over the complex plane cut along the real axis to the left (or right) of 1. Furthermore, it may be established that L_{12}^- never loops around L_{12}^+. (Fig. 13 illustrates how the values of L_{12}^- are restricted when L_{12}^+ is fixed.) Since L_{12}^- never crosses the real line between 1 and $+\infty$ it can never lie between this line and the x' contour. Since L_{12}^+ lies above the real axis in the normal region, and crosses the real axis through the cut only, the x' contour must always pass below L_{12}^+. Thus is the relationship between $L_{12}^{\pm}$ and the x' contour determined unambiguously for every k_1, k_2, independently of the paths through which k_1, k_2 had been continued through the upper half planes. In addition we may require that y_+' be infinite on the real positive axis only, which means that, as x' approaches 1, it must go around this point in the clockwise sense till the phase of $x'-1$ is that of a, and then reach 1 from this direction. Of course, arg $a = 0$ for $k_1 + k_2 < 0$, and arg $a = -2\pi$ for $k_1 + k_2 > 0$.

The y' contour has already been determined for $x' > x_0$, and can now be found for all relevant

values of x' by continuously moving x' along its contour and demanding that y_-' never cross the y' contour. Thus we have established that the path of continuation is immaterial; the final values of k_1 and k_2 uniquely determine the continuation of $F(x, y)$.

The case $k_1 + k_2 < 0$ is the simplest, and we dispose of this first. The anomalous region is

$$k_2 \leq -1, \qquad 1 \leq k_1 \leq -k_2.$$

In this case L_{12}^{*} are real and larger than 1, with $L_{12}^{+}(L_{12}^{-})$ above (below) the x' contour, which follows the real axis from 1 to $+\infty$. As x' is decreased it first encounters L_{12}^{+}. As x' passes under L_{12}^{+}, y_-' passes over y_+', and the spectral function is $i\rho'$ for $y_+' < y' < y_-'$. As x' passes over L_{12}^{-}, y_-' returns to the left of y_+', again passing over y_+', so that the y' contour does not get snagged, and the spectral function returns to a positive real value. Therefore the representation (2) is modified as follows:

$$F(x, y) = (\pi i)^{-2}\left\{\iint_H dx'\, dy' \frac{\rho(x', y')}{(x' - x)(y' - y)} + \iint_I dx'\, dy' \frac{i\rho'(x', y')}{(x' - x)(y' - y)}\right\}. \tag{36}$$

The integration regions are shown in Fig. 12.

Next consider the case

$$-1 \leq k_2 \leq +1, \qquad -k_2 \leq k_1 \leq 1. \tag{37}$$

The form of the x' contour is determined by the rules given above, with L_{12}^{-} on the real axis and L_{12}^{+} on the underside of the real axis, with $0 < L_{12}^{-} < L_{12}^{+} < 1$. It is illustrated in Fig. 14, where the loci of y_-' are also shown. The radii of the two circles may be taken equal to $\frac{1}{2}(k_1 + k_2)$. In that case, $0 \leq (1 - k_1)(1 - k_3) \leq x'y_+' - b$, so that y_-' never enters the y_+' circle, and the y' contour may run on the straight lines from y_+' to 1 to $+\infty$. In order to compute $F(x, y)$ for x or y inside the circles, however, the appropriate contour must be indented in a way that is similar to the situation shown in Fig. 9. This produces a branch point at

$$(x - 1)(y - 1) = a.$$

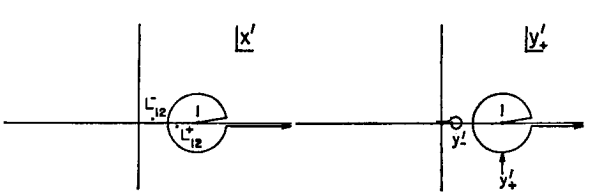

FIG. 14. Possible deformed contours for Eq. (2) in the case $k_1 = k_3$, $k_2 = k_4$, $k_2^2 < 1$, $-k_2 \leq k_1 \leq 1$.

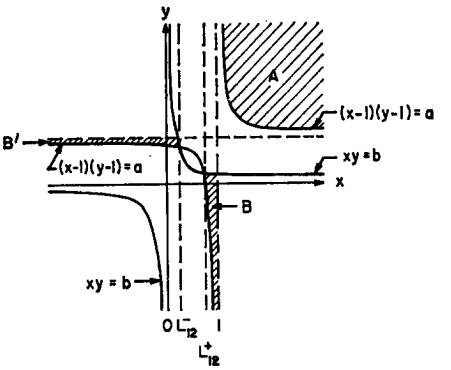

FIG. 15. Integration domain for Eq. (39).

We try to run the y' contour from y_+' to 1 on a straight line, so that the cut be given by

$$(x - 1)(y - 1) = ra, \qquad 0 \leq r \leq 1. \tag{38}$$

This is possible only if (38) is incompatible with $xy = b$. It is, in fact, easy to show that xy equals b on (38) only if x is real and either $0 < x < L_{ij}^{-}$ or $L_{ij}^{+} < x < 1$. Therefore the x' contour must avoid these portions of the real axis, which clearly means that $F(x, y)$ has a cut in x for real x between L_{ij}^{+} and 1. There is a similar cut in y, because if we try to continue in y through this line, the x' contour is forced to cross the real axis to the left of L_{ij}^{-}. Thus we see that the situation is a simple extension of the case $k_1 = k_2$ discussed in a previous paper[5]; the only difference being that the role that was played by the origin in that special case is now taken over by the point L_{ij}^{-}. A representation that is valid in the region (37) is therefore

$$F(x, y) = (\pi i)^{-2}\left\{\iint_A dx'\, dy' \frac{\rho(x', y')}{(x' - x)(y' - y)} + \left[\iint_B dx'\, dy' \frac{(x + x' - 2)\rho(x', y')}{(x' - x)N} + (x \leftrightarrow y)\right]\right\}, \tag{39}$$

where

$$N = (x' - 1)(y' - 1) - (x - 1)(y - 1),$$

and the integration regions are shown in Fig. 15.

The simplest representation valid in (37) is obtained by rotating both contours so that they run from $-\infty$ to 1. The simplest way to find the modified Mandelstam representation for that case is to introduce the variables

$$u' = -2(x' - 1)/(k_1 + k_2),$$

$$v' = -2(y' - 1)/(k_1 + k_2)$$

in (2), to obtain

$$F(x, y) = (\pi i)^{-2} \int_0^\infty du' \int_{1/u'}^\infty dv' \times \frac{a\rho(u', v')}{[-\frac{1}{2}(k_1 + k_2)u' + 1 - x][-\frac{1}{2}(k_1 + k_2)v' + 1 - y]},$$

$$\rho(u', v') = ((u'v' - 1)a\{[\tfrac{1}{2}(k_1 + k_2)u' - 1] \times [\tfrac{1}{2}(k_1 + k_2)v' - 1] - b\})^{-\frac{1}{2}}.$$

If x and y are in the upper half planes, then the denominators have no zeros in the upper half k_1, k_2 planes. When $k_1 + k_2$ is fixed, and k_1, k_2 are in the upper half planes, it may be shown by the methods of the Appendix that the possible values of b are

$$b = [\tfrac{1}{2}(k_1 + k_2) + r]^2 + s,$$

with r, s real and s positive. When this is inserted into the curly bracket,

$$\{[\tfrac{1}{2}(k_1 + k_2)u' - 1] \times [\tfrac{1}{2}(k_1 + k_2)v' - 1] - b\} = Z \quad \text{(say)}, \qquad (40)$$

and we set Im $Z = 0$, we get

$$\text{Im}\,(k_1 + k_2)\{\text{Re}\,(k_1 + k_2)(u'v' - 1) - (2r + u' + v')\} = 0.$$

If the second factor vanishes we find that the Re $Z \leq 0$. Therefore, as k_1, k_2 vary in the upper half planes, Z does not cross the positive real axis. This means that the phase of Z depends on the final values of k_1 and k_2 only, and not on the path of continuation.

In terms of x' and y', $Z = x'y' - b$. Three regions must be distinguished:

I. $x'y' - b > 0$, $x', y' < 0$, $x' < 1, y' < y'_+$,

II. $x'y' - b < 0$, ,,

III. $x'y' - b > 0$, $x', y' > 0$, ,,

Region I is the region of large u', v' and here the phase of Z is the same as that of a, namely -2π. In region II the phase of Z is $-\pi$. Region III overlaps the region of integration only in the case $k_1^2 < 1$, $k_2^2 < 1$. It may be shown that the phase of Z is 0 is this case. This result gives the phases of ρ as $+1$, $-i$ and -1 in I, II, and III, respectively. Reverting to the integration variables x' and y', we get the modified Mandelstam representation

$$F(x, y) = (\pi i)^{-2}\left\{\iint_{\text{I}} dx'\, dy' \frac{\rho(x', y')}{(x' - x)(y' - y)} + \iint_{\text{II}} dx'\, dy' \frac{-i\rho'(x', y')}{(x' - x)(y' - y)} + \iint_{\text{III}} dx'\, dy' \frac{-\rho(x', y')}{(x' - x)(y' - y)}\right\}, \qquad (41)$$

which is valid for all $k_1 + k_2 > 0$, including the cases k_1^2 and/or $k_2^2 > 1$.

IV. CONCLUDING REMARK

In all of the examples treated in the preceding section, it was found possible to obtain a Mandelstam-type representation by suitably choosing the continuation into the lower half planes. It is worth pointing out that, although the existence of such representations is not guaranteed by the reality of the distinguished boundary, they are of much more general applicability than shown here.

ACKNOWLEDGMENT

One of us (R.E.N.) would like to express his gratitude for the hospitality offered by the Physics Division of the Aspen Institute of Humanistic Studies, Aspen, Colorado, where part of his contribution to this work was performed.

APPENDIX

In this appendix we illustrate the methods that we have used to prove several statements made in the text. The following theorem was needed in Sec. III A.

Theorem. If Im $k_1 \geq 0$, Im $k_3 \geq 0$ and

$$(x - 1)(y - 1) = ra, \qquad 0 \leq r \leq 1,$$

with the restriction

$$\arg(x - 1) + \arg(y - 1) = 2\pi + \arg a,$$

then $xy \neq b$ except in the limiting cases Im $k_1 = 0$ or Im $k_3 = 0$. If we introduce

$$\alpha = \tfrac{1}{2}(1 + k_1), \qquad \beta = \tfrac{1}{2}(1 + k_3),$$

then

$$a = (\alpha - 1)(\beta - 1), \qquad b = \alpha\beta.$$

Proof. Let a be a fixed complex number. We shall determine the possible values of b, and afterwards the possible values of xy.

A. The domain of b. Our method is a simple application of that used by one of the authors in another connection.[11] Assuming that b cannot take every value in the complex plane, we look at values of b on the boundary of the allowed domain. Here there exists a complex number $N(b)$ normal to the

boundary at b, and pointing into the allowed domain. If b makes an allowed variation we have

$$\text{Re}\,\{N^*\,\text{db}\} \geq 0.$$

Now

$$db = \alpha\,d\beta + \beta\,d\alpha,$$
$$da = (\alpha - 1)\,d\beta + (\beta - 1)\,d\alpha = 0,$$

so

$$db = \{\beta - \alpha[(\beta - 1)/(\alpha - 1)]\}\,d\alpha.$$

If $\text{Im}\,\alpha \neq 0$, then $d\alpha$ is absolutely arbitrary, and no N can exist. Thus, on the boundary,

$$\text{Im}\,\alpha = 0, \qquad \text{Im}\,d\alpha \geq 0.$$

In that case, the condition on N becomes

$$N^*\{\beta - \alpha[(\beta - 1)/(\alpha - 1)]\} = -i\lambda, \qquad \lambda \geq 0.$$

Set $\alpha = r + 1$, with r real. Then

$$\beta = 1 + a/r, \qquad r\,\text{Im}\,a \geq 0,$$

and

$$b = (r + 1)(1 + a/r)$$

on the boundary. To parametrize the allowed domain note that

$$\text{Re}\,N = \lambda\,\text{Im}\,\{\beta - \alpha[(\beta - 1)/(\alpha - 1)]\}^{-1} \sim \text{Im}\,a.$$

Thus the allowed domain for b is

$$b = (r + 1)(1 + a/r) + s, \quad r\,\text{Im}\,a \geq 0, \quad s\,\text{Im}\,a \geq 0.$$

B. The domain of xy. We have

$$d(xy) = x\,dy + y\,dx,$$
$$(x - 1)\,dy + (y - 1)\,dx = a\,dr,$$

and hence, on the boundary,

$$\text{Re}\,\{N^*\,d(xy)\} = \text{Re}\,\{N^*[(x - y)/(x - 1)\,dx + ax/(x - 1)\,dr]\} \geq 0.$$

The restrictions on the variations are

$$\text{Im}\,dr = 0,$$
$$\text{Re}\,dr > 0, \quad \text{if} \quad r = 0,$$
$$\text{Re}\,dr < 0, \quad \text{if} \quad r = 1,$$
$$\text{Im}\,dx < 0, \quad \text{if} \quad \text{Im}\,x = 0, \quad \text{Re}\,x \geq 1,$$
$$\text{Im}\,dx/a > 0, \quad \text{if} \quad \text{Im}\,y = 0, \quad \text{Re}\,y \geq 1,$$
$$\text{or if} \quad y = 1, x = 1 + sa, s > 0.$$

This gives

$$\text{Re}\,\{N^*[ax/(x - 1)]\} > 0, \quad \text{if} \quad r = 0,$$
$$< 0, \quad \text{if} \quad r = 1,$$
$$= 0, \quad \text{otherwise,}$$

and

$$N^*[(x - y)/(x - 1)] = i\lambda, \quad \lambda > 0, \quad \text{if} \quad x = s + 1, \quad s > 0.$$
$$aN^*[(x - y)/(x - 1)] = -i\lambda, \quad \lambda > 0, \quad \text{if} \quad y = s + 1, \quad s > 0,$$
$$\text{or if} \quad y = 1, \quad x = 1 + sa, \quad s > 0.$$

For $\text{Im}\,a < 0$, the result is simple, and could easily have been obtained directly without the apparatus that we have developed. In that case, xy is bounded by the real axis above 1 and the line $1 + ra$, $r \geq 0$, and it may easily be seen that $xy \neq b$ except in the limiting case $a < 0$, $\text{Im}\,k_1 = \text{Im}\,k_3 = 0$. But for $\text{Im}\,a > 0$, the situation is more delicate. In this case the true boundary is found by taking the particular case $r = 1$, and either $x = s + 1$, $s > 0$ or $y = s + 1$, $s > 0$. Taking $x = s + 1$, $s > 0$ we get $\text{Im}\,a > 0$ from the two conditions

$$\text{Re}\,\{N^*[ax/(x - 1)]\} > 0, \quad N^*(x - y)/(x - 1) = i\lambda,$$

and

$$xy = (s + 1)(1 + a/s)$$

on the boundary. To determine the allowed domain, note that

$$\text{Re}\,N^* = -\lambda\,\text{Im}\,[(x - 1)/(x - y)] \sim -\text{Im}\,a.$$

We see that the boundary for b is the same as the boundary for xy, and that the inward normals have opposite directions. Therefore the domain for xy is exactly the complement of the allowed domain for b, so $xy = b$ only when both xy and b are on the boundaries; that is, when $\alpha = r = s = x - 1 > 0$. Q.E.D.

Acnodes and Cusps and the Mandelstam Representation

JAMAL N. ISLAM
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 21 December 1962)

Complex singularities on the physical sheet associated with acnodes (isolated real points) and cusps are found for a particular Feynman diagram. These singularities emerge before the appearance of anomalous thresholds, and cause a breakdown of the Mandelstam representation.

1. INTRODUCTION

MUCH attention has been directed towards examining the validity of the Mandelstam representation.[1] In this connection, perturbation theory has proved to be a valuable guide. In perturbation theory it has been shown that the Mandelstam representation ceases to be valid for sufficiently large values of the external masses.[2,3] Two kinds of behavior have so far been discovered through which the breakdown occurs, the first of which requires the existence of anomalous thresholds.[3,4] These eventually cause the boundary of the spectral region to have a crunode on it, after which complex singularities appear on the physical sheet. We shall be concerned with the second kind of behavior, in which acnodes (isolated real points) appear within the support of the spectral function. The breakdown occurs due to the fact that the complex surface attached to the acnodes is singular on the physical sheet. Two of the Feynman diagrams so far studied have been shown to display the second kind of behavior.[5,6] In these diagrams, either a pair of acnodes appear in the spectral region on the line $s = t$, or an acnode enters at infinity on the same line. We shall study here the diagram of Fig. 1, which will be shown to possess the second kind of singularity. A new feature arises in the appearance of two pairs of acnodes in the spectral region not on the line $s = t$.

The importance of studying acnode-type singularities derives from the fact that the lack of knowledge of these singularities for a general Feynman diagram makes the proof of the Mandelstam representation in perturbation theory incomplete.[5] It is hoped that by studying particular examples in detail some insight will be gained into the general problem.

The Feynman diagram of Fig. 1 is the simplest scattering one which has three branches of singular curves on the physical sheet. The Landau equations for this diagram have been studied previously by various people.[6,7,8] It has been shown[6] that, for equal external and equal internal masses, the diagram has no acnode type singularity on $s = t$ in the physical sheet for values of the external masses which exclude anomalous thresholds. It will be shown here that when the internal masses are taken to be not all equal, acnode type singularities do appear for certain values of the masses.

2. THE LANDAU EQUATIONS

We take the external masses to be M, the masses associated with the internal momenta q_1, q_2, q_3 and q_4 (see Fig. 1) to be 1 and those associated with q_5 and q_6 to be μ. Following reference 7, we derive the Landau equations[9] by imposing certain symmetry conditions on the $q_{ij} = q_i q_j$:

$$q_{16} = q_{36} = -q_{15} = -q_{35} = x,$$
$$q_{26} = q_{46} = q_{25} = q_{45} = y,$$
$$q_{23} = q_{14} = -q_{12} = -q_{34} = z.$$

The Landau surface is given by

$$s = 2 + 2\mu^2 + 8x + 2\left(\frac{x^2 - y^2}{1 - z^2} + \frac{x^2 - \mu^2 z^2}{\mu^2 - y^2}\right),$$
$$t = 2 + 2\mu^2 + 8y + 2\left(\frac{y^2 - \mu^2 z^2}{\mu^2 - x^2} + \frac{y^2 - x^2}{1 - z^2}\right), \quad (1)$$
$$u = 4 + 8z + 2\left(\frac{\mu^2 z^2 - x^2}{\mu^2 - y^2} + \frac{\mu^2 z^2 - y^2}{\mu^2 - x^2}\right),$$

[1] S. Mandelstam, Phys. Rev. **112**, 1344 (1958).
[2] S. Mandelstam, Phys. Rev. **115**, 1741 (1959).
[3] J. Tarski, Journal of Math. Phys. **1**, 149 (1960).
[4] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, Phys. Rev. **122**, 307 (1961).
[5] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. **2**, 656 (1961).
[6] D. I. Olive and J. C. Taylor, Nuovo Cimento **24**, 814 (1962).
[7] V. A. Kolkunov, L. B. Okun, and A. P. Rudik, Zh. Eksperim. i Teor. Fiz. **38**, 877 (1960) [English transl.: Soviet Phys.—JETP **11**, 634 (1960)].
[8] V. A. Kolkunov, Zh. Eksperim. i Teor. Fiz. **40**, 678 (1961) [English transl.: Soviet Phys.—JETP **13**, 2, 474 (1961)].
[9] L. D. Landau, Nucl. Phys. **13**, 181 (1959).

where x, y and z satisfy the constraints

$$M^2 = 2 + \mu^2 + 2(x + y + z), \tag{2}$$

$$\mu^2 = x^2 + y^2 + \mu^2 z^2 + 2xyz, \tag{3}$$

obtained from conservation at the vertices and the loop conditions, respectively. We also have

$$\frac{ds}{dt} = -\frac{(\mu^2 - x^2)(1 - \mu^2 + y^2 - z^2)}{(\mu^2 - y^2)(1 - \mu^2 + x^2 - z^2)}. \tag{4}$$

From symmetry between s and t, we expect some of the nodes to lie on $s = t$. Following reference 6, we obtain the nodes from solutions of $s = t$ with $x \neq y$. If such a solution yields a real value of s, it is a crunode if ds/dt is real and an acnode otherwise. From Eq. (1) we obtain, for $s = t$ with $x \neq y$ using Eq. (3),

$$2 + (x + y)\left(\frac{1}{1 - z^2} + \frac{z}{xy + \mu^2 z}\right) = 0. \tag{5}$$

Setting $\zeta = 2(z + 1)$ and $\omega = (x + y)/(1 - z)$, Eq. (5) reduces to

$$\omega\zeta(\zeta - 2) + 2(\omega^2 - \mu^2)(\omega + \zeta) = 0. \tag{6}$$

In the variables ω and ζ, Eq. (2) becomes

$$(\zeta - 4)(\omega - 1) = 4 + \mu^2 - M^2. \tag{7}$$

The value of $s = t$ at these solutions is given by

$$s = t = 2(\zeta + \mu^2 + 4\omega + \omega^2). \tag{8}$$

It is clear from Eq. (8) that the real solutions of Eqs. (6) and (7) yield real values of s and t. These solutions can, however, yield either real or complex-conjugate values of x and y. In the former case, ds/dt is real and we have a crunode; in the latter case we have an acnode. The condition for an acnode can easily be seen to be the following:

$$\zeta[\omega^2(\zeta - 4) + 4\mu^2] < 0. \tag{9}$$

The real solutions of Eqs. (6) and (7) can be found by plotting these curves for any set of values of M and μ.

In addition to the acnodes on $s = t$, there may be others not lying on this line. There seems to be no convenient method of obtaining these analytically. They can, however, be discovered by the method of searchlines (described in Sec. 4) in suitable circumstances.

3. APPEARANCE OF SINGULARITIES

We shall now study the Landau curve for certain values of M and μ, fixing our attention on the boundary of the spectral region which lies normally within the normal thresholds s, $t = 4(1 + \mu)^2$. We call this boundary the spectral curve. We note first that anomalous thresholds occur in s and t at $M^2 = 2 + 2\mu + \mu^2 = M_A^2$, and in u at $M^2 = 4 + \mu^2$. We note also that at $M^2 = 2 + 2\mu + \mu^2$, Eqs. (6) and (7) have a solution $\omega = \mu$, $\zeta = 2$ which yields a crunode at [from Eq. (8)]

$$s = t = 4(1 + \mu)^2,$$

so that at the anomalous threshold value, we always have a crunode at the intersection of the normal thresholds. As we shall see, in some cases this crunode forms a part of branches of the Landau curve other than the spectral curve and in other cases it forms a part of the spectral curve itself. This occurs due to a singular (in the geometric sense) distortion of the spectral curve, which in turn is due to acnodes appearing within it. We shall not be concerned with the behavior when M^2 exceeds the anomalous threshold value.

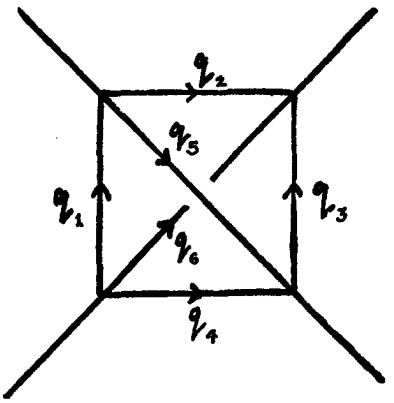

FIG. 1. The Feynman diagram under consideration.

The nodes on $s = t$ for the case $\mu = 1$ have been studied in reference 6. We mention here the relevant details. There is always one acnode present outside the normal threshold cuts and as M^2 is increased towards 5, the anomalous threshold value, two crunodes are formed on $s = t$, a little outside the threshold cuts. These are formed by the intersection of two cusps which belong to branches of the Landau curve other than the spectral curve and are quite harmless, although one of them moves into the threshold-cut region as M^2 approaches 5. For $\mu > 1$, there is no essential difference as far as the spectral curve in the threshold-cut region in s, t is concerned. As μ is decreased from 1, the same kind of behavior persists till $\mu = \frac{1}{2}(5^{\frac{1}{2}} - 1) \approx 0.618$ when (on increasing M^2 towards $2 + 2\mu + \mu^2$) the crunodes mentioned above begin to be formed *at* $s = t = 4(1 + \mu)^2$. As μ is further decreased, the crunodes

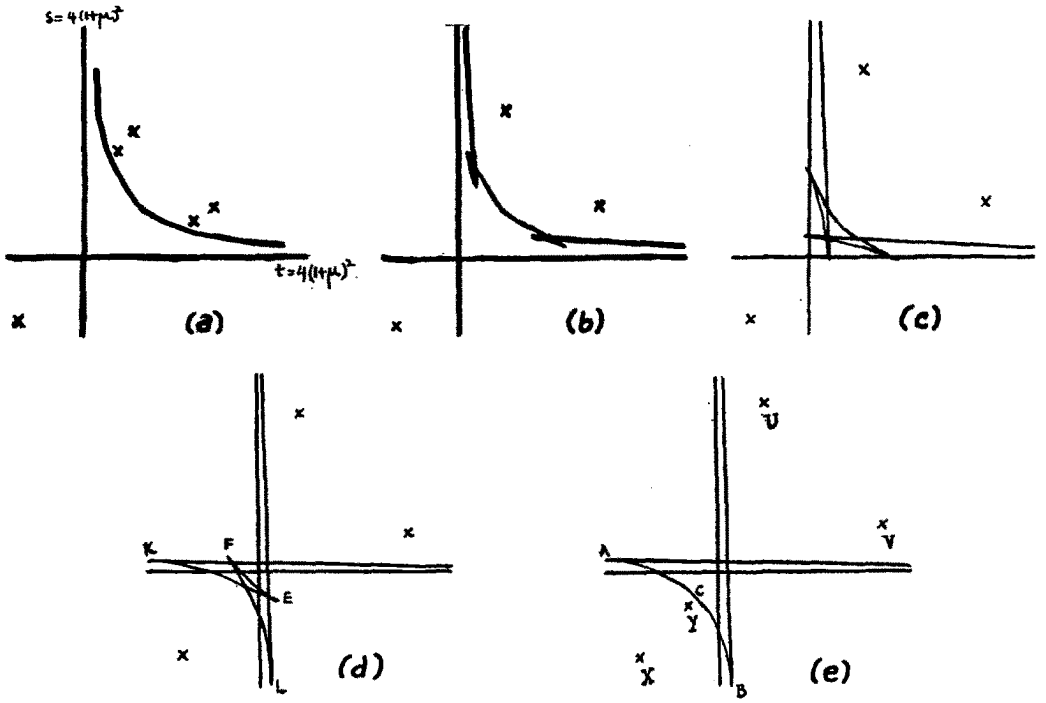

FIG. 2. Part of the Landau curve (in the real s, real t plane) associated with the diagram of Fig. 1, showing acnodes and cusps for five values of the external mass M, and for a value a little less than 0.5 of the internal mass μ.

are formed *inside* the normal threshold cuts on increasing M^2. We note that we are always considering values of M^2 less than M_A^2. For $\mu < 1$ this excludes also the anomalous threshold in u, which is at $M^2 = 4 + \mu^2$. The formation of the crunodes inside the threshold cuts occurs as follows: For $\mu < \frac{1}{2}(5^{\frac{1}{2}} - 1)$, the two cusps which formed the crunodes for $\mu > \frac{1}{2}(5^{\frac{1}{2}} - 1)$ have receded from the threshold-cut region and the crunodes are now formed by a singular distortion of the spectral curve itself. To see how this arises, we consider the case $\mu = 0.5$ in detail. We note that in this case $M_A^2 = 3.25$. The distortion of the spectral curve takes place as follows: At a little beyond $M^2 = 3.21$, two pairs of acnodes are formed within the spectral curve [Fig. 2(a)]. As we increase M^2 towards 3.25, we get successively the situations depicted in Fig. 2(b), (c), (d), and (e), i.e., two acnodes, one from each pair, hit the spectral curve producing two pairs of cusps [(b)]. These two pairs of cusps intersect, producing the crunodes [(c)]. Finally two of the cusps shrink to zero to produce an acnode [the acnode y in Fig. 2(e)]. At $M^2 = 3.25$, one crunode lies at $s = t = 4(1 + \mu)^2 = 9$ as expected, and the new acnode is just beginning to be formed. In Fig. 2(e) the arc AB corresponds to positive Feynman parameters and hence it is singular in the inappropriate limit, i.e., in $\lim_{\epsilon,\epsilon' \to 0} (s \pm i\epsilon,\ t \pm i\epsilon')$, and so also in the appropriate limit since it lies outside the threshold cuts. Thus the complex surface attached to it must also be singular on the physical sheet, giving complex

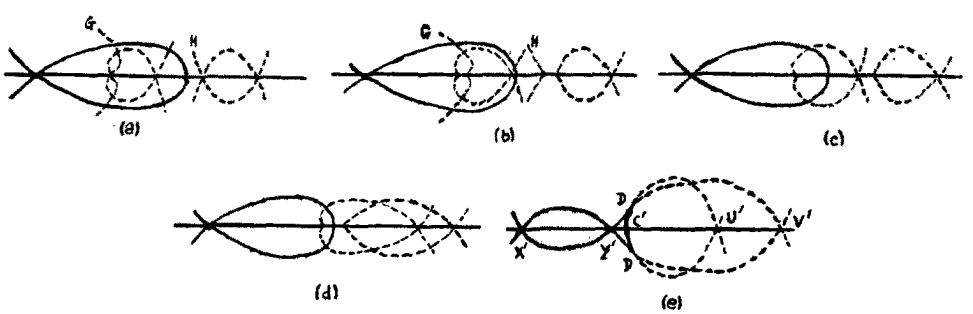

FIG. 3. The projection in the complex s (or t) plane of the five situations in Fig. 2 through the equation $s + t = \lambda$.

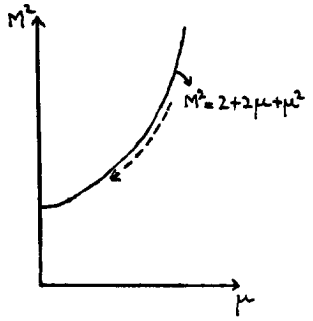

FIG. 4. One of the variations of masses considered is along the dotted path, where the solid curve represents the anomalous threshold value of M.

singularities before the appearance of anomalous thresholds. In Fig. 2(e) the singular surface is in fact the one joining the arc AB to the two acnodes U and V. At these acnodes these surfaces go on to an unphysical sheet. There is also a complex surface which joins the arc AB to the two acnodes x and y but the singularity is in fact dissolved at a complex cusp which occurs between AB and the acnode y. We see all this more clearly in the next section.

4. COMPLEX PARTS OF THE LANDAU SURFACE

We employ the method of searchlines[3,5] to study the complex surfaces attached to the relevant parts of the real Landau curves. For this we consider the intersections of the line

$$s + t = \lambda, \tag{10}$$

with the Landau surface as λ varies through real values. This enables us to obtain complex points of the Landau surface, e.g., to get the complex surface attached to the spectral curve we consider values of λ for which Eq. (10) successively intersects the spectral curve in two points, touches it and fails to intersect it, i.e., intersects it in complex points. These then give us the required surface. To get the complete complex surface we should of course consider such intersections with lines of all slopes, but the searchline (10) makes the calculation the simplest and the essential features are present in the corresponding complex cross section.

We now derive the points of intersection of Eq. (10) and the Landau surface [Eqs. (1), (2), (3)]. We get

$$4 + 4\mu^2 + 8(x + y) + 2\left(\frac{x^2 - \mu^2 z^2}{\mu^2 - y^2} + \frac{y^2 - \mu^2 z^2}{\mu^2 - x^2}\right) = \lambda.$$

Using Eq. (3), this reduces to

$$2(x + y) + \left(\frac{z(x^2 + y^2) + xy - \mu^2 z}{\mu^2 z + xy}\right) = \frac{\lambda}{4} - 1 - \mu^2 = k. \tag{11}$$

We set $x + y = p$, $xy = q$ and $a = \frac{1}{2}(M^2 - \mu^2) - 1$ so that from Eqs. (2) and (3), we get

$$z + p = a, \tag{12}$$

and

$$q = [p^2 + \mu^2(a - p)^2 - 1]/2(p + 1 - a), \tag{13}$$

respectively. Substituting in Eq. (11) for z and $xy = q$ from Eqs. (12) and (13), we obtain, after some reduction,

$$\begin{aligned} &p^4 - (1 + 2a)p^3 + p^2 \\ &\times [\tfrac{1}{2}k(1 - \mu^2) + \tfrac{1}{2}(\mu^2 + 2a\mu^2 + 2a^2 - 1)] \\ &+ p[k\mu^2(a - 1) + \mu^2(1 + a - 2a^2)] \\ &- \mu^2\tfrac{1}{2}k(a - 1)^2 + \tfrac{1}{2}\mu^2(2a^3 - 3a^2 + 1) = 0. \end{aligned} \tag{14}$$

The points of intersection for any value of λ, i.e., for any k, are thus obtained by solving the quartic (14) for this k, evaluating the corresponding q from Eq. (13), and x, y from

$$x, y = \tfrac{1}{2}p \pm \tfrac{1}{2}(p^2 - 4q)^{\frac{1}{2}},$$

and substituting in (1). We note that real solutions of p and q yield either real or complex-conjugate

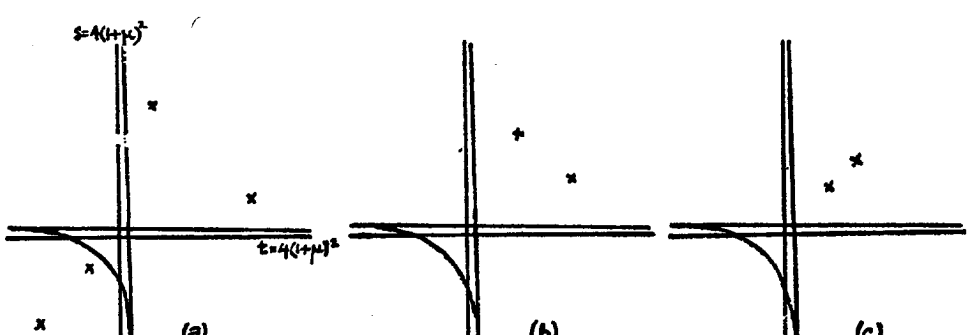

FIG. 5. The changes in the acnodes near the spectral curve when the masses are decreased from $\mu = 0.5$ along the dotted path in Fig. 4.

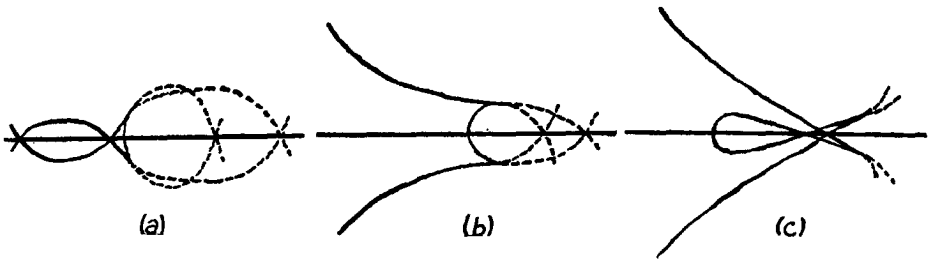

FIG. 6. The projections in the complex s (or t) plane corresponding to the three situations in Fig. 5, in the manner of Fig. 3.

values of x and y. These give parts of the complex surface for which s and t are complex conjugates. This part can be obtained more easily from Eq. (1) as it corresponds to real values of z. The acnodes on the line $s = t$ lie on this part of the surface. It is clear that acnodes which lie on the complex cross sections and which do not necessarily lie on $s = t$ are also revealed by the above analysis.

In Fig. 3 are drawn the five complex cross sections in the complex s (or t) plane corresponding to the five situations of Fig. 2. The solid lines indicate portions of the surface for which s and t are complex conjugates. The acnodes lying on $s = t$ lie on this part of the surface and those not lying on $s = t$ lie on the dotted parts. We note, e.g., in Fig. 3(b), which we recall is drawn in the complex s plane, the points of the surface with Im $s = 0$ give the values of s in Fig. 2(b) at which the acnodes and the cusps occur. In Figs. 3(a) and (b) the portions G and H of the surface are the ones that join the acnodes and cusps to the other parts of the real Landau curve and are omitted in Figs. 3(c), (d), and (e), as they are irrelevant for our discussion. Each diagram in Fig. 3 should be considered in conjunction with a similar one in the complex t plane, e.g. in Fig 3(e), if we follow a path along one of the upper two dotted curves in the s plane, the corresponding t goes along one of the lower curves in the t plane. The two surfaces that join C′ to U′ and V′ have, in fact, a symmetrical configuration in the s, t complex space.

We now study the singularities in the situation corresponding to Figs. 2(e) and 3(e). The points X′, Y′, U′, and V′ in Fig. 3(e) correspond to the acnodes of Fig. 2(e) and C' corresponds to the midpoint C of the arc AB. In this case, the surfaces joining C' to U′ and V′ are singular on the physical sheet since they are attached to the arc AB which corresponds to positive Feynman parameters and lies outside the threshold cuts. As mentioned earlier, these two surfaces have a symmetrical configuration in the complex s, t space. We note from Fig. 3(e) that C′ is also attached to the two acnodes X′ and Y′ outside the threshold cuts by a complex surface, but in going from C′ to Y′ we have to go through a complex cusp D at which the singularity is dissolved. The manner in which the singularity is dissolved is discussed in detail in reference 5 and is essentially a change in the pairing of zeros of the denominator of the Feynman integral at a double point (such as a cusp) of the Landau surface. We thus conclude that the surface joining X and Y is not singular on the physical sheet. This is to be expected since the surface has not grown from zero size.

We note that the surfaces which join D to U′ and V′ in Fig. 3(e) have grown successively from the ones that join the real cusps to the same two

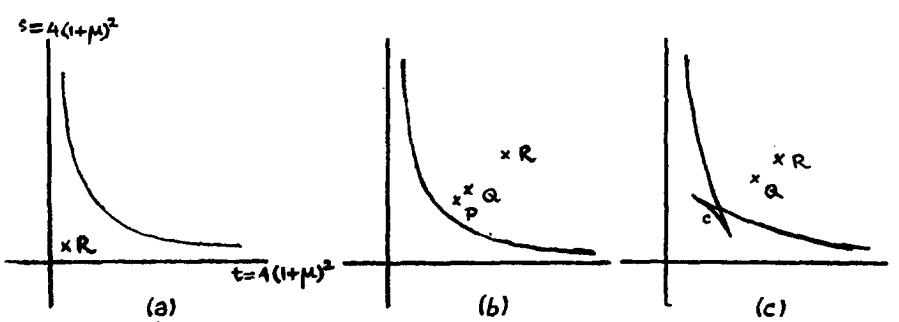

FIG. 7. The changes in the acnodes and the spectral curve for $\mu = 0.1$ as M is increased towards the anomalous threshold value.

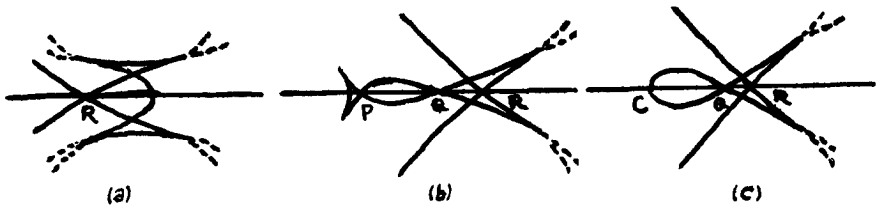

FIG. 8. The projections in the complex s (or t) plane corresponding to three situations in Fig. 7 in the manner of Fig. 3.

acnodes in Figs. 3(d), (c), and (b), and those that join the pairs of acnodes in Fig. 3(a). These are thus also singular on the physical sheet. We conclude that complex singularities appear as soon as the two pairs of acnodes are formed, which occurs at a little after $M^2 = 3.21$.

An apparent paradox might arise from the fact that in Fig. 2(d), the arc EF corresponds to positive Feynman parameters and lies outside the normal threshold cuts which would imply that the surface joining this arc to the acnode outside is singular, This is impossible since this surface has grown into the one joining the acnodes X and Y in Fig. 2(e), which we know to be nonsingular. The resolution of the paradox lies in the fact that the arc EF is reached through cuts attached to the singular surfaces joined to the arcs EK and FL , and so the surface attached to EF is not singular on the physical sheet, since EF is not singular in its appropriate limit, i.e., in $\lim_{\epsilon,\epsilon'\to 0} (s \pm i\epsilon, t \mp i\epsilon')$.

As we decrease μ, going along the dotted path in Fig. 4, i.e., keeping M^2 close to but less than the anomalous threshold value, the two acnodes X and Y approach each other until they meet and disappear. Decreasing μ further causes the two acnodes within the spectral curve to approach and meet on the line $s = t$ and separate again along $s = t$. These three situations, together with the complex cross sections, are depicted in Figs. 5 and 6, the situation (a) in these being the same as (e) in Figs. 2 and 3. It is clear from the complex cross sections how the acnodes outside disappear and how the ones inside meet on $s = t$ and separate along it; e.g., in Fig. 6(c), the acnodes now lie on the solid part of the curve and are therefore on $s = t$ in Fig. 5(c).

The situation in Fig. 5(c) arises for $\mu \approx 0.1$. It is instructive to see how the singularities arise for this value of μ as M^2 is increased. For $M^2 \ll 2.21$ (the anomalous threshold value), we have just one acnode (R in Fig. 7) outside on $s = t$. On increasing M^2, this acnode walks into the region within the spectral curve. This does not cause any distortion of the spectral curve since the acnode is joined to it through complex cusps in the manner of Fig. 8(a). On increasing M^2 further, a pair of acnodes are formed within the spectral curve,[10] one of which hits the curve to yield a pair of cusps. These situations are depected in Figs. 7 (a), (b), and (c). Again it is clear how the acnodes arise by looking at the corresponding complex cross sections in Fig. 8. In the situation of Figs. 7(b) and 8(b), the surface joining the acnodes P and Q which becomes the surface joining C and Q in 7(c) and 8(c), is singular on the physical sheet. In 7(b), 8(b) the singularity is dissolved at the complex cusps near C. These turn into real cusps in 7(c), 8(c). The singular surface in this last situation leaves the physical sheet at the acnode Q. We thus get a picture consistent with the case $\mu = 0.5$, The same kind of behavior persists as μ approaches zero.

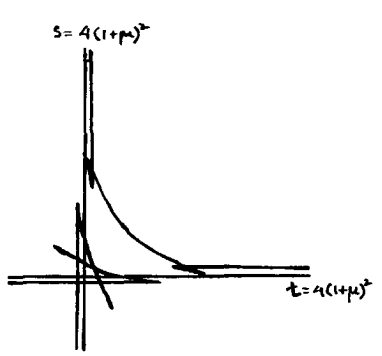

FIG. 9. Behavior of the spectral curve and other relevant parts of the Landau curve for $\frac{1}{2}(5^{\frac{1}{2}} - 1) < \mu < 1$, and for M less than but near the anomalous threshold value.

5. BEHAVIOR NEAR $\mu = 1$

We recall that for $\mu > \frac{1}{2}(5^{\frac{1}{2}} - 1)$, the pair of crunodes on $s = t$ are formed outside the normal threshold cuts on increasing M towards M_A. It turns out that the spectral curve suffers a singular distortion as before as depicted in Fig. 2(b), but cusps attached to it do not now intersect. Instead as M is increased towards M_A, two cusps belonging to other parts of the Landau curve come up to meet two of the cusps attached to the spectral curve in the manner of Fig. 9. The pairs of cusps on the

[10] This is in fact the same phenomenon as the formation of a pair of crunodes inside the threshold cuts but outside the spectral curve for $\mu = 0.5$. As μ is decreased, after a certain value we get a pair of acnodes forming inside the spectral curve, instead of a pair of crunodes outside.

spectral curve remain finite as M approaches M_A. At $M = M_A$ there is a degeneracy of a part of the Landau curve into a pair of straight lines. As is evident from Fig. 9, the whole of the spectral curve does not contribute to the straight lines. The formation of the pair of cusps seems to be the mechanism through which only a part of the spectral curve reduces to straight lines. This behavior continues as μ is increased towards 1, the cusps being formed further away and for values of M nearer M_A. As μ approaches 1, the value of M at which the cusps are formed tends to M_A and the positions of the cusps move towards infinity, i.e., for each of the pairs of cusps one of the variables s, t approaches infinity. We note that for all values of $\mu < 1$ there are complex singularities before the appearance of anomalous thresholds. The same kind of phenomena as described above occur in the spectral curves associated with the other two channels for $\mu > 1$.

We remark in conclusion that the appearance of acnodes and cusps associated with the spectral curve before the appearance of anomalous thresholds may be quite a general feature of Feynman diagrams symmetric in the two relevant variables. In such a case, anomalous thresholds in the two variables appear simultaneously, which causes a part of the Landau curve to degenerate into a pair of straight lines along the normal threshold, parts of which are supplied by the spectral curve. Of the cases studied so far, only in the fourth-order case does the whole of the spectral curve contribute to the straight lines, and in this case there are no acnodes or cusps. In the other cases there is a part of the spectral curve which does not reduce to straight lines at the anomalous threshold value. This seems to be achieved by the formation of cusps on the spectral curve, which in turn is due to the appearance of acnodes within it.

ACKNOWLEDGMENTS

I am grateful to Dr. J. C. Taylor for constant help and encouragement during the course of this work. Many grateful thanks are due to Mrs. Mutch, who performed the computations for this work on EDSAC II. I would also like to thank Cambridge University for a research maintenance grant.

Singularities of the Four-Point Function in Perturbation Theory with Variable Masses

FRANCIS R. HALPERN
University of California, San Diego, La Jolla, California
(Received 14 December 1962)

The Landau surface for the four-point function with one or more variable external masses is parameterized. The real singularities of the four-point function with one external mass are explicitly determined. The location of the complex singularities of the function depend on the choice of the cuts for the three-point function. Several suggestions are made as to how these could be taken.

I. INTRODUCTION

THE analytic properties of the Feynman amplitude arising from the box diagram (four-point function) were first studied as a function of energy and momentum transfer by Mandelstam[1] and Karplus, Sommerfield, and Wichmann.[2] Tarski gave a complete discussion of the analytic properties of the four point function as a function of two complex variables.[3] A. C. T. Wu has treated some aspects of the four-point function in six complex variables.[4] For a discussion of the scattering problem, it is sufficient to study the four-point function in only the energy (s) and momentum-transfer (t) variables. However, the four-point function as a function of s, t, and a mass variable m appears as a contracted singularity of the five-point function, with more masses variable as a contracted singularity of the higher functions. Thus to study the analytic properties of these functions, it is necessary to treat the four-point function in 2, 3, 4, 5, and 6 variables. (In the case of four variables there are two distinct choices of four variables that may occur.) Since the determination of the intersection of the manifold on which one or more of these variables is constant with the singularity manifold is not a simple problem, a summary is given in this paper of some properties for the singularity surface in each of these cases. Most of the effort is applied to the case of s, t and a single mass variable since this is relevant to the five-point function and may have some immediate application to explaining some of the more dubious resonances.[5]

The general method employed is to parameterize the Landau surface using the integration variables of the parametric integrals as parameters, and then applying previously given criteria to distinguish those portions that are singular on the physical sheet.[6] There is one problem in this procedure. The singularity region of the four-point function is *not uniquely* specified but depends on how the cuts for the three-point function are taken. The same difficulty would exist for the three-point function, but the "correct" location of the two-point cuts is very strongly indicated by physical considerations. They are taken in the physical region so that the discontinuity across the two-point cuts may be evaluated in terms of physical amplitudes. For the three-point and higher functions, this is no longer possible because the singularities no longer occur only at the edge of the physical region, but they occur for complex values of the variable and at points arbitrarily remote from the physical region. In this work, several choices for three-point cuts are considered none of which is very compelling.

II. PARAMETERIZATION OF THE LANDAU SURFACE

The four-point function F is defined by the integral

$$F = \int_0^1 du_1 du_2 du_3 du_4 \, \delta(1 - u_1 - u_2 - u_3 - u_4)/D^2,$$

where

$$D = \sum x_{ij} u_i u_j,$$

and $x_{ii} = 1$ and $x_{ij} = x_{ji}$. The energy, momentum transfer, and masses are related to the x_{ij} by

$$p_{ij}^2 = m_i^2 + m_j^2 - 2m_i m_j x_{ij}.$$

The m_i are the four internal masses, p_{ij}^2 is the mass of the external particle attached at the vertex where the lines with masses m_i and m_j join, p_{13}^2 is s, and p_{24}^2 is t.[3] The properties of F will be studied as

[1] S. Mandelstam, Phys. Rev. **112**, 1344 (1958).
[2] R. Karplus, C. Sommerfield, and E. Wichmann, Phys. Rev. **114**, 376 (1959).
[3] J. Tarski, J. Math. Phys. **1**, 154 (1960).
[4] A. C. T. Wu, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd **33**, 3 (1961).
[5] Peaks in cross sections that are dependent on the initial energy s may be the result of singularities in the n-point functions for $n > 2$. This is a possible explanation for ζ, *ABC*, and f_0. See in this connection, S. Treiman and P. Landshoff, Phys. Rev. **127**, 649 (1962).
[6] F. R. Halpern, J. Math. Phys. **3**, 1130 (1962); Phys. Rev. **127**, 1819 (1962).

functions of x_{13}, x_{24} and one or more of the remaining variables.

The singularities of F are known to lie on the variety $\det(x_{ij}) = 0$. It is useful to introduce a parameterization of this set. This may be done conveniently by observing that if $\det(x_{ij}) = 0$, then there must exist a vector u_i such that $x_{ij}u_j = 0$; that is, the matrix x_{ij} has a characteristic vector u_i belonging to the eigenvalue 0. It is simplest to treat the case of three and four variables separately from the case of five and six variables. Thus we first consider x_{13}, x_{24}, x_{12} and x_{23} to be variables and x_{14} and x_{34} to be fixed masses. The equation $x_{ij}u_i = 0$ may be rewritten as

$$\begin{bmatrix} u_2 & u_3 & 0 & 0 \\ u_1 & 0 & u_3 & u_4 \\ 0 & u_1 & u_2 & 0 \\ 0 & 0 & 0 & u_2 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{13} \\ x_{23} \\ x_{24} \end{bmatrix} = - \begin{bmatrix} u_1 + u_4\, x_{14} \\ u_2 \\ u_3 + u_4\, x_{34} \\ u_4 + u_1\, x_{14} + u_3\, x_{34} \end{bmatrix}. \tag{1}$$

This has the solution

$$\begin{aligned} x_{12} &= (-u_1^2 - u_2^2 + u_3^2 + u_4^2 + 2u_3u_4x_{34})/2u_1u_2, \\ x_{13} &= (-u_1^2 + u_2^2 - u_3^2 - u_4^2 - 2u_1u_4x_{14} \\ &\qquad - 2u_3u_4x_{34})/2u_1u_3, \\ x_{23} &= (u_1^2 - u_2^2 - u_3^2 + u_4^2 + 2u_1u_4x_{14})/2u_2u_3, \\ x_{24} &= -(u_4 + u_1x_{14} + u_3x_{34})/u_2, \end{aligned} \tag{2}$$

and is the required parametric representation of the singularity surface. To consider one of the mass variables x_{23} as fixed, it is necessary to impose a constraint on the parameters. That is, if x_{23} is to be fixed, the following relation must hold:

$$u_1^2 + 2u_1u_4x_{14} + u_4^2 = u_2^2 + 2u_2u_3x_{23} + u_3^2.$$

This may be achieved by replacing the u's by parameters A, θ, and φ in the following manner:

$$\begin{aligned} u_1 &= A[\sin\theta(1 + x_{14})^{-\frac{1}{2}} + \cos\theta(1 - x_{14})^{-\frac{1}{2}}], \\ u_2 &= A[\sin\varphi(1 + x_{23})^{-\frac{1}{2}} + \cos\varphi(1 - x_{23})^{-\frac{1}{2}}], \\ u_3 &= A[\sin\varphi(1 + x_{23})^{-\frac{1}{2}} - \cos\varphi(1 - x_{23})^{-\frac{1}{2}}], \\ u_4 &= A[\sin\theta(1 + x_{14})^{-\frac{1}{2}} - \cos\theta(1 - x_{14})^{-\frac{1}{2}}]. \end{aligned} \tag{3}$$

The form of the radicals suggest the trigonometric representation

$$x_{14} = \cos\alpha, \qquad x_{23} = \cos\beta, \tag{4}$$

with real α and β such that $0 \leq \alpha \leq \pi$ and $0 \leq \beta \leq \pi$, for the fixed masses x_{14} and x_{23}, and these will be used. If these results for the u's are substituted into the expressions for the variable x's, these are given by

$$\begin{aligned} x_{12} &= [2x_{34}\cos(\theta + \tfrac{1}{2}\alpha) \\ &\quad \times \cos(\varphi + \tfrac{1}{2}\beta) - \sin\beta\sin 2\theta - \sin\alpha \\ &\quad \times \sin 2\varphi]/[2\cos(\theta - \tfrac{1}{2}\alpha)\cos(\varphi - \tfrac{1}{2}\beta)], \\ x_{13} &= [\sin\alpha\sin(\tfrac{1}{2}\beta - \varphi) + x_{34} \\ &\quad \times \cos(\tfrac{1}{2}\alpha + \theta)]/\cos(\theta - \tfrac{1}{2}\alpha), \\ x_{24} &= [\sin\beta\sin(\tfrac{1}{2}\alpha - \theta) + x_{34} \\ &\quad \times \cos(\tfrac{1}{2}\beta + \varphi)]/\cos(\varphi - \tfrac{1}{2}\beta). \end{aligned} \tag{5}$$

The domain of variation of the parameters θ and φ will be discussed later. There is another possibility with two masses variable. Only the case of two adjacent mass variables x_{12} and x_{23} has been treated. The case of opposite variables x_{12} and x_{34} is more complicated since the matrix of u's on the left-hand side of the equation analogous to (1) is singular. This imposes a nonlinear constraint between the u's and fixed x's.

In the case of five or six variables, the same general procedure may be used, although there are modifications since there are now more unknown x's than equations. For six variables, the equations analogous to (1) are

$$\begin{bmatrix} u_2 & u_3 & u_4 & 0 & 0 & 0 \\ u_1 & 0 & 0 & u_3 & u_4 & 0 \\ 0 & u_1 & 0 & u_2 & 0 & u_4 \\ 0 & 0 & u_1 & 0 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{13} \\ x_{14} \\ x_{23} \\ x_{24} \\ x_{34} \end{bmatrix} = - \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}. \tag{6}$$

A solution to this system of equations is

$$\begin{aligned} x_{12} &= [(a - b) - 2u_1^2 - 2u_2^2 + u_3^2 - u_4^2]/6u_1u_2, \\ x_{13} &= [(b - c) - 2u_1^2 + u_2^2 - 2u_3^2 + u_4^2]/6u_1u_3, \\ x_{14} &= [(c - a) - 2u_1^2 + u_2^2 + u_3^2 - 2u_4^2]/6u_1u_4, \\ x_{23} &= [(c - a) - u_1^2 - 2u_2^2 - 2u_3^2 - u_4^2]/6u_2u_3, \\ x_{24} &= [(b - c) + u_1^2 - 2u_2^2 + u_3^2 - 2u_4^2]/6u_2u_4, \\ x_{34} &= [(a - b) + u_1^2 + u_2^2 - 2u_3^2 - 2u_4^2]/6u_3u_4, \end{aligned} \tag{7}$$

where $a - b$, $b - c$, and $c - a = -(b - c) - (a - b)$ are two arbitrary parameters in addition of the u's to give the most general solution of $\det x_{ij} = 0$. For the case of 5 variables, one of these additional parameters may be fixed. For example, to make $x_{34} = \xi_{34}$ constant, put $a - b = 6u_3u_4\xi_{34} - u_1^2 - u_2^2 + 2u_3^2 + 2u_4^2$ in all the preceding equations.

III. THE THREE-POINT MINORS ON THE LANDAU SURFACE

In order to apply the criteria it is necessary to calculate the value of the 3×3 principal minors L_i of $\det(x_{ij})$. The minors L_i arise by crossing out the ith row and ith column of x_{ij}. It has been shown that if the Landau surface has a parameterization of the type exhibited above, then the minors L_i have the form $L_i = f_i^2 g$.[6] For the case of three variables, substitution of the (5) in the expressions for L_i yields

$$\begin{aligned} g^{(3)} &= \{-x_{34}^2 + 2x_{34}\sin(\tfrac{1}{2}\alpha - \theta)\sin(\varphi - \tfrac{1}{2}\beta) \\ &\quad + \cos^2(\tfrac{1}{2}\alpha - \theta) - \sin^2(\varphi - \tfrac{1}{2}\beta)\}, \\ f_1^{(3)} &= \sin\beta/\cos(\varphi - \tfrac{1}{2}\beta), \\ f_2^{(3)} &= \sin\alpha/\cos(\theta - \tfrac{1}{2}\alpha), \\ f_3^{(3)} &= \cos(\varphi + \tfrac{1}{2}\beta)\sin\alpha/\cos(\theta - \tfrac{1}{2}\alpha) \\ &\quad \times \cos(\varphi - \tfrac{1}{2}\beta), \qquad (8) \\ f_4^{(3)} &= \cos(\theta + \tfrac{1}{2}\alpha)\sin\beta/\cos(\theta - \tfrac{1}{2}\alpha) \\ &\quad \times \cos(\varphi - \tfrac{1}{2}\beta). \end{aligned}$$

A similar set of expressions in the case of four variables is

$$\begin{aligned} g^{(4)} &= -u_1^4 - u_2^4 - u_3^4 - u_4^4 + 2u_1^2u_2^2 + 2u_1^2u_3^2 \\ &\quad - 2u_1^2u_4^2 + 2u_2^2u_3^2 + 2u_2^2u_4^2 - 2u_3^2u_4^2 - 4(u_1u_4x_{14} \\ &\quad + u_3u_4x_{34} + u_1u_3x_{13}x_{34})(u_1^2 - u_2^2 + u_3^2 + u_4^2) \\ &\quad - 4u_1^2x_{14}^2(u_3^2 + u_4^2) - 4u_3^2x_{34}^2(u_1^2 + u_4^2) \\ &\quad - 8u_1u_3u_4x_{14}x_{34}(u_1x_{14} + u_3x_{34}), \\ f_1^{(4)} &= 1/2u_2u_3, \\ f_2^{(4)} &= 1/2u_1u_3, \qquad (9) \\ f_3^{(4)} &= 1/2u_1u_2, \\ f_4^{(4)} &= u_4/2u_1u_2u_3. \end{aligned}$$

When all six x's are variables, the appropriate expressions are

$$\begin{aligned} 108u_1^2u_2^2u_3^2u_4^2g^{(6)} &= \sum_i u_i^6 - 6\sum_{i\neq j} u_i^4u_j^2 \\ &\quad + 33\sum_{i\neq j\neq k} u_i^2u_j^2u_k^2 + 3(a-b)[-2u_1^2u_2^2 + u_1^2u_3^2 \\ &\quad + u_1^2u_4^2 + u_2^2u_3^2 + u_2^2u_4^2 - 2u_3^2u_4^2] + 3(b-c)[u_1^2u_2^2 \\ &\quad - 2u_1^2u_3^2 + u_1^2u_4^2 + u_2^2u_3^2 - 2u_2^2u_4^2 + u_3^2u_4^2] \\ &\quad + 3(c-a)[u_1^2u_2^2 + u_1^2u_3^2 - 2u_1^2u_4^2 - 2u_2^2u_3^2 \\ &\quad + u_2^2u_4^2 + u_3^2u_4^2] + (u_1^2 + u_2^2 + u_3^2 + u_4^2) \\ &\quad \times (-a^2 - b^2 - c^2 + ab + bc + ca) \\ &\quad + (a-b)(b-c)(c-a), \qquad (10) \\ f_i^{(6)} &= u_i. \end{aligned}$$

The case of five variables may be obtained again by fixing one of the parameters. [The f's are unchanged and $(a - b)$ in g is replaced.]

TABLE I. The column heading (ij) gives the equation in question, and the row label indicates the case under consideration (3, 4, 5, or 6 variables).

	(12)	(13)	(14)	(23)	(24)	(34)
3	+	−	−	−	−	+
4	+	+	+	+	+	+
5	+	+	+	+	+	+
6	+	+	+	+	+	+

TABLE II. The signs of $(L_i)^{\frac{1}{2}} = \pm f_i g^{\frac{1}{2}}$. The column heading indicates the index i, and the row label indicates the case.

	1	2	3	4
3	+	+	−	−
4	+	+	+	+
5	+	+	+	+
6	+	+	+	+

IV. APPLICATION OF THE SINGULARITY CRITERION

The condition $\det(x_{ij}) = 0$ is a quadratic equation in each of the six x's with coefficients dependent on the other five x's. Each of these six quadratic equations has two roots. At any particular point on the Landau surface only one of these two roots is relevant. It may be shown that the discriminant Δ_{ij} of the equation for x_{ij} is L_iL_j.[3] Since $L_i = f_i^2 g$, it follows that

$$(\Delta_{ij})^{\frac{1}{2}} = (f_i^2 g f_j^2 g)^{\frac{1}{2}} = \pm f_i f_j g,$$

and in order to apply the criterion, it is necessary to determine which sign is applicable at each point of the Landau surface. It is straightforward to substitute the parameterizations back into the determinant and decide which is the proper choice. The correct sign for each of the cases under consideration is indicated in Table I.

Since $\Delta_{ij} = L_iL_j$, one might wonder whether an appropriate choice of $(L_i)^{\frac{1}{2}}$ is possible so that $(\Delta_{ij})^{\frac{1}{2}}$ occurs with the sign given in the table. That this question may be answered in the affirmative has been proven in general.[6] In the cases under consideration, $(L_i)^{\frac{1}{2}} = \pm f_i g^{\frac{1}{2}}$, and signs chosen in accordance with Table II reproduce all the results of Table I. The square root of g must be taken the same way wherever it occurs.

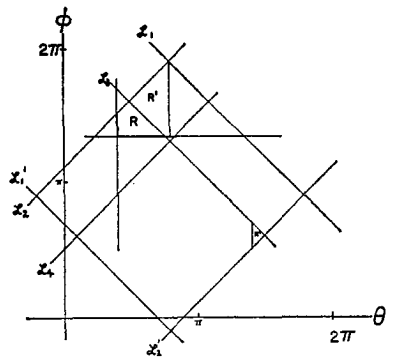

FIG. 1. The square in the real θ, φ plane over which the parameters vary. The region R is singular in all limits on the physical sheet. The regions R' adjacent to R are singular in some limits, but not in others.

The criterion that is now employed states that if the roots $(L_i)^{\frac{1}{2}}$ are all on the same sheet, the portion of the Landau surface under consideration corresponds to a physical-sheet singularity and otherwise to a singularity on some other sheet. In order to apply this criterion it is necessary to give a definition to the sheets of $(L_i)^{\frac{1}{2}}$. When $L_i = 0$ there is a three-point singularity and the cuts of $(L_i)^{\frac{1}{2}}$ are thus the cuts of the three-point function. The question thus arises as to how these cuts shall be taken. Before attempting to treat this question in general, it is convenient to discuss the situation for real values of the variables x_{ij}. The different regions in this case are separated not by cuts of the three-point function, but the actual three-point singularities.

If the x_{ij} are real, all the components u_k of the eigenvector must have the same phase or else the vector could be separated into linearly independent real and imaginary parts, each of which would be an eigenvector belonging to eigenvalue zero. The determinant would then have rank two, which occurs on a subset of the points of the variety $\det x_{ij} = 0$. Real values of the x occur only for real u's.

Consider now the range of variability of the parameter θ and φ in the case; x_{12}, x_{13}, and x_{24} are the variables and are restricted to be real. According to the formulas (3) for the u's, it might appear that a range of 2π is required for both φ and θ. However, it is clear from (5) that increasing φ and θ each by π leaves the x's invariant. This invariance permits the 2π by 2π square in the φ, θ plane to be reduced to a square having diagonals of length 2π parallel to the φ and θ axes, and sides of length $\sqrt{2}\pi$ at 45° to these axes. Again, by use of the invariance of the x's under the transformation $\varphi \to \varphi + \pi$ $\theta \to \theta + \pi$, it is easy to see that the rest of the 2π by 2π square is equivalent to some portion of this smaller region. Finally, using this same transformation, the smaller square may be translated parallel to itself into the following position; the edges of the fundamental square are chosen as the following lines:

$$\mathfrak{L}_1: \quad \varphi + \theta = \tfrac{1}{2}(\alpha + \beta) - \gamma + \pi,$$

$$\mathfrak{L}_1': \quad \varphi + \theta = \tfrac{1}{2}(\alpha + \beta) - \gamma - \pi,$$

$$\mathfrak{L}_2: \quad \varphi - \theta = \tfrac{1}{2}(\beta - \alpha) - \gamma,$$

$$\mathfrak{L}_2': \quad \varphi - \theta = \tfrac{1}{2}(\beta - \alpha) - \gamma - 2\pi,$$

where $-\cos\gamma = x_{34}$, $-\pi \leq \gamma \leq 0$, and the other quantities have been defined previously (4).

The function g vanishes on each of these four lines and also on the two additional lines

$$\mathfrak{L}_3: \quad \varphi + \theta = \tfrac{1}{2}(\alpha + \beta) + \gamma + \pi,$$

$$\mathfrak{L}_4: \quad \varphi - \theta = \tfrac{1}{2}(\beta - \alpha) + \gamma.$$

These are the only roots of g in the fundamental square. The square is shown in Fig. 1. It is divided into two squares and two rectangles by the pair of lines $\mathfrak{L}_3$ and $\mathfrak{L}_4$. The sign of g, and consequently $L_i = f_i^2 g$, is fixed inside these four regions. A calculation at the center of each of these regions shows that g is positive in the two rectangles and negative in the two squares.

The lines on which g vanishes are parts of the three-point Landau surface. They are of a very singular nature and correspond to only one point in x space. On these lines the x's have the values given in Table III.

Besides the points $g = 0$ the three-point minor L_3 vanishes for $\varphi = -\frac{1}{2}\beta + \frac{1}{2}\pi$ and $\varphi = -\frac{1}{2}\beta + \frac{3}{2}\pi$. The minor L_4 similarly vanishes for $\theta = -\frac{1}{2}\alpha + \frac{1}{2}\pi$ and $-\frac{1}{2}\alpha + \frac{3}{2}\pi$. All four of these lines have portions inside the square, and, together with the lines on which $g = 0$, they exhaust the real intersection of the three- and four-point surfaces. The significant portions of these lines are shown in Fig. 1.

It is easy to examine the three-point surfaces and see which of the intersections examined above are singular on the physical sheet. For example, when

TABLE III. The values of the x's on the lines $\mathfrak{L}$.

	x_{12}	x_{13}	x_{24}
$\mathfrak{L}_1$ or $\mathfrak{L}_1'$	$-\cos(\alpha+\beta-\gamma)$	$-\cos(\alpha-\gamma)$	$-\cos(\beta-\gamma)$
$\mathfrak{L}_2$ or $\mathfrak{L}_2'$	$-\cos(\alpha-\beta+\gamma)$	$-\cos(\alpha+\gamma)$	$-\cos(\beta-\gamma)$
$\mathfrak{L}_3$	$-\cos(\alpha+\beta+\gamma)$	$-\cos(\alpha+\gamma)$	$-\cos(\beta+\gamma)$
$\mathfrak{L}_4$	$-\cos(-\alpha+\beta+\gamma)$	$-\cos(\alpha-\gamma)$	$-\cos(\beta+\gamma)$

$\varphi = \frac{1}{2}(\pi - \beta)$,

$$x_{24} = \sin(\tfrac{1}{2}\alpha - \theta) = \cos[\tfrac{1}{2}(\pi - \alpha) + \theta],$$

$$x_{12} = -\sin(\theta + \tfrac{1}{2}\alpha) = \cos[\tfrac{1}{2}(3\pi - \alpha) - \theta],$$

and by definition $x_{14} = \cos\alpha$. The condition that $L_3 = 0$ be singular on the physical sheet is that the arguments of these three cosines add up to 2π, and that each argument be between 0 and π. We have

$$\tfrac{1}{2}(\pi - \alpha) + \theta + \tfrac{1}{2}(3\pi - \alpha) - \theta + \alpha = 2\pi$$

and if $\pi - \frac{1}{2}\alpha \leq \theta \leq \pi + \frac{1}{2}\alpha$, the second condition is met. In a similar way the line $\theta = \frac{1}{2}(\pi - \alpha)$ is singular on the physical sheet from $\frac{1}{2}(\pi - \beta) \leq \varphi \leq \frac{1}{2}(\pi + \beta)$. The other lines may be tested and found to be not singular on the physical sheet. It is reassuring that the singular portions begin at the common point $\theta = \frac{1}{2}(\pi - \alpha)$, $\varphi = \frac{1}{2}(\pi - \beta)$, but they appear to end at the peculiar points $\theta = \pi + \frac{1}{2}\alpha$, $\varphi = \pi + \frac{1}{2}\beta$. These points correspond to infinity in x space, however, and are thus satisfactory end points for these singular lines.

The four-point real singularities are in the region R bounded by the three-point singularities. For x's in this region, the integration parameters are real and positive and this is the extent of the unambiguous singularities. The right-hand rectangle is free of physical-sheet singularities. The two squares in which g is negative have portions that are singular in some limits but not in others. These are the portions R' adjacent to R as indicated in Fig. 1. Crossing the line $g = 0$ makes a similar change in all the L_i and thus the radicals $(L_i)^{\frac{1}{2}}$ are all on the same sheet. Alternatively, the entire region $g < 0$ is part of all the four three-point cuts, and those portions of the four-point surface lying in the three-point cut will be singular in some limits and not in others. The region R is, of course, singular in all limits.

If x_{12} is fixed, this corresponds to a curve in the square. The intersection of this curve with R is the region Γ_5 of Tarski's treatment,[3] and the existence of such a region gives rise to complex singularities since no matter how a complex surface leaves this region, it remains singular. The intersection of the curve x_{12} = constant with the provisionally singular region in the squares, corresponds to the usual Mandelstam curve in the s–t plane. It should be noted that the curve x_{12} = constant will not, in general, cross the lines $\mathfrak{L}$ since, on these lines, x_{12} has a fixed value. Thus, appearances to the contrary, Γ_5 does not touch the other region.

To extend the present treatment into the complex space of x_{12}, x_{13}, and x_{24} requires some specification of the cuts of three-point functions since these are just the boundaries of the four-point function singularities. The simplest way of choosing these cuts would be to have their intersection with the planes Im $\theta = \theta_0$, Im $\varphi = \varphi_0$ be independent of θ_0 and φ_0. That is to extend the real singularities as "polycylinders" into the complex space. Although this choice is appealing because of the simplicity of its intersections with the four-point surface, it is not clear how to choose the cuts to give this simple intersection. A different scheme would choose the three-point cut so that Im $(L) \leq 0$, Re $(L) < 0$. This is a generalization of the two-point cut Im $(1 - x^2) = 0$, Re $(1 - x^2) < 0$. This can be worked out in detail. It is clear that the cut does not intersect the surface $L = 0$ again. Its intersection with the four-point surface may be found and is sufficiently complicated so that it holds little promise of being a satisfactory choice. If this choice is made then, and if all the $(L_i)^{\frac{1}{2}}$ have the same sign for the real part, there is a physical-sheet singularity. There is a second possible generalization of the two-point cut that holds some promise. The two-point branchpoint is at $x = -1$ and the cut is taken along $x = -1 - A$, where A is real and positive. In a similar way, the three-point cut attached to $L_4 = 0$ might be taken as

$$x_{12} = \bar{x}_{12} - A \qquad x_{13} = \bar{x}_{13} - A \qquad x_{23} = \bar{x}_{23} - A,$$

where the $\bar{x}$'s are points for which $L_4 = 0$, and A is a real positive number. This is simpler than the previously suggested cut. The cut may intersect the Landau surface again and it has not been possible to show that such intersections only occur at points that are not physical-sheet singularities. It is easy to show that this is the case for the cuts attached to real portions of $L_4 = 0$.

A similar treatment may be given for the four-point function in a larger number of variables. Until a good choice is made for the location of the three-point cuts, this does not appear to be worthwhile.

There is yet another possibility for the four-point function as a function of six variables. The parameterization given has a projective invariance property. That is, if $u \to \lambda u$, $a \to \lambda^2 a$, the x's are unaltered. This may be taken advantage of to remove one parameter and set $g^{(6)} = 1$. This can be done, for example, by solving for $a - b$ in terms of $b - c$. It requires only the solution of a quadratic equation and may be carried through explicitly. It is not apparently possible to eliminate a, b, c

in favor of a single parameter in a symmetric way. If g has been made equal to one, then $(L_i)^{\frac{1}{2}} = u_i$, and the physical-sheet singularities correspond to the region in which all the u_i have positive real parts. The three-point cuts play another role in these considerations. Their intersections with each other and the two- and four-point cuts will form the distinguished boundary of the physical sheet; thus, they should be chosen with a view to making these intersections simple, since the representation integral that is the ultimate goal of these considerations will extend over this intersection.

The five-point function may be treated in an exactly analogous manner and it will be the four-point cuts that are the issue. The five-point function has a real region similar to R. It is contained in the intersection of several four-point R's, and is thus more remote from the physical region than R. Similarly, R is contained in the intersection of several three-point regions, and is further from the physical region than they are. It thus seems highly unlikely that the real singularities of the higher functions that come from regions like R have any easily detectable influence on scattering processes, since they are successively further removed from the physical region. Whether or not the complex singularities manage to bend back and approach the physical region, or whether the singularities from the R' region have a detectable influence, is an open question.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 7 JULY 1963

Density Matrices Arising from Incomplete Measurements

EYVIND H. WICHMANN*
Department of Physics, University of California, Berkeley, California
(Received 11 March 1963)

The problem of how to associate a statistical ensemble of a quantum mechanical system with an incomplete set of measured ensemble averages is discussed. The view adhered to in this note is that the most chaotic ensemble consistent with the measured ensemble averages is a reasonable choice of ensemble. The properties of this ensemble are studied rigorously for the case when the quantum mechanical system is associated with a finite-dimensional Hilbert space. The results are extended heuristically to some particular ensembles of many-boson and many-fermion systems.

INTRODUCTION

1

IN this note we shall consider the question of how a statistical ensemble may be defined in terms of an incomplete set of measurements. We shall restrict our rigorous discussion to the case of statistical ensembles of quantum mechanical systems, the state vectors of which are vectors in a finite-dimensional Hilbert space $\mathcal{H}$. We will, however, also apply our results heuristically to more general systems when we consider some examples of particular physical interest.

Let N be the dimensionality of $\mathcal{H}$. A statistical ensemble can then be described by an N-dimensional density matrix ρ. We shall review some well-known facts about density matrices and at the same time explain the notation and terminology which we shall follow.

(a) Let V_h be the real vector space of all Hermitian N-dimensional matrices: This vector space is of dimensionality N^2. On V_h we define a positive-definite real symmetric scalar product by

$$(X, Y) = \mathrm{Tr}\,(XY), \tag{1}$$

where X and Y are any two elements of V_h.

(b) Let V_d be the set of all N-dimensional density matrices, i.e., the set of all matrices in V_h which are of unit trace, and which are nonnegative definite. (c) We shall, in general, denote proper Hermitian idempotents in V_h by the letter F, suitably indexed. Such an operator thus satisfies the conditions

$$F^\dagger = F; \quad F^2 = F; \quad \mathrm{Tr}\,(F) = \text{a positive integer}.$$

We denote primitive Hermitian idempotents in V_h by the letter E, suitably indexed. Such an idempotent thus satisfies the additional condition $\mathrm{Tr}\,(E) = 1$.

* Alfred P. Sloan Foundation Fellow.

in favor of a single parameter in a symmetric way. If g has been made equal to one, then $(L_i)^{\frac{1}{2}} = u_i$, and the physical-sheet singularities correspond to the region in which all the u_i have positive real parts. The three-point cuts play another role in these considerations. Their intersections with each other and the two- and four-point cuts will form the distinguished boundary of the physical sheet; thus, they should be chosen with a view to making these intersections simple, since the representation integral that is the ultimate goal of these considerations will extend over this intersection.

The five-point function may be treated in an exactly analogous manner and it will be the four-point cuts that are the issue. The five-point function has a real region similar to R. It is contained in the intersection of several four-point R's, and is thus more remote from the physical region than R. Similarly, R is contained in the intersection of several three-point regions, and is further from the physical region than they are. It thus seems highly unlikely that the real singularities of the higher functions that come from regions like R have any easily detectable influence on scattering processes, since they are successively further removed from the physical region. Whether or not the complex singularities manage to bend back and approach the physical region, or whether the singularities from the R' region have a detectable influence, is an open question.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 4, NUMBER 7 JULY 1963

Density Matrices Arising from Incomplete Measurements

EYVIND H. WICHMANN*
Department of Physics, University of California, Berkeley, California
(Received 11 March 1963)

The problem of how to associate a statistical ensemble of a quantum mechanical system with an incomplete set of measured ensemble averages is discussed. The view adhered to in this note is that the most chaotic ensemble consistent with the measured ensemble averages is a reasonable choice of ensemble. The properties of this ensemble are studied rigorously for the case when the quantum mechanical system is associated with a finite-dimensional Hilbert space. The results are extended heuristically to some particular ensembles of many-boson and many-fermion systems.

INTRODUCTION

1

IN this note we shall consider the question of how a statistical ensemble may be defined in terms of an incomplete set of measurements. We shall restrict our rigorous discussion to the case of statistical ensembles of quantum mechanical systems, the state vectors of which are vectors in a finite-dimensional Hilbert space $\mathcal{H}$. We will, however, also apply our results heuristically to more general systems when we consider some examples of particular physical interest.

Let N be the dimensionality of $\mathcal{H}$. A statistical ensemble can then be described by an N-dimensional density matrix ρ. We shall review some well-known facts about density matrices and at the same time explain the notation and terminology which we shall follow.

(a) Let V_h be the real vector space of all Hermitian N-dimensional matrices: This vector space is of dimensionality N^2. On V_h we define a positive-definite real symmetric scalar product by

$$(X, Y) = \mathrm{Tr}\,(XY), \tag{1}$$

where X and Y are any two elements of V_h.

(b) Let V_d be the set of all N-dimensional density matrices, i.e., the set of all matrices in V_h which are of unit trace, and which are nonnegative definite. (c) We shall, in general, denote proper Hermitian idempotents in V_h by the letter F, suitably indexed. Such an operator thus satisfies the conditions

$$F^\dagger = F; \quad F^2 = F; \quad \mathrm{Tr}\,(F) = \text{a positive integer}.$$

We denote primitive Hermitian idempotents in V_h by the letter E, suitably indexed. Such an idempotent thus satisfies the additional condition $\mathrm{Tr}\,(E) = 1$.

* Alfred P. Sloan Foundation Fellow.

(d) It is well-known that the set V_d of all density matrices is a closed convex $(N^2 - 1)$-dimensional set.[1] The set of extreme points of V_d consists precisely of the set of all primitive Hermitian idempotents in V_h. The boundary of V_d consists precisely of the set of all singular density matrices ρ, i.e., for which $\det(\rho) = 0$. A density matrix is an interior point of V_d if and only if $\det(\rho) > 0$.

(e) There is a one–one correspondence between all density matrices in V_d, and all ensembles of the system. If an ensemble is described by the density matrix ρ, then the ensemble average of any observable X (i.e., any matrix in V_h) is given by

$$\mathrm{Av}(X) = \mathrm{Tr}(\rho X).$$

The ensemble is completely known if the corresponding ensemble averages are known for any set of N^2 linearly independent matrices in V_h. Let the matrices B_n, $n = 1, 2, \cdots, N^2$, form an orthonormal basis of V_h. Then

$$\rho = \sum_n (\rho, B_n)B_n = \sum_n B_n \mathrm{Tr}(\rho B_n) = \sum_n B_n \mathrm{Av}(B_n),$$

which formula expresses ρ explicitly in terms of the ensemble averages of the B_n. Since $\mathrm{Av}(I) = 1$ it actually suffices to determine the ensemble averages of a suitable set of $(N^2 - 1)$ observables.

(f) The entropy, $s(\rho)$, of a statistical ensemble described by the density matrix ρ is defined by

$$s = s(\rho) = -\mathrm{Tr}(\rho \ln \rho).$$

The negative of the entropy is a measure of the amount of information possessed about the elements of the ensemble. The smallest value of s equals 0; it is assumed if and only if ρ is a primitive Hermitian idempotent. The largest value of s equals $\ln(N)$, and it is assumed only for the uniform ensemble, $\rho = I/N$, which represents the most chaotic ensemble possible.

(g) The negative of the entropy is a strictly convex function on V_d, which property is expressed by the Von Neumann Mixing Theorem,[2] as follows: Let the real numbers $\theta_n > 0$ for $n = 1, 2, \cdots, p$, satisfy the condition

$$\sum_{n=1}^{p} \theta_n = 1,$$

and let

$$\rho = \sum_{n=1}^{p} \theta_n \rho_n,$$

where the ρ_n are density matrices. Then ρ is also a density matrix, and if s_n is the entropy of ρ_n, and if s is the entropy of ρ, then the theorem asserts that

$$s \geq \sum_{n=1}^{p} \theta_n s_n,$$

and that the equality can obtain if and only if $\rho = \rho_n$ for all n.

(h) From the theorem just quoted we can draw the following conclusion:

Lemma I. If V_d' is any closed convex subset of V_d, then there exists a *unique* density matrix ρ_0 in V_d' with the property that the entropy of ρ_0 is *larger* than the entropy of any other density matrix in V_d', or in other words: V_d' contains a *unique* most chaotic ensemble. We shall call ρ_0 *the most chaotic density matrix in* V_d', or the density matrix of maximum chaos in V_d'.

This lemma is a trivial consequence of Von Neumann's Theorem. A similar lemma would hold for an "entropy" defined by any other convex function in place of $x \ln x$, but it may be noted that given V_d' the "most chaotic" density matrix ρ_0 does depend on how the "entropy" is defined.

2

We may now ask the following question: How should we describe an ensemble for which less than $(N^2 - 1)$ ensemble averages have been determined? We may thus assume that we have determined (measured) the ensemble averages of only the p operators $X_1, X_2, \cdots, X_p$. Given these averages there exists a closed convex subset V_d' of V_d such that all the matrices in V_d', but only these, give the measured averages for the X_n. In this paper we shall adhere to the view that the most rational and natural assumption that can be made about the ensemble in such a case is that the ensemble is the most chaotic one in V_d'. In other words: *We associate with any incomplete set of measurements the unique statistical ensemble which is the most chaotic one consistent with the ensemble averages which have in fact been measured.*

We shall discuss the determination of this ensemble and its properties. We know beforehand that the problem of finding the ensemble always has a unique solution.

It is to be noted that the matrices X_n need by no means commute among themselves. The solution is well-known for the case when the matrices do

[1] For basic facts and definitions in the theory of convex sets, we refer to H. G. Eggleston, *Convexity* (Cambridge University Press, London, England, 1958).

[2] J. Von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Dover Publications, Inc., New York, 1943), Chap. V, Sec. 3.

commute, a particular case of which is the case when $p = 1$.

In Secs. 3 and 4 we shall review the theory for the case $p = 1$. In Secs. 5 and 6 we shall determine the region of possible simultaneous ensemble averages of a set of p operators. In Secs. 7–14 we shall consider the determination of the most chaotic ensemble given the ensemble averages of p operators. In Secs. 15 and 16 we shall consider the application of our results to the case of a particle in one-dimensional motion where the operators X_n are linear or quadratic expressions in the position and momentum variables: In this case the Hilbert space $\mathfrak{H}$ is infinite-dimensional. In Secs. 17–19 we shall finally apply our theory to ensembles of many-boson and many-fermion systems for the case when the operators X_n are bilinear expressions in creation and destruction operators.

THE MOST CHAOTIC ENSEMBLE DETERMINED BY A SINGLE ENSEMBLE AVERAGE

3

Let X be a proper dynamical variable, i.e., a matrix in V_h not a multiple of the identity I. Let x_M be the largest eigenvalue of X, and let x_m be the smallest eigenvalue of X. Let μ_M be the multiplicity of x_M and let μ_m be the multiplicity of x_m, and let the Hermitian idempotents associated with x_M and x_m be denoted by F_M, respectively F_m. We shall state two lemmas:

Lemma II. (a) The set $S(X)$ of all possible ensemble averages of X is identical with the closed interval (x_m, x_M).

(b) Let, for any real number y, the matrix $\omega(y)$ be defined by

$$\omega(y) = \{\mathrm{Tr}\,[\exp(yX)]\}^{-1} \exp(yX). \tag{2}$$

Then, if x is any number such that $x_M > x > x_m$, the equation

$$x = x(y) = \mathrm{Tr}\,[X\omega(y)] \tag{3}$$

has a *unique* real solution y. The mapping $x = x(y)$ is a one–one mapping, analytic for all real y, of the real line onto the *open* interval (x_m, x_M).

(c) Furthermore,

$$\lim_{y\to+\infty} \omega(y) = F_M/\mu_M; \qquad \lim_{y\to-\infty} \omega(y) = F_m/\mu_m, \tag{4}$$

and, consequently,

$$\lim_{y\to+\infty} \mathrm{Tr}\,[X\omega(y)] = x_M; \quad \lim_{y\to-\infty} \mathrm{Tr}\,[X\omega(y)] = x_m. \tag{5}$$

(d) Let the inverse of the mapping $x = x(y)$, defined in Eq. (3), be denoted by $y = y(x)$. Let the matrix $\rho(x)$ be defined for all x in $S(X)$, i.e., all x in the closed interval (x_m, x_M), by

$$\rho(x) = \omega(y(x)), \text{ for all } x \text{ such that } x_M > x > x_m, \tag{6a}$$

$$\rho(x_M) = F_M/\mu_M; \qquad \rho(x_m) = F_m/\mu_m. \tag{6b}$$

Let $R(X)$ be the set of all matrices $\omega(y)$ as y takes all real values, and let $R_c(X)$ be the closure of $R(X)$. Then $R_c(X)$ is the set of all matrices $\rho(x)$ as x assumes all values in $S(X)$, and the mapping of $S(X)$ onto $R_c(X)$, as defined by Eqs. (6), is a bicontinuous one–one mapping.

(e) For every y, the matrix $\omega(y)$ is a density matrix, and in fact every matrix in $R_c(X)$ is a density matrix.

Lemma III. With the notation in the preceding lemma, let x be any point in $S(X)$, i.e., let x be any possible ensemble average of the operator X. Then the most chaotic density matrix subject to the restriction $\mathrm{Av}\,(X) = x$ is the matrix $\rho(x)$. In other words, if ρ' is any other density matrix such that $\mathrm{Tr}\,(X\rho') = x$, then the entropy of ρ' is *smaller* than the entropy of $\rho(x)$.

These two lemmas are easily proved, and are well-known, even if not always stated in exactly the above form, and we will not repeat the proofs here.[3] We remark that if X is interpreted as the Hamiltonian of a quantummechanical system, then the ensemble of maximum chaos given in Lemma III is, as is well-known, the canonical ensemble.

Furthermore, we may remark that if we consider the entropy $s = s(\rho)$ as a function of the density matrix ρ, where ρ varies over the set of all density matrices satisfying $\mathrm{Tr}\,(X\rho) = x$, x being fixed and in the interior of the interval (x_m, x_M), then $s(\rho)$ assumes a stationary value at one and only one point, namely at the point $\rho(x)$. As we have said before, this stationary value is always a strict maximum.

4

The facts presented in the preceding section enable us to draw a conclusion which will be useful later. We thus state

Lemma IV. Let A and B be any two Hermitian N-dimensional matrices. Then the two equations

$$\mathrm{Tr}\,(Ae^A)\,\mathrm{Tr}\,(e^B) = \mathrm{Tr}\,(Ae^B)\,\mathrm{Tr}\,(e^A), \tag{7a}$$

$$\mathrm{Tr}\,(Be^A)\,\mathrm{Tr}\,(e^B) = \mathrm{Tr}\,(Be^B)\,\mathrm{Tr}\,(e^A) \tag{7b}$$

are both satisfied if and only if $A = B + cI$, for some real number c.

[3] See reference 2, Chap. V, Sec. 3. The derivation may be simplified a bit by the application of the considerations in Sec. 9 of the present paper.

Proof: Suppose that both equations (7) hold. Let s_a be the entropy of the density matrix $e^A/\mathrm{Tr}\ (e^A)$ and let s_b be the entropy of the density matrix $e^B/\mathrm{Tr}\ (e^B)$. Then, by Lemma III, Eq. (7a) implies that $s_a \geq s_b$, whereas the Eq. (7b) similarly implies that $s_b \geq s_a$. Hence we have $s_a = s_b$, and it follows from Lemma III that $e^A/\mathrm{Tr}\ (e^A) = e^B/\mathrm{Tr}\ (e^B)$. This is possible only if $(A - B)$ is a multiple of the identity, in which case Eq. (7) are trivially satisfied.

REGION OF SIMULTANEOUS ENSEMBLE AVERAGES OF p OBSERVABLES

5

Let $X = \{X_1, X_2, \cdots, X_p\}$ be a set of p observables (p matrices in V_h) such that the matrices $I, X_1, X_2, \cdots, X_p$ are linearly independent.

With every ρ of V_d we associate a vector $x = \{x_1, x_2, \cdots, x_p\}$ in the space R^p of p real variables by defining

$$x_k = \mathrm{Av}\ (X_k) = \mathrm{Tr}\ (X_k\rho); \quad k = 1, 2, \cdots, p. \tag{8a}$$

We may also denote this *linear* mapping of V_d into R^p by

$$x = \mathrm{Tr}\ (X\rho). \tag{8b}$$

Furthermore, we will employ the notations

$$y\cdot X = \sum_{k=1}^{p} y_k X_k; \qquad y\cdot y' = \sum_{k=1}^{p} y_k y_k',$$

where y and y' are any two vectors of R^p.

Let $S(X)$ be the set of all x in R^p which are images of points in V_d under the mapping defined by Eq. (8). We shall collect some interesting properties of $S(X)$ in a theorem.

Theorem I. (a) The set $S(X)$, defined as above, is a bounded, closed convex region in R^p.

(b) Every interior point of V_d is mapped into an interior point of $S(X)$, and hence every point on the boundary of $S(X)$ is the image of some point, or points, on the boundary of V_d but never the image of an interior point.

(c) Let $S_E(X)$ be the image of the set of all extreme points of V_d, i.e., the image of the set of all primitive Hermitian idempotents. The extreme points of $S(X)$ all lie in $S_E(X)$, and $S(X)$ is identical with the convex envelope of $S_E(X)$.

(d) Let u be any unit vector in R^p, i.e., any real vector u satisfying $u\cdot u = 1$. Let $\lambda_M(u)$ be the largest eigenvalue of $u\cdot X$, and let $\lambda_m(u)$ be the smallest eigenvalue of $u\cdot X$. Then every interior point x of $S(X)$ satisfies the conditions

$$\lambda_M(u) > u\cdot x > \lambda_m(u)$$

for all unit vectors u. For any unit vector u, the two planes in R^p described by the equations $u\cdot x = \lambda_M(u)$ and $u\cdot x = \lambda_m(u)$ are support hyperplanes of the set $S(X)$, and each one of these planes thus intersects the boundary of $S(X)$ in at least one point. The set $S(X)$ is the set of all points x which satisfy the relations

$$\lambda_M(u) \geq u\cdot x \geq \lambda_m(u),$$

for *all* unit vectors u. The plane $u\cdot x = \lambda_M(u)$ is identical with the plane $-u\cdot x = \lambda_m(-u)$.

(e) Let $F_M(u)$ be the Hermitian idempotent associated with the largest eigenvalue of $u\cdot X$, and let $x'(u) = \mathrm{Tr}\ [XF_M(u)/\mu_M(u)]$, where $\mu_M(u)$ is the multiplicity of $\lambda_M(u)$. Let $S'(X)$ be the set of all points $x'(u)$, as u ranges through the set of all unit vectors. Then all the extreme points of $S(X)$ are included in $S'(X)$, and $S(X)$ is thus the convex envelope of $S'(X)$.

Note: (a) It may perfectly well happen that some of the points of $S_E(X)$ lie in the interior of $S(X)$.

(b) It may perfectly well happen that $S(X)$ has boundary points which are not extreme points; on the other hand, it may also happen that every boundary point of $S(X)$ is an extreme point.

(c) The actual construction of $S(X)$ given the set X may be based on the facts listed in parts (c), (d), and (e) of the theorem.

Proof: The assertions in parts (a), (b), and (c) are trivial and need not be proved here. To prove the parts (d) and (e) we note the inequalities

$$\lambda_M(u) \geq \mathrm{Tr}\ [(u\cdot X)\rho] \geq \lambda_m(u),$$

which must hold for any density matrix ρ, and any unit vector u. If now $x = \mathrm{Tr}\ (X\rho)$ is an interior point of $S(X)$, we must have

$$\lambda_M(u) > u\cdot x > \lambda_m(u),$$

for all unit vectors u. On the other hand, we have

$$\mathrm{Tr}\ [(u\cdot X)F_M(u)/\mu_M(u)] = \lambda_M(u)$$

for the density matrix $F_M(u)/\mu_M(u)$, and the point $x'(u)$ defined in part (e) thus satisfies $u\cdot x'(u) = \lambda_M(u)$. It follows that $u\cdot x = \lambda_M(u)$ is the equation for a support hyperplane of $S(X)$, and that $x'(u)$ lies in this plane as well as an the boundary of $S(X)$. As u varies over the set of all unit vectors we obtain *all* support hyperplanes of $S(X)$. If x is an extreme point of $S(X)$ there exists at least one support hyperplane which intersects $S(X)$ only at x, and we conclude that the set $S'(X)$ must include all extreme points of $S(X)$. The proofs of the remaining assertions are obvious.

6

It may be of interest to consider some examples and special cases of the sets $S(X)$ before we proceed with the general discussion.

(a) Suppose that $p = 3$, and that the operators X_1, X_2, and X_3 are angular-momentum operators. Let j be the largest eigenvalue of X_3; then j is also the largest eigenvalue of $u \cdot X$ for any unit vector u. It follows that the region $S(X)$ is a sphere of radius j, centered at the origin. (j must of course be larger than $\frac{1}{2}$ since the operators X_1, X_2 and X_3 are assumed to be linearly independent.) In this example, every boundary point of $S(X)$ is also an extreme point.

(b) Suppose that all the matrices X_k, $k = 1, 2, \cdots, p$ commute among themselves. We note that in this case p must be less than N. Let $x^{(s)} = \{x_1^{(s)}, x_2^{(s)}, \cdots, x_p^{(s)}\}$, $s = 1, 2, \cdots, N$, be the points of R^p corresponding to the *simultaneous* eigenvalues of the operators X_k. The extreme points of $S(X)$ must be included in this set. There can be at most N different such points. On the other hand at least $(p + 1)$ points in the set $x^{(s)}$, $s = 1, 2, \cdots, N$, must be different and noncoplanar, as there would otherwise exist a linear relation between the matrices $I, X_1, X_2, \cdots, X_p$. The region $S(X)$ is thus a convex polytope of at least $(p + 1)$, but at most N, vertices. In the special case of $p = 2$, $S(X)$ is a convex polygon of at least 3, but at most N, vertices. In the special case of $p = (N - 1)$, which is the largest possible p, the polytope has N vertices and is the $(N - 1)$-dimensional analogue of the tetrahedron (or the triangle).

(c) Suppose that the Hilbert space $\mathcal{H}$ is the tensor product of p Hilbert spaces $\mathcal{H}_k$, $k = 1, 2, \cdots, p$, and suppose furthermore that for every k the operator X_k is the tensor product of an operator $\bar{X}_k$ acting on $\mathcal{H}_k$ and the identity operators acting on the remaining factors $\mathcal{H}_m$, $m \neq k$. Under these circumstances, the region $S(X)$ is a rectangular parallelotope, the 2^p vertices of which correspond to the 2^p different combinations of smallest and largest eigenvalues of the $\bar{X}_k$.

(d) Finally we may consider the case when $p = N^2 - 1$. Without loss of generality we may assume that the matrices X_k form a complete orthonormal basis of the p-dimensional real vector space of all traceless Hermitian matrices. The set $S(X)$ is then the set of all x such that the matrix

$$B(x) = \frac{I}{N} + \sum_{n=1}^{p} x_n X_n$$

is nonnegative definite. As x varies over $S(X)$, $B(x)$ varies over the set of all density matrices, and the two sets are in one–one correspondence. The boundary points of $S(X)$, to which correspond the boundary points of V_d, satisfy the equation $\det [B(x)] = 0$.

The author does not know of any simple geometrical characterization of the region $S(X)$. Suffice it to say that for $N = 2$ the region is a sphere, whereas for $N > 2$ the region is *not* a sphere, although all extreme points of $S(X)$ lie on the surface of a sphere of radius $(1 - 1/N)^{\frac{1}{2}}$.

THE MOST CHAOTIC ENSEMBLE DETERMINED BY p ENSEMBLE AVERAGES

7

We will next state and prove a generalization, to the case of p operators, of the statements contained in Lemmas II and III.

Theorem II. Let $X = \{X_1, X_2, \cdots, X_p\}$ be a set of p observables (a set of p N-dimensional Hermitian matrices) such that the matrices $I, X_1, X_2, \cdots, X_p$ are linearly independent. Let $S(X)$ be the region of all simultaneous ensemble averages of the matrices X. Let $y = \{y_1, y_2, \cdots, y_p\}$ be any vector in R^p, the space of p real variables, and let

$$\omega(y) = [\mathrm{Tr}\,[\exp(y \cdot X)]]^{-1} \exp(y \cdot X), \tag{9}$$

where

$$y \cdot X = \sum_{k=1}^{p} y_k X_k.$$

Let, furthermore,

$$x(y) = \mathrm{Tr}\,[X\omega(y)]; \quad \{\text{or} \quad x_k(y) = \mathrm{Tr}\,[X_k\omega(y)]\}. \tag{10}$$

Then

(a) For every y in R^p, the matrix $\omega(y)$ is a density matrix in the *interior* of V_d, the set of all density matrices.

(b) The mapping $x = x(y)$, as defined in Eq. (10), is a one–one mapping, analytic for all real y, of R^p onto the interior of $S(X)$.

(c) Let $R(X)$ be the set of all $\omega(y)$ as y varies throughout R^p, and let $R_c(X)$ be the closure of $R(X)$. For every x of $S(X)$ there exists a unique density matrix in $R_c(X)$, denoted $\rho(x)$, such that

$$x = \mathrm{Tr}\,[X\rho(x)], \tag{11}$$

and such that the entropy of $\rho(x)$ is larger than the entropy of any other ρ in $R_c(X)$ which satisfies $\mathrm{Tr}\,(\rho X) = x$. Let $R_c^*(X)$ be the set of all matrices $\rho(x)$. The one–one mapping of $S(X)$ onto $R_c^*(X)$ is such that if $y = y(x)$ denotes the inverse of the mapping defined by Eq. (10); then

$$\rho(x) = \omega[y(x)], \tag{12}$$

for any x in the interior of $S(X)$.

(d) For any x in $S(X)$, the density matrix $\rho(x)$ in $R_c^*(X)$ is the most chaotic one among all density matrices which satisfy the condition Av $(X) = x$, i.e., if ρ is any other density matrix such that Tr $(X\rho) = x$, then the entropy of $\rho(x)$ is larger than the entropy of ρ.

Note: (a) The most important part of the above theorem has to do with the interior points of $S(X)$, namely, if x is in the interior of $S(X)$, then the equation $x = x(y)$ has a *unique* solution y in R^p, and the density matrix $\omega(y) = \rho(x)$ is then the most chaotic one among all density matrices having x as the ensemble average of X. For the sake of completeness we also include the case when x is on the boundary of $S(X)$ in our statement of the theorem.

(b) The assertion in part (a) of the theorem is trivial. Before we prove the remaining assertions we shall state and prove three lemmas.

Lemma V. Let $x = x(y)$ be defined as in Eq. (10). Then $x(y') = x(y'')$ implies that $y' = y''$. Hence if $S'(X)$ is the image of R^p under the mapping $x = x(y)$, then the mapping of R^p onto $S'(X)$ is one-one. $S'(X)$ is contained in the interior of $S(X)$.

Proof: Suppose that for two vectors y' and y'' we have $x(y') = x(y'')$. Let $A = y' \cdot X$ and let $B = y'' \cdot X$. The two matrices A and B then satisfy Eqs. (7) in Lemma IV, and it follows from this lemma that $(A - B)$ must be a multiple of the identity. Since we have assumed that $I, X_1, X_2, \cdots, X_p$ are linearly independent, it follows that $A = B$, and hence $y' = y''$, as asserted. That $S'(X) = x(R^p)$ is included in the interior of $S(X)$ follows from part (b) of Theorem I, and from the trivial fact that $\omega(y)$ is in the *interior* of V_d for all y in R^p.

Lemma VI. Let $x = x(y)$ be defined as in Eq. (10).

(a) The symmetric p-dimensional real matrix M defined by

$$M_{mn} = \frac{\partial x_m}{\partial y_n} = \frac{\partial^2}{\partial y_n \, \partial y_m} \ln \{\text{Tr } [\exp (y \cdot X)]\} \qquad (13)$$

is positive definite for all y, and the Jacobian $\partial(x)/\partial(y)$ is thus positive for all y.

(b) The set $S'(X) = x(R^p)$ is accordingly *open* relative to $S(X)$.

Proof: Let $w = \{w_1, w_2, \cdots, w_p\}$ be any nonzero vector in R^p. Let y be any vector in R^p and let $W = w \cdot X$ and $H = y \cdot X$. Then, by a simple computation,[4]

$$\sum_{m,n=1}^{p} w_m M_{mn} w_n = [\text{Tr } (e^H)]^{-2} \Big\{ \text{Tr } (e^H) \times \text{Tr} \Big[\int_0^1 dt \, e^{tH} W e^{(1-t)H} W \Big] - [\text{Tr } (We^H)]^1 \Big\}. \qquad (14)$$

The denominator $[\text{Tr } (e^H)]^2$ in the right-hand side of the above expression is positive for any y. To study the numerator we consider, for a real variable λ,

$$g(\lambda) = \int_0^1 dt \, \text{Tr} \, \{[e^{-t\frac{1}{2}H}(\lambda - W)e^{t\frac{1}{2}H}] \times [e^{t\frac{1}{2}H}(\lambda - W)e^{-t\frac{1}{2}H}]e^H\},$$

which is manifestly positive unless $\lambda = W$, i.e., $g(\lambda)$ is positive for *all* λ because of our assumptions about w. The condition that the equation $g(\lambda) = 0$ shall have no real solution is however equivalent to the condition that the numerator in the expression (14) is positive, which proves part (a) of the present lemma. Part (b) is a trivial consequence of part (a): If $x_0 = x(y_0)$, then a neighborhood of y_0 in R^p is mapped onto a neighborhood of x_0 in $S(X)$.

Lemma VII. Let $y^{(s)}$, $s = 1, 2, \cdots$, be an infinite sequence of points in R^p such that

$$\lim_{s \to \infty} x(y^{(s)}) = x_0,$$

where x_0 is an *interior* point of $S(X)$. Then the points $y^{(s)}$ converge to a point y_0 in R^p such that $x_0 = x(y_0)$, and the density matrices $\omega(y^{(s)})$ converge to the density matrix $\omega(y_0)$.

Proof: We must prove that the points $y^{(s)}$ stay within some bounded region in R^p. Suppose that this is *not* the case. We may then select a subsequence, $y^{(s')}$, where the integers s' tend to infinity, such that

$$\lim_{s' \to \infty} |y^{(s')}| = \infty; \qquad \lim_{s' \to \infty} \{y^{(s')}/|y^{(s')}|\} = u,$$

where u is a suitable unit vector. For this subsequence we then have

$$\lim_{s' \to \infty} \text{Tr } \{(u \cdot X)\omega[y^{(s')}]\} = \lambda_M(u),$$

where, as in Theorem I(d), the number $\lambda_M(u)$ stands for the largest eigenvalue of the matrix $u \cdot X$. On the other hand, we have assumed that

$$\lim_{s \to \infty} \text{Tr } \{X\omega[y^{(s)}]\} = x_0,$$

and we thus obtain

$$u \cdot x_0 = \lambda_M(u).$$

By Theorem I(d) it follows that x_0 must then be on the boundary of $S(X)$ which contradicts the

[4] In deriving this expression, one may make use of a well-known expansion of the trace of the exponential of a sum of two matrices. This expansion is quoted in M. L. Goldberger and E. N. Adams, J. Chem. Phys. **20**, 240 (1952), Appendix II.

assumption that x_0 is an interior point. Consequently all the points $y^{(n)}$ lie within some bounded region in R^p, and must then have at least one limit point y_0 such that $x_0 = x(y_0)$. By Lemma V it follows that there can be only one such point, as asserted in the present lemma. The proof of the remaining statement is trivial.

We are now in a position to prove the assertions in parts (b)–(d) of Theorem II.

8

Proof of Theorem II: (a) Let, as in Lemma V, $S'(x)$ be the image of R^p under the mapping $x = x(y)$, i.e., $S'(X) = x(R^p)$. Lemma VI asserts that $S'(X)$ is open relative to $S(X)$, and is contained in the interior of $S(X)$. On the other hand, Lemma VII asserts that $S'(X)$ contains every limit point of itself which lies in the interior of $S(X)$, and it follows that $S'(X)$ is precisely the interior of $S(X)$. Since, by Lemma V, the mapping $S'(X) = x(R^p)$ is one–one, and since the analyticity of this mapping for all real y is obvious, we have thus proved part (b) of the Theorem II.

(b) We next prove that if x is an interior point of $S(X)$ then the entropy of $\rho(x) = \omega[y(x)]$, where $y = y(x)$ is the inverse of the mapping defined in Eq. (10), and is *larger* than the entropy of any other density matrix ρ such that Tr $(X\rho) = x$. In order to prove this, we consider the matrix $A = y \cdot X$, and define $a = \text{Tr}\,[A\omega(y)]$. If now ρ is any other density matrix such that $x = \text{Tr}\,(X\rho)$, then we have Tr $(A\rho) = a$, and by Lemma III it follows that the entropy of $\rho(x) = \omega(y)$ is larger than the entropy of ρ, as asserted. We note that we have thereby proved part (d) of Theorem II for the special case that x is an interior point of $S(X)$.

(c) Let us now consider the case when x is a point on the boundary of $S(X)$. Let ρ_c be the density matrix of maximum chaos in the set of all density matrices ρ which satisfy the condition Tr $(X\rho) = x$. By Lemma I there exists such a unique density matrix ρ_c; we denote the entropy of ρ_c by $s_c = s(\rho_c)$. Let $\rho^{(n)}$, $n = 1, 2, \cdots$ be an infinite sequence of density matrices in the interior of V_d which converges to ρ_c. We note that since Tr $(X\rho_c) = x$ is on the boundary of $S(X)$ it follows, by Theorem I(b), that ρ_c must be on the boundary of V_d. Let s'_n be the entropy of $\rho^{(n)}$ and let $x^{(n)} = \text{Tr}\,[X\rho^{(n)}]$. We then have

$$\lim_{n\to\infty} x^{(n)} = x; \qquad \lim_{n\to\infty} s'_n = s_c. \tag{15}$$

Since $\rho^{(n)}$ is, for every n, in the interior of V_d, the points $x^{(n)}$ are all in the interior of $S(X)$, and we may construct the density matrices of maximum chaos $\rho[x^{(n)}] = \omega[y^{(n)}]$ where $y^{(n)}$ is the unique solution of the equation $x^{(n)} = x(y)$. Let s_n be the entropies of the matrices $\rho[x^{(n)}]$.

Consider now the infinite sequence of matrices $\rho[x^{(n)}]$. Since V_d is closed and bounded, this sequence must have at least one limit point in V_d, say ρ''. Without loss of generality, we may assume that ρ'' is the *only* limit point of the sequence; if this is not the case we can always select a subsequence of the integers n such that for this subsequence, $\rho[x^{(n)}]$ converges to ρ''. Let now s'' be the entropy of ρ''. We thus have

$$\lim_{n\to\infty} \rho(x^{(n)}) = \rho''; \qquad \lim_{n\to\infty} s_n = s''. \tag{16}$$

In view of part (b) of the present proof, we have, for every n, $s_n \geqq s'_n$, and hence, by Eqs. (15) and (16), $s'' \geqq s_c$. On the other hand, we must have Tr $(X\rho'') = x$, and by the construction of ρ_c, it follows that $s_c \geq s''$. Hence $s_c = s''$, and by Lemma I it follows that $\rho'' = \rho_c$. We have thus shown that for every x on the boundary of $S(X)$ the density matrix of maximum chaos, subject to the condition $\text{Av}(X) = x$, lies in the closure of $R(X)$ and actually on the boundary of $R_c(X)$.

This thus completes the proof of Theorem II.

9

We may remark here that for the case when x is such that there exists a y in R^p such that $x = \text{Tr}\,[X\omega(y)]$, it is very easy to show that the entropy is stationary at the point $\omega(y)$. To do this we simply write $\rho_H = \exp(H)$ where the variable H is a Hermitian matrix, and we then maximize the entropy $s = s(\rho_H)$ subject to the $(p + 1)$ auxiliary conditions Tr $(\rho_H X) = x$, and Tr $(\rho_H) = 1$, taking these conditions into account by the method of Lagrange's multipliers. We then find that the entropy is stationary for a matrix H such that $\rho_H = \omega(y)$. However, if we are given an arbitrary vector $\boldsymbol{x}$ we must first determine whether there actually exists a vector y such that Tr $[X\omega(y)] = x$, as it could well happen that no such y will exist. We may therefore say that the difficult part of the proof of Theorem II is the proof of the assertions contained in parts (b) and (c) of the theorem.

It should be pointed out that a well-known special application of the Theorem II is the determination of the grand canonical ensemble for an identical-particle system. In this case, $p = 2$, and the variables X_1 and X_2 are the two commuting operators H

and N, where H is the one-particle Hamiltonian and N is the number operator.[5]

In the particular case that all the operators in the set X commute among themselves, the proof of Theorem II can be simplified considerably. The simplification is the following: Suppose that we are given a point x in $S(X)$. Let ρ_0 be the density matrix of maximum chaos among all density matrices ρ which satisfy the condition $\mathrm{Tr}\,(X\rho) = x$. Then ρ_0 must commute with *every* operator which commutes with all the operators X_k. Suppose that this were *not* the case for some unitary U which commutes with all the X_k. We may then form the density matrix $\rho_0' = U\rho_0 U^\dagger$. This density matrix also satisfies $\mathrm{Tr}\,(X\rho_0') = x$, and it has the same entropy as the density matrix ρ_0, which is a contradiction since, by Lemma I, the density matrix of maximum chaos is unique. The conclusion that ρ_0 must commute with every operator which commutes with all the operators X_k is valid in general, irrespective of whether the X_k commute or do not commute among themselves. If, however, we now consider the case when the operators X_k do commute among themselves, we may select a basis in $\mathfrak{K}$ with respect to which all the X_k are diagonal. By what has been said we conclude that ρ_0 must then be diagonal with respect to this basis and the diagonal elements of ρ_0 may be determined by a fairly simple application of calculus. This, then is a considerable simplification. However, we still have to show that, given any x in the interior of $S(X)$, there exists a unique y such that $\mathrm{Tr}\,[X\omega(y)] = x$.

10

We may here elaborate on and generalize the principle that the density matrix of maximum chaos must commute with any unitary operator which commutes with all the X_k. We present this extension in the form of a lemma.

Lemma VIII. Let $\mathcal{G}$ be any group of (real orthogonal) transformations on the real vector space V_h of all Hermitian matrices, such that each element of $\mathcal{G}$ is induced by either a unitary, or else an antiunitary transformation on the Hilbert space $\mathfrak{K}$, and let furthermore $\mathcal{G}$ have the property that for any $\mathcal{R}$ of $\mathcal{G}$ we have $\mathcal{R}X_k = X_k$ for all $k = 1, 2, \cdots, p$. Thus: If $\mathcal{R}_u$ is induced by a unitary transformation on $\mathfrak{K}$, we have $\mathcal{R}_u X = UXU^\dagger$ where U is unitary, whereas if $\mathcal{R}_a$ is induced by an antiunitary transformation on $\mathfrak{K}$, we have $\mathcal{R}_a X = A\tilde{X}A^\dagger$, where A is unitary, for every X in V_h.

(a) If ρ_0 is a density matrix of maximum chaos defined by any set of possible ensemble averages of the X_k, then $\mathcal{R}\rho_0 = \rho_0$ for all $\mathcal{R}$ in $\mathcal{G}$.

(b) Let V_h^I be the subspace of V_h consisting of all Hermitian operators Y such that $\mathcal{R}Y = Y$ for all $\mathcal{R}$ in $\mathcal{G}$. Let Z be any Hermitian operator orthogonal to V_h^I in the sense of the scalar product defined in Eq. (1). Then $\mathrm{Tr}\,(Z\rho_0) = 0$.

The proof of this lemma is trivial and need not be given in detail here. It is based on the fact that the three matrices ρ_0, $U\rho_0 U^\dagger$, and $A\tilde{\rho}_0 A^\dagger$, for any unitary matrices U and A, are of the same entropy, and have the same ensemble average of X.

We may add the remark that if Z is a matrix in the orthogonal complement of V_h^I, then the linear span of $\mathcal{R}Z$, as $\mathcal{R}$ ranges throughout $\mathcal{G}$, is the carrier space of a representation of $\mathcal{G}$ which does not contain the identity representation.

Lemma VIII is of a certain interest in practice because it first of all gives us some information about ρ_0, and secondly because it may aid us in the evaluation of expectation values of new operators in the most chaotic ensemble for the case when we can find, by inspection, a suitable group $\mathcal{G}$ satisfying the conditions of the lemma.

11

We may note the following circumstance: Let us imagine that we perform a sequence of measurements on the ensemble such that we first measure only Av (X_1); in the next step we also measure Av (X_2) but no other average, and so on, and that we construct a corresponding sequence of ensembles of maximum chaos subject to the restriction that Av (X_1), Av (X_2), $\cdots$, Av (X_k), where $k = 1, 2, \cdots, p$, are known and given. Then the corresponding entropies satisfy the conditions

$$s_1 \geq s_2 \geq \cdots \geq s_p,$$

as should be the case for a sensible physical interpretation. The determination of additional ensemble averages cannot lead to a decrease in the amount of information possessed about the states in the ensemble.

12

It is of interest to quote an expression for the entropy $s[\omega(y)]$ associated with the ensemble of maximum chaos. A simple computation gives the result

$$s[\omega(y)] = \ln\{\mathrm{Tr}\,[\exp(y\cdot X)]\} - y\cdot x. \qquad (17)$$

If we regard s as a function of x, then the change

[5] C. Kittel, *Elementary Statistical Physics* (John Wiley & Sons, Inc., New York, 1958), Part 1, Sec. 14.

in entropy of the ensemble of maximum chaos due to a change in x is given by the differential form

$$ds = -\sum_{k=1}^{p} y_k \, dx_k. \tag{18}$$

In the above formulas (17) and (18), $y = y(x)$ is the function of x given by the inverse of the mapping defined in Eq. (10).

We may also find an expression for the entropy $s(\rho)$ when ρ is in a neighborhood of the ensemble of maximum chaos $\omega(y)$, but subject to the constraints that $\mathrm{Tr}\ (X\rho) = \mathrm{Tr}\ [X\omega(y)] = x$, where x is a fixed point of $S(X)$. We may parametrize the matrices ρ in a neighborhood of $\omega(y)$ by writing

$$D(\Delta) = \{\mathrm{Tr}\ [\exp\ (y\cdot X + \Delta)]\}^{-1} \times \exp\ (y\cdot X + \Delta), \tag{19a}$$

where the otherwise arbitrary Hermitian matrix Δ is a parameter subject to the constraints

$$x_k\ \mathrm{Tr}\ [\exp\ (y\cdot X + \Delta)] = \mathrm{Tr}\ [X_k \exp\ (y\cdot X + \Delta)], \tag{19b}$$

for $k = 1, 2, \cdots, p$.

If we denote the entropy of $D(\Delta)$ by $s(\Delta) = s[D(\Delta)]$, we obtain, by a simple computation,[6]

$$s(\Delta) = s(0) + \frac{1}{2}\left\{\left[\frac{\mathrm{Tr}\ (\Delta e^{H})}{\mathrm{Tr}\ (e^{H})}\right]^2 - \frac{\mathrm{Tr}\left[\int_0^1 dt\ \Delta e^{tH}\ \Delta e^{(1-t)H}\right]}{\mathrm{Tr}\ (e^{H})}\right\} + O(\Delta^3), \tag{20}$$

where we have written $H = y\cdot X$. We have, of course, $D(0) = \omega(y)$, and $s(0) = s[\omega(y)]$, the entropy associated with the ensemble of maximum chaos $\omega(y)$. We note that in the expression (20), the term linear in Δ is missing, as it should be. It is also easily shown that the term quadratic in Δ is indeed negative for every Δ which is not a multiple of the identity, which confirms the fact that the entropy has a maximum at $\omega(y)$.

13

Let us finally consider the special case of Theorem II which arises in the situation discussed in Sec. 6(c). We shall follow the notation of that section. Suppose that x is any point in the interior of $S(X)$. Then the ensemble of maximum chaos, $\omega(y)$, is described by the density matrix

$$\omega(y) = \prod_{k=1}^{p} \{[\mathrm{Tr}\ (\exp\ [y_k\bar{X}_k])]^{-1} \exp\ [y_k\bar{X}_k]\} \otimes, \tag{21a}$$

where the components of the vector y are the unique solutions to the equations

$$x_k\ \mathrm{Tr}\ [\exp\ (y_k\bar{X}_k)] = \mathrm{Tr}\ [\bar{X}_k \exp\ (y_k\bar{X}_k)]. \tag{21b}$$

In the above formulas (21), the symbol "Tr" when applied to operators indexed by the integer k refers to the trace operation on the subspace $\mathfrak{K}_k$.

The ensemble of maximum chaos, $\omega(y)$, is therefore, not suprisingly, the tensor product of p ensembles of maximum chaos acting on the subspaces $\mathfrak{K}_k$ each one being defined by the average x_k of the operator $\bar{X}_k$. The same is true for the case when x is on the boundary of $S(X)$.

It may be noted that if P_k is any operator in the form of a tensor product of an operator $\bar{P}_k$ acting on $\mathfrak{K}_k$, and the identity operators acting on the remaining factors $\mathfrak{K}_m$, $m \neq k$, then $\mathrm{Tr}\ [P_k\omega(y)]$ depends on x_k only, but not on any other x_m for $m \neq k$.

DISCUSSION

14

We have thus found a complete and reasonably explicit solution to the problem which we posed in Sec. 2. for the special case that the Hilbert space $\mathfrak{K}$ of the state vectors of the quantum mechanical system is finite-dimensional. In the remainder of this note we will apply our results, heuristically, to some cases in which $\mathfrak{K}$ is not finite-dimensional. The author's excuse for the lapse in rigor is that, whereas the infinite-dimensional case certainly presents some entirely new problems, there will nevertheless be situations in which the almost straightforward generalization from the finite-dimensional case could be properly justified. Among the difficulties, one may mention the technical question of how to characterize, mathematically, a physically meaningful statistical ensemble. Physically such an ensemble is defined by a particular set of averages of the set of all possible observables, and here we actually mean by an observable, an operator which represents a bona fide measuring instrument which can, at least in principle, be realized physically.

It is clear that when measurements are performed in practice on any ensemble, say, for instance, in the case when the elements of the ensemble consist of "repeated" experiments, we never measure the ensemble averages of *all* imaginable dynamical variables. We may measure the averages of some, and have some vague ideas of reasonable limits on other variables. The variables about which we have only a vague knowledge might really be quite irrelevant to the situation at hand in the sense that

[6] See reference 4.

their averages are of no importance *provided* they stay within "reasonable limits." In such a situation we must find some rational way of describing the ensemble in terms of our knowledge about it. The author has tried to simplify the problem by considering the idealized situation in which *all* of our information about the ensemble can be expressed in the form of ensemble averages of a suitable set of operators, and under these circumstances the author wishes to advocate the view that the most rational choice of ensemble is the ensemble which is the most chaotic one consistent with our information, and that our prediction of ensemble averages of other operators which have not been measured should be based on this ensemble. The author believes that this view is neither radical nor new and original, but that it corresponds to fairly commonly held beliefs, and that it is in the spirit of the approach to statistical mechanics followed by, for instance, Tolman in his well-known text on the subject.[7]

A rigorous mathematical formulation of this principle is, however, not entirely trivial if $\mathfrak{K}$ is infinite-dimensional, as it is clear that unless the set of operators, the ensemble averages of which are given, satisfy a number of restrictive conditions, there will be no sensible ensemble of maximum chaos associated with the given averages.

To prevent any possible misunderstanding, the author wishes to state emphatically that the theory presented so far must be applied with judgment and proper caution in any practical case; otherwise it is meaningless and misleading. If we have measured a number of ensemble averages and then wish to make the assumption that the ensemble is the most chaotic one consistent with these averages, we can do this only if we are reasonably certain that we have measured all the "important" variables. i.e., those which appear to be relevant for the application at hand, and only if we honestly do not have any knowledge of the maybe very large number of remaining variables, none of which seems crucial to the effect we wish to predict. The same applies, of course, in the case where we wish to construct an ensemble *theoretically*, such that the ensemble is parametrized by the ensemble averages of all the "important" observables.

We must, in other words, exercise the same kind of caution in the general case as we have to exercise in the special case of, say, the canonical ensemble.

[7] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, London, England, 1962), Chap. IX.

APPLICATION TO A PARTICLE IN ONE-DIMENSIONAL MOTION

15

We shall now apply our general theory to the case of a single particle in one-dimensional motion. Let Q be the position variable and let P be the momentum variable of the particle. These variables satisfy the commutation relation

$$[Q, P] = i. \tag{22}$$

The spectrum of each one of these variables consists of the real line. We thus deal with an irreducible representation of the commutation relation (22) such that Q and P are Hermitian, or, more precisely, we deal with an irreducible unitary representation of the corresponding group of translations in coordinate and momentum space. As is well-known, this representation is unique up to unitary equivalence.

Let us now assume that we have an ensemble for which we have measured the following ensemble averages:

$$\text{Av}\,(Q) = x_q, \qquad \text{Av}\,(P) = x_p, \tag{23a}$$

$$\text{Av}\,(Q^2) = x_{qq}, \qquad \text{Av}\,(P^2) = x_{pp},$$

$$\text{Av}\,(QP + PQ) = 2x_{qp}, \tag{23b}$$

but no other averages. The ensemble of maximum chaos, denoted by ρ_0, defined by these measurements must then be of the form

$$\rho_0 = C \exp\,(G), \tag{24}$$

where C is a normalization constant, and where G is second-degree polynomial in Q and P; G must of course be a Hermitian operator.

Depending on the form of G, we may have several cases: G has a nondegenerate point spectrum (equally spaced) bounded above or bounded below; G has a continuous spectrum consisting of a half line, or G has the entire real line as its spectrum. Only in the first case will the trace of ρ_0 exist, and a moment's reflection shows that G can then be written in the form

$$G = -g\mathbf{a}^\dagger\mathbf{a}; \qquad g > 0, \tag{25}$$

where we have omitted an additive constant which may be absorbed in the constant C, and where the operator $\mathbf{a}$ is an expression linear in Q and P satisfying the condition

$$[\mathbf{a}, \mathbf{a}^\dagger] = 1. \tag{26}$$

The operator $\mathbf{a}$ is thus a lowering operator, whereas its Hermitian conjugate $\mathbf{a}^\dagger$ is a raising

operator. By a simple computation we obtain the following results:

$$\rho_0 = (1 - e^{-g}) \exp(-g\mathbf{a}^\dagger\mathbf{a}), \tag{27a}$$

$$\mathrm{Tr}\,(\rho_0\mathbf{a}) = \mathrm{Tr}\,(\rho_0\mathbf{a}^\dagger) = 0, \tag{27b}$$

$$\mathrm{Tr}\,(\rho_0\mathbf{a}\mathbf{a}) = \mathrm{Tr}\,(\rho_0\mathbf{a}^\dagger\mathbf{a}^\dagger) = 0, \tag{27c}$$

$$\mathrm{Tr}\,(\rho_0\mathbf{a}^\dagger\mathbf{a}) = K = (e^g - 1)^{-1}, \tag{27d}$$

where, of course, the normalization constant C has been selected such that ρ_0 is of unit trace.

The operators Q and P are Hermitian linear expressions in $\mathbf{a}$ and $\mathbf{a}^\dagger$, and we may write

$$Q = u^*\mathbf{a}^\dagger + u\mathbf{a} + x_q; \quad P = w^*\mathbf{a}^\dagger + w\mathbf{a} + x_p, \tag{28}$$

where u and w are (complex) constants. We note that the ensemble averages of Q and P satisfy Eqs. (23a) if and only if these operators are of the form given in Eqs. (28). In order that Eqs. (22) and (26) shall both hold, u and w must satisfy the condition

$$uw^* - wu^* = i. \tag{29a}$$

We may now compute the ensemble averages of Q^2, P^2 and $(QP + PQ)$, making use of Eqs. (27), and we obtain

$$\mathrm{Av}\,(Q^2) - [\mathrm{Av}\,(Q)]^2 = |u|^2\,(2K + 1) = x_{qq} - x_q^2, \tag{29b}$$

$$\mathrm{Av}\,(P^2) - [\mathrm{Av}\,(P)]^2 = |w|^2\,(2K + 1) = x_{pp} - x_p^2, \tag{29c}$$

$$\mathrm{Av}\,(QP + PQ) - 2[\mathrm{Av}\,(Q)][\mathrm{Av}\,(P)] = (uw^* + wu^*)(2K + 1) = 2(x_{qp} - x_q x_p). \tag{29d}$$

The condition that the four equations (29) in the unknowns u, w, and K shall have solutions such that $K > 0$ is equivalent to the condition that the ensemble averages x_q, x_p, x_{qq}, x_{pp}, and x_{qp} are in the set S of possible simultaneous ensemble averages. By inspection of the equations we note that the argument of the complex number u is arbitrary, and without loss of generality we may select u to be positive. We then get

$$\begin{aligned} 2\,|uw|\sin\theta &= -1, \\ u &= [(x_{qq} - x_q^2)/(2K + 1)]^{\frac{1}{2}}, \\ w &= e^{i\theta}[(x_{pp} - x_p^2)/(2K + 1)]^{\frac{1}{2}}, \\ [(x_{qq} - x_q^2)(x_{pp} - x_p^2)]^{\frac{1}{2}}\cos\theta &= x_{qp} - x_q x_p, \end{aligned} \tag{30}$$

and

$$(x_{qq} - x_q^2)(x_{pp} - x_p^2) - (x_{qp} - x_q x_p)^2 = \tfrac{1}{4}(2K + 1)^2 = \tfrac{1}{4}[(e^g + 1)/(e^g - 1)]^2. \tag{31}$$

The conditions that $\{x_q, x_p, x_{qq}, x_{pp}, x_{qp}\}$ be an *interior* point of S are thus

$$x_{qq} - x_q^2 > 0, \qquad x_{pp} - x_p^2 > 0, \tag{32a}$$

$$(x_{qq} - x_q^2)(x_{pp} - x_p^2) - (x_{qp} - x_q x_p)^2 > \tfrac{1}{4}. \tag{32b}$$

If, and only if these conditions are satisfied do we have $g > 0$ and $K > 0$, and a real argument θ of w. That the inequalities (32a) must hold is obvious. The inequality (32b) is a generalization of the familiar uncertainty relation for Q and P.

Computation of the entropy $s(\rho_0)$ gives the result

$$s(\rho_0) = (K + 1)\ln(K + 1) - K\ln K. \tag{33}$$

As K tends to zero, ρ_0 tends to a primitive Hermitean idempotent; the parameter g tends to infinity and the entropy tends to zero. The state vector defined by this idempotent is a Kennard packet, and in this limiting case we have an equality instead of the inequality (32b).

16

Let us briefly consider what happens if for the physical system described in the preceding section we only measure the ensemble averages of Q^2 and P^2. Hence

$$\mathrm{Av}\,(Q^2) = x_{qq} > 0, \qquad \mathrm{Av}\,(P^2) = x_{pp} > 0. \tag{34}$$

The problem of finding the ensemble of maximum chaos is easy to solve directly. We may, however, also find the solution from the results of the preceding section provided we apply the considerations of Sec. 10 and the facts stated in Lemma VIII. We know that there may be defined a unitary reflection transformation R such that when R acts on Q and P, it changes the sign of these operators. On the other hand, the operators Q^2 and P^2 are invariant under R, and it follows, by Lemma VIII, that the ensemble averages of Q and P must vanish in the ensemble of maximum chaos defined by the averages given in Eqs. (34). Furthermore, we may define an antiunitary time-reversal transformation T which leaves Q, Q^2, and P^2 invariant, but changes the sign of P and $(QP + PQ)$. By Lemma VIII, the ensemble average of $(QP + PQ)$ must therefore vanish in the ensemble of maximum chaos. The results of the preceding section are thus directly applicable provided we set $x_q = x_p = x_{qp} = 0$. The region S of possible simultaneous ensemble averages of Q^2 and P^2 is the region

$$x_{qq} > 0, \qquad x_{pp} > 0, \qquad x_{qq}x_{pp} \geq \tfrac{1}{4}, \tag{35}$$

where, in the last inequality, the equality holds if

and only if the ensemble is the pure state corresponding to a Kennard packet.

We may note that the problem posed in Sec. 15 can always be reduced to the problem of the present section by performing a suitable inhomogeneous orthogonal transformation on the variables Q and P.

The trivial examples presented in this and the preceding section illustrate the modifications which arise when the Hilbert space $\mathfrak{H}$ is infinite-dimensional. The region S need no longer be closed and bounded although it must of course be convex. It can still be found by examining the spectrum of linear combinations of the operators $X_1, X_2, \cdots, X_p$, although the detailed prescription given in Theorem I must be appropriately modified.

It is clear that our discussion can be generalized to the case of N operators Q, and N operators P. The corresponding ensembles of maximum chaos may be regarded as the quantum mechanical analogues of classical Gaussian ensembles.

APPLICATION TO MANY-BOSON SYSTEMS

17

Let us consider a many-boson theory, and let $\mathbf{a}_k^\dagger$, $k = 1, 2, \cdots$, be a complete orthonormal set of creation operators which thus satisfy the commutation relations

$$[\mathbf{a}_n, \mathbf{a}_k^\dagger] = \delta_{nk}. \tag{36}$$

Suppose that, for a given ensemble, we have measured the ensemble averages of *all* operators of the form $\mathbf{a}_m^\dagger \mathbf{a}_n$, but no other ensemble averages. Let

$$\mathrm{Av}\,(\mathbf{a}_m^\dagger \mathbf{a}_n) = D^{(1)}_{nm}, \tag{37}$$

for all m and n. We call operators of this kind *one-particle operators*. The (infinite) matrix $D^{(1)}$ is Hermitian and nonnegative definite. We shall determine the most chaotic ensemble, ρ_0, among all ensembles satisfying the conditions (37).

To simplify the discussion, we select a basis in the Hilbert space of creation operators with respect to which $D^{(1)}$ is diagonal. Let the creation operators forming this basis be denoted by $\bar{\mathbf{a}}_k^\dagger$, $k = 1, 2, \cdots$. This basis is related to the original one by a unitary transformation U, and we have, explicitly,

$$\mathbf{a}_k^\dagger = \sum_n U_{nk} \bar{\mathbf{a}}_n^\dagger \quad \text{where} \quad UU^\dagger = U^\dagger U = I, \tag{38a}$$

$$D^{(1)} = U^\dagger \bar{D}^{(1)} U \tag{38b}$$

$$\bar{D}^{(1)}_{mn} = \mathrm{Av}\,(\bar{\mathbf{a}}_n^\dagger \bar{\mathbf{a}}_m) = \delta_{mn}\alpha_n; \qquad \alpha_n \geq 0. \tag{38c}$$

The ensemble of maximum chaos, subject to the constraints expressed by Eq. (38c), must be of the form

$$\rho_0 = C \exp\left(-\sum_{m,n} \bar{\mathbf{a}}_m^\dagger \bar{G}_{mn} \bar{\mathbf{a}}_n\right) = C \exp\left(-\sum_{m,n} \mathbf{a}_m^\dagger G_{mn} \mathbf{a}_n\right), \tag{39a}$$

where $\bar{G}$ and G are related by U, i.e.,

$$G = U^\dagger \bar{G} U. \tag{39b}$$

The matrix $\bar{G}$ must be Hermitian and positive definite. If we now apply the considerations of Sec. 10, we see that $\bar{G}$ must commute with every matrix which commutes with $\bar{D}^{(1)}$, and hence $\bar{G}$ must be a diagonal matrix,

$$\bar{G}_{mn} = \delta_{mn}\beta_n; \qquad \beta_n > 0. \tag{40}$$

It is now a trivial problem to determine the normalization constant C as well as the numbers β_n. We shall merely list the results:

$$C = \prod_n (1 - e^{-\beta_n}) = \prod_n (1 + \alpha_n)^{-1}, \tag{41a}$$

$$\alpha_n = (e^{\beta_n} - 1)^{-1}, \qquad \beta_n = \ln\,[(1 + \alpha_n)/\alpha_n], \tag{41b}$$

$$\bar{D}^{(2)}_{(mn)(pq)} = \mathrm{Tr}\,(\rho_0 \bar{\mathbf{a}}_p^\dagger \bar{\mathbf{a}}_q^\dagger \bar{\mathbf{a}}_m \bar{\mathbf{a}}_n) = \bar{D}^{(1)}_{mp}\bar{D}^{(1)}_{nq} + \bar{D}^{(1)}_{mq}\bar{D}^{(1)}_{np}. \tag{41c}$$

These formulas, of course, make sense only if the infinite products in the formulas (41a) converge.

If we now go back to the original basis system we may write our results in the form

$$\rho_0 = C \exp\left(-\sum_{m,n} \mathbf{a}_m^\dagger G_{mn} \mathbf{a}_n\right), \tag{42a}$$

$$\ln C = -\mathrm{Tr}\,\{\ln\,[I + D^{(1)}]\}, \tag{42b}$$

$$G = \ln\,[(D^{(1)} + I)/D^{(1)}]; \qquad D^{(1)} = [\exp(G) - I]^{-1}, \tag{42c}$$

$$D^{(2)}_{(mn)(pq)} = \mathrm{Tr}\,(\rho_0 \mathbf{a}_p^\dagger \mathbf{a}_q^\dagger \mathbf{a}_m \mathbf{a}_n) = D^{(1)}_{mp}D^{(1)}_{nq} + D^{(1)}_{mq}D^{(1)}_{np}. \tag{42d}$$

The formula (42d) is very interesting. We shall call operators of the form $\mathbf{a}_p^\dagger \mathbf{a}_q^\dagger \mathbf{a}_m \mathbf{a}_n$ *two-particle operators*. We see that all operators of this form are given by the matrix $D^{(2)}$. This matrix again, is, for the ensemble of maximum chaos, equal to twice the symmetrized Kronecker product of $D^{(1)}$ with itself. This result can easily be generalized, and the corresponding n-particle operator, being a product of n creation operators standing to the right of n destruction operators, is essentially given by the n-fold symmetrized Kronecker product of $D^{(1)}$ with itself. The ensemble average of any product of n' creation operators and n'' destruction operators vanishes trivially for all $n' \neq n''$.

18

We wish to comment briefly on the case when the ensemble averages of some, but not all one-particle operators have been measured. Suppose thus that we have measured the ensemble averages of the operators $\mathfrak{A}^{(k)}$, $k = 1, 2, \cdots$, where

$$\mathfrak{A}^{(k)} = \sum_{m,n} \mathbf{a}_m^\dagger A^{(k)}_{mn} \mathbf{a}_n. \tag{43}$$

The ensemble of maximum chaos is then of the form

$$\rho_0 = C \exp\left(-\sum_{m,n} \mathbf{a}_m^\dagger G_{mn} \mathbf{a}_n\right), \tag{44}$$

where the matrix G is a linear combination of the matrices $A^{(k)}$. The actual form of G may in principle be determined from the equations

$$\mathrm{Tr}(\rho_0 \mathfrak{A}^k) = \mathrm{Tr}\,\{A^{(k)}[\exp(G) - I]^{-1}\}. \tag{45}$$

The matrix $D^{(1)}$ and the matrix $D^{(2)}$ are then given by Eqs. (42c) and 42d).

As a familiar example of this procedure, we may consider the case when we have determined the ensemble averages of the two operators

$$\mathfrak{N} = \sum_k \mathbf{a}_k^\dagger \mathbf{a}_k, \qquad \mathfrak{K} = \sum_{m,n} \mathbf{a}_m^\dagger H_{mn} \mathbf{a}_n,$$

but no other ensemble averages. The matrix H here stands for the Hamiltonian in the one-particle theory. The ensemble of maximum chaos is then, as is well-known, the grand canonical ensemble for a "gas" of noninteracting bosons.

As a possible application of the theory presented in this and the preceding section, we may consider a beam of particles, and in particular a beam of photons. Suppose that we have measured a certain number of one-particle operators, and, among other things, determined the spectral distribution of the number of one-particle operators, and determined the spectral distribution of the photons, as well as the intensity in various regions in space. For the case when the photon beam is produced by a gas discharge tube, a star, or any "ordinary" light source, the author would be willing to assume that the beam can be described by the ensemble of maximum chaos defined by the ensemble averages of all the one-photon operators which have in fact been measured. There are certainly photon beams which cannot be so described. However, if no special effort has been made to build strong many-photon correlations into the beam, then the assumption stated seems reasonable. It should be pointed out that the formula (42d) expresses the correlations between *two* bosons typical of Bose statistics.

APPLICATION TO MANY-FERMION SYSTEMS

19

We shall next present, for the case of a many-fermion system, results analogous to those presented in Secs. 17 and 18. Let the fermion operators $\mathbf{a}_k^\dagger$, $k = 1, 2, \cdots$, form a complete orthonormal basis of creation operators. These operators thus satisfy the anticommutation relations

$$\{\mathbf{a}_m, \mathbf{a}_n^\dagger\} = \mathbf{a}_m \mathbf{a}_n^\dagger + \mathbf{a}_n^\dagger \mathbf{a}_m = \delta_{mn}. \tag{46}$$

Suppose now that we have determined the ensemble averages of all one-particle operators, and that these averages are given by the Hermitian matrix $D^{(1)}$:

$$D^{(1)}_{mn} = \mathrm{Av}\,(\mathbf{a}_n^\dagger \mathbf{a}_m). \tag{47}$$

The (infinite) matrix $D^{(1)}$ is Hermitian and nonnegative definite. Since we deal with fermions, the matrix $(I - D^{(1)})$ is also nonnegative definite.

The ensemble of maximum chaos, denoted by ρ_0, is given by

$$\rho_0 = C \exp\left(-\sum_{m,n} \mathbf{a}_m^\dagger G_{mn} \mathbf{a}_n\right), \tag{48a}$$

where

$$\ln C = \mathrm{Tr}\,\{\ln [I - D^{(1)}]\}, \tag{48b}$$

$$D^{(1)} = [I + \exp(G)]^{-1}; \qquad G = \ln\{[I - D^{(1)}]/D^{(1)}\}. \tag{48c}$$

The averages of two-particle operators in the ensemble of maximum chaos are given by

$$D^{(2)}_{(mn)(pq)} = \mathrm{Tr}\,(\rho_0 \mathbf{a}_q^\dagger \mathbf{a}_p^\dagger \mathbf{a}_m \mathbf{a}_n) = D^{(1)}_{mp} D^{(1)}_{nq} - D^{(1)}_{mq} D^{(1)}_{np}. \tag{49}$$

The above results are easily derived by a method analogous to the one employed in Sec. 17, and we may omit the details.

In the case of a fermion system, the matrix $D^{(2)}$ is thus twice the antisymmetrized Kronecker product of the matrix $D^{(1)}$ with itself, as we might have expected. In a similar manner we may find the ensemble averages of any product of n creation operators and n destruction operators. The ensemble averages of any product of n' creation operators and n'' destruction operators vanish for all $n' \neq n''$.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor B. Friedman, Dr. A, Glassgold, and Dr. G. Baym of the University of California, Berkeley, for advice and helpful discussion.

Journal
of
Applied Physics

Volume 37, Number 11 October 1966

Selected Papers in High-Polymer Physics

The following eighteen papers were presented as part of the program of the Division of High-Polymer Physics at the meeting of The American Physical Society 28–31 March 1966, Durham, North Carolina

Effect of Chain Conformations on the Frequency Distribution of an Idealized Chain*

G. JANNINK AND G. C. SUMMERFIELD
Department of Nuclear Engineering, The University of Michigan, Ann Arbor, Michigan

The frequency distribution of a simplified carbon skeleton chain is calculated for several conformations, given by successions of trans- and gauche structures in various orders. It is found that the phonon frequency distribution is sensitive to these orders. Validity of the model and use of the results are discussed.

I. INTRODUCTION

An important characteristic of a large system of atoms is the phonon frequency distribution, that is, the number of vibrational normal modes per unit frequency interval. This function will reflect both the interatomic forces and the configuration of the atomic system.

This function is directly related to a number of measurable properties of an atomic system, for example, the specific heat[1] and the cross section for inelastic neutron scattering.[2]

When the atomic system is arranged in a regular lattice with a few defects, the frequency distribution can be determined fairly easily by using the method of Green's functions.[3] The frequency distribution can also be computed by other methods[4] when certain types of defects occur at many locations in the lattice, as for example, in a binary alloy.

Here we determine the dependance of the frequency distribution on the conformation for a chain of atoms. The variation of infrared and Raman active modes as a function of conformation has been given for a number of models.[5] The model of the chain which is considered here looks like the carbon skeleton of polyethylene with the following simplifying assumptions (Fig. 1): The "in-plane" motion is approximated by the superposition of three independent motions—two longitudinal motions of the even and odd rows, respectively, and one zigzag motion along C–C bonds. The "out-of-plane" motion is approximated by the superposition of two independent transverse motions of even and odd rows, respectively.

The angle of internal rotation ϕ_j is defined with respect to the plane of vertices, j, $j-1$, $j-2$. From the geometry of the figure we have

$$\cos\beta_j = 1-\sin^2\alpha(1-\cos\phi_j),$$

where 2α is the valence angle.

To each vertex j one associates a Cartesian coordinate system: The x axis joins the $j-2$ and j vertices, the y and z axes are, respectively, in and orthogonal to the j, $j-1$, $j-2$ plane.

The equations for the harmonic motion of the vertices

* Work partially supported by the National Science Foundation.

[1] B. Wunderlich, J. Chem. Phys. 37, 1203 (1962).

[2] W. Myers, G. C. Summerfield, and J. S. King, J. Chem. Phys. 44, 184 (1966).

[3] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, New York, 1963), pp. 137, 178.

[4] P. Dean, Proc. Phys. Soc. (London) 73, 413 (1959).

[5] D. E. Witenhafer and J. L. Koenig, Bull. Am. Phys. Soc. 11, 232 (1966).

are written in internal coordinates. For the bending mode, the passage from Cartesian displacement x to internal displacement l is illustrated in Fig. 2. Similar transformations define s and t in:

$$m\ddot{l}_j = -\gamma_b \cos\beta_{j-2} l_{j-2} + 2\gamma_b l_j - \gamma_b \cos l_{j+2} \text{ (bending)},$$
$$m\ddot{s}_j = -\gamma_s \cos 2\alpha s_{j-1} + 2\gamma_s s_j - \gamma_s \cos 2\alpha s_{j+1} \text{ (stretching)},$$
$$m\ddot{t}_j = -\gamma_t |\cos\phi_{j-2}| t_{j-2} + 2\gamma_t t_j - \gamma_t |\cos\phi_{j+2}| t_{j+2} \text{ (out of plane)}. \quad (1)$$

A conformation is given by a sequence $\{\phi_i\}$, $i=1, \cdots, N$. The effect of a change in conformation is to modify the orientation of the interatomic forces; we assume, however, that the magnitude of binary interaction is invariant under changes in conformations. We consider two cases: (1) ϕ_i can have one of the two values, 0 or ϕ, and (2) ϕ_i can have any value in the interval $(0, 2\pi)$. In case (1) we expect results analogous to single mass defects and binary alloys. In case (2) we expect results similar to level distributions for a random matrix.[6]

When every ϕ_i is zero, the chain is extended. When only one element is different from zero, it is called singly folded. When more than one element is different from zero, it is called multiply folded.

II. CALCULATIONS FOR SINGLY FOLDED CHAINS

If $M_0(\omega^2) = A_0 - \omega^2 I$, where A_0 is the dynamical matrix of the extended structure with cyclic boundary conditions, there corresponds to the singly folded chain a matrix M which, according to the Green's function method, is written as

$$M(\omega^2) = M_0(I + M_0^{-1} D).$$

The frequencies are the roots of

$$|I + M_0^{-1}(\omega^2) D| = 0. \quad (2)$$

In our approximation the "defect" matrix has the form

$$D = \begin{bmatrix} \cdots & \vdots & \vdots & \cdots \\ \cdots & 0 & d_{j-1,j} & \cdots \\ \cdots & d_{j,j-1} & 0 & \cdots \\ & \vdots & \vdots & \end{bmatrix}. \quad (3)$$

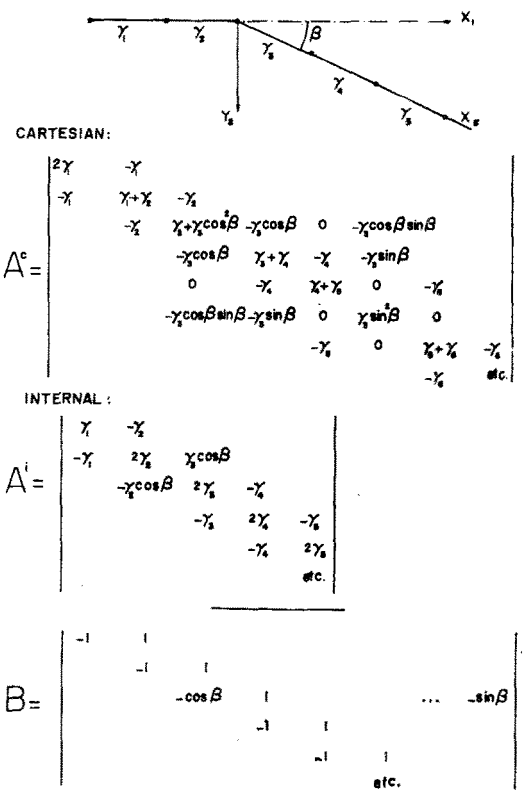

FIG. 2. Dynamical matrix in Cartesian and internal displacements coordinates for the bending motion. B matrix.

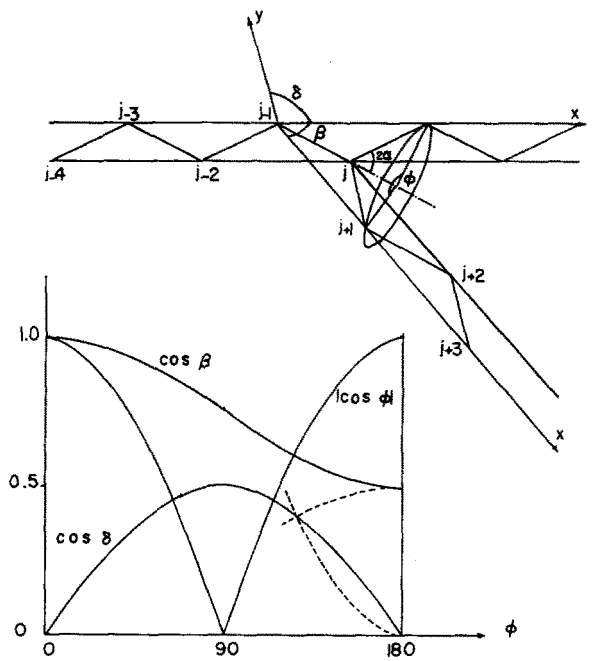

FIG. 1. Trans- and gauche conformations of the carbon skeleton for $2\alpha = 60°$. The diagram at the bottom is the variation of force orientations as a function of internal rotation angle; the dotted lines are drawn symmetrically to $\cos\beta$ and $|\cos\phi|$, respectively, and intersect on the $\cos\delta$ curve.

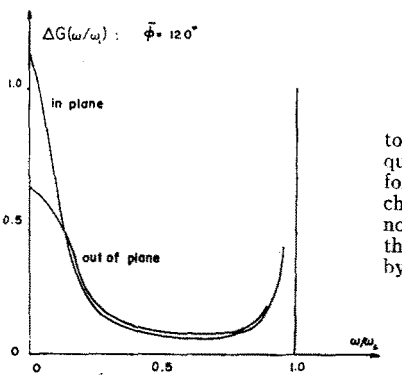

FIG. 3. Correction to the extended frequency distribution for a singly folded chain. The exact normalization of these curves is given by Eq. (7).

[6] C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic Press, New York, 1965), pp. 239, 298.

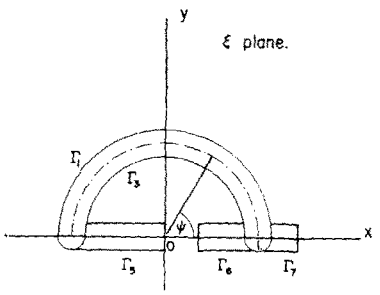

FIG. 4. Integration contour for Eq. (11).

For the bending mode:

$$d_{j-1,j}=d_{j,j-1}=\gamma_b(1-\cos\beta_j).$$

For the transverse mode, this matrix element is

$$d_{j-1,j}=d_{j,j-1}=\gamma_t(1-|\cos\phi_j|).$$

The stretch mode is not affected by changes in conformation in this model. The dependance of the stretch frequency on a valence-angle defect is given later. The elements of the matrix are the Green's function of the extended conformation; they are written:

$$M_{0ij}^{-1}=g(n), \quad n=|i-j|.$$

In this case,

$$g(n)=\begin{cases}\dfrac{\cot gN\theta/4}{2\gamma_b\sin\theta}\cos n\theta/2-\dfrac{\sin|n|\theta/2}{2\gamma_b\sin\theta}, & n \text{ even}\\ 0, & n \text{ odd}\end{cases} \quad (4)$$

$$\sin^2\theta/2=\omega^2/\omega_L^2; \quad \omega_L^2=4\gamma.$$

For the bending and out-of-plane modes, the frequency distribution for the extended structure is

$$G_0(\omega)=\sum_{mr} G_0^{mr}, \quad (5)$$

where m is the index of the mode (in and out of plane), and r is the index of the row (upper and lower).

$$G_0^{mr}=1/\omega_L(1-\omega^2/\omega_L^2)^{\frac{1}{2}}. \quad (6)$$

A single fold introduces a correction to the frequency distribution of the form:

$$\Delta G^{mr}(\omega)=f_{ms}G_0^{mr}-(\omega/N)[\delta(\omega^2-\omega_2^2)+\delta(\omega^2)]$$
$$+\begin{cases}0, & \nu<0\\ \dfrac{2\omega(1+\nu)}{N}\delta\{\omega^2-(\omega_2^2/2)[1+(1+\nu/2)(1+\nu)^{-\frac{1}{2}}]\}, & \nu>0,\end{cases} \quad (7)$$

so that the frequency distribution of a singly folded chain excluding the stretch mode is

$$G(\omega)=G_0+\sum_{mr}\Delta G^{mr}, \quad (8)$$

TABLE I. Correspondence between ω and ξ.

ω^2	ξ	Contour
0 to $(\omega_L^2/2)(1-\cos2\alpha)$	$1/\cos2\alpha\pm(1/\cos^2 2\alpha-1)$ to 1	Γ_6,Γ_7
$(\omega_L^2/2)(1-\cos2\alpha\cos\psi)$	$e^{i\psi}$; $0\leq\psi\leq\pi$	$\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4$
$(\omega_L^2/2)(1+\cos2\alpha)$ to ∞	-1 to 0^-	Γ_5

where

$$f_{mr}=-(1+\nu/2)\nu/N\pi[\nu^2+32(\omega^2/\omega_L^2)(1-\omega^2/\omega_L^2)] \quad (9)$$

and

$$\nu=-\sin^2\alpha(1-\cos\phi)[2-\sin^2\alpha(1-\cos\phi)]$$

for bending motion, and

$$\nu=\cos^2\phi-1$$

for transverse motion. In all cases $\nu\leq 0$.

The effect of folding is to reduce the singularities of $G(\omega^2)$ at $\omega=0$ and $\omega=\omega_L$, and to raise the frequency distribution in the central part of the $(0,\omega_L)$ interval (Fig. 3).

For the transverse mode, this effect reaches a maximum at $\phi=\pi/2$; the left and right branches move independently. For the bend mode the effect is greatest at $\phi=\pi$.

The effect of a valence-angle defect on the stretch mode is determined in a similar way; the defect matrix elements are

$$d_{j-1,j}=d_{j,j-1}=\gamma_s(\cos2\alpha-\cos2\alpha'),$$

where α' is the "defect" value.

Relations (2) and (3) remain valid. The Green's function, however, has an additional factor $\cos2\alpha$ in the denominator which modifies the integration path given in Mahanty *et al.*[7]

The Fourier transform of the square frequency distribution is determined in this reference from

$$\Delta F(\alpha)=\int_c e^{i\alpha\omega^2}d\ln|I+M_0^{-1}D|. \quad (10)$$

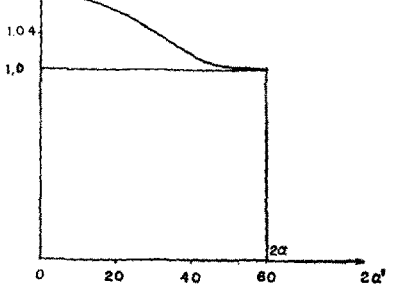

FIG. 5. Separation of the out-of-band stretching modes from the continuum as a function of valence-angle defect See Eq. (14).

[7] J. Mahanty, A. A. Maradudin, and G. H. Weiss, Progr. Theoret. Phys. (Kyoto) **20**, 369 (1958).

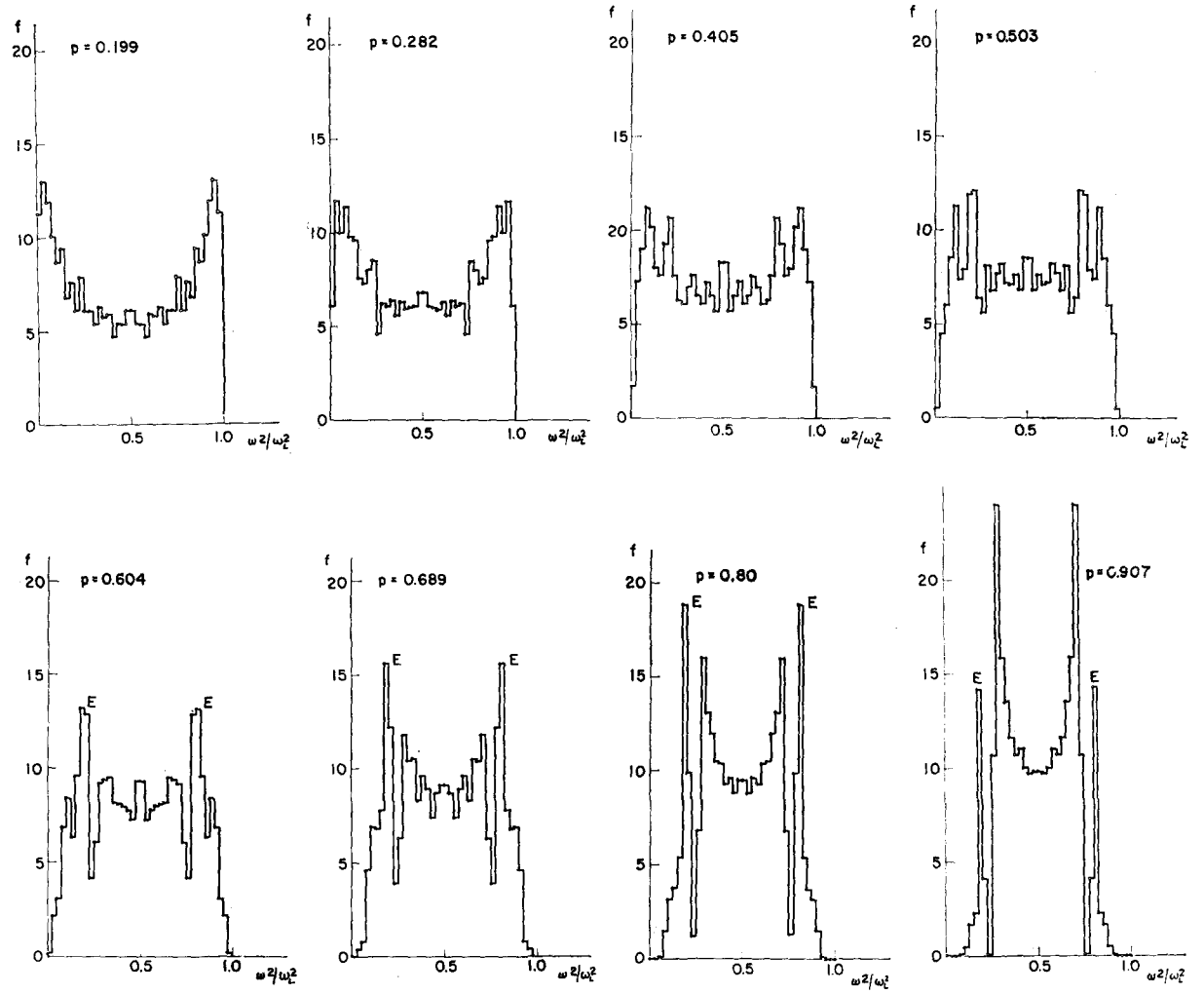

FIG. 6. Frequency distribution for several concentrations of randomly distributed gauche conformations.

The appropriate change of variable is here

$$\omega^2=(\omega_L{}^2/4)[2-\cos 2\alpha(\xi-\xi^{-1})]. \tag{11}$$

The correspondence is given in Fig. 4 and Table I.

The out-of-band mode above the continuum arises in the Γ_5 integration, and the out-of-band mode below the continuum arises in the Γ_6 integration. Thus the Green's function is

$$g(n)=\begin{cases}-(1/\gamma_s\cos 2\alpha)[\xi^n/(\xi-\xi^{-1})], & |\xi|<1\\ -(1/\gamma_s\cos 2\alpha)[\xi^{-n}/(\xi^{-1}-\xi)], & |\xi|>1.\end{cases} \tag{12}$$

The frequency distribution of the unperturbed chain is

$$G_{0s}=(\omega/\pi\omega_L{}^2)[\cos^2 2\alpha-(1-2\omega^2/\omega_L{}^2)^2]^{-\frac{1}{2}}. \tag{13}$$

The additional term due to the defect is

$$\Delta G_s(\omega)=fG_{0s}-\frac{\omega}{2N}\left\{\delta[\omega^2-\omega_L{}^2(1+\cos 2\alpha)]+\delta\left[\omega^2-\frac{\omega_L{}^2}{2}(1-\cos 2\alpha)\right]\right\} \tag{14}$$

$$+\begin{cases}0, & \mu<0\\ \dfrac{2\omega(1+\mu)}{N}\left(\delta\left\{\omega^2-\dfrac{\omega_L{}^2}{2}[1+\cos 2\alpha(1+\mu/2)(1+\mu)^{-\frac{1}{2}}]\right\}+\delta\left\{\omega^2-\dfrac{\omega_L{}^2}{2}[1-\cos 2\alpha(1+\mu/2)(1+\mu)^{-\frac{1}{2}}]\right\}\right), & \mu>0\end{cases}$$

$$f=(1+\mu/2)2\mu/N[\mu^2+(8/\cos^2 2\alpha)][\cos^2 2\alpha-(1-2\omega^2/\omega_L{}^2)^2]$$

$$\mu=\cos^2 2\alpha'/\cos^2 2\alpha-1. \tag{15}$$

Then the contribution to the frequency distribution from the stretch mode for a singly folded chain is

$$G_s(\omega)=G_{0s}+\Delta G_s.$$

For $\alpha'<\alpha$, the valence defect reduces the singularities at $(\omega_L/2)(1\pm\cos 2\alpha)$. In the range $\alpha'>\alpha$, the singularities are "enhanced." Also, two out-of-band modes appear: one lower than the lowest, the other higher than the greatest unperturbed modes.[8] The separations $(\omega_0/\omega_L)^2$ are given as a function of α' in Fig. 5.

III. CALCULATIONS FOR MULTIPLY FOLDED CHAINS

Our model includes only nearest-neighbor forces and we neglect interactions between modes. Consequently, the dynamical matrix for any conformation is in tridiagonal form (Fig. 2). The eigenvalue distribution f of such matrices is known to be easily calculated by use of Sturm sequences. This method is explained in detail elsewhere,[4] and we do not go into the details of its application. However, it simply is a method of computing the number of eigenvalues of the dynamical matrix in an interval of the variable ω^2, in terms of the principal determinants evaluated at the end points of the interval.

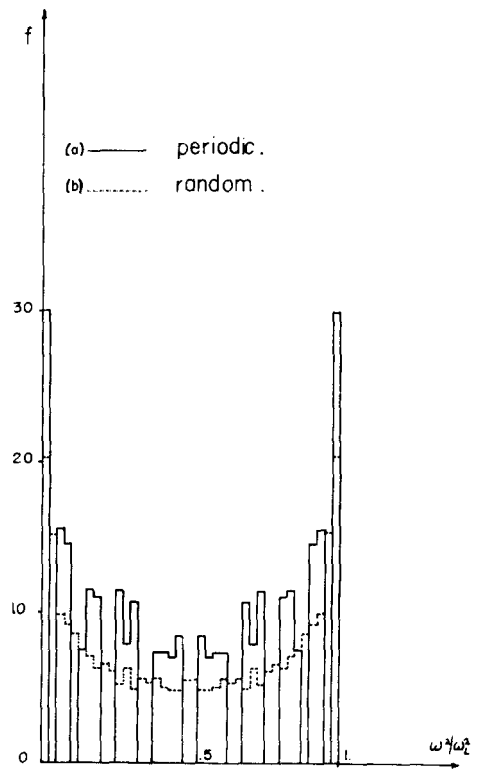

FIG. 7. Comparison of frequency distributions for (a) regularly spaced folds (ten $\phi_i=0$ followed by one $\phi_i=120°$) and (b) a random distribution of 120° folds occurring with a frequency $p=0.089$.

[8] It is interesting to note that if we had used fixed, rather than cyclic, boundary conditions, we would have obtained two extra out-of-band modes. Both of these modes appear well separated from and below the continuum and they appear even for the extended conformation. Fixed boundary conditions correspond to an infinite mass defect; and it is well known[3] that this defect does not produce out-of-band modes above the continuum. However due to the zigzag geometry, we have two modes below the continuum. If we use free boundary conditions, we get no extra out-of-band modes.

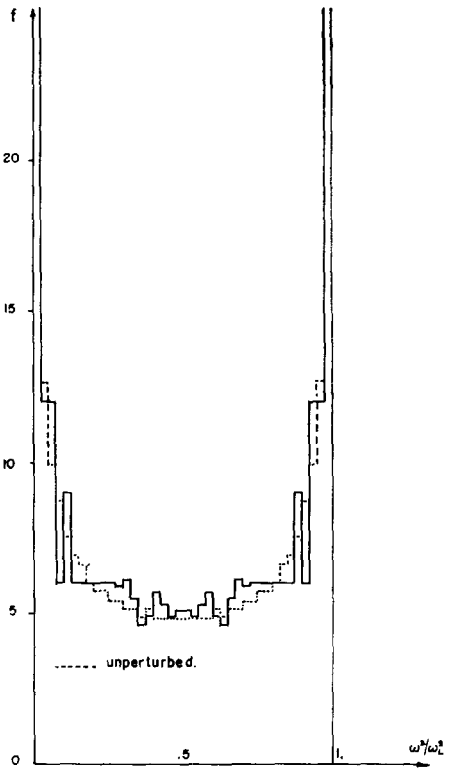

FIG. 8. Frequency distribution for a periodically folded chain one-hundred $\phi_i=0$ followed by five $\phi_i=120°$).

We first consider the case where ϕ_i takes one of the two values 0 and $\tilde{\phi}$; we chose $\tilde{\phi}=120°$, corresponding to "gauche" conformations. We define the concentration p as the number of internal rotation angles different from zero, divided by the total number of angles. We shall determine the frequency distribution of the transverse mode for various values of p. Another parameter is the order in which the ϕ_i are distributed along the chain; we have considered two extreme cases, namely, random and periodic distribution.

Figure 6 shows the spectrum for the out-of-plane mode for eight concentrations of gauche conformations, randomly distributed. We can recognize the inverse square law in the vicinities $p=0$ and $p=1$. Also recognizable is the occurrence of the sequences $\tilde{\phi}$, $\tilde{\phi}$, $\tilde{\phi}$, 0, $\tilde{\phi}$, $\tilde{\phi}$, $\tilde{\phi}$ in the diagrams for $p>0.5$, by the peaks labeled E; this is done according to the analysis of Ref. 4. It would be interesting to identify the island 0, 0, 0, $\tilde{\phi}$, 0, 0, 0, and others by a similar technique.

The effects of ordered sequences of angles 0 and $\tilde{\phi}$ are shown in Figs. 7 and 8. In Fig. 7(a), we show the frequency distribution for the out-of-plane mode for repeated sequences of ten zero angles followed by one $\tilde{\phi}$. Comparing this to Fig. 7(b) for $p=0.089$, we see that

there is a considerable difference between random and regular ordering of defects. This certainly suggests an investigation of the frequency distribution for Markov chains.

In Fig. 8, we show the out-of-plane frequency distribution for an ordered sequence of one-hundred zero angles followed by five $\tilde{\phi}=120°$. This configuration should correspond to the folding of chains in crystallites. While there are differences between this frequency distribution and the frequency distribution for the extended chain, it seems unlikely that an experiment could be devised to measure the differences.

In Fig. 9, we show the frequency distribution for the out-of-plane and bending modes for a random distribution of the values of the ϕ's in the interval zero to 2π. These frequency distributions are considerably different from both the ordered and disordered sequences of trans- and gauche angles. They also show a characteristic difference among themselves; this is to be expected since the matrix elements for in- and out-of-plane motions do not have the same dependence on the internal rotation angle.

IV. DISCUSSION

Using a simple model for the motions of the carbon skeleton of a polyethylene chain, we have computed the frequency distribution as a function of chain conformation. For this model there are variations in the frequency distribution that should be observable experimentally. For example, one could distinguish between certain conformations in the variation of the specific heat[1] or the neutron-scattering cross section.[2]

Our model is certainly too simple to describe the detailed nature of the frequency distribution as a function of chain conformation. However, the model does give the general features of the frequency distribution quite well for the extended chain. Certainly the ease with which calculations can be made for this model makes it at least a good starting point for more detailed and realistic calculations.

We have not attempted a detailed calculation of any measurable quantities such as the specific heat or neutron cross section. We have limited ourselves here to examining the variation of the frequency distribution for various chain conformations and, in particular, which of these should be differentiable experimentally.

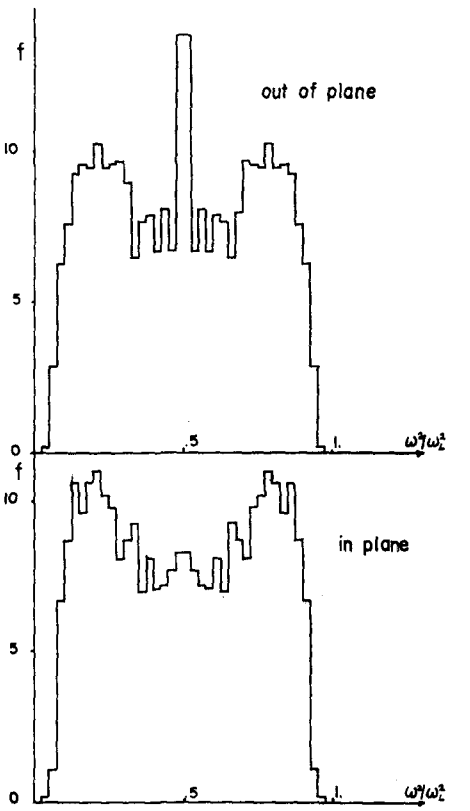

FIG. 9. Frequency distributions for in-plane bending and out-of-plane motions for a random distribution of angles in the range $(0, 2\pi)$.

From these results we would expect to find a measurable difference in the specific heats and neutron cross sections for certain conformations. In a future work, we will perform an analysis of neutron-scattering data and specific heats to distinguish between the various possibilities.

Insertion of the neglected interactions due to the folding itself may well modify the present results. Other effects such as change of magnitude of binary forces may also have a nonnegligible contribution.

However, we have shown that each conformation has a typical frequency distribution. If a more rigorous correspondence could be established between conformations and frequency distributions, it could help to decide such questions as structures of crystallite, amorphous, and intermediate regions.

N-Dimensional Total Orbital Angular-Momentum Operator*

KENNETH D. GRANZOW
Sandia Corporation, Albuquerque, New Mexico
(Received 20 February 1963)

A set of generalized polar coordinate systems are defined in N-dimensional space. The total orbital angular-momentum operator as defined by Louck is found to be a tensor invariant on the $(N-1)$-dimensional unit sphere; hence it is easily explicitly determined in any of the possible generalized polar coordinate systems. Commuting operators can be found and the eigenvalue problem solved in many coordinate systems. Two examples are given: (1) a coordinate system of $3M$ dimensions where M is an integer, and (2) a coordinate system of 8 dimensions.

I. INTRODUCTION

SOLUTIONS of the eigenvalue problem of the N-dimensional total orbital angular-momentum operator are of direct interest in the study of the N-dimensional harmonic oscillator[1] and of some interest in theoretical atomic physics. The theory of the N-dimensional harmonic oscillator is useful in the study of molecular vibration; see, for instance, the review articles by Nielsen[2] and by Shaffer.[3] Hence, the generalizations presented here should prove useful.

II. INVARIANT DEFINTITION OF THE TOTAL-ANGULAR-MOMENTUM OPERATOR

Louck[1] has defined components of orbital angular momentum in N-dimensional Euclidean space as

$$L_{ij} = x_i p_j - x_j p_i. \tag{1}$$

He then defines the orbital angular-momentum operators L_k^2 in terms of a particular polar coordinate system and finds that they are given by

$$L_k^2 = \sum_{j=2}^{k+1} \sum_{i=1}^{j-1} L_{ij}^2, \qquad k = 1, 2, \cdots, N-1. \tag{2}$$

If $N = 3$, L_1^2, and L_2^2 are the well-known operators L_z^2 and L^2, respectively. In general, the operator L_k^2 with $k = N-1$ (where N is the dimension of the Euclidean space) is the N-dimensional analog of the total orbital angular-momentum operator. Thus L_{N-1}^2 will be referred to as the total angular-momentum operator and will be denoted simply by L^2.

Consider now the Euclidean space in which these operators are defined to be the Riemannian space with metric $ds^2 = \sum_{i=1}^{N} dx_i^2$. A generalized polar coordinate system is defined as follows:

$$R^2 = \sum_{i=1}^{N} x_i^2, \tag{3}$$

$$\theta^j = f^j(x_i), \quad i = 1, \cdots, N, \quad j = 1, \cdots, N-1.$$

In this transformation, the functions $f^j(x_i)$ are restricted such as to transform the metric to the form $ds^2 = dR^2 + R^2 ds^{\star 2}$, where $ds^{\star 2} = g^{\star}_{\alpha\beta}\, d\theta^{\alpha}\, d\theta^{\beta}$, $\alpha, \beta = 1, \cdots, N-1$, and where the functions $g^{\star}_{\alpha\beta}$ are independent of R. The $(N-1)$-dimensional Riemannian space with the metric $ds^{\star 2}$ will be referred to as ★ space.[4]

The Laplacian in the new coordinate system can be written

$$\nabla^2 = \frac{1}{g^{\frac{1}{2}}} \frac{\partial}{\partial x^i} \left(g^{\frac{1}{2}} g^{ij} \frac{\partial}{\partial x^j} \right),$$

where $x^1 = R$, $x^n = \theta^{n-1}$, $1 < n \leq N$; $g^{1i} = \delta^{1i}$, $g^{\frac{1}{2}} = R^{N-1} g^{\star \frac{1}{2}}$, $g^{ij} = g^{\star (i-1)(j-1)}/R^2$, $1 < i \leq N$, $1 < j \leq N$; and $g = \det(g_{ij})$, $g^{\star} = \det(g^{\star}_{\alpha\beta})$. Separating the R term and remembering that $g^{\star \frac{1}{2}}$ is independent of R, the following expression for the Laplacian is obtained:

$$\nabla^2 = \frac{1}{R^{N-1}} \frac{\partial}{\partial R} \left(R^{N-1} \frac{\partial}{\partial R} \right) + \sum_{i,j=2}^{N} \frac{1}{g^{\frac{1}{2}}} \frac{\partial}{\partial x^i} \left(g^{\frac{1}{2}} g^{ij} \frac{\partial}{\partial x^j} \right),$$

$$\nabla^2 = \frac{1}{R^{N-1}} \frac{\partial}{\partial R} \left(R^{N-1} \frac{\partial}{\partial R} \right) + \frac{\nabla^{\star 2}}{R^2}, \tag{4}$$

where

$$\nabla^{\star 2} = \frac{1}{g^{\star \frac{1}{2}}} \frac{\partial}{\partial \theta^{\alpha}} \left(g^{\star \frac{1}{2}} g^{\star \alpha\beta} \frac{\partial}{\partial \theta^{\beta}} \right). \tag{5}$$

Louck[1] has shown that the Laplacian is

$$\nabla^2 = \frac{1}{R^{N-1}} \frac{\partial}{\partial R} \left(R^{N-1} \frac{\partial}{\partial R} \right) - \frac{L^2}{\hbar^2 R^2} \tag{6}$$

in the particular polar coordinate system that he

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

[1] J. D. Louck, J. Mol. Spectr. **4**, 298 (1960).
[2] H. H. Nielsen, Rev. Mod. Phys. **23**, 90 (1951).
[3] W. H. Shaffer, Rev. Mod. Phys. **16**, 245 (1944).

[4] In the above equation and in the rest of the paper, the summation over identical upper and lower indices in a term will be assumed, Latin indices being summed from 1 to N, Greek indices being summed from 1 to N-1.

chose. Hence, in that coordinate system, the relationship $L^2 \equiv -\hbar^2\nabla^{\star 2}$ holds. Since $\nabla^{\star 2}$ is a tensor invariant with respect to coordinate transformations in ★ space, so is L^2 since they differ only by a multiplicative constant. Any quantity which is a tensor invariant in one coordinate system, is a tensor invariant in all coordinate systems.[5] Changing coordinates in ★ space is synonymous with choosing different $f^i(x_i)$ functions. Therefore L^2 $(= -\hbar^2\nabla^{\star 2})$ can be easily obtained explicitly from Eq. (5) for any generalized polar coordinate system once the metric for the coordinate system of interest has been determined. The ★ space need not even be orthogonal although in practice this would simplify the algebra.

III. POLAR COORDINATE SYSTEMS

The polar coordinate system defined by Louck[1] is a natural extension of the ordinary three-dimensional polar coordinate system. There are, however, many other reasonable extensions that may prove useful in solving particular problems.

Suppose, for instance, that the Euclidean coordinate system of interest has $3M$ dimensions, where M is an integer. It could be visualized as representing a system of M particles which have three degrees of freedom each. (The analysis which follows has obvious extensions for systems of particles with any number of linear degrees of freedom.) The $3M$ Euclidean coordinates are denoted x^i, $i = 1, \cdots, 3M$. The coordinates x_n, y_n, z_n can be viewed as cartesian coordinates of the nth particle. The transformation is:

$$
\begin{aligned}
x^1 &= x_1 = Ra^1 \sin\theta^1 \cos\phi^1,\\
x^2 &= y_1 = Ra^1 \sin\theta^1 \sin\phi^1,\\
x^3 &= z_1 = Ra^1 \cos\theta^1,\\
x^4 &= x_2 = Ra^2 \sin\theta^2 \cos\phi^2,\\
x^5 &= y_2 = Ra^2 \sin\theta^2 \sin\phi^2,\\
x^6 &= z_2 = Ra^2 \cos\theta^2;\\
&\vdots\\
x^{3n-2} &= x_n = Ra^n \sin\theta^n \cos\phi^n,\\
x^{3n-1} &= y_n = Ra^n \sin\theta^n \sin\phi^n,\\
x^{3n} &= z_n = Ra^n \cos\theta^n;\\
&\vdots\\
x^{3M-2} &= x_M = Ra^M \sin\theta^M \cos\phi^M,\\
x^{3M-1} &= y_M = Ra^M \sin\theta^M \sin\phi^M,\\
x^{3M} &= z_M = Ra^M \cos\theta^M,
\end{aligned}
\tag{7}
$$

where

$$
\begin{aligned}
a^1 &= \sin\xi^1 \sin\xi^2 \cdots \sin\xi^{M-1},\\
a^2 &= \cos\xi^1 \sin\xi^2 \cdots \sin\xi^{M-1},\\
a^3 &= \cos\xi^2 \sin\xi^3 \cdots \sin\xi^{M-1};\\
&\vdots\\
a^n &= \cos\xi^{n-1} \sin\xi^n \cdots \sin\xi^{M-1};\\
&\vdots\\
a^M &= \cos\xi^{M-1},\\
0 &\leqslant \xi^n \leqslant \tfrac{1}{2}\pi,\\
0 &\leqslant \theta^n \leqslant \pi,\\
0 &\leqslant \phi^n < 2\pi,
\end{aligned}
$$

The metric in this coordinate system is

$$
ds^2 = \sum_{n=1}^{3M} (dx^n)^2 = dR^2 + R^2\, ds^{\star 2},
$$

where

$$
ds^{\star 2} = \sum_{n=1}^{M-1} \sin^2\xi^{n+1} \cdots \sin^2\xi^{M-1} (d\xi^n)^2 + \sum_{n=1}^{M} (a^n)^2 [(d\theta^n)^2 + \sin^2\theta^n (d\phi^n)^2]. \tag{8}
$$

For $n = M - 1$ in the above expression and elsewhere in this paper, the factor $\sin^2\xi^{n+1} \cdots \sin^2\xi^{M-1}$ is taken to be equal to one. The transformation retains as coordinates the angle coordinates of the ordinary polar coordinates of each particle; the radial coordinate of the nth particle is easily expressed as $r_n = Ra^n$; the resulting coordinate system is orthogonal. These features make physical interpretation of the results obtained with this coordinate system easy. The system has been used by Morse and Feshbach[6] in discussing the case $M = 2$.

IV. EIGENVALUES AND EIGENVECTORS OF L^2

The eigenvalue problem can now be written

$$
\begin{aligned}
L^2\psi &= -\hbar^2\nabla^{\star 2}\psi = \lambda\hbar^2\psi,\\
\nabla^{\star 2}\psi &= -\lambda\psi.
\end{aligned}
\tag{9}
$$

This equation will be solved in the coordinates

[5] See, for instance, J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, Ontario, Canada, 1949).

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1730.

defined by Eqs. (7). Since the ★ space in which we are working is orthogonal, $\nabla^{\star 2}$ can be written simply

$$\nabla^{\star 2} = \frac{1}{g^{\star\frac{1}{2}}} \sum_{\alpha=1}^{N-1} \frac{\partial}{\partial\theta^\alpha}\left(\frac{g^{\star\frac{1}{2}}}{h_\alpha^2}\frac{\partial}{\partial\theta^\alpha}\right), \tag{10}$$

where $h_\alpha^2 = g^\star_{\alpha\alpha}(g^\star_{\alpha\beta} = 0,\ \alpha \neq \beta)$. Evaluating $\nabla^{\star 2}$ from Eq. (10), using the metric, Eq. (8), one obtains

$$\nabla^{\star 2} = \sum_{n=1}^{M} \frac{\Omega_n}{(a^n)^2} + \sum_{n=1}^{M-1}\left\{\frac{1}{\sin^2\xi^{n+1}\cdots\sin^2\xi^{M-1}\sin^{(3n-1)}\xi^n\cos^2\xi^n} \times \frac{\partial}{\partial\xi^n}\left[\sin^{(3n-1)}\xi^n\cos^2\xi^n\frac{\partial}{\partial\xi^n}\right]\right\},$$

where

$$\Omega_n = \frac{1}{\sin\theta^n}\frac{\partial}{\partial\theta^n}\left(\sin\theta^n\frac{\partial}{\partial\theta^n}\right) + \frac{1}{\sin^2\theta^n}\frac{\partial^2}{(\partial\phi^n)^2}.$$

One recognizes Ω_n as $(-1/\hbar^2)$ times the total angular-momentum operator of the nth particle. Now a new set of angular-momentum operators will be defined:

$$-\frac{\Lambda_1^2}{\hbar^2} = \frac{\Omega_1}{\sin^2\xi^1} + \frac{\Omega_2}{\cos^2\xi^1} + \frac{1}{\sin^2\xi^1\cos^2\xi^1}\frac{\partial}{\partial\xi^1}\left(\sin^2\xi^1\cos^2\xi^1\frac{\partial}{\partial\xi^1}\right),$$

$$-\frac{\Lambda_2^2}{\hbar^2} = -\frac{\Lambda_1^2}{\hbar^2}\frac{1}{\sin^2\xi^2} + \frac{\Omega_3}{\cos^2\xi^2} + \frac{1}{\sin^5\xi^2\cos^2\xi^2}\frac{\partial}{\partial\xi^2}\left(\sin^5\xi^2\cos^2\xi^2\frac{\partial}{\partial\xi^2}\right);$$

$$\vdots$$

$$-\frac{\Lambda_n^2}{\hbar^2} = -\frac{\Lambda_{n-1}^2}{\hbar^2}\frac{1}{\sin^2\xi^n} + \frac{\Omega_{n+1}}{\cos^2\xi^n} + \frac{1}{\sin^{(3n-1)}\xi^n\cos^2\xi^n}\frac{\partial}{\partial\xi^n}\left[\sin^{(3n-1)}\xi^n\cos^2\xi^n\frac{\partial}{\partial\xi^n}\right];$$

$$\vdots$$

$$-\frac{\Lambda_{M-1}^2}{\hbar^2} = \nabla^{\star 2}. \tag{11}$$

The M operators representing the z component of angular momentum for each particle, the M total angular-momentum operators $(-\hbar^2\Omega_n)$ and the $(M-1)$ operators just defined (Λ_n^2) form a set of $(3M-1)$ operators which all commute with each other and thus can take on simultaneous eigenvalues.[7] The Ω_n of course have eigenfunctions, $Y_{l_n}^{m_n}(\theta^n, \phi^n)$, and eigenvalues, $-l_n(l_n+1)$. The solution of Eq. (9) can take on the form

$$\psi = \prod_{n=1}^{M} Y_{l_n}^{m_n}(\theta^n, \phi^n) \prod_{i=1}^{M-1}\gamma_i(\xi^i),$$

where $\gamma_n(\xi^n)$ are eigenfunctions of the operators defined by Eqs. (11). Let λ_n be the eigenvalues satisfying the equation

$$\Lambda_n^2\gamma_n(\xi^n) = \lambda_n\hbar^2\gamma_n(\xi^n).$$

To find the λ_n and γ_n, the following equations must be solved:

$$\left\{\frac{1}{\sin^{(3n-1)}\xi^n\cos^2\xi^n}\frac{d}{d\xi^n}\left[\sin^{(3n-1)}\xi^n\cos^2\xi^n\frac{d}{d\xi^n}\right] - \frac{l_{n+1}(l_{n+1}+1)}{\cos^2\xi^n} - \frac{\lambda_{n-1}}{\sin^2\xi^n} + \lambda_n\right\}\gamma_n = 0, \quad n = 1, 2, \cdots, M-1, \tag{12}$$

where λ_0 is taken to be $l_1(l_1+1)$, the eigenvalues of $(-\Omega_1)$, and $\lambda_{M-1} = \lambda$ [of Eq. (9)]. In Eq. (12) we change independent variables by letting $z = \cos^2\xi^n$. This yields the equation

$$\frac{d^2\gamma_n}{dz^2} + \left[\frac{\frac{3}{2}}{z} + \frac{\frac{3}{2}n}{z-1}\right]\frac{d\gamma_n}{dz} - \left[\frac{\lambda_{n-1}/4}{z-1} - \frac{l_{n+1}(l_{n+1}+1)/4}{z} + \tfrac{1}{4}\lambda_n\right]\frac{\gamma_n}{z(z-1)} = 0. \tag{13}$$

Morse and Feshbach[8] give the general solution of this equation in terms of hypergeometric functions. The solution which is analytic at $z = 0$ is

$$\gamma_n = z^{l_{n+1}/2}(z-1)^{h_{n-1}/2}F(a, b\ |c|\ z), \tag{14}$$

where

$$F(a, b\ |c|\ z) = 1 + \frac{ab}{c}z + \frac{a(a+1)b(b+1)}{2!\,c(c+1)}z^2 + \cdots,$$

$$a = \tfrac{1}{2}(l_{n+1} + h_{n-1} - h_n),$$

$$b = \tfrac{1}{2}(l_{n+1} + h_n + h_{n-1} + 3n + 1),$$

$$c = l_{n+1} + \tfrac{3}{2},$$

$$\lambda_k = h_k(h_k + 3k + 1).$$

[7] Commuting operators in the form of functions of a single variable and the other operators (as above) can be constructed once a coordinate system is chosen (if they exist) by the well-known separation-of-variables technique. The separation constants become the eigenvalues.

[8] Reference 6, pp. 539, 542.

The number h_k is the quantum number for the eigenvalue, λ_k. Since we have $\lambda_0 = l_1(l_1 + 1)$, h_0 is identical to l_1. In order to make γ_n analytic at $z = 1$, $F(a, b\ |c|\ z)$ must be a polynomial; it will be a polynomial if either a or b is a negative integer or zero. Setting b equal to a negative integer, results in the same eigenvalues and eigenfunctions as setting a equal to a negative integer; however, making b a negative integer, also results in the quantum number h_k being negative. Hence, let a equal $-\kappa$. Then

$$h_n = 2\kappa + h_{n-1} + l_{n+1}, \quad \kappa = 0, 1, 2, \cdots . \tag{15}$$

Thus the quantum number h_n can take on only even or odd values depending on the even or odd property of $(h_{n-1} + l_{n+1})$, and it must take on values equal to or greater than $(h_{n-1} + l_{n+1})$.

The eigenvalues of L^2 are $\hbar^2\lambda_{M-1}$ and are thus given by

$$\hbar^2\lambda_{M-1} = \hbar^2\lambda_{N/3-1} = \hbar^2 h_{M-1}(h_{M-1} + N - 2), \tag{16}$$

where

$$h_{M-1} = 2\kappa + h_{M-2} + l_M, \qquad \kappa = 0, 1, 2, \cdots .$$

This is an interesting result when compared to the work of Louck.[1] He found the same eigenvalues in another coordinate system using a different set of $(N - 1)$ commuting operators. Louck has pointed out that there is a multiplicity of sets of commuting operators[9] but has not connected them with the choice of polar coordinate systems.

V. CONCLUSION

The eigenfunctions of the generalized total angular-momentum operator defined by Louck are the N-dimensional analog of spherical harmonics. They provide a set of orthogonal functions on the $(N - 1)$-dimensional unit sphere imbedded in N-dimensional Euclidean space. When N is greater than three, there is always at least one other orthogonal polar coordinate system besides the one given by Louck. Every orthogonal polar coordinate system which the author has tried, has yielded a differential equation in which variables are separable. It seems reasonable to suppose that every orthogonal polar coordinate system will yield a separable equation.

When N is prime, the polar coordinate systems (other than Louck's) lack the obvious symmetry shown by the coordinate system presented in Sec. III. Louck's coordinates, in the terminology of Sec. III, can be viewed as representing N particles of one degree of freedom each. Hence, in his coordinates it is of no consequence whether N is prime or not.

Notice that if N is factorable into three factors, the functions, a^n, can take on the general pattern of the x^n's of Eq. (7), giving a new kind of symmetry. For instance, if $N = 8 = 2 \times 2 \times 2$, the coordinates can be defined:

$$\begin{aligned} x^1 &= Ra^1 \sin\theta^1, & x^5 &= Ra^3 \sin\theta^3, \\ x^2 &= Ra^1 \cos\theta^1, & x^6 &= Ra^3 \cos\theta^3, \\ x^3 &= Ra^2 \sin\theta^2, & x^7 &= Ra^4 \sin\theta^4, \\ x^4 &= Ra^2 \cos\theta^2, & x^8 &= Ra^4 \cos\theta^4, \end{aligned}$$

where

$$\begin{aligned} a^1 &= \sin\sigma \sin\xi^1, & a^3 &= \cos\sigma \sin\xi^2, \\ a^2 &= \sin\sigma \cos\xi^1, & a^4 &= \cos\sigma \cos\xi^2, \end{aligned}$$

The metric for this coordinate system is given by

$$\begin{aligned} ds^{\bigstar 2} = {} & d\sigma^2 + \sin^2\sigma(d\xi^1)^2 + \cos^2\sigma(d\xi^2)^2 \\ & + (a^1)^2(d\theta^1)^2 + (a^2)^2(d\theta^2)^2 \\ & + (a^3)^2(d\theta^3)^2 + (a^4)^2(d\theta^4)^2. \end{aligned}$$

The seven commuting operators are given by

$$\Theta_n = -\hbar^2\,\partial^2/(\partial\theta^n)^2, \qquad n = 1, 2, 3, 4,$$

$$\Xi_1 = \frac{\Theta_1}{\sin^2\xi^1} + \frac{\Theta_2}{\cos^2\xi^1} - \frac{\hbar^2}{\sin\xi^1\cos\xi^1}\frac{\partial}{\partial\xi^1}\left(\sin\xi^1\cos\xi^1\frac{\partial}{\partial\xi^1}\right),$$

$$\Xi_2 = \frac{\Theta_3}{\sin^2\xi^2} + \frac{\Theta_4}{\cos^2\xi^2} - \frac{\hbar^2}{\sin\xi^2\cos\xi^2}\frac{\partial}{\partial\xi^2}\left(\sin\xi^2\cos\xi^2\frac{\partial}{\partial\xi^2}\right),$$

$$L^2 = -\hbar^2\nabla^{\bigstar 2} = \frac{\Xi_1}{\sin^2\sigma} + \frac{\Xi_2}{\cos^2\sigma} - \frac{\hbar^2}{\sin^3\sigma\cos^3\sigma}\frac{\partial}{\partial\sigma}\left(\sin^3\sigma\cos^3\sigma\frac{\partial}{\partial\sigma}\right).$$

Clearly there is a wide choice of polar coordinate systems available; each one yields a set of commuting operators. The fact that L^2 is an invariant of ★ space conveniently leaves the choice of coordinate systems up to the investigator.

ACKNOWLEDGMENT

The author is indebted to Harold L. Davis for a key suggestion which led to this study, and for many helpful discussions.

[9] James D. Louck, "Theory of Angular Momentum in N-Dimensional Space," Los Alamos Scientific Laboratory report LA-2451 (1960) (unpublished).

A Remarkable Representation of the 3 + 2 de Sitter Group

P. A. M. DIRAC
Cambridge University, Cambridge, England and
*Institute for Advanced Study, Princeton, New Jersey**
(Received 20 February 1963)

Among the infinitesimal operators of the 3 + 2 de Sitter group, there are four independent cyclic ones, one of which is separate from the other three. A representation is obtained for which this one has integral eigenvalues while the other three have half-odd eigenvalues, or vice versa. The representation is of a specially simple kind, with the wavefunctions involving only two variables.

INTRODUCTION

We consider the group of rotations of five real variables x_1, x_2, x_3, x_4, x_5 which leave the quadratic form

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 - x_5^2 \tag{1}$$

invariant. The infinitesimal operators of the group are $m_{ab} = -m_{ba} (a, b = 1, 2, 3, 4, 5)$. There are ten independent ones. They satisfy the following commutation relations, in the notation $[\xi, \eta] = \xi\eta - \eta\xi$, with a, b, c, d all different:

$$[m_{ab}, m_{cd}] = 0. \tag{2}$$

With a, b, c, all different and no summation over a,

$$\begin{aligned} [m_{ab}, m_{ac}] &= m_{bc} \quad \text{for} \quad a = 4, 5, \\ &= -m_{bc} \quad \text{for} \quad a = 1, 2, 3. \end{aligned} \tag{3}$$

We shall consider only unitary representations, so that im_{ab} has real eigenvalues, for all a, b.

Some of the basic rotations m_{ab} are cyclic and some are hyperbolic. The cyclic ones are m_{12}, m_{23}, m_{31}, m_{45}, and the hyperbolic ones are m_{14}, m_{24}, m_{34}, m_{15}, m_{25}, m_{35}. For m_{ab} hyperbolic, im_{ab} has a continuous range of eigenvalues and for m_{ab} cyclic, im_{ab} has discrete eigenvalues, which may be either integers or half-odd integers.

If we set up wave equations in the 3 + 2 de Sitter space,

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 - x_5^2 = -R^2, \tag{4}$$

with various tensor or spinor kinds of wavefunctions, we get various representations of the de Sitter group, but they are such that, for all the cyclic operators, m_{ab}, im_{ab} has integral eigenvalues, or else for all of them it has half-odd integral eigenvalues. For these straightforward representations, a mixing of integral and half-odd integral eigenvalues does not occur.

The present paper is concerned with a more primitive representation, for which im_{12}, im_{23}, im_{31} have half-odd integral eigenvalues while im_{45} has integral eigenvalues. There is nothing inconsistent in such a mixing, because the m_{45} rotation is completely detached from the m_{12}, m_{23}, m_{31} rotations. In fact, if one goes over to the covering group of the 3 + 2 de Sitter group, the detachment of the m_{45} rotation allows im_{45} to have any real eigenvalues, independently of what eigenvalues the other rotations have. A general theory of the representations of this covering group has been given by Ehrman.[1]

THE γ MATRICES

We consider 4 × 4 matrices whose elements are all real. There are 16 independent ones. We choose them in a certain way and call them γ matrices. One of them is the unit matrix γ_0. The other 15 are chosen to be all symmetric or skew matrices, with the square of each symmetric one equal to γ_0 and the square of each skew one equal to $-\gamma_0$, and with the product of any two equal to ± a third one. These 15 may be written $\gamma_{AB} = -\gamma_{BA}$, where A and B are two different suffixes going from 1 to 6.

The rules for multiplying the γ_{AB} are as follows: We use the notation $\gamma_{AB}\gamma_{CD} = \gamma_{ABCD}$, and so on for products of more than two factors. Thus any product appears as a γ with an even number of suffixes. There are two general rules: (i) Any two different suffixes may be interchanged, if one brings in the factor -1. (ii) A suffix A occurring in two consecutive positions may be suppressed but one must then bring in the factor -1 for $A = 4$, 5 or 6. Thus for example, with $a, b = 1, \cdots 5$,

$$\gamma_{a6}\gamma_{b6} = \gamma_{a6b6} = -\gamma_{ab66} = \gamma_{ab}. \tag{5}$$

The suppression of all the suffixes yields γ_0.

As a consequence of these rules, we find $\gamma_{AB}^2 = \gamma_0$ if one of the suffixes A, B is in the set 1, 2, 3, and

* The author's stay in Princeton was supported by the National Science Foundation.

[1] J. B. Ehrman, Proc. Cambridge Phil. Soc. **53**, 290 (1957).

the other in the set 4, 5, 6, so γ_{AB} is then symmetric. Otherwise $\gamma_{AB}^2 = -\gamma_0$ and γ_{AB} is skew. We find also that γ_{AB} and γ_{CD} commute if A, B, C, D are all different and anticommute if one of the suffixes A, B is the same as one of the suffixes C, D. Thus a set of five γ_{AB}'s with one suffix in common, all anticommute. Further we find that

$$\gamma_{123456}^2 = \gamma_0,$$

so that

$$\gamma_{123456} = \pm\gamma_0.$$

We arrange that

$$\gamma_{123456} = \gamma_0. \tag{6}$$

This equation enables us to reduce γ_{ABCD} to a two-suffix γ when A, B, C, D are all different. For example, multiplying (6) by $-\gamma_{56}$, we get

$$\gamma_{1234} = -\gamma_{56}.$$

The reduction of γ_{ABCD} to a two-suffix γ when two of the suffixes A, B, C, D are the same is given directly by the rules (i) and (ii), as is illustrated by the example (5).

The result of interchanging rows and columns in any matrix α will be written $\alpha^\dagger$. We have $\gamma_{AB}^\dagger = \pm\gamma_{AB}$, the $+$ sign occurring when $\gamma_{AB}^2 = \gamma_0$ and the $-$ sign when $\gamma_{AB}^2 = -\gamma_0$.

Each γ_{AB} has the diagonal sum zero;

$$\langle\gamma_{AB}\rangle = 0. \tag{7}$$

To prove this, let σ be any 4×4 matrix with a reciprocal. Then

$$\langle\gamma_{AB}\rangle = \langle\gamma_{AB}\sigma\sigma^{-1}\rangle = \langle\sigma^{-1}\gamma_{AB}\sigma\rangle.$$

By taking σ to be one of the γ matrices that anticommutes with γ_{AB}, we get

$$\langle\gamma_{AB}\rangle = -\langle\gamma_{AB}\rangle.$$

A REPRESENTATION OF THE CONFORMAL GROUP

Introduce four variables u_1, u_2, u_3, u_4 to correspond to the four rows of the γ matrices. They may be written as a column matrix and we then denote them by the symbol u. Alternatively, they may be written as a row matrix and are then denoted by $u^\dagger$.

Let v_1, v_2, v_3, v_4 be a second such set of four variables. With α any 4×4 matrix, we may form, by matrix multiplication, $v^\dagger\alpha u$, which is a 1×1 matrix or a number. We have evidently

$$v^\dagger\alpha u = u^\dagger\alpha^\dagger v. \tag{8}$$

Introduce the four differential operators $\partial_n = \partial/\partial u_n (n = 1, 2, 3, 4)$. They satisfy the commutation relations

$$[\partial_n, u_{n'}] = \delta_{n,n'}. \tag{9}$$

We may write them as a column matrix ∂ or as a row matrix $\partial^\dagger$. If we substitute ∂ for v in (8), the equation is no longer valid, on account of the lack of commutation of the ∂'s and u's. One easily finds from (9),

$$\partial^\dagger\alpha u = u^\dagger\alpha^\dagger\partial + \langle\alpha\rangle. \tag{10}$$

From (8) and (9) we find also, for α and β any two 4×4 matrices,

$$[u^\dagger\alpha\partial, u^\dagger\beta\partial] = u^\dagger[\alpha, \beta]\partial. \tag{11}$$

For each matrix γ_{AB}, define the differential operator

$$\chi_{AB} = \tfrac{1}{2}u^\dagger\gamma_{AB}\partial. \tag{12}$$

From (11) we get

$$[\chi_{AB}, \chi_{CD}] = \tfrac{1}{4}u^\dagger(\gamma_{ABCD} - \gamma_{CDAB})\partial.$$

It follows that, for A, B, C, D all different,

$$[\chi_{AB}, \chi_{CD}] = 0, \tag{13}$$

and for A, B, C all different, with no summation over A,

$$\begin{aligned} [\chi_{AB}, \chi_{AC}] &= \chi_{BC} \quad \text{for} \quad A = 4, 5, 6, \\ &= -\chi_{BC} \quad \text{for} \quad A = 1, 2, 3. \end{aligned} \tag{14}$$

The commutation relations (13), (14) are like the commutation relations (2), (3), with the difference that the suffixes can take on six values instead of five. They show that the χ_{AB} may be looked upon as infinitesimal operators of the group of rotations of six real variables $x_1 \cdots x_6$ which leave

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 - x_5^2 - x_6^2 \tag{15}$$

invariant. This is the conformal group of four-dimensional space. It contains the $3 + 2$ de Sitter group as a subgroup.

With the u's restricted to be real variables, the χ_{AB} provide a representation of the conformal group for which the wavefunctions are functions $\psi(u_1, u_2, u_3, u_4)$ of the four real u's. The representation is unitary, because the adjoint of χ_{AB} is

$$-\tfrac{1}{2}\partial^\dagger\gamma_{AB}^\dagger u = -\chi_{AB},$$

from (10) and (7). If we choose ψ so that $\int |\psi|^2 \, d^4u$ converges, the integral is left invariant by the application of any infinitesimal rotation to ψ.

Corresponding to (12), we define

$$\chi_0 = \tfrac{1}{2}u^\dagger\gamma_0\partial = \tfrac{1}{2}u^\dagger\partial.$$

We see from (11) that χ_0 commutes with every χ_{AB}. Thus we may put χ_0 equal to a number, a, say. The condition that $\chi_0 = a$ means that the wavefunction ψ is homogeneous in the u's of degree $2a$. Let us introduce polar variables ρ, θ_1, θ_2, θ_3 in the u space, with $\rho = (u^\dagger u)^{\frac{1}{2}}$, and θ_1, θ_2, θ_3, three independent functions of the ratios of the u's. Then ψ is of the form

$$\psi = \rho^{2a}\psi_1(\theta_1, \theta_2, \theta_3). \tag{16}$$

To secure the unitary character of the representation, we require

$$i(u\partial^\dagger + \partial^\dagger u) = \text{real number}.$$

Now,

$$u^\dagger\partial + \partial^\dagger u = 2u^\dagger\partial + 4 = 4(\chi_0 + 1).$$

Thus,

$$\chi_0 = -1 + ib, \tag{17}$$

where b is a real number. Equation (16) becomes

$$\psi = \rho^{-2+2ib}\psi_1(\theta). \tag{18}$$

The wavefunction ψ is now not square integrable, being an eigenfunction of an operator χ_0 with a continuous range of eigenvalues. We have

$$\int |\psi|^2 \, d^4u = \int |\psi|^2 \, \rho^3 \, d\Omega \, d\rho,$$

where $d\Omega$ is an element of 3-dimensional solid angle in the u space and is some multiple of $d\theta_1 \, d\theta_2 \, d\theta_3$. Substituting the form (18) for ψ, we get

$$\int_{-\infty}^{\infty} |\psi|^2 \, d^4u = \int_0^\infty \rho^{-1} \, d\rho \int |\psi_1|^2 \, d\Omega. \tag{19}$$

We may now drop the variable ρ from the representation, which involves working with the wave function $\psi_1(\theta)$ and dropping the infinite factor $\int \rho^{-1} \, d\rho$ from the formula (19) for the squared length of a wavefunction.

Going back to general wavefunctions ψ that are not eigenfunctions of χ_0, let us consider wavefunctions of the form

$$\psi = P(u_1, u_2, u_3, u_4)f(\rho), \tag{20}$$

where P is a power series in the u's and $f(\rho)$ is chosen so that ψ is square integrable. One finds readily

$$[\chi_{AB}, \rho^2] = u^\dagger\gamma_{AB}u. \tag{21}$$

If the rotation χ_{AB} is cyclic, γ_{AB} is skew, and so χ_{AB} commutes with ρ. Thus a cyclic rotation applied to the wavefunction (20) affects only the power series P and leaves the factor $f(\rho)$ invariant.

Take P to be linear in the u's and suppose it is an eigenfunction of a particular cyclic χ_{AB}. The eigenvalue is then an eigenvalue of $\frac{1}{2}\gamma_{AB}$, and is thus $\pm\frac{1}{2}i$. One can readily infer that, if P is homogeneous of the nth degree in the u's and is an eigenfunction of a particular cyclic χ_{AB}, the eigenvalue of $i\chi_{AB}$ is an integer or half-odd integer according to whether n is even or odd.

The parity of the power series P is invariant under all the rotations, hyperbolic as well as cyclic. This parity determines the integral or half-odd integral character of the eigenvalues of each of the cyclic rotations. It follows that in the present representation there is no mixing of integral and half-odd integral eigenvalues, such as occurs in the remarkable representation we are seeking.

SOME PROPERTIES OF THE OPERATORS χ

Lemma. For any symmetrical 4×4 matrix S,

$$\gamma_{12}S\gamma_{12} + \gamma_{23}S\gamma_{23} + \gamma_{31}S\gamma_{31} = S - \langle S\rangle\gamma_0, \tag{22}$$

and similarly,

$$\gamma_{45}S\gamma_{45} + \gamma_{56}S\gamma_{56} + \gamma_{64}S\gamma_{64} = S - \langle S\rangle\gamma_0. \tag{23}$$

To prove (22), we note that any symmetric γ_{AB} has one of its suffixes equal to 1, 2, or 3 and the other equal to 4, 5, or 6 and therefore it commutes with one of the three quantities γ_{12}, γ_{23}, γ_{31} and anticommutes with the other two. Thus if such a γ_{AB} is substituted for S in (22), two of the terms on the left become S and the third one $-S$, and the $\langle S\rangle$ on the right is zero. Also (22) holds with $S = \gamma_0$, so it holds for any symmetric S. The proof of (23) is similar.

The lemma enables one to deduce a number of quandratic relations between the χ's. From (10),

$$2\chi_{AB} = u^\dagger\gamma_{AB}\partial = \partial^\dagger\gamma_{AB}^\dagger u.$$

Hence

$$-4(\chi_{12}^2 + \chi_{23}^2 + \chi_{31}^2) = \partial^\dagger\gamma_{12}uu^\dagger\gamma_{12}\partial + \partial^\dagger\gamma_{23}uu^\dagger\gamma_{23}\partial + \partial^\dagger\gamma_{31}uu^\dagger\gamma_{31}\partial.$$

The multiplication is associative, so we can pick out the $uu^\dagger$ in the middle of each term on the right and consider it as a 4×4 matrix. It is symmetric, so we can take it to be the S of (22) and get

$$-4(\chi_{12}^2 + \chi_{23}^2 + \chi_{31}^2) = \partial^\dagger uu^\dagger\partial - \partial^\dagger\langle uu^\dagger\rangle\partial = (u^\dagger\partial + 4)u^\dagger\partial - u^\dagger u\partial^\dagger\partial - 2u^\dagger\partial.$$

Thus,

$$\chi_{12}^2 + \chi_{23}^2 + \chi_{31}^2 = -\chi_0(\chi_0 + 1) + \tfrac{1}{4}u^\dagger u\partial^\dagger\partial. \tag{24}$$

Similarly,

$$\chi_{45}^2 + \chi_{56}^2 + \chi_{64}^2 = -\chi_0(\chi_0 + 1) + \tfrac{1}{4}u^\dagger u\partial^\dagger\partial. \quad (25)$$

Hence,

$$\chi_{12}^2 + \chi_{23}^2 + \chi_{31}^2 = \chi_{45}^2 + \chi_{56}^2 + \chi_{64}^2. \quad (26)$$

A more general result may be obtained as follows: Let γ_P and γ_Q be any two γ matrices, i.e., each of them is either γ_0 or a γ_{AB}. Putting $S = uu^\dagger$ in (22), and multiplying by $\partial^\dagger\gamma_P^\dagger$ on the left and $\gamma_Q\partial$ on the right, we get

$$\begin{aligned}\partial^\dagger\gamma_P^\dagger\gamma_{12}uu^\dagger\gamma_{12}\gamma_Q\partial + \partial^\dagger\gamma_P^\dagger\gamma_{23}uu^\dagger\gamma_{23}\gamma_Q\partial \\ + \partial^\dagger\gamma_P^\dagger\gamma_{31}uu^\dagger\gamma_{31}\gamma_Q\partial - \partial^\dagger\gamma_P^\dagger uu^\dagger\gamma_Q\partial \\ = -\partial^\dagger\gamma_P^\dagger\langle uu^\dagger\rangle\gamma_Q\partial \\ = -u^\dagger u\partial^\dagger\gamma_{P\dagger Q}\partial - 2u^\dagger\gamma_{P\dagger Q}\partial,\end{aligned}$$

where $\gamma_{P\dagger Q}$ denotes $\gamma_P\dagger\gamma_Q$. With the help of (10) and (12), this becomes

$$\begin{aligned}&\{\chi_{12P} + \tfrac{1}{2}\langle\gamma_{12P}\rangle\}\chi_{12Q} + \{\chi_{23P} + \tfrac{1}{2}\langle\gamma_{23P}\rangle\}\chi_{23Q} \\ &+ \{\chi_{31P} + \tfrac{1}{2}\langle\gamma_{31P}\rangle\}\chi_{31Q} + \{\chi_P + \tfrac{1}{2}\langle\gamma_P\rangle\}\chi_Q \\ &= \tfrac{1}{4}u^\dagger u\partial^\dagger\gamma_{P\dagger Q}\partial + \chi_{P\dagger Q}. \quad (27)\end{aligned}$$

Now $\langle\gamma_{12P}\rangle$ vanishes unless $\gamma_{12P} = \pm\gamma_0$, in which case $\gamma_{12} = \mp\gamma_P = \pm\gamma_{P\dagger}$. Thus,

$$\langle\gamma_{12P}\rangle\chi_{12Q} = |\langle\gamma_{12P}\rangle|\,\chi_{P\dagger Q}.$$

Treating the other $\langle\gamma\rangle$ terms in (27) in the same way, we get finally

$$\begin{aligned}&\chi_{12P}\chi_{12Q} + \chi_{23P}\chi_{23Q} + \chi_{31P}\chi_{31Q} + \chi_P\chi_Q \\ &= \{1 - \tfrac{1}{2}\,|\langle\gamma_{12P} + \gamma_{23P} + \gamma_{31P} + \gamma_P\rangle|\}\chi_{P\dagger Q} \\ &\quad + \tfrac{1}{4}u^\dagger u\partial^\dagger\gamma_{P\dagger Q}\partial. \quad (28)\end{aligned}$$

Similarly, it may be shown that

$$\begin{aligned}&\chi_{45P}\chi_{45Q} + \chi_{56P}\chi_{56Q} + \chi_{64P}\chi_{64Q} + \chi_P\chi_Q \\ &= \{1 - \tfrac{1}{2}\,|\langle\gamma_{45P} + \gamma_{56P} + \gamma_{64P} + \gamma_P\rangle|\}\chi_{P\dagger Q} \\ &\quad + \tfrac{1}{4}u^\dagger u\partial^\dagger\gamma_{P\dagger Q}\partial. \quad (29)\end{aligned}$$

It should be noted that the last term in (28) or (29) vanishes if $\gamma_{PQ}^\dagger$ or γ_{PQ} is skew.

There are many applications of the general formulas (28) and (29). With $\gamma_P = \gamma_Q = \gamma_0$, we get back to (24) and (25). With $\gamma_P = \gamma_0$, $\gamma_Q = \gamma_{56}$ in (28), we get

$$\chi_{12}\chi_{34} + \chi_{23}\chi_{14} + \chi_{31}\chi_{24} = -(1 + \chi_0)\chi_{56}. \quad (30)$$

With $\gamma_P = \gamma_{12}$, $\gamma_Q = \gamma_{45}$ in (28), we get

$$\begin{aligned}&-\chi_0\chi_{36} + \chi_{31}\chi_{16} + \chi_{32}\chi_{26} + \chi_{12}\chi_{45} \\ &\qquad = \chi_{36} - \tfrac{1}{4}u^\dagger u\partial^\dagger\gamma_{36}\partial. \quad (31)\end{aligned}$$

With $\gamma_P = \gamma_{31}$, $\gamma_Q = \gamma_{16}$ in (29), we get

$$\begin{aligned}&\chi_{26}\chi_{32} + \chi_{24}\chi_{51} + \chi_{25}\chi_{14} + \chi_{31}\chi_{16} \\ &\qquad = -\chi_{36} - \tfrac{1}{4}u^\dagger u\partial^\dagger\gamma_{36}\partial. \quad (32)\end{aligned}$$

Subtracting (32) from (31), we get (taking into account the noncommutativity of χ_{26} and χ_{32}),

$$\chi_{12}\chi_{45} + \chi_{24}\chi_{15} + \chi_{41}\chi_{25} = (1 + \chi_0)\chi_{36}. \quad (33)$$

Equations (30) and (33) and similar equations lead to the general formula, for A, B, C, D all different:

$$\chi_{AB}\chi_{CD} + \chi_{BC}\chi_{AD} + \chi_{CA}\chi_{BD} = (1 + \chi_0)\chi_{ABCD}. \quad (34)$$

With $\gamma_P = \gamma_Q = \gamma_{45}$ in (28), we get

$$\chi_{36}^2 + \chi_{16}^2 + \chi_{26}^2 + \chi_{45}^2 = \chi_0 + \tfrac{1}{4}u^\dagger u\partial^\dagger\partial. \quad (35)$$

Subtracting (25) from (35), we get

$$\chi_{16}^2 + \chi_{26}^2 + \chi_{36}^2 - \chi_{46}^2 - \chi_{56}^2 = \chi_0(\chi_0 + 2). \quad (36)$$

More generally, we have for each value of B,

$$\sum_{A\neq B}\chi_{AB}\chi^{AB} = -\chi_0(\chi_0 + 2), \quad (37)$$

where the raising of a suffix involves a minus sign if the suffix is 4, 5, or 6. Summing for all B, we get

$$\sum_A\sum_{B<A}\chi_{AB}\chi^{AB} = -3\chi_0(\chi_0 + 2), \quad (38)$$

THE REMARKABLE REPRESENTATION

With $\mu, \nu = 1, 2, 3, 4$, define

$$\begin{aligned}m_{\mu\nu} &= \chi_{\mu\nu}, \\ m_{\mu 5} &= C(\chi_{\mu 5} - i\chi_{\mu 6}),\end{aligned} \quad (39)$$

where C is the operator which converts numbers into their conjugate complexes. It has the algebraic properties

$$C^2 = 1, \qquad Ci = -iC, \quad (40)$$

and C commutes with all the ∂'s, u's and γ's. Thus it commutes with all the χ's.

We must check whether the m_{ab} defined by (39) satisfy the commutation relations (2), (3). It is evident from (14) that all these relations are satisfied except the ones for $[m_{\mu 5}, m_{\nu 5}]$ with $\mu \neq \nu$. We have

$$\begin{aligned}[m_{\mu 5}, m_{\nu 5}] &= C(\chi_{\mu 5} - i\chi_{\mu 6})C(\chi_{\nu 5} - i\chi_{\nu 6}) \\ &\quad - C(\chi_{\nu 5} - i\chi_{\nu 6})C(\chi_{\mu 5} - i\chi_{\mu 6}) \\ &= (\chi_{\mu 5} + i\chi_{\mu 6})(\chi_{\nu 5} - i\chi_{\nu 6}) \\ &\quad - (\chi_{\nu 5} + i\chi_{\nu 6})(\chi_{\mu 5} - i\chi_{\mu 6}) \\ &= [\chi_{\mu 5}, \chi_{\nu 5}] + [\chi_{\mu 6}, \chi_{\nu 6}] \\ &\quad + 2i(\chi_{\mu 6}\chi_{\nu 5} - \chi_{\mu 5}\chi_{\nu 6}) \\ &= 2\chi_{\mu\nu} + 2i\{-\chi_{\mu\nu}\chi_{56} + (1 + \chi_0)\chi_{\mu\nu 56}\}\end{aligned}$$

from (34), with $A, B, C, D = \mu, \nu, 5, 6$. Thus,

$$[m_{\mu 5}, m_{\nu 5}] = \chi_{\mu\nu} - 2i\chi_{\mu\nu}(\chi_{56} + \tfrac{1}{2}i) + 2i\chi_{\mu\nu 56}(1 + \chi_0). \quad (41)$$

Consider wavefunctions ψ that satisfy the two supplementary conditions

$$(\chi_0 + 1)\psi = 0, \quad (42)$$

$$(\chi_{56} + \tfrac{1}{2}i)\psi = 0. \quad (43)$$

The conditions are consistent since χ_0 and χ_{56} commute. The operators $\chi_0 + 1$ and $\chi_{56} + \frac{1}{2}i$ occurring in these conditions commute with all the m_{ab}. This is obvious for the operator $\chi_0 + 1$. It is also obvious that $\chi_{56} + \frac{1}{2}i$ commutes with $m_{\mu\nu}$, and it can be proved to commute with $m_{\mu 5}$ as follows. We have

$$\begin{aligned} m_{\mu 5}(\chi_{56} + \tfrac{1}{2}i) &= C(\chi_{\mu 5} - i\chi_{\mu 6})(\chi_{56} + \tfrac{1}{2}i) \\ &= C\{\chi_{56}(\chi_{\mu 5} - i\chi_{\mu 6}) - \chi_{\mu 6} - i\chi_{\mu 5} + \tfrac{1}{2}i(\chi_{\mu 5} - i\chi_{\mu 6})\} \\ &= C(\chi_{56} - \tfrac{1}{2}i)(\chi_{\mu 5} - i\chi_{\mu 6}) \\ &= (\chi_{56} + \tfrac{1}{2}i)m_{\mu 5}. \end{aligned}$$

The supplementary conditions thus pick out a set of ψ's that is invariant under all the operations m_{ab}.

For this set of ψ's, (41) reduces to

$$[m_{\mu 5}, m_{\nu 5}] = m_{\mu\nu},$$

so the m_{ab} satisfy all the commutation relations (2), (3). They thus provide a representation of the 3 + 2 de Sitter group. The representation is unitary, because the m_{ab} are all equal to minus their adjoints, as follows in the case of $m_{\mu 5}$ from C and Ci being self-adjoint.

Some properties of the representation will be worked out. We have, from (30) and (42),

$$m_{12}m_{34} + m_{23}m_{14} + m_{31}m_{24} = 0. \quad (44)$$

The corresponding relations involving other sets of four of the suffixes $1, \cdots 5$ also hold.

The condition (43) means that $i\chi_{56}$ has just the one eigenvalue $\frac{1}{2}$. It follows from the parity discussion in connection with the form (20) that i times any cyclic χ_{AB} has half-odd integral eigenvalues. Thus im_{12}, im_{23}, im_{31} have half-odd integral eigenvalues. It follows then that

$$m_{12}^2 + m_{23}^2 + m_{31}^2 = -k(k + 1), \quad (45)$$

where k has half-odd integral eigenvalues.

We have, from (39) and (14),

$$\begin{aligned} m_{45}^2 &= (\chi_{45} + i\chi_{46})(\chi_{45} - i\chi_{46}) \\ &= \chi_{45}^2 + \chi_{46}^2 - i\chi_{56} \\ &= \chi_{12}^2 + \chi_{23}^2 + \chi_{31}^2 - \chi_{56}^2 - i\chi_{56}, \end{aligned} \quad (46)$$

with the help of (26). Thus, with the supplementary condition (43),

$$m_{45}^2 = m_{12}^2 + m_{23}^2 + m_{31}^2 - \tfrac{1}{4}, \quad (47)$$

$$= -(k + \tfrac{1}{2})^2, \quad (48)$$

with the help of (45). It follows that im_{45} has integral eigenvalues. Thus the representation gives a mixing of integral and half-odd integral eigenvalues for the cyclic rotations.

The k introduced by Eq. (45) may be restricted to have only positive eigenvalues. It follows from (48) that m_{45} does not have the eigenvalue zero.

The absence of the zero eigenvalue has the consequence that the representation can be reduced to two component representations, involving the eigenfunctions of im_{45} with positive and negative eigenvalues, respectively. To prove the result, we note that m_{12}, m_{23}, or m_{31} applied to an eigenfunction of im_{45} does not change the eigenvalue, while $m_{r4}(r = 1, 2, 3)$ applied to it changes the eigenvalue by ± 1, on account of the commutation relation

$$[[m_{r4}, m_{45}], m_{45}] = -m_{r4}. \quad (49)$$

Thus, any application of operators $m_{\mu\nu}$ to an eigenfunction with a positive eigenvalue produces a linear combination of eigenfunctions with positive eigenvalues, as they cannot make the jump from 1 to -1. One can easily check that the two components are irreducible.

The eigenfunctions of im_{45} may be obtained in the following way: We have, from (46) and (25),

$$\begin{aligned} m_{45}^2\psi &= \{-\chi_{56}^2 - i\chi_{56} - \chi_0(\chi_0 + 1) + \tfrac{1}{4}u^\dagger u\partial^\dagger\partial\}\psi \\ &= \tfrac{1}{4}(-1 + u^\dagger u\partial^\dagger\partial)\psi. \end{aligned} \quad (50)$$

Now suppose ψ is of the form

$$\psi = (u^\dagger u)^{-\frac{1}{2}n-1}\phi, \quad (51)$$

where ϕ is homogeneous of degree n. This ψ satisfies (42). Suppose further that ϕ is harmonic, i.e.,

$$\partial^\dagger\partial\phi = 0.$$

Equation (50) now leads to

$$m_{45}^2\psi = -\tfrac{1}{4}(n + 1)^2\psi. \quad (52)$$

The condition (43) leads to

$$(\chi_{56} + \tfrac{1}{2}i)\phi = 0. \quad (53)$$

Thus any harmonic ϕ satisfying (53) provides, according to (51), a ψ that is a linear combination of the eigenfunctions of im_{45} with eigenvalues $\pm\frac{1}{2}(n + 1)$. Of course n has to be an odd integer.

One can infer this directly by noting that (51) is of the form (20) with ϕ for P and the condition (43) requires the function ϕ to be of odd parity.

INHOMOGENEOUS FORM OF THE REPRESENTATION

The representation given by (39) involves the four u's in a homogeneous manner, each $m_{\mu\nu}$ being of degree zero in the u's. One may use the two supplementary conditions (42), (43) to eliminate two of the variables in the wavefunctions, so that there are only two left, which then appear in a nonhomogeneous manner. The resulting form of the representation is more convenient for problems requiring detailed calculation.

Let us first pass from the four real u's to two complex variables. Put

$$v_1 = u_1 + iu_2, \qquad v_2 = u_3 + iu_4.$$

The conjugate complex equations are

$$v_1^* = u_1 - iu_2, \qquad v_2^* = u_3 - iu_4.$$

The operators $\partial/\partial v_1$, $\partial/\partial v_2$ will be written d_1, d_2 for brevity, and they have the values

$$d_1 = \tfrac{1}{2}(\partial_1 - i\partial_2), \qquad d_2 = \tfrac{1}{2}(\partial_3 - i\partial_4).$$

Their conjugates are

$$d_1^* = \tfrac{1}{2}(\partial_1 + i\partial_2), \qquad d_2^* = \tfrac{1}{2}(\partial_3 + i\partial_4).$$

We may choose the γ matrices so that, with the χ's given by (12),

$$\begin{aligned}
\chi_{14} - i\chi_{23} &= -v_2 d_1 - v_1 d_2,\\
\chi_{15} - i\chi_{16} &= v_1^* d_1 - v_2^* d_2,\\
\chi_{24} - i\chi_{31} &= -i(v_2 d_1 - v_1 d_2),\\
\chi_{25} - i\chi_{26} &= i(v_1^* d_1 + v_2^* d_2),\\
\chi_{34} - i\chi_{12} &= v_1 d_1 - v_2 d_2,\\
\chi_{35} - i\chi_{36} &= v_2^* d_1 + v_1^* d_2,\\
\chi_0 + i\chi_{56} &= v_1 d_1 + v_2 d_2,\\
\chi_{45} - i\chi_{46} &= v_2^* d_1 - v_1^* d_2.
\end{aligned} \tag{54}$$

These eight complex equations give sixteen real equations which determine the sixteen γ's, and one finds that the γ's so determined have all the desired properties.

The supplementary conditions (42), (43) may be written

$$\begin{aligned}
(\chi_0 + i\chi_{56})\psi &= -\tfrac{1}{2}\psi,\\
(\chi_0 - i\chi_{56})\psi &= -\tfrac{3}{2}\psi.
\end{aligned} \tag{55}$$

In the new notation they become

$$\begin{aligned}
(v_1 d_1 + v_2 d_2)\psi &= -\tfrac{1}{2}\psi,\\
(v_1^* d_1^* + v_2^* d_2^*)\psi &= -\tfrac{3}{2}\psi.
\end{aligned} \tag{56}$$

They show that ψ is homogeneous of degree $-\frac{1}{2}$ in v_1, v_2 and homogeneous of degree $-\frac{3}{2}$ in v_1^*, v_2^*. Thus ψ is of the form

$$\psi = v_2^{-\frac{1}{2}} v_2^{*-\frac{3}{2}} f(z, z^*), \tag{57}$$

where

$$z = v_1/v_2.$$

Resolving z into its real and pure imaginary parts,

$$z = x_1 + ix_2,$$

we have

$$\psi = v_2^{-\frac{1}{2}} v_2^{*-\frac{3}{2}} g(x_1, x_2). \tag{58}$$

The wavefunction ψ is determined by the function f or g of two variables. Thus we may take f or g to be the wavefunction in a new form of the representation.

Let us see what is the squared length of a wavefunction in the new form. The condition (42) shows that ψ is of the form (18) with $b = 0$, so we have again a squared length like (19) and have to drop out the infinite factor $\int \rho^{-1}\, d\rho$. We may define the variables θ_1, θ_2, θ_3 by means of the equations

$$\begin{aligned}
v_1 &= |v_1| e^{i\theta_1} = \rho \sin\theta_3 e^{i\theta_1},\\
v_2 &= |v_2| e^{i\theta_2} = \rho \cos\theta_3 e^{i\theta_2},
\end{aligned}$$

Thus,

$$|z| = |v_1/v_2| = \tan\theta_3.$$

We now have

$$\begin{aligned}
d^4u &= |v_1|\, d|v_1|\, d\theta_1\, |v_2|\, d|v_2|\, d\theta_2\\
&= |v_1|\, |v_2|\, \rho\, d\rho\, d\theta_3\, d\theta_1\, d\theta_2\\
&= \rho^3\, d\rho \sin\theta_3 \cos\theta_3\, d\theta_1\, d\theta_2\, d\theta_3\\
&= \rho^3\, d\rho \cos^4\theta_3\, d\theta_1\, d\theta_2\, |z|\, d|z|.
\end{aligned}$$

Hence, with ψ of the form (58),

$$\int |\psi|^2\, d^4u = \int \rho^{-1}\, d\rho \int |g|^2_{\cdot}\, d\theta_1\, d\theta_2\, |z|\, d\, |z|.$$

The function g involves only $|z|$ and $\theta_1 - \theta_2$, so we may pass from the variables θ_1, θ_2 to $\theta_1 + \theta_2$ and $\theta_1 - \theta_2$, and then carry out the integration with respect to $\theta_1 + \theta_2$. The result is, with omission of the infinite factor $\int \rho^{-1}\, d\rho$,

$$\begin{aligned}
\tfrac{1}{2} \int |g|^2\, d(\theta_1 + \theta_2)\, d(\theta_1 - \theta_2)\, |z|\, d|z|\\
= \pi \int |g|^2\, dx_1\, dx_2.
\end{aligned} \tag{59}$$

Thus we must take the squared length of the new wavefunctions in the usual way, with weight factor unity.

Define the operators

$$D = v_2 d_1, \qquad D^* = v_2^* d_1^*, \qquad C_1 = C v_2^*/v_2. \tag{60}$$

Note that $C_1^2 = 1$,

$$C_1 z C_1 = z^*, \qquad C_1 D C_1 = D^*. \tag{61}$$

With ψ of the form (57), we have

$$D\psi = v_2^{-\frac{1}{2}} v_2^{*-\frac{3}{2}}\, \partial f/\partial z,$$
$$D^*\psi = v_2^{-\frac{1}{2}} v_2^{*-\frac{3}{2}}\, \partial f/\partial z^*, \tag{62}$$
$$C_1\psi = v_2^{-\frac{1}{2}} v_2^{*-\frac{3}{2}} f^*.$$

Thus the operators D, D^*, C_1 applied to a ψ of the form (57) give other ψ's of this form, and are equivalent to the operators $\partial/\partial z$, $\partial/\partial z^*$, and taking the conjugate complex, applied to f. So these three operators occur in the expressions for the m_{ab} in the new form of the representation.

One finds from (54), with the help of (56),

$$\begin{aligned}
m_{14} - im_{23} &= (z^2 - 1)D - z(v_1 d_1 + v_2 d_2) \\
&= (z^2 - 1)D + \tfrac{1}{2}z, \\
m_{14} + im_{23} &= (z^{*2} - 1)D^* - z^*(v_1^* d_1^* + v_2^* d_2^*) \\
&= (z^{*2} - 1)D^* + \tfrac{3}{2}z^*, \\
m_{24} - im_{31} &= -i(z^2 + 1)D + iz(v_1 d_1 + v_2 d_2) \\
&= -i(z^2 + 1)D - \tfrac{1}{2}iz, \\
m_{24} + im_{31} &= i(z^{*2} + 1)D^* - iz^*(v_1^* d_1^* + v_2^* d_2^*) \\
&= i(z^{*2} + 1)D^* + \tfrac{3}{2}iz^*, \\
m_{34} - im_{12} &= 2zD - v_1 d_1 - v_2 d_2 = 2zD + \tfrac{1}{2}, \\
m_{34} + im_{12} &= 2z^*D^* - v_1^* d_1^* - v_2^* d_2^* \\
&= 2z^*D^* + \tfrac{3}{2}, \\
m_{15} &= C_1\{(z + z^*)D - v_1 d_1 - v_2 d_2\} \\
&= C_1\{(z + z^*)D + \tfrac{1}{2}\}, \\
m_{25} &= -C_1 i\{(z - z^*)D - v_1 d_1 - v_2 d_2\} \\
&= -C_1 i\{(z - z^*)D + \tfrac{1}{2}\}, \\
m_{35} &= C_1\{(1 - zz^*)D + z^*(v_1 d_1 + v_2 d_2)\} \\
&= C_1\{(1 - zz^*)D - \tfrac{1}{2}z^*\}, \\
m_{45} &= C_1\{(1 + zz^*)D - z^*(v_1 d_1 + v_2 d_2)\} \\
&= C_1\{(1 + zz^*)D + \tfrac{1}{2}z^*\}.
\end{aligned} \tag{63}$$

These equations give the m_{ab} in the new form. One can easily check that they are equal to minus their adjoints, as they have to be with the form (59) for the squared length.

TRANSFORMATION OF VARIABLES

We shall make a transformation of the operators z, z^*, D, D^* occurring in the expressions (63) for the m_{ab} and get a new representation. Define the operators ζ, ζ^*, Δ, Δ^* according to

$$\zeta^2 = 2D, \qquad \zeta^{*2} = -2D^*, \tag{64}$$
$$\Delta = -z\zeta, \qquad \Delta^* = \zeta^* z^*. \tag{65}$$

ζ and ζ^* commute, and will be taken as the basic variable in terms of which the new wavefunctions are expressed. They are adjoint operators, so they form a pair of conjugate complex variables. We have

$$[\Delta, \zeta^2] = -2[z\zeta, D] = 2\zeta,$$

so

$$[\Delta, \zeta] = 1.$$

Also,

$$[\Delta, \zeta^*] = 0.$$

Thus, Δ may be considered as the operator $\partial/\partial\zeta$. Δ^* is minus the adjoint of Δ, so it may be considered as $\partial/\partial\zeta^*$.

From (61),

$$\zeta^{*2} = -C_1\zeta^2 C_1 = -(C_1\zeta C_1)^2.$$

We may assume

$$\zeta^* = iC_1\zeta C_1, \tag{66}$$

this being consistent with the other equations. Equations (65) now give

$$\begin{aligned}
\Delta^* &= iC_1\zeta z C_1 \\
&= -iC_1\zeta\Delta\zeta^{-1}C_1.
\end{aligned}$$

Since Δ^* commutes with ζ, we get

$$\begin{aligned}
\Delta^* &= -i\zeta^{-1}C_1\zeta\Delta\zeta^{-1}C_1\zeta \\
&= -iC_2\Delta C_2,
\end{aligned} \tag{67}$$

where

$$C_2 = \zeta^{-1}C_1\zeta. \tag{68}$$

C_2 has the properties

$$C_2^2 = 1, \qquad C_2 i = -iC_2. \tag{69}$$

From (67),

$$C_2\Delta C_2 = i\Delta^*, \qquad C_2\Delta^* C_2 = i\Delta. \tag{70}$$

From (66),

$$C_2\zeta C_2 = -i\zeta^*, \qquad C_2\zeta^* C_2 = -i\zeta. \tag{71}$$

From (69), (70), (71), it follows that C_2 commutes with each of the following:

$$i\Delta\Delta^*, \quad i\zeta\zeta^*, \quad i(\zeta^*\Delta - \zeta\Delta^*), \quad \zeta^*\Delta + \zeta\Delta^*. \tag{72}$$

From (66) again,

$$C_1\zeta^* = -i\zeta C_1 = \zeta C_1 i,$$

or

$$\zeta^{-1}C_1 = C_1 i\zeta^{*-1}.$$

Thus an alternative form for (68) is

$$C_2 = C_1 i\zeta\zeta^{*-1}. \tag{73}$$

In terms of the new operators ζ, ζ^*, Δ, Δ^*, C_2, Eqs. (63) become

$$\begin{aligned} m_{14} - im_{23} &= \tfrac{1}{2}\{(\Delta\zeta^{-1})^2 - 1\}\zeta^2 - \tfrac{1}{2}\Delta\zeta^{-1} \\ &= \tfrac{1}{2}(\Delta^2 - \zeta^2), \\ m_{24} - im_{31} &= -\tfrac{1}{2}i(\Delta^2 + \zeta^2), \\ m_{34} - im_{12} &= -\zeta\Delta - \tfrac{1}{2}, \end{aligned}$$

and the three adjoint equations. Also,

$$\begin{aligned} m_{15} &= \tfrac{1}{2}C_1\{(-\Delta\zeta^{-1} + \zeta^{*-1}\Delta^*)\zeta^2 + 1\} \\ &= \tfrac{1}{2}C_2 i(\zeta^*\Delta - \zeta\Delta^*), \end{aligned}$$

with the help of (73), and similarly,

$$\begin{aligned} m_{25} &= \tfrac{1}{2}C_2(\zeta^*\Delta + \zeta\Delta^*), \\ m_{35} &= -\tfrac{1}{2}C_2 i(\Delta\Delta^* + \zeta\zeta^*), \\ m_{45} &= \tfrac{1}{2}C_2 i(\Delta\Delta^* - \zeta\zeta^*). \end{aligned}$$

Since C_2 commutes with the expressions (72), it commutes with the $m_{\mu 5}$. It must therefore also commute with the $m_{\mu\nu}$, which are just the commutators of the $m_{\mu 5}$. Let us define new infinitesimal operators m'_{ab} by

$$\begin{aligned} m'_{\mu\nu} &= m_{\mu\nu}, \\ m'_{\mu 5} &= C_2 m_{\mu 5}. \end{aligned}$$

One sees at once that the m'_{ab} also satisfy the commutation relations (2), (3) and thus provide a representation of the 3 + 2 de Sitter group. The C_2 operator does not occur in the m'_{ab}, each of them being merely a quadratic function of the variables ζ, ζ^*, Δ, Δ^*.

The new representation can be better understood if we refer it to real variables instead of the complex ζ, ζ^*. Put

$$\zeta = 2^{-\frac{1}{2}}(q_1 + iq_2), \qquad \zeta^* = 2^{-\frac{1}{2}}(q_1 - iq_2),$$

$$-i\Delta = 2^{-\frac{1}{2}}(p_1 - ip_2), \qquad -i\Delta^* = 2^{-\frac{1}{2}}(p_1 + ip_2).$$

Thus q_1, q_2, p_1, p_2 are Hermitian operators satisfying

$$[q_1, p_1] = [q_2, p_2] = i,$$

with their other commutators vanishing, so they are like canonical variables in quantum mechanics. We now get

$$\begin{aligned} im'_{12} &= \tfrac{1}{2}(q_1p_2 - q_2p_1), \\ im'_{23} &= \tfrac{1}{4}(p_1^2 + q_1^2 - p_2^2 - q_2^2), \\ im'_{31} &= -\tfrac{1}{2}(q_1q_2 + p_1p_2), \\ im'_{14} &= \tfrac{1}{2}(q_1q_2 - p_1p_2), \\ im'_{24} &= \tfrac{1}{4}(q_1^2 - p_1^2 - q_2^2 + p_2^2), \\ im'_{34} &= \tfrac{1}{2}(q_1p_1 + p_2q_2), \\ im'_{15} &= -\tfrac{1}{2}(q_1p_2 + q_2p_1), \\ im'_{25} &= -\tfrac{1}{2}(q_1p_1 - q_2p_2), \\ im'_{35} &= \tfrac{1}{4}(q_1^2 + q_2^2 - p_1^2 - p_2^2), \\ im'_{45} &= \tfrac{1}{4}(p_1^2 + q_1^2 + p_2^2 + q_2^2). \end{aligned}$$

There are ten independent quadratic functions of the four variables q_1, q_2, p_1, p_2, and suitable linear combinations of these ten provide the expressions for the m'_{ab}.

The algebraic connection between the operators in the new representation and those of the previous representation (39) or (63) ensures that the new operators satisfy the same algebraic relations as before, e.g., the commutation relations (2), (3) and the various quadratic relations such as (47). However, the new representation is not equivalent to the previous one.

The new representation can be reduced to two component representations, for which the wavefunctions $\psi(q_1, q_2)$ are even or odd functions of the q's, respectively. One can easily check that these components are irreducible.

One sees that im'_{12} has integral or half-odd integral eigenvalues according to whether ψ is even or odd. Further, the expression for im'_{45} is like half the energy of two harmonic oscillators. Its eigenvalues are all positive and the lowest eigenvalue is $\frac{1}{2}$, coming from the zero-point energy of each oscillator. Even wavefunctions lead to half-odd integral eigenvalues for im'_{45}, and odd wavefunctions lead to integral eigenvalues. Thus for each of the component representations, one of the operators im'_{12}, im'_{45} has integral eigenvalues, and the other half-odd integral eigenvalues. The component with im'_{12} half-odd and

im'_{45} integral is equivalent to the component of (39) or (63) for which im_{45} has positive eigenvalues.

The new representation is thus more fundamental than the previous one in two respects—it separates the negative eigenvalues of im_{45} away from the positive ones and it allows both the alternatives of integral and half-odd eigenvalues for im_{45}. The greater power of the new representation comes from the introduction, by (64), of new variables equal to the square roots of previous variables.

The quadratic functions of the four variables q_1, q_2, p_1, p_2 are the infinitesimal generators of the group of linear transformations of these variables that leave their commutation relations invariant. It has been pointed out to me by R. Jost that this group is just the 4-dimensional simplectic group, which is equivalent to the 3 + 2 de Sitter group. These results are easily established by the following argument:

For convenience, put $p_1 = q_3$, $p_2 = q_4$. We then have to consider linear transformations of the four q's which leave the commutators $[q_m, q_n]$ invariant ($m, n = 1, 2, 3, 4$). Let a_m be a set of four numbers which transform contravariantly to the four q's, so that $a_m q_m$ is invariant, and let b_m be another such set of four numbers. Put

$$W_{mn} = a_m b_n - a_n b_m.$$

Thus, each linear transformation of the q's gives rise to a linear transformation of the a's and b's, which in turn gives rise to a linear transformation of the W's.

We have

$$[a_m q_m, b_n q_n] = a_m b_n [q_m, q_n] = 2i(W_{13} + W_{24}).$$

This is invariant. Also,

$$W_{12}W_{34} + W_{23}W_{14} + W_{31}W_{24} = 0.$$

Put

$$W_{12} = x_1 + x_4, \quad W_{23} = x_2 + x_5, \quad W_{31} = x_3 + x_6,$$

$$W_{34} = x_1 - x_4, \quad W_{14} = x_2 - x_5, \quad W_{24} = x_3 - x_6.$$

Then the six x's are subject to linear transformations which leave x_6 invariant, and which preserve the quadratic equation

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 - x_5^2 - x_6^2 = 0.$$

This is just the 3 + 2 de Sitter group.

Particle Charge Symmetries as R_8 Subgroups

D. C. PEASLEE
Australian National University, Canberra, Australia
(Received 27 February 1963)

It is shown how the formalism of 8 × 8 Dirac matrices can be extended to include all the groups so far proposed as relevant to elementary-particle charge symmetries: R_4, R_7, G_2, SU_3. These are all treated as subgroups of R_8, which appears to determine the eightfold structure of the particle families, even before the particle interactions are "switched on". Since these subgroups of R_8 are incompatible, they will lead to a "clash of symmetries", as observed experimentally. It is pointed out that if the plausible association is made between real and charge-space statistics in representations of R_8, the group SU_3 satisfies charge-conjugation invariance for 3-boson interactions but not for two-fermion–one-boson interactions.

An argument is given that the representations **8** and $\bar{\mathbf{8}}$ of SU_3 plus the representations (**7** + **1**) of G_2 and R_7 completely span the symmetries obeying the limited invariance implied by conservation of isotopic spin I and hypercharge Y.

1. INTRODUCTION

STRONG interactions of elementary particles have been classified according to various groups[1] under charge transformation; the most popular symmetries have been[2–9] those of R_4, R_7, G_2, and SU_3, in approximate historical order. Such simple schemes are generally too symmetric to accommodate the particle mass spectrum. Moreover, it has long been known from reaction data that some "clash of symmetries" must occur.[10,11] To this end we can mix two or more of the groups listed but will need a single notation to encompass them all.

The purpose of the present note is to exhibit such a notation in terms of the 8 × 8 Dirac matrices. Explicit representations are obtained for SU_3 and G_2; R_4 is a trivial reduction from R_7 for these matrices and is not discussed in detail. In terms of this formalism one can also inquire whether any additional symmetries appear necessary on physical grounds; the answer is negative.

Two assumptions are made to begin with:

(i) The relevant symmetries are all subgroups of R_8.

(ii) The strong interactions conserve isotopic spin I and hypercharge Y.

The first assumption is prompted by the observation that all elementary particles of minimum mass for given spatial characteristics come in groups of 8: baryons, leptons; pseudoscalar and vector bosons. This property seems to be inherent in the particles themselves, even before the strong interactions are "switched on", so that it is even more fundamental than assumption (ii).

The desired notation should therefore be provided by a representation of R_8—in the first instance one of dimension 8. It happens that R_8 and R_4 are the only rotation groups R_n with two inequivalent representations of dimension n. For R_8 they are most easily seen from the breakup of irreducible representations of R_9 upon restriction to R_8. In terms of the structure parameters $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$, two of the simplest representations of R_9 are $(\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2})$ and (1000), of dimensions 16 and 9, respectively.[1] They are analogues of the well-known representations of R_5 denoted by $(\frac{1}{2}\ \frac{1}{2})$ and (1, 0), corresponding to the Dirac and Kemmer matrices of dimensions 4 and 5. Upon restriction to R_8, the spinor representation $(\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2})$ breaks up into two representations of dimension 8, denoted by R_8^s and between which we need not distinguish; the vector representation (1000) decomposes into (**8** + **1**), the representation **8** to be denoted by R_8^v. Nevertheless, R_8^s and R_8^v provide *equivalent* representations of the subgroups $(R_7 + G_2 + SU_3)$, so that explicit construction will be necessary only for R_8^s, with occasional comparisons to the less easily manipulated R_8^v.

We shall consider only the simplest combinations of fields: linear φ and bilinear $(\bar{\varphi}B\varphi)$ for bosons,

[1] F. D. Murnaghan, *The Theory of Group Representations* (Johns Hopkins Press, Baltimore, Maryland, 1938); R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962). We follow the notation R_7, R_8 for the rotational groups sometimes denoted by O_7^+, O_8^+.
[2] A. Salam and J. C. Polkinghorne, Nuovo Cimento **2**, 685 (1955).
[3] J. Tiomno, Nuovo Cimento **6**, 69 (1957).
[4] R. E. Behrends, Nuovo Cimento **11**, 424 (1959).
[5] D. C. Peaslee, Phys. Rev. **117**, 873 (1960).
[6] J. M. Souriau and D. Kastler, Aix-en-Provence Conference (C. E. N. Saclay, 1961), Vol. I, p. 169.
[7] A. Pais, Phys. Rev. Letters **7**, 291 (1961).
[8] Y. Ne'eman, Nuclear Phys. **26**, 222 (1961).
[9] M. Gell-Mann, Phys. Rev. **125**, 1067 (1962).
[10] J. Schwinger, Ann. Phys. **2**, 407 (1957).
[11] A. Pais, Phys. Rev. **110**, 574 (1958).

bilinear only ($\bar{\psi}F\psi$) for fermions, where B and F are charge space operators, and real space operators are suppressed. These restrictions are imposed by the spin-statistics properties of these fields in real space, but they are exactly echoed by R_8^v and R_8^a, respectively, in charge space. For R_8^v, the operator B is composed of products and sums of $\Sigma_{CD} = -\Sigma_{DC}$ and is hence a two-index (or zero-index) tensor; one-index tensors are available by writing φ_A with a charge subscript. In this way one can build up charge tensors of any order. To use R_8^a for fermions would restrict the construction of charge tensors to even order, while for R_8^a the most elementary operators F are Dirac-type Γ_A, from whose products, tensors of all orders can be made. Of course R_8^a could also be employed for $(\bar{\varphi}B\varphi)$, but this would make for redundancy with φ_A. It may therefore seem plausible to introduce a third assumption[12]:
(iii) Real and charge space statistics are correlated;

$$\{B\} = R_8^v, \qquad \{F\} = R_8^a.$$

This assumption is also a primary one, made at the level of the particles themselves before switching on any interactions, which tend to obscure the correlation.

The column matrices corresponding to assumption (iii) are $\varphi = (\varphi_A)$, with $\varphi_0 = \eta$, $(\varphi_1 \mp i\varphi_2)/\sqrt{2} = \pi^{\pm}$, $\varphi_3 = \pi^0$, $(\varphi_4 \mp i\varphi_5)/\sqrt{2} = K^{\pm}$, $(\varphi_6 \pm i\varphi_7)/\sqrt{2} = (K^0, \bar{K}^0)$; and $\psi = \begin{pmatrix} \mathfrak{Y} \\ \mathfrak{N} \end{pmatrix}$, corresponding to $\rho_3 = \pm 1$, $\mathfrak{Y} = \begin{pmatrix} Y \\ Z \end{pmatrix}$ and $\mathfrak{N} = \begin{pmatrix} N \\ \Xi \end{pmatrix}$ corresponding to $\sigma_3 = \pm 1$, and

$$\sqrt{2}\,Y = \begin{pmatrix} \sqrt{2}\,\Sigma^+ \\ \Sigma^0 + \Lambda \end{pmatrix},$$

$$\sqrt{2}\,Z = \begin{pmatrix} \Sigma^0 - \Lambda \\ \sqrt{2}\,\Sigma^- \end{pmatrix}, \quad \mathfrak{N} = \begin{pmatrix} p \\ n \end{pmatrix}, \quad \Xi = \begin{pmatrix} \Xi^0 \\ \Xi^- \end{pmatrix},$$

corresponding to $\tau_3 = \pm 1$. Here ϱ, δ and τ are three independent Pauli spin operators, from which a set of 8×8 Dirac matrices can be constructed:

$$\Gamma_1 = \rho_3\tau_1, \quad \Gamma_2 = \rho_3\tau_2, \quad \Gamma_3 = \rho_3\tau_3, \quad (1)$$
$$\Gamma_4 = \rho_1\sigma_1, \quad \Gamma_5 = \rho_1\sigma_2, \quad \Gamma_6 = \rho_1\sigma_3, \quad \Gamma_7 = \rho_2.$$

These satisfy

$$\{\Gamma_A, \Gamma_B\} = 2\delta_{AB}, \qquad (2)$$
$$[\Gamma_A, \Gamma_B] = 4i\Sigma_{AB},$$
$$\Gamma_{1234567} = -i.$$

We shall discuss in terms of the Γ_A various reductions of F from R_8^a; to obtain corresponding reductions for B, replace the Σ_{AB} of Eq. (2) by the R_8^v rotation operator, which has $+i(-i)$ in the Ath row (column) and Bth column (row), and zeros elsewhere.

2. REPRESENTATIONS OF SU_3

To display the representations **8** and $\bar{\mathbf{8}}$ of SU_3 as subgroups of R_8^a, it is convenient to recall that R_8^a itself is a reduction from R_9^a. Thus, we introduce a fourth Pauli matrix υ and write

$$\Upsilon_0 = \upsilon_1, \quad \Upsilon_A = \upsilon_2\Gamma_A \ (A = 1 \cdots 7), \quad \Upsilon_8 = \upsilon_3. \quad (3)$$

Reduction to R_8^a is obtained by projection operators $P_{\pm} = \frac{1}{2}(\Upsilon_8 \pm 1)$: that is, by setting $\upsilon_3 = \pm 1$. Now use the structure factors[13] f_{ABC} for SU_3 to define the quantity

$$W = \frac{1}{12i} f_{ABC}\Upsilon_{ABC} \qquad (4)$$
$$= \tfrac{1}{2}\sqrt{3}\, Y\upsilon_1 + L\upsilon_2,$$

where hypercharge is

$$Y = \tfrac{1}{2}(1 - \rho_3)\sigma_3, \qquad (5a)$$

and

$$L = \tfrac{1}{4}[(1 + \rho_3)(1 + \delta\cdot\tau) - (1 - \rho_3)]. \qquad (5b)$$

Since isotopic spin is given by

$$I(I+1) = \tfrac{1}{4}[(1 + \rho_3)(3 + \delta\cdot\tau) + \tfrac{3}{2}(1 - \rho_3)], \qquad (5c)$$

we can also write

$$L = \{I(I+1) - \tfrac{9}{8}\} - \tfrac{1}{4}\{Y^2 - \tfrac{1}{2}\}, \qquad (5d)$$

where $\frac{9}{8}$ and $\frac{1}{2}$ are the expectation values of $I(I+1)$ and Y^2 over all eight particles. An important feature of W is that

$$W^2 = L^2 + \tfrac{3}{4}Y^2 = 1. \qquad (6)$$

Viewed in retrospect, the requirement that W have eigenvalues ± 1 only, and not depend on any physical quantities like I or Y, may be seen as a reason for the occasional $\sqrt{3}$ factors in the f_{ABC}.

Now since W and Υ_A for $A = 0 \cdots 7$ depend only on υ_1 and υ_2, the combinations $\{W, \Upsilon_A\}$ and $[W, \Upsilon_A]$ are linear in υ_3, and hence reducible to 8×8 matrices by the projection operators $P_{\pm}$. Consider, for example,

[12] This point was made in reference 5 and has recently been emphasized by L. de Broglie, D. Bohm, P. Hillion, F. Halbwachs, T. Takabayashi, and J. P. Vigier, Phys. Rev. **129**, 451 (1963).

[13] The f_{ABC} are taken from Table II of reference 9, with $8 \to 0$ and reflection of one axis, $6 \to -6$.

$$\Lambda_A = \tfrac{1}{2}\{W, \Upsilon_A\}; \qquad \Lambda_0 = \tfrac{1}{2}\sqrt{3}\, Y,$$
$$\Lambda_A = \tfrac{1}{2}\{L, \Gamma_A\} + i\tfrac{1}{4}\sqrt{3}\,[Y, \Gamma_A]v_3, \qquad A = 1 \cdots 7. \tag{7}$$

On projection with P_+, the 8×8 forms are

$$\begin{aligned}
\lambda_0 &= \tfrac{1}{2}\sqrt{3}\,(\Sigma_{45} - \Sigma_{67}),\\
\lambda_1 &= \tfrac{1}{2}(\Sigma_{56} + \Sigma_{47}) + \Sigma_{23},\\
\lambda_2 &= \tfrac{1}{2}(\Sigma_{64} + \Sigma_{57}) + \Sigma_{31},\\
\lambda_3 &= \tfrac{1}{2}(\Sigma_{45} + \Sigma_{67}) + \Sigma_{12},\\
\lambda_4 &= \tfrac{1}{2}(\Sigma_{26} + \Sigma_{53} + \Sigma_{71}) - \tfrac{1}{4}\sqrt{3}\,\Gamma_5,\\
\lambda_5 &= \tfrac{1}{2}(\Sigma_{34} + \Sigma_{61} + \Sigma_{72}) + \tfrac{1}{4}\sqrt{3}\,\Gamma_4,\\
\lambda_6 &= \tfrac{1}{2}(\Sigma_{15} + \Sigma_{42} + \Sigma_{73}) + \tfrac{1}{4}\sqrt{3}\,\Gamma_7,\\
\lambda_7 &= \tfrac{1}{2}(\Sigma_{14} + \Sigma_{25} + \Sigma_{36}) - \tfrac{1}{4}\sqrt{3}\,\Gamma_6.
\end{aligned} \tag{8}$$

Equations (8) can be summarized by

$$\lambda_A = \tfrac{1}{2} f_{ABC}\Sigma_{BC}, \tag{9}$$

with summation over repeated indices, provided that we define

$$\Sigma_{0A} = -\Sigma_{A0} \equiv \tfrac{1}{2}\Gamma_A, \qquad A = 1 \cdots 7. \tag{10}$$

This is the well-known procedure for extending the Σ_{AB} of Eq. (2), which generate R_7, to a complete set of generators for R_8. Use of P_- instead of P_+ in extracting Eq. (8) from Eq. (7) would only have changed the sign convention in Eq. (10), which is of no physical consequence.

From Eq. (9), the standard commutation rules of the Σ_{BC}, and the Jacobi identity

$$f_{ABE}f_{CDE} + f_{BCE}f_{ADE} + f_{CAE}f_{BDE} = 0, \tag{11}$$

it follows that

$$[\lambda_A, \lambda_B] = if_{ABC}\lambda_C. \tag{12}$$

Equation (12) states that λ_B is an object transforming under SU_3, and that λ_A is the transformation operator as well; thus, λ_A is the representation **8**. The independent representation $\bar{\mathbf{8}}$ is obtainable from

$$K_A = \tfrac{1}{2}i[W, \Upsilon_A]; \qquad K_0 = Lv_3, \tag{13}$$
$$K_A = \tfrac{1}{2}i[L, \Gamma_A] - \tfrac{1}{4}\sqrt{3}\,\{Y, \Gamma_A\}v_3, \quad A = 1 \cdots 7,$$

by the projection P_-. To show this, note that Eq. (12) holds for the Λ_A as well, and since $W^2 = 1$,

$$W\Lambda_A = \tfrac{1}{2}(\Upsilon_A + W\Upsilon_A W) = \Lambda_A W. \tag{14}$$

Hence it is also true that

$$[\Lambda_A, \Upsilon_B] = if_{ABC}\Upsilon_C, \tag{15a}$$
$$[\Lambda_A, K_B] = if_{ABC}K_C. \tag{15b}$$

Although Eq. (15a) is not diagonal in v_3, Eq. (15b) is, so that

$$[\lambda_A, \kappa_B] = if_{ABC}\kappa_C, \tag{16}$$

where κ_A is the 8×8 matrix obtained by putting $v_3 = \pm 1$ in Eq. (13). Thus κ_A is also an eightfold representation of SU_3; since

$$\Upsilon_A\Lambda_A = \Lambda_A\Upsilon_A = 3W, \tag{17}$$

and W commutes with Λ_A by Eq. (14), it follows that

$$\begin{aligned}
\tfrac{1}{2}i[W, \Upsilon_A\Lambda_A] &= 0 = K_A\Lambda_A,\\
\lambda_A\kappa_A &= \kappa_A\lambda_A = 0.
\end{aligned} \tag{18}$$

Since κ_A is orthogonal to λ_A, it must be $\bar{\mathbf{8}}$.

It can also be made apparent that κ_A comes from the symmetric part of $(\lambda \times \lambda)$; Eq. (9) shows that $\lambda \sim \Sigma_{AB}$, and $[W, \Upsilon] \sim \{\Sigma_{AB}, \Sigma_{CD}\}$. The connecting coefficients can be extracted explicitly and are of course the d_{ABC} given in Table II of reference 9 (with $8 \to 0$ and $6 \to -6$). To obtain λ_A from the representation R_8^v, it is only necessary to substitute the appropriate Σ_{AB} in Eq. (8), together with Eq. (10). The corresponding κ_A can be constructed directly from these λ_A and the d_{ABC}. This discussion makes it appear that λ_A is somehow a more basic representation of SU_3 than κ_A; for λ_A is transformation operator as well as operand, and κ_A is in general to be derived from $(\lambda \times \lambda)$. In assigning SU_3 invariance to physical interactions, one would therefore be inclined to select the representation λ_A as primary.

At this point some physical consequence can be seen for assumption (iii). The primary interactions

$$(\bar{\psi}\lambda_A\psi)\varphi_A, \qquad (\bar{\psi}\kappa_A\psi)\varphi_A, \tag{19}$$

display SU_3 invariance; but consider their properties under charge conjugation C, for which

$$\begin{aligned}
C\varphi_A &= c_A\varphi_A \quad \text{(no summation)},\\
c_A &= +1, \qquad A = 01346\\
&= -1 \qquad\quad = 257.
\end{aligned} \tag{20}$$

If ψ is R_8^s, then

$$C\lambda_A, C\kappa_A = \tilde{\lambda}_A, \tilde{\kappa}_A. \tag{21}$$

But Eqs. (7) and (13) show that $\tilde{\lambda}_A$, $\tilde{\kappa}_A$ contain terms of opposite signatures arising from the commutator and anticommutator; and as λ_A and κ_A are independent, no linear combination of them can remove this difficulty. For the boson forms

$$(\bar{\varphi}\lambda_A\varphi)\varphi_A, \qquad (\bar{\varphi}\kappa_A\varphi)\varphi_A, \tag{22}$$

if φ is R_8^v, $C\Sigma_{AB} = c_Ac_B\Sigma_{AB}$, and comparison with

Eq. (8) shows that

$$C\lambda_A = -c_A\lambda_A, \qquad C\kappa_A = c_A\kappa_A, \tag{23}$$

where the second relation is inferred from the symmetry of κ_A as a component of $\{\lambda_B, \lambda_C\}$. There is no difficulty with charge-conjugation invariance in this case. One thus concludes that under assumption (iii) only the three-boson form of SU_3 interaction is permissible, as the two-fermion–one-boson SU_3 interaction violates C invariance.

This restriction is (implicitly) avoided in the standard references[8,9] by ignoring assumption (iii) and using R_8^v also for ψ in Eq. (19). In this case,

$$C\lambda_A, C\kappa_A = \lambda_A^\dagger, \kappa_A^\dagger = \lambda_A, \kappa_A, \tag{24}$$

so that **8** and $\bar{\mathbf{8}}$ have C properties not only satisfactory but identical, with no distinction as in Eq. (23).

3. REPRESENTATIONS OF G_2 AND R_7

The structure constants of G_2 are given[2] by the defining coefficients for octonions,[7] so that the first step is to form a quantity similar to W:

$$\begin{aligned} V &= \tfrac{1}{12}ie_{ABC}\Gamma_{ABC} \\ &= \tfrac{1}{2}[(1+\rho_3)\boldsymbol{\vartheta}\cdot\boldsymbol{\tau} - \rho_3] \\ &= 2[I(I+1) - \tfrac{9}{8}] + \tfrac{5}{2}(Y^2 - \tfrac{1}{2}). \end{aligned} \tag{25}$$

Here the e_{ABC} are obtained from the $\epsilon_{\alpha\beta\gamma}$ of reference 7 by the axis transformations $3 \to -1$, $1 \to 3$, $2 \to -2$; the indices of e_{ABC} run only from 1 to 7, so it is not necessary to go to R_9 and the Υ_A in this case. The eigenvalues of V are $(\frac{1}{2})^7$, $(-\frac{7}{2})$; to obtain simple eigenvalues as in Eq. (6), define a related quantity U:

$$U = -\tfrac{1}{2}(V + \tfrac{3}{2}), \qquad U^2 = 1. \tag{26}$$

If we define, in analogy with Eq. (7),

$$\partial_A = \tfrac{1}{2}\{U, \Gamma_A\} = -\tfrac{3}{4}\Gamma_A - \tfrac{1}{4}e_{ABC}\Sigma_{BC}, \tag{27}$$

we have

$$[\partial_A, \Gamma_B] = [\Gamma_A, \partial_B] = -3i\Sigma_{AB} - \tfrac{1}{2}ie_{ABC}\Gamma_C, \tag{28a}$$

and by use of the special rule

$$\begin{aligned} &e_{ABC}e_{ADE} + e_{ABE}e_{ADC} \\ &\qquad = 2\delta_{BD}\delta_{CE} - \delta_{BC}\delta_{DE} - \delta_{BE}\delta_{CD}, \end{aligned} \tag{29}$$

also

$$[\partial_A, \partial_B] = 3i\Sigma_{AB} + ie_{ABC}(\partial_C + \tfrac{3}{2}\Gamma_C). \tag{28b}$$

Adding Eqs. (28a) and (28b) yields

$$\begin{aligned} [\partial_A, \epsilon_B] &= ie_{ABC}\epsilon_C, \\ \epsilon_A &= \partial_A + \Gamma_A = \tfrac{1}{4}\{\tfrac{1}{2}(1+U), \Gamma_A\}. \end{aligned} \tag{30}$$

Equation (30) states that ∂_A is the G_2 transformation operator under which ϵ_A (but not ∂_A itself) behaves as the representation **7**. There is no other representation of this dimension.

The last form of ϵ_A shows its physical meaning: $\frac{1}{2}(1+U)$ is a projection operator that selects Λ from the other seven hyperons when applied to ψ. The written-out forms of ϵ_A are

$$\begin{aligned} \epsilon_1 &= \tfrac{1}{4}\Gamma_1 + \tfrac{1}{2}\Sigma_{23} - \tfrac{1}{2}(\Sigma_{56} + \Sigma_{47}), \\ \epsilon_2 &= \tfrac{1}{4}\Gamma_2 + \tfrac{1}{2}\Sigma_{31} - \tfrac{1}{2}(\Sigma_{64} + \Sigma_{57}), \\ \epsilon_3 &= \tfrac{1}{4}\Gamma_3 + \tfrac{1}{2}\Sigma_{12} - \tfrac{1}{2}(\Sigma_{45} + \Sigma_{67}), \\ \epsilon_4 &= \tfrac{1}{4}\Gamma_4 - \tfrac{1}{2}(\Sigma_{26} + \Sigma_{63} + \Sigma_{71}), \\ \epsilon_5 &= \tfrac{1}{4}\Gamma_5 - \tfrac{1}{2}(\Sigma_{34} + \Sigma_{61} + \Sigma_{72}), \\ \epsilon_6 &= \tfrac{1}{4}\Gamma_6 - \tfrac{1}{2}(\Sigma_{15} + \Sigma_{42} + \Sigma_{73}), \\ \epsilon_7 &= \tfrac{1}{4}\Gamma_7 - \tfrac{1}{2}(\Sigma_{14} + \Sigma_{25} + \Sigma_{36}). \end{aligned} \tag{31}$$

The seven Γ_A themselves of course give the representation of R_7; the interaction form for two fermions and one boson is

$$g_0(\bar{\psi}\psi)\varphi_0 + g_1(\bar{\psi}\Gamma_A\psi)\varphi_A, \tag{32}$$

so that a distinction occurs between the coupling constants for η and the other seven mesons π and K. Addition of the G_2 form

$$g_2(\bar{\psi}\epsilon_A\psi)\varphi_A \tag{33}$$

makes an exactly corresponding distinction between Λ and the other baryons.

For the representation R_8^v, the expression in Eq. (31) can be taken over directly by the usual substitution Eq. (10), and similarly for R_7 in Eq. (32). The complementary behavior of G_2 and R_7 as subgroups of R_8 is seen in the following curious fact: Viewed simply as a collection of 8×8 matrices, the $\epsilon_A(R_8^s)$ are obtainable by a similarity transformation on Σ_{0A}, the R_8^v representation of R_7; and conversely, the Γ_A which represent R_7^s are related by a similarity transformation to $\epsilon_A(R_8^v)$.

There are no difficulties with charge-conjugation invariance for any of these cases.

4. COMPLETENESS OF THE SUBGROUPS

The interactions discussed above all have the property of conserving I spin and hypercharge Y. It is natural to inquire whether any other interaction forms should be introduced besides those invariant under SU_3, G_2, and R_7. The following argument suggests a negative answer. Consider the form $(\bar{\psi}F\psi)$, where F is a charge-space operator constructed from the Γ_A. There are 64 of these operators, and

TABLE I. Charge Properties of Operators[a]

$\lvert\Delta Y\rvert$	$\lvert\Delta I\rvert$	Operator
0	0	$1,\ \rho_3,\ \rho^+\boldsymbol{\sigma}\cdot\boldsymbol{\tau},\ \rho^-\sigma_3$
0	1	$\boldsymbol{\tau},\ \rho_3\boldsymbol{\tau},\ \rho^+\boldsymbol{\sigma},\ i\rho^+\boldsymbol{\sigma}\times\boldsymbol{\tau}$
0	2	$\rho^+\{\boldsymbol{\sigma}\boldsymbol{\tau}\}$
1	$\frac{1}{2}$	$\rho_{\pm},\ \rho_{\pm}\boldsymbol{\sigma}\cdot\boldsymbol{\tau},\ \rho_{\pm}\boldsymbol{\sigma},\ \rho_{\pm}(\boldsymbol{\sigma}+\boldsymbol{\tau}+i\boldsymbol{\sigma}\times\boldsymbol{\tau})$
1	3/2	$\rho_{\pm}(2\boldsymbol{\tau}\mp i\boldsymbol{\sigma}\times\boldsymbol{\tau}),\ \rho_{\pm}\{\boldsymbol{\sigma}\boldsymbol{\tau}\}$
2	0	$\rho^-\sigma_1,\ \rho^-\sigma_2$
	1	$\rho^-\boldsymbol{\sigma}\cdot\boldsymbol{\tau},\ i\rho^-\boldsymbol{\sigma}\times\boldsymbol{\tau},\ \rho^-\{\boldsymbol{\sigma}\boldsymbol{\tau}\}$

Here $\rho^{\pm}=(1\pm\rho_3)$, $\rho_{\pm}=(\rho_1\pm i\rho_2)$, $\{\boldsymbol{\sigma}\boldsymbol{\tau}\}$ is the traceless, symmetric tensor.

[a] Table I was originally prepared by Mr. B. A. Faithfull.

since the $\Gamma_A+\Sigma_{AB}$) provide a representation of R_8, the F can be regarded as representing R_8 in various orders. If full R_8 invariance is required, the 64 different F include only one scalar, the unit operator. If full R_8 invariance is not required, additional members of $\{F\}$ will appear as scalars under the transformations that have been retained. Each such scalar generates an associated subgroup of R_8. To display the scalars under I and Y, Table I lists the charge properties of all 64 F, expressed in terms of $\boldsymbol{\rho}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\tau}$.

It is seen that there are just three scalars in addition to the unit operator, which is distinguished upon $R_8\rightarrow R_7$. The additional scalars are equivalent to $I(I+1)$, Y and Y^2; or alternatively to Y, L and U. But Y and L are just λ_0 and κ_0, while U is a suitable candidate for a scalar ϵ_0 to associate with the seven ϵ_A. The two representations of SU_3 and one each of G_2 and R_7 thus complete the span of operators conserving I and Y and derivable from the 8×8 representation of R_8. The interactions observed in nature may be complicated by the advent of higher representations, but the principle of extraction would be the same: from any representation of R_8 to find the (incompatible) representations of subgroups SU_3, G_2, and R_7. Note that R_4 is extracted from the Γ_A simply by the use of $\rho_3\sim(Y^2-\frac{1}{2})$, and therefore contributes nothing additional to the other three groups.

Glass Transitions of Ethylene Oxide Polymers

J. A. FAUCHER, J. V. KOLESKE, E. R. SANTEE, JR., J. J. STRATTA, AND C. W. WILSON, III*
Chemicals Division, Research and Development, Union Carbide Corporation, South Charleston, West Virginia

Glass transition temperatures of a very wide molecular-weight range of ethylene oxide polymers were measured by mechanical loss and by broadline NMR. Starting at about −95°C for ethylene glycol, the glass transition rises to a maximum of −17° for molecular weight 6000 and then drops off to −53° for polymers with molecular weights greater than 200 000. This unusual behavior is probably caused by the high crystallinity of the intermediate molecular-weight polymers.

INTRODUCTION

THEORETICAL predictions and experimental data pertaining to the relation of glass transition temperature (Tg) and molecular weight (M) have been discussed in detail by Fox and Loshaek[1]. They primarily considered addition polymers such as polystyrene, polymethylmethacrylate, and polyisobutylene and showed that the expected relation is

$$Tg(M)=Tg^{\circ}-K/M'$$

in which Tg° is the limiting value of glass transition (for a polymer of very high molecular weight) and K is a positive constant whose value is of the order of 10^5. Read[2,3] first showed that poly(ethylene oxide) exhibited a rather different behavior, in which the glass transition temperature apparently goes through a maximum near $M=10\,000$. We have been interested for some time in the physical properties of this polymer, and hence decided to investigate this anomalous behavior in more detail.

Most of the polymers in this particular series are highly crystalline. It is well known that the widely accepted "thermodynamic" methods for measuring Tg, such as dilatometry and calorimetry, fail to give well-defined results for such materials. We have, therefore, decided to use the methods of mechanical loss and broadline nuclear magnetic resonance (NMR). Strictly speaking, one should refer in these cases to the "temperature of maximum loss" or "temperature of line narrowing"; in fact, Read[2] scrupulously observed this terminology. However, we feel little hesitation in identifying our temperatures of maximum loss or of line narrowing with the glass transition. Ever since the classic paper of Schmieder and Wolf[4], it has been apparent that temperature of maximum loss measured mechanically in the range of one cycle per second correlates quite well with the glass transition temperature determined by dilatometry, for example. It would be too dogmatic to insist on an identity here; to a certain extent every polymer must be examined on its own merits. Cases are known in which the loss spectrum is complex enough so that there is considerable difficulty in assigning the glass transition to any one peak; nylon and polytetrafluoroethylene are examples of this. However, when a polymer has only one distinct maximum in loss it is a reasonable inference that this corresponds to the glass transition, and if we could do the dilatometric or calorimetric experiment sufficiently accurately we would arrive at substantially the same result. The polymers considered here are of this type, if we except the obvious loss maximum due to the melting point.

The situation with regard to broadline NMR measurements is somewhat similar. As a matter of experiment, the temperature of line narrowing for many polymers is reasonably close to the glass transition temperature as determined by other methods. The results of the present work in fact show rather remarkable agreement between mechanical and NMR "glass transition temperatures."

EXPERIMENTAL

Materials

The lower members of the series (mono, di-, tri-, and tetraethylene glycols) were redistilled from commercial samples of Union Carbide Chemicals Division. The Carbowax polyethylene glycols 400 to 10 000 were commercial samples from the same source. They were not further treated except for drying. The numbers of these samples correspond to molecular weights as determined by end-group analysis, and are therefore number average molecular weights. Experience indicates that the molecular weight distributions of the Carbowax polyethylene glycol series are fairly sharp.

For the molecular-weight range 22 000 to 180 000 we have used samples based on Polyox WSR 301, a commercial product of Union Carbide Chemicals Division. These samples were irradiated in a ^{60}cobalt source for varying periods of time, resulting in degradation of the resin. Very little cross-linking occurred, as indicated by complete solubility of the products. The molecular weights were estimated by solution viscosity measurements in water and using the relation established by Bailey and co-workers.[5] These numbers are therefore weight average molecular weights.

* Present address: Department of Physics, University of Akron, Akron, Ohio.

[1] T. G. Fox and S. Loshaek, J. Polymer Sci. 15, 371 (1955).

[2] B. E. Read, Polymer 3, 529 (1962).

[3] T. M. Connor, B. E. Read, and G. Williams, J. Appl. Chem. 14, 74 (1964).

[4] K. Schmieder and K. Wolf, Kolloid Z. 134, 149 (1953).

[5] F. E. Bailey, J. K. Kucera, and L. G. Imhof, J. Polymer Sci. 32, 517 (1958).

TABLE I. Summary of glass transition data for ethylene oxide polymers.

Material	Molecular weight	Glass transition temperature: Torsion pendulum	Glass transition temperature: Broadline NMR
Ethylene glycol	62	−93°C	−85°C
Diethylene glycol	106	−83°	−76°
Triethylene glycol	150	−75°	−80°
Tetraethylene glycol	194	−70°	−69°
Carbowax polyethylene glycol 400	400	−50°	−49°
Carbowax polyethylene glycol 600	600	−42°	−43°
Carbowax polyethylene glycol 1540	1540	−28°	...
Carbowax polyethylene glycol 6000	6000	−17°	−20°
Carbowax polyethylene glycol 10 000	10 000	−20°	...
Degraded Polyox resin	22 000	−28°	...
Carbowax polyethylene glycol 30 000	30 000	−35°	−38°
Degraded Polyox resin	48 000	−39°	...
Degraded Polyox resin	76 000	−44°	...
Degraded Polyox resin	115 000	−47°	...
Degraded Polyox resin	180 000	−48°	...
Polyox resin 5L1	225 000	−53°	−52°
Polyox resin 5M3	1 000 000	−53°	−55°
Polyox resin coagulant grade	7 800 000	−53°	−52°

Experiments showed that water is capable of plasticizing ethylene oxide polymers. Consequently, each sample was carefully dried before measurements were made. Liquid samples were treated with Linde molecular sieves and solid samples were put in a vacuum oven at 50°C and 5-Torr pressure. The water content of each sample was then determined by the Karl Fischer titration method. In no case was more than 0.1% water found.

Mechanical Loss Measurements

Mechanical loss data were obtained from a recording torsion pendulum similar to that described by Nielsen.[6] The frequency range was about 3 to 5 cps in the vicinity of the maximum. For the higher members (above 20 000 in molecular weight) the procedure is quite straightforward since the polymer is self-supporting. A specimen was molded in a press at about 70°C and 1000 psi. From a 20 to 50-mil plaque a small strip about 1.0 in. long and 0.25 wide was cut and mounted in the torsion pendulum. Damping curves were taken at 10° intervals (5° near the maximum) starting at −180°C.

However, this procedure cannot be followed for lower molecular-weight samples. The solid Carbowax polyethylene glycols can be pressed into a plaque but specimens obtained thereby are too brittle to make suitable measurements. The lowest members, of course, are liquids and consequently not self-supporting. For all of these materials a supporting technique was devised which makes use of a cellulose blotter.[7] In brief, a blotter strip of dimensions given above is soaked in the liquid (solid samples are first melted), mounted in the torsion pendulum, and then cooled rapidly to liquid-nitrogen temperature. Damping curves can then be taken in the usual way. For poly(ethylene oxides), at least, the substrate has no effect on the position of the damping peak although its magnitude may vary. Glass cloth and asbestos paper have both been tried successfully; however, cellulose blotter paper is the most convenient support used so far.

NMR Measurements

Broadline proton nuclear magnetic resonance (NMR) spectra were taken on a modified Varian 4300-2 spectrometer at 40.0 Mc/sec. The temperature range covered was from −140°C to at least 10° above the melting point. NMR crystallinity was estimated for a few samples by quantitative separation of the narrow (mobile, amorphous) portion of the NMR spectrum from the broad (immobile crystalline) portion.[8] *Tg* was taken to be the lowest temperature at which the narrow component of the linewidth could be clearly seen.

DISCUSSION

Several mechanical loss[2,3,9] and NMR[10–13] measurements of various poly(ethylene oxide) samples have been previously reported in the literature. However, the present study encompasses a greater number of samples and an increased range of molecular weights examined. The NMR linewidth-transition method of measuring *Tg*, used in this work is also regarded as being more truly indicative of the onset of polymer chain motions in amorphous regions than the NMR spin-lattice

[6] L. E. Nielsen, Rev. Sci. Inst. **22**, 690 (1951).
[7] J. J. Stratta, F. P. Reding, and J. A. Faucher, J. Polymer Sci. **A2**, 5017 (1964).
[8] C. W. Wilson and G. E. Pake, J. Polymer Sci. **10**, 503 (1953); J. Chem. Phys. **27**, 115 (1957).
[9] N. G. McCrum, J. Polymer Sci. **54**, 565 (1961).
[10] K. Hikichi and J. Furuichi, J. Phys. Soc. Japan **18**, 742 (1963).
[11] W. P. Slichter, Makromol. Chem. **34**, 83 (1959).
[12] G. Allen, T. M. Connor, and H. Pursey, Trans. Faraday Soc. **59**, 1525 (1963).
[13] T. M. Connor, Trans. Faraday Soc. **60**, 1574 (1964).

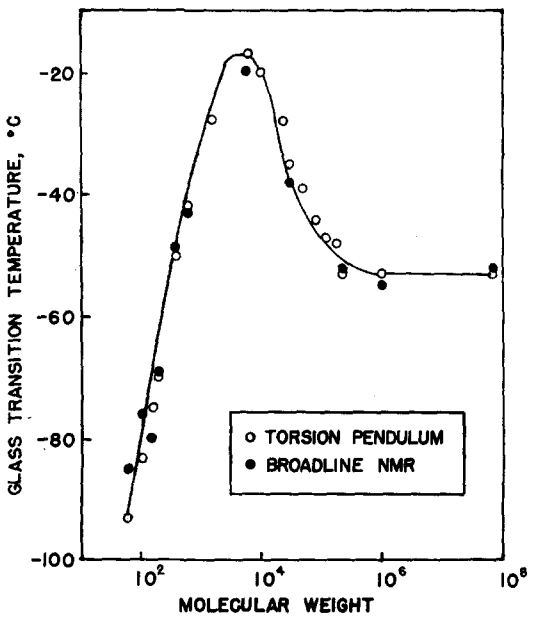

FIG. 1. Glass transition temperature vs molecular weight for poly(ethylene oxide.)

relaxation time measurements reported by Connor and co-workers.[3,12,13]

The results of our determinations are given in Table I and plotted in Fig. 1. A distinct maximum is evident in glass transition at about 6000 molecular weight. This confirms the results of Read and co-workers[2,3] who also observed a maximum, although at approximately 10 000 molecular weight. This rather unusual behavior was considered by Read[2] to be due to the high crystallinity of the ethylene oxide polymers; confirming evidence was found when annealing raised the temperature of maximum loss. We are in accord with these observations and would like to add some additional information. The first four members in Table I can be readily supercooled. They then behave as true glasses and their transitions are all lower than −53°C, characteristic of the very high molecular-weight polymer. However, Carbowax polyethylene glycol 400 and higher polymers can only be obtained in crystalline form. This fundamental difference is apparent, for example, in the shape of the mechanical loss curves. The area under the peak is much greater for glassy materials than for crystalline ones. Further, the data of Connor, Read, and Williams[3] suggest that the absolute amount of crystallinity is a maximum somewhat between 4000 and 30 000 molecular weight. Our crystallinity data from NMR measurements show that above 30 000 molecular weight the crystallinity decreases considerably, falling to about 50% for the "coagulant grade" Polyox resin. However, below 30 000 all samples are so crystalline that a relative comparison is very difficult to make.

A number of attempts were made to quench the intermediate molecular-weight materials by dipping the molten, supported polymer immediately into liquid nitrogen. Although a semicrystalline sample always resulted, these experiments did, in general, produce a transition temperature significantly lower than the corresponding value in Table I. The transitions of such quenched samples were very erratic and unreproducible, and hence are not included in the Table. Nevertheless, these results suggest that truly glassy samples of poly(ethylene oxide) would follow a monotonically increasing curve of glass transition temperature vs molecular weight.

Whatever its cause, this maximum in glass transition temperature may seem very unusual; actually it is apparently not a unique case. In a paper dealing with mechanical properties of polyoxymethylene,[14] Linton and Goodman have noted that the brittle temperature decreases with molecular weight in the general range 20 000 to 100 000. Lower molecular-weight polymers were not investigated but it is probable that a similar maximum exists for that case. Such behavior may be confined to polyethers, or possibly may be even more general. Unfortunately at the present time poly(ethylene oxide) seems to be the only crystalline polymer for which extensive data are available on the glass-transition–molecular weight relation.

[14] W. H. Linton and H. H. Goodman, J. Appl. Polymer Sci. **1**, 179 (1959).

Empty-Space Generalization of the Schwarzschild Metric*

E. NEWMAN, L. TAMBURINO, AND T. UNTI
University of Pittsgurgh, Pittsburgh, Pennsylvania
(Received 5 November 1962)

A new class of empty-space metrics is obtained, one member of this class being a natural generalization of the Schwarzschild metric. This latter metric contains one arbitrary parameter in addition to the mass. The entire class is the set of metrics which are algebraically specialized (contain multiple-principle null vectors) such that the propagation vector is not proportional to a gradient. These metrics belong to the Petrov class type I degenerate.

I. INTRODUCTION

PRINCIPAL null vectors[1,2] of the Riemann curvature tensor (sometimes called Ruse, Debever, or Penrose vectors), which are tangent to a congruence of null geodesics, are defined as geodesic rays. A principal null vector satisfies the algebraic equation $l_{[\mu}R_{\alpha]\beta\gamma[\delta}l_{\nu]}l^{\beta}l^{\gamma} = 0$. Any vacuum metric possesses at least one, and at most four, distinct principal null vectors. When two or more of these vectors coincide, they satisfy $R_{\alpha\beta\gamma[\delta}l_{\mu]}l^{\beta}l^{\gamma} = 0$ in which case one refers to the metric as being algebraically special (Petrov type ID, II, or III).[3] If a vacuum metric is algebraically special, it always contains a geodesic ray.[4] If the principal null vectors are not multiple, the metric is algebraically general (Petrov type I nondegenerate). These metrics in general do not contain geodesic rays except in a small number of cases. Metrics which contain geodesic rays appear to be good approximations in the asymptotic region to exact radiation metrics, the geodesic ray being thought of as the propagation vector. Analytically, the existence of geodesic rays corresponds to a considerable simplification of the initial data necessary to characterize space–time.[5]

A geodesic field, l^{μ}, may be described by three scalars, the divergence, curl, and shear, which are defined by the following:

$$\rho + \bar{\rho} = \text{div } l^{\mu} \equiv l^{\mu}{}_{;\mu}, \tag{1.1}$$

$$|\rho - \bar{\rho}| = \text{curl } l^{\mu} \equiv (l_{[\mu;\nu]}l^{\mu;\nu})^{\frac{1}{2}}, \tag{1.2}$$

$$\sigma\bar{\sigma} = (\text{shear } l^{\mu})^2 \equiv \tfrac{1}{2}[l_{(\mu;\nu)}l^{\mu;\nu} - \tfrac{1}{2}(l^{\mu}{}_{;\mu})^2]. \tag{1.3}$$

Robinson and Trautman[6] have presented all metrics containing nonshearing and noncurling geodesic rays with nonvanishing divergence. The metrics containing noncurling geodesic rays but with shear and divergence were obtained by Newman and Tamburino.[5] Kundt[7] has considered the class of metrics containing geodesic rays with vanishing divergence, curl, and shear.

In this paper we present a subset of the class of metrics possessing nonshearing but curling geodesic rays with nonvanishing divergence. A total of three metrics and their groups of motion are obtained. All these solutions are determined up to several constants.

One solution is of particular interest for it is a generalization of the Schwarzschild metric. There is a discussion of the bound orbits of test particles in the field by means of perturbation theory.

II. CALCULATIONS

The spin coefficient formalism[8] [the reference for this will be denoted by (NP)] introduces into Riemannian space, with signature -2, a tetrad system l_{μ}, n_{μ}, m_{μ}, and $\bar{m}_{\mu}$; l_{μ} and n_{μ} being real null vectors, and m_{μ} and its complex conjugate $\bar{m}_{\mu}$ being complex null vectors. These vectors satisfy the following orthogonality conditions:

$$l_{\mu}n^{\mu} = -m_{\mu}\bar{m}^{\mu} = 1,$$
$$l_{\mu}m^{\mu} = n_{\mu}m^{\mu} = m_{\mu}m^{\mu} = 0. \tag{2.1}$$

It follows that

$$g_{\mu\nu} = l_{\mu}n_{\nu} + l_{\nu}n_{\mu} - m_{\mu}\bar{m}_{\nu} - \bar{m}_{\mu}m_{\nu}. \tag{2.2}$$

In empty space, the physical components of the Riemann tensor are equal to the five (complex) physical components of the Weyl tensor which are

* This work was supported by the Aeronautical Research Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio.
[1] R. Sachs, Proc. Roy. Soc. (London) **264**, 309 (1961).
[2] R. Penrose, Ann. Phys. **10**, 171 (1960).
[3] A. Z. Petrov, Sci. Not. Kazan Univ., **114**, 55 (1954).
[4] J. Goldberg and R. Sachs, (to be published).
[5] E. Newman and L. Tamburino, J. Math. Phys. **3**, 902 (1962).
[6] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **265**, 463 (1962).
[7] W. Kundt, Z. Physik **163**, 77 (1961).
[8] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

defined as follows:

$$\begin{aligned}
\Psi_0 &= -R_{\alpha\beta\gamma\delta}l^\alpha m^\beta l^\gamma m^\delta,\\
\Psi_1 &= -R_{\alpha\beta\gamma\delta}l^\alpha n^\beta l^\gamma m^\delta,\\
\Psi_2 &= -\tfrac{1}{2}R_{\alpha\beta\gamma\delta}(l^\alpha n^\beta l^\gamma n^\delta + l^\alpha n^\beta m^\gamma \bar{m}^\delta),\\
\Psi_3 &= R_{\alpha\beta\gamma\delta}l^\alpha n^\beta n^\gamma \bar{m}^\delta,\\
\Psi_4 &= -R_{\alpha\beta\gamma\delta}n^\alpha \bar{m}^\beta n^\gamma \bar{m}^\delta.
\end{aligned} \tag{2.3}$$

It can be easily shown that if l^μ is a geodesic ray, then Ψ_0 vanishes. [It should be noted that in (NP), Ψ_0 is part of the initial data to be specified over a null hypersurface.] Furthermore, if the geodesic ray is shear free, it can be shown that Ψ_1 must also vanish.[4,8]

The field equations developed in (NP), which are equivalent to the Einstein equations, are a set of nonlinear coupled differential equations (intrinsic derivatives) interrelating a system of variables consisting of the tetrad components (metric components), spin coefficients (defined below), and the physical components of the Riemann tensor. These equations are greatly simplified by suitable choices of the tetrad system and coordinate conditions. Specifically, the vector l^μ will be taken as a normalized geodesic ray and the remainder of the tetrad system is propagated in a parallel manner along this congruence. Coordinate conditions are chosen such that the components of the vector l^μ equal δ_2^μ (the coordinate $x^2 \equiv r$ becomes the affine parameter along the null geodesics). With this choice, the pertinent nonvanishing spin coefficients and their definitions are:

$$\begin{aligned}
&\rho = l_{\mu;\nu}m^\mu \bar{m}^\nu, \quad \sigma = l_{\mu;\nu}m^\mu m^\nu, \quad \tau = l_{\mu;\nu}m^\mu n,^\nu\\
&\lambda = -n_{\mu;\nu}\bar{m}^\mu \bar{m}^\nu, \quad \mu = n_{\mu;\nu}\bar{m}^\mu m^\nu, \quad \nu = -n_{\mu;\nu}\bar{m}^\mu n^\nu,\\
&2\alpha = l_{\mu;\nu}n^\mu \bar{m}^\nu - m_{\mu;\nu}\bar{m}\,^\mu \bar{m}\,^\nu,\\
&2\beta = l_{\mu;\nu}n^\mu m^\nu - m_{\mu;\nu}\bar{m}^\mu m^\nu,\\
&2\gamma = l_{\mu;\nu}n^\mu n^\nu - m_{\mu;\nu}\bar{m}^\mu n^\nu.
\end{aligned} \tag{2.4}$$

It has been shown in (NP) that ρ and σ are related to the divergence, curl, and shear as indicated in Eqs. (1.1), (1.2), and (1.3). In the problem considered in this paper, the shear σ is equal to zero. The tetrad components will be denoted by

$$\begin{aligned}
l^\mu &= \delta_2^\mu,\\
n^\mu &= U\delta_2^\mu + X^a\delta_a^\mu, \quad a = 1, 3, 4,\\
m^\mu &= \omega\delta_2^\mu + \xi^a\delta_a^\mu,\\
\bar{m}^\mu &= \bar{\omega}\delta_2^\mu + \bar{\xi}^a\delta_a^\mu;
\end{aligned} \tag{2.5}$$

and the intrinsic derivatives by

$$\begin{aligned}
D\phi &\equiv \phi_{;\mu}l^\mu = \partial\phi/\partial r,\\
\Delta\phi &\equiv \phi_{;\mu}n^\mu = U\,\partial\phi/\partial r + X^a\phi_{,a}\\
\delta\phi &\equiv \phi_{;\mu}m^\mu = \omega\,\partial\phi/\partial r + \xi^a\phi_{,a},\\
\bar{\delta}\phi &\equiv \phi_{;\mu}\bar{m}^\mu = \bar{\omega}\,\partial\phi/\partial r - \bar{\xi}^a\phi_{,a}.
\end{aligned} \tag{2.6}$$

With the above remarks, we now display the (NP) empty-space field equations that describe a space containing shear-free but curling geodesic rays ($\rho - \bar{\rho} \neq 0$).

$$D\rho = \rho^2, \tag{2.7a}$$

$$D\tau = \rho\tau, \tag{2.7b}$$

$$D\alpha = \rho\alpha, \tag{2.7c}$$

$$D\beta = \bar{\rho}\beta, \tag{2.7d}$$

$$D\gamma = \tau\alpha + \bar{\tau}\beta + \Psi_2, \tag{2.7e}$$

$$D\lambda = \rho\lambda, \tag{2.7f}$$

$$D\mu = \bar{\rho}\mu + \Psi_2, \tag{2.7g}$$

$$D\nu = \bar{\tau}\mu + \tau\lambda + \Psi_3; \tag{2.7h}$$

$$D\Psi_2 - 3\rho\Psi_2 = 0, \tag{2.7i}$$

$$D\Psi_3 - 2\rho\Psi_3 = \bar{\delta}\Psi_2, \tag{2.7j}$$

$$D\Psi_4 - \rho\Psi_4 = \bar{\delta}\Psi_3 + 2\alpha\Psi_3 - 3\lambda\Psi_2; \tag{2.7k}$$

$$DU = \tau\bar{\omega} + \bar{\tau}\omega - (\gamma + \bar{\gamma}), \tag{2.7l}$$

$$DX^a = \tau\bar{\xi}^a + \bar{\tau}\xi^a \qquad a = 1, 3, 4, \tag{2.7m}$$

$$D\omega = \bar{\rho}\omega - (\bar{\alpha} + \beta), \tag{2.7n}$$

$$D\xi^a = \bar{\rho}\xi^a; \tag{2.7o}$$

$$\delta\rho = \rho(\bar{\alpha} + \beta) + (\rho - \bar{\rho})\tau, \tag{2.8a}$$

$$\delta\alpha - \bar{\delta}\beta = \rho\mu + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta + \gamma(\rho - \bar{\rho}) - \Psi_2, \tag{2.8b}$$

$$\delta\lambda - \bar{\delta}\mu = (\rho - \bar{\rho})\nu + \mu(\alpha + \bar{\beta}) + \lambda(\bar{\alpha} - 3\beta) - \Psi_3, \tag{2.8c}$$

$$\Delta\lambda - \bar{\delta}\nu = \lambda(\bar{\gamma} - 3\gamma - \mu - \bar{\mu}) + \nu(3\alpha + \bar{\beta} - \bar{\tau}) - \Psi_4, \tag{2.8d}$$

$$\delta\nu - \Delta\mu = \mu^2 + \lambda\bar{\lambda} + \mu(\gamma + \bar{\gamma}) + \nu[\tau - \bar{\alpha} - 3\beta], \tag{2.8e}$$

$$\delta\tau = \rho\bar{\lambda} + \tau(\tau + \beta - \bar{\alpha}), \tag{2.8f}$$

$$\Delta\rho - \bar{\delta}\tau = \rho(\gamma + \bar{\gamma} - \bar{\mu}) + \tau(\bar{\beta} - \alpha - \bar{\tau}) - \Psi_2, \tag{2.8g}$$

$$\Delta\alpha - \bar{\delta}\gamma = \rho\nu - \lambda(\tau + \beta) + \alpha(\bar{\gamma} - \bar{\mu}) + \gamma(\bar{\beta} - \bar{\tau}) - \Psi_3, \quad (2.8h)$$

$$\delta\gamma - \Delta\beta = \alpha\bar{\lambda} + \mu\tau + \gamma(\tau - \bar{\alpha} - \beta) + \beta(\mu + \bar{\gamma} - \gamma); \quad (2.8i)$$

$$\delta\Psi_2 = 3\tau\Psi_2, \quad (2.8j)$$

$$\Delta\Psi_2 - \delta\Psi_3 = -3\mu\Psi_2 + 2(\beta - \tau)\Psi_3, \quad (2.8k)$$

$$\Delta\Psi_3 - \delta\Psi_4 = 3\nu\Psi_2 - (2\gamma + 4\mu)\Psi_3 + (4\beta - \tau)\Psi_4; \quad (2.8l)$$

$$\delta U - \Delta\omega = -\bar{\nu} + U(\tau - \bar{\alpha} - \beta) + \bar{\omega}\bar{\lambda} + \omega(\mu - \gamma + \bar{\gamma}), \quad (2.8m)$$

$$\delta X^a - \Delta\xi^a = X^a(\tau - \bar{\alpha} - \beta) + \bar{\xi}^a\bar{\lambda} + \xi^a(\mu - \gamma + \bar{\gamma}), \quad (2.8n)$$

$$\bar{\delta}\omega - \delta\bar{\omega} = (\bar{\mu} - \mu) + U(\bar{\rho} - \rho) + \bar{\omega}(\beta - \bar{\alpha}) + \omega(\alpha - \bar{\beta}), \quad (2.8o)$$

$$\bar{\delta}\xi^a - \delta\bar{\xi}^a = (\bar{\rho} - \rho)X^a + (\beta - \bar{\alpha})\bar{\xi}^a + (\alpha - \bar{\beta})\xi^a. \quad (2.8p)$$

There are a number of coordinate and tetrad transformations that preserve the coordinate conditions and the orthogonality conditions as well as the parallel propagation for the tetrad system. The coordinate transformations are

$$r' = r + R(x^a),$$
$$x^{a'} = x^{a'}(1, 3, 4), \qquad a = 1, 3, 4. \quad (2.9)$$

The allowed tetrad transformations are

$$l^{\mu'} = l^\mu,$$
$$n^{\mu'} = n^\mu + \bar{B}m^\mu + B\bar{m}^\mu + B\bar{B}l^\mu,$$
$$m^{\mu'} = m^\mu + Bl^\mu, \quad (2.10)$$

where B is a complex scalar independent of r, and

$$l^{\mu'} = l^\mu, \qquad n^{\mu'} = n^\mu, \qquad m^{\mu'} = m^\mu e^{iC}, \quad (2.11)$$

where C is real and independent of r.

A permissable combination of tetrad and coordinate transformation is

$$l^{\mu'} = A^{-1}\delta^\mu_2, \qquad r' = Ar,$$
$$n^{\mu'} = An^\mu, \qquad x^{a'} = x^a,$$
$$m^{\mu'} = m^\mu, \qquad \text{where} \quad A = A(x^a). \quad (2.12)$$

We are now ready to integrate the (NP) equations. The following simplifying assumptions will be made:

$$\psi_3 = \psi_4 = 0, \quad (2.13)$$

$$\nu = 0, \quad (2.14)$$

$$\tau = 0. \quad (2.15)$$

(R. Kerr has found the more general solutions without these simplifications.)

In the remainder of this paper, a degree sign shall indicate a function is independent of r. The solution to $D\rho = \rho^2$ [Eq. (2.7a)] is

$$\rho = -(r + \rho^\circ)^{-1}. \quad (2.16)$$

By using coordinate transformation (2.9), one can set the real part of the constant of integration ρ° equal to zero. Under the combined transformation (2.12), ρ° transforms as

$$\rho^{\circ\prime} = A^\circ\rho^\circ. \quad (2.17)$$

The scalar A° is chosen to make ρ° a constant,

$$\rho^\circ = i\,|\rho^\circ| = \text{constant}. \quad (2.18)$$

The constant could be given a definite value, however it is perferable to let ρ° be a parameter. It follows directly from Eqs. (2.15) and (2.8f) that

$$\lambda = 0. \quad (2.19)$$

From Eqs. (2.7c)–(2.7o) one can readily obtain the r dependence of the remaining variables. The solutions to these equations are summarized below.

$$\alpha = \rho\alpha^\circ, \quad (2.20)$$

$$\beta = \bar{\rho}\beta^\circ, \quad (2.21)$$

$$\Psi_2 = \rho^3\Psi_2^\circ, \quad (2.22)$$

$$\gamma = \gamma^\circ + \tfrac{1}{2}\rho^2\Psi_2^\circ, \quad (2.23)$$

$$\mu = \mu^\circ\bar{\rho} + \tfrac{1}{2}\Psi_2^\circ(\rho^2 + \rho\bar{\rho}), \quad (2.24)$$

$$U = U^\circ - (\gamma^\circ + \bar{\gamma}^\circ)r - \tfrac{1}{2}(\Psi_2^\circ\rho + \bar{\Psi}_2^\circ\bar{\rho}), \quad (2.25)$$

$$\omega = \bar{\rho}\omega^\circ + (\bar{\alpha}^\circ + \beta^\circ), \quad (2.26)$$

$$\chi^a = \chi^{\circ a}, \qquad a = 1, 3, 4, \quad (2.27)$$

$$\xi^a = \bar{\rho}\xi^{\circ a}, \qquad a = 1, 3, 4. \quad (2.28)$$

These expressions are now substituted into the remaining equations [(2.7j) and (2.8j)]. Relationships involving the constants of integration are obtained by comparing the coefficients of like powers of r.

From Eqs. (2.7j) and (2.8j), we obtain the following relationships:

$$\omega = 0, \quad (2.29)$$

$$\alpha = -\bar{\beta}, \quad (2.30)$$

$$\xi^{\circ a}\Psi^\circ_{2,a} = \bar{\xi}^{\circ a}\Psi^\circ_{2,a} = 0. \quad (2.31)$$

Equation (2.8k) yields the following four relations:

$$\gamma^\circ + \bar{\gamma}^\circ = 0, \tag{2.32}$$

$$\mu^\circ = \bar{\mu}^\circ = -U^\circ, \tag{2.33}$$

$$4\mu^\circ\rho^\circ = \Psi_2^\circ - \bar{\Psi}_2^\circ, \tag{2.34}$$

$$X^{\circ a}\Psi_{2,a}^\circ = 0. \tag{2.35}$$

From Eqs. (2.8c) and (2.8e), we obtain

$$\xi^{\circ a}\mu_{,a}^\circ = X^{\circ a}\mu_{,a}^\circ = 0. \tag{2.36}$$

Equation (2.8n) yields

$$\xi^{\circ a}X_{,a}^{\circ b} - X^{\circ a}\xi_{,a}^{\circ b} = 2\bar{\gamma}^\circ\xi^{\circ b}. \tag{2.37}$$

At this stage it is convenient to use the available coordinate and tetrad transformations to simplify the above results. Transformation (2.11), $m^{\mu'} = m^\mu \exp[iC^\circ]$, and Eq. (2.32) are used to set

$$\gamma^\circ = 0. \tag{2.38}$$

Under the coordinate transformation (2.9), $x^a = \theta^a(x^b)$, the variables $X^{\circ a}$ and $\xi^{\circ a}$ transform as follows:

$$X^{\circ a'} = X^{\circ b}\theta_{,b}^a, \tag{2.39}$$

$$\xi^{\circ a'} = \xi^{\circ b}\theta_{,b}^a. \tag{2.40}$$

Equations (2.39) can be used to set

$$X^{\circ 1} = 1, \tag{2.41}$$

$$X^{\circ i} = 0, \qquad i = 3, 4. \tag{2.42}$$

The remaining coordinate freedom is now

$$x^{1'} = x^1 + \theta^1(3, 4), \qquad \xi^{\circ 1'} = \xi^{\circ 1} + \xi^{\circ i}\theta_{,i}^1, \tag{2.43}$$

$$x^3 = \theta^3(3, 4), \qquad \xi^{\circ 3'} = \xi^{\circ i}\theta_{,i}^3, \tag{2.44}$$

$$x^4 = \theta^4(3, 4), \qquad \xi^{\circ 4'} = \xi^{\circ i}\theta_{,i}^4. \tag{2.45}$$

Because Eq. (2.37) becomes, using Eqs. (2.41) and (2.42),

$$\xi_{,1}^{\circ a} = 0, \tag{2.46}$$

it is possible to use Eqs. (2.44) and (2.45) to set

$$\xi^{\circ 3} = -i\xi^{\circ 4} = P(3, 4), \tag{2.47}$$

and the tetrad transformation $m^{\mu'} = m^\mu \exp[iC(3, 4)]$ to set

$$P = \bar{P} \equiv p(3, 4). \tag{2.48}$$

The coordinate freedom now available is

$$x^{1'} = x^1 + f(3, 4), \qquad \xi^{\circ 1'} = \xi^{\circ 1} + p\nabla f, \tag{2.49}$$

and

$$\zeta' = g(\zeta), \tag{2.50}$$

where

$$\zeta = x^3 + ix^4, \qquad \nabla \equiv 2\,\partial/\partial\bar{\zeta}.$$

Putting results (2.41), (2.42), and (2.47) into Eqs. (2.31)–(2.36), one finds that μ° and Ψ_2° are constants which satisfy

$$4\mu^\circ\rho^\circ = \Psi_2^\circ - \bar{\Psi}_2^\circ. \tag{2.51}$$

From Eqs. (2.8p) and (2.8b), we obtain the following three relations:

$$2\alpha^\circ = p\bar{\nabla}\ln p, \tag{2.52}$$

$$2\rho^\circ = p^2[\nabla(\bar{\xi}^{\circ 1}/p) - \bar{\nabla}(\xi^{\circ 1}/p)], \tag{2.53}$$

$$2\mu^\circ = (\sqrt{2}\,p)^2\nabla\bar{\nabla}\ln(\sqrt{2}\,p). \tag{2.54}$$

This completes the information obtainable from the field equations. Equation (2.52) is taken as the definition of α°. The remaining task is to integrate Eqs. (2.53) and (2.54).

The constant $2\mu^\circ$ in Eq. (2.54) is the Gaussian curvature of a two-space with the metric $(\sqrt{2}p)^2\delta^{ij}$. This metric is equal to the $\lim_{r\to\infty} r^2 g^{ij}$, where $g^{ij}(i, j = 3, 4)$ are the solutions [Eq. (2.61)]. With the coordinate transformation (2.50) it is possible to choose coordinates such that

$$\sqrt{2}\,p = 1 + \tfrac{1}{2}\mu^\circ\zeta\bar{\zeta}. \tag{2.55}$$

The curvature $2\mu^\circ$ may be reduced by means of a scale change in r to one of the three values 1, 0, or -1. It is for this reason that the constant ρ° is regarded as a parameter.

Putting this expression for p into Eq. (2.53), one can obtain the following solution to the inhomogeneous equation:

$$\xi^{\circ 1} = -\rho^\circ\zeta/\sqrt{2}. \tag{2.56}$$

Let η° be the most general solution to the homogeneous equation so that the most general solution to Eq. (2.53) is

$$\xi^{\circ 1} = \eta^\circ - \rho^\circ\zeta/\sqrt{2}. \tag{2.57}$$

The coordinate transformation (2.49) yields

$$\xi^{\circ 1'} = \eta^\circ + p\nabla f(3, 4) - \rho^\circ\zeta/\sqrt{2}; \tag{2.58}$$

f shall be chosen such that $\xi^{\circ 1'} = -\rho^\circ\zeta/\sqrt{2}$ which means that Eq. (2.56) becomes the most general solution to Eq. (2.53). One can always satisfy

$$\nabla f = -\eta^\circ/p, \tag{2.59}$$

because the integrability condition $\nabla\bar{\nabla}f = \nabla\bar{\nabla}f$ is equivalent to the homogeneous equation itself and hence satisfied. This completes the formal calculations. The metric is obtained by using Eqs. (2.22) and (2.5).

A summary of the results follows.

The tetrad has the following form:

$$l^\mu = \delta^\mu_2, \qquad l_\mu = \delta^1_\mu - A^\circ\delta^3_\mu - B^\circ\delta^4_\mu,$$
$$n^\mu = U\delta^\mu_2 + \delta^\mu_1,$$
$$n_\mu = \delta^2_\mu + U[-\delta^1_\mu + A^\circ\delta^3_\mu + B^\circ\delta^4_\mu],$$
$$m^\mu = \bar\rho\xi^{\circ a}\delta^\mu_a, \qquad m_\mu = -(\delta^3_\mu + i\delta^4_\mu)/2\rho p. \tag{2.60}$$

The contravariant and covariant metrics are

$$g^{\mu\nu} = \begin{vmatrix} -(A^{\circ 2}+B^{\circ 2})R^{-2} & 1 & -A^\circ R^{-2} & -B^\circ R^{-2} \\ 1 & 2U & 0 & 0 \\ -A^\circ R^{-2} & 0 & -R^{-2} & 0 \\ -B^\circ R^{-2} & 0 & 0 & -R^{-2} \end{vmatrix}, \tag{2.61}$$

and

$$g_{\mu\nu} = \begin{vmatrix} -2U & 1 & 2A^\circ U & 2B^\circ U \\ 1 & 0 & -A^\circ & -B^\circ \\ 2A^\circ U & -A^\circ & -(R^2+2A^{\circ 2}U) & -2A^\circ B^\circ U \\ 2B^\circ U & -B^\circ & -2A^\circ B^\circ U & -(R^2+2B^{\circ 2}U) \end{vmatrix}, \tag{2.62}$$

where

$$U = -\mu^\circ + \rho\bar\rho(r\Psi^\circ_{2x} + 2\mu^\circ\,|\rho^\circ|^2), \qquad \rho = -(r+\rho^\circ)^{-1},$$
$$2\mu^\circ = 1, 0, -1, \qquad \rho\bar\rho = 1/(r^2 + |\rho^\circ|^2),$$

ρ° is pure imaginary and constant,

Ψ°_{2x} is a real constant,

$$A^\circ = |\rho^\circ|\,x^4/\sqrt{2}\,p, \qquad B^\circ = -|\rho^\circ|\,x^3/\sqrt{2}\,p,$$
$$\sqrt{2}\,p = 1 + \tfrac{1}{2}\mu^\circ\zeta\bar\zeta, \qquad \mu^\circ = \tfrac{1}{2}, 0, -\tfrac{1}{2},$$
$$R^{-2} = 2\rho\bar\rho p^2 = (-\mathrm{Det}\, g^{\mu\nu})^{\frac{1}{2}},$$
$$\xi^{\circ 1} = p(A^\circ + iB^\circ), \qquad \xi^{\circ 3} = -i\xi^{\circ 4} = p.$$

There is a four-parameter group of motions which is described by the following solutions to Killing's equation ($\xi_{\mu;\nu} + \xi_{\mu;\nu} = 0$):

$$\xi^\mu_t = [1, 0, 0, 0], \qquad \mu = 1, 2, 3, 4, \tag{2.63}$$
$$\xi^\mu_a = [-2\,|\rho^\circ|, 0, -y, x], \tag{2.64}$$
$$\xi^\mu_b = [-2\,|\rho^\circ|\,y, 0, 2 + \mu^\circ(x^2 - y^2), 2\mu^\circ xy], \tag{2.65}$$
$$\xi^\mu_c = [2\,|\rho^\circ|\,x, 0, 2\mu^\circ xy, 2 - \mu^\circ(x^2 - y^2)], \tag{2.66}$$

where

$$x \equiv x^3 \quad \text{and} \quad y \equiv x^4.$$

These metrics are an extension of metrics obtained by Taub.[9] The solution with $\mu^\circ = \frac{1}{2}$ passes to the Schwarzschild metric in the limit $\rho^\circ = 0$. The remainder of this paper will be devoted to this metric.

III. GENERALIZED SCHWARZSCHILD SOLUTION

The solution (2.61) for $2\mu^\circ = 1$ yields the Schwarzschild metric in the limit $\rho^\circ = 0$. This solution can be expressed in spherical coordinates by use of the following transformation:

$$r' = r, \qquad t = x^1 - \int (2U)^{-1}\,dr,$$
$$\sin\theta = \zeta\bar\zeta/(1 + \tfrac{1}{4}\zeta\bar\zeta), \qquad \tan\phi = x^4/x^3,$$
$$\Psi^\circ_{2x} \to \tfrac{1}{2}\Psi^\circ_{2x}. \tag{3.1}$$

The transformed covariant and contravariant metric are

$$ds^2 = -U[dt + 4\,|\rho^\circ|\,(\sin^2\tfrac{1}{2}\theta)\,d\phi]^2 + dr^2/U - (d\theta^2 + \sin^2\theta\,d\phi^2)/\rho\bar\rho, \tag{3.2}$$

and

$$g^{\mu\nu} = \begin{vmatrix} g^{11} & 0 & 0 & 4\,|\rho^\circ|\,\rho\bar\rho\,\csc^2\theta\sin^2\tfrac{1}{2}\theta \\ 0 & U & 0 & 0 \\ 0 & 0 & -\rho\bar\rho & 0 \\ g^{14} & 0 & 0 & -\rho\bar\rho\,\csc^2\theta \end{vmatrix}, \tag{3.3}$$

where

$$x^1 \equiv t, \qquad x^2 \equiv r, \qquad x^3 \equiv \theta, \qquad x^4 \equiv \phi,$$
$$\rho\bar\rho = (r^2 + |\rho^\circ|^2)^{-1},$$
$$U = -1 + \rho\bar\rho(r\Psi^\circ_{2x} + 2\,|\rho^\circ|^2),$$
$$g^{11} = -U^{-1} - 16\,|\rho^\circ|^2\,\rho\bar\rho(\sin^4\tfrac{1}{2}\theta)\,\csc^2\theta,$$
$$\Psi^\circ_{2x} \text{ and } |\rho^\circ| \text{ are constants.}$$

Expressed in the new coordinate system the Killing vectors are

$$\xi^\mu_t = (1, 0, 0, 0), \tag{3.4}$$
$$\xi^\mu_a = (-2\,|\rho^\circ|, 0, 0, 1), \tag{3.5}$$
$$\xi^\mu_b = [-2\,|\rho^\circ|\sin\phi(\tan\tfrac{1}{2}\theta), 0, \cos\phi, -\sin\phi\,\cot\theta], \tag{3.6}$$
$$\xi^\mu_c = [2\,|\rho^\circ|\cos\phi(\tan\tfrac{1}{2}\theta), 0, \sin\phi, \cos\phi\,\cot\theta]. \tag{3.7}$$

[9] A. H. Taub, Ann. Math., **53**, 472 (1951). For a discussion of the relation of the metrics of the present paper to those of Taub, and for an analysis of these metrics, see C. Misner, J. Math. Phys. 4, 924 (1963) (following paper). A special class of fields of this type has been constructed by J. Ehlers in 1957 (thesis, Hamburg University Sec. 32); his method has been described at the Royaumont Conference.

The Killing fields associated with the Schwarzschild solution are obtained by setting $\rho^\circ = 0$.

In order to gain physical insight, it is convenient to consider a test particle whose Lagrangian, four-momentum, and Hamiltonian are:

$$L = \tfrac{1}{2} g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu, \tag{3.8}$$

$$P_\mu = \partial L/\partial \dot{x}^\mu = g_{\mu\nu}\dot{x}^\nu, \tag{3.9}$$

$$H = \tfrac{1}{2} g_{\mu\nu} P^\mu P^\nu, \tag{3.10}$$

where $\dot{x}^\mu$ is the proper time derivative of x . Another form of Killing's equation is

$$\begin{aligned}\tilde{\delta} g^{\mu\nu} &\equiv g^{\mu\nu\prime}(x^\rho) - g^{\mu\nu}(x^\rho) \\ &\equiv g^{\mu\rho}\xi^\nu_{A,\rho} + g^{\nu\rho}\xi^\mu_{A,\rho} - g^{\mu\nu}_{,\rho}\xi^\rho_A = 0.\end{aligned} \tag{3.11}$$

Killing generators are defined by

$$G_A \equiv P_\mu \xi^\mu_A. \tag{3.12}$$

With definitions (3.10) and (3.12), one can readily readily show that Eq. (3.11) is equivalent to the Poisson bracket relation

$$\{G_A, H\} = 0. \tag{3.13}$$

Thus the Killing generators are constants of the motion.

The Killing generators obtained from Eqs. (3.4)–(3.7) are:

$$G_t = P_1, \tag{3.14}$$

$$G_a = -2\,|\rho^\circ|\,P_1 + P_4, \tag{3.15}$$

$$G_b = -2\,|\rho^\circ|\,P_1 \sin\phi(\tan\tfrac{1}{2}\theta) + P_3\cos\phi - P_4 \sin\phi\cot\theta, \tag{3.16}$$

$$G_c = 2\,|\rho^\circ|\,P_1\cos\phi(\tan\tfrac{1}{2}\theta) + P_3\sin\phi + P_4\cos\phi\cot\theta. \tag{3.17}$$

These four Killing generators are constants of the motion. In particular, the energy of the test particle, P_1, and the ϕ component of angular momentum, P_4, are constants of the motion.

The following Poisson bracket relations are readily obtained:

$$\{G_t, G_A\} = 0, \qquad A = a, b, c \tag{3.18}$$

$$\{G_a, G_b\} = G_c, \qquad a, b, c \text{ in cyclic order.} \tag{3.19}$$

A natural definition for the angular momenta is the generators of rotations about the three axes, which happen to be the Killing generators of the Schwarzschild solution:

$$M_a = P_4, \tag{3.20}$$

$$M_b = P_3\cos\phi - P_4\sin\phi\cot\theta, \tag{3.21}$$

$$M_c = P_3\sin\phi - P_4\cos\phi\cot\theta. \tag{3.22}$$

From

$$G_a - M_a = -2\,|\rho^\circ|\,P_1, \tag{3.23}$$

$$G_b - M_b = -2\,|\rho^\circ|\,P_1\sin\phi\tan\tfrac{1}{2}\theta, \tag{3.24}$$

$$G_c - M_c = 2\,|\rho^\circ|\,P_1\cos\phi\tan\tfrac{1}{2}\theta, \tag{3.25}$$

one sees that the components of angular momentum differ from the constants of motion by terms proportional to $|\rho^\circ|\,P_1$.

IV. TEST-PARTICLE TRAJECTORIES

Since Ψ°_{2x} is proportional to the mass in the Schwarzschild limit $\rho^\circ = 0$, and corresponds to the Bondi[10] and Newman–Unti[11] definitions of mass, it will be assumed here that Ψ°_{2x} is proportional to the mass of the source even when $\rho^\circ \neq 0$;

$$\Psi^\circ_{2x} = 2\kappa M,$$

where κ is the gravitational constant. It is of interest to find the motion of a test particle in this generalized Schwarzschild field. The orbit of the test particle of mass m, will be calculated by a perturbation method often used in celestial mechanics.

First, the method will be illustrated by application to the special case $|\rho^\circ| = 0$; i.e., to the Schwarzschild metric. When $|\rho^\circ|$ is set equal to zero, the metric (3.2) becomes

$$g_{\mu\nu} = \begin{vmatrix} e^\nu & \cdot & \cdot & \cdot \\ \cdot & -e^{-\nu} & \cdot & \cdot \\ \cdot & \cdot & -r^2 & \cdot \\ \cdot & \cdot & \cdot & -r^2\sin^2\theta \end{vmatrix}, \tag{4.1}$$

where $e^\nu \equiv 1 - (2\kappa M/r)$. The "action" for the test-particle problem is

$$S = -m\int ds^2 = -m\int (g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu)^{\frac{1}{2}}\,dt, \quad \dot{x}^\mu \equiv dx^\mu/dt.$$

[The development here differs from that given in an earlier section, Eqs. (3.8)–Eqs. (3.10). The three-dimensional notation, $i = 2, 3, 4$ will be used in this section and $x^1 \equiv t$ will be used as the parameter along the test particle instead of the proper time.] The Lagrangian is then

$$L = -m(g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu)^{\frac{1}{2}} = -m(e^\nu + v^2)^{\frac{1}{2}}, \tag{4.2}$$

where $v^2 \equiv g_{ij}\dot{x}^i\dot{x}^j$. The generalized momenta are defined by

$$p_k = \frac{\partial L}{\partial \dot{x}^k} = -m\,\frac{g_{kj}\dot{x}^j}{(e^\nu + v^2)^{\frac{1}{2}}}. \tag{4.3}$$

[10] H. Bondi, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[11] E. Newman and T. Unti, J. Math. Phys. **3**, 891 (1962).

This may be solved for v^2 in terms of $p^2 \equiv g^{ij}p_ip_j$;

$$v^2 = e^{\nu}p^2/(m^2 - p^2). \tag{4.4}$$

The Hamiltonian is given by

$$H = p_kv^k - L = me^{\nu}(e^{\nu} + v^2)^{-\frac{1}{2}} = e^{\frac{1}{2}\nu}(m^2 - p^2)^{\frac{1}{2}}, \tag{4.5}$$

The expansion parameters will be

$$p^2, \quad \kappa M = O(\epsilon), \tag{4.6}$$

where ϵ is small. (This corresponds to slow motion as well as a small gravitational constant κ.) The Hamiltonian, expanded through terms of $O(\epsilon^2)$, becomes

$$H = m - \frac{p^2}{2m} - \frac{\kappa mM}{r} - \frac{p^4}{8m^3} + \frac{\kappa Mp^2}{2mr} - \frac{m\kappa^2M^2}{2r^2} + O(\epsilon^3), \tag{4.7}$$

with

$$p^2 = -[(1 - 2\kappa Mr^{-1})p_r^2 + r^{-2}(p_\theta^2 + \csc^2\theta p_\phi^2)].$$

Hence,

$$-\frac{p^2}{2m} = \frac{p_0^2}{2m} - \frac{\kappa M}{mr}p_r^2 + O(\epsilon^3), \qquad -\frac{p^4}{8m^3} = -\frac{p_0^4}{8m^3} + O(\epsilon^3), \tag{4.8}$$

where $p_0^2/2m$ is the usual Newtonian kinetic energy. The Hamiltonian is now

$$H = m + \left(\frac{p_0^2}{2m} - \frac{\kappa mM}{r}\right) - \left[\frac{p_0^4}{8m^3} + \frac{\kappa Mp_r^2}{mr} + \frac{\kappa Mp_0^2}{2mr} + \frac{m\kappa^2M^2}{2r^2}\right] = m + H_0 + H_1. \tag{4.9}$$

H_0 represents the ordinary nonrelativistic Hamiltonian

$$H_0 = T + V,$$

$$T \equiv \frac{p_0^2}{2m}, \qquad V \equiv -\frac{\kappa mM}{r}.$$

The relativistic effects are carried in the perturbation term

$$H_1 = -\left[\frac{p_0^4}{8m^3} + \frac{\kappa Mp_r^2}{mr} + \frac{\kappa Mp_0^2}{2mr} + \frac{m\kappa^2M^2}{2r^2}\right]. \tag{4.10}$$

The action-angle variables associated with the unperturbed orbits are defined by

$$J_r = \oint p_r\,dr, \quad J_\theta = \oint p_\theta\,d\theta, \quad J_\phi = \oint p_\phi\,d\phi, \tag{4.11}$$

the integrals being taken over the unperturbed paths. Of more interest are the following linear combinations:

$$J_1 = J_r + J_\theta + J_\phi, \; J_2 = J_\theta + J_\phi, \; J_3 = J_\phi. \tag{4.12}$$

In physical terms,[12] J_1 is a unique function of the total unperturbed energy of the two-body interaction (neglecting the rest energy). J_2 is proportional to the total angular momentum of the particle, and J_3 is 2π times the component of the angular momentum along the polar axis. The coordinates conjugate to $J_k (k = 1, 2, 3)$, the "angle variables", are given in Hamilton–Jacobi theory by the transformation equations

$$\omega_k = \partial W/\partial J_k, \tag{4.13}$$

where W is the unperturbed characteristic function. The equations of motion for the angle variables are

$$\dot{\omega}_k = \partial H(J_m)/\partial J_k. \tag{4.14}$$

The importance of the angle variables lies in their simple geometrical interpretation. ω_2 is the angle between the perihelion and the line of nodes (in the plane of the orbit). The line of nodes itself makes an angle with the fixed axis $\phi = 0$ in the equatorial plane; this is given by ω_3. If the time derivatives of these variables [given by Eq. (4.14)] are multiplied by $2\pi T$, where T is one period of revolution of the unperturbed test particle, one obtains

$2\pi T\dot{\omega}_2$, the advance of the perihelion of the orbit per revolution;

$2\pi T\dot{\omega}_3$, the rotation of the line of nodes in one revolution (precession of the plane of the orbit).

It is assumed for these results, that the system has not changed much due to the perturbation, in one period.

The details of this calculation follow. Equations (4.11) are put into Eqs. (4.12), and solved[13] for the momenta P_i. The momenta in turn are substituted into the perturbation term of the Hamiltonian [Eq. (4.10)]. H_1 is then averaged over one period of the motion,

$$\langle H_1\rangle = \frac{1}{T}\int_0^T H_1\,dt.$$

The necessary relations, listed below, for taking

[12] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Company Inc., Reading, Massachusetts, 1951), Chap. 9.

[13] M. Born, *The Mechanics of the Atom* (Frederick Ungar Publishing Company, New York, 1960), Chap. 3.

the time average $\langle H_1\rangle$ are derived, for example, by Born[13]:

$$dt = (r^2T/2\pi ab)\,d\Psi, \quad \cos\theta = \sin\gamma\sin\Psi,$$
$$\cos\gamma = J_3/J_2, \qquad (4.15)$$
$$\langle r^{-1}\rangle = 1/a, \quad \langle r^{-2}\rangle = 1/ab, \quad \langle r^{-3}\rangle = 1/b^3,$$

where Ψ is the angular distance (in the plane of the orbit) of the moving test particle from the line of nodes, γ is the angle between the normal to the plane and the polar axis,

$$a \equiv J_1^2/4\pi^2\kappa m^2 M, \qquad b \equiv J_1J_2/4\pi^2\kappa m^2 M.$$

When the momenta are found as functions of the coordinates and the J_k, and the coordinates are averaged over one period of the motion, the perturbation term [Eq. (4.10)] becomes

$$\langle H_1\rangle = -\frac{\kappa M}{2ab}\left[8\kappa mM - \frac{J_2^2}{2\pi^2 m}\left(\frac{a}{b^2}\right)\right]$$
$$+ \text{(terms independent of } J_2 \text{ and } J_3). \qquad (4.16)$$

Thus for a Schwarzschild source, the advance of the perihelion of a test particle is

$$2\pi T\dot{\omega}_2 = 2\pi T\frac{\partial\langle H_1\rangle}{\partial J_2} = \frac{24\pi^3 m^2}{J_2^2}\kappa^2 M^2. \qquad (4.17)$$

Expressed in terms of the orbit parameters this becomes

$$2\pi T\dot{\omega}_2 = 24\pi^3 a^2/(1-e^2)T^2, \qquad (4.18)$$

where a is the semimajor axis of the test-particle orbit, e is the numerical eccentricity, and T is the period of revolution. Equations (4.17) or (4.18) describe, for example, the advance of the perihelion of Mercury.

Because J_3 does not appear in $\langle H_1\rangle$, there is no precession of the plane of the orbit.

The same approach will now be applied to the case where the parameter $|\rho^\circ|$, which is a measure of the curl of the geodesic ray l_μ, is not zero. The more general metric is Eq. (3.2);

$$g_{\mu\nu} = \begin{vmatrix} -U & 0 & 0 & g_{14} \\ 0 & 1/U & 0 & 0 \\ 0 & 0 & -1/\rho\bar{\rho} & 0 \\ g_{14} & 0 & 0 & g_{44} \end{vmatrix}, \qquad (4.19)$$
$$g_{14} = -4U\,|\rho^\circ|\sin^2\tfrac{1}{2}\theta,$$
$$g_{44} = -16U\,|\rho^\circ|^2\sin^4\tfrac{1}{2}\theta - \sin^2\theta/\rho\bar{\rho}.$$

The equations leading from the Lagrangian to the Hamiltonian are as follows:

$$L = -m[g_{11} + 2g_{1k}\dot{x}^k + g_{mn}\dot{x}^m\dot{x}^n]^{\frac{1}{2}},$$
$$p_k = \partial L/\partial\dot{x}^k = (m^2/L)(g_{kj}\dot{x}^j + g_{1k}), \qquad (4.20)$$
$$\dot{x}_k = (L/m^2)\tilde{g}^{kn}p_n - \tilde{g}^{kn}g_{n1},$$

where

$$\tilde{g}^{ij} = \begin{vmatrix} U & 0 & 0 \\ 0 & -\rho\bar{\rho} & 0 \\ 0 & 0 & 1/g_{44} \end{vmatrix}, \qquad \tilde{g}^{ij}g_{ik} = \delta^j_k, \qquad (4.21)$$

$$H = p_k\dot{x}^k - L = (g_{11} - \tilde{g}^{ij}g_{1i}g_{1j})^{\frac{1}{2}}(m^2 - p^2)^{\frac{1}{2}} - (1/m)\tilde{g}^{ij}g_{1i}p_j.$$

In this case there are three expansion parameters,

$$p^2, \quad \kappa M, \quad |\rho^\circ| = O(\epsilon),$$

where ϵ is small. The Hamiltonian, expanded through terms of $O(\epsilon^2)$, is now

$$H = m - \frac{p^2}{2m} - \frac{\kappa mM}{r} - \frac{p^4}{8m^3} + \frac{\kappa Mp^2}{2mr} - \frac{m\kappa^2M^2}{2r^2}$$
$$+ \frac{m}{r^2}|\rho^\circ|^2(2\tan^2\tfrac{1}{2}\theta - 1)$$
$$+ \frac{p_\phi|\rho^\circ|}{r^2\cos^2\tfrac{1}{2}\theta} + O(\epsilon^{5/2}). \qquad (4.22)$$

It is seen that

$$H = H^s + \frac{m}{r^2}|\rho^\circ|^2(2\tan^2\tfrac{1}{2}\theta - 1)$$
$$+ \frac{p_\phi|\rho^\circ|}{r^2\cos^2\tfrac{1}{2}\theta} + O(\epsilon^{5/2}), \qquad (4.23)$$

where H^s is the Schwarzschild Hamiltonian, [Eq. (4.7)]. Thus the perturbation term is

$$H_1 = H_1^s + \frac{m}{r^2}|\rho^\circ|^2(2\tan^2\tfrac{1}{2}\theta - 1)$$
$$+ \frac{p_\phi|\rho^\circ|}{r^2\cos^2\tfrac{1}{2}\theta}. \qquad (4.24)$$

The time average of H_1^s is given by Eq. (4.16). The time average of the remaining terms in Eq. (4.24) may be calculated by use of the relations (4.15). One obtains

$$\frac{1}{T}\int_0^T\frac{dt}{r^2\cos^2\tfrac{1}{2}\theta} = \frac{1}{\pi ab}\int_0^{2\pi}\frac{d\Psi}{1+\sin i\sin\Psi} = \frac{2}{ab}\sec i = \frac{2J_2}{abJ_3},$$

$$\frac{1}{T}\int_0^T\frac{1}{r^2}\tan^2\tfrac{1}{2}\theta\,dt = \frac{1}{ab}(2\sec i - 1) = \frac{1}{ab}\left(\frac{2J_2}{J_3} - 1\right).$$

When these relations are used in Eq. (4.24), the result is

$$\langle H_1\rangle = \langle H_1^s\rangle + \frac{(4\pi^2\kappa m^2 M)^2}{J_1^3 J_2} \times \left\{\frac{J_2}{J_3}\left[4m\,|\rho^\circ|^2 + \frac{|\rho^\circ|\,J_3}{\pi}\right] - 3m\,|\rho^\circ|^2\right\}. \quad (4.25)$$

Hence the advance of perihelion of the orbit for a test particle in the field of a "generalized Schwarzschild" source is

$$2\pi T\,\frac{\partial\langle H_1\rangle}{\partial J_2} = \frac{24\pi^3 m^2}{J_2^2}\,(\kappa^2 M^2 + |\rho^\circ|^2). \quad (4.26)$$

This may be compared to the static Schwarzschild result [Eq. (4.17)] to see the effect of the "curling" term, $|\rho^\circ|$.

An additional effect which occurs in the generalized case is the rotation of the line of nodes. The advance of the line of nodes in one revolution is

$$2\pi T\,\frac{\partial\langle H_1\rangle}{\partial J_3} = -\frac{32\pi^3 m^2}{J_3^2}\,|\rho^\circ|^2 = -\frac{32\pi^3 m^2}{J_2^2\cos^2\gamma}\,|\rho^\circ|^2. \quad (4.27)$$

ACKNOWLEDGMENTS

The authors are indebted to A. Janis, I. Robinson, R. Sachs, and J. Stachel for many stimulating discussions and suggestions, and to R. Kerr and A. Thompson for having caught a serious error in the original manuscript.

The Flatter Regions of Newman, Unti, and Tamburino's Generalized Schwarzschild Space*

CHARLES W. MISNER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 17 December 1962)

The "generalized Schwarzschild" metric discovered by Newman, Unti, and Tamburino, which is stationary and spherically symmetric, is investigated. We find that the orbit of a point under the group of time translations is a circle, rather than a line as in the Schwarzschild case. The timelike hypersurfaces r = const which are left invariant by the group of motions are topologically three-spheres S^3, in contrast to the topology $S^2 \times R$ (or $S^2 \times S^1$) for the r = const surfaces in the Schwarzschild case. In the Schwarzschild case, the intersection of a spacelike surface t = const and an r = const surface is a sphere S^2. If σ is any spacelike hypersurface in the generalized metric, then its (two-dimensional) intersection with an r = const surface is not any closed two-dimensional manifold, that is, the generalized metric admits no reasonable spacelike surfaces. Thus, even though all curvature invariants vanish as $r \to \infty$, in fact $R_{\mu\nu\alpha\beta} = O(1/r^3)$ as in the Schwarzschild case, this metric is not asymptotically flat in the sense that coordinates can be introduced for which $g_{\mu\nu} - \eta_{\mu\nu} = O(1/r)$. An apparent singularity in the metric at small values of r, which appears to be similar to the spurious Schwarzschild singularity at $r = 2m$, has not been studied. If this singularity should again be spurious, then the "generalized Schwarzschild" space would represent a terminal phase in the evolution of an entirely nonsingular cosmological model which, in other phases, contains closed spacelike hypersurfaces but no matter.

I. INTRODUCTION

THE primary purpose of this paper is to study and describe geometrically the stationary, spherically symmetric solution of the Einstein equations recently discovered by Newman, Unti, and Tamburino[1] which I shall refer to as NUT space. A second important purpose of this work is to provide an example of the recognition and elimination of a spurious singularity in a Riemannian line element with the Lorentz $-+++$ signature. No general method is known for eliminating coordinate singularities in a metric, nor are there adequate criteria to determine that a singularity is not merely a coordinate singularity, and I expect that further examples beyond the Kruskal–Fronsdal[2,3] elimination of the Schwarzschild singularity will be helpful in leading to an understanding of these problems. A third, minor, aim of this paper is to provide an example of the use of orthonormal frames (tetrads) in a style more economical with indices than is usual in the literature of physics, and in particular a method of computing the curvature tensor very rapidly (cf. Appendix A).

The question of singularities in metrics is broader and more important than the study of one particular metric to which most of this paper is devoted, so I will briefly summarize the present state of the art.[4] The first step is to find some clearly stated problems, and the clue to clarity is to refuse ever to speak of a singularity but instead to phrase everything in terms of the properties of differentiable metric fields on manifolds. If one is given a manifold, and on it a metric which does not at all points satisfy the necessary differentiability requirements, one simply throws away all the points of singularity. The starting point for any further discussion is then the largest submanifold on which the metric *is* differentiable. This is done because there is not known any useful way of describing the singularities of a function except by describing its behavior at regular points near the singularity. The first problem then is to select a criteria which will identify in an intuitively acceptable way a "nonsingular space." Evidently, differentiability is only a minimum prerequisite, since everything becomes differentiable when the singular points are discarded. The problem is rather to recognize the holes left in the space where singular (or even regular) points have been omitted. For a connected Riemannian manifold

* Supported in part by the U. S. Air Force Office of Scientific Research, Air Research and Development Command.

[1] E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. 4, 915 (1963). I wish to thank these authors for sending me a preprint of their paper.

[2] M. D. Kruskal, Phys. Rev. 119, 1743 (1960).

[3] C. Fronsdal, Phys. Rev. 116, 778 (1959).

[4] I wish to thank Mr. L. Shepley for preparing this review and for correcting numerous errors in an earlier draft. We have borrowed heavily from L. Marcus' lectures on this topic at the American Mathematical Society's 1962 Summer Institute at the University of California at Santa Barbara.

with a *positive-definite* metric, the following criteria are equivalent[5]:

(1) (Complete metric topology) *Every Cauchy sequence converges* (to a point in the space). Here the distance $d(x, y)$ between two points is defined as the greatest lower bound of the lengths of curves joining the two points, and a Cauchy sequence (or fundamental sequence) $x_1, x_2, \cdots$ is one where, for any $\epsilon > 0$, one can satisfy $d(x_n, x_m) < \epsilon$ for all sufficiently large m, n. This is the criterion which shows that the space of all rational numbers is full of holes which should be filled in with real numbers.

An equivalent statement of this metric topology definition of completeness is that *every bounded closed set in the manifold is compact.*

(2) (Complete affine connection) *Every geodesic can be continued* in both directions to infinite values of the affine path parameter.

The reasonableness of the above definitions of nonsingular positive-definite Riemannian spaces is reinforced by their equivalence, by the fact that if one point be removed from any Riemannian manifold, the result is certainly not complete, and by the following theorems:

Theorem I[6]. A complete connected Riemannian manifold is *inextensible,* i.e., it is not a proper open submanifold of any connected Riemannian manifold.

Theorem II[5]. A compact, connected manifold with a positive-definite Riemannian metric is complete.

Theorem III[5]. A complete, connected, positive-definite Riemannian metric manifold, is *geodesically convex,* i.e., any two points can be joined by a geodesic.

If we turn now to the case of the Lorentz $(-+++)$ signature, the results are as follows: The metric does not define a (positive-definite) distance function, so the criterion (1) of completeness is an obvious metric topology is not available.[7] Affine completeness [criterion (2)] continues to make sense and is a sufficient condition for a space to be "nonsingular," since Theorem I is true with this sense of completeness and without the positive-definite requirement on the metric. (The proof[6] involves only the idea of affine completeness: If M is a proper open submanifold of a connected affine manifold M', then a point p on the boundary of M in M' has geodesics emanating from it which enter M. These geodesics, followed backwards from a point in M cannot have infinite length in M since p is not in M. Hence M is not affinely complete.) However affine completeness cannot be considered necessary in order that a space be "nonsingular" since Theorem II fails without the positive definite signature assumption; I will give one of several known counterexamples below. Finally, Theorem III is also false in the indefinite-metric cases; Calabi and Markus[8] note that de Sitter space, which is affinely complete, has many pairs of points which cannot be joined by geodesics.

As a counterexample to Theorem II in the case of an indefinite metric, consider the metric

$$ds^2 = \cos x\,(dy^2 - dx^2) + 2 \sin x\, dx\, dy \qquad (1)$$

which is analytic on the torus $0 \leq x \leq 2\pi$, $0 \leq y \leq 2\pi$. Since the metric components are independent of y, the geodesic equation is easily integrated. In addition to the constant of motion

$$g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = \epsilon = 0, \pm 1,$$

there is a second constant,

$$p = \dot{y} \cos x + \dot{x} \sin x.$$

Integration of the relationship

$$\dot{x}^2 + \epsilon \cos x = p^2$$

then yields all geodesics explicitly. One finds that there are exactly four closed geodesics on each of which x is constant, namely $x = 0$ (spacelike), $x = \frac{1}{2}\pi$ (null), $x = \pi$ (timelike), and $x = \frac{3}{2}\pi$ (null). *No* other geodesic in this space can be extended to infinite values of the affine path parameter. In fact, the affine parameter λ on the two closed null geodesics is given by $y = \ln \lambda^2$, so even these two closed null geodesics are not extendable from positive to negative values of the affine parameter.

Although we lack a definition of "nonsingular" for Lorentz-signature Riemannian manifolds, we have two sufficient conditions: compactness and affine completeness. In addition, one criteria for an "essentially singular" space is commonly accepted; it presumes lack of affine completeness and implies that the space cannot be extended to either an affinely complete or a compact space. This criteria requires that some scalar polynomial in

[5] H. Hopf and W. Rinow, Comment. Math. Helv. **3**, 209 (1931). See also T. J. Willmore, *An Introduction to Differential Geometry* (Oxford University Press, England, 1959), p. 133 (although the theorem is here stated for a 2-surface, the proof given does not use this assumption); S. S. Chern, *Differentiable Manifolds* (Department of Mathematics, University of Chicago, 1959), p. 121; S. Sternberg, *Lectures on Differential Geometry* (Department of Mathematics, Harvard University; Prentice-Hall, Inc., 1961), Sec. 4–7; S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press, Inc., New York, 1962), p. 55.

[6] See, for example, S. S. Chern (reference 5), p. 122.

[7] Avez has considered some ways of introducing a completeness idea analogous to criterion (1). As yet unpublished, they supercede his proposal in Compt. Rend. **240**, 485 (1955).

[8] E. Calabi and L. Marcus, Ann. Math. **75**, 63 (1962).

the curvature tensor and its covariant derivatives be unbounded on an (open) geodesic segment of finite length. It is not presumed that all spaces which should be called "essentially singular" are identified by this criteria.

Let us now turn to the example which will occupy our attention for the remainder of the paper, and recapitulate the properties which Newman, Unti, and Tamburino[1] have established for it. The line element presented by Newman et al. is

$$ds^2 = -f^2(r)[dt + 4l\sin^2\tfrac{1}{2}\theta\, d\phi]^2 + (1/f^2)\, dr^2 + (r^2 + l^2)(d\theta^2 + \sin^2\theta\, d\phi^2), \quad (2)$$

where

$$f^2 = 1 - 2[(mr + l^2)/(r^2 + l^2)]. \quad (3)$$

This metric satisfies the empty-space Einstein field equations

$$R_{\mu\nu} = 0, \quad (4)$$

and was obtained in a systematic search for all solutions where the principle null vectors of the Riemann curvature tensor are tangent to a congruence of null geodesics which has vanishing shear but nonzero divergence and curl. The Riemann tensor is of Petrov type D (type I degenerate), or in Penrose's more informative notation, type [2, 2].[9] (This search for solutions of a specified character is an excellent example of the power of the Newman–Penrose spin coefficient formalism.) If $l = 0$, the NUT metric (2) reduces obviously to the Schwarzschild metric; that these metrics are distinct when $l \neq 0$ can be seen from the curvature tensor. [Equation (2.51) in reference 1 shows that the only nonzero curvature component, ψ_2, in the spinor formalism is real in the Schwarzschild case, and complex when $l \neq 0$. Since the spinor basis (tetrad) has been picked out by geometric conditions, the curvature components are effectively invariants.] The line element (2) has a Schwarzschild-like singularity when $f^2 = 0$, but this singularity is not observed in the curvature tensor. There are four linearly independent Killing vectors ξ_x, ξ_y, ξ_z, and ξ_t which Newman *et al.* exhibit and which satisfy the same commutation relations as in the Schwarzschild case, namely

$$[\xi_a, \xi_b] = -\epsilon_{abc}\xi_c, \quad (5)$$

and

$$[\xi_a, \xi_t] = 0, \quad (6)$$

where each a, b, c has the range of values x, y, z.

[9] See, for an exposition and further references, F. A. E. Pirani's article in *Gravitation: an introduction to current research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962), Sec. 6-5.

In Sec. II below, we discuss how the $\theta = \pi$ singularity in the line element in Eq. (2) can be identified and blamed on the time coordinate. In Sec. III we sidestep this singularity by group theory techniques (thanks to the high symmetry of this metric) and find a nonsingular presentation of the metric. In Sec. IV this same singularity is eliminated by means of a singular coordinate transformation, thus illustrating a technique which does not rely on symmetry. In Sec. V a theorem about the nonexistence of spacelike surfaces is stated and proven. Section VI contains further discussion, particularly concerning the possible relationship of this metric to a cosmological model.

II. RECOGNITION OF SINGULARITIES

In studying the metric (2) I will use as basis vectors to which tensor components will normally refer, not the coordinate axes, but an orthonormal tetrad. As covariant basis vectors[10] I take the following four vectors:

$$\begin{aligned} \omega^0 &= f(r)[dt + 4l\sin^2\tfrac{1}{2}\theta\, d\phi], \\ \omega^1 &= f^{-1}\, dr, \\ \omega^2 &= (r^2 + l^2)^{\frac{1}{2}}\, d\theta, \\ \omega^3 &= (r^2 + l^2)^{\frac{1}{2}}\sin\theta\, d\phi. \end{aligned} \quad (7)$$

Here f is as given in Eq. (3). [Equations (7) show covariant vectors σ_μ by displaying the corresponding differential form, $\sigma \equiv \sigma_\mu\, dx^\mu$. In fact, differential forms and covariant vectors are commonly identified by the device of interpreting d as the gradient operator. Then, for instance, dx^0 is the vector which is the gradient of a function $x^0(P)$ whose value at a point P is used as one of the coordinates of that point.] Evidently, the NUT metric of Eq. (2) can be written

$$ds^2 \equiv g_{\mu\nu}\omega^\mu\omega^\nu = -(\omega^0)^2 + (\omega^1)^2 + (\omega^2)^2 + (\omega^3)^2. \quad (8)$$

In the ω^μ frame then, the metric components $g_{\mu\nu}$ take on the Lorentz values, $g_{\mu\nu} = \eta_{\mu\nu}$; this is equivalent to the statement that the base vectors ω^μ are orthonormal. A computation of the Riemann tensor components in this orthonormal frame is given in Appendix A. From this computation we reproduce certain results of Newman *et al.* We can

[10] The notation and computational methods being used here are due to E. Cartan; see his text, *Leçons sur la Geometrie des Espaces de Riemann*, (Gauthier-Villars, Paris, 1951), 2nd Ed. The conceptual basis of this approach has been much clarified by C. Chevalley in his *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), Chap. III. For an exposition, see T. J. Willmore [cf. reference 5, Chaps. V and VI].

verify that NUT space is distinct from the Schwarzschild metric when $l \neq 0$, since then

$$R_{\mu\ \nu}^{\ \alpha\beta}(-g)^{\frac{1}{2}}\epsilon_{\alpha\beta\sigma\tau}R^{\mu\nu\sigma\tau} \neq 0, \tag{9}$$

while this invariant vanishes for the Schwarzschild metric. We can also see from the curvature components given in Eq. (A14) that NUT space is asymptotically flat in the sense that the Riemann tensor vanishes (like $1/r^3$) as $r \to \infty$ [from Eq. (2) a curve of constant t, θ, ϕ is spacelike, and along this curve r is asymptotically a measure of proper distance s]. Note particularly that the curvature components in this orthonormal frame, and therefore all invariants formed from the Riemann tensor, depend only on r, and not on the other coordinates. Note also that these components are analytic functions of r for $-\infty < r < +\infty$, so we have no evidence of any intrinsic singularities in this metric. (If $l = 0$, there is a singularity apparent at $r = 0$, the singular end point of an idealized collapsing neutron star, for example, which can be described[11] by the Schwarzschild solution.)

Although no singularities are evident in the curvature, it is still necessary to find what range of values for the coordinates leaves the metric nonsingular. Before attacking this question systematically, it is useful to look first at the time coordinate. Equations (7) are readily inverted to give

$$dt = f^{-1}\omega^0 - 2l(r^2 + l^2)^{-\frac{1}{2}} \tan \tfrac{1}{2}\theta\, \omega^3. \tag{10}$$

From this expansion of grad t in terms of the orthonormal basis vectors ω^μ, we compute directly

$$-(\text{grad } t)^2 = f^{-2} - (2l)^2(r^2 + l^2)^{-1} \tan^2 \tfrac{1}{2}\theta. \tag{11}$$

Thus $(\text{grad } t)^2$ becomes infinite at $\theta = \pi$, so either the metric or the t coordinate must be considered singular there. Note also that the hypersurfaces $t = \text{const}$ are spacelike only for sufficiently small θ,

$$\theta < \theta_0 = 2 \tan^{-1} [(r^2 + l^2)^{\frac{1}{2}}/2lf], \tag{12}$$

since only then does the vector dt normal to these surfaces remain timelike.

Because of standard singularities in spherical coordinate systems, we of course cannot expect that the metric components with respect to the tr $\theta\phi$ coordinate frame be regular. We must rather introduce "rectangular" coordinates xyz related to $r\theta\phi$ in the standard way before discussing differentiability. The computations are shortened by recalling that

$$r^2(d\theta^2 + \sin^2\theta\, d\phi^2) = dx^2 + dy^2 + dz^2 - dr^2. \tag{13}$$

Then we have

$$ds^2 = -(\omega^0)^2 + f^{-2}\, dr^2 + r^{-2}(r^2 + l^2) \times (dx^2 + dy^2 + dz^2 - dr^2). \tag{14}$$

Since xyz are differentiable functions (by definition), so is r and, in a certain range, $f(r)$; the tensor $dr^2 \equiv dr \otimes dr$ then is differentiable, and the only possible singularities in ds^2 at large r are the singularities of ω^0. From $\tan\phi = yx^{-1}$ and $\cos\theta = zr^{-1}$ we compute

$$f^{-1}\omega^0 = dt + \frac{2l}{r(r+z)}(x\, dy - y\, dx), \tag{15}$$

so ω^0, and hence ds^2, is singular with respect to the coordinates $txyz$ on the surface $z = -r$. To show that this is a spurious coordinate singularity as the curvature invariants suggest, we must find a way to present the metric in a singularity-free way. This is done by group theoretic methods in the next section, and in Sec. IV by methods which could be available even if a perturbation destroyed the symmetries of the metric now under consideration.

III. THE GROUP OF MOTIONS

The four Killing vectors for NUT space given by Newman *et al.* are[10,12]

$$\begin{aligned}
\xi_t &= \partial/\partial t, \\
\xi_x &= -\sin\phi \frac{\partial}{\partial\theta} - \cos\phi \times \left[\cot\theta \frac{\partial}{\partial\phi} + \lambda \tan\tfrac{1}{2}\theta \frac{\partial}{\partial t}\right], \\
\xi_y &= \cos\phi \frac{\partial}{\partial\theta} - \sin\phi \times \left[\cot\theta \frac{\partial}{\partial\phi} + \lambda \tan\tfrac{1}{2}\theta \frac{\partial}{\partial t}\right], \\
\xi_z &= (\partial/\partial\phi) - \lambda\, \partial/\partial t,
\end{aligned} \tag{16}$$

where $\lambda = 2l$. Here we have represented contravariant vectors v^μ by the associated differential operator $\mathbf{v} = v^\mu\, \partial/\partial x^\mu$, so that the Lie bracket of two vectors,

[11] For a clear geometrical picture of the collapsing star metric of J. R. Oppenheimer and H. Snyder [Phys. Rev. **56**, 455 (1939)], see D. L. Beckedorff (senior thesis, Mathematics Department, Princeton University, Princeton, New Jersey, 1962), and D. L. Beckedorff and C. W. Misner (in preparation).

[12] Some expository illustrations of the idea of a contravariant vector as a differential operator, and of the Lie bracket, can be found in F. K. Manasse and C. W. Misner, J. Math. Phys. **4**, 751 (1963).

$$[\mathbf{u}, \mathbf{v}]^\mu = u^\alpha \, \partial v^\mu/\partial x^\alpha - v^\alpha \, \partial u^\mu/\partial x^\alpha, \tag{17}$$

is just the commutator of the differential operators. The following commutation relations, given by Newman et al., can thus be directly verified:

$$[\xi_a, \xi_b] = -\epsilon_{abc}\xi_c, \tag{18}$$

$$[\xi_a, \xi_t] = 0. \tag{19}$$

In these equations a, b, c each take the values xyz, and ϵ_{abc} is the completely antisymmetric symbol with $\epsilon_{xyz} = +1$. As these commutation relations do not depend on l, the groups of motions of NUT space ($l \neq 0$) and of Schwarzschild space ($l = 0$) have isomorphic covering groups. We shall shortly see, however, that the action of these groups on the space is very different. First note, however, that no terms in $\partial/\partial r$ occur in Eqs. (16) so that every trajectory of the Killing vectors lies in a hypersurface of constant r. It will therefore be sufficient to restrict our attention to one such hypersurface on which the induced metric is

$$d\sigma^2 = -(\omega^0)^2 + (\omega^2)^2 + (\omega^3)^2, \tag{20}$$

with

$$\begin{aligned} \omega^0 &= dt + 2\lambda \sin^2 \tfrac{1}{2}\theta \, d\phi, \\ \omega^2 &= d\theta, \\ \omega^3 &= \sin\theta \, d\phi. \end{aligned} \tag{21}$$

This reduced form was obtained from Eq. (2) by setting $r = \text{const}$, choosing $(r^2 + l^2)^{\frac{1}{2}}$ as a unit of proper length σ, $(r^2 + l^2)^{\frac{1}{2}} f^{-1}$ as a unit of coordinate time t, and abbreviating the remaining constant as

$$2lf(r^2 + l^2)^{-\frac{1}{2}} = \lambda. \tag{22}$$

The entire structure we wish to study, namely the singularity at $\theta = \pi$ and the Killing vector fields (16), appears already in this reduced problem.

The probable cause of the singularity has been identified as a bad choice of the coordinate t, cf. Eq. (11). Let us then ask how the good t coordinate used in the Schwarzschild solution can be constructed. One way of generating the standard $t = \text{const}$ surface (within an $r = \text{const}$ hypersurface) of the Schwarzschild metric is to pick one point P and then move it in all possible ways under the action of the rotation group generated by ξ_x, ξ_y and ξ_z. This yields a two-dimensional, nonsingular, spherical surface as an orbit of a point P under the rotation group. What then is an orbit of a point in NUT space under the rotations generated by ξ_x, ξ_y, and ξ_z there? The crucial fact is that this orbit is *three*-dimensional as can be seen by computing the volume element spanned by the vectors ξ_a, namely

$$\xi_x \wedge \xi_y \wedge \xi_z = -\lambda \frac{\partial}{\partial t} \wedge \frac{\partial}{\partial \theta} \wedge \frac{1}{\sin\theta} \frac{\partial}{\partial \phi}. \tag{23}$$

This volume element evidently vanishes only if $\lambda = 0$. For $\lambda \neq 0$, the entire $r = \text{const}$ hypersurface described by the metric (20) is generated by the action of the rotation group alone, without employing, in addition, the time translations generated by ξ_t. This possibility can also be suggested by the relation

$$\xi_z \cos\theta + (\xi_x \cos\phi + \xi_y \sin\phi) \sin\theta = -\lambda \xi_t \tag{24}$$

which follows from Eqs. (16). Although we fail by this study of orbits to construct a good time coordinate, a new method of approach is opened up which yields a complete solution.[13] Let us choose one point P_0 as a reference; then any other point P (in the same $r = \text{const}$ hypersurface) can be identified by the rotation A which carries P_0 into P. Where there two such rotations, A and B, then $A^{-1}B$ would leave P_0 invariant. As no infinitesimal rotation leaves P_0 invariant, according to Eq. (23), there is at most a discrete set of group elements A, B, etc. which correspond to P, and we have the group as a covering space over the orbit. But we began with merely a local line element (2) or (20), rather than a manifold with a definite topological structure, so we shall be content to study the simply connected universal covering space from which all other possibilities arise by appropriate identifications. The group characterized by Eqs. (18) is the covering group of the rotation group $SO(3)$, it is isomorphic to $SU(2)$, but I will call it S^3 to emphasize its topology and will represent it as the quaternions of unit norm, $|q| = 1$. We can then take this group space as the three-dimensional manifold whose geometry the metric of Eq. (20) was attempting to describe. I will construct a metric equivalent to Eq. (20) on this manifold by using the Killing vectors as point of contact between the two spaces, but without attempting to find the previous $t\theta\phi$ coordinates on S^3.

It is easiest to just guess what the metric $d\sigma^2$ on S^3 must be, and then verify the guess. Three vector fields ξ_a on S^3 with the commutation relation (18) are those tangent to the curve systems indicated:

$$\begin{aligned} \xi_x\colon &\quad q(\alpha) = e^{i\alpha/2} q_0, \\ \xi_y\colon &\quad q(\alpha) = e^{j\alpha/2} q_0, \\ \xi_z\colon &\quad q(\alpha) = e^{k\alpha/2} q_0. \end{aligned} \tag{25}$$

[13] N. Steenrod, *The Topology of Fibre Bundles* (Princeton University Press, Princeton, New Jersey, 1951), Theorem 7.3.

Here i, j, k are the standard unit imaginary quaternions. The curves in Eqs. (25) each satisfy $|q| = |q_0| = \text{const}$, so their tangent vectors ξ_i are tangent to the sphere S^3 of unit quaternions at all points of S^3. A verification of the commutation relations (18) is given in Appendix B. A fourth vector, which commutes with all these, is

$$\mathbf{n}_z\colon \quad q(\alpha) = q_0 e^{-k\alpha/2}. \tag{26}$$

We have not prejudged the exact equality of $\mathbf{n}_z$ and ξ_t since the commutation relations (19) do not restrict the normalization of ξ_t. This normalization can be obtained from Eq. (24), where it is apparent that ξ_z and ξ_t are parallel at one point, namely $\theta = 0$, and there $\xi_z = -\lambda\xi_t$. Similarly, from Eqs. (25) and (26), one sees that the curves tangent to ξ_z and $\mathbf{n}_z$ are parallel at the point $q_0 = 1$, and there $\xi_z = -\mathbf{n}_z$. Thus, as our guess for ξ_t on S^3, we take

$$\xi_t = \lambda^{-1}\mathbf{n}_z. \tag{27}$$

Now that we have chosen a manifold, S^3, and on that manifold, four vector fields ξ_a and ξ_t with the appropriate commutation rules (18) and (19), the next step is to find a metric invariant under the motions generated by these vectors, i.e., a metric on S^3 for which they will be Killing vectors. It has been essential in the previous discussion that ξ_a and ξ_t were taken as contravariant vectors, otherwise we could not have discussed the motions they generate without making use of a metric. It will be convenient now to regard the metric as a contravariant tensor, so it can more readily be expressed in terms of ξ_a and ξ_t. There is no standard notation for the contravariant metric (as distinct from its components g^{ab}); I will write it as

$$\left(\frac{\partial}{\partial \mathbf{P}}\right)^2 \equiv g^{ab}\frac{\partial}{\partial x^a}\otimes\frac{\partial}{\partial x^b}. \tag{28}$$

The standard group invariant metric on S^3 is then

$$\xi_x\otimes\xi_x + \xi_y\otimes\xi_y + \xi_z\otimes\xi_z. \tag{29}$$

Since this metric is positive definite, it is not the one we are looking for, even though it makes ξ_a and ξ_t Killing vectors. The vector $\mathbf{n}_z$ is a unit vector for the metric (29) and the symmetric tensor $\mathbf{n}_z\otimes\mathbf{n}_z$ is invariant under the ξ_a and $\mathbf{n}_z$ motions, so the metric

$$(\partial/\partial\mathbf{P})^2 = \xi_x\otimes\xi_x + \xi_y\otimes\xi_y + \xi_z\otimes\xi_z - (1+\Lambda^{-2})\mathbf{n}_z\otimes\mathbf{n}_z \tag{30}$$

has the desired Killing vectors, and also the proper $++-$ signature when $\Lambda^2 > 0$. In the next paragraph, we stop guessing and verify that Eqs. (30) and (20) represent the same metric, and in fact with $\Lambda = \lambda$; but let me conclude the heuristics with a calculation showing that ξ_a, and ξ_t or $\mathbf{n}_z$, are indeed Killing vectors for the metric (30), that is, that the metric (30) is invariant under the motions generated by these vector fields. If $\mathbf{v}$ is any vector field, then under a motion ξ, the change in $\mathbf{v}$ is measured by its Lie derivative along ξ, $\pounds_\xi\mathbf{v}$, defined by

$$\pounds_\xi\mathbf{v} \equiv [\xi, \mathbf{v}]. \tag{31}$$

The change in a tensor $\mathbf{v}\otimes\mathbf{w}$ under the ξ motion is measured by $\pounds_\xi(\mathbf{v}\otimes\mathbf{w})$ which is defined by requiring the standard product law of differentiation. It then is obvious that the metric (30) is invariant under $\mathbf{n}_z$ motions, so let us sketch the ξ_x-invariance computation. The $\xi_x\otimes\xi_x$ and $\mathbf{n}_z\otimes\mathbf{n}_z$ terms in (30) clearly have vanishing Lie derivatives $\pounds_x$ along ξ_x; for the terms $\xi_y\otimes\xi_y$, we compute

$$\pounds_x(\xi_y\otimes\xi_y) = (\pounds_x\xi_y)\otimes\xi_y + \xi_y\otimes(\pounds_x\xi_y).$$

Then, since $\pounds_x\xi_y \equiv [\xi_x, \xi_y] = -\xi_z$, this is

$$\pounds_x(\xi_y\otimes\xi_y) = -\xi_z\otimes\xi_y - \xi_y\otimes\xi_z.$$

These terms are just cancelled by $\pounds_x(\xi_z\otimes\xi_z)$, to verify that $\pounds_x(\partial/\partial\mathbf{P})^2 = 0$.

We have made a guess, in Eq. (30), of the form of the metric NUT space induces on an $r = \text{const}$ hypersurface. We have not been able to guess reasonably at the value of the constant Λ^2 which appears in Eq. (30), or at a possible constant multiplying the entire metric. To determine these constants and verify that our guesses were correct, we must compare Eq. (30) and Eq. (20). The first step will be to rewrite Eq. (20) in contravariant form. The set of basis vectors dual[13a] to the ω^a of Eq. (21) is

$$\begin{aligned}\mathbf{e}_0 &= \partial/\partial t,\\ \mathbf{e}_2 &= \partial/\partial\theta,\\ \mathbf{e}_3 &= \frac{1}{\sin\theta}\frac{\partial}{\partial\phi} - \lambda\tan\tfrac{1}{2}\theta\frac{\partial}{\partial t}.\end{aligned} \tag{32}$$

Thus the contravariant form of Eq. (20) is

[13a] Duality of a basis ω^μ for covariant vectors and a basis $\mathbf{e}_\mu$ for contravariant vectors means that they satisfy the "inner product" relations $\langle\omega^\mu, \mathbf{e}_\nu\rangle = \delta_\nu{}^\mu$. These relations are verified between Eqs. (21) and (32) by evaluating them in terms of the relations $\langle dx^\mu, \partial/\partial_{x^\nu}\rangle = \delta_\nu{}^\mu$ which follow from the definition $\langle df, \mathbf{v}\rangle = \partial f/\partial\mathbf{v}$ of the gradient operator d.

$$(\partial/\partial\mathbf{P})^2 = -\mathbf{e}_0 \otimes \mathbf{e}_0 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3$$
$$= (\lambda^2 \tan^2 \tfrac{1}{2}\theta - 1)\frac{\partial}{\partial t} \otimes \frac{\partial}{\partial t} + \frac{\partial}{\partial \theta} \otimes \frac{\partial}{\partial \theta}$$
$$+ \frac{1}{\sin^2 \theta}\frac{\partial}{\partial \phi} \otimes \frac{\partial}{\partial \phi} - \frac{\lambda}{2\cos^2(\theta/2)}$$
$$\times \left(\frac{\partial}{\partial t} \otimes \frac{\partial}{\partial \phi} + \frac{\partial}{\partial \phi} \otimes \frac{\partial}{\partial t}\right). \tag{33}$$

Now, by substituting Eqs. (16) into Eq. (30), or equivalently into

$$(\partial/\partial\mathbf{P})^2 = \xi_x \otimes \xi_x + \xi_y \otimes \xi_y + \xi_z \otimes \xi_z$$
$$- (1 + \lambda^2)\xi_t \otimes \xi_t, \tag{34}$$

we verify that the metrics (34) and (33) are identical, i.e., that (30) with $\Lambda = \lambda$ and (20) agree.

Equation (34) expresses the metric of a three-dimensional space in terms of its four Killing vectors. For some purposes, an expression in terms of three orthonormal vectors is more convenient. This can be obtained as follows: We introduce two vector fields $\mathbf{n}_x$ and $\mathbf{n}_y$ tangent, respectively, to the curves $q_0 e^{-i\alpha/2}$ and $q_0 e^{-i\alpha/2}$. With $\mathbf{n}_z$ and the ξ_a, these satisfy the commutation rules

$$[\mathbf{n}_a, \mathbf{n}_b] = -\epsilon_{abc}\mathbf{n}_c, \tag{35a}$$
$$[\mathbf{n}_a, \xi_b] = 0, \tag{35b}$$
$$[\xi_a, \xi_b] = -\epsilon_{abc}\xi_c, \tag{35c}$$

characteristic of the generators of $SO(4)$. The relation

$$\xi_x \otimes \xi_x + \xi_y \otimes \xi_y + \xi_z \otimes \xi_z$$
$$= \mathbf{n}_x \otimes \mathbf{n}_x + \mathbf{n}_y \otimes \mathbf{n}_y + \mathbf{n}_z \otimes \mathbf{n}_z \tag{36}$$

can be established by noting that the symmetric tensors on either side of this equation are invariant under the ξ_a motions (also the $\mathbf{n}_a$ motions) and therefore determined by their values at any one point; but we see from the curves which define these vectors that $\xi_a = -\mathbf{n}_a$ at the point $q = 1$. Introducing the $\mathbf{n}_a$ in Eq. (30) or (34) by use of Eq. (36), then, we find

$$(\partial/\partial\mathbf{P})^2 = \mathbf{n}_x \otimes \mathbf{n}_x + \mathbf{n}_y \otimes \mathbf{n}_y - (1/\lambda^2)\mathbf{n}_z \otimes \mathbf{n}_z, \tag{37}$$

so $\mathbf{n}_x$, $\mathbf{n}_y$, and $\xi_t = \lambda^{-1}\mathbf{n}_z$ form an orthonormal basis. This orthonormal basis, however, is composed of globally defined differentiable vector fields, in contrast to the basis $\mathbf{e}_a$ of Eqs. (32) which shared the singularities of the $t\theta\phi$ coordinates, and, in particular, a familiar singularity of spherical coordinates at the pole $\theta = 0$. Of course neither $\mathbf{n}_x$ nor $\mathbf{n}_y$ is a Killing vector for the metric (37); however this orthonormal frame, $\mathbf{n}_x$, $\mathbf{n}_y$, $\lambda^{-1}\mathbf{n}_z$, has a number of interesting properties. In the first place, each vector $\mathbf{n}_x$, $\mathbf{n}_y$, and $\lambda^{-1}\mathbf{n}_z$ is, according to Eqs. (35b), invariant under the "rotation" symmetries of the space generated by the Killing vectors ξ_x, ξ_y, ξ_z. Secondly, although $\xi_t = \lambda^{-1}\mathbf{n}_z$ is of course invariant under the "time translation" symmetries it generates, the equations $[\xi_t, \mathbf{n}_x] = -\lambda^{-1}\mathbf{n}_y$ and $[\xi_t, \mathbf{n}_y] = +\lambda^{-1}\mathbf{n}_x$ show that if the vectors fields $\mathbf{n}_x$ and $\mathbf{n}_y$ are time-translated, the result is a pair of vectors fields which is a rotation of the original $\mathbf{n}_x$, $\mathbf{n}_y$ fields in the $\mathbf{n}_x$, $\mathbf{n}_y$ plane. The third of Eqs. (35a), $[\mathbf{n}_x, \mathbf{n}_y] = -\mathbf{n}_z$, shows there is no two-dimensional surface whose tangent plane is parallel to $\mathbf{n}_x \wedge \mathbf{n}_y$, i.e. that the Killing vector $\xi_t = \lambda^{-1}\mathbf{n}_z$ is *not* hypersurface orthogonal. (The condition that $\mathbf{n}_x \wedge \mathbf{n}_y$ be surface forming is[14] that at each point, $[\mathbf{n}_x, \mathbf{n}_y]$ be a linear combination of $\mathbf{n}_x$ and $\mathbf{n}_y$.)

IV. NONSINGULAR COORDINATES

In this section I want, again, to show that those singularities in the NUT space metric of Eq. (2) which are not associated with $f^2 = 0$ can be eliminated by assigning to the $r = \text{const}$ surfaces the topology S^3. However I shall do this now by methods which might be more widely applicable than the group theoretic approach of Sec. III, since I shall now make no use of the existence of Killing vectors. As all the singularities of the metric were to be found in ω^0, according to the discussion in Sec. II, and were of the same character on every $r = \text{const}$ hypersurface, it is sufficient to assume $r = \text{const}$ and study only

$$\omega^0 = dt_N + 2\lambda \sin^2 \tfrac{1}{2}\theta \, d\phi. \tag{38}$$

For later convenience, I have now set $t \equiv t_N$, since we found this time coordinate to be regular at the north pole, $\theta = 0$, although singular at $\theta = \pi$ by Eqs. (11) or (15). A haphazard guessing of singular coordinate transformations, such as $t = t' - 2\lambda\phi \sin^2 \frac{1}{2}\theta$, in an attempt to cancel out by this means the singularity in Eq. (38), soon leads to

$$t_N = t_S - 2\lambda\phi, \tag{39}$$

so that

$$\omega^0 = dt_S - 2\lambda \cos^2 \tfrac{1}{2}\theta \, d\phi. \tag{40}$$

From this formula, or from

$$dt_S = \omega^0 - \lambda \cot \tfrac{1}{2}\theta \, \omega^3 \tag{41}$$

[14] This is known as Frobenius' Theorem. In Willmore [cf. reference 5, Eqs. (VII-14.2) and (VII-14.8)], it is reduced to its perhaps more familiar statement about completely integrable Pfaffian systems. A complete proof can be found, for instance, in Sternberg [cf. reference 5, Sec. III-4].

as compared to Eq. (10), we see that t_S is singular at $\theta = 0$ but regular at the south pole $\theta = \pi$. At first glance the introduction of t_S seems to be no gain, as we have traded away a singularity in t_N at the south pole for one equally bad in t_S at the north pole. But then let us use t_N only in a coordinate patch $0 \leq \theta < \pi$ where it is regular, and t_S only in a patch $0 < \theta \leq \pi$ where t_S is regular; the union of these two patches could then be a manifold on which the metric is everywhere nonsingular. Certainly the metric is differentiable with respect to the coordinates appropriate to each patch, so the only thing that remains to be verified is that the coordinates satisfy the appropriate regularity conditions[15] with respect to each other in the overlap region $0 < \theta < \pi$. The Jacobian condition is readily verified. However ϕ itself, in Eq. (39), is not a differentiable function on the sphere S^2. Even away from its worst singularities at the poles, one can only treat ϕ as a regular function if its multiple valuedness, $\phi \equiv \phi + 2\pi$, is respected. Thus the time coordinates in Eq. (39) can *only* be thought of as angles with

$$t_N \equiv t_N + 4\pi\lambda, \tag{42}$$

and similarly for t_S. With this understanding, then, the overlap of the coordinate patches is proper, and the metric is free from singularities.

Although the preceding discussion constructs a manifold on which the metric of Eq. (20) is free from singularities, we do not yet have a good idea what it looks like. One important characteristic has been shown, namely that the timelike curves of constant θ, ϕ are closed. Related to this is the fact that the manifold is compact (closed). For a further discussion, I prefer to view this manifold not as two overlapping coordinate patches, but as two abutting pieces. Each piece is a solid torus; for instance the north piece is described by t_N, ϕ and θ with $0 \leq \theta \leq \frac{1}{2}\pi$. For each fixed t_N, then, $\theta \leq \frac{1}{2}\pi$ and ϕ describe a disc or cell, E^2, while t_N, due to its periodicity, describes a circle S^1. The north piece is then the solid torus $S^1 \times E^2_N$, and its boundary, $\theta = \frac{1}{2}\pi$, is the torus $S^1 \times S^1_N$ with coordinates t_N, ϕ. Similarly, the south piece with coordinates t_S, ϕ, and θ with $\frac{1}{2}\pi \leq \theta \leq \pi$ is a solid torus $S^1 \times E^2_S$ whose boundary is the torus $S^1 \times S^1_S$ with coordinates t_S, ϕ. To join these two pieces together, we identify their boundaries using Eq. (39). This identification is illustrated by Fig. 1. A presentation of the manifold in this way, which is effectively a cell decomposition of the space, allows topological computations to be carried out by elementary means. One can, for instance, compute the homology groups and find that they are the same as those of a sphere, S^3. This does not prove our manifold is S^3, although it points in that direction. The most prominent feature of the mathematical literature about S^3 is the Poincaré conjecture that every simply connected, closed, 3-manifold is S^3. A computation of the fundamental group of our manifold, given at the end of this section, shows that the manifold is indeed simply connected. Although the Poincaré conjecture is still unproven, we find in the literature concerning it a theorem[16] that a simply connected 3-manifold is topologically S^3 if it can be formed by identifying the boundaries of two (otherwise disjoint) solid tori. Thus the manifold we have constructed is in fact S^3.

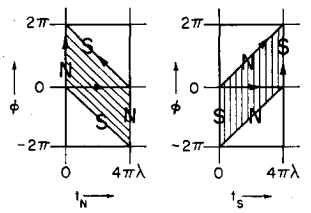

FIG. 1. The torus T^2 is represented as either the t_N, ϕ plane or the t_S, ϕ plane in which points are identified for which ϕ differs by a multiple of 2π, or t_N, t_S differ by a multiple of $4\pi\lambda$. A convenient fundamental rectangle to choose is that, shown shaded, bounded by closed curves N where $t_N = 4\pi\lambda n$, and closed curves S where $t_S = 4\pi\lambda n (n = 0, 1, 2, \cdots)$.

In some cases it is not difficult to visualize the union of two solid tori as forming S^3. Imagine S^3 as consisting of ordinary Euclidean space E^3 with a point at infinity added. Then the standard imbedding of the torus T^2 in E^3 gives in this way an imbedding of a solid torus—the interior of the surface T^2—in E^3 and hence in S^3. What does the outside of T^2 look like in S^3? This is most easily discovered by turning things inside out, i.e., imagine that the unit sphere, $|\mathbf{x}| = 1$, lies in the interior of T^2 and perform an inversion $\mathbf{x} \rightarrow \mathbf{x}/|x|^2$. The image of T^2 under this inversion is again the standard imbedding of T^2 in E^3, except now situated on the interior of the unit sphere. The interior of this new T^2 is evidently again a solid torus, and is homeomorphic (under the inversion) to the exterior of the old T^2. Thus we see S^3 as the union of two solid tori. To see the boundaries identified, not crosswise as just described, but with a twist as in

[15] The axioms for defining the idea of a differentiable manifold can be found in Willmore [reference 5, Sec. VI-1]. Several elementary examples of the application of these axioms can be found in C. W. Misner and J. A. Wheeler, Ann. Phys. (N. Y.) **2**, 553 (1957).

[16] R. H. Bing, Ann. Math. **68**, 17 (1958), Theorem 3.

Fig. 1, we replace one of the solid tori by its image under the following homeomorphism: Let the solid torus $S^1 \times E^2$ be cut open into $E^1 \times E^2$, twist it as many times as desired, and then rejoin the ends. Since no previously nearby points of the solid torus are now separated (the ends were rejoined point for point as they were separated), and the operation is invertible, it describes a homeomorphism of $S^1 \times E^2$ onto itself.

To conclude this section, I verify a result we have already used, namely the fact that the manifold we constructed with the aid of Eq. (39) is simply connected. I must show that any closed curve in the manifold can be deformed continuously into a point. Given any closed curve in the manifold, let us first deform it onto the interface T^2 between the two pieces. This is possible, since by at most an infinitesimal deformation, we can arrange that no point on the curve lie along $\theta = 0$ or $\theta = \pi$, and then a point on the curve at t_N, θ, ϕ in the north piece is moved to the interface $\theta = \frac{1}{2}\pi$ by increasing θ keeping t_N, ϕ fixed. Similarly, for points in the south piece, decrease θ to $\frac{1}{2}\pi$ keeping t_S, ϕ fixed. Now the closed curve lies on T^2, which we can picture as the shaded rectangle in Fig. 1, with the boundaries identified. The curve can now be pushed away from some point (not on the curve) in the center of the rectangle, thus deforming it continuously onto the boundary of the rectangle. A closed curve on the boundary of the shaded rectangle in Fig. 1 must consist of the curves marked N and S followed any number of times in any order, e.g., $NNSNS^{-1}$. But the curve N, which is the boundary of the cell $t_N = 0$, $0 \leq \theta \leq \frac{1}{2}\pi$ in the north solid torus, can clearly be contracted across this cell to a point, and similarly the curve S, which bounds the cell $t_S = 0$, $\frac{1}{2}\pi \leq \theta \leq \pi$, is contractible to a point. Consequently, any closed curve can be deformed continuously to a point; the manifold is simply connected.

V. SPACELIKE SURFACES[16a]

In this section, I want to discuss spacelike surfaces in NUT space, but without getting into questions of the Schwarzschild-like singularity $f^2 = 0$ at small r. Consequently we shall look only at the intersection of a three-dimensional spacelike hypersurface in NUT space with the timelike hypersurface $r = \text{const}$. In the Schwarzschild solution, the intersection of a $t = \text{const}$ spacelike surface with the $r = \text{const}$ hypersurface $R \times S^2$ with coordinates t, θ, ϕ is the spacelike closed manifold S^2 on which both t and r are constant. Suppose σ^3 is a spacelike hypersurface in NUT space. Then if p is any point of σ^3, the orbit of p under the group of motions is, as we have seen, a timelike hypersurface $r = \text{const}$ whose topology is S^3. Since σ^3 is spacelike, while S^3 is timelike, they are not tangent, hence intersect in a two-dimensional spacelike surface σ. I will show that σ does *not* have the topology S^2 (surface of a sphere), and, in fact, that σ is not *any* closed manifold. It is evidently possible for σ to be an open manifold, for instance by taking t and r constant in Eq. (2), and restricting θ by $\theta < \theta_0$ as in Eq. (12). Whether σ can be a complete Riemannian manifold is an open question, but if it were, the imbedding of σ in S^3 would have to be pathological, analogous, for example, to an infinite-length line immersed in S^2 as a spiral which asymptotically approaches two latitude circles.

To prove that σ is not closed, I assume that σ is a connected closed manifold and obtain a contradiction; cf. Fig. 2. A standard application[17] of Alexander duality theory shows that a closed connected $(n-1)$ manifold σ imbedded in S^n divides S^n into two disconnected pieces; for us, $n = 3$. Thus we may identify part of S^3 as the "inside of σ" by choosing arbitrarily one of these two pieces. Then there exists a unit timelike outward normal vector $\mathbf{n}$, defined continuously over σ. But the Killing vector field ξ_t is, by Eqs. (27) and (37), a continuous unit timelike vector field on S^3, so the inner product $\mathbf{n}\cdot\xi_t$ is continuous on σ and cannot vanish there. (No two timelike vectors are orthogonal.) For any point p_0 of σ, we now follow the trajectory $p(t)$ tangent to ξ_t with $p(0) = p_0$. Assume, for definiteness, that ξ_t points outward at p_0, so $\mathbf{n}\cdot\xi_t < 0$ at p_0. For sufficiently small positive t, $p(t)$ is in the piece of S^3 outside σ, while for small negative t, $p(t)$ is inside σ. But $p(t)$, we know, is a closed curve so $p(T) = p_0$ for $T = 4\pi\lambda > 0$, and therefore $p(T-\epsilon) = p(-\epsilon)$ is inside σ for small positive ϵ. Hence, for some t_1 with $0 < t_1 < T$, the trajectory must pass from the outside, as at $p(0^+)$, to the inside, as at $p(T^-)$. Thus $p(t_1) \equiv p_1$ lies on σ and ξ_t points inward there so that $\mathbf{n}\cdot\xi_t > 0$ at p_1. But $\mathbf{n}\cdot\xi_t$ is continuous on σ, positive at p_1, negative at p_0, and therefore, since σ is connected, $\mathbf{n}\cdot\xi_t$ vanishes somewhere on σ, which is impossible.

[16a] A theorem similar to that proven in this section has been given by A. Avez, Compt. Rend. **254**, 3984 (1962).

[17] See, for example, S. Lefshetz, *Topology* (Chelsea Publishing Company, New York, 1956), 2nd Ed., p. 144. I wish to thank D. Lowdenslager for directing me to this proof.

VI. DISCUSSION

Most of the results have been referred to a single r = const surface, so I will recapitulate by reinstating the radial dependence. Using the methods of Sec. IV, we see that a new time coordinate which is nonsingular at $\theta = \pi$ can be introduced into the NUT space metric of Eq. (2) only if we assume both the new and the old time coordinates are periodic with period

$$\Delta t = 8\pi l. \tag{43}$$

The closed, timelike, curves of constant r, θ, ϕ in the coordinates of Eq. (2), which are trajectories of the Killing vector field ξ_t, then have a length, measured in proper time, of

$$\Delta\tau = 8\pi l f(r). \tag{44}$$

Although it was possible to see by coordinate-patching methods in Sec. IV that the hypersurfaces r = const should be assigned the topology S^3 of the three-dimensional spherical surface of a ball in four space, a detailed study of the properties of NUT space is facilitated by the group theoretic methods of Sec. III. In particular, the three Killing vectors ξ_x, ξ_y, and ξ_z generate a group of motions which is simply transitive on each r = const surface. When this surface is represented as the surface $|q|$ = const in the space of quaternions, $q = w + ix + jy + kz$, the metric induced on it can be written in terms of a convenient, globally defined, orthonormal tetrad, as in Eq. (37). The covariant form of this equation is

$$d\sigma^2 = \sigma_x^2 + \sigma_y^2 - \lambda^2\sigma_z^2, \tag{45}$$

where the σ_a are the differential forms dual to the tangent vector fields $\mathbf{n}_a$ of Sec. III and Appendix B, namely

$$\begin{aligned}
\sigma_x &= 2\,|q|^{-2}\,(x\,dw - w\,dx - z\,dy + y\,dz),\\
\sigma_y &= 2\,|q|^{-2}\,(y\,dw + z\,dx - w\,dy - x\,dz),\\
\sigma_z &= 2\,|q|^{-2}\,(z\,dw - y\,dx + x\,dy - w\,dz).
\end{aligned} \tag{46}$$

In this way the NUT space metric can be written

$$ds^2 = f^{-2}\,dr^2 + (r^2 + l^2)(\sigma_x^2 + \sigma_y^2) - (2lf)^2\sigma_z^2. \tag{47}$$

Here r could be taken as an independent coordinate, with the other three chosen from among $wxyz$ subject to $|q| = 1$. However, it is probably simpler to assume that $r = \ln |q|^2$, so

$$|q|^2 = w^2 + x^2 + y^2 + z^2 = e^r. \tag{48}$$

The set of vectors displayed in Eq. (46) can then be completed to an orthogonal tetrad by including the vector

$$dr = 2\,|q|^{-2}\,(w\,dw + x\,dx + y\,dy + z\,dz). \tag{49}$$

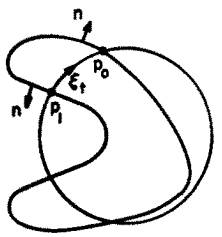

FIG. 2. A closed, two-dimensional, spacelike surface σ in S^3 divides S^3 into two pieces and admits, therefore, a continuous nonzero outward normal vector $\mathbf{n}$. But a trajectory of the Killing vector field ξ_t is a closed curve, hence after leaving the inside of σ at p_0 (where $\mathbf{n}\cdot\xi_t < 0$) must enter again at p_1 (where $\mathbf{n}\cdot\xi_t > 0$). As $\mathbf{n}\cdot\xi_t$ is continuous on σ, it vanishes somewhere between p_0 and p_1 on σ. At this point, σ cannot be spacelike since ξ_t is timelike.

The single coordinate patch $wxyz$ now covers the entire space we are discussing, namely the region $|q|^2 > e^R$ where R defines the Schwarzschild-like singular surface where $f^2 = 0$, i.e.,

$$R = m + (m^2 + l^2)^{\frac{1}{2}}.$$

Computations involving the metric in the form (47) can be much simplified by using the relation

$$d\sigma_z = \sigma_x \wedge \sigma_y \tag{50}$$

and its cyclic permutations, which follow by a standard rule from Eqs. (35a). For instance, the connection components $\Gamma^a_{\mu\nu}$ in the obvious orthonormal frame derived from dr and the σ_a can be quickly computed by the Cartan methods of Appendix A. The group of motions generated by the three Killing vectors ξ_x, ξ_y, ξ_z is just the group corresponding to unit quaternions p with $|p| = 1$. That the motion

$$q \to pq \tag{51}$$

is an isometry when $|p|^2 = p^*p = 1$ follows from the fact that it leaves invariant the coordinate $r = \ln |q|^2$ and the vector fields σ_a and dr which appear in Eq. (47). This can be seen from the relation

$$q^*\,dq \to (pq)^*\,d(pq) = q^*p^*p\,dq = q^*\,dq,$$

and the formulas

$$\begin{aligned}
\sigma_a &= 2\,|q|^{-2}\,\mathrm{Re}\,(q^*\,dq\,i_a),\\
dr &= 2\,|q|^{-2}\,\mathrm{Re}\,(q^*\,dq),
\end{aligned} \tag{52}$$

where $i_x = i$, $i_y = j$, $i_z = k$. To verify that

$$q \to qe^{-kt/2} \tag{53}$$

is an isometry (generated by $\mathbf{n}_z$) requires a little more computation; one notes that it leaves invariant r, dr, σ_z, and $|q^*dq|^2$, consequently also

$$\sigma_x^2 + \sigma_y^2 = 4\,|q|^{-4}\,|q^*\,dq|^2 - \sigma_z^2 - dr^2. \tag{54}$$

Because the Riemann tensor vanishes as r^{-3} for

$r \to \infty$ (cf. Newman et al. or Appendix A), we can say that NUT space is asymptotically flat for $|q| \to \infty$. However this does *not* mean that a set of asymptotically rectangular coordinates exists. In a coordinate system x^μ where $g_{\mu\nu} \sim \eta_{\mu\nu}$ the hypersurface $x^0 = \text{const}$, $(x^1)^2 + (x^2)^2 + (x^3)^2 > L^2$, is certainly spacelike for sufficiently large L^2. But in Sec. V, I showed that the intersection of this spacelike hypersurface with the closed timelike hypersurface $r = \text{const}$ cannot be a closed surface, even for arbitrarily large r. It seems probable, therefore, that a coordinate system where $g_{\mu\nu} - \eta_{\mu\nu} \to 0$ as $L^2 \to \infty$ is impossible. This excludes then any interpretation of NUT space as the gravitational field produced by a localized source introduced into a previously flat part of the universe. The reason no analogue of NUT space was found in studies of the linearized gravitational field equations is simply that, even if it had been found, it would be rejected as failing to satisfy reasonable boundary conditions at infinity. Unless it is interpretable as a cosmological model, it must be similarly rejected in general relativity theory. It remains, of course, interesting as a counterexample: a space which does not admit an interpretation without a periodic time coordinate, a space without reasonable spacelike surfaces, and an asymptotically zero curvature space which apparently does not admit asymptotically rectangular coordinates.

Dieter Brill suggested to me that perhaps NUT space would behave more reasonably on the other side of its "Schwarzschild singularity". This we readily find to be correct. In the region

$$m - (m^2 + l^2)^{\frac{1}{2}} < r \equiv \ln |q|^2 < k + (m^2 + l^2)^{\frac{1}{2}}, \qquad (55)$$

where f^2 is negative, we see from Eq. (47) that the 3-spheres $|q| = \text{const}$ are closed spacelike hypersurfaces, while r becomes a time coordinate. This is, therefore, a cosmological model[18] with homogeneous (but not isotropic) closed spacelike hypersurfaces. At the two endpoints, $r = m \pm (m^2 + l^2)^{\frac{1}{2}}$, each trajectory of the Killing vector $\mathbf{n}_z$ collapses to a point (or becomes a null curve), so the spacelike hypersurface S^3 collapses to an S^2 (or becomes a null hypersurface, still S^3). The proper time between these end points is $\tau = \int (-f^2)^{-\frac{1}{2}}\, dr$, or

$$\tau = (m^2 + l^2)^{\frac{1}{2}} \times \int_{-1}^{1} \frac{[z^2 + 2zm(m^2 + l^2)^{-\frac{1}{2}} + 1]^{\frac{1}{2}}}{[1 - z^2]^{\frac{1}{2}}}\, dz, \qquad (56)$$

and is finite.

The most intriguing aspect of NUT space then is its Schwarzschild singularity, for if the singularities where $f^2 = 0$ are in fact spurious,[19] it would be possible to join together the cosmological solution with $f^2 < 0$ and the two asymptotically zero curvature spaces where $f^2 > 0$ [i.e., $r > m + (m^2 + l^2)^{\frac{1}{2}}$ and $r < m - (m^2 + l^2)^{\frac{1}{2}}$] to form a single, singularity-

[18] This metric has been given, in another coordinate system covering only the region limited by Eq. (55), by A. Taub, Ann. Math. **53**, 472 (1951). This can be recognized by the similarity of Eq. (47) here and the form of Taub's metric given by O. Heckmann and E. Schüking in Eqs. (11-1.21) of their article in *Gravitation: an introduction to current research*, edited by L. Witten (John Wiley and Sons, Inc., New York, 1962). Taub (private communication) has found that the transformation $\theta = x^1$, $\phi = -x^2$, and $t + 4l\phi = 2lx^3$ relates the coordinates used in Eq. (2) and those used in Eq. (7.3) of Taub's article. The time coordinates, r here, and t in Taub's article, are related in an obvious way by equating proper time along curves of fixed spatial coordinates.

[19] It has been previously noted that the curvature remains finite at the boundaries [Eq. (55)] of the cosmological region by V. Joseph, Proc. Cambridge Phil. Soc. **53**, 836 (1957). I am also informed that this metric, at least in the form of Eq. (2), has been studied by J. Ehlers in his unpublished thesis "Konstructionen und Charakterisierungen der Einsteinschen Gravitationsfeldgleichungen" (Hamburg, 1957).

Note added in proof. The spurious nature of the singularities $f^2 = 0$ has been demonstrated by Taub and the author who note that the transformation

$$t \equiv t_N = -2l(\phi + \psi) + \int (1/f^2)\, dr$$

leaves the metric in the form

$$ds^2 = -f^2(2l)^2(d\psi + \cos\theta\, d\phi)^2 + 2\, dr(2l)(d\psi + \cos\theta\, d\phi) + (r^2 + l^2)(d\theta^2 + \sin^2\theta\, d\phi^2)$$

which is analytic for $-\infty < r < +\infty$, i.e. is everywhere analytic on the manifold $R \times S^3$ provided $\psi\theta\phi$ are interpreted as the Euler angle coordinates on S^3, [see Corben and Stehle, *Classical Mechanics*, (John Wiley & Sons, Inc., New York, 1960) 2nd ed., Eqs. IV-34, for the relation of $\psi\theta\phi$ to quaternion coordinates $wxyz$]. They furthermore show that this metric is nonsingular in a newly defined sense, namely that every open geodesic arc which approaches the boundaries of the manifold (here $r = \pm\infty$) is infinitely long as measured by any affine path parameter. This definition seems mathematically satisfactory since it includes both affinely complete and closed (compact) manifolds, and it implies that the manifold is inextensible. The physically singular aspects of this solution as well as the above results and various families of geodesics in this space are more fully described in a paper by Misner and Taub which will be submitted to JETP.

Another result from the paper mentioned above is an example of an analytic solution of the Einstein equations in a neighborhood of a closed spacelike hypersurface which can be extended in two *inequivalent* ways as an *analytic* Riemannian manifold satisfying $R_{\mu\nu} = 0$. Consider the metric

$$ds^2 = -f^2(2l)^2(d\Psi + \cos\theta\, d\phi)^2 - 2\, dr(2l)(d\Psi + \cos\theta\, d\phi) + (r^2 + l^2)(d\theta^2 + \sin^2\theta\, d\phi^2)$$

which is everywhere analytic ($-\infty < r < +\infty$) and satisfies $R_{\mu\nu} = 0$ on the manifold $R \times S^3$ provided $\Psi\theta\phi$ are interpreted as the Euler angle coordinates on S^3. In the region $f^2 < 0$, which is a neighborhood of the closed spacelike hypersurface $r = 0$ where $f^2 = -1$, the two metrics above are analytically isometric under the isometry which identifies the coordinates $r\theta\phi$ for the two metrics but sets

$$\psi = \Psi + \int_0^r (lf^2)^{-1}\, dr.$$

This isometry cannot be extended analytically to a global isometry of the two manifolds, since some of the geodesic arcs which are identified under this isometry can be extended to infinite length in one manifold but not in the other.

free space–time. As yet, no example is known of a nonflat singularity-free cosmological solution with vanishing cosmological constant. The behavior we are led to conjecture for an extended NUT space would also be quite remarkable. The unequal expansion rates in different directions of the closed spacelike hypersurface $r = 0$ would smoothly develop into a situation where closed spacelike hypersurfaces no longer could exist, while the evolution in time would smoothly resolve itself into a state of affairs which was periodic in time. A difficulty which will arise in attempting to eliminate the Schwarzschild-like singularities at $f^2 = 0$ using Kruskal's methods is the time periodicity of NUT space. For instance, in the Schwarzschild solution for $r > 2m$ (or for $0 < r < 2m$), we can easily identify points to give a periodic time, $t \equiv t + T$. In Kruskal's extended Schwarzschild solution,[2] we may then attempt to make this same identification. The singularity in the coordinate t causes no difficulty, since it is not the t coordinate, but motions along the Killing vector field $\partial/\partial t$ which define the identifications we wish to make, and $\partial/\partial t$ is an analytic vector field, everywhere free from singularities, as is evident from its expression in terms of Kruskal's nonsingular coordinates u, v:

$$\frac{\partial}{\partial t} = \frac{1}{4m}\left(v\frac{\partial}{\partial u} + u\frac{\partial}{\partial v}\right). \tag{57}$$

What happens, then, when we identify points which differ by a motion of amount T, i.e. when we identify points P and $\exp\{T\ \partial/\partial t\}P$? In fact, this introduces a singularity, not everywhere along the null 3-surfaces $r = 2m$ (or $u = \pm v$), but only at the single 2-surface $u = v = 0$ where, by Eq. (57), $\partial/\partial t = 0$ so the Killing motion has fixed points.

APPENDIX A. COMPUTATION OF THE CURVATURE

We shall compute the Riemann tensor by methods due to Cartan[20] which, at least for metrics with a considerable amount of symmetry, are much more efficient than the methods usually employed by physicists. If $\omega^1, \omega^2, \cdots \omega^n$ are a set of covariant basis vectors, then the metric tensor is written

$$ds^2 = g_{\mu\nu}\omega^\mu\omega^\nu. \tag{A1}$$

Covariant derivatives are formed with the aid of the connection forms $\omega^\mu{}_\nu$, or their components $\Gamma^\mu{}_{\alpha\beta}$:

$$\omega^\mu{}_\alpha = \Gamma^\mu{}_{\alpha\beta}\,\omega^\beta. \tag{A2}$$

[20] E. Cartan [cf. reference 10, Chap. VII]; T. J. Willmore [cf. reference 5, Secs. VII-16 and VII-19]; H. Flanders, Trans. Am. Math. Soc. **75**, 311 (1953).

The following two sets of equations determine the $\omega^\mu{}_\alpha$ uniquely:

$$dg_{\mu\nu} = \omega_{\mu\nu} + \omega_{\nu\mu} \tag{A3}$$

$$d\omega^\mu = -\omega^\mu_\nu \wedge \omega^\nu. \tag{A4}$$

In the more familiar case of a *coordinate* frame $\omega^\mu = dx^\mu$, the second equation here gives $\Gamma^\mu{}_{\alpha\beta} = \Gamma^\mu{}_{\beta\alpha}$ (using the property $d^2 = 0$ of the exterior derivative, and the antisymmetry of the exterior product $\wedge$), while the first is a standard relation between the metric derivatives and the $\Gamma_{\mu\alpha b}$, which is solved to show that $\Gamma^\mu{}_{\alpha\beta}$ is a Christoffel symbol. We will use these equations in a different case, that of an orthonormal frame, where, since $g_{\mu\nu} = \eta_{\mu\nu} = \text{const}$, Eq. (A3) states that the forms $\omega_{\mu\nu}$ are anti-symmetric

$$\omega_{\mu\nu} + \omega_{\nu\mu} = 0. \tag{A5}$$

With the aid of this antisymmetry, Eqs. (A4) can now be solved for $\omega^\mu{}_\nu$ when ω^μ is given. Although a formula like the Christoffel relation exists also in this case, *the computation is most efficient when Eqs. (A) can be solved by inspection*, as we shall shortly illustrate. Once the connection forms have been computed, the curvature forms $\theta^\mu{}_\nu$ are obtained from the formula

$$\theta^\mu{}_\nu = d\omega^\mu{}_\nu + \omega^\mu{}_\alpha \wedge \omega^\alpha{}_\nu. \tag{A6}$$

The components of the Riemann tensor $R^\mu{}_{\nu\alpha\beta}$ are then read out of these curvature froms:

$$\theta^\mu{}_\nu = \tfrac{1}{2}R^\mu{}_{\nu\alpha\beta}\,\omega^\alpha \wedge \omega^\beta, \tag{A7}$$

and the Ricci tensor is formed by contraction

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}. \tag{A8}$$

Note that in 4-space, with an *orthonormal* basis ω^μ, there are only six connection forms $\omega^\mu{}_\nu$ in contrast to forty Christoffel symbols, and only six curvature forms $\theta^\mu{}_\nu$ in contrast to twenty components $R_{\mu\nu\alpha b}$ of the Riemann tensor or ten for the Ricci tensor. In simple cases, such as NUT space, a savings of labor on a scale suggested by these numbers is actually attained, and the Ricci tensor can be computed much more rapidly by these methods (which provide the Riemann tensor as a bonus along the way) than from the usual formula in terms of Christoffel symbols.

The computation begins by writing the metric in terms of an orthonormal frame, as has been done in Eqs. (7) and (8). Next the curl, $d\omega^\mu$, of each base vector ω^μ must be computed. Let us compute, for example, $d\omega^0$ where

$$\omega^0 = f(r)[dt + 4l\sin^2\tfrac{1}{2}\theta\, d\phi]. \tag{A9}$$

First note that $d(f[\ ]) = df \wedge [\ ] + f d[\ ]$, and $df = f'dr = f'f\omega^1$ since $\omega^1 = f^{-1}dr$. Thus $d\omega^0 = f'\omega^1 \wedge \omega^0 + f d[\ ]$. Using $d^2 = 0$ we then find

$$d[\ ] = 4l \sin \tfrac{1}{2}\theta \cos \tfrac{1}{2}\theta\, d\theta \wedge d\phi = 2l \sin\theta\, d\theta \wedge d\phi$$
$$= 2l(r^2 + l^2)^{-1}\omega^2 \wedge \omega^3.$$

Proceeding in this manner we obtain

$$\begin{aligned} d\omega^0 &= f'\omega^1 \wedge \omega^0 + 2lf(r^2 + l^2)^{-1}\omega^2 \wedge \omega^3, \\ d\omega^1 &= 0, \\ d\omega^2 &= rf(r^2 + l^2)^{-1}\omega^1 \wedge \omega^2, \\ d\omega^3 &= rf(r^2 + l^2)^{-1}\omega^1 \wedge \omega^3 \\ &\qquad + (r^2 + l^2)^{-\frac{1}{2}} \cot\theta\, \omega^2 \wedge \omega^3. \end{aligned} \tag{A10}$$

Now we must compare this set of equations with Eqs. (A4) and pick out the $\omega^\mu{}_\nu$. The first equation of (A10), for instance, must take the form

$$d\omega^0 = -\omega^0{}_1 \wedge \omega^1 - \omega^0{}_2 \wedge \omega^2 - \omega^0{}_3 \wedge \omega^3,$$

so we guess that $\omega^0{}_1 = +f'\omega^0$. By antisymmetry we have $\omega^0{}_1 = -\omega_{01} = \omega_{10} = \omega^1{}_0$, and can verify that this choice is consistent with $0 = d\omega^1 = -\omega^1{}_0 \wedge \omega^0 + \cdots$ since $\omega^0 \wedge \omega^0 = 0$. The remaining term in $d\omega^0$ of Eqs. (A10) could arise either from $\omega^0{}_2$ or $\omega^0{}_3$, and this choice can most conveniently be settled later, so we proceed to the $d\omega^2$ equation. From $d\omega^2 = -\omega^2{}_0 \wedge \omega^0 - \omega^2{}_1 \wedge \omega^1 - \omega^2{}_3 \wedge \omega^3$ we guess that $\omega^2{}_1 = rf(r^2 + l^2)^{-1}\omega^2$, so that $\omega^2{}_0$ and $\omega^2{}_3$ terms must cancel here. In this manner, one proceeds to fill out the list below. As the solution $\omega^\mu{}_\nu$ of Eqs. (A4) and (A5) is known to be unique, the proof that a guess for a set of $\omega^\mu{}_\nu$ is correct is simply that it satisfies these equations. We have then

$$\begin{aligned} \omega^0{}_1 &= +\omega^1{}_0 = f'\omega^0, \\ \omega^0{}_2 &= +\omega^2{}_0 = lf(r^2 + l^2)^{-1}\omega^3, \\ \omega^0{}_3 &= +\omega^3{}_0 = -lf(r^2 + l^2)^{-1}\omega^2, \\ \omega^2{}_3 &= -\omega^3{}_2 = lf(r^2 + l^2)^{-1}\omega^0 - (r^2 + l^2)^{-\frac{1}{2}} \cot\theta\, \omega^3, \\ \omega^3{}_1 &= -\omega^1{}_3 = rf(r^2 + l^2)^{-1}\omega^3, \\ \omega^1{}_2 &= -\omega^2{}_1 = -rf(r^2 + l^2)^{-1}\omega^2. \end{aligned} \tag{A11}$$

Of a possible 24 connection components $\Gamma^\mu{}_{\alpha\beta}$, only the seven nonvanishing ones which appear via Eq. (A2) in Eqs. (A11) caused us any labor. That the others vanish we discovered by finding no need for additional terms in Eqs. (A11), not by explicitly evaluating a formula for $\Gamma^\mu{}_{\alpha\beta}$ and obtaining a zero result. The curvature computation is now purely mechanical. The first of Eqs. (A6), for instance, reads

$$\theta^0{}_1 = d\omega^0{}_1 + \omega^0{}_2 \wedge \omega^2{}_1 + \omega^0{}_3 \wedge \omega^3{}_1. \tag{A12}$$

Only two terms appear in the sum because of the antisymmetry of $\omega_{\mu\nu}$. Substituting in this and similar formulas from Eq. (A11) yields

$$\begin{aligned} \theta_{01} &= -2A\omega^0 \wedge \omega^1 - 2D\omega^2 \wedge \omega^3, \\ \theta_{02} &= +C\omega^0 \wedge \omega^2 + D\omega^3 \wedge \omega^1, \\ \theta_{03} &= +C\omega^0 \wedge \omega^3 + D\omega^1 \wedge \omega^2, \\ \theta_{23} &= +2B\omega^2 \wedge \omega^3 - 2D\omega^0 \wedge \omega^1, \\ \theta_{31} &= -C\omega^3 \wedge \omega^1 + D\omega^0 \wedge \omega^2, \\ \theta_{12} &= -C\omega^1 \wedge \omega^2 + D\omega^0 \wedge \omega^3, \end{aligned} \tag{A13}$$

where

$$\begin{aligned} A &= \tfrac{1}{4}(f^2)'', \\ B &= \tfrac{1}{2}[-f^2 + 1 + 4l^2f^2/(r^2 + l^2)](r^2 + l^2)^{-1}, \\ C &= [\tfrac{1}{2}r(f^2)' + l^2f^2/(r^2 + l^2)](r^2 + l^2)^{-1}, \\ D &= [\tfrac{1}{2}l(f^2)' - rlf^2/(r^2 + l^2)](r^2 + l^2)^{-1}. \end{aligned} \tag{A14}$$

The first line of Eqs. (A13), for example, tells us, by comparison with Eq. (A6), that $R_{0101} = -2A$, $R_{0123} = -2D$, $R_{0102} = 0$, $R_{0131} = 0$, etc. The contractions necessary to form $R_{\mu\nu}$, for instance

$$\begin{aligned} R_{11} &= R_1{}^0{}_{10} + R_1{}^2{}_{12} + R_1{}^3{}_{13} \\ &= -R_{1010} + R_{1212} + R_{1313} \\ &= 2A - 2C, \end{aligned}$$

are readily performed by scanning Eqs. (A13). Thus we find

$$R \equiv R^\mu_\mu \equiv R^{\mu\nu}{}_{\mu\nu} = 4(A + B - 2C), \tag{A15}$$

while for

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R,$$

the only nonvanishing components are

$$\begin{aligned} G_{11} &= -G_{00} = 2(C - B), \\ G_{22} &= G_{33} = 2(C - A). \end{aligned} \tag{A16}$$

The empty-space Einstein equations thus require $A = B = C$.

Each of the quantities A, B, C, and D is effectively an invariant since the basis vectors ω^0 and ω^1 we used can be characterized geometrically, while the form (A13) of the curvature tensor is invariant under rotations in the 23 plane. We can characterize the vector ω^1 as the unit normal to the orbits $r =$ const of the group of motions; similarly, the contravariant vector $\mathbf{e}_0$ from the dual basis can be characterized as the unit vector parallel to that unique Killing vector which commutes with all the Killing vectors. These four invariants reduce

to two when the Einstein equations are satisfied, and for the Schwarzschild solution, where $l = 0$, evidently $D = 0$ so there remains only one invariant, $A = m/r^3$. For NUT space we have

$$\begin{aligned} A &= B = C = (r^2 + l^2)^{-3} \\ &\quad \times (mr^3 + 3l^2r^2 - 3ml^2r - l^4), \\ D &= -l(r^2 + l^2)^{-3}(r^3 - 3mr^2 - 3l^2r + ml^2). \end{aligned} \tag{A17}$$

APPENDIX B. VECTOR FIELDS ON S^3

We collect here some well known properties of S^3, including a computation of the commutation relations in Eq. (35). As coordinates for a quaternion q, we take $w(q)$, $x(q)$, $y(q)$, $z(q)$, defined by

$$q = w + ix + jy + kz. \tag{B1}$$

The unit quaternions satisfying

$$|q|^2 \equiv qq^* \equiv w^2 + x^2 + y^2 + z^2 = 1 \tag{B2}$$

constitute S^3. They form a group under multiplication since $|qp| = |q| \cdot |p|$. This group S^3 can be related to the rotation group $SO(3)$ as follows: Identify the imaginary quaternions r which satisfy $r + r^* = 0$ [i.e. $w(r) = 0$] with Euclidean 3-space R^3. Then the transformations

$$r \to qrq^* \tag{B3}$$

map R^3 into itself [check $qrq^* + (qrq^*)^* = 0$ using the rule $(pq)^* = q^*p^*$] and, when $|q| = 1$, preserve $|r|$ which is just the Euclidean norm since $w(r) = 0$. Both q and $-q$ give the same transformation (B3), so S^3 covers $SO(3)$ twice.

Both the transformations

$$q \to pq \tag{B4ξ}$$

and

$$q \to qp^*, \tag{B4η}$$

where $|p| = 1$ preserve the quaternion norm $|q|$ and therefore map S^3 into itself. The infinitesimal generators of these transformations are respectively the vector fields ξ_a and $\mathbf{n}_a$ defined in Table I. For instance ξ_x generates the one-parameter subgroup of (B4ξ) transformations obtained by taking $p = e^{i\alpha/2}$. We now want to represent ξ_x in terms of its components along the base vectors $\partial_w = \partial/\partial w$, etc. associated with the coordinate system we are using. Differentiation of $f(q)$ along ξ_x is defined by

$$\xi_x[f] = [(d/d\alpha)f(e^{i\alpha/2}q)]_{\alpha=0}. \tag{B5}$$

TABLE I. The vector fields ξ_a and $\mathbf{n}_a$ on S^3 are defined as the infinitesimal generators $\partial/\partial\alpha$ of certain one-parameter groups of motions of S^3 as shown.

Group of motions	One-parameter subgroup $p = e^{i\alpha/2}$	$p = e^{j\alpha/2}$	$p = e^{k\alpha/2}$
left translations $q \to pq$	ξ_x	ξ_y	ξ_z
right translations $q \to qp^*$	$\mathbf{n}_x$	$\mathbf{n}_y$	$\mathbf{n}_z$

To compute the components of ξ_x, namely $\xi_x[w]$ etc., we note that $d(e^{i\alpha/2}q)/d\alpha = (\frac{1}{2})iq$; so, upon taking the various real and imaginary parts of this relation we find $\xi_x[w] \equiv \xi_x[\mathrm{Re}\ q] = \mathrm{Re}\ [(\frac{1}{2})iq] = -\frac{1}{2}x$, etc. The results of this and similar computations are summarized as follows:

$$\begin{aligned} 2\xi_x &= -x\partial_w + w\partial_x - z\partial_y + y\partial_z, \\ 2\xi_y &= -y\partial_w + z\partial_x + w\partial_y - x\partial_z, \\ 2\xi_z &= -z\partial_w - y\partial_x + x\partial_y + w\partial_z, \\ 2\mathbf{n}_x &= +x\partial_w - w\partial_x - z\partial_y + y\partial_z, \\ 2\mathbf{n}_y &= +y\partial_w + z\partial_x - w\partial_y - x\partial_z, \\ 2\mathbf{n}_z &= +z\partial_w - y\partial_x + x\partial_y - w\partial_z. \end{aligned} \tag{B6}$$

The commutators of these differential operators are now straightforwardly computed with the result given in Eq. (35). For these computations, we have regarded ξ_a and $\mathbf{n}_a$ as defined over the entire space R^4 of all quaternions, which permitted the convenience of using rectangular coordinates. Since the results [Eqs. (35)], do not refer to any coordinate system, and since all the vectors involved are tangent to S^3, the method of extending the definition of the vector fields beyond S^3 was irrelevant.

Initial-Value Problem on Einstein–Rosen Manifolds

RICHARD W. LINDQUIST
Adelphi College, Garden City, New York
(Received 11 March 1963)

A class of solutions of the time-symmetric initial-value equations for gravitation and electromagnetism is obtained on a two-sheeted manifold containing N Einstein–Rosen bridges. The initial metric tensor and electric field are expressed in terms of a pair of harmonic functions, called "metric potentials," which are required to be analytic and asymptotically flat and to satisfy certain match-up conditions; these potentials are then determined by applying the method of inversion images. The particular case of two identical Einstein–Rosen bridges is examined in detail, and is shown to correspond to a wormhole in an asymptotically flat (one-sheeted) universe. A charge and "renormalized mass" are defined for each Einstein–Rosen bridge, as well as an interaction energy (gravitational plus electrostatic potential energy) for the system as a whole. Expressions for these and other physical characteristics are evaluated explicitly in the two-body case.

I. INTRODUCTION

MUCH can be learned about the physical implications of general relativity by looking for solutions of the *source-free* field equations for gravitation plus electromagnetism. This approach, first advocated by Einstein and Rosen[1] in 1935, has recently received considerable attention, particularly by Wheeler,[2] who gave it the name "geometrodynamics." Wheeler showed that many of the features one normally associates with classical charged particles can be extracted from a metric tensor which is *everywhere continuous* (and even analytic), on a manifold with a non-Euclidean topology.

In the development of geometrodynamics, particular attention has been focused on the *time-symmetric initial-value problem*, since this leads to physically significant results with a minimum of mathematical labor.[3] (Misner has suggested the term "geometrostatics" for the study of this important problem. In the absence of any nontrivial and regular static solutions of the Einstein–Maxwell equations, this is perhaps the closest analogue in general relativity to the familiar special case of electrostatics within electromagnetic theory.) One assumes, in this very restricted analysis, that there exists a spacelike hypersurface Σ on which the magnetic field components f_{ij} of the Maxwell tensor, and the second fundamental form K_{ij} vanish. The remaining dynamical variables, $f_{i0} \equiv E_i$, and g_{ij}, cannot then be chosen arbitrarily, but must satisfy a system of coupled elliptic equations, the *initial-value equations*:

$$R = 2E_iE^i, \tag{1.1a}$$

$$E^{i;i} = 0. \tag{1.1b}$$

Here g^{ij} is the 3-metric inverse to g_{ij}, $E^i = g^{ij}E_j$, and R is the curvature scalar (on Σ) computed from g_{ij}. The Riemannian metric g_{ij} and the vector field E_i must be, respectively, of class C^3 and C^2 on Σ,[4] and must satisfy suitable boundary conditions. (The latter usually take the form of a prescription of the way in which the fields fall off "at infinity.") While the continuity assumptions given above are all that is necessary for the initial-value problem in general relativity, we shall further assume that Σ is an *analytic* manifold, and that g_{ij} and E_i are everywhere *analytic* on Σ.

Brill and Lindquist[3] have analyzed one particular class of solutions of these equations—previously obtained by Misner and Wheeler[2]—that describes the geometry on a three-dimensional Riemannian manifold containing $N + 1$ "sheets", on each of which the metric becomes asymptotically flat (Fig. 1). One can identify N of these sheets with N "particles", each having a well-defined mass and charge. (The remaining sheet is then taken to represent the "rest of the universe.") Thus these solutions form a natural generalization of the one-body metrics of Schwarzschild and Reissner–Nordström, both of which can be extended[5] to

[1] A. Einstein and N. Rosen, Phys. Rev. **48**, 73 (1935).

[2] C. W. Misner and J. A. Wheeler, Ann. Phys. (N. Y.) **2**, 525 (1957). This and other important papers are collected in J. A. Wheeler, *Geometrodynamics* (Academic Press Inc., New York, 1962).

[3] See for example D. R. Brill, Ann. Phys. (N. Y.) **7**, 466 (1959) and references given there. Recent investigations of the time-symmetric initial-value problem, which are especially pertinent to the present work, will be found in papers by D. R. Brill and R. W. Lindquist, Phys. Rev. **131**, 471 (1963) (hereinafter referred to as BL); and by C. W. Misner, Phys. Rev. **118**, 1110 (1960), and Ann. Phys. (N. Y.) (to be published) (referred to as MI and MII, respectively).

[4] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson et Cie., Paris, 1955), p. 48.

[5] M. D. Kruskal, Phys. Rev. **119**, 1743 (1960); J. C. Graves and D. R. Brill, Phys. Rev. **120**, 1507 (1960).

give a regular metric on a 3-manifold with two sheets, joined together at the Schwarzschild radius.

However, this is not the only possible model of a universe containing N particles consistent with the viewpoint of geometrodynamics. An alternative is the manifold containing two sheets joined together by N handles or "bridges", as shown in Fig. 2. This is in fact the model for the N-body problem originally proposed by Einstein and Rosen,[1] who were the first to point out the corresponding two-sheeted character for the Schwarzschild problem. We shall accordingly refer to these double-sheeted models as "Einstein–Rosen (E–R) manifolds." In MII, Misner has shown how to construct a solution of the time-symmetric initial-value equation for a pure gravitational field on an arbitrary E–R manifold. In the present paper we extend his method to include an electric field as well.

The construction relies on the reflection symmetry of the manifold[6] M_N about a horizontal plane in the imbedding space of Fig. 2. Expressed in intrinsic terms, this symmetry implies the existence of N distinct homeomorphisms $J_k : M_N \to M_N$ ($k = 1, \cdots N$), each of which is an isometry of M_N onto itself, and each of which leaves invariant a 2-sphere S_k, which we shall call the kth "throat" or the kth inversion sphere.

In Sec. II we investigate the structure of analytic scalar and vector fields on M_N, and formulate the conditions that must be imposed on a metric and electric field on M_N in order that they shall (a) satisfy the time-symmetric initial-value equations, and (b) be invariant under each of the mappings J_k (in a sense to be made precise later). These conditions express the requirement that the metric and electric fields shall match up smoothly (in fact, analytically) when the two sheets are joined together at each throat.

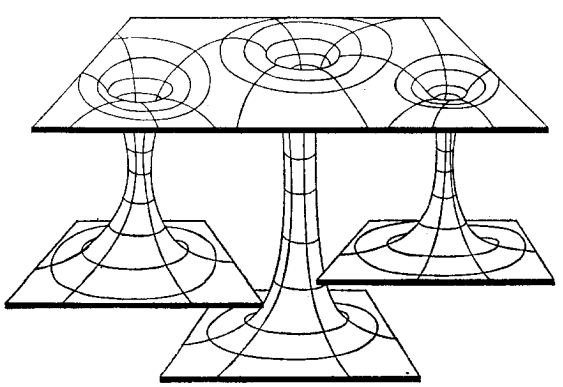

FIG. 1. A two-dimensional section of the hypersurface of time symmetry for a manifold containing 3 + 1 sheets is shown imbedded in flat 3-space to suggest its curvature and topology. Each of the three lower sheets becomes asymptotically flat and represents a single Reissner–Nordström particle.

[6] Throughout this paper we denote the Einstein–Rosen (or briefly, E–R) manifold containing N bridges by M_N.

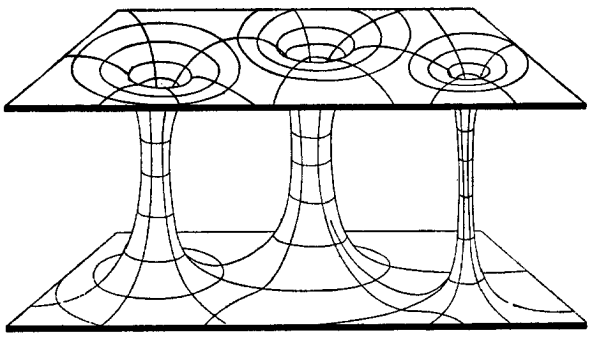

FIG. 2. A two-dimensional section of the Einstein–Rosen manifold M_3 is shown imbedded in flat 3-space. This manifold is obtained from the simply-connected space of Fig. 1 by joining the three lower sheets smoothly together. Note the symmetry between the two sheets under reflection in a horizontal plane in the imbedding space.

In Sec. III we exhibit a solution of the time-symmetric initial-value equations satisfying these match-up conditions. The solution is first obtained as a formal power series, by a method which parallels closely the well-known method of images of electrostatics. We then examine the conditions under which this formal series actually converges to an analytic solution of the initial-value equations.

The physical characteristics of this solution are not transparent, nor are they easily read out of the formal solution in the general case. We therefore, in Secs. IV and V, specialize our results to the case of a manifold containing two isometric E–R bridges, and investigate some of its properties. This solution provides a model of a universe containing two particles of equal mass and equal but opposite charge. In this special case the manifold M_N forms an isometric double covering of the Wheeler–Misner "wormhole" manifold; hence the solution we obtain here generalizes Misner's initial-value metric for a purely gravitational wormhole (given in MI) to one containing a divergence-free electric field as well.

II. FORMULATION OF THE INITIAL-VALUE PROBLEM

The Metric Potentials

Instead of trying to solve the initial-value equations (1.1) on M_N directly, we shall build up the desired solution out of initial-value data on E^3. Following the notation of BL, we set

$$ds^2 = (\chi\psi)^2\, ds^2_\flat, \tag{2.1}$$

$$E_i = -\phi_{,i} = [\ln(\chi/\psi)]_{,i}. \tag{2.2}$$

Here $ds_\flat^2$ is the flat-space metric, and the "metric potentials" χ and ψ are linearly independent function which satisfy

$$\begin{aligned} &\text{(i)} \quad \chi, \psi \to 1 \quad \text{as} \quad r \to \infty, \\ &\text{(ii)} \quad \chi > 0 \quad \text{and} \quad \psi > 0 \quad \text{everywhere.} \end{aligned} \tag{2.3}$$

Eqs. (1.1), together with these boundary conditions, imply that χ and ψ must be solutions of the (flat-space) Laplace equation, of the particular form

$$\begin{aligned} \chi &= 1 + \sum_{i=1}^{N} \alpha_i |\mathbf{r} - \mathbf{r}_i|^{-1}, \\ \psi &= 1 + \sum_{i=1}^{N} \beta_i |\mathbf{r} - \mathbf{r}_i|^{-1}, \end{aligned} \tag{2.4}$$

with

$$\alpha_i > 0 \quad \text{and} \quad \beta_i > 0.$$

These potentials describe an everywhere regular metric and electric field on a 3-manifold with $N + 1$ sheets, obtained by deleting from E^3 the N points $\mathbf{r} = \mathbf{r}_i$.[7]

If we construct the manifold M_N from E^3 by following the procedure outlined in the next section, we find that the potentials (2.4) induce on M_N a metric structure which is continuous, but which is not in general of class C^1 at the N throats S_k. (The only exception occurs when $N = 1$; the metric in this case reduces to that of Reissner–Nordström.) To obtain a potential with a continuous first derivative, we must add additional terms to the series (2.4), corresponding to the addition of simple poles at the images of the points $\mathbf{r}_i$. The resulting potential has a discontinuity in its second derivative, so we add still other poles at the images of these image points, and so on, until we have built up an infinite sequence of images. If the original spheres are not chosen too close together, this sequence of potentials converges to a limiting function, which satisfies the initial-value equations and is analytic on M_N.

Such an approach is reminiscent of the classical problem of determining the potential distribution due to N charged conducting metal spheres. There is in fact a close formal resemblance between these two problems: in place of the demand that the S_k be equipotential surfaces, one here requires that the metric potentials satisfy certain "match-up conditions," which will ensure that the metric and electric fields be analytic on M_N. The precise form of these match-up conditions is only trivially different from the equipotential requirement of the electrostatic problem. It is thus not surprising that the method of inversion images should also play an important role in solving the initial-value problem.

[7] It is easy to show that the metric becomes flat as $\mathbf{r} \to \mathbf{r}_i$. See, e.g., BL or MII.

Analytic Structure of M_N

In order to formulate these match-up conditions precisely, we must first give M_N the structure of an analytic manifold. This consists of choosing a family U_i of open sets covering M_N, together with a set of homeomorphisms f_i of each U_i into an open set of Euclidean 3-space (these U_i's are the coordinate patches, and f_i the corresponding coordinate functions), such that whenever U_i and U_j have a nonzero intersection, the mapping $f_i \circ f_j^{-1}$ of $f_j(U_j)$ onto $f_i(U_i)$ is analytic.

The essential idea behind the construction is to start with two copies of E^3, cut out the interiors of N distinct spheres in each of them, and identify corresponding points on their boundaries. To make this idea precise, let E_x^3 and E_y^3 be two euclidean 3-spaces whose points are labelled $\mathbf{x} = (x^1, x^2, x^3)$ and $\mathbf{y} = (y^1, y^2, y^3)$, respectively. (The components x^i and y^i thus form a set of standard coordinates on E_x^3 and E_y^3.) Draw in each space N nonintersecting spheres S_k, with radii a_k and centers at $\mathbf{C}_k$, and form the subsets

$$\Delta^+ = \{\mathbf{x} \in E_x^3 \mid |\mathbf{x} - \mathbf{C}_k| > a_k;\ k = 1, \cdots N\}$$

$$\Delta^- = \{\mathbf{y} \in E_y^3 \mid |\mathbf{y} - \mathbf{C}_k| > a_k;\ k = 1, \cdots N\}$$

Then M_N can be regarded—as a point set—as the union $\bar{\Delta}^+ \cup \Delta^-$, with Δ^+ and Δ^- representing the upper and lower sheets respectively, and the boundary of $\Delta^+ (= \bar{\Delta}^+ - \Delta^+)$ representing the N "throats."

Since Δ^+ is open, each point $\mathbf{x}$ in Δ^+ has a neighborhood lying in Δ^+, which we can use to define a family of neighborhoods in M_N. The same applies to each point $\mathbf{y} \in \Delta^-$. However, for points on the boundary of Δ^+ the construction of a family of neighborhoods is rather more involved. Indeed, it is just this construction which joins the two sets $\bar{\Delta}^+$ and Δ^- together and makes M_N a connected manifold.

First introduce the inversion mapping in S_k, defined by

$$\mathbf{r}_k \to J_k(\mathbf{r}_k) = (a_k/r_k)^2 \mathbf{r}_k, \tag{2.5}$$

with

$$\mathbf{r}_k \equiv \mathbf{x}_k - \mathbf{C}_k \quad \text{and} \quad r_k = |\mathbf{r}_k|. \tag{2.6}$$

J_k maps the punctured 3-plane $E_x^3 - (\mathbf{C}_k)$ homeo-

morphically onto itself. It is clearly analytic (except at the point $\mathbf{x} = \mathbf{C}_k$, which we exclude), and in addition satisfies

$$J_k^2 = I \quad \text{(identity map)}, \tag{2.7a}$$

$$J_k(S_k) = S_k. \tag{2.7b}$$

In particular, therefore, J_k maps the open set Δ^+ onto an open set Γ_k in the interior of S_k (see Fig. 3):

$$\Gamma_k = \{\mathbf{z} \in E_x^3 \mid \mathbf{z} = J_k\mathbf{x}, \mathbf{x} \in \Delta^+\}.$$

Let D_k be the set $\Delta^+ \cup S_k \cup \Gamma_k$; D_k is evidently an open subset of E_x^3. (It represents the upper and lower sheets of M_N, joined together at the kth sphere.) Let $J_k^* : E_x^3 \to E_y^3$ be the combined mapping $\eta \circ J_k$, where η assigns to each point $\mathbf{x}$ of E_x^3 the point $\mathbf{y}$ of E_y^3 with the same coordinate values. Then J_k^* sends Γ_k homeomorphically onto Δ^-. Since Δ^+, S_k, and Γ_k are disjoint subsets of E_x^3, this defines a mapping φ_k of D_k *into* M_N, defined by

$$\varphi_k(\mathbf{x}) = \begin{matrix} \mathbf{x} & \text{if} & \mathbf{x} \in \Delta^+ \cup S_k, \\ J_k^*(\mathbf{x}) & \text{if} & \mathbf{x} \in \Gamma_k. \end{matrix} \tag{2.8}$$

We now assign a topology to M_N by requiring that each map $\varphi_k : D_k \to M_N (k = 1, \cdots N)$ shall be continuous. That is, we call a subset $U \subset M_N$ *open* if and only if U is either: (a) the image $\varphi_k(U^+)$ of an open set $U^+ \subset D_k$ for some k, or (b) the union of such images.

It is clear that the N sets $\varphi_k(D_k) \equiv U_k$ cover M_N; these can therefore be used as coordinate patches, with the corresponding maps $\varphi_k^{-1}(U_k)$ serving as coordinate functions. To prove that the coordinates so defined give M_N the structure of an analytic manifold, we must show that whenever two coordinate patches overlap, the transformation functions $\varphi_{k'}^{-1} \circ \varphi_k$ are analytic. This is, however, a trivial consequence of the definition (2.8), together with the analyticity of J_k.

Analytic Scalar and Vector Fields on M_N

Let f^* be any real-valued function defined on M_N, and let φ_k be the coordinate functions defined by (2.8). Then f^* is analytic if and only if the function $f : E_x^3 \to R$ defined by $f = f^* \circ \varphi_k$ is analytic in the ordinary sense (i.e., can be expanded in a convergent Taylor series within each coordinate patch D_k in each of its arguments x^1, x^2, x^3).

Conversely, suppose that $f(\mathbf{x})$ is a real-valued analytic function defined on E_x^3. Under what conditions will f induce an analytic function f^* on M_N? It is clear that if such an f^* exists it must satisfy $f^* = f \circ \varphi_k^{-1}$ for $k = 1, \cdots N$. Let $\mathbf{z}$ be any point of M_N, and let $\mathbf{x}_k$ be the image of $\mathbf{z}$ in $D_k : \mathbf{x}_k = \varphi_k^{-1}(\mathbf{z})$. Then f must satisfy the conditions

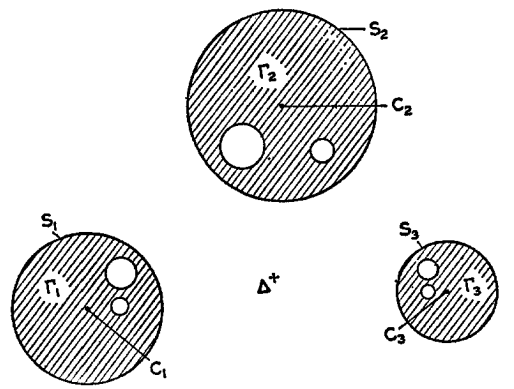

FIG. 3. Illustration of the construction of M_3 from Euclidean 3-space. The three spheres are represented by the circles S_1, S_2, S_3 with centers $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$. Δ^+ is the region outside all three circles. The open set Γ_1 consists of the shaded region inside S_1 excluding the point $\mathbf{C}_1$, while D_1 is the union of Δ^+, Γ_1 and their common boundary S_1. Γ_2, Γ_3, D_2, and D_3 are defined similarly.

$$f(\mathbf{x}_k) = f(\mathbf{x}_{k'}) \qquad (k, k' = 1, \cdots N), \tag{2.9}$$

if f^* is to be single-valued. This in turn implies

$$f(J_k\mathbf{x}) = f(J_{k'}\mathbf{x}), \tag{2.10a}$$

or

$$f(\mathbf{x}) = f(J_kJ_{k'}\mathbf{x}). \tag{2.10b}$$

[Strictly speaking, Eq. (2.9) only implies (2.10a) for values of $\mathbf{x}$ in Δ^+. However, it follows by analytic continuation that (2.10a) and consequently (2.10b) must hold for all $\mathbf{x}$ in the domain of analyticity of f.]

It is not necessary that f be analytic at every point of E^3 in order that it lead to an analytic function f^* on M_N. Indeed, suppose that f is defined and analytic in $\cup D_k$, and satisfies Eq. (2.10a) for $\mathbf{x} \in \Delta^+$; then f can be extended, using (2.10b), to a function $\hat{f}$ which is analytic in a domain D defined as follows[8]:

(i) $\bar{\Delta}^+ \subset D$,

(ii) if $\mathbf{x} \in D$, then $J_k(\mathbf{x}) \in D$.

The function $\hat{f}$ will in general be the maximal analytic extension of f. Thus $\hat{f}$ may have isolated singularities (poles) at points in the complement $\bar{G}$ of D.

Lemma 1. Let $f(\mathbf{x})$ be a real-valued analytic function on the domain D of E_x^3 defined above. In order that $f(\mathbf{x})$ shall induce an analytic function f^* on M_N it is necessary and sufficient that

[8] The following definition is equivalent: Let G be the set of points of E^3 obtained by applying the maps $J_{k'}$ to $\{\mathbf{C}_k\}$ a finite number of times (including the points $\mathbf{C}_k$ themselves), and let $\bar{G}$ be its closure; then D is the complement of $\bar{G}$.

$$f(\mathbf{x}) = f(J_k J_{k'}\mathbf{x}) \qquad (k, k' = 1, \cdots N)$$

for all $\mathbf{x} \in D$.

Proof: The necessity of this condition has already been established. To prove sufficiency, define a function f^* on M_N by

$$f^*(\mathbf{z}) = \begin{cases} f(\mathbf{z}) & \text{if } \mathbf{z} \in \bar{\Delta}^+, \\ f(J_k^{*-1}\mathbf{z}) & \text{if } \mathbf{z} \in \Delta^-. \end{cases} \tag{2.11}$$

The choice of k in this definition is clearly immaterial if the above condition is satisfied. Then $f^* \circ \varphi_k = f(\mathbf{x})$ for $\mathbf{x} \in D_k$, whence f^* is analytic.

A similar result applies to the case of analytic vector fields. It is most convenient to adopt the point of view of modern differential geometry, in which an (analytic) tangent vector at the point $\mathbf{x}$ is regarded as a linear map $\mathbf{V}_x$ of the (analytic) functions at $\mathbf{x}$ into the real numbers, satisfying

$$\mathbf{V}_x[fg] = (\mathbf{V}_x[f])g(\mathbf{x}) + f(\mathbf{x})\mathbf{V}_x[g]$$

for any pair of (analytic) functions f and g.[9]

If one introduces a system of coordinates x^i and takes the tangent vectors to these coordinate lines at $\mathbf{x}$ as a basis, he can represent $\mathbf{V}_x$ as the operator

$$\mathbf{V}_x = (V^i\, \partial/\partial x^i)_x, \tag{2.12}$$

where V^i are the usual contravarient components. Thus $\partial/\partial x^i$ may be identified with the ith basis vector $\mathbf{e}_i$.

Let $\mathbf{V}[f]$ be an analytic vector field on D. The mapping $\varphi_k : D_k \to M_N$ induces a mapping of the tangent space at $\mathbf{x}$ onto the tangent space at $\mathbf{z} = \varphi_k(\mathbf{x})$, given by

$$\mathbf{V}^*[f^*] = \mathbf{V}[f^* \circ \varphi_k] = \mathbf{V}[f]. \tag{2.13}$$

In order that $\mathbf{V}^*[f^*]$ be well-defined over M_N, its value at the point $\varphi_k(\mathbf{x})$ must not depend on the particular coordinate patch (i.e., the particular value of k) chosen. This imposes on $\mathbf{V}[f]$ the condition[10]

$$\mathbf{V}(J_k\mathbf{x}) = \mathbf{V}(J_{k'}\mathbf{x}), \tag{2.14a}$$

or equivalently,

$$\mathbf{V}(\mathbf{x}) = \mathbf{V}(J_k J_{k'}\mathbf{x}). \tag{2.14b}$$

Up to this point we have not made any assumptions about the symmetry between the two sheets of M_N. We now impose the further restriction that these two sheets shall be *physically equivalent*. This means in particular that the inversion map J_k shall be an isometry, and more generally that any scalar field of physical interest shall have the same value at corresponding points $\mathbf{x}$ and $J_k(\mathbf{x})$. Thus the match-up condition (2.10b)—which must hold for *any* scalar field on M_N—is to be replaced by the more stringent requirement

$$f(\mathbf{x}) = f(J_k\mathbf{x}), \qquad (k = 1, \cdots N). \tag{2.15}$$

A similar restriction applies to the physically acceptable vector fields on M_N. Here, however, two possibilities arise, according as $\mathbf{V}$ is even or odd under the mapping J_k:

$$\mathbf{V}(\mathbf{x}) = \pm\mathbf{V}(J_k\mathbf{x}). \tag{2.16}$$

Remember that $\mathbf{V}_x[f]$ is a number; it can be thought of as the directional derivative of f in the direction of the vector $\mathbf{V}_x$. The statement that $\mathbf{V}$ is even under the mapping J_k means that $\mathbf{V}$ makes the same angle with the gradient of f at corresponding points $\mathbf{x}$ and $J_k\mathbf{x}$, while vectors which are odd make angles with ∇f at corresponding points that are supplementary. Suppose that the gradient of f points away from S_k at points in Δ^+ (corresponding to the upper sheet of M_N); then it follows from (2.15) that it points away from S_k at points in Γ_k and hence on the lower sheet of M_N as well. Thus a vector field is even on M_N if it has the same magnitude and direction at corresponding points on the two sheets; it is odd if it has the same magnitude but *opposite* direction.

We shall be interested in one particular vector field on M_N, viz., an *electric field* $\mathbf{E}^*$ satisfying the time-symmetric initial-value equation (1.1b). This particular field is odd under J_k; that is, it satisfies

$$\mathbf{E}(\mathbf{x}) = -\mathbf{E}(J_k\mathbf{x}). \tag{2.17}$$

This symmetry condition—which is readily checked in the case of the Reissner–Nordström solution—is necessary if the Einstein–Rosen bridge is to describe a model of charged particles. An even field would have zero normal component at the throat S_k, so that the integral over S_k of the normal component of $\mathbf{E}$—which we associate with 4π times the charge of the kth throat—would have to vanish. For an odd field no such difficulty arises: electric lines of

[9] See, e.g., C. Chevalley, *Theory of Lie Groups* (Princeton University Press, Princeton, New Jersey, 1946), p. 77, or S. I. Goldberg, *Curvature and Homology* (Academic Press Inc., New York, 1962), p. 98. The importance of this coordinate-free description of tangent vectors for analyzing problems in general relativity has been stressed by F. K. Manasse and C. W. Misner, J. Math. Phys. **4**, 735 (1963).

[10] Eq. (2.14a) should, strictly speaking, be written as

$$\mathbf{V}_{J_k(\mathbf{x})}[f] = \mathbf{V}_{J_{k'}(\mathbf{x})}[f],$$

but for typographical reasons (and to bring our notation more into line with that familiar to physicists), we shorten it to the form given in the text. We shall treat other vector equations similarly. However, it must always be borne in mind that such vector equations are in fact *operator* equations, and hold only for a suitable class of real-valued functions. In the present case, this class consists of all analytic functions in D satisfying (2.10).

force enter the throat from one side and emerge from the other, thereby giving the illusion of a negative charge to an observer on one sheet and of a positive charge to an observer on the other. (Thus the two sheets of the E–R manifold, while physically equivalent, are not identical, because of this interchange of positive and negative charge. This is clearly not a fundamental distinction, since one can always interchange the labels "positive" and "negative" for all particles without altering their interactions.)

Match-Up Conditions on the Metric Potentials

Let $\chi(\mathbf{x})$ and $\psi(\mathbf{x})$ be two independent solutions of Laplace's equation, which are analytic in the domain D of E_x^3, defined above, and approach unity as $\mathbf{x}$ approaches infinity. These "metric potentials" can be used to define a metric tensor and electric field on D—via Eqs. (2.1) and (2.2)—which automatically satisfy the time-symmetric initial-value equations. The metric tensor so obtained will induce a metric on M_N if and only if each inversion map J_k is an isometry:

$$ds^2(\mathbf{x}) = ds^2(J_k\mathbf{x}). \tag{2.18}$$

Furthermore, the electric field $\mathbf{E}$ will induce an acceptable field on M_N if and only if it satisfies Eq. (2.17). In the present section we investigate the restrictions which these match-up conditions impose on the potentials χ and ψ.[11]

The computations are simplified if one introduces in place of the cartesian coordinates (x^1, x^2, x^3) a set of spherical coordinates (r, θ, φ) centered about the point $\mathbf{C}_k$. This is legitimate, since the latter also forms a set of analytic coordinates on D, and hence the match-up conditions (2.17)–(2.18) maintain the same form when expressed in this coordinate system. The inversion map J_k takes the particularly simple form

$$(r, \theta, \varphi) \rightarrow J_k(r, \theta, \varphi) = (r', \theta', \varphi'), \tag{2.19}$$

with

$$r' = a^2/r, \qquad \theta' = \theta, \qquad \varphi' = \varphi,$$

where we have temporarily set $a_k = a$. It is readily checked that

$$\begin{aligned} ds_{\flat}^2(J_k\mathbf{x}) &\equiv dr'^2 + r'^2(d\theta'^2 + \sin^2\theta'\, d\varphi'^2) \\ &= (a/r)^4(dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2) \\ &= (a/r)^4\, ds_{\flat}^2(\mathbf{x}). \end{aligned} \tag{2.20}$$

Inserting this result, together with the definition (2.1), into Eq. (2.18), one obtains the first match-up condition on χ and ψ:

$$\chi(\mathbf{x})\psi(\mathbf{x}) = (a_k/r_k)^2\chi(J_k\mathbf{x})\psi(J_k\mathbf{x}). \tag{2.21}$$

The remaining match-up condition follows from Eq. (2.17). We use the formulas

$$\begin{aligned} \left(r\frac{\partial f}{\partial r}\right)\Big|_{r'} &= -\left(r\frac{\partial f}{\partial r}\right)\Big|_{r}; \\ \frac{\partial f}{\partial\theta}\Big|_{\theta'} &= \frac{\partial f}{\partial\theta}\Big|_{\theta}; \qquad \frac{\partial f}{\partial\varphi}\Big|_{\varphi'} = \frac{\partial f}{\partial\varphi}\Big|_{\varphi}, \end{aligned} \tag{2.22}$$

for any scalar field f satisfying (2.15)[12], together with the basic definitions (2.2) and (2.12):

$$\begin{aligned} \mathbf{E}[f] = E^i\frac{\partial f}{\partial x^i} &= -(\chi\psi)^{-2}\frac{\partial\phi}{\partial x^i}\frac{\partial f}{\partial x^i}, \\ \phi &= -\ln(\chi/\psi), \end{aligned} \tag{2.23}$$

to obtain

$$\begin{aligned} \left(r\frac{\partial\phi}{\partial r}\right)\Big|_{\mathbf{x}'} &= \left(r\frac{\partial\phi}{\partial r}\right)\Big|_{\mathbf{x}}; \\ \frac{\partial\phi}{\partial\theta}\Big|_{\mathbf{x}'} &= \frac{\partial\phi}{\partial\theta}\Big|_{\mathbf{x}}; \qquad \frac{\partial\phi}{\partial\varphi}\Big|_{\mathbf{x}'} = \frac{\partial\phi}{\partial\varphi}\Big|_{\mathbf{x}}. \end{aligned} \tag{2.24}$$

The first of these relations implies [cf. reference 12]

$$\ln\left[\frac{\chi(r,\ \theta,\ \varphi)}{\psi(r,\ \theta,\ \varphi)}\right] = -\ln\left[\frac{\chi(r',\ \theta',\ \varphi')}{\psi(r',\ \theta',\ \varphi')}\right] + g(\theta, \varphi),$$

while the remaining two imply that $g(\theta, \varphi)$ is a constant, which we write as $2\lambda_k$. This gives the second match-up condition

$$\chi(\mathbf{x})/\psi(\mathbf{x}) = e^{2\lambda_k}\psi(J_k\mathbf{x})/\chi(J_k\mathbf{x}). \tag{2.25}$$

Combining Eqs. (2.21) and (2.25) yields the formulas

$$\chi(\mathbf{x}) = (a_k/r_k)e^{\lambda_k}\psi(J_k\mathbf{x}), \tag{2.26a}$$

$$\psi(\mathbf{x}) = (a_k/r_k)e^{-\lambda_k}\chi(J_k\mathbf{x}). \tag{2.26b}$$

[11] Note that we cannot apply Eq. (2.15) to χ and ψ directly, for they are not true scalar fields. (They do not transform as scalars under arbitrary coordinate transformations.) We cannot apply it to the electrostatic potential $\phi = \ln(\psi/\chi)$ either, since the latter is not a single-valued continuous function on M_N. This follows from the existence of closed lines of force which thread through two or more throats; if ϕ were continuous, the line integral of $\mathbf{E}$ along any such closed path would have to vanish. We have here an illustration of the well-known theorem that if curl $\mathbf{E} = 0$ on a manifold $\mathfrak{M}$, then $\mathbf{E} = -\text{grad}\ \phi$ locally but not globally—unless the first Betti number of $\mathfrak{M}$ vanishes.

[12] To derive the first equation, make the change of variable $u = \ln(r/a)$. Then $J_k(u) = \ln(r'/a) = -u$, so that (2.15) implies $f(u, \theta, \varphi) = f(-u, \theta, \varphi)$. Consequently $(\partial f/\partial u)\,|_u = -(\partial f/\partial u)\,|_{u'}$, from which (2.22) follows.

III. SERIES SOLUTION FOR THE METRIC POTENTIALS

It is convenient at this point to introduce the *linear operator* $\mathcal{J}_k[f]$, defined by

$$\mathcal{J}_k[f](\mathbf{x}) = (a_k/r_k)f(J_k\mathbf{x}). \tag{3.1}$$

Note that $\mathcal{J}_k$, like J_k, satisfies $\mathcal{J}_k^2 = I$. Furthermore,

$$\Delta(\mathcal{J}_k f) = (a_k/r_k)^4 \mathcal{J}_k[\Delta f], \tag{3.2}$$

where Δ is the flat-space Laplacian, so that $\Delta f = 0$ implies $\Delta(\mathcal{J}_k f) = 0$ as well. (It is just this fact which provides the basis for the method of inversion in electrostatics.) In terms of this operator the match-up conditions (2.26) take the form[13]

$$\begin{aligned} \chi &= e^{\lambda_k}\mathcal{J}_k[\psi], \\ \psi &= e^{-\lambda_k}\mathcal{J}_k[\chi]. \end{aligned} \tag{3.3}$$

Consider now the operators

$$\mathcal{J}^+ = \sum_{\{k_j\}}{}' \left\{ \prod_{j=1}^{n} \exp\,[(-1)^j \lambda_{k_j}] \mathcal{J}_{k_j} \right\}, \tag{3.4a}$$

$$\mathcal{J}^- = \sum_{\{k_j\}}{}' \left\{ \prod_{j=1}^{n} \exp\,[(-1)^{j+1} \lambda_{k_j}] \mathcal{J}_{k_j} \right\}, \tag{3.4b}$$

where each index k_j can take on any of the values $1, \cdots N$, and where the sum extends over all finite sequences $\{k_1, k_2, \cdots k_n\}$ of length $n = 0, 1, 2, \cdots$, subject to the restriction

$$k_{j+1} \neq k_j.$$

These operators satisfy the identities

$$\mathcal{J}^+ = e^{+\lambda_k}\mathcal{J}_k\mathcal{J}^-, \qquad \mathcal{J}^- = e^{-\lambda_k}\mathcal{J}_k\mathcal{J}^+. \tag{3.5}$$

Indeed,

$$\begin{aligned} e^{\lambda_k}\mathcal{J}_k\mathcal{J}^- &= \sum{}' \left\{ \exp\,(\lambda_k)\mathcal{J}_k \prod_{j=1}^{n} \exp\,[(-1)^{j+1}\lambda_{k_j}]\mathcal{J}_{k_j} \right\} \\ &= \sum_{\{k_j'\}}{}' \left\{ \prod_{j'=1}^{n} \exp\,[(-1)^{j'}\lambda_{k_{j'}}]\mathcal{J}_{k_{j'}} \right\}, \end{aligned} \tag{3.6}$$

where

$$\{k_j'\} = \begin{cases} \{k, k_1, \cdots, k_n\} & \text{if} \quad k \neq k_1 \\ \{k_2, \cdots, k_n\} & \text{if} \quad k = k_1 \end{cases}.$$

Since $k_2 \neq k_1$ by hypothesis, each sequence $\{k_1, \cdots, k_n\}$ in the original set occurs once and only once among the sequences $\{k_j'\}$, and thus the right-hand side of (3.6) coincides with $\mathcal{J}^+$. The proof of the second identity follows similarly.

Let us now examine the functions χ and ψ defined formally by the series

$$\chi = \mathcal{J}^+[1], \qquad \psi = \mathcal{J}^-[1]. \tag{3.7}$$

It follows immediately from (3.5) that these functions (if they exist) satisfy the match-up conditions (3.3). Furthermore, each term in the series is a harmonic function which is analytic in D; indeed, a short calculation shows that

$$\begin{aligned} \chi &= 1 + \sum_P \alpha_P\, |\mathbf{x} - \mathbf{x}_P|^{-1}, \\ \psi &= 1 + \sum_P \beta_P\, |\mathbf{x} - \mathbf{x}_P|^{-1}, \end{aligned} \tag{3.8}$$

where the sum is extended over all points $\mathbf{x}_P \in G$ [i.e., over all points of the form

$$\mathbf{C}_k,\ J_{k'}(\mathbf{C}_k),\ J_{k'}J_{k'}(\mathbf{C}_k),\ \cdots],$$

and where the coefficients α_P and β_P are all greater than zero. According to Harnack's Theorem,[14] if either series converges at a single point in D, it converges everywhere in any closed subset of D to a function which is analytic in D. Furthermore the functions χ and ψ so obtained are positive-definite and approach unity as $\mathbf{x}$ approaches infinity, so that they satisfy the boundary conditions (2.3) as well.

Since D is open, the convergence of (3.8) is equivalent to the convergence of the series $\Sigma\alpha_P$ and $\Sigma\beta_P$. One readily checks that

$$\mathcal{J}_k[1] = a_k\, |\mathbf{x} - \mathbf{C}_k|^{-1}, \tag{3.9a}$$

while

$$\begin{aligned} &\mathcal{J}_k[\alpha_j\, |\mathbf{x} - \mathbf{x}_j|^{-1}] \\ &\quad = [\alpha_k\alpha_j\, |\mathbf{C}_k - \mathbf{x}_j|^{-1}]\, |\mathbf{x} - J_k\mathbf{x}_j|^{-1}. \end{aligned} \tag{3.9b}$$

Since $\mathbf{x}_j$ is itself the image of a pole $\mathbf{x}_i$ under the map J_i, with $j \neq k$, $\mathbf{x}_j$ lies inside the sphere S_i, and consequently,

$$|\mathbf{C}_k - \mathbf{x}_j| > |\mathbf{C}_k - \mathbf{C}_j| - a_j \geq d,$$

where

$$d = \min_{\substack{k,k' \\ (k' \neq k)}} [|\mathbf{C}_k - \mathbf{C}_{k'}| - a_{k'}]. \tag{3.10a}$$

Hence the residues α_P, β_P associated with the sequence $\{k_1, \cdots, k_n\}$ are bounded from above by

$$\exp\,[|\lambda_{k_1}| + \cdots + |\lambda_{k_n}|]a_{k_1} \cdots a_{k_n}/d^{n-1} \leq e^{n\lambda}a^n/d^{n-1},$$

where

$$a = \max a_k, \qquad \lambda = \max |\lambda_k|. \tag{3.10b}$$

[13] In the uncharged case, where $\lambda_k = 0$ and $\chi = \psi$, this reduces to the single equation $\chi = \mathcal{J}[\chi]$ derived in MII.

[14] See, e.g., O. D. Kellogg, *Foundations of Potential Theory* (Frederick Ungar Publishing Company, New York, 1929), p. 263.

But there are $N(N-1)^{n-1}$ distinct sequences of length n, and therefore

$$\sum \binom{\alpha_P}{\beta_P} < Nae^{\lambda} \sum_n [(N-1)e^{\lambda}(a/d)]^{n-1},$$

which converges whenever the expression inside brackets is less than unity. A similar argument shows that

$$\sum \binom{\alpha_P}{\beta_P} > Na'e^{-\lambda} \sum_n [(N-1)e^{-\lambda}(a'/d')]^{n-1}$$

with

$$a' = \min a_k$$

and

$$d' = \max_{k,k'} [|\mathbf{C}_k - \mathbf{C}_{k'}| + a_{k'}]. \qquad (3.10c)$$

We have thus proved

Lemma 2: If

$$(N-1)e^{\lambda}(a/d) < 1, \qquad (3.11a)$$

the series defined by $\mathcal{J}^{*}[1]$ *converge in every closed subset of* D *to a pair of functions,* χ *and* ψ*, which are harmonic in* D *and which satisfy the boundary conditions* (2.3) *and the match-up conditions* (3.3)*; these are therefore the desired metric potentials on* M_N*. Furthermore, the series* $\mathcal{J}^{*}[1]$ *converge only if*

$$(N-1)e^{-\lambda}(a'/d') < 1. \qquad (3.11b)$$

In the following section we examine in detail the special case of a manifold containing two identical E–R bridges, and obtain a single convergence criterion for this case [Eq. (4.14)] which is both necessary and sufficient. The two separate conditions given in the above lemma are the best we have been able to deduce for the general case.

IV. ILLUSTRATION: THE SYMMETRIC TWO-BODY PROBLEM

The formal sums (3.7) or (3.8) admit a simple explicit representation when $N = 2$ and

$$a_1 = a_2 = a; \qquad \lambda_1 = -\lambda_2 = \lambda(>0). \qquad (4.1)$$

The metric potentials for this case describe an Einstein–Rosen manifold containing two identical bridges, in which the electric field through the two throats is equal and opposite; thus they can be used to construct a model of an asymptotically flat universe containing two charged particles, of equal mass and equal but opposite charge.

There are here only two sequences of length n (for $n \geq 1$), viz. $\{1, 2, 1, 2, \cdots\}$ and $\{2, 1, 2, 1, \cdots\}$, so that

$$\mathcal{J}^{+} = 1 + \sum_{n=1}^{\infty} [e^{-n\lambda}\mathcal{J}_1\mathcal{J}_2\mathcal{J}_1 \cdots \text{(n factors)} + e^{n\lambda}\mathcal{J}_2\mathcal{J}_1\mathcal{J}_2 \cdots \text{(n factors)}], \qquad (4.2)$$

and similarly

$$\mathcal{J}^{-} = 1 + \sum_{n=1}^{\infty} [e^{n\lambda}\mathcal{J}_1\mathcal{J}_2\mathcal{J}_1 \cdots + e^{-n\lambda}\mathcal{J}_2\mathcal{J}_1\mathcal{J}_2 \cdots].$$

We choose the origin of a cartesian coordinate system at the midpoint between the two spheres S_1 and S_2, and let $\mathbf{d}_n$ be the coordinates of the nth image point inside S_2, with $d_n = |\mathbf{d}_n|$. It follows from symmetry that there is a corresponding image at the point $-\mathbf{d}_n$ inside S_1. (In particular, $\mathbf{C}_2 = \mathbf{d}_1$, $\mathbf{C}_1 = -\mathbf{d}_1$, while $2d_1 = |\mathbf{C}_2 - \mathbf{C}_1|$ is the distance between the two spheres.) The terms $\mathcal{J}_2\mathcal{J}_1 \cdots$ (n factors) [1] and $\mathcal{J}_1\mathcal{J}_2 \cdots$ [1] have simple poles at the image points $\mathbf{d}_n$ and $-\mathbf{d}_n$ respectively, *with the same coefficient* (again because of symmetry), and accordingly

$$\chi = 1 + \sum_{n=1}^{\infty} c_n[e^{-n\lambda} |\mathbf{x} + \mathbf{d}_n|^{-1} + e^{n\lambda} |\mathbf{x} - \mathbf{d}_n|^{-1}],$$
$$\psi = 1 + \sum_{n=1}^{\infty} c_n[e^{n\lambda} |\mathbf{x} + \mathbf{d}_n|^{-1} + e^{-n\lambda} |\mathbf{x} - \mathbf{d}_n|^{-1}]. \qquad (4.3)$$

Now $\mathbf{d}_n$ is the image (in S_2) of $-\mathbf{d}_{n-1}$; i.e.,

$$d_n = d_1 - a^2/(d_1 + d_{n-1}) \qquad (n > 1), \qquad (4.4)$$

while from Eq. (3.9),

$$c_1 = a,$$
$$c_n = [a/(d_1 + d_{n-1})]c_{n-1} \qquad (n > 1). \qquad (4.5)$$

To solve the recursion relations (4.4) and (4.5), introduce a new pair of parameters, c and μ_0, related to the original set (a, d_1) by

$$a = c \operatorname{csch} \mu_0, \qquad (4.6a)$$
$$d_1 = c \coth \mu_0. \qquad (4.6b)$$

It is then readily verified by induction that

$$d_n = c \coth n\mu_0, \qquad (4.7a)$$

and

$$c_n = c \operatorname{csch} n\mu_0. \qquad (4.7b)$$

The resulting expressions for the metric potentials assume their neatest form if one transforms to bispherical coordinates[15]

[15] See, e.g., H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand, Inc., New York, 1943), p. 182.

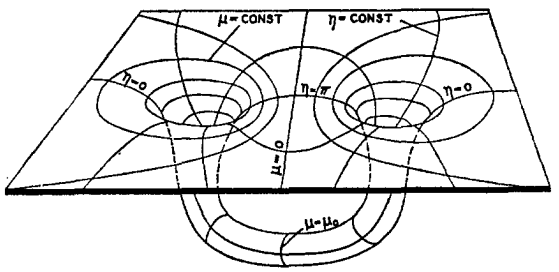

FIG. 4. A two-dimensional section of the "wormhole" manifold W is shown imbedded in flat 3-space. At large distances, the coordinate lines μ = const and η = const become arcs of circles.

$$\coth \mu = (x^2 + y^2 + z^2 + c^2)/(2cz),$$
$$\cot \eta = (x^2 + y^2 + z^2 - c^2)/[2c(x^2 + y^2)^{\frac{1}{2}}], \quad (4.8)$$
$$\cot \varphi = x/y.$$

Choosing the z-axis to point along the line of symmetry joining the two spheres, one computes

$$c_n |\mathbf{x} \pm \mathbf{d}_n|^{-1} = [\cosh \mu - \cos \eta]^{\frac{1}{2}} \times [\cosh (\mu \pm 2n\mu_0) - \cos \eta]^{-\frac{1}{2}}, \quad (4.9)$$

and therefore

$$\begin{pmatrix} \psi \\ \chi \end{pmatrix} = [\cosh \mu - \cos \eta]^{\frac{1}{2}} \times \sum_{n=-\infty}^{\infty} e^{\pm n\lambda} [\cosh (\mu + 2n\mu_0) - \cos \eta]^{-\frac{1}{2}}. \quad (4.10)$$

In these coordinates the flat-space metric $ds_\flat^2$ takes the form

$$ds_\flat^2 = [c/(\cosh \mu - \cos \eta)]^2 \times [d\mu^2 + d\eta^2 + \sin^2 \eta \, d\varphi^2]. \quad (4.11)$$

It is convenient to absorb the factor $[c/(\cosh \mu - \cos \eta)]^2$ into the definitions of χ and ψ, by defining a new pair of metric potentials

$$\begin{pmatrix} \bar{\psi} \\ \bar{\chi} \end{pmatrix} = c^{\frac{1}{2}} \sum_{n=-\infty}^{\infty} e^{\pm n\lambda} [\cosh (\mu + 2n\mu_0) - \cos \eta]^{-\frac{1}{2}}. \quad (4.12)$$

Then Eqs. (2.1) and (2.2) become

$$ds^2 = (\bar{\chi}\bar{\psi})^2 (d\mu^2 + d\eta^2 + \sin^2 \eta \, d\varphi^2), \quad (4.13a)$$
$$E_i = [\ln (\bar{\chi}/\bar{\psi})]_{,i}. \quad (4.13b)$$

Lemma 2a: The infinite series (4.12) *converge if and only if*

$$|\lambda| < \mu_0. \quad (4.14)$$

The proof of this lemma is an elementary application of the ratio test, and is therefore omitted. (When $|\lambda| = \mu_0$, both series diverge, since the nth term does not approach zero.) Note that this single condition replaces both parts of Lemma 2, and gives a much stronger result in this case.

The Wormhole Manifold

In bispherical coordinates the equations for the two throats S_1 and S_2 become simply

$$\mu = -\mu_0 \quad \text{and} \quad \mu = +\mu_0.$$

Furthermore,

$$\bar{\chi}(\mu + 2\mu_0, \eta) = e^{\lambda} \bar{\chi}(\mu, \eta);$$
$$\bar{\psi}(\mu + 2\mu_0, \eta) = e^{-\lambda} \bar{\psi}(\mu, \eta),$$

and consequently

$$ds^2(\mu + 2\mu_0, \eta) = ds^2(\mu, \eta), \quad \mathbf{E}(\mu + 2\mu_0, \eta) = \mathbf{E}(\mu, \eta). \quad (4.15)$$

This result allows one to interpret (4.13) as the metric tensor and electric field on a manifold W, which is obtained from E^3 by cutting out the two spheres $\mu = \pm\mu_0$, throwing away their interiors and identifying corresponding points on the boundaries. This manifold—called a "wormhole" by Misner and Wheeler[2]—has only a single asymptotic region (one "sheet"). Like the E–R manifold, it is multiply connected; electric lines of flux can be trapped in the wormhole, thereby giving the illusion of two "particles" of equal and opposite charge, even though the manifold is everywhere free of singularities (Fig. 4).

Note that the Einstein–Rosen manifold M_2 is an isometric double covering of W. (W may be obtained from M_2 by identifying corresponding points on the two sheets Δ^+ and Δ^-, while at the same time identifying corresponding points on S_1 and S_2.) The direction of the electric field through the two throats on M_2—down in one case, up in the other—assures that after this identification the electric field will still be continuous. Thus it is not surprising that the potentials (4.12), derived originally for M_2, apply to W as well.

V. PHYSICAL CHARACTERISTICS

In this section, formulas for some important physical properties of the Einstein–Rosen manifold are obtained. In particular, it is shown that one can associate a "renormalized" mass and charge with each "particle" (i.e., with each E–R bridge), as well as a total mass-energy and charge with the entire system of N interacting particles. While the total charge is just the algebraic sum of the N individual charges, the same does not apply to the total mass; one finds an additional (negative)

contribution to the total mass, which can be identified with the gravitational and electrostatic interaction energy. Calculations are carried through in detail for the symmetric two-body problem discussed in the preceding section.

Mass and Charge

Formulas for the mass and charge associated with metric potentials of the form (3.8) are given in BL.[16] Thus, to determine the mass of the system, one picks off the coefficient of $1/r$ in an expansion of ds^2 in inverse powers of r; the result is

$$m_{\text{tot}} = \sum_{P\in G} (\alpha_P + \beta_P). \tag{5.1}$$

For the particular case of the symmetric two-body problem this sum reduces to

$$m_{\text{tot}} = 4c \sum_{n=1}^{\infty} \cosh n\lambda \operatorname{csch} n\mu_0. \tag{5.2}$$

To determine the charge associated with either throat, one computes the electric flux through this throat and divides by 4π:

$$q_i = \left(\frac{1}{4\pi}\right) \int_{S_i} E_i n^i \, dS$$

$$= \left(\frac{1}{4\pi}\right) \int_{S_i} [\ln (\chi/\psi)]_{,i} n^i \, dS. \tag{5.3}$$

One can evaluate the integral directly, but it is somewhat easier to compute the charge associated with each image point $\mathbf{x}_P$ inside S_i and then add up the contributions from all such images to get the total charge. Using Eq. (16) of BL and letting G_i denote the points of G inside S_i, one finds

$$q_i = \sum_{P\in G_i} \left[\beta_P - \alpha_P - \sum_{P'\in G}{}' \frac{(\alpha_P\beta_{P'} - \alpha_{P'}\beta_P)}{|\mathbf{x}_P - \mathbf{x}_{P'}|} \right]. \tag{5.4}$$

It is clear that the two terms inside the double sum will cancel whenever $P' \in G_i$, so that one can restrict the sum over P' to those points which lie in G but not in G_i. Then $|\mathbf{x}_P - \mathbf{x}_{P'}| > d > 0$, and the convergence of the series for q_i follows from the (assumed) convergence of the series $\Sigma\alpha_P$ and $\Sigma\beta_P$.

In the symmetric two-body case one obtains, using (4.3) and (4.7),

$$q_1 = -q_2 = 2c \sum_{k=1}^{\infty} \sum_{m=0}^{\infty} \sinh (m + k)\lambda \times \operatorname{csch}(m + k)\mu_0. \tag{5.5}$$

Let $n = m + k$; clearly all terms having the same value of n will contribute an equal amount $\sinh n\lambda \operatorname{csch} n\mu_0$ to the sum. These terms are obtained by letting k take on all integer values from 1 to n, while at the same time, m runs from $n - 1$ to 0; consequently there will be just n of them:

$$q_1 = 2c \sum_{n=1}^{\infty} \sinh n\lambda \operatorname{csch} n\mu_0. \tag{5.6}$$

Since the charge associated with each throat is an additive function, one obtains a formula for the total charge of all N particles by summing (5.4) over i. This gives

$$q_{\text{tot}} = \sum_{P\in G} (\beta_P - \alpha_P), \tag{5.7}$$

in analogy with Eq. (5.1) for the total mass.

In order to complete the description, one would like a formula—analogous to (5.4)—for the individual mass m_i to be associated with each Einstein–Rosen bridge. Of course, in the symmetric case one can obviate the problem by assigning half the total mass to each particle. However, this is not really legitimate, for the total mass is *not* an additive function (i.e., $m_{\text{tot}} \neq \Sigma_i m_i$) because of the additional contribution to m_{tot} from the gravitational and electromagnetic interaction energy. One would like to split off this interaction energy in some unambiguous way, leaving a residue which could be identified with the "bare mass" of each throat. BL have shown how to achieve this for the initial-value metric (2.4) on a manifold with $N + 1$ sheets, and it is tempting to try to apply their results to the present case. Using Eq. (13) of BL, one associates a bare mass with each image point $\mathbf{x}_P$ inside S_i, and adds up the contribution from all these images to give

$$m_{0i} = \sum_{P\in G_i} \left[\beta_P + \alpha_P + \sum_{P'\in G}{}' \frac{(\alpha_P\beta_{P'} + \alpha_{P'}\beta_P)}{|\mathbf{x}_P - \mathbf{x}_{P'}|} \right]. \tag{5.8}$$

The double sum must now be taken over all points $P'(\neq P)$ in G, including those in G_i, and since $|\mathbf{x}_P - \mathbf{x}_{P'}|$ is no longer bounded away from zero, one cannot infer the convergence of this series from that of $\Sigma\alpha_P$ and $\Sigma\beta_P$ alone. In fact, the double series diverges,[17] so that this approach fails.

The failure can be traced to the fact that one has not included the interaction energy between the various bare masses which together make up m_i. This interaction energy, which is given by[16]

[16] See in particular Eqs. (14), (16), and (19) of BL.

[17] This is most easily seen in the symmetric two-body case, where one finds explicitly the divergent series $\Sigma_k \Sigma_{m(\neq k)} \cosh (k - m)\lambda \operatorname{csch} |k - m| \mu_0$.

$$\delta m_i = -\sum_{\substack{P,P'\in G_i \\ (P'\neq P)}}{}' \frac{(\alpha_P\beta_{P'} + \alpha_{P'}\beta_P)}{|\mathbf{x}_P - \mathbf{x}_{P'}|}, \tag{5.9}$$

is negatively infinite, and indeed is just sufficient to cancel the (positively) infinite contribution from m_{0i}. We shall borrow some language from quantum field theory, and call

$$m_i = m_{0i} + \delta m_i \tag{5.10}$$

the "renormalized mass" of the ith particle. Clearly

$$m_i = \sum_{P\in G_i}\left[\alpha_P + \beta_P + \sum_{P'}{}' \frac{(\alpha_P\beta_{P'} + \alpha_{P'}\beta_P)}{|\mathbf{x}_P - \mathbf{x}_{P'}|}\right], \tag{5.11}$$

where the sum of P' is now taken over those points which lie in G but not in G_i. Applying the same method used to derive (5.6), one obtains for the symmetric case

$$\begin{aligned} m_1 &= m_2(=m) \\ &= 2c\sum_{n=1}^{\infty} n\cosh n\lambda\ \mathrm{csch}\ n\mu_0, \end{aligned} \tag{5.12}$$

and therefore

$$\begin{aligned} m_{\mathrm{int}} &\equiv m_{\mathrm{tot}} - (m_1 + m_2) \\ &= -4c\sum (n-1)\cosh n\lambda\ \mathrm{csch}\ n\mu_0. \end{aligned} \tag{5.13}$$

Note that, in general, $|q_i/m_i| < 1$—an inequality which also occurs in the one-body (Reissner–Nordström) problem, and which, as is well known, is far from satisfied by the naturally occurring elementary particles.

It is not at all evident that the renormalized mass defined above is a constant of the motion (i.e. that it has the same value at later times), as one would expect of any sensible definition of the term "mass." Indeed, as time goes on, the throat of each E–R bridge collapses—somewhat like the gravitational collapse of a star whose pressure is zero. The "bare particles" approach closer together and their mutal potential energy becomes more negative, leading to a decrease in the value for m_i one would compute from Eq. (5.11). However, this decrease in potential energy is presumably compensated for by an increase in kinetic energy, which is in principle calculable from the geometry on any given hypersurface, and which ought also to be included in the expression for m_i. Such a deep-seated investigation, requiring the analysis of interaction energies on an arbitrary spacelike hypersurface, has not yet been carried through.

Invariant Distance of Separation

On the time-symmetric initial surface, one can give an unambiguous definition of the invariant separation distance between any pair of particles, as half the length of the *shortest closed path* which threads through their corresponding E–R bridges. It is clear that each such curve is a 3-geodesic; the condition of time symmetry implies that it is a 4-geodesic as well.[18]

Consider, by way of illustration, the case of two identical E–R bridges. Expressed in bispherical coordinates, the equations for the minimal geodesic are simply

$$\eta = \pi, \qquad \varphi = 0, \tag{5.14}$$

with μ varying between $-\mu_0$ and $+\mu_0$. Hence the distance of separation in this case[19] is

$$\begin{aligned} L &= \int_{-\mu_0}^{\mu_0} [g_{\mu\mu}(\mu,\pi)]^{\frac{1}{2}}\, d\mu \\ &= \tfrac{1}{2}c\int_{-\mu_0}^{\mu_0}\sum_{m=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}[e^{(k-m)\lambda}\,\mathrm{sech}\,(\tfrac{1}{2}\mu + m\mu_0) \\ &\qquad \times \mathrm{sech}\,(\tfrac{1}{2}\mu + k\mu_0)]\, d\mu. \end{aligned}$$

Each of the series appearing above is uniformly convergent over the range $-\mu_0 \leq \mu \leq \mu_0$. Thus we may interchange the orders of summation and integration, to obtain (after setting $n = k - m$)

$$L = 2c\left[1 + 2\mu_0\sum_{n=1}^{\infty} n\cosh n\lambda\ \mathrm{csch}\ n\mu_0\right] \tag{5.15}$$

The infinite series for L and m_1—in contrast to those for m_{tot} and q_1—can be evaluated in closed form in terms of elliptic functions. One needs the formula[20]

$$\begin{aligned} &\sum_{n=1}^{\infty} n\,\mathrm{csch}\,(n\pi K/K')\cosh\,(n\pi\alpha/K') \\ &= [K'E - \tfrac{1}{2}\pi + k'^2K'K\,\mathrm{sc}^2\,\alpha][K'/(\pi^2K)], \end{aligned} \tag{5.16}$$

where $\mathrm{sc}\ \alpha = \mathrm{sn}\ \alpha/\mathrm{cn}\ \alpha$, and k' is the complete elliptic integral of the first kind for the complementary modulus $k' = (1 - k^2)^{\frac{1}{2}}$. If one chooses the modulus k to satisfy the transcendental equation

$$\mu_0 = \pi K/K', \tag{5.17}$$

[18] One can also define a distance of separation at other times, by choosing the length of the corresponding 3-geodesic lying in the hypersurface t = const. However, the length of this curve will depend on the choice of time coordinate (i.e., it will not in general be a 4-geodesic). Its physical significance, therefore, is questionable.

[19] Note that this is also the length of the shortest closed geodesic which threads through the two mouths of the wormhole manifold.

[20] See H. Hancock, *Lectures on the Theory of Elliptic Functions* (Dover Publications, Inc., New York, 1958), Sec. 252, for an essentially equivalent expression.

Eq. (5.15) reduces to

$$L = (4c/\pi)K'[E + k'^2 K \operatorname{sc}^2 (K'\lambda/\pi)]. \quad (5.18)$$

When $\lambda = 0$, $\operatorname{sc}(K'\lambda/\pi) = 0$, and this expression reduces to one given by Misner.[21]

Deformation of the Throat

When viewed in terms of the Euclidean background metric ds_0^2, the N "throats" at which the two sheets of the manifold M_N are joined together appear as spheres. To an observer performing measurements with the initial-value metric ds^2, however, these throats deviate from spherical symmetry, being elongated slightly along the lines connecting them. This effect has been pointed out by Wheeler[22] in a discussion of Misner's initial data for an uncharged wormhole; it demonstrates in a vivid way the mutual interaction of the two masses. We shall here obtain the corresponding result for a wormhole (or pair of E–R bridges) containing an electric field.

The initial-value metric (4.13a) induces on the throat $\mu = \mu_0$ a two-dimensional submetric

$$ds^2_{\text{throat}} = [\chi(\mu_0, \eta)\psi(\mu_0, \eta)]^2(d\eta^2 + \sin^2 \eta\, d\varphi^2). \quad (5.19)$$

Introducing for convenience the auxiliary function

$$\xi(\eta) = \sqrt{2} \sum_{n=0}^{\infty} \cosh(n + \tfrac{1}{2})\lambda \times [\cosh(2n + 1)\mu_0 - \cos\eta]^{-\frac{1}{2}}, \quad (5.20)$$

one finds

$$ds^2_{\text{throat}} = 4c^2\xi^4(\eta)(d\eta^2 + \sin^2 \eta\, d\varphi^2). \quad (5.21)$$

To investigate the deviation of the metric (5.21) from spherical symmetry, it suffices to compute the length (i.e., circumference) of two closed geodesics which run around the throat, one in the direction of increasing φ (azimuthal geodesic), the other in the direction of increasing η (polar geodesic). Clearly,

$$C_\eta = 4c \int_0^\pi \xi^2(\eta)\, d\eta, \quad (5.22)$$

while

$$C_\varphi = 4\pi c\xi^2(\eta_0) \sin \eta_0, \quad (5.23a)$$

with η_0 chosen to minimize C_φ:

$$(d/d\eta)[\xi^2 \sin \eta]_{\eta=\eta_0} = 0. \quad (5.23b)$$

Neither of the expressions (5.22) or (5.23a) can be evaluated in closed form, and one must therefore turn to approximation methods. Particularly important is the case $\mu_0 \gg 1$, since this describes two weakly interacting particles (see the following section). If one expands $\xi(\eta)$ in a power series in $e^{-\mu_0}$ and keeps terms through third order, he finds

$$C_\varphi = 16\pi c e^{-\mu_0} \cosh^2(\tfrac{1}{2}\lambda)[1 + 2(2\cosh\lambda - 1)e^{-\mu_0} + 4\cosh\lambda(3\cosh\lambda - 2)e^{-2\mu_0} + (32\cosh^3\lambda - 24\cosh^2\lambda - 10\cosh\lambda + 5)e^{-3\mu_0} + \cdots]. \quad (5.24a)$$

For C_η, one obtains terms identical to those in Eq. (5.24a) through second order; however the third-order term is, instead,

$$C_\eta = 16\pi c e^{-\mu_0} \cosh^2(\tfrac{1}{2}\lambda)[\cdots + (32\cosh^3\lambda - 24\cosh^2\lambda - 7\cosh\lambda + \tfrac{7}{2})e^{-3\mu_0} + \cdots]. \quad (5.24b)$$

Thus,

$$C_\eta - C_\varphi = 16\pi c e^{-\mu_0} \cosh^2(\tfrac{1}{2}\lambda) \times [\tfrac{3}{2}(2\cosh\lambda - 1)e^{-3\mu_0} + \cdots]. \quad (5.25)$$

In terms of the deformation

$$\delta \equiv (C_\eta - C_\varphi)/C_\eta, \quad (5.26)$$

this becomes

$$\delta \approx \tfrac{3}{2}e^{-3\mu_0}(2\cosh\lambda - 1), \quad (5.27)$$

This expression is clearly always positive,[23] showing that the throat is deformed from a spherical shape to that of a prolate spheriod, with its long axis along the line of centers joining the two particles.

Approximate Formulas

The infinite series for m, m_{tot}, q and L [Eqs. (5.12), (5.2), (5.6), and (5.15)] can be evaluated approximately in the limiting case $\mu_0 \gg 1$,[24] thereby providing simple closed expression for these constants which are much easier to interpret. One sets

$$\operatorname{csch} n\mu_0 \approx 2e^{-n\mu_0}, \quad (5.28)$$

and then evaluates the resulting geometric series using standard formulas. For example, from Eq. (5.2),

$$m_{\text{tot}} \approx 4c \sum_{n=1}^{\infty} (e^{n\lambda} + e^{-n\lambda})e^{-n\mu_0} = 4c(\cosh\lambda - e^{-\mu_0})(\cosh\mu_0 - \cosh\lambda)^{-1},$$

[21] Equation (11) of MI.

[22] J. A. Wheeler, Rev. Mod. Phys. **33**, 63 (1961).

[23] A similar calculation in the other limiting case, $\mu_0 \ll 1$, leads to a very different expression for δ, which is also positive definite.

[24] These series can also be summed approximately when $\mu_0 \lesssim 1$; the corresponding results for this case are given in a thesis by the author (Princeton University Press, Princeton, New Jersey, 1962).

which, to first order, reduces to

$$m_{\text{tot}} \approx 4c \cosh\lambda(\cosh\mu_0 - \cosh\lambda)^{-1}. \quad (5.29)$$

Similarly,

$$m \approx ce^{\mu_0} \cosh\lambda(\cosh\mu_0 - \cosh\lambda)^{-2}, \quad (5.30)$$

$$q \approx ce^{\mu_0} \sinh\lambda(\cosh\mu_0 - \cosh\lambda)^{-2}, \quad (5.31)$$

m and q being, respectively, the renormalized mass and charge of particle 1, while

$$m_{\text{int}} \approx -2c(2\cosh^2\lambda - 1) \times (\cosh\mu_0 - \cosh\lambda)^{-2}, \quad (5.32)$$

and

$$L \approx 2c[1 + \mu_0 e^{\mu_0} \cosh\lambda \times (\cosh\mu_0 - \cosh\lambda)^{-2}]. \quad (5.33)$$

These formulas are easily inverted to give μ_0 and λ in terms of the physical parameters m, q, and L:

$$\tanh\lambda \approx q/m, \quad (5.34)$$

and

$$e^{-\mu_0} \approx \tfrac{1}{2}(m/L)\,\text{sech}\,\lambda \approx \tfrac{1}{2}(m/L)[1 - \tfrac{1}{2}(q/m)^2]. \quad (5.35)$$

This last expression is valid when $|q| \ll m \ll L$ (or equivalently, when $\mu_0 \gg 1$ and $|\lambda| \ll \mu_0$). This limiting case may be termed the "quasi-Newtonian" approximation; here the metric potentials (4.3) reduce to those for a pair of well-separated Reissner–Nordström particles, and several of the formulas of this section reduce to ones familiar from classical physics. Thus the interaction energy [Eq. (5.13)] becomes approximately

$$m_{\text{int}} \approx (q_1 q_2/L) - (m_1 m_2/L), \quad (5.36)$$

while the expression for the deformation of the two throats [Eq. (5.27)] simplifies to

$$\delta \approx \tfrac{3}{16}(m/L)^3[1 - \tfrac{1}{2}(q/m)^2]. \quad (5.37)$$

Note that in this limit the deformation is proportional to the inverse cube of the separation distance, as one would expect by analogy with tidal forces.[22]

ACKNOWLEDGMENTS

I am grateful to Professors J. A. Wheeler, C. W. Misner, and D. R. Brill for many valuable and stimulating discussions.

Scattering of Light by Films Having Nonrandom Orientation Fluctuations*

R. S. STEIN, P. F. ERHARDT,† S. B. CLOUGH,‡ AND G. ADAMS§
Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts

A theory is developed for the scattering of light by a two-dimensional solid having correlated fluctuations in density and in the optic axis orientation of anisotropic scattering elements. The case of nonrandom orientation correlations is considered where the probability of a given angle between two optic axes depends upon the separation of the scattering elements r and the angle β which the optic axis makes with the interconnecting vector. It is shown that in addition to the density correlation function $\gamma(r)$, three correlation functions $T_0(r)$, $[T_2(r)-S_2(r)]$, and $[T_4(r)-S_4(r)]$ are usually sufficient to describe the orientation correlations. The values of the latter two characterize the kind and amount of dependence of the scattered intensity upon the polarization directions ψ_1 of the incident and ψ_2 of the scattered light, while the dependence of these functions upon r characterizes the dependence of scattered intensity upon the angle θ between the incident and scattered beam.

PREVIOUS analyses of the scattering of light by films of crystalline polymers such as polyethylene have demonstrated that the scattering arises in part from fluctuations in the average refractive index, and in part from fluctuations in the orientation and anisotropy of anisotropic entities.[1–3] The theory developed was based upon the so-called "random orientation fluctuation" model in which it is assumed that the probability of having two optic axes with correlated orientation depends only upon the separation of the axes and is independent of the angle which these axes make with the separation vector. This gives rise to correlated regions which are, on the average, spherical in shape. It is quite possible, however, that such regions can be extended in space to give nonrandom correlations (Fig. 1)

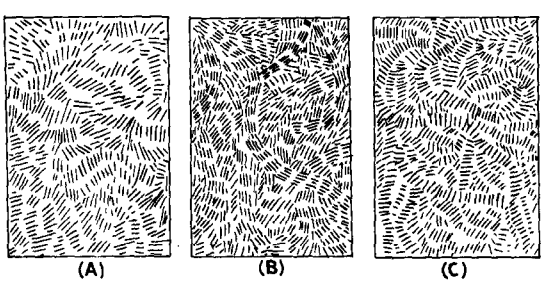

FIG. 1. (A) Random, (B) fiber, and (C) disk-like correlation in orientation.

in which correlation may depend upon the angle β between one of the axes and the vector $\mathbf{r}_{ij}$ (Fig. 2). Fiber or rod-like correlations are those for which correlation is greater for values of β close to 0° while disk-like correlation is where correlation is greater for β close to 90°.

An example of a system exhibiting nonrandom orientation correlations is seen in the photomicrograph of Fig. 3 obtained by George Adams of our laboratory for a recrystallized film of a polychlorotrifluorethylene copolymer (supplied by Allied Chemical Company). This is viewed between crossed polars with the polarization directions vertical and horizontal. Under these conditions, crystals should best transmit light when their optic axes are at 45° to the axes of the polars. Close examination of the photograph reveals that such crystals are not randomly arranged but tend to lie with linear correlation of their centers along lines parallel to the polars and sometimes as long as 20 μ. The phenomonon is not a consequence of macroscopic sample orientation since the appearance is independent of rotation of the sample between the polars.

The interpretation is that crystals tend to be best correlated in orientation to each other when their optic axes are oriented at 45° to the vector connecting the crystals. As illustrated in Fig. 4, this type of orientation would not be rod- or disk-like and would be describable only in more general terms.

A theory for nonrandom orientation correlations has

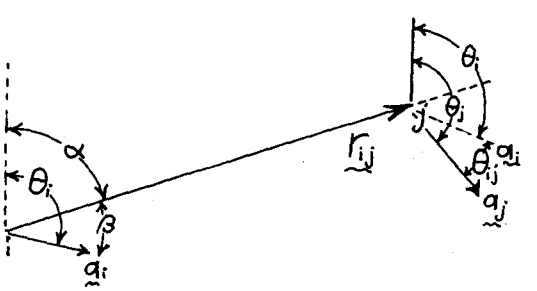

FIG. 2. The angles α, β, and θ_{ij} for two scattering elements with their optic axes oriented at angles θ_i and θ_j and separated by $\mathbf{r}_{ij}$ in two dimensions.

* Partly supported by a contract with the Office of Naval Research and in part by grants from the Army Research Office (Durham) and the Petroleum Research Fund. This was presented in part at the Second Interdisciplinary Conference on Electromagnetic Scattering, Amherst, Massachusetts, June, 1965.

† On leave from General Electric Research Laboratory. Present address: Chemical Development Operations, General Electric Company, Pittsfield, Mass.

‡ Partly supported by an NSF Cooperative Fellowship. Present address: Army Natick Laboratory, Natick, Massachusetts.

§ Partly supported by a fellowship from the General Chemistry Division of Allied Chemical Company. Present address: E. I. du Pont de Nemours, Buffalo, N. Y.

[1] R. S. Stein and R. P. Wilson, J. Appl. Phys. 33, 1914 (1962).

[2] R. S. Stein, in *Proceedings of the Interdisciplinary Conference on Electromagnetic Scattering*, M. Kerker, Ed. (Pergamon Press, New York, 1963), pp. 430–458.

[3] R. S. Stein, J. J. Keane, F. H. Norris, F. A. Bettelheim, and P. R. Wilson, Ann. N. Y. Acad. Sci. 83, 37 (1959).

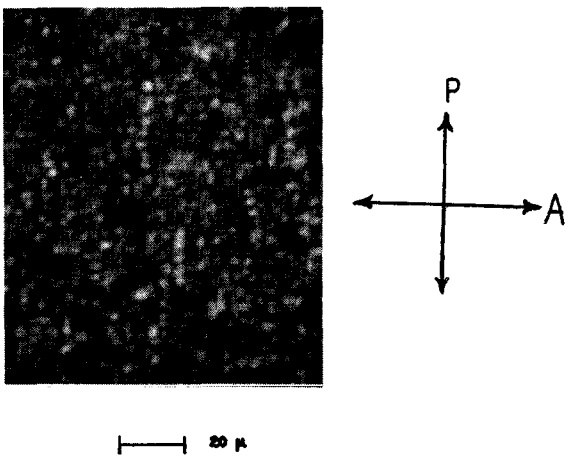

FIG. 3. A photomicrograph of an unoriented film of a polychlorotrifluoroethylene (PTFE) copolymer viewed between vertical polarizer (P) and horizontal analyzer (A).

been proposed,[4,5] which involves an interpolation between the scattering from rod-like structures (characterized by a nonrandomness parameter $\epsilon=1$) and disk-like structures ($\epsilon=-1$). Random fluctuations correspond to $\epsilon=0$. Such a theory would not be adequate for describing the scattering from systems of the sort shown in Figs. 3 and 4 and is not capable of accounting for the dependence of scattering upon azimuthal angle as illustrated by the low-angle scattering photograph for this sample shown in Fig. 5 (for a sample between horizontal and vertical polars using a gas laser light source).

Even higher-order type of nonrandomness of orientation may exist in structures such as spherulites where,

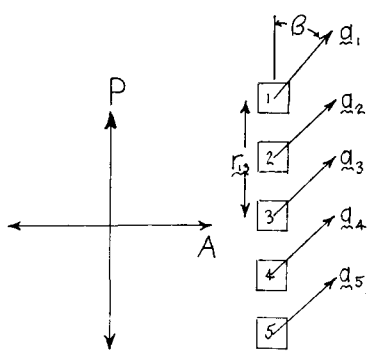

FIG. 4. The type of orientation correlation required to account for the observations of Fig. 3. Two crystals are most likely to have their optic axes $\mathbf{a}_i$ and $\mathbf{a}_j$ parallel if their optic axes are at $\beta=45°$ to the vector $\mathbf{r}_{ij}$ interconnecting them. Maximum transmission occurs for those crystals oriented with their optic axes at 45° to the polarizer and analyzer.

[4] R. S. Stein, P. Erhardt, J. J. van Aartsen, S. Clough, and M. B. Rhodes J. Polymer Sci., Pt. C, No. 18, 1 (1966).

[5] R. S. Stein, P. Erhardt, and G. Adams, ONR Tech. Rept. No. 74, Project NR356-378, Contract: Nonr 3357(01), Polymer Research Institute, University of Massachusetts, Amherst, Mass., July 1965.

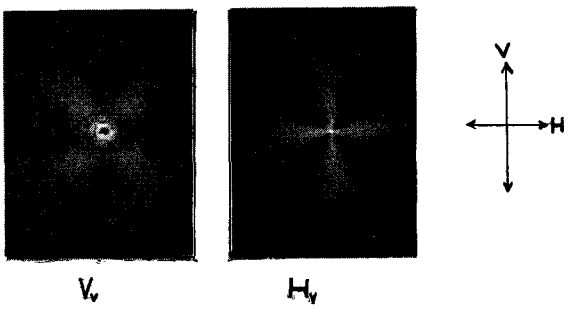

FIG. 5. An experimental V_V and H_V low-angle light scattering pattern for polytetrafluoroethylene (PTFE).

as shown in Fig. 6, the optic axes of two volume elements tend to lie parallel to each other with rod-like order if they are close together or far apart but may tend to be perpendicular to each other with a tendency for 45° orientation to the vector at intermediate distances.

An experimental test for randomness of the orientation correlations may be seen from qualitative considerations. The intensity of scattering may be obtained from[1-3]

$$I=K\int_{\mathrm{r}}(\mathbf{M}_i\cdot\mathbf{O})(\mathbf{M}_j\cdot\mathbf{O})\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_{ij}, \qquad (1)$$

where $\mathbf{M}_i$ and $\mathbf{M}_j$ are the induced dipole moments of scattering elements i and j separated by distance $\mathbf{r}_{ij}$. $k=2\pi/\lambda$, $\mathbf{s}=\mathbf{s}'-\mathbf{s}_0$ where $\mathbf{s}'$ and $\mathbf{s}_0$ are unit-scattered and incident ray propagation vectors. $\mathbf{O}$ is a unit vector perpendicular to the scattered ray and in the direction of the analyzing polaroid.

$\mathbf{M}_i$ is given by

$$\mathbf{M}_i=|\alpha_i|\,\mathbf{E}=\delta_i(\mathbf{E}\cdot\mathbf{a}_i)\mathbf{a}_i-(\alpha_i-\tfrac{1}{3}\delta_i)\mathbf{E}, \qquad (2)$$

where $|\alpha_i|$ is the polarizability tensor, $\mathbf{E}$ is the incident field strength, $\mathbf{a}_i$ is a unit vector in the direction of the optic axis of the ith scattering element (assumed uniaxial), α_i is its average polarizability, and δ_i its anisotropy (difference between principal polarizabilities).

It is apparent that terms $(\mathbf{M}_i\cdot\mathbf{O})$ will be largest when $\mathbf{M}_i$ and $\mathbf{O}$ lie close to the same direction. If a scattering experiment is done as shown in Fig. 7 where the polarizer

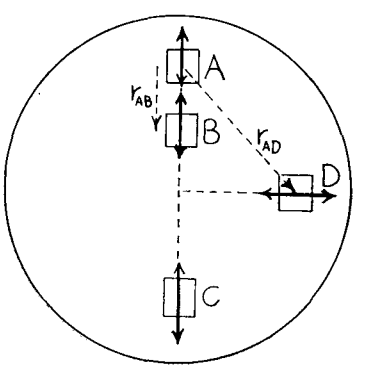

FIG. 6. Nonrandom orientation correlations for a spherulite. A and B as well as A and C have their optic axes correlated in orientation in a fiberlike manner, but A and D tend to be perpendicular to each other if their optic axes are at about 45° to the interconnecting vector.

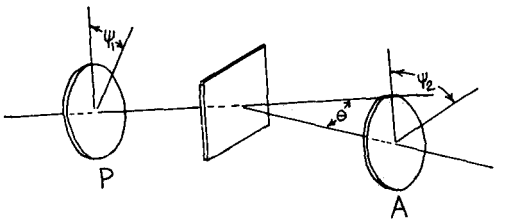

FIG. 7. Scattering experiment with rotating polarizer (P) and analyzer (A) at angles ψ_1 and ψ_2 to the vertical.

in the incident beam is rotated through angle ψ_1 and the analyzer in the scattered beam through angle ψ_2, the scattered intensity at a given scattering angle θ (the reader should not confuse the scattering angle θ in the XY plane with θ_i, the angle between the vector a_i and the vertical axis, as illustrated in Fig. 2) depends upon these polarization angles. Consider the particular cases of $I_{\parallel}$ where $\psi_1=\psi_2$ and $I_{\perp}$ where $\psi_1=90°-\psi_2$. That is, the polarizer and analyzer are rotated, keeping them parallel or perpendicular.

For $I_{\parallel}$, at small values of θ, it may be seen from Eqs. (1) and (2) that the scattered intensity is greatest when $\mathbf{a}_i$ lies close to the direction of polarization. Thus in a polymer with no macroscopic orientation, the scattering occurs mainly from those dipoles oriented at angle $\psi=\psi_1=\psi_2$. If these occur in a region lacking correlation in orientation, then the intensity of light scattered from the strongly scattering cluster of dipoles will be independent of ψ. However, if rod-like orientation correlations occur, a dipole oriented at a particular ψ tends to lie in a cluster which has greatest size along the axis oriented at ψ, and scattering at a particular θ will be greatest in a direction perpendicular to this. If ψ is changed, different clusters of dipoles aligned in the new direction will be the principal contributers to scattering, and the direction of maximum scattering will be perpendicular to this. Therefore, it is apparent that if scattering is measured at small θ's using a photometer in the horizontal plane (Fig. 7), for random orientation correlations, $I_{\parallel}$ will be almost independent of ψ, but for rod-like correlations, the intensity will be greatest at $\psi=0°$, whereas for disk-like correlations, the intensity may be shown to be greatest at $\psi=90°$.[1,4]

For $I_{\perp}$, it follows that the principal contributors to the scattering are those elements with their $\mathbf{a}_i$'s close to 45° to the polarizer and at −45° to the analyzer, and similar conclusions are reached concerning the variation of $I_{\perp}$ with ψ except that the variation is displaced by 45°.

The experimental results for the variation of $I_{\parallel}$ and $I_{\perp}$ for low-density polyethylene are shown in Figs. 8 and 9 where they are compared with the scattering from frosted glass which show no ψ variation.[4] Appreciable ψ variation is seen indicating appreciable nonrandomness. This may be contrasted with the results of Keijzers, van Aartsen, and Prins[6] on shock-cooled polypropylene which shows no ψ variation and has random correlations.

It is possible for preferential correlation to occur for β's closer to other angles than 0° or 90°. For example, for a form of polytetrafluorethlene similar to that of Fig. 3, correlation appears most probable when the optic axis lies at 45° to $\mathbf{r}_{ij}$. Consequently, the variation of $I_{\parallel}$ with ψ for this polymer [Fig. 10] is displaced by 45° from that for polyethylene. The nature of the preferential correlation is dependent upon the details of the crystal-growth mechanism and can lead to complex deviations from randomness as are encountered with spherulite structure.

It is quite likely that the degree and type of nonrandomness of the correlations will depend upon separation of scattering elements in a different manner than does the correlation function itself. A consequence

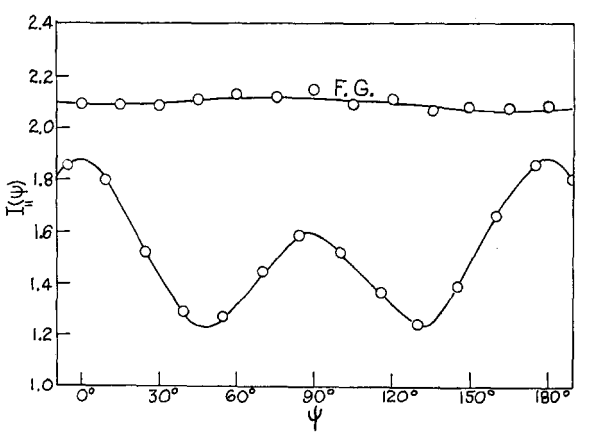

FIG. 8. Variation of $I_{\parallel}$ with ψ for low-density polyethylene at $\theta=15°$. This is compared with the measured variation of the random scattering from frosted glass (F.G.).

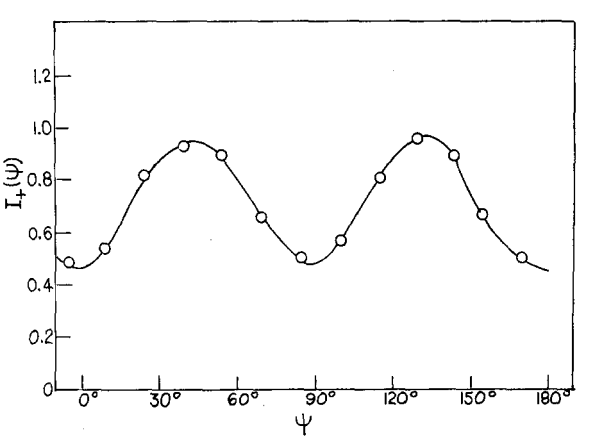

FIG. 9. Variation of $I_{\perp}$ with ψ for low-density polyethylene at $\theta=15°$.

[6] A. E. M. Keijzers, J. J. van Aartsen, and W. Prins, J. Appl. Phys. **36**, 2874 (1965); *Proceedings of the Second Interdisciplinary Conference on Electromagnetic Scattering*, R. L. Rowell and R. S. Stein, Eds. (Gordon & Breach Science Publishers, New York, to be published), and private communications.

of this is that the ψ dependence of scattering will depend upon the scattering angle θ at which it is measured. The variation at smaller values of θ corresponds to the state of randomness at larger values of scattering element separation $\mathbf{r}_{ij}$, so that a complete study requires the determination of the intensity variation upon both θ and ψ. Such will obviously be the situation for the scattering from spherulitic samples as in Fig. 6 or the PTFE film illustrated in Fig. 3.

For films having macroscopic orientation one must, in addition, determine the variation of the scattered intensity upon the angle Ω between the stretching direction and the normal to the plane of measurement of θ (the vertical direction).[7,8]

THEORY

For simplicity the theory[9] is developed in two dimensions, but extension to three dimensions follows the same principles. If the incident light is polarized at angle ψ and travels in the X direction, then (Fig. 11)

$$\mathbf{E}=E_0[(\cos\psi)\mathbf{k}+(\sin\psi)\,\mathbf{j}]. \tag{3}$$

For light scattered at angle θ in the XY plane

$$\mathbf{O}_{\parallel}=-(\sin\psi\sin\theta)\mathbf{i}+(\sin\psi\cos\theta)\,\mathbf{j}+(\cos\psi)\mathbf{k}, \tag{4}$$

$$\mathbf{O}_{\perp}=-(\cos\psi\sin\theta)\mathbf{i}+(\cos\psi\cos\theta)\,\mathbf{j}-(\sin\psi)\mathbf{k}. \tag{5}$$

The sample and the optic axes of the scattering elements are confined to the XZ plane for the two-dimensional problem, so that

$$\mathbf{a}_i=(\cos\theta_i)\mathbf{k}+(\sin\theta_i)\,\mathbf{j}. \tag{6}$$

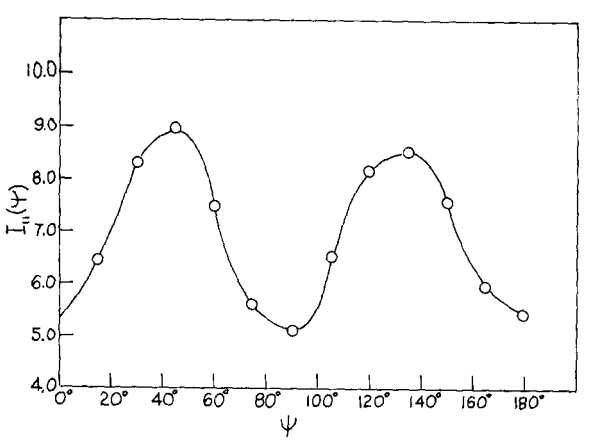

FIG. 10. Variation of $I_{\parallel}$ with ψ for a polytetrafluorethylene film at $\theta=4°$.

[7] F. H. Norris and R. S. Stein, J. Polymer Sci. **27**, 87 (1958).
[8] R. S. Stein and T. Hotta, J. Appl. Phys. **35**, 2237 (1964).
[9] R. S. Stein, P. Erhardt, S. Clough, J. J. van Aartsen, and M. B. Rhodes, in *Proceedings of the Second Interdisciplinary Conference on Electromagnetic Scattering*, R. L. Rowell and R. S. Stein, Eds. (Gordon & Breach Science Publishers, New York, to be published).

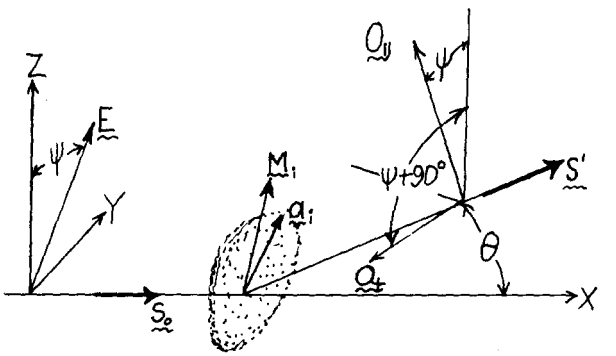

FIG. 11. Vectors **E**, **M**, $\mathbf{a}_i$, $\mathbf{O}_{\parallel}$, and $\mathbf{O}_{\perp}$ defining the scattering measurement.

Combining Eqs. (1) through (6) gives

$$I_{\perp}=KE_0^2\{\delta^2[(K_1+K_2\cos^2\theta-2K_3\cos\theta)\sin^2\psi\cos^2\psi+K_4(\cos^2\psi\cos\theta-\sin^2\psi)^2-2(\cos^2\psi\cos\theta-\sin^2\psi)\sin\psi\cos\psi(K_5-K_6\cos\theta)]+K_7[\sin^2\psi\cos^2\psi(1-\cos\theta)^2]\} \tag{7}$$

and

$$I_{\parallel}=KE_0^2\{\delta^2[K_1\cos^4\psi+K_2\cos^2\theta\sin^4\psi+2K_3\cos\theta\times\sin^2\psi\cos^2\psi+2K_5\cos^3\psi\sin\psi(1+\cos\theta)+2K_6\sin^3\psi\times\cos\psi\cos\theta(1+\cos\theta)+K_4(1+\cos\theta)^2\sin^2\psi\cos^2\psi]+K_7(\cos^2\psi+\cos\theta\sin^2\psi)^2\}, \tag{8}$$

where

$$K_1=\int\int\langle\cos^2\theta_i\cos^2\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \tag{9}$$

$$K_2=\int\int\langle\sin^2\theta_i\sin^2\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \tag{10}$$

$$K_3=\int\int\langle\cos^2\theta_i\sin^2\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \tag{11}$$

$$K_4=\int\int\langle\sin\theta_i\cos\theta_i\sin\theta_j\cos\theta_j\rangle_{r_{ij}}\times\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \tag{12}$$

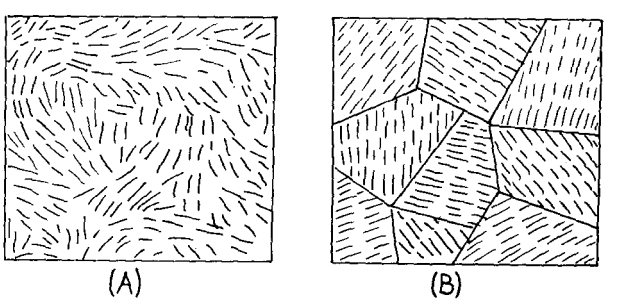

FIG. 12. Comparison of an (A) statistically homogeneous and (B) statistically heterogeneous medium.

$$K_5=\int\int\langle\cos^2\theta_i\sin\theta_j\cos\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \quad (13)$$

$$K=_6\int\int\langle\sin^2\theta_i\sin\theta_j\cos\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \quad (14)$$

$$K_7=\int\int\langle\alpha_i\alpha_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j, \quad (15)$$

where the symbol $\langle\ \rangle r_{ij}$ designates an average at constant $\mathbf{r}_{ij}$. It has been assumed that the anisotropy δ is the same for all scattering elements and that integrals involving cross correlations between fluctuations in density and orientation of the sort

$$\delta\int\langle\alpha_i\cos^2\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_i d\mathbf{r}_j \quad (16)$$

are zero.

At this stage we might introduce the concept of a "statistically homogeneous" medium for which the correlation and statistical distribution of scattering elements appears (on the average) identical from the point of view of any scattering element which might be taken as the origin. This should be distinguished from a "statistically inhomogeneous" medium which contains definite structures having boundaries such as rods or spherulites where the absolute location of a scattering element must be considered. For example, the orientation distribution appears different when viewed from a scattering element in the center of a spherulite as compared with what is seen when viewed from a scattering element toward the outside. Figure 12 compares a statistically homogeneous and heterogeneous medium possessing orientation correlation.

For statistically homogeneous systems, the double integral over $\mathbf{r}_i$ and $\mathbf{r}_j$ in the expressions for the K_i's may be replaced by an integral over $\mathbf{r}_i$ and one over $\mathbf{r}_{ij}$. The results of the integration over $\mathbf{r}_{ij}$ should be independent of $\mathbf{r}_i$. Thus the integration over $\mathbf{r}_i$ just gives the area A of the film (or the volume in three dimensions) so Eq. (9) becomes, for example,

$$K_1=A\int\langle\cos^2\theta_i\cos^2\theta_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_{ij}. \quad (17)$$

Now

$$\langle\cos^2\theta_i\cos^2\theta_j\rangle_{r_{ij}}$$
$$=\tfrac{1}{4}\langle(1+\cos2\theta_i+\cos2\theta_j+\cos2\theta_i\cos2\theta_j\rangle_{r_{ij}}. \quad (18)$$

If, as in Fig. 2,

$$\theta_i=\alpha+\beta, \quad (19)$$

then

$$\langle\cos2\theta_i\rangle_{r_{ij}}=\cos2\alpha\langle\cos2\beta\rangle_{r_{ij}}-\sin2\alpha\langle\sin2\beta\rangle_{r_{ij}}=0, \quad (20)$$

since if one averages over all positions of the ith scattering element, all values of β will be represented equally and

$$\langle\sin2\beta\rangle_{r_{ij}}=\langle\cos2\beta\rangle_{r_{ij}}=0. \quad (21)$$

However, realizing $\theta_j=\theta_i+\theta_{ij}$,

$$\langle\cos2\theta_j\rangle_{r_{ij}}=\langle\cos2\theta_i\cos2\theta_{ij}\rangle_{r_{ij}}-\langle\sin2\theta_i\sin2\theta_{ij}\rangle_{r_{ij}}. \quad (22)$$

Let us now define two correlation functions

$$F_1(\beta,\mathbf{r}_{ij})=\langle\cos2\theta_{ij}\rangle_{r_{ij},\beta} \quad (23)$$

and

$$F_2(\beta,\mathbf{r}_{ij})=\langle\sin2\theta_{ij}\rangle_{r_{ij},\beta}. \quad (24)$$

Then

$$\langle\cos2\theta_j\rangle_{r_{ji}}=\frac{1}{2\pi}\int_{\beta=0}^{2\pi}[\cos(2\alpha+2\beta)F_1$$
$$-\sin(2\alpha+2\beta)F_2]d\beta. \quad (25)$$

The evaluation of these integrals requires a knowledge of how F_1 and F_2 depend on β. These may be generally expressed in a Fourier expansion in two-dimensions (or in spherical harmonics in three) as

$$F_1=T_0+\sum_n[T_n\cos n\beta+Z_n\sin n\beta] \quad (26)$$

and

$$F_2=S_0+\sum_n[R_n\cos n\beta+S_n\sin n\beta]. \quad (27)$$

The Fourier coefficients generally depend upon $\mathbf{r}_{ij}$. However, in the case of systems having no macroscopic orientation, they are independent of the orientation of $\mathbf{r}_{ij}$ specified by the angle α and depend only upon its magnitude. For random orientation fluctuations, F_1 and F_2 are independent of β, and T_0 and S_0 are the only nonvanishing coefficients. In this case, T_0 is identical with the previously defined two-dimensional orientation correlation functions designated as $f(\mathbf{r})$.[1,2]

$$T_0=f(\mathbf{r})=\langle2\cos^2\theta_{ij}-1\rangle_r. \quad (28)$$

It may be useful at this stage to give these correlation functions $F_1(\mathbf{r})$ and $F_2(\mathbf{r})$ some physical interpretation. If there is no correlation in the angle between the optic axes of elements i and j, θ_{ij} assumes all values with equal frequency and both F_1 and F_2 are zero. If there is a strong tendency for the optic axes to be parallel so θ_{ij} tends toward zero, F_1 approaches one while F_2 approaches zero. If the tendency for them to be parallel is greatest when they lie parallel to the interconnecting vector $\mathbf{r}_{ij}$ so that F_1 is large when β is small, terms like T_2 will be larger than the Z_2 terms (corresponding to rod-like correlation), whereas if F_1 is large when β is close to 90° (disk-like correlation), the reverse holds. For more complex correlation, say greater tendency for parallelness when the optic axes are at 45° to β, terms like T_4 are large. Since correlation is unaffected by rotating an optic axis through 180°, T_n and Z_n are nonzero

only for even values of n. If correlation is symmetrical in β [so there is no "handedness" in symmetry and $F_1(\beta)=F_1(-\beta)$], Z_n terms will be zero.

F_2 will be large if correlation is greatest for optic axes lying at 45° to each other (in which case F_1 tends toward zero). Symmetry with respect to rotation through 180° again requires vanishing of terms with odd n, whereas, in this case, lack of "handedness" requires antisymmetry in $\beta[F_2(\beta)=-F_2(-\beta)]$ so that only sine terms in F_2 remain, requiring that the R_n's are zero.

In all cases the coefficients will be decreasing functions of r and approach zero (not necessarily uniformly or at the same rate as each other) as r goes to infinity.

From Eqs. (25), (26), and (27), it follows that

$$\langle\cos2\theta_j\rangle_{r_{ij}}=\tfrac{1}{2}(T_2-S_2)\cos2\alpha-(Z_2+R_2)\sin2\alpha. \quad (29)$$

It is of interest to account for the result that $\langle\cos2\theta_i\rangle_{r_{ij}}=0$ but $\langle\cos2\theta_j\rangle_{r_{ij}}$ does not, since they might be expected to be equal on the basis of symmetry. The difference is a consequence of the order of summation and may be understood upon realizing that in both cases, one is averaging over all positions of element i for a fixed $\mathbf{r}_{ij}$. The nonvanishing $\langle\cos2\theta_j\rangle_{r_{ij}}$ is a consequence of the fact that θ_j will not be random once element i is selected. For a given i, optic axis $\mathbf{a}_i$ will be oriented at some (random) angle β with respect to $\mathbf{r}_{ij}$. The degree of correlation of θ_j with θ_i depends upon both $\mathbf{r}_{ij}$ and β. Therefore, even though θ_i varies randomly, θ_j will assume preferential values if correlation is nonrandom.

The last term of Eq. (18) is evaluated similarly.

$$\langle\cos2\theta_i\cos2\theta\rangle_{r_{ij}}=\tfrac{1}{2}[T_0+\tfrac{1}{2}(T_4-S_4)\cos4\alpha-\tfrac{1}{2}(R_4+Z_4)\sin4\alpha]. \quad (30)$$

Thus

$$K_1=\frac{A\pi}{4}\int_{r=0}^{\infty}\{T_0J_0(w)+[T_2-S_2]J_2(w)+\tfrac{1}{2}[T_4-S_4]J_4(w)\}rdr \quad (31)$$

since $w=kr\sin\theta$ and

$$\int_0^{2\pi}\cos(n\alpha)\cos(w\sin\alpha)d\alpha=2\pi J_n(w), \quad (32)$$

$$\int_0^{2\pi}\sin(n\alpha)\cos(w\sin\alpha)d\alpha=0. \quad (33)$$

By a similar procedure to that for K_1

$$K_2=\frac{A\pi}{4}\int_{r=0}^{\infty}\{T_0J_0(w)-[T_2-S_2]J_2(w)+\tfrac{1}{2}[T_4-S_4]J_4(w)\}rdr. \quad (34)$$

For K_3, since the order of integration over i and j matters, in transforming the double integral to a single one, one obtains

$$K_3=\frac{A}{2}\int[\langle\cos^2\theta_i\sin^2\theta_j\rangle_{r_{ij}}+\langle\cos^2\theta_j\sin^2\theta_i\rangle_{r_{ij}}]\cos(w\sin\alpha)d\mathbf{r}_{ij}. \quad (35)$$

Now

$$\langle\cos^2\theta_i\sin^2\theta_j\rangle_{r_{ij}}=\tfrac{1}{4}\langle1+\cos2\theta_i-\cos2\theta_j-\cos2\theta_i\cos2\theta_j\rangle_{r_{ij}} \quad (36)$$

and

$$\langle\cos^2\theta_j\sin^2\theta_i\rangle_{r_{ij}}=\tfrac{1}{4}\langle1+\cos2\theta_j-\cos2\theta_j-\cos2\theta_i\cos2\theta_j\rangle_{r_{ij}}. \quad (37)$$

Thus

$$K_3=-\frac{A\pi}{4}\int_{r=0}^{\infty}\{T_0J_0(w)+\tfrac{1}{2}[T_4-S_4]J_4(w)\}rdr. \quad (38)$$

Also

$$K_4=\frac{A\pi}{4}\int_{r=0}^{\infty}\{T_0J_0(w)-\tfrac{1}{2}[T_4-S_4]J_4(w)\}rdr. \quad (39)$$

Similarly, considering the two orders of integration,

$$K_5=\frac{A\pi}{4}\int_{r=0}^{\infty}\{[Z_2+R_2]J_2(w)+[Z_4+R_4]J_4(w)\}rdr. \quad (40)$$

Similarly

$$K_6=\frac{A\pi}{4}\int_{r=0}^{\infty}\{[Z_2+R_2]J_2(w)-[Z_4+R_4]J_4(w)\}rdr. \quad (41)$$

The K_7 terms represent the contribution arising from density fluctuations and may be expressed in terms of a density correlation function[1,9] and

$$K_7=\int\int\langle\Delta\alpha_i\Delta\alpha_j\rangle_{r_{ij}}\cos k(\mathbf{r}_{ij}\cdot\mathbf{s})d\mathbf{r}_id\mathbf{r}_j=2\pi\langle(\Delta\alpha_i)^2\rangle A\int_0^{\infty}\gamma(r)J_0(w)rdr. \quad (42)$$

Upon substituting back into Eqs. (7) and (8) one obtains

$$I_\perp=KA\pi\int_0^{\infty}\{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r)\Phi_i+(\delta^2/4)\times[T_0J_0(w)\phi_2+(T_2-S_2)J_2(w)\Phi_3+\tfrac{1}{2}(T_4-S_4)J_4(w)\Phi_4+(Z_2+R_2)J_2(w)\Phi_{12}+(Z_4+R_4)J_4(w)\Phi_{13}]\}rdr, \quad (43)$$

$$I_{\parallel}=KA\pi\int_0^\infty \{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r)\Phi_5 + (\delta^2/4)[T_0J_0(w)\Phi_6+(T_2-S_2)J_2\Phi_7 - \tfrac{1}{2}(T_4-S_4)J_4(w)\Phi_8+(Z_2+R_2)J_2(w)\Phi_9 + (Z_4+R_4)J_4(w)\Phi_{10}]\}rdr, \quad (44)$$

where

$$\Phi_0=\sin^2\psi\cos^2\psi(1+\cos\theta)^2 \quad (45)$$

$$\Phi_1=\sin^2\psi\cos^2\psi(1-\cos\theta)^2 \quad (46)$$

$$\Phi_2=\Phi_0+(\cos^2\psi\cos\theta-\sin^2\psi)^2 \quad (47)$$

$$\Phi_3=\sin^2\theta\sin^2\psi\cos^2\psi \quad (48)$$

$$\Phi_4=\Phi_0-(\cos^2\psi\cos\theta-\sin^2\psi)^2 \quad (49)$$

$$\Phi_5=(\cos^2\psi+\cos\theta\sin^2\psi)^2 \quad (50)$$

$$\Phi_6=\Phi_0+(\cos^2\psi-\cos\theta\sin^2\psi)^2 \quad (51)$$

$$\Phi_7=\cos^4\psi-\sin^4\psi\cos^2\theta \quad (52)$$

$$\Phi_8=\Phi_0-(\cos^2\psi-\cos\theta\sin^2\psi)^2 \quad (53)$$

$$\Phi_9=(\Phi_5)^{\frac{1}{2}}\sin\psi\cos\psi(1+\cos\theta) \quad (54)$$

$$\Phi_{10}=(\cos^2\psi-\sin^2\psi\cos\theta)\sin\psi\cos\psi(1+\cos\theta) \quad (55)$$

$$\Phi_{11}=(\sin^2\psi-\cos^2\psi\cos\theta)\sin\psi\cos\psi \quad (56)$$

$$\Phi_{12}=\Phi_{11}(1-\cos\theta) \quad (57)$$

$$\Phi_{13}=\Phi_{11}(1+\cos\theta). \quad (58)$$

The terms in these equations involving the Z_n's and R_n's are multiplied by $\sin\psi\cos\psi$ which is an odd function of ψ. Consequently these terms would be of different sign for positive and negative values of ψ and would lead to a predicted $I_{\parallel}(\psi)$ variation which would not be symmetrical about $\psi=0$. As previously discussed, such terms are associated with asymmetry or "handedness" of structure in which the probability for correlation in orientation is different for an optic axis oriented at $+\beta$ with respect to $\mathbf{r}_{ij}$ than for $-\beta$. In three dimensions, this could correspond to helicoidal orientation of optic axes where right-handed and left-handed helices are not equally probable. This asymmetry is not found for synthetic polymers (unless made by asymmetric reagents or resolved by an asymmetric fractionation technique) but may be found in natural polymers. Since such structures are unusual and such asymmetric scattering patterns have not so far been reported, the contribution of these terms will be neglected in the rest of this paper.

The ψ dependence of scattering which is predicted by these equations at very small θ is of interest. As θ approaches zero, Φ_1, Φ_3, and Φ_{12} become zero; Φ_2, Φ_5, and Φ_6 approach unity; Φ_7 approaches $\cos2\psi$; Φ_4 and Φ_8 approach $-\cos4\psi$; Φ_9 approaches $\sin2\psi$; Φ_{10} approaches $+\frac{1}{2}\sin4\psi$; and Φ_{13} approaches $-\frac{1}{2}\sin4\psi$. [In this group of equations, the description $\theta=0°$ is intended to imply "small θ" since at actual $\theta=0°$ conditions, $J_2(0)=J_4(0)=0$ and there will be no ψ dependence of intensity.] Thus, neglecting the asymmetric terms,

$$I_{\perp}(\theta=0°)=KA\pi\int_0^\infty \{(\delta^2/4)[T_0J_0(w) - \tfrac{1}{2}(T_4-S_4)J_4(w)\cos4\psi]\}rdr \quad (59)$$

and

$$I_{\parallel}(\theta=0°)=KA\pi\int_0^\infty \{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r) + (\delta^2/4)[T_0J_0(w)+(T_2-S_2)J_2(w)\cos2\psi + \tfrac{1}{2}(T_4-S_4)J_4(w)\cos4\psi]\}rdr. \quad (60)$$

It is seen that the density fluctuations contribute only to the $I_{\parallel}(\theta=0°)$ scattering and do so in a ψ-independent manner. $I_{\perp}(\theta=0°)$ consists of two terms: the one arising from random orientation correlations is ψ independent and characterized by the correlation function $T_0(r)$ [and corresponds to the $f(r)$ term which was the only one in the "$I_{Hv}(\theta=0°)$" equation of the previous theory][1]. The other is multiplied by $\cos4\psi$ and has fourfold symmetry in ψ and corresponds to the "four-leaf clover" symmetry of the experimentally observed low-angle patterns. This is characterized by the coefficient $[T_4(r)-S_4(r)]$ and may add or subtract from the constant term depending upon the sign of the coefficient. Thus the relative amount of ψ variation of the $I_{\perp}$ pattern at low angles characterizes the relative values of $(T_4-S_4)/T_0$.

The $I_{\parallel}(\theta=0°)$ scattering also contains these two contributions plus two more. By subtracting $I_{\perp}(\theta=0°)$ from $I_{\parallel}(\theta=0°)$, one obtains

$$I_{\parallel}(\theta=0°)-I_{\perp}(\theta=0°)=KA\pi\int_0^\infty \{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r) + (\delta^2/4)[(T_2-S_2)J_2(w)\cos2\psi + (T_4-S_4)J_4(w)\cos4\psi]\}rdr \quad (61)$$

in which the random orientation term subtracts out. This equation contains the density fluctuation term characterized by the correlation function $\gamma(r)$, and an additional nonrandom orientation correlation term having twofold symmetry (multiplied by $\cos2\psi$) and characterized by the coefficient $[T_2(r)-S_2(r)]$. One sees that

$$I_{\parallel}(\theta=0°,\psi=0°)-I_{\perp}(\theta=0°,\psi=0°) = KA\pi\int_0^\infty \{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r)-(\delta^2/4) \times[(T_2-S_2)J_2(w)+(T_4-S_4)J_4(w)]\}rdr, \quad (62)$$

$$I_{\parallel}(\theta=0°,\psi=22.5°)-I_{\perp}(\theta=0°,\psi=22.5°)$$

$$=KA\pi\int_0^\infty\{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r)$$

$$-(\delta^2/4)[0.707(T_2-S_2)J_2(w)]\}rdr, \quad (63)$$

$$I_{\parallel}(\theta=0°,\psi=45°)-I_{\perp}(\theta=0°,\psi=45°)$$

$$=KA\pi\int_0^\infty\{2\langle(\Delta\alpha_i)^2\rangle J_0(w)\gamma(r)$$

$$-(\delta^2/4)[(T_4-S_4)J_4(w)]\}rdr. \quad (64)$$

Therefore, measurement at these values of ψ permits one to separate the three terms contributing to $(I_{\parallel}-I_{\perp})$ and then subtracting from $I_{\parallel}$ permits determination of the fourth. In this way the four contributions to scattering may be resolved. Carrying out the analogous procedure as a function of scattering angle θ as is discussed in detail later, provides data which may be inverted to obtain the functions $\gamma(r)$, T_0, (T_2-S_2), and (T_4-S_4) as a function of r. These results together with the values of $\langle(\Delta\alpha_i)^2\rangle$ and δ^2 which could be obtained from absolute intensity measurements then provide the most complete description of the system obtainable by scattering measurements. These functions may have different r dependence which results in a different ψ dependence of scattering at different θ.

It is apparent that $I_{\parallel}$ will have a fourfold symmetry with superposed twofold symmetry of relative amount determined by the ratio of (T_2-S_2) to (T_4-S_4). This is in agreement with the type of results found, for example, with polyethylene exhibited in Fig. 8. The twofold symmetry term may exhibit intensity maxima at either $\psi=0°$ and 180° or at 90° and 270° depending on whether the sign of (T_2-S_2) is positive or negative. The experimentally found maxima at $\psi=0°$ and 180° for a low-density polyethylene sample at $\theta=15°$ in Fig. 8 indicates that (T_2-S_2) makes a positive contribution for this material characteristic of a greater tendency for optic axes to lie parallel in the vicinity of $\beta=0°$ as for rod-like correlation. This is in contrast to the ψ variation found for a medium-density spherulitic polyethylene plotted in Fig. 13 for which the intensity is greatest at $\psi=90°$ and 270°, requiring a negative (T_2-S_2) representing greater correlation toward $\beta=90°$ as with disk-like correlation. Thus, as might be expected, (T_2-S_2) is dependent upon the relationship between r and a characteristic size of the structure and and may be positive for samples of low spherulitic order where the crystals are aggregated in bundle-like miscelles with the optic-axis parallel to the miscelle axis, but may be negative for spherulitic samples containing folded-chain lamellae, with the optic (chain) axis perpendicular to the lamellae axis.

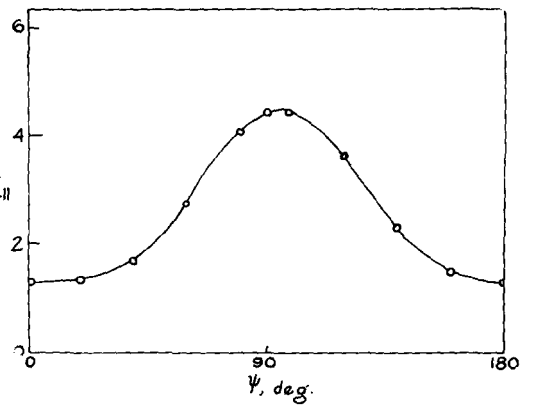

FIG. 13. Variation of $I_{\parallel}$ with ψ for a medium density spherulitic polyethylene at $\theta=15°$.

The fact that polyethylene $I_{\parallel}$ scattering exhibits an appreciable twofold symmetry contribution to its ψ dependence but PTFE does not [Fig. (10)] indicates a smaller (T_2-S_2) contribution for PTFE. Since the (T_2-S_2) term does not appear in the expression for $I_{\perp}$, fourfold symmetry in $I_{\perp}$ is predicted for both polyethylene and PTFE as is observed.

The specific evaluation of scattering requires a knowledge of the r dependence of those coefficients. For purposes of illustration, one may assume that they may be approximated by Gaussian functions

$$T_i=\rho_i\exp[-r^2/a_i^2] \quad (65)$$

and

$$S_i=\sigma_i\exp[-r^2/b_i^2]. \quad (66)$$

In this case, the coefficient a_0 may be interpreted as a correlation distance for random orientation correlations whereas, for example, a_4 and b_4 represent correlation distances for fourfold symmetrical nonrandomness. Generally these are different so that, for example, scattering elements close together may have their optic axes correlated in orientation in a nonrandom manner with rod and disk-like symmetry, but correlation may become more random as the separation increases. This would give rise to decreasing ψ dependence of scattering with decreasing scattering angle θ. Substitution of these functions into Eqs. (43) and (44) gives (neglecting asymmetric terms)

$$I_{\perp}=KA\pi\{\Lambda_1\Phi_1+\Lambda_2\Phi_2+\Lambda_3\Phi_3+\Lambda_4\Phi_4\} \quad (67)$$

$$I_{\parallel}=KA\pi\{\Lambda_1\Phi_5+\Lambda_2\Phi_6+\Lambda_3\Phi_7-\Lambda_4\Phi_8\} \quad (68)$$

where

$$\Lambda_1=2\langle(\Delta\alpha_i)^2\rangle\int_0^\infty J_0(w)\gamma(r)rdr=\langle(\Delta\alpha_i)^2\rangle a_d^2\exp(-W_d^2/4) \quad (69)$$

$$\Lambda_2=(\delta^2/4)\int_0^\infty T_0(r)J_0(w)rdr=(\delta^2/8)\rho_0 a_0^2\exp(-W_0^2/4) \tag{70}$$

$$\Lambda_3=(\delta^2/4)\int_0^\infty (T_2-S_2)J_2(w)rdr=(\delta^2/8)\{\rho_2 a_2^2((4/W_2^2)(1-\exp[-W_2^2/4])-\exp[-W_2^2/4])$$
$$-\sigma_2 b_2^2((4/V_2^2)(1-\exp[-V_2^2/4])-\exp[-V_2^2/4])\} \tag{71}$$

$$\Lambda_4=(\delta^2/8)\int_0^\infty (T_4-S_4)J_4(w)rdr=(\delta^2/8)\{\rho_4 a_4^2((48/W_4^4)(\exp[-W_4^2/4]-1+[W_4^2/4])$$
$$-(8/W_4^2)(1-\exp[-W_4^2/4])+\tfrac{1}{2}\exp[-W_4^2/4])-\sigma_4 b_4^2((48/V_4^4)(\exp[-V_4^2/4]-1+[V_4^2/4])$$
$$-(8/V_4^2)(1-\exp[-V_4^2/4])+\tfrac{1}{2}\exp[-V_4^2/4])\}, \tag{72}$$

where a_d is the Gaussian correlation distance for density fluctuations and

$$W_i=2\pi(a_i/\lambda)\sin\theta \tag{73}$$

$$V_i=2\pi(b_i/\lambda)\sin\theta. \tag{74}$$

More general equations where the ρ_i and σ_i's have been expanded in power series in r have been obtained.[9]

The predictions of Eqs. (67) and (68) are plotted in Figs. 14 through 16 for the case where the density term is zero. The results are given for $\theta=30°$, $a_0=a_2=a_4=b_2=b_4=\lambda$. In Figs. 14(a) and 14(b), $\rho_0=1.0$, $(\rho_2-\sigma_2)=0$, and $(\rho_4-\sigma_4)=0.8$, 0, and -0.8. Positive values of $(\rho_4-\sigma_4)$ are seen to lead to $I_{\parallel}$ maxima at 0° and 90° and $I_\perp$ maxima at 45° and 135°, as found for polyethylene, while negative $(\rho_4-\sigma_4)$ reverses this as for PTFE. The case where $(\rho_4-\sigma_4)=0$ gives only minor variation of $I_{\parallel}$ and $I_\perp$ resulting from the slight ψ dependence of Φ_2 and Φ_6 at these finite θ's.

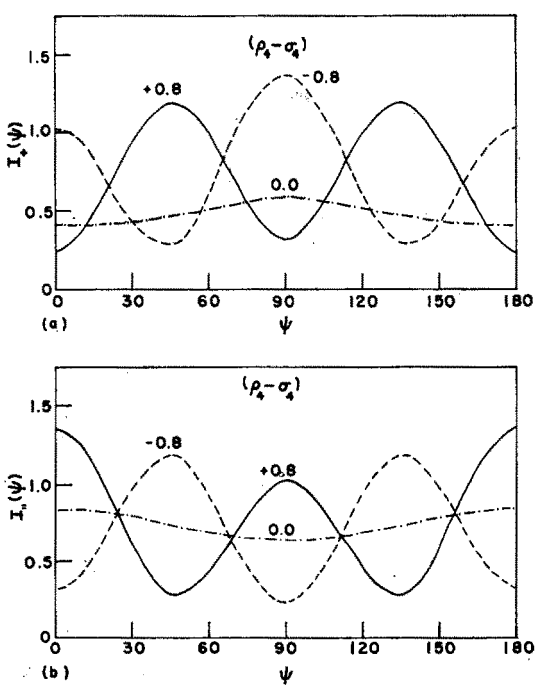

FIG. 14. Predicted variation of (a) $I_\perp$ and (b) $I_{\parallel}$ with ψ for $a_0=b_0=b_4=\lambda$, $\rho_0=1.0$, $(\rho_2-\sigma_2)=0$, and $(\rho_4-\sigma_4)=0.8$, 0, and -0.8.

In Fig. 15 an $I_\perp$ plot is given at $\theta=40°$ and $\theta=50°$ for $\rho_0=0.7$, $(\rho_2=\sigma_2)=0$, $\rho_4=0.2$, $\sigma_4=0.1$, $a_0=a_4=\lambda$, and $b_4=(5)^{\frac{1}{2}}\lambda$. It is seen that the different correlation distances for T_4 and S_4 can lead to a reversal of the ψ dependence of $I_\perp$ with changing θ.

Figure 16 corresponds to the case of $(\rho_4-\sigma_4)=0.8$ of Fig. 14(b) but where $(\rho_2-\sigma_2)$ is allowed to equal 0.4, 0, and -0.4. This is seen to give rise to a twofold variation of $I_{\parallel}(\psi)$ superposed on the fourfold variation corresponding to the differences observed for polyethylene and PTFE. Corresponding plots are not given for $I_\perp$ since these are almost unaffected by $(\rho_2-\sigma_2)$ and would be identical with Fig. 14(a).

The study of the intensity dependence on the angles ψ and θ permits a determination of the density and orientation correlation functions. It may be seen that

$$\frac{\Lambda_1(\theta)+2\Lambda_4(\theta)}{KA\pi}=\frac{1}{2}\left[I_{\parallel}(\theta,0)+\frac{I_{\parallel}(\theta,90)}{\cos^2\theta}\right]-I_\perp(\theta,90) \tag{75}$$

$$\frac{\Lambda_3(\theta)}{KA\pi}=\frac{1}{2}\left[I_{\parallel}(\theta,0)-\frac{I_{\parallel}(\theta,90)}{\cos^2\theta}\right] \tag{76}$$

$$\frac{\Lambda_2(\theta)-\Lambda_4(\theta)}{KA\pi}=I_\perp(\theta,90). \tag{77}$$

The separation of Λ_1, Λ_2, and Λ_4 can be carried out by measurement of the intensity at some other ψ than 0° or 90°, say 45° to give, for example

$$I_\perp(\theta,45°)=\tfrac{1}{4}\{\Lambda_1(\theta)(1-\cos\theta)^2+2\Lambda_2(\theta)(1+\cos^2\theta)$$
$$+\Lambda_3(\theta)\sin^2\theta+4\Lambda_4(\theta)\cos\theta\}. \tag{78}$$

If the value of Λ_3 from (76) is substituted in this and the result is solved simultaneously with Eqs. (75) and (77), values of Λ_1, Λ_2, and Λ_4 are obtained.

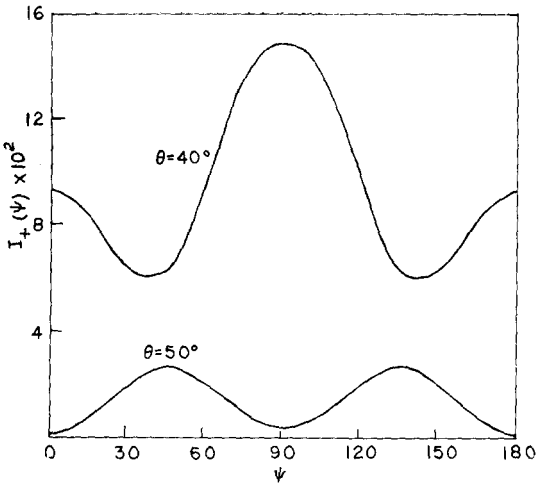

FIG. 15. Predicted variation of $I_\perp$ with ψ for $\theta=40°$ and 50°, where $\rho_2=0.7$, $(\rho_2-\sigma_2)=0.0$, $\rho_4=0.2$, $\sigma_4=0.1$, $a_0=a_4=\lambda$, and $b_4=(5)^{\frac{1}{2}}\lambda$.

Values of the correlation functions may then be obtained by use of the Fourier–Bessel integral[10]

$$f(z)=\int_0^\infty J_m(bz)bdb\int_0^\infty f(\xi)J_m(b\xi)d\xi. \tag{79}$$

If for example $z=\xi=r$, $f(\xi)=[T_2(r)-S_2(r)]r$, $m=2$, $b=l=k\sin\theta$, so $bz=lr=w$.

$$[T_2(r)-S_2(r)]r=\int_0^\infty J_2(lr)ldl\int_0^\infty \times[T_2(r)-S_2(r)]J_2(lr)rdr. \tag{80}$$

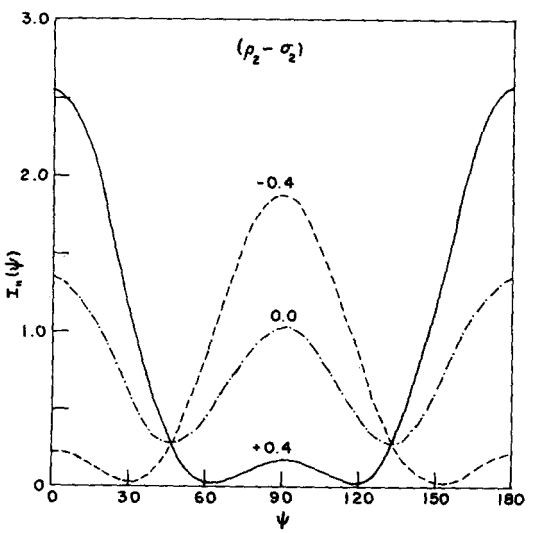

FIG. 16. Variation of $I_\perp$ with ψ for $a_0=a_2=a_4=b_4=\lambda$, $\rho_0=1$, $(\rho_4-\sigma_4)=0.8$, and $(\rho_2-\sigma_2)=0.4$, 0, and -0.4.

[10] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., Inc., New York, 1953), Part I, p. 766.

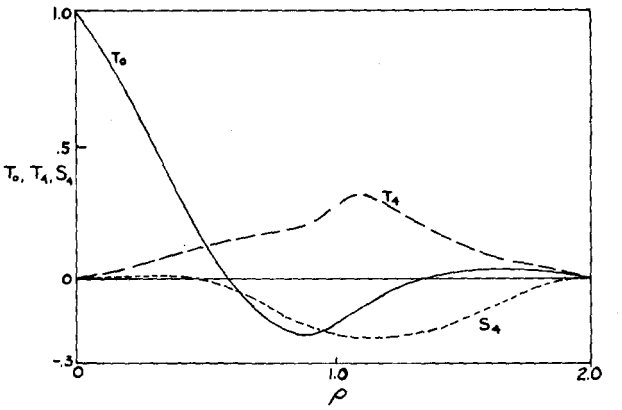

FIG. 17. Calculated variation of $T_0(\rho)$, $T_4(\rho)$, and $S_4(\rho)$ (where $\rho=r/R$) for a two-dimensional spherulite of radius R having radially directed optic axes.

Now from (71)

$$\Lambda_3=(\delta^2/4)\int_0^\infty (T_2-S_2)J_2(lr)rdr \tag{81}$$

so

$$[T_2(r)-S_2(r)](\delta^2/4)=\frac{1}{r}\int_0^\infty \Lambda_3(l)J_2(lr)ldl$$
$$=\frac{1}{r}\int_0^\infty \Lambda_3(\theta)J_2(w)ldl. \tag{82}$$

The other correlation functions may be determined in a similar manner.

For structures of known form, it is possible to calculate the correlation functions by numerically summing over the structure. For example, in Fig. 17 the functions $T_0(r)$, $T_4(r)$, and $S_4(r)$ are plotted as a function of $\rho=r/R$ for a two-dimensional isolated spherulite of uniform density having radially directed optic axes. These are calculated[11] by numerical summation of quantities like $F_1=\langle\cos2\theta_{ij}\rangle_{r_{ij},\beta}$ over all pairs of optic axis vectors lying within the spherulite and then obtaining quantities like $T_4(r)$ by integrating

$$T_4(r)=\frac{1}{\pi}\int_0^{2\pi} F_1(r,\beta)\cos4\beta d\beta. \tag{83}$$

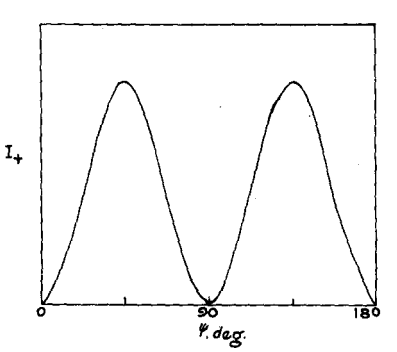

FIG. 18. Variation of $I_\perp$ with ψ calculated from the correlation functions of Fig. 17 at $\theta=5°$ for $R=7.5\,\mu$ and $\lambda=0.5\,\mu$.

The function $T_0(r)$ obtained in this manner is identical with the function $h(r)$ previously obtained in connection with the calculation of the transmission of light by spherulitic polymers.[11,12] The function T_0 is seen to oscillate through positive and negative values as it approaches zero (discretely at $r=2R$), rather than monotonically approaching zero as described by the empirical Gaussian functions while T_4 and S_4 are zero at $r=0$ and pass through a maximum. It is through such functions that information about the shape and anisotropy distribution of the structure is conveyed by the scattering experiment.

In Fig. 18, the calculated variation of the $I_1(\psi)$ scattering intensity with ψ at $\theta=5°$ for a spherulite of radius $R=1.5\mu$, $\lambda=0.5\mu$ is given using the above calculated correlation functions for the two-dimensional spherulite using Eq. (54). The agreement with experiment and with the scattering calculation made directly by the amplitude summation method[13] is very good, even though this method for statistically homogeneous systems has been applied to a statistically heterogeneous one.

Thus this theory may be applied to systems ranging from random collections of isolated crystals to highly ordered structures such as spherulites and offers a general statistical description of the structure of a complex system in terms of which other properties of the system may be described.

[11] S. Clough, Ph.D. thesis, University of Massachusetts, Amherst, Massachusetts, 1966.

[12] S. Clough and R. S. Stein, ONR Tech. Rept. No. 75, Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts, Project NR356-378, Contract: Nonr 3357(01), August 1963.

[13] P. Wilson and R. S. Stein, ONR Tech. Rept. No. 35, Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts, Project NR 356-378, Contract: Nonr 3357(01), August 1961.

JOURNAL OF APPLIED PHYSICS VOLUME 37, NUMBER 11 OCTOBER 1966

Measurement of Pole Figures and Orientation Functions for Polyethylene Films Prepared by Unidirectional and Oriented Crystallization

C. R. DESPER*† AND R. S. STEIN

Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts

The x-ray pole figure technique has been applied to the study of orientation in polyethylene. Wilchinsky-type orientation functions of the form $\langle\cos^2\phi_{hkl,q}\rangle$ are determined as well as pole figures. These functions are shown to be related to the crystalline contributions to birefringence and infrared dichroism. A new method of presenting these functions on an equilateral triangle plot is introduced. Equations are presented for calculating the orientation functions for weakly diffracting [*hkl*] planes for materials of the orthorhombic, tetragonal, hexagonal, and cubic systems. The method of analysis is applied to various polyethylene samples, including a unidirectionally recrystallized sample (I) and to crosslinked film recrystallized at low orientation (II), and at higher orientation (III). The resulting data are interpreted in terms of various morphological models and are also correlated with birefringence. The data on (I) are consistent with a model of random orientation of *a* and *c* about the *b* axis. In (II) and (III) the row nucleation model and the *a*-axis orientation model are both inadequate. For (II) a diffusely oriented helix model is suggested; for (III) a screw dislocation model of crystal growth, with the screw axes parallel to the stretch direction, is proposed.

I. INTRODUCTION

In recent years a number of workers[1–5] have applied the x-ray pole figure technique to characterize crystallite orientation in polymers. Birefringence,[2,6–10] infrared dichroism,[8,10] and light scattering[11] have also been used to study the morphology of biaxially oriented films. All of these methods have been brought to bear in this work, and in addition, a new method of graphically plotting crystalline orientation functions is proposed and demonstrated.

Pole figure data is presented on a number of polyethylene samples, including unidirectionally recrystallized film and films crystallized in the oriented state. The resulting data are interpreted in terms of crystallization and orientation mechanisms.

* Work supported in part by a contract with the Office of Naval Research and in part by grants from the Petroleum Research Fund, the Packaging Division of the Monsanto Company, and the Mobil Chemical Company.

† Present address: Materials Research Branch, U. S. Army Natick Laboratories, Natick, Mass.

[1] Z. W. Wilchinsky, J. Appl. Phys. **31**, 1969 (1960).

[2] C. J. Heffelfinger and R. L. Burton, J. Polymer Sci. **47**, 289 (1960).

[3] J. W. Jones, *Advances in X-Ray Analysis*, W. M. Mueller and M. Fay, Eds. (Plenum Press, Denver, 1962), Vol. 6, p. 223.

[4] Z. W. Wilchinsky, J. Appl. Polymer Sci. **7**, 923 (1963).

[5] P. H. Lindenmeyer and S. Lustig, J. Appl. Polymer Sci. **9**, 227 (1965).

[6] J. Spence, J. Phys. Chem. **43**, 865 (1939).

[7] R. S. Stein, J. Polymer Sci. **24**, 383 (1957).

[8] R. S. Stein, J. Polymer Sci. **31**, 335 (1958).

[9] H. de Vries, Angew. Chem. **74**, 574 (1962).

[10] W. J. Dulmage and A. L. Geddes, J. Polymer Sci. **31**, 499 (1958).

[11] R. S. Stein and T. Hotta, J. Appl. Phys. **35**, 2237 (1964).

Undulatory Riemannian Spaces

C. LANCZOS
Dublin Institute for Advanced Studies, Dublin, Ireland
(Received 7 March 1963)

A demonstration is given that Riemannian spaces of very high curvature in submicroscopic domains do not contradict the existence of a macroscopic line element which is nearly Minkowskian. The signature of the microscopic line element is positive definite and the wave property of the metric in macroscopic domains comes about by a peculiar "wave-guide action" of a strongly curved, two-dimensional line element, in harmony with the particlelike behavior of the photon. The four-dimensional lattice structure of the metrical vacuum field does not establish an absolute frame of reference and can be harmonized with the macroscopic validity of the Lorentz transformations.

1. INTRODUCTION

NUMEROUS probings of the author into the possibility of arriving at a rational theory of matter on the basis of singularity-free solutions of a Riemannian space satisfying a quadratic action principle,[1] led to the conclusion that such a program cannot be achieved on the basis of a line element which is not more than a more or less pronounced modification of the Minkowskian metric. An earlier paper by the author (cf. I, p. 716) enunciated for the first time the possibility that there exists an undulatory metrical plateau of submicroscopic wavelengths which in fact cause exceedingly high local curvatures, with the result that the constant λ in the "cosmological equations"

$$R_{ik} + \lambda g_{ik} = 0 \qquad (1.1)$$

is, in fact, not an astronomical but a subatomic constant, having in its wake an exceedingly *large* rather than exceedingly *small* average curvature. That this undulatory metrical plateau does not come phenomenologically into appearance, and is macroscopically smoothed out to a quasi-Euclidean metric, is caused by the immense smallness of the basic wavelength of these vibrations. In a later investigation the same thought returned once more in modified form (cf. II, p. 348), but with the same result, namely that the cosmological constant must be considered as an exceedingly large constant, thus providing a universal gauge of subatomic dimensions, which explains the extraordinary smallness of the elementary particles. Finally, in a new formulation of the quadratic action principle, which reduced the basic differential equations from fourth to second order with the help of a new tensor of the order three, the "spintensor" (cf. III, p. 386), the necessity for a large value of the scalar curvature R, was once more demonstrated.

In spite of these results, the investigation of an undulatory metrical plateau as the possible substructure of the geometry of the space–time world remained in a rudimentary stage. There existed a fundamental difficulty, pointed out to the author by Einstein,[2] which militated against these concepts, namely the spectre of an "absolute reference system" which would be established by such an undulatory field, in contradiction to the postulates of special relativity.

No progress could occur in this problem, on account of a "missing link" which only recently came in the author's possession. In the author's previous efforts he has always assumed that the metrical fields encountered are "nearly Minkowskian", that is the deviation of the g_{ik} from the Minkowskian values η_{ik} is not large. The high curvature of the Riemannian space was attributed to the highly oscillatory character of the deviations from the average values η_{ik}. High curvatures could be obtained on account of the first and second partial derivatives of the g_{ik} with respect to the x_i, even if the g_{ik} themselves were not far from constant values. Under these circumstances, the macroscopic line element did not differ fundamentally from the microscopic line element, but represented its average value, taken over a small, (although large compared with the submicroscopic-wavelength) domain of the space–time world. Hence in spite of the "macroscopic deception" which gives the illusion of a smooth, instead of highly agitated line element, it seemed that at least the *signature* of the line element cannot be a macroscopic deception, but must be assumed to be valid even for microscopic dimensions. This, in fact, was a fundamental mistake. If the undulatory background of

[1] C. Lanczos, Phys. Rev. **39**, 716, (1932); *ibid.* **61**, 713, (1942) (quoted as I); Rev. Mod. Phys. **29**, 337, (1947) (quoted as II); *ibid.* **34**, 379, (1962) (quoted as III).

[2] Private communication.

the metric is strongly Riemannian (and with that we mean that the values of the g_{ik} show a strong variation), then we cannot assume any *a priori* relation between the macroscopic and the microscopic line element, and the possibility is given that even the signature $+++-$ of the Minkowskian line element is not more than a macroscopic deception. By assigning to the true line element of nature the $++++$ signature, we do away with all the obviously unsatisfactory geometrical features of the Minkowskian world, particularly the fact that geodetic lines are not shortest lines, (in fact, any two points of a Minkowskian manifold can be connected by lines of zero length, and in infinitely many ways). To this has to be added the failure of the "triangular inequality"

$$\overline{AC} \leq \overline{AB} + \overline{BC}, \tag{1.2}$$

which is such a natural corollary of any reasonable metric, and the failure of the coincidence principle

$$\overline{AB} = 0 \quad \text{means} \quad A = B. \tag{1.3}$$

In fact, the Minkowskian geometry cannot be considered as a consistent differential geometry in which everything is based on the line element alone. The points which belong to the "distance zero" form in themselves a three-dimensional manifold (the "null cone") which has to be added to the geometry from extraneous sources. Nor are these geometrical inadequacies the only shortcomings. Another source of serious worry is the four-dimensional potential equation

$$\Delta\varphi - (1/c^2)\, \partial^2\varphi/\partial t^2 = \rho, \tag{1.4}$$

which in consequence of the hyperbolic character in t is not solvable in unique terms, because the homogeneous equation has an infinity of eigensolutions. We exclude these solutions by the arbitrary assumption that the solution should occur solely in terms of "retarded potentials". This entails a peculiar onesidedness of time events, in strange contrast to the time reversibility of the given differential operator. All these discrepancies are accepted on the basis that they are dictated by our physical observations.

The present investigation takes a different starting point. It restores to nature's line element the natural $++++$ signature, thus uniting space and time to a truly homogeneous manifold. We now have a genuinely Riemannian geometry, based on a positive-definite line element. The geodesics become shortest lines, the triangular inequality (1.2) is restored and so is the coincidence principle (1.3). We have a genuine differential geometry which is self-contained, without any extraneous elements. The four-dimensional potential equation (1.4) changes to an equation which is elliptic in all the four coordinates and has no longer any eigensolutions; (except φ = const). The basic unity and mathematical rationality of the physical universe is thus established on a previously unattained scale.

We will demonstrate that such a line element is in no contradiction to the macroscopic radiation phenomena. In fact, it is a strongly Riemannian, although positive-definite, line element by which we arrive at the transmission of light signals with light velocity, although the transmission does not occur any more in terms of waves which spread out in spheres, but in localized energy bundles traveling with light velocity, as demanded by Einstein's "light quantum" or "photon" hypothesis. Furthermore, a closer investigation of the question of coordinate transformations reveals that the danger of an "absolute reference system" is in fact not present. It is exactly the undulatory nature of the basic metrical field which makes a certain limited group of transformations available. The basic transformations are: rotation of the space coordinates, plus that special (x, t) Lorentz transformation from which Einstein derived the fundamental relativistic effects of special relativity.

2. PERIODICITY HYPOTHESIS

We assume that the postulated "undulatory metrical plateau" has the following structure: We divide the space–time continuum into an infinity of cubic cells of identical dimensions, and we assume that the values of the g_{ik} inside of each cell are strictly repeated from cell to cell. Hence the g_{ik} can be conceived as functions of a fourfold periodicity. If we know the distribution of the g_{ik} within one cell, we know in fact the metric everywhere, because the same values return in corresponding points. We thus obtain the picture of a four-dimensional *lattice*, filled with radiation of a certain kind. This radiation is *stationary*, inasmuch as the ten g_{ik} assume the same values in corresponding lattice points. Since a constant is a special case of a periodic function, the Euclidean values g_{ik} = const satisfy the periodicity hypothesis, with a period which can be chosen as large or as small as we wish. We want to assume, however, that in fact the g_{ik} within the fundamental cell are far from constants, but belong to a strongly Riemannian field of high curvature. The second assumption is that the fundamental lattice cell has immensely small dimensions. In any

observable physical phenomenon a tremendously large number of elementary cells are involved.

To handle a strongly Riemannian space mathematically is obviously a task which surpasses our present capabilities. If we consider that even the very weak gravitational fields in Einstein's theory are already of high complexity, we can envisage that strongly curved Riemannian spaces, whose metric changes radically from point to point, lie completely beyond our present methamatical competence. Yet it is possible that we can draw conclusions even *without* possessing any detailed knowledge of the metrical field contained in the elementary cell. Two circumstances come to our rescue: one is that we know the g_{ik} are strictly *periodic* functions of the lattice coordinates; the other is that the dimensions of the elementary cell are *extremely small.*

The assumption of a metrical background of this structure establishes a microstructure in our physical world which remains latent under ordinary physical conditions and reconstructible only by inference. The metrical space which takes cognizance of our usual physical experiences, may be vastly different from the true metrical space of the submicroscopic regions.

3. POTENTIAL EQUATION IN TWO DIMENSIONS

The strongest motivating force in the development of special relativity was the invariance of the Maxwell equations with respect to Lorentz transformations. But the Maxwell equations are reducible to the solution of the four-dimensional wave equation (1.4) and it is this equation which must be conceived of as the fundamental building block in the erection of classical electrodynamics. If we can deduce this equation on the basis of our assumptions, we have good reason to assume that we are on the right road toward the interpretation of relativistic phenomena.

In fact, even less is sufficient. In his celebrated paper on the "Electrodynamics of moving bodies,"[3] Einstein succeeded in deducing all the decisive relativistic effects—the Doppler effect, aberration, Fresnel convection coefficient, and the dynamical behavior of the electron—by considering the two-dimensional Lorentz transformation $(x, t) \rightarrow (x', t')$. This transformation is associated with the two-dimensional wave equation

$$\frac{\partial^2\varphi}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} = \rho. \tag{3.1}$$

[3] A. Einstein, Ann. Phys. **17**, 891 (1905).

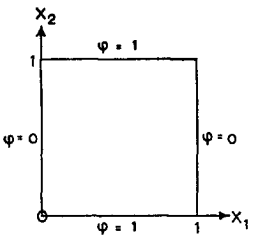

FIG. 1. Boundary-value problem in potential theory.

It will be our first aim to study the mathematical nature of this particular equation.

We consider the potential equation of an arbitrary Riemannian space, written in covariant form:

$$\Delta\varphi = \frac{1}{g^{\frac{1}{2}}}\frac{\partial}{\partial x_i}\left(g^{\frac{1}{2}}g^{ik}\frac{\partial\varphi}{\partial x_k}\right) = 0. \tag{3.2}$$

We restrict ourselves to two dimensions and assume that our line element is given in the orthogonal form

$$ds^2 = g_1(dx^1)^2 + g_2(dx^2)^2 \tag{3.3}$$

(replacing for the sake of simplicity g_{11} by g_1 and g_{22} by g_2). We assume a positive-definite metric, and thus

$$g_1 > 0, \qquad g_2 > 0. \tag{3.4}$$

Equation (3.2) can be conceived as the result of minimizing the following action integral:

$$A = \int \frac{\partial\varphi}{\partial x_i}\frac{\partial\varphi}{\partial x_k}\, g^{ik}g^{\frac{1}{2}}\, d\tau, \tag{3.5}$$

extended over a certain domain, on the boundaries of which the values of $\varphi(x_i)$ are freely prescribed. In our case, the action integral takes the following form:

$$A = \int \left[\left(\frac{g_2}{g_1}\right)^{\frac{1}{2}}\left(\frac{\partial\varphi}{\partial x_1}\right)^2 + \left(\frac{g_1}{g_2}\right)^{\frac{1}{2}}\left(\frac{\partial\varphi}{\partial x_2}\right)^2\right] dx^1\, dx^2. \tag{3.6}$$

In the simple Euclidean plane,

$$g_1 = g_2 = 1, \tag{3.7}$$

and the potential equation is reduced to the standard form,

$$(\partial^2\varphi/\partial x_1^2) + (\partial^2\varphi/\partial x_2^2) = 0. \tag{3.8}$$

We will now compare the solution of a simple boundary-value problem in the Euclidean plane (3.8) and in the Riemannian plane (3.3), whose structure we will prescribe in a definite way. This prescription may not correspond to an actual case realized in nature, but it can be used in the sense of a *model* to demonstrate what can be expected of an oscillatory Riemannian space of the positive signature $++$.

The boundary-value problem we want to solve is of a particularly simple kind. We choose a square

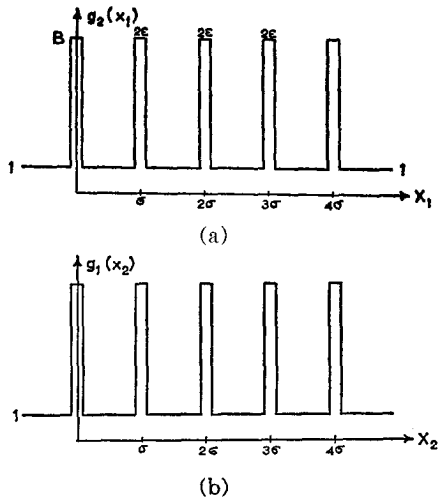

FIG. 2. Model of strongly Riemannian space; (a) plot of $g_2(x_1)$, (b) plot of $g_1(x_2)$.

region of the coordinates x_1, x_2 and prescribe the values $\varphi = 1$ on the upper and lower edges, and $\varphi = 0$ on the other two edges. In order to avoid doublevaluedness in the corners, we assume that the value $\varphi = 1$ on the two edges drops steeply to zero in the immediate neighborhood of the two corners of the edge. The region we have chosen shall be a *macroscopic* region, i.e., it shall contain a large number of lattice points of the Riemannian plane (3.3).

We define the metrical components of the line element (3.3) of the Riemannian space as follows:

$$g_1 = g_1(x_2), \qquad g_2 = g_2(x_1), \tag{3.9}$$

and picture these two functions according to accompanying graph (Fig. 2) which, on account of the periodicity, continues without end. The steep increase of $g_2(x_1)$ near the lattice points $x_1 = \alpha\sigma$ occurs with a finite but very steep tangent. Then $g_2(x_1)$ attains, along a narrow ridge, a constant high value B, which is maintained on a ridge of the width 2ϵ, after which $g_2(x_1)$ drops steeply down again to the normal value 1. Quite similar is the behavior of $g_1(x_2)$ in the lattice points $x_2 = \beta\sigma$. (Our graph is schematic since the first and second derivatives of the functions $g_1(x_2)$ and $g_2(x_1)$ must remain finite at every point. This, however, is irrelevant for our present purposes). If we plot the position of the high ridges of $g_1(x_2)$ and $g_2(x_1)$ in our domain, we obtain a figure as indicated in Fig. 3.

Let us now consider the variational principle (3.6), in whose minimization we are interested. Compared with the Euclidean plane, we observe the following difference: In the Euclidean case the weight factors of $(\partial_1\varphi)^2$ and $(\partial_2\varphi)^2$ are uniformly distributed, while in the Riemannian case the partial derivatives of φ obtain particularly large weight factors near to the ridges of Fig. 3. Let us first pay attention to the two edges $x_1 = 0$ and $x_1 = 1$. In the Euclidean case we have to consider our domain up to the very boundaries and minimize the integral

$$\int_0^1 \int_0^1 \left[\left(\frac{\partial\varphi}{\partial x_1}\right)^2 + \left(\frac{\partial\varphi}{\partial x_2}\right)^2\right] dx_1\, dx_2. \tag{3.10}$$

By reducing our domain to a smaller one, we restrict our potentialities and obtain a less favorable minimum. In the Riemannian case the situation is different. If we reduce the domain of minimization from $x_1 = [0, 1]$ to $x_1 = [\epsilon, 1 - \epsilon]$ and make φ to zero on the edges $x_1 = \epsilon$ and $x_1 = 1 - \epsilon$ (instead of $x_1 = 0$ and $x_1 = 1$), and then leave $\varphi = 0$ along the narrow strip $x_1 = [0, \epsilon]$ and $x_1 = [1 - \epsilon, 1]$, we have apparently not obtained the best possible minimum, but in fact we are in a particularly favorable position because we made the otherwise very large contribution

$$\int_0^\epsilon \left(\frac{g_2}{g_1}\right)^{\frac{1}{2}}\left(\frac{\partial\varphi}{\partial x_1}\right)^2 dx_1 + \int_{1-\epsilon}^1 \left(\frac{g_2}{g_1}\right)^{\frac{1}{2}}\left(\frac{\partial\varphi}{\partial x_1}\right)^2 dx_1 \tag{3.11}$$

[large on account of the excessive weight factor $(g_2/g_1)^{\frac{1}{2}}$ within this narrow strip], to zero, thus obtaining a much better minimum than if we had allowed any finite value of $\partial\varphi/\partial x_1$ in this strip. The result is that *not only* φ *but even the normal derivative* $\partial\varphi/\partial n$ *becomes zero at the two null edges* $x_1 = 0$ *and* $x_1 = 1$.

A similar consideration holds concerning the attenuation of the function $\varphi = 1$, as we move from the lower edge to the upper edge. In the Euclidean plane, the minimum of the variational integral is obtained by attenuating the function gradually, until a minimum is reached at $x_2 = \frac{1}{2}$, after which the function increases again, reaching the value $\varphi = 1$ on the upper edge. The solution of the potential equation (3.8) under the given boundary conditions is obtained by the usual separation into products of functions which depend on x_1 and x_2 only, and then determining the free

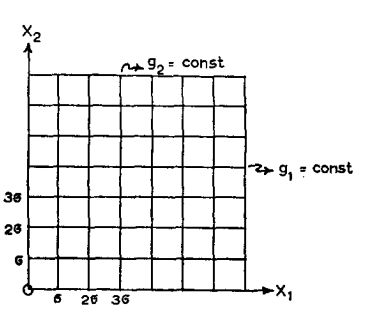

FIG. 3. High ridges of $g_1(x_2)$ and $g_2(x_1)$.

constants of the solution by satisfying the given boundary conditions. In our present problem, the solution of the boundary-value problem of Fig. 1 is given by the following formula:

$$\varphi(x_1, x_2) = \frac{4}{\pi} \sum \frac{\sin k\pi x_1}{k} \frac{\cosh k\pi(x_2 - \frac{1}{2})}{\cosh \frac{1}{2}(k\pi)} \qquad (k = 1, 3, 5, 7, \cdots). \tag{3.12}$$

In Fig. 4, the transmitted profiles are graphically indicated for the values $x_2 = \frac{1}{4}, \frac{2}{4}, \frac{3}{4}$.

In the Riemannian case the situation is quite different. By avoiding the attenuation and thus keeping the same profile of φ from one edge to the other, we make $\partial\varphi/\partial x_2 = 0$, and thus avoid the large contributions which would otherwise accrue from the integral

$$\int_{\beta\sigma-\epsilon}^{\beta\sigma+\epsilon} \left(\frac{g_1}{g_2}\right)^{\frac{1}{2}} \left(\frac{\partial\varphi}{\partial x_2}\right)^{\frac{1}{2}} dx_2 \tag{3.13}$$

on account of the high ridges of g_1.

The result of this discussion is that the high ridges of the functions $g_1(x_2)$ and $g_2(x_1)$ operate like a *wave guide* in transmitting radiation in the vertical direction, (cf. Fig. 4b), practically without any attenuation. Exactly the same holds, however, in the horizontal direction, since we could have formulated our boundary value problem with exchanged x_1 and x_2.

We thus notice two characteristic differences between the spread of the potential in the Euclidean and the Riemannian case. In the first case the potential is attenuated as we move into the inside of the domain; in the second case the outer profile remains unchanged. Furthermore, in the Euclidean plane the normal derivative of φ, as we arrive at the zero edge, is not zero, and thus the function cannot be terminated there but spills over into the surrounding domain. In the Riemannian case the potential is completely contained between the two edges $\varphi = 0$ and does not spread into the outer domain.

All these statements have to be taken as *approximations* only, since they would hold exactly only if the ridges of g_1 and g_2 become infinitely narrow, while the values of g_1 and g_2 become infinite on the ridges, i.e., if g_1 and g_2 turned into true delta functions. Such a model is not permissible, of course, but we can approach the asymptotic stage with a high degree of accuracy if $g_1(x_2)$ and $g_2(x_1)$ behave *very nearly* like delta functions.

Under these circumstances we can write down the solution of the potential equation (3.2) for our model, illustrated by Fig. 2, with a high degree of accuracy as follows:

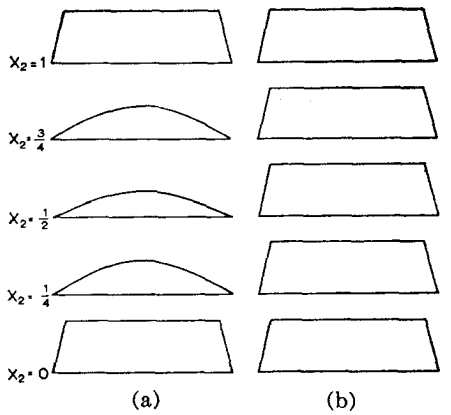

FIG. 4. (a) Solution of potential problem in Euclidean plane; (b) Wave-guide action of Riemannian space.

$$\varphi(x_1, x_2) = F(x_1) + G(x_2) \tag{3.14}$$

(holding, of course, only for *macroscopic* domains), which means

$$\partial^2\varphi/\partial x_1\, \partial x_2 = 0, \tag{3.15}$$

and if we now rotate our axes by 45° according to the transformation

$$\begin{aligned} \xi_1 &= (1/\sqrt{2})(x_1 + x_2), \\ \xi_2 &= (1/\sqrt{2})(-x_1 + x_2), \end{aligned} \tag{3.16}$$

then Eq. (3.15) becomes, in the new reference system,

$$\tfrac{1}{2}[(\partial^2\varphi/\partial\xi_1^2) - (\partial^2\varphi/\partial\xi_2^2)] = 0, \tag{3.17}$$

which is the *two-dimensional wave equation*, if the light velocity is normalized to 1.

We have thus demonstrated that the potential equation of a positive-definite Riemannian space of highly undulatory character can *in a macroscopic domain imitate the behavior of the wave equation*, i.e., *it can transmit signals to the right and to the left with light velocity, without appreciable attenuation.*

Without relying on any particular model, we will assume the following property of the metrical tensor enclosed in the fundamental unit cell. We can certainly solve the boundary-value problem associated with the potential equation (3.2) by prescribing the value $\varphi = 1$ on two neighboring edges and $\varphi = 0$ on the other two edges. This problem has a unique solution. Without knowing any further properties of the fundamental cell, we will assume that the solution of this boundary-value problem has the property that, together with $\varphi = 0$, also $\partial\varphi/\partial n$ (n = normal) is seen to become zero (exactly speaking, very small). This cell cannot exist in

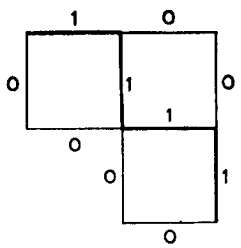

FIG. 5. Zero edges ($\varphi = 0$, $\partial\varphi/\partial n = 0$) in Riemannian space.

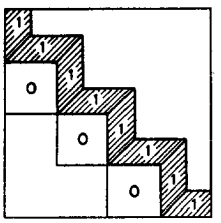

FIG. 7. Wave-guide effect in 3 dimensions.

isolation, on account of the two edges on which $\varphi = 1$. But it is possible that a continuation can be obtained by putting a second cell above the first cell. The edges denoted by zero are such that not only φ but also $\partial\varphi/\partial n$ vanishes there to a high degree of accuracy. This construction does not allow a termination but we can continue the staircase by going both upward and downward to infinity. The staircase is now flanked by zero edges everywhere and this means that the radiation is contained (to a high degree of accuracy) in a narrow strip of the width of two cells. Inside the cell we have a "black box" of unknown structure, except for the fact that the black boxes in corresponding positions are exactly equal to each other.

A termination of the staircase is possible only if a *right side* appears in the potential equation, that is if we solve the inhomogeneous equation

$$\Delta\varphi = \rho. \tag{3.18}$$

The finite staircase AB has thus the physical significance of a brief (practically infinitesimal) light pulse emitted to the right by a particle at A and absorbed by a particle at B. By the superposition principle of linear operators, we can add any number of such staircases, thus obtaining the transmission of an arbitrary (possibly time-dependent) time signal of arbitrary duration.

Similar staircase constructions are possible in higher dimensions, as Fig. 7 illustrates. The remarkable feature of these transmissions is that the radiation is restricted to a narrow tube enclosed between very narrow parallel walls, generated by the wave-guide property of an undulatory strongly Riemannian space.

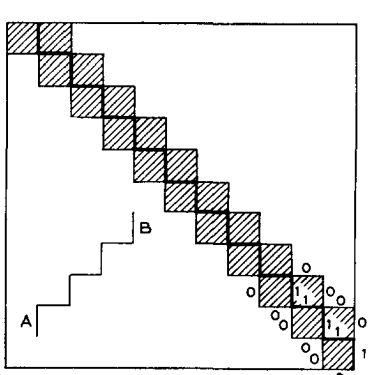

FIG. 6. Signal propagation in Riemannian space.

We have thus obtained a completely consistent, field-theoretic description of Einstein's celebrated "light quantum" or "photon" hypothesis,[4] which assigns to the spread of radiation a localized property of the nature of a particle which moves in a definite direction with light velocity, without spreading into a wave. Indeed, we would not be able to obtain anything resembling the wave equation on the basis of a positive-definite line element. But an undulatory and strongly curved Riemannian space has the power of providing the high potential walls which prevent radiation from spreading, thus confining the radiation within a narrow tube in which the signal propagates without attenuation—although this cannot hold with absolute accuracy since a certain amount of leakage is inevitable.

Even Einstein's fundamental equation concerning the energy of a photon,

$$E = h\nu, \tag{3.19}$$

which has baffled any field-theoretic explanation, can now be interpreted in simple terms. Let us assume that the transmitted profile of a light excitation is not assumed in the form of a sinusoidal vibration but in the form of a *square wave* (cf. Fig. 8). The majority of the transmitting cells are now filled with the potential value $\varphi = \text{const}$, and this is *exactly* true, no matter how strongly curved the Riemannian metric of the elementary cells may be, since $\varphi = \text{const}$ is *always* an exact solution of the potential equation—and thus they carry *no energy*. The energy is concentrated in those cells, which transmit the abrupt *change* of φ from the value $+1$ to -1. In view of the identity of the metrical cells, the radiation energy received at a certain point per unit time will thus be simply proportional to the number of sign changes, and that means the *frequency* of the transmitted radiation. Hence two emitted photons will transmit energy *proportional to their frequency*, if the amplitude

[4] A. Einstein, Ann. Phys. **17**, 132 (1905).

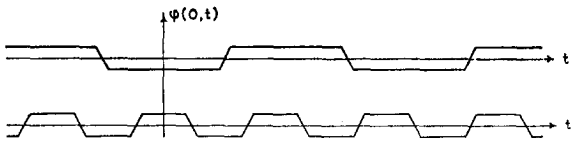

FIG. 8. Signal profile in the form of square wave.

of the transmitted profile remains the same constant for both photons. If we assume, in addition, that even the *length* of the transmitted profile is the same for all photons, then the energy carried by the photons becomes equal to the frequency, *multiplied by a universal constant.*

4. THE LORENTZ TRANSFORMATION

In personal discussions with the author, Einstein expressed his doubts concerning the possibility of an undulatory metrical plateau. He argued that such a metrical plateau would establish an *absolute frame of reference*, which cannot be harmonized with the experimentally proved validity of the Lorentz transformations. We would have to expect phenomena which are in contradiction to the relativistic effects demanded by special relativity. It will be our task to examine the validity of this argument and the possibility of removing the alleged contradiction.

It is certainly true that if the undulatory plateau establishes a crystallike lattice structure in the space–time world, due to a strongly Riemannian metric which is repeated as we go from cell to cell, the very existence of such a lattice distinguishes a definite set of coordinates. It is meaningless to analyze a crystal in arbitrary coordinates since the periodic nature of the crystal lattice automatically prescribes the coordinates in which we have to study the physical properties of the crystal. Yet we cannot be sure that this system will be the *only* one to do justice to the structure of the crystal.

We must keep in mind that we are interested in phenomena which involve a *very large number* of elementary cells, since the customarily encountered lengths are large in comparison to the distance between two lattice points. Under these circumstances, it is not the elementary cell but the concerted action of a very large number of cells which comes to our attention in the observable physical phenomena. It is possible that we can modify the dimensions or the shape of the elementary cells and still derive identical, or equivalent, results.

The first characteristic property of the elementary cell is the *periodicity* of the g_{ik} with respect to a translation by the unit distance, along any one of the four axes X_1, X_2, X_3, X_4. But the same property remains valid if we choose the elementary cells in infinitely many different ways. In fact, if we had no other condition to meet but the periodicity, we could draw a straight line from the origin of the reference system to any four not-coplanar lattice points and consider the parallelepiped thus generated as the new elementary cell. This would mean a general *linear transformation* of our original coordinates. The coefficients of this transformation would not form a continuous manifold; yet, considering the immense smallness of the basic length, they could be considered as *practically* continuous. Such a transformation would be of little value, however, if it did not preserve the fundamental *symmetry patterns* which characterize the geometry of the basic cell. We are interested solely in such linear transformations which will preserve not only the periodicity, but also the basic symmetry elements of the fundamental cell.

The first transformation which is at our disposal is that we enlarge the fundamental cell by multiplying every one of the edges by an arbitrary (not too excessive) integer. The new, larger cell is in every respect similar to the original one and can be considered as a substitute for it. Hence the first transformation at our disposal is an arbitrary *scale factor* by which the four coordinates may be multiplied.

Next we consider once more the two-dimensional metric associated with the coordinates ξ_1, ξ_2, with which we have dealt before [cf. Eq. (3.16)]. The high ridges of g_1 and g_2, which we considered in our Riemannian model, were at 45° to the coordinate lines. Hence we can assume that the basic cell exhibits a symmetry pattern for which the two cross diagonals are of vital importance. We preserve this symmetry pattern if we choose our transformed cell as a *rhombus* which is symmetrically oriented with respect to the direction of the cross diagonals, as demonstrated in Fig. 9.

If we denote the original coordinates by (x, t)

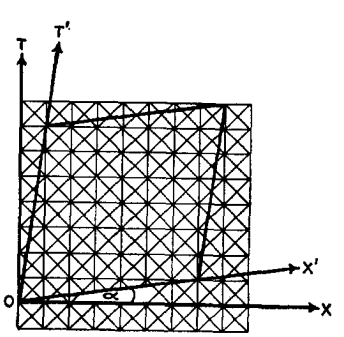

FIG. 9. Transformation in the (x, t) plane.

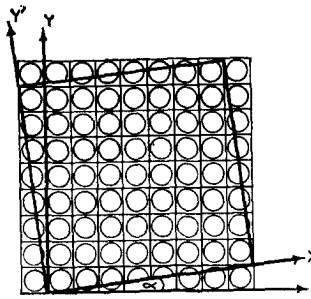

FIG. 10. Transformation in the (x, y) plane.

(instead of ξ_1, ξ_2), and the new coordinates by (x', t'), we obtain the following relation between the old and the new coordinates:

$$x = C(x' \cos\alpha + t' \sin\alpha), \qquad t = C(x' \sin\alpha + t' \cos\alpha). \tag{4.1}$$

The constant C arises from the previously observed freedom of a scale factor. This transformation is not far from that special Lorentz transformation that Einstein established in his paper of 1905. Our transformation (4.1) is characterized by the property that $x^2 - t^2$ is transformed proportionally;

$$x^2 - t^2 = C^2(\cos^2\alpha - \sin^2\alpha)(x'^2 - t'^2). \tag{4.2}$$

This expresses already the fact that the light velocity (here normalized to 1) is transformed as an invariant quantity, since $x^2 - t^2 = 0$ involves $x'^2 - t'^2 = 0$. But let us assume that we have the right to make the further assertion that the metric of the elementary cell is such that any macroscopic phenomenon described in the (x, t) system can equally be described in the (x', t') system, *with no preference for the one or the other system.* In that case we come to the determination of the free constant C of the transformation (4.1), in exactly the same manner in which Einstein obtained that constant, demanding the reversibility of the transformation (4.1).[5] In that case the factor on the right side of (4.2) must become 1, which gives

$$C = 1/(\cos^2\alpha - \sin^2\alpha)^{\frac{1}{2}}. \tag{4.3}$$

The demanded Lorentz transformation is then established.

We see that the supposed undulatory metrical plateau, far from contradicting the existence of the Lorentz transformation, may possibly yield its deeper significance. Since all the relativistic effects of Einstein's special relativity are solely based on the equivalence of two reference systems which are connected by a Lorentz transformation, we need not fear that any of the relativistic effects may contradict the underlying hypothesis of a fundamental submicroscopic metrical radiation field.

Instead of considering an (x, t) cross section of the four-dimensional world, we can equally consider an (x, y)—or (x, z)—cross section. Here the symmetry pattern is fundamentally different and we will assume that we have a metrical structure of the elementary cell which is of essentially *circular* (in the x, y, z space spherical) symmetry. In this case the transformed parallelogram is merely *rotated*, compared with the fundamental parallelogram. The relation between the original and the transformed coordinates now becomes

$$x = x' \cos\alpha - y' \sin\alpha, \qquad y = x' \sin\alpha + y' \cos\alpha, \tag{4.4}$$

which is an *orthogonal* transformation. These transformations are now by no means self-evident and of vital importance by abolishing the privileged direction of the fundamental axes and allowing the propagation of a signal in arbitrary directions.

5. SUMMARY AND OUTLOOK

The present investigation departs in some vital points from the traditionally accepted concepts of theoretical physics. It assumes an unadulterated Riemannian geometry with a positive definite line element of the signature $++++$, thus abandoning the Minkowskian indefinite signature $+++-$. Consequently the geodesics become shortest lines, the null lines cease to exist, and the triangular inequality (1.2) is restored. The four-dimensional potential equation becomes purely elliptic and ceases to have eigensolutions, making the solution of the four-dimensional Poisson equation unique. Space and time are thus united to a truly homogeneous four-dimensional manifold.

In this purely Riemannian world, the existence of a submicroscopic metrical radiation field of fourfold periodicity is postulated. If we investigate the properties of this metrical field by its two-dimensional cross sections, we find that the symmetry properties of the line element are fundamentally different in the cross sections which exclude, and cross sections which include the fourth coordinate (the time). While the Riemannian line element is *a priori* homogeneous in all the four coordinates, the actually existing undulatory plateau establishes *a posteriori* an asymmetry between space and time coordinates, which finds its macroscopic expression in the Minkowskian line element. But

[5] Cf. reference 3, p. 902.

the indefiniteness of the macroscopic line element is not more than a deception, brought about by the immense smallness of the fundamental lattice structure—a structure which remains hidden under macroscopic conditions, giving the illusion of a smooth line element with constant values of the g_{ik}, of the signature $+++-$. We can prove that, in consequence of a peculiar wave-guide action of a strongly Riemannian field, the propagation of signals with light velocity becomes possible, but that this propagation occurs in narrow channels of high-energy concentration, which have very little leakage. In this manner, Einstein's "photon hypothesis" receives its field-theoretical justification.

In view of the positive definiteness of nature's line element, the eigenvalues of the matter tensor become real and the quadratic action principle employed for the derivation of the basic field equations changes from a stationary value principle to a genuine minimum principle.

Although these ideas differ radically from the Einsteinian concepts, they are natural developments of his immortal thought constructions. The *a priori* character of the theory is in harmony with Einstein's metaphysical principle of the rationality of cosmic happenings, so vividly demonstrated by the success of his gravitational speculations. The basic difference between the theory here pursued and Einstein's own attempts toward a unified field theory can be stated in the following terms: Einstein was constantly looking for static, spherically symmetric solutions of the field equations which should represent the fields of material particles. Here we are looking for solutions which satisfy *fourfold periodicity conditions*. It is possible that the "cosmological equations"(1.1), which are exact solutions of the field equations, provide in themselves (with a tremendously high value of the cosmological constant λ) the desired undulatory metrical plateau of the universe, without the necessity of invoking the *general* solutions which follow from the quadratic action principle. The equations (1.1) have not as yet been sufficiently investigated to tell whether or not the demanded regular solutions with fourfold periodicity (and a positive-definite metric) exist.

But even after finding this solution, we are still not further than at a point which now corresponds to the matterless world of special relativity, macroscopically characterized by the Minkowskian line element. We now have to find, by a superimposed perturbation, some other solutions of the field equations (derived by minimizing the quadratic Lagrangian action principle), which will yield the fields of material particles. Here the boundary conditions are no longer periodic, and in fact no boundary conditions are demanded, except regularity of the solution and gradually *de*creasing, rather than *in*creasing, perturbation of the periodic radiation field. Although these solutions will not be locally static, they will, in *macroscopic* relations, appear as static fields of spherical symmetry. The previous investigations of the author have demonstrated that a four-dimensional Riemannian world contains, in the vector potential and the "spin tensor", those elements which seem to be in harmony with the fundamental building blocks demanded for the erection of material particles.

These extrapolations into future possibilities may appear from the purely empirical standpoint as premature, if not futile. And yet, the fact is that the basic structure of the proposed theory is not contradicted by empirical evidence. Admitting Einstein's marvelous postulate that the true geometry of the space–time world is not Euclidean but Riemannian, the added dictum: "But it must be very nearly Euclidean, to remain in harmony with our everyday experiences", is in fact a *non sequitur*, if we include in our statement the submicroscopic metrical world. Our macroscopic experiences do not contradict the possibility of a four-dimensional periodic lattice structure, if the lattice constant of this structure is tremendously small. Such a structure cannot come into evidence under standard macroscopic conditions, because it becomes averaged out to a plain Euclidean platform. But the quantum theory of the vacuum field clearly indicates that this field is in fact the seat of very high energy density, and thus in a state of high agitation, in contradiction to a passive Euclidean structure of vanishing energy content. If, in addition, the possibility is given that we can base the geometry of the universe on a genuine Riemannian metric with a positive-definite line element, and still explain the indefinite Minkowskian line element as a result of macroscopic averaging, one can hardly deny that we have gained immensely in the inner consistency and rationality of our physical world picture.

The function $T_0(r)$ obtained in this manner is identical with the function $h(r)$ previously obtained in connection with the calculation of the transmission of light by spherulitic polymers.[11,12] The function T_0 is seen to oscillate through positive and negative values as it approaches zero (discretely at $r=2R$), rather than monotonically approaching zero as described by the empirical Gaussian functions while T_4 and S_4 are zero at $r=0$ and pass through a maximum. It is through such functions that information about the shape and anisotropy distribution of the structure is conveyed by the scattering experiment.

In Fig. 18, the calculated variation of the $I_1(\psi)$ scattering intensity with ψ at $\theta=5°$ for a spherulite of radius $R=1.5\mu$, $\lambda=0.5\mu$ is given using the above calculated correlation functions for the two-dimensional spherulite using Eq. (54). The agreement with experiment and with the scattering calculation made directly by the amplitude summation method[13] is very good, even though this method for statistically homogeneous systems has been applied to a statistically heterogeneous one.

Thus this theory may be applied to systems ranging from random collections of isolated crystals to highly ordered structures such as spherulites and offers a general statistical description of the structure of a complex system in terms of which other properties of the system may be described.

[11] S. Clough, Ph.D. thesis, University of Massachusetts, Amherst, Massachusetts, 1966.

[12] S. Clough and R. S. Stein, ONR Tech. Rept. No. 75, Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts, Project NR356-378, Contract: Nonr 3357(01), August 1963.

[13] P. Wilson and R. S. Stein, ONR Tech. Rept. No. 35, Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts, Project NR 356-378, Contract: Nonr 3357(01), August 1961.

JOURNAL OF APPLIED PHYSICS VOLUME 37, NUMBER 11 OCTOBER 1966

Measurement of Pole Figures and Orientation Functions for Polyethylene Films Prepared by Unidirectional and Oriented Crystallization

C. R. DESPER*† AND R. S. STEIN

Polymer Research Institute, University of Massachusetts, Amherst, Massachusetts

The x-ray pole figure technique has been applied to the study of orientation in polyethylene. Wilchinsky-type orientation functions of the form $\langle\cos^2\phi_{hkl,q}\rangle$ are determined as well as pole figures. These functions are shown to be related to the crystalline contributions to birefringence and infrared dichroism. A new method of presenting these functions on an equilateral triangle plot is introduced. Equations are presented for calculating the orientation functions for weakly diffracting [*hkl*] planes for materials of the orthorhombic, tetragonal, hexagonal, and cubic systems. The method of analysis is applied to various polyethylene samples, including a unidirectionally recrystallized sample (I) and to crosslinked film recrystallized at low orientation (II), and at higher orientation (III). The resulting data are interpreted in terms of various morphological models and are also correlated with birefringence. The data on (I) are consistent with a model of random orientation of *a* and *c* about the *b* axis. In (II) and (III) the row nucleation model and the *a*-axis orientation model are both inadequate. For (II) a diffusely oriented helix model is suggested; for (III) a screw dislocation model of crystal growth, with the screw axes parallel to the stretch direction, is proposed.

I. INTRODUCTION

In recent years a number of workers[1–5] have applied the x-ray pole figure technique to characterize crystallite orientation in polymers. Birefringence,[2,6–10] infrared dichroism,[8,10] and light scattering[11] have also been used to study the morphology of biaxially oriented films. All of these methods have been brought to bear in this work, and in addition, a new method of graphically plotting crystalline orientation functions is proposed and demonstrated.

Pole figure data is presented on a number of polyethylene samples, including unidirectionally recrystallized film and films crystallized in the oriented state. The resulting data are interpreted in terms of crystallization and orientation mechanisms.

* Work supported in part by a contract with the Office of Naval Research and in part by grants from the Petroleum Research Fund, the Packaging Division of the Monsanto Company, and the Mobil Chemical Company.

† Present address: Materials Research Branch, U. S. Army Natick Laboratories, Natick, Mass.

[1] Z. W. Wilchinsky, J. Appl. Phys. **31**, 1969 (1960).

[2] C. J. Heffelfinger and R. L. Burton, J. Polymer Sci. **47**, 289 (1960).

[3] J. W. Jones, *Advances in X-Ray Analysis*, W. M. Mueller and M. Fay, Eds. (Plenum Press, Denver, 1962), Vol. 6, p. 223.

[4] Z. W. Wilchinsky, J. Appl. Polymer Sci. **7**, 923 (1963).

[5] P. H. Lindenmeyer and S. Lustig, J. Appl. Polymer Sci. **9**, 227 (1965).

[6] J. Spence, J. Phys. Chem. **43**, 865 (1939).

[7] R. S. Stein, J. Polymer Sci. **24**, 383 (1957).

[8] R. S. Stein, J. Polymer Sci. **31**, 335 (1958).

[9] H. de Vries, Angew. Chem. **74**, 574 (1962).

[10] W. J. Dulmage and A. L. Geddes, J. Polymer Sci. **31**, 499 (1958).

[11] R. S. Stein and T. Hotta, J. Appl. Phys. **35**, 2237 (1964).

II. EXPERIMENTAL METHODS

A. Measurement of Pole Figures

X-ray pole figure data were obtained with our automated x-ray diffractometer,[12] using the geometrical configuration of Decker, Asp, and Harker.[13] Nickel-filtered Cu$K\alpha$ radiation was used. The Bragg angle 2Θ was set at a particular [hkl] diffraction peak, and the sample was moved by pulse motors about the ω and χ axes, as shown in Fig. 1. Provision was made for stretching the sample on the pole figure goniometer in one direction. The raw data were obtained at each (ω,χ) position at a prefixed count of 1000 counts, automatically punched onto IBM cards, and processed by computers.[14]

The intensity data were corrected for background, polarization, absorption, incoherent scattering, and overlap of adjacent diffraction peaks. The correction equations were checked by running scans on an unoriented sample, obtained by casting powdered polyethylene in epoxy resin, running scans on the amorphous, [110], [200], and [020] peaks. The method of Fujino *et al.*[15] was used to correct for overlap between the various peaks. It was found necessary to apply an instrumental correction to the [020] data obtained in the reflection region. This correction is due to narrowing of the peak as parafocusing geometry is obtained. A similar correction was found necessary by Wilchinsky.[4] These corrections are discussed more fully in the Appendix.

B. Measurement of Birefringences

Two components of birefringence were measured, using the method of Stein,[7] using a Babinet compensator. The birefringences obtained were:

$$\Delta_x = n_z - n_y \text{ (in-plane birefringence)} \quad (1)$$

$$\Delta_y = n_z - n_x \text{ (out-of-plane birefringence).} \quad (2)$$

The directions x, y, and z are defined to be the film normal, transverse direction, and machine or stretch direction, respectively.

C. Measurement of Crystallinity

The crystallinity of various polyethylene samples were measured both by density and by x-ray diffraction. Densities were obtained using an alcohol–water density gradient column at 23°; crystallinity was calculated using values[16,17] of 0.805 and 1.002 g/cc for the amorphous and crystalline densities, respectively. The procedure of Matthews, Peiser, and Richards[18] was used to calculate crystallinity from x-ray diffraction data, with suitable modification to average out orientation by integration over all possible orientations. This procedure is intimately related to the procedure for correcting for overlap and is discussed in the Appendix.

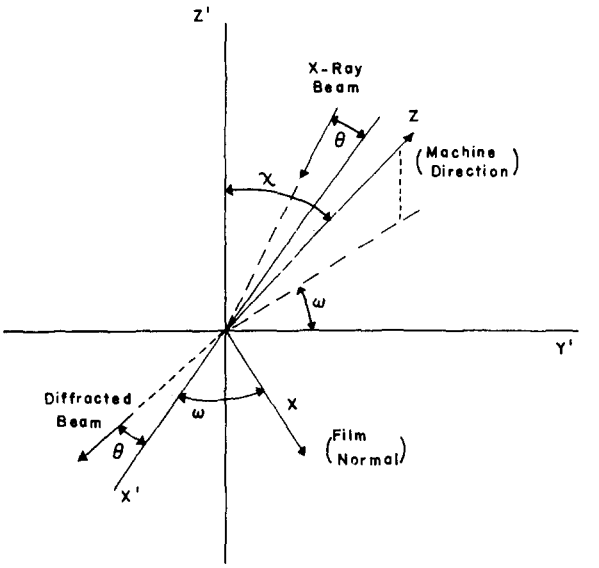

FIG. 1. Motions of the pole figure goniometer.

III. GRAPHICAL REPRESENTATION OF ORIENTATION FUNCTIONS IN BIAXIALLY ORIENTED SAMPLES

A. Definition of Orientation Function Coordinate System

In the past, a number of workers[19–21] have used orientation functions to describe the state of orientation of uniaxially oriented samples. In this method, a sample is characterized by functions such as

$$f_a = \tfrac{3}{2}\langle\cos^2\phi_{a,z}\rangle - \tfrac{1}{2}. \quad (3)$$

The angle $\phi_{a,z}$ is the angle between the a axis of a given crystallite and the symmetry axis of the film (z direction). The brackets indicate averaging over all crystallites. Similar functions f_b and f_c are defined for the b and c axes.

Wilchinsky[4] has expanded the concept of the orientation function to biaxially oriented films. We are

[12] C. R. Desper, ONR Tech. Rept. No. 78, Project NR: 056-378, contract Nonr 3357(01), University of Massachusetts, 15 January 1965.
[13] B. F. Decker, E. T. Asp, and D. Harker, J. Appl. Phys. **19**, 388 (1948).
[14] C. R. Desper, ONR Tech. Rept. No. 80, Project NR: 056-378, contract Nonr 3357(01), University of Massachusetts, October 1965.
[15] K. Fujino, H. Kawai, T. Oda, and H. Maeda, *Proceedings of the Fourth International Congress on Rheology*, E. H. Lee, Ed. (Interscience Publishers, New York, 1963), Vol. 3, p. 501.
[16] J. B. Nichols, J. Appl. Phys. **25**, 840 (1954).
[17] H. Kojima and A. Abe, Kobunshi Kagaku **18**, 728 (1961).
[18] J. L. Matthews, H. S. Peiser, and R. B. Richards, Acta Cryst. **2**, 85 (1949).
[19] P. H. Hermans, *Physics and Chemistry of Cellulose Fibers* (Elsevier Publ. Co., New York, 1949), Chap. 5.
[20] R. S. Stein, J. Polymer Sci. **31**, 327 (1958).
[21] S. Hoshino, J. Powers, D. G. LeGrand, H. Kawai, and R. S. Stein, J. Polymer Sci. **58**, 185 (1962).

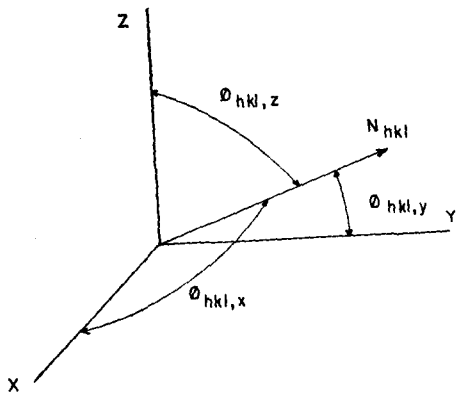

FIG. 2. Orientation of the [*hkl*] normal for a single crystallite.

proposing a new way of graphing the Wilchinsky functions which has a number of advantages to offer. Let us specify the orientation of the [*hkl*] normal of a particular crystallite by its direction cosines with respect to the film coordinates x,y,z, as shown in Fig. 2. Wilchinsky[4] then defines three quantities $\langle\cos^2\phi_{hkl,x}\rangle$, $\langle\cos^2\phi_{hkl,y}\rangle$, and $\langle\cos^2\phi_{hkl,z}\rangle$ the three average squared direction cosines, obtained by integration over all possible orientations. (The procedure for obtaining these quantities is discussed more fully in Sec. IIIC). These quantities may then be used to specify the location of a point on an equilateral triangle diagram, as shown in Fig. 3. Only two of the three quantities are independent, the third being fixed by the dependency relationship:

$$\langle\cos^2\phi_{hkl,x}\rangle+\langle\cos^2\phi_{hkl,y}\rangle+\langle\cos^2\phi_{hkl,z}\rangle=1. \quad (4)$$

The point *hkl* assumes different locations within the *X,Y,Z* triangle depending upon the Miller indices *h*, *k*, and *l*, and depending upon the state of orientation of the sample. Several possible limiting cases are shown in Fig. 3:

Point 1: Perfect orientation of the [*hkl*] normals parallel to the x direction of the film, which is the film normal. At this vertex, labeled X,

$$\langle\cos^2\phi_{hkl,x}\rangle=1 \quad (5)$$

and

$$\langle\cos^2\phi_{hkl,y}\rangle=\langle\cos^2\phi_{hkl,z}\rangle=0. \quad (6)$$

Point 2: Orientation of the [*hkl*] normals perpendicular to the x direction, where

$$\langle\cos^2\phi_{hkl,x}\rangle=0. \quad (7)$$

The [*hkl*] normals are constrained to lie in the y,z plane of the film, and the remaining two orientation functions are constrained by

$$\langle\cos^2\phi_{hkl,y}\rangle+\langle\cos^2\phi_{hkl,z}\rangle=1. \quad (8)$$

Point 3: Random orientation of the [*hkl*] normals would map into the center of triangle X, Y, Z, where

$$\langle\cos^2\phi_{hkl,x}\rangle=\langle\cos^2\phi_{hkl,y}\rangle=\langle\cos^2\phi_{hkl,z}\rangle=\tfrac{1}{3}. \quad (9)$$

Point 4: Uniaxial orientation about the z direction maps onto the line bisecting the 60° angle at vertex Z. In this case,

$$\langle\cos^2\phi_{hkl,x}\rangle=\langle\cos^2\phi_{hkl,y}\rangle=(1-\langle\cos^2\phi_{hkl,z}\rangle)/2 \quad (10)$$

Uniaxial orientation about the x or y axes maps onto the bisectors of the corresponding X and Y vertices of triangle X, Y, Z.

It should be noted that we have not made any assumptions concerning the symmetry of the orientation distribution. Thus, the film coordinates x, y, z, could be arbitrarily chosen. If the sample possesses biaxial symmetry—and by this we mean that the orientation distribution of the sample possesses three mutually perpendicular mirror planes of symmetry—then the coordinate system x, y, z, should be chosen to reflect the symmetry of the sample.

B. Characterization of Polyethylene Orientation by *a*, *b*, and *c* Axis Orientation Functions

So far we have treated the orientation functions of [*hkl*] planes in general. It is of particular interest to consider the orientation functions of the *a*, *b*, and *c* crystal axes. Figure 4 shows the points *a*, *b*, and *c* plotted to depict the orientation of these three crystal axes for a hypothetical polyethylene sample. The locations of these three points are not independent but are bound by these relationships, which arise from the orthogonality of the three crystal axes:

$$\langle\cos^2\phi_{a,x}\rangle+\langle\cos^2\phi_{b,x}\rangle+\langle\cos^2\phi_{c,x}\rangle=1 \quad (11)$$

$$\langle\cos^2\phi_{a,y}\rangle+\langle\cos^2\phi_{b,y}\rangle+\langle\cos^2\phi_{c,y}\rangle=1 \quad (12)$$

$$\langle\cos^2\phi_{a,z}\rangle+\langle\cos^2\phi_{b,z}\rangle+\langle\cos^2\phi_{c,z}\rangle=1. \quad (13)$$

These restrictions are satisfied if, and only if, the center of gravity of the triangle *abc* is at $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$, coinciding with the center of gravity of the large triangle XYZ. If two of the points (e.g., *a* and *b*) are known, the third may be plotted by a simple graphical construction, shown in Fig. 4. The procedure is as follows: draw the line *ab* and find its midpoint *m*. Draw another line *mw*, from *m* to the center of gravity *w*, and extend this line beyond *w* a distance 2 *mw*. This will be the point *c*.

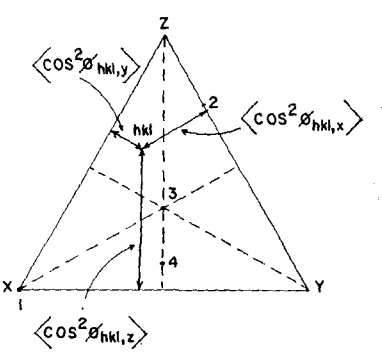

FIG. 3. Location of the [*hkl*] orientation function in the equilateral triangle plot.

This construction may be algebraically verified as follows: let m_z be the z coordinate of m. Since m is the midpoint of ab, m_z is given by

$$m_z=(\langle\cos^2\phi_{a,z}\rangle+\langle\cos^2\phi_{b,z}\rangle)/2. \tag{14}$$

By the method of construction, the z component of point c is given by

$$c_z=m_z+3(w_z-m_z). \tag{15}$$

Since $w_z=\frac{1}{3}$, this reduces to

$$c_z=1-2m_z. \tag{16}$$

Combining with Eqs. (14) and (15),

$$c_z=1-\langle\cos^2\phi_{a,z}\rangle-\langle\cos^2\phi_{b,z}\rangle=\langle\cos^2\phi_{c,z}\rangle. \tag{17}$$

Identical derivatives for c_x and c_y show that the coordinates of c are $(\langle\cos^2\phi_{c,x}\rangle, \langle\cos^2\phi_{c,y}\rangle, \langle\cos^2\phi_{c,z}\rangle)$, completing the proof.

Once the points a, b, and c are located, the point for any $[hkl]$ crystallographic direction in an orthorhombic unit cell may be easily plotted. This may be done either algebraically or graphically. The algebraic method would be to use the following equation, which is derived in Sec. IIIC2:

$$\langle\cos^2\phi_{hkl,q}\rangle=D^{-2}(h^2a^{-2}\langle\cos^2\phi_{a,q}\rangle+k^2b^{-2}\langle\cos^2\phi_{b,q}\rangle+l^2c^{-2}\langle\cos^2\phi_{c,q}\rangle), \tag{18}$$

where

$$D=(h^2a^{-2}+k^2b^{-2}+l^2c^{-2})^{-\frac{1}{2}}. \tag{19}$$

Alternatively, one may also plot the point hkl by a graphical construction. By examining Eq. (18), it is evident that the coordinates of the point hkl are the same as the coordinates of the center of gravity of a three point mass system, where the masses m_a, m_b, and m_c at the points a, b, and c are unequal:

$$m_a=h^2/a^2 \tag{20}$$

$$m_b=k^2/b^2 \tag{21}$$

$$m_c=l^2/c^2. \tag{22}$$

The center of gravity can be found as follows: mark three points r, s, and t, as shown in Fig. 5, so that the following ratios are obtained:

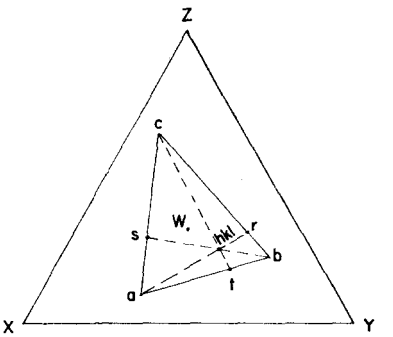

FIG. 5. Locating point $[hkl]$ from the locations of points a, b, and c.

$$rc/rb=(k^2b^{-2})/(l^2c^{-2})=m_b/m_c \tag{23}$$

$$sc/sa=(h^2a^{-2})/(l^2c^{-2})=m_a/m_c \tag{24}$$

$$tb/ta=(h^2a^{-2})/(k^2b^{-2})=m_a/m_b. \tag{25}$$

The center of gravity is the common intersection of ar, bs, and ct. The proof is as follows: r is the center of gravity of the two-mass system $b+c$. Hence, we may replace m_b and m_c with a single mass (m_b+m_c) at r. The center of gravity of the three-mass system a, b, c is identical with that of the two-mass system $a+r$, and therefore must lie on line ar. Similarly, the center of gravity must be on lines bs and ct; hence it is the common intersection of all three lines. This construction can also be verified algebraically.

A more definite interpretation may now be given to the sides of the triangle abc. The sides ab, ac, and bc are the locus of all points of the form $hk0$, $h0l$, and $0kl$, respectively. The straight line ar from the vertex a through the point hkl is the locus of points having a constant k/l ratio. Similarly, h/k is constant along ct, and h/l along bz.

The entire triangle abc may be constructed for polyethylene from a axis and [110] orientation data. The procedure is shown in Fig. 6: draw a line at from the point a through t of the [110] point. Extend this line to the point b, where the ratio at/bt is given by

$$at/bt=a^2/b^2=2.25. \tag{26}$$

Once point b is located, c may be found by the method previously described.

Alternately, one may calculate the a, b, and c axis orientation functions from [110] and [200] data, using methods described in Sec. IIIC. This is probably more accurate than graphical construction. However, a knowledge of how the trigonometric formulas can be translated into graphical construction would probably give one a greater insight into the workings of the orientation function diagram.

This method of plotting can be used for crystal systems other than the orthorhombic system. However, in such cases the simple geometric relations involving the triangle abc may no longer be valid. Three difficulties

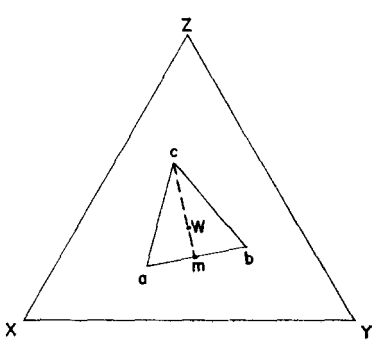

FIG. 4. Orientation function plot of a hypothetical polyethylene sample.

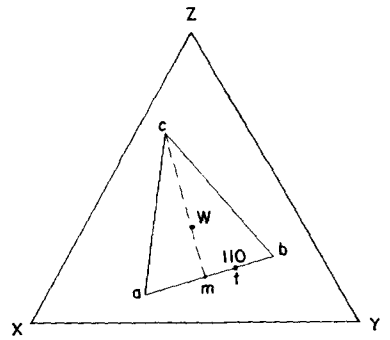

FIG. 6. Construction of triangle *abc* from points *a* and [110].

occur: (1) The crystal axes may not be normal to crystallographic planes and may not be mutually perpendicular. (2) In systems of low symmetry, i.e., the triclinic and monoclinic systems, planes differing only by the signs of h, k, and l may not be symmetrically equivalent and must be plotted separately. (3) In systems of higher symmetry, such as the tetragonal, hexagonal, or cubic systems, the symmetry may act to reduce the triangle *abc* to a simpler form. For instance, in the tegragonal system, the *a* and *b* axes are equivalent so that the separate data cannot be obtained for $[hk0]$ and $[kh0]$ planes. This causes the triangle *abc* to reduce to a line, with points *a* and *b* coalescing. The same effect occurs in the hexagonal system. In the cubic system, the high symmetry reduces the triangle *abc* to a single point $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$, in effect rendering the method useless.

This degeneration of the triangle *abc* results from the presence of different $[hkl]$ planes which have the same spacing because of the symmetry of the unit cell. This point is more fully developed in Sec. IIIC.

C. Evaluation of Orientation Functions by Indirect Means in Polycrystalline Materials

1. Effect of Crystal Symmetry

In considering orientation determinations with materials in the orthorhombic, hexagonal, tetragonal, or cubic systems, one must take multiplicity into account. A given $[hkl]$ crystal plane may be related by symmetry operations to a number of other $[hkl]$ planes, all having the same spacing d_{hkl}, the same Bragg angle for diffraction, and the same structure factor for scattering x rays. Let us consider a particular $[hkl]$ plane having multiplicity m; i.e., there are a total of m distinct planes in every unit cell related by the above-mentioned symmetry. In a polycrystalline sample the intensities diffracted by different $[hkl]$ planes having the same spacing are experimentally inseparable, so at any sample orientation (ω,χ) the intensity is a sum over m different components:

$$I_{hkl}(\omega,\chi)=\sum_{i=1}^{m} I_{hkl}{}^{i}(\omega,\chi). \tag{27}$$

Here we define $I_{hkl}{}^{i}(\omega,\chi)$ to be one of the m components of intensity, attributable to diffraction from a particular one of the m different planes which could diffract at the Bragg angle in question. Of course, for a particular crystallite, only one such plane could be in the proper orientation to diffract. However, we have no means of knowing which type of plane it is so we must sum over all types. We denote the total intensity by leaving off the superscript i as in Eq. (27).

The experimentally measured orientation function for the $[hkl]$ reflection is given by

$$\langle\cos^2\phi_{hkl,q}\rangle=\frac{\int_{-\pi/2}^{\pi/2}\int_{0}^{2\pi} I_{hkl}(\omega,\chi)\cos^2\phi_{hkl,q}\cos\omega d\omega d\chi}{\int_{-\pi/2}^{\pi/2}\int_{0}^{2\pi} I_{hkl}(\omega,\chi)\cos\omega d\omega d\chi}. \tag{28}$$

The reference direction q may be any arbitrary direction in the sample. For the sake of brevity we use

$$d\tau=\cos\omega d\omega d\chi. \tag{29}$$

Also, we omit the limits of integration. Substituting (27) and (28) and interchanging the integral and summation signs we obtain

$$\langle\cos^2\phi_{hkl,q}\rangle=\frac{\sum_{i=1}^{m}\int\int I_{hkl}{}^{i}(\omega,\chi)\cos^2\phi_{hkl,q}d\tau}{\sum_{i=1}^{m}\int\int I_{hkl}{}^{i}(\omega,\chi)d\tau}. \tag{30}$$

Because the m different planes have the same scattering power, the total intensity scattered from each of them is equal:

$$\int\int I_{hkl}{}^{i}(\omega,\chi)d\tau=\int\int I_{hkl}{}^{j}(\omega,\chi)d\tau \tag{31}$$

for $i=1$ to m and $j=1$ to m. Thus,

$$\langle\cos^2\phi_{hkl,q}\rangle=\frac{\sum_{i=1}^{m}\int\int I_{hkl}{}^{i}(\omega,\chi)\cos^2\phi_{hkl,q}d\tau}{m\int\int I_{hkl}{}^{j}(\omega,\chi)d\tau} \tag{32}$$

and

$$\langle\cos^2\phi_{hkl,q}\rangle=\frac{1}{m}\sum_{i=1}^{m}\frac{\int\int I_{hkl}{}^{i}(\omega,\chi)\cos^2\phi_{hkl,q}d\tau}{\int\int I_{hkl}{}^{j}(\omega,\chi)d\tau}. \tag{33}$$

But by Eq. (31), we may replace the superscript j in the denominator with an i.

$$\langle\cos^2\phi_{hkl,q}\rangle=\frac{1}{m}\sum_{i=1}^{m}\frac{\int\int I_{hkl}{}^{i}(\omega,\chi)\cos^2\phi_{hkl,q}d\tau}{\int\int I_{hkl}{}^{i}(\omega,\chi)d\tau}. \tag{34}$$

Let us define

$$\langle\cos^2\phi_{hkl,q}{}^{i}\rangle=\frac{\int\int I_{hkl}{}^{i}(\omega,\chi)\cos^2\phi_{hkl,q}d\tau}{\int\int I_{hkl}{}^{i}(\omega,\chi)d\tau}. \tag{35}$$

This quantity is the orientation function that would be measured if the ith plane were diffracting and the other $(m-1)$ planes were silent. By substitution into Eq. (34), we obtain

$$\langle\cos^2\phi_{hkl,q}\rangle=\frac{1}{m}\sum_{i=1}^{m}\langle\cos^2\phi_{hkl,q}{}^{i}\rangle. \tag{36}$$

Equation (36) is found useful in the next section of this paper.

According to Sack,[22] the maximum number of diffraction planes having independent orientation functions is two for the orthorhombic system, one for the hexagonal and tetragonal systems, and zero for the cubic system. In this work we consider the dependency relationships between the orientation functions of an $[hkl]$ plane in general and the a, b, and c axis orientation functions for all these systems.

2. *Orientation Functions in the Orthorhombic System*

Let us designate $\mathbf{u}_q$ as a unit vector in some arbitrary reference direction q in the sample. We do not further specify $\mathbf{u}_q$ and require no particular symmetry (except the unavoidable center of inversion) of the orientation distribution with respect to the macroscopic sample coordinates. Also, we define $\mathbf{u}_a$, $\mathbf{u}_b$, $\mathbf{u}_c$, and $\mathbf{u}_{hkl}$ to be unit vectors in the direction of the a, b, and c axes and the $[hkl]$ normal, respectively, for a single orthorhombic crystalline, as shown in Fig. 7.

In addition let ϕ_{mn} be the angle between vectors $\mathbf{u}_i$ and $\mathbf{u}_j$, where i and j are variable dummy subscripts. We use only the quantity $\cos\phi_{ij}$, so the order of the subscripts is immaterial, since

$$\phi_{ij}=-\phi_{ji} \tag{37}$$

and

$$\cos\phi_{ij}=\cos(-\phi_{ji})=\cos\phi_{ji}. \tag{38}$$

We may decompose vectors $\mathbf{u}_q$ and $\mathbf{u}_{hkl}$ into a, b, and c axis components as follows:

$$\mathbf{u}_q=\sum_{i=a,b,c}\cos\phi_{i,q}\mathbf{u}_i \tag{39}$$

$$\mathbf{u}_{hkl}=\sum_{j=a,b,c}\cos\phi_{j,hkl}\mathbf{u}_j. \tag{40}$$

Consider the dot product of these two vectors:

$$\mathbf{u}_q\cdot\mathbf{u}_{hkl}=\sum_{i=a,b,c}\sum_{j=a,b,c}\cos\phi_{i,q}\cos\phi_{j,hkl}(\mathbf{u}_i\cdot\mathbf{u}_j). \tag{41}$$

Since the a, b, and c crystal axes are orthogonal, we may write

$$\mathbf{u}_i\cdot\mathbf{u}_j=\delta_{ij}, \tag{42}$$

where δ_{ij} is the Kronecker delta, resulting in

$$\mathbf{u}_{hkl}\cdot\mathbf{u}_q=\sum_{j=a,b,c}\cos\phi_{j,hkl}\cos\phi_{j,q}. \tag{43}$$

[22] R. A. Sack, J. Polymer Sci. **54**, 543 (1961).

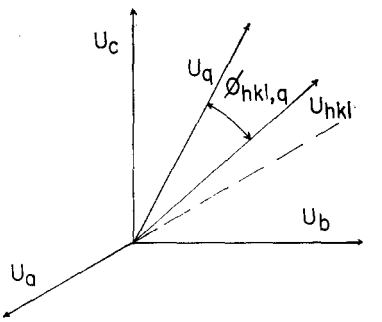

FIG. 7. Orientation of the orthorhombic unit cell with respect to reference direction Q.

But the dot product can also be expressed as

$$\mathbf{u}_{hkl}\cdot\mathbf{u}_q=\cos\phi_{hkl,q} \tag{44}$$

which gives

$$\cos\phi_{hkl,q}=\sum_{j=a,b,c}\cos\phi_{j,hkl}\cos\phi_{j,q}. \tag{45}$$

Squaring and averaging over all crystallites, we obtain

$$\langle\cos^2\phi_{hkl,q}{}^{i}\rangle=\sum_{j=a,b,c}\sum_{n=a,b,c}\cos\phi_{j,hkl}\cos\phi_{n,hkl}\times\langle\cos\phi_{j,q}\cos\phi_{n,q}\rangle. \tag{46}$$

The superscript i is added to emphasize that this average refers to just one of the m equivalent planes. According to Wilchinsky,[23] the symmetry of the orthorhombic unit cell should cause the cross terms from the m planes to cancel each other. We may obtain the experimentally measured orientation function by summing as indicated in (36). We then obtain

$$\langle\cos^2\phi_{hkl,q}\rangle=\tfrac{1}{4}\sum_{h=\pm h}\sum_{k=\pm k}\langle\cos^2\phi_{hkl,q}{}^{i}\rangle. \tag{47}$$

Substituting Eq. (46) into (47) and rearranging the terms,

$$\langle\cos^2\phi_{hkl,q}\rangle=\tfrac{1}{4}\sum_{j=a,b,c}\sum_{n=a,b,c}\langle\cos\phi_{j,q}\cos\phi_{n,q}\rangle\times\sum_{h=\pm h}\sum_{k=\pm k}\cos\phi_{j,hkl}\cos\phi_{n,hkl}. \tag{48}$$

We now wish to show that

$$\sum_{h=\pm h}\sum_{k=\pm k}\cos\phi_{j,hkl}\cos\phi_{n,hkl}$$

vanishes for $n\neq j$. Since $n\neq j$, we may assume that either n or j is not equal to c. Assume without loss of generality that $n\neq c$. There are only two possibilities, $n=a$ or $n=b$.

Case 1. $n=a$, then

$$\cos\phi_{n,hkl}=\cos\phi_{a,hkl}=h/aD, \tag{49}$$

[23] Z. W. Wilchinsky, in *Advances in X-Ray Analysis*, W. M. Mueller and M. Fay, Eds. (Plenum Press, Denver, 1963), Vol. 6, p. 231.

where

$$D=[(h/a)^2+(k/b)^2+(l/c)^2]^{\frac{1}{2}}. \quad (50)$$

Then

$$\sum_{h=\pm h}\sum_{k=\pm k}\cos\phi_{j,hkl}\cos\phi_{n,hkl} = \sum_{k=\pm k}\cos\phi_{j,hkl}\sum_{h=\pm h}\cos\phi_{n,hkl}. \quad (51)$$

By substitution,

$$\sum_{h=\pm h}\sum_{k=\pm k}\cos\phi_{j,hkl}\cos\phi_{n,hkl} = \sum_{k=\pm k}\cos\phi_{j,hkl}\sum_{h=\pm h}(h/aD)=0. \quad (52)$$

Case 2. $n=b$, then

$$\cos\phi_{n,hkl}=\cos\phi_{b,hkl}=k/bD; \quad (53)$$

$$\sum_{h=\pm h}\sum_{k=\pm k}\cos\phi_{j,hkl}\cos\phi_{n,hkl} = \sum_{h=\pm h}\cos\phi_{j,hki}\sum_{k=\pm k}(k/bD)=0. \quad (54)$$

Since all cross terms are found to vanish, we may rewrite Eq. (48) as follows:

$$\langle\cos^2\phi_{hkl,q}\rangle=\tfrac{1}{4}\sum_{j=a,b,c}\langle\cos^2\phi_{j,q}\rangle\cdot 4\cos^2\phi_{j,hkl} \quad (55)$$

which simplifies to

$$\langle\cos^2\phi_{hkl,q}\rangle=\cos^2\phi_{a,hkl}\langle\cos^2\phi_{a,q}\rangle+\cos^2\phi_{b,hkl}\langle\cos^2\phi_{b,q}\rangle +\cos^2\phi_{c,hkl}\langle\cos^2\phi_{c,q}\rangle \quad (56)$$

or

$$\langle\cos^2\phi_{hkl,q}\rangle=D^{-2}(h^2a^{-2}\langle\cos^2\phi_{a,q}\rangle+k^2b^{-2}\langle\cos^2\phi_{b,q}\rangle +l^2c^{-2}\langle\cos^2\phi_{c,q}\rangle). \quad (57)$$

These two equations are alternate forms of the desired dependency relationship in the orthorhombic system. This equation, in combination with the orthogonality relationship

$$\langle\cos^2\phi_{a,q}\rangle+\langle\cos^2\phi_{b,q}\rangle+\langle\cos^2\phi_{c,q}\rangle=1, \quad (58)$$

restricts the number of planes which may have independent orientation functions to two.

Let us consider the case $[hkl]=[110]$. By Eq. (57) we obtain

$$\langle\cos^2\phi_{110,q}\rangle=(a^2+b^2)^{-1}(b^2\langle\cos^2\phi_{a,q}\rangle+a^2\langle\cos^2\phi_{b,q}\rangle). \quad (59)$$

Solving for the b-axis orientation functions:

$$\langle\cos^2\phi_{b,q}\rangle=\sec^2\rho_{110}\langle\cos^2\phi_{110,q}\rangle-\tan^2\rho_{110}\langle\cos^2\phi_{a,q}\rangle, \quad (60)$$

where ρ_{110} is the angle between the b axis and the [110] normal. Combining this equation with the orthogonality condition (11) we obtain an expression for the c-axis orientation function:

$$\langle\cos^2\phi_{c,q}\rangle=1-(1-\tan^2\rho_{110})\langle\cos^2\phi_{a,q}\rangle -\sec^2\rho_{110}\langle\cos^2\phi_{110,q}\rangle. \quad (61)$$

Equation (61) is identical to the result obtained by using Wilchinsky's[1] formula.

3. *Orientation Functions in the Tetragonal System*

In treating the tetragonal and hexagonal systems it is convenient to make use of the mathematical formulation of Roe.[24] In this method the solution of the pole figure inversion problem is obtained by expanding the distribution functions for various diffracting planes as a series of associated Legendre polynomials. The relations between the orientation functions are obtained from the Roe formulation by considering the second spherical harmonic term of the expansion. For the sake of brevity we do not explain the Roe formulation further, but indicate which identities we are using from the Roe paper and then translate them into the nomenclature used here. The basic identity to be used is Eq. (13) of Roe's paper, here denoted as (62):

$$Q_{lm}{}^{i}=(2\pi)\sqrt{2}(2l+1)^{-\frac{1}{2}}\sum_{n=-l}^{l}W_{lmn}P_l{}^{n}(\Xi_i)e^{in\Phi_i}. \quad (62)$$

Setting $l=2$ and $m=0$, we translate this equation into our nomenclature as follows:

$$f_{hkl,q}=4\pi^2(\tfrac{2}{5})^{\frac{1}{2}}\sum_{n=-2}^{2}W_{20n}f_{hkl,c}e^{in\Phi_i}, \quad (63)$$

where

$$f_{hkl,q}=\tfrac{3}{2}\langle\cos^2\phi_{hkl,q}\rangle-\tfrac{1}{2} \quad (64)$$

and

$$f_{hkl,c}=\tfrac{3}{2}\cos^2\phi_{hkl,c}-\tfrac{1}{2}. \quad (65)$$

The problem of evaluating the constants W_{20n} are greatly simplified by the symmetry of the tetragonal unit cell. The tetragonal system contains a fourfold symmetry axis parallel to the c axis. According to Roe,[24] this results in

$$W_{lmn}=0 \quad (66)$$

unless

$$n=4k; \quad k=0, \pm1, \pm2, \text{ etc.} \quad (67)$$

Thus, in Eq. (63), the summation reduces to one term since only W_{200} does not vanish:

$$f_{hkl,q}=4\pi^2(\tfrac{2}{5})^{\frac{1}{2}}W_{200}f_{hkl,c}. \quad (68)$$

The constant W_{200} is evaluated by setting $[hkl]=[001]$, recalling that $f_{001,c}=1$:

$$f_{c,q}=4\pi^2(\tfrac{2}{5})^{\frac{1}{2}}W_{200}. \quad (69)$$

Combining (68) and (69) we obtain:

$$f_{hkl,q}=f_{c,q}f_{hkl,c}. \quad (70)$$

After suitable manipulation, this equation may be expressed as follows:

$$\langle\cos^2\phi_{hkl,q}\rangle=\cos^2\phi_{hkl,c}\langle\cos^2\phi_{c,q}\rangle +\tfrac{1}{2}\sin^2\phi_{hkl,c}\langle\sin^2\phi_{c,q}\rangle. \quad (71)$$

[24] R.-J. Roe, J. Appl. Phys. **36**, 2024 (1965).

Equation (70) or (71) provides a means of calculating the orientation functions of any $[hkl]$ plane from experimental data on one such plane. Two specific cases are of especial interest, the [001] and the $[hk0]$ reflections. In the former case Eq. (70) results in

$$f_{001,q}=f_{c,q} \tag{72}$$

while in the latter case

$$f_{hk0,q}=-\tfrac{1}{2}f_{c,q}. \tag{73}$$

The value of $f_{hkl,q}$ for all other cases must lie between these two limits.

4. Orientation Functions in the Hexagonal System

In the hexagonal system, as in the tetragonal, it is convenient to use the Roe[24] formulation of the problem. Since the hexagonal unit cell possesses sixfold symmetry about the c axis we may write

$$W_{lmn}=0 \tag{74}$$

except when

$$n=6k;\quad k=0,\ \pm1,\ \pm2\cdots. \tag{75}$$

As in the tetragonal case we obtain

$$f_{hkil,q}=f_{c,q}f_{hkil,c} \tag{76}$$

or

$$\langle\cos^2\phi_{hkil,q}\rangle=\cos^2\phi_{hkil,c}\langle\cos^2\phi_{c,q}\rangle +\tfrac{1}{2}\sin^2\phi_{hkil,c}\langle\sin^2\phi_{c,q}\rangle, \tag{77}$$

where $[hkil]$ are the Miller–Bravais indices. Again, all possible orientation functions may be calculated from data on one $[hkil]$ plane, the limiting cases are

$$f_{0001,q}=f_{c,q} \tag{78}$$

and

$$f_{hki0,q}=-\tfrac{1}{2}f_{c,q} \tag{79}$$

and all other $[hkil]$ orientation functions are bounded by these two limits.

5. Orientation Functions in the Cubic System

We now show that orientation functions are not useful in the cubic system, since all orientation functions must identically be equal to the random orientation value:

$$\langle\cos^2\phi_{hkl,q}\rangle=\tfrac{1}{3} \tag{80}$$

or

$$f_{hkl,q}=0. \tag{81}$$

This result follows immediately from the treatment of the pole figure inversion for cubic systems given by Roe.[25] In this paper Roe concludes that for cubic systems the high symmetry of the unit cell results in

$$W_{200}=0. \tag{82}$$

[25] R.-J. Roe, J. Appl. Phys. **37**, 2069 (1966).

The fourfold symmetry about c causes W_{201}, W_{202}, $W_{20\bar{1}}$, and $W_{20\bar{2}}$ to vanish as in the tetragonal case. We may apply Eq. (63), and since all values of W_{20n} vanish, we are left with

$$f_{hkl,q}=0 \tag{83}$$

or

$$\langle\cos^2\phi_{hkl,q}\rangle=\tfrac{1}{3}. \tag{84}$$

IV. DETERMINATION OF CRYSTALLINE BIREFRINGENCE AND INFRARED DICHROISM FROM POLE FIGURE DATA

A. Calculation of Birefringence

Using the two-phase model of polymer morphology, the birefringence of a uniaxially oriented polymer sample may be regarded as the sum of three terms arising from crystalline, amorphous, and form birefringence[22]:

$$\Delta=x_{c,v}\Delta_c+(1-x_{c,v})\Delta_A+\Delta_f, \tag{85}$$

where $x_{c,v}$ = volume fraction crystallinity, Δ_c = crystallite birefringence, Δ_A = amorphous birefringence, and Δ_f = form birefringence. We may apply this equation to biaxially oriented films provided we distinguish between the two independent birefringences Δ_x and Δ_y, defined in Eqs. (1) and (2). We obtain

$$\Delta_x=x_{c,v}\Delta_{c,x}+(1-x_{c,v})\Delta_{A,x}+\Delta_{f,x} \tag{86}$$

and

$$\Delta_y=x_{c,v}\Delta_{c,y}+(1-x_{c,v})\Delta_{A,y}+\Delta_{f,y}. \tag{87}$$

We now concern ourselves with the calculation of the quantities $\Delta_{c,x}$ and $\Delta_{c,y}$, which we call the "crystalline birefringences." The reader should keep in mind that these quantities are not the birefringences of a single crystal, but instead depend upon the degree of orientation of the crystals.

The unit cell of any crystal form has associated with it a polarizability ellipsoid, characterized by three principle polarizabilities for light polarized in three mutually perpendicular directions. Depending upon the symmetry of the crystal system these directions may or may not coincide with specific crystallographic directions. Let us consider a polycrystalline film sample, having associated with it a Cartesian coordinate system defined as follows: x is the film normal, z is a preferred direction (machine or stretch direction) in the plane of the film, and y, the transverse direction, lies in the plane of the film perpendicular to both x and z. The indices of refraction of the collection of crystals in these three directions are given by Stein[8]:

$$n_{x,c}=n_a\langle\cos^2\phi_{a,x}\rangle+n_b\langle\cos^2\phi_{b,x}\rangle+n_c\langle\cos^2\phi_{c,x}\rangle \tag{88}$$

$$n_{y,c}=n_a\langle\cos^2\phi_{a,y}\rangle+n_b\langle\cos^2\phi_{b,y}\rangle+n_c\langle\cos^2\phi_{c,y}\rangle \tag{89}$$

$$n_{z,c}=n_a\langle\cos^2\phi_{a,z}\rangle+n_b\langle\cos^2\phi_{b,z}\rangle+n_c\langle\cos^2\phi_{c,z}\rangle. \tag{90}$$

n_a, n_b, and n_c are the indices of refraction of the unit cell in the three principle directions.

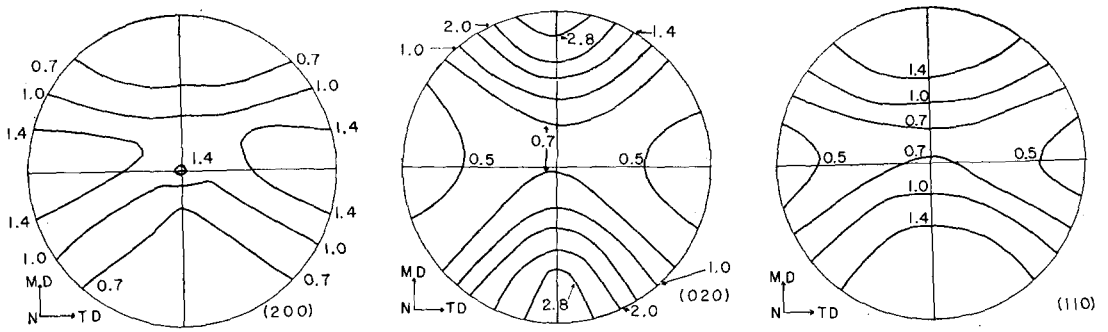

FIG. 8. Pole figure of unidirectionally recrystallized polyethylene film. In these and subsequent pole figures, there is a factor of $\sqrt{2}$ between the intensity levels at adjacent contour lines. Thus the intensity doubles at every second contour line.

Next, let us define two components of the crystalline birefringence as follows:

$$\Delta_{c,x}=n_{z,c}-n_{y,c}, \tag{91}$$

$$\Delta_{c,y}=n_{z,c}-n_{x,c}, \tag{92}$$

where $n_{x,c}$, $n_{y,c}$, and $n_{z,c}$ are the hypothetical indices of refraction which would be observed if the sample were 100% crystalline. By combining Eqs. (88) through (92), along with the orthogonality conditions (11) through (13), we obtain the following equations for $\Delta_{c,x}$ and $\Delta_{c,y}$:

$$\Delta_{c,x}=(n_a-n_c)(\langle\cos^2\phi_{a,z}\rangle-\langle\cos^2\phi_{a,y}\rangle)+(n_b-n_c)(\langle\cos^2\phi_{b,z}\rangle-\langle\cos^2\phi_{b,y}\rangle) \tag{93}$$

$$\Delta_{c,y}=(n_a-n_c)(\langle\cos^2\phi_{a,z}\rangle-\langle\cos^2\phi_{a,x}\rangle)+(n_b-n_c)(\langle\cos^2\phi_{b,z}\rangle-\langle\cos^2\phi_{b,x}\rangle). \tag{94}$$

In the orthorhombic, tetragonal, and hexagonal systems n_a, n_b, and n_c may be taken along the a, b, and c crystal axes. However, one must be careful in applying these equations to the monoclinic or triclinic systems since the directions of the principle polarizabilities need not coincide with crystal axes in these cases.

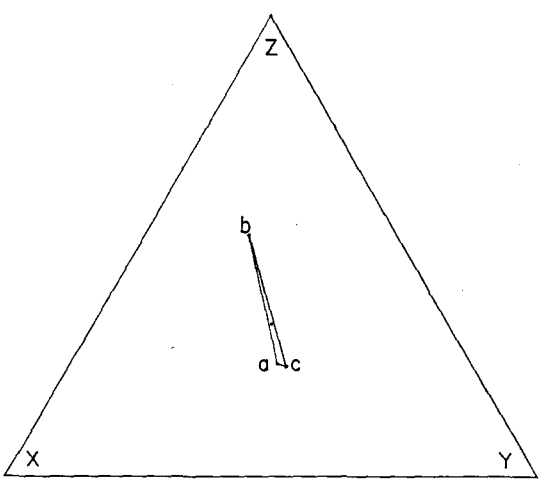

FIG. 9. Orientation function diagram—unidirectionally recrystallized polyethylene.

Equations (93) and (94) for $\Delta_{c,x}$ and $\Delta_{c,y}$ are preferable to those previously obtained[8] since they are derived with no assumptions concerning the form of the orientation distribution.

B. Calculation of Infrared Dichroism

Let us consider a crystalline infrared absorption band whose transition moment lies along (for example) the b axis of the unit cell. Using the previous nomenclature[8] we may define the dichroism D_A (at normal incidence) as follows:

$$D_A(0^0)=(a_c^0)_{A\parallel}/(a_c^0)_{A\perp}. \tag{95}$$

The two absorbancies are given by Stein[8]:

$$(a_c^0)_{A\parallel}=K_b\langle\cos^2\phi_{b,z}\rangle \tag{96}$$

$$(a_c^0)_{A\perp}=K_b\langle\cos^2\phi_{b,y}\rangle. \tag{97}$$

Thus the dichroism D_A is given by

$$D_A=\langle\cos^2\phi_{b,z}\rangle/\langle\cos^2\phi_{b,y}\rangle. \tag{98}$$

In analogy to the birefringence case, we may define two independent dichroisms for a biaxially oriented film. However, the second dichroism is difficult to measure so we do not treat this case.

V. EXPERIMENTAL RESULTS

A. Crystal Orientation Measurements by X-Rays

1. *Unidirectionally Recrystallized Polyethylene Film*

This film was prepared by passing a sheet of compression-molded polyethylene between two sets of heaters, one set kept above and the second set below the melting temperature of polyethylene. The actual conditions of

TABLE I. Conditions of preparation: unidirectionally recrystallized polyethylene.

Heater temperature	185°C
Cooler temperature	13°C
Separation of heater and cooler	1.0 mm
Sample speed	1.3 mm/h

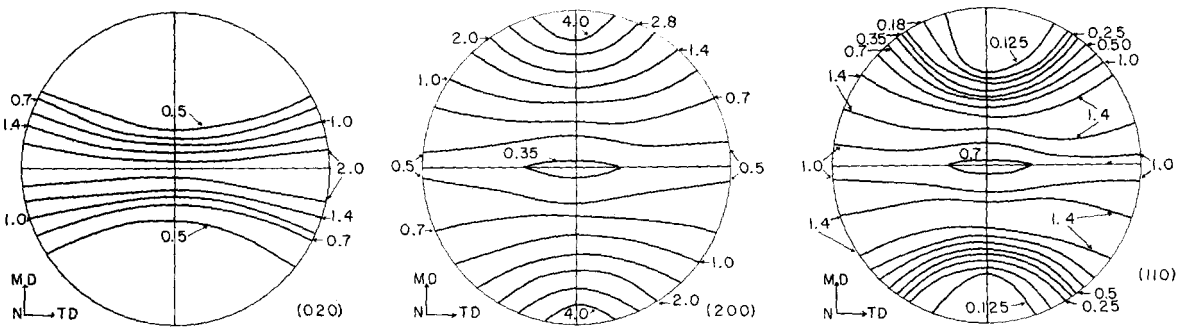

FIG. 10. Pole figures of stretched and recrystallized polyethylene (low elongation).

preparation of this sample are shown in Table I. By this technique, previously discussed by Hearle[26] and demonstrated by Takayanagi,[27] it is possible to produce orientation by constraining the crystal aggregates to grow in one particular direction, the direction of maximum temperature gradient, in preference to other directions. Once a spherulite starts growing, different regions of the spherulite tend to grow at different rates, due to the temperature gradient across the spherulite. In addition, one side of the spherulite is growing into uncrystallized material, while the other side soon impinges on previously crystallized material. The net result of both effects is to give a highly eccentric shape to the spherulites.

The pole figures for the [200], [020], and [110] planes are shown in Fig. 8, and the orientation functions in Fig. 9. The [020] pole figure shows a definite tendency for the *b* axis to orient parallel to the machine direction. The [110] pole figure shows a mild orientation of this reciprocal lattice vector towards the machine direction, and the [200] pole figure shows a tendency for the *a* axis to concentrate in the equatorial regions. The three pole figures do not exhibit the usual biaxial symmetry, i.e., the distributions are not characterized by three mutually perpendicular symmetry planes. There is symmetry with respect to reflection across the vertical diameter of the pole figure, but not with respect to the horizontal diameter. This asymmetry is probably due to a small component of temperature gradient in the direction of the film normal. It is interesting that Cullity[28] notes that exactly the same type of asymmetry occurs when a metal sheet is rolled over and over in the same direction, without reversing end-for-end between passes.

The resulting orientation function diagram is shown in Fig. 9. The *b* axis shows a fairly strong orientation parallel to the machine direction. This reflects the fact that the *b* axis is the radial direction of a growing spherulite, and the spherulites are elongated in the machine direction. The *a*- and *c*-axis orientation functions are nearly identical; the difference could be due to experimental error. This would agree with a model in which these two axes are randomly oriented about the *b* axis. The asymmetry of the pole figures shows up in the orientation function diagram as a small deviation from uniaxial symmetry about *z*. The *b* axis shows a slight tendency to orient towards the film normal rather than the transverse direction. The *a* axis exhibits the opposite tendency.

2. *Crosslinked, Melted, Stretched, and Recrystallized Polyethylene*

In order to study the effect of amorphous orientation on crystallization, two polyethylene samples were crosslinked with a small dose of beta radiation, melted, stretched in the molten state, and then crystallized. The degree of crosslinking was sufficient to prevent the molten polymer from flowing. Two samples were prepared in this manner, one crystallized at low elongation (100%) and one at high elongation (200%). The conditions of preparation of these samples are listed in Table II.

The pole figures obtained for these two samples are shown in Figs. 10 and 11. Both samples show strong orientation of the *b* axis perpendicular to the stretch direction. However, in the high-elongation sample the *b* axes are not evenly distributed along the equator of the pole figure but are concentrated near the transverse direction. The [110] pole figures show a similar trend. In both samples, the [110] normal vectors concentrate at an angle of approximately 65°–75° to the stretch direction. However, in the low-elongation sample the distribution is nearly uniaxial, while in the high-elongation case there is a tendency for the [110] vectors to concentrate towards the plane of the film.

TABLE II. Conditions of preparation: crosslinked, melted, stretched, and recrystallized polyethylene samples.

Melt temperature	160°C
Crystallization temperature	104°C
Low-elongation sample: dose = 15 mR, $l/l_0 = 2$	
High-elongation sample: dose = 30 mR, $l/l_0 = 3$	

[26] J. W. S. Hearle, J. Appl. Polymer Sci. 7, 1193 (1963).
[27] M. Takayanagi, Mem. Fac. Eng. Kyushu Univ. 23, 41 (1963).
[28] B. D. Cullity, *Elements of X-Ray Diffraction* (Addison-Wesley Publ. Co., Reading, Mass., 1956), p. 293.

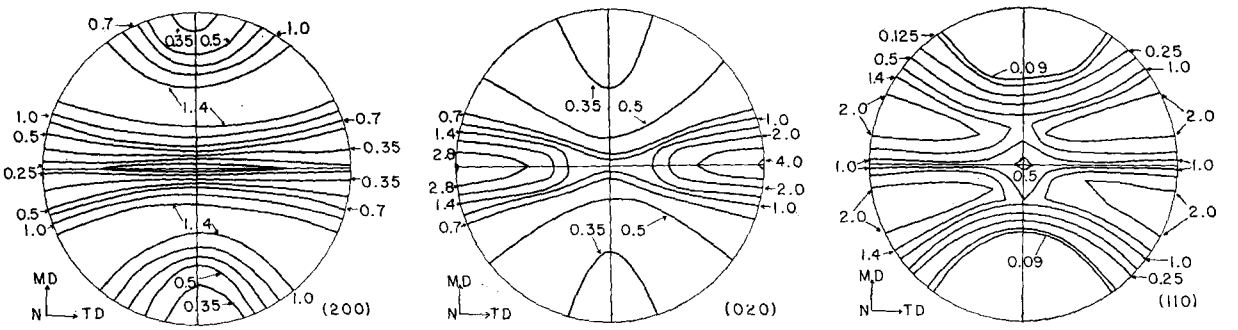

FIG. 11. Pole figures of stretched and recrystallized polyethylene (high elongation).

The most striking differences are in the *a*-axis pole figures. The *a*-axis distribution is nearly uniaxial in both samples, but in the low-elongation sample the *a* axes are oriented towards the stretch direction, while in the high-elongation sample the *a* axis is tilted at an angle of about 50° to the stretch direction. These results agree, in general, with previous results[29,30] obtained by flat-film photograph and azimuthal scans on similar samples. Such techniques do not, however, indicate biaxial orientation.

The orientation function diagram (Fig. 12) shows nearly uniaxial orientation for the low-elongation sample and considerable biaxial character for the high-elongation sample. The latter may be due to a small component of tension in the transverse direction during the process of stretching and crystallization. The diagram also indicates that the *c* axis is oriented closer to the stretch direction in the high-elongation sample than in the low-elongation sample. No indication of the actual *c*-axis distributions are given, but in the case of the high-elongation sample we may presume that the *c* axis has a rather sharp distribution, as do the *a* and *b* axes. Judging from the orientation functions, the *c* axis may be tilted at an angle of 30°–40° in this case. This combination of *a*, *b*, and *c* orientations places the [201] pole close to the stretch direction. For the low-elongation sample the *c*-axis distribution is probably not sharp, so there is no meaning to such an angle.

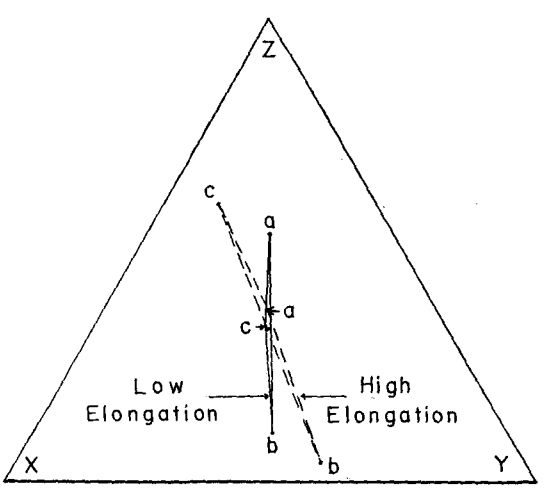

FIG. 12. Orientation function diagram—stretched and recrystallized polyethylene samples.

A model proposed by Kobayashi[31] could explain the pole figures obtained for the high-elongation sample. Kobayashi suggests that crystal growth may proceed by a screw dislocation mechanism; the *b* axis lies perpendicular to the screw axis and the *a* and *c* axes are tilted at fixed angles to the screw axis, the two angles being complements of each other. In our sample the screw axes appear to be oriented quite close to the stretch direction. The biaxial character could be explained by a small distortion of the structure in the transverse direction, caused by the transverse tension. The cross section of the resulting structures would then be, on the average, elliptical, the major axis of the ellipse lying in the transverse direction.

Let us consider two possible models for orientation in the low-elongation film: the row nucleation model, originally proposed by Keller,[32] and *a*-axis orientation, proposed by Holmes and Palmer.[33] In the row nucleation model the *b* axis is preferentially oriented into the plane perpendicular to *z*, and the *a* and *c* axes are randomly oriented about *b*. In the *a*-axis orientation model, the *a* axes have a preferred orientation towards the *z* direction, while *b* and *c* are randomly oriented about *a*. Thus, with the row nucleation model we must conclude that the *a*- and *c*-axis orientation functions will be identical, while the *a*-axis model would require that the *b*- and *c*-axis orientation functions be identical. It is obvious that neither model agrees with the x-ray data. Moreover, the birefringence Δ_x of the film is close to zero, while the row nucleation model would predict considerable positive birefringence and the *a*-axis model would predict a strong negative birefringence.

[29] J. T. Judge and R. S. Stein, J. Appl. Phys. **32**, 2357 (1961).
[30] W. R. Krigbaum and R.-J. Roe, J. Chem. Phys. **41**, 737 (1964).
[31] K. Kobayashi, in *Polymer Single Crystals*, P. H. Geil, Ed. (Interscience Publishers, New York, 1963), pp. 465–475.
[32] A. Keller, J. Polymer Sci. **15**, 31 (1955).
[33] D. R. Holmes and R. P. Palmer, J. Polymer Sci. **31**, 345 (1958).

The most promising model would be a helical morphology, another alternative proposed by Keller.[32] In order that the a, b, and c axes all have different orientation functions it is necessary to assume that none of the crystallographic axes are parallel to the helix axis. However, the model must be consistent with the pole figures as well, which show the a axis maximum in the stretch direction and the b axis maximum perpendicular to that direction. This data can be explained by assuming that the helix axes are not perfectly oriented parallel to z, but are distributed in a broad maximum centered about the z direction. Also, there could be a distribution of the pitch of the helix rather than one value. Both effects would smear out the a-axis distribution, giving maximum [200] intensity at z rather than at some preferred tilt angle.

These conclusions point up the great advantage of the orientation function method. Lindenmeyer and Lustig[5] have found that the [002] reflection is so weak that c-axis pole figures may be obtained only with difficulty in high-density polyethylene and not at all in low-density samples. Such a measurement was not possible for these samples, due to the low crystallinity. However, the c-axis orientation functions may be quantitatively obtained from data on the stronger [110] and [200] reflections, and may be used to decide between alternate models of morphology.

B. Crystalline and Amorphous Birefringence Values

Table III shows values of the crystalline and amorphous birefringences for a number of polyethylene samples. Crystalline birefringences were calculated from the [110] and [200] plane orientation functions and the amorphous values were obtained by difference, using Eqs. (86) and (87).

The two crosslinked, stretched, and recrystallized samples have rather high degrees of amorphous orientation, probably because of the crosslinks, which prevent the amorphous chains from disorienting. The low-elongation sample has a negative $\Delta_{A,x}$, indicating amorphous orientation away from the stretch direction. This is in agreement with previously published data.[29] For the high-elongation sample, the value of $\Delta_{A,y}$ is very large, showing a strong tendency for amorphous orientation parallel to the plane of the film. The quantity $\Delta_{A,x}$ is positive for this sample. Thus, in the high elongation sample, the most preferred direction of amorphous orientation is the stretch direction, and the least is the film normal.

TABLE III. Birefringence values for films prepared by oriented crysrallization.

Sample	$X_{c,v}$	Δ_x	$\Delta_{c,x}$	$\Delta_{A,x}$	Δ_y	$\Delta_{c,y}$	$\Delta_{A,y}$
Low elongation	0.41	−2	48	−37	a	5	a
High elongation	0.38	186	300	115	235	185	266

Unit of birefringence = 0.0001

[a] Measurement not possible.

ACKNOWLEDGMENTS

The authors wish to thank Hisashi Suzuki, who prepared the unidirectionally recrystallized films, and Stuart Clough, who prepared the crosslinked and oriented films. Also, we wish to acknowledge the assistance of the Research Computing Center at the University of Massachusetts and the MIT Computation Center for use of the CDC 3600 and IBM 7094 computers.

APPENDIX

AI. Correction of X-Ray Intensity

The corrections applied to the raw intensity data included corrections for polarization, background, incoherent scattering, absorption, and variation of scattering volume with tilt angle ω. The method of Krimm and Tobolsky[34] was used to estimate incoherent scattering. The equations for the absorption and scattering volume corrections are equivalent to those of Decker *et al.*,[13] except that we make an absolute correction rather than normalizing to $\omega=0$.

The amorphous, [110], and [200] data are corrected for overlap by the method used by Fujino *et al.*[15] The procedure is:

(1) Run a 2θ scan across these three peaks at $\omega=\chi=0$.

(2) Fit the corrected intensity curve to a series of three Lorentzian functions, adjusting the linewidth of the three peaks until agreement is reached with the experimental curve.

(3) Using these linewidths one may resolve the true intensity of each peak by solving a set of three simultaneous equations at every (ω,χ) position.

No overlap correction is necessary for the [020] peak.

AII. Averaging Out Orientation in Calculating Crystallinity from X-Ray Data

The method of Matthews *et al.*[18] for calculating crystallinity involves calculations with the integrated intensities under the amorphous, [110], and [200] curves. In an oriented sample we must integrate over two angles as well average out orientation. Assume the [hkl] peak has a Lorentzian line shape:

$$I_{hkl}(2\theta,\omega,\chi)=I_p(\omega,\chi)\cdot L(\Delta 2\theta,\beta_{hkl}). \qquad \text{(A1)}$$

L is the Lorentzian function and $I_p(\omega,\chi)$ is the peak intensity at a fixed orientation. When we perform a triple integration, the following is obtained when the integrals are evaluated:

$$\int\int\int I_{hkl}(2\theta,\omega,\chi)d2\theta d\tau=4\pi^2\beta_{hkl}\bar{I}_{hkl}. \qquad \text{(A2)}$$

[34] S. Krimm and A. V. Tobolsky, J. Polymer Sci. 7, 57 (1951).

The quantity $\bar{I}_{hkl}$ is the average peak intensity over the pole figure, and is a by-product of the orientation function calculation. Putting this into Matthews' formula (A3)

$$x_{c,w}/x_{A,w}=(a_{110}I_{110}+a_{200}I_{200})/a_{\text{amorph}}I_{\text{amorph}}, \quad \text{(A3)}$$

we obtain

$$x_{c,w}/x_{A,w}=(a_{110}\beta_{110}\bar{I}_{110}+a_{200}\beta_{200}I_{200})/ a_{\text{amorph}}\beta_{\text{amorph}}\bar{I}_{\text{amorph}}. \quad \text{(A4)}$$

AIII. Instrumental Considerations in Obtaining Pole Figure Data by the Decker Method

In order to obtain good data with this method it is necessary to use the proper instrumental conditions. The main problem is that the finite size of the scattering volume of the sample tends to broaden the diffracted beam. The effect is especially important when the angle between the plane of the sample and the incident x-ray beam is small.

Several steps may be taken to alleviate this difficulty: use a narrow incident beam, a thin sample, and a wide receiving slit. We use an incident beam of less than 0.1° horizontal divergence and limit our sample thicknesses to about 0.7 mm to keep the scattering volume small. The receiving slit is slightly over 1 mm wide. Such a wide slit would usually result in poor resolution, but we use an unusually large sample-to-receiving-slit distance, 280 mm. At this distance the receiving slit covers a width of 0.2°, viewed from the center of the diffractometer, resulting in acceptable resolution of the peaks. With these conditions the absorption correction is valid for the amorphous, [110], and [200] peaks from $\omega=-70°$ to $\omega=0°$ in the transmission region. For these peaks a good match is obtained on unoriented samples between the transmission and reflection regions.

With the [020] peak it was found necessary to apply an instrumental correction factor to the $\omega=90°$ datum point to obtain a match with transmission data obtained in the region from $\omega=-60°$ to $\omega=0°$. The $\omega=90°$ point corresponds to parafocusing geometry, the conditions under which instrumental broadening is minimized. In our experiment this focusing is undesirable since it occurs only at one ω value, giving an anomalously high peak intensity at $\omega=90°$, owing to the narrowing of the peak. Since we should be comparing integrated intensities, the instrumental correction factor F_{90} is

$$F_{90}=\beta_{90}/\beta_0, \quad \text{(A5)}$$

where β_{90} and β_0 are the linewidths at $\omega=90°$ and $\omega=0°$, respectively. Our value of 0.73 for F_{90} is found to check with this ratio.

It is reasonable to ask why this narrowing occurs only for the [020] plane. The reason is this: The [020] peak occurs at $2\Theta=36°$, much higher than the [110] and [200] Bragg angle. It is well known that the instrumental broadening decreases with increasing 2Θ in parafocusing geometry. This occurs because the x-ray beam spreads out too far along the surface of the sample at low 2Θ values. Thus, the focusing is negligible for the [110] and [200] peaks, but not for the [020].

On the Number of Self-Avoiding Walks

HARRY KESTEN
Department of Mathematics, Cornell University, Ithaca, New York
(Received 26 February 1963)

Let χ_n be the number of self-avoiding walks on the integral points in Euclidean d space and γ_n the number of n-stepped self-avoiding polygons. It is shown that $\chi_{n+2}/\chi_n - \beta^2$ and $\gamma_{2n+3}/\gamma_{2n+1} - \beta^2$ tend to zero as $n \to \infty$ where $\beta = \lim_{n\to\infty} \chi_n^{1/n}$. Asymptotic estimates for these differences are given. β is also characterized as the unique positive root of $\sum_1^\infty \lambda_k x^{-k} = 1$ where the λ_k are the number of certain k-stepped self-avoiding walks.

1. INTRODUCTION

THE self-avoiding walks which we consider are paths without double points on the integral points in d-dimensional space, in which each step takes us from an integral point to one of its $2d$ nearest neighbors. The interest in such walks arises, e.g., because for $d = 3$ they would give us an impression of the shape of the main chain of a polymer in solution.[1] Knowledge about the number of self-avoiding walks could be useful for the study of percolation processes[2] and there exist analogies between some problems on self-avoiding walks and the Ising model.[3] Self-avoiding walks on other lattices than the integral points in Euclidean space have also been considered, and our methods are applicable to some of these as well.

X will always stand for a point in d dimensions with integer coordinates, which are denoted by $X^{(1)}, \cdots, X^{(d)}$. To avoid trivialities we always take $d \geq 2$. The symbols $w[w^{(n)}]$ are reserved for self-avoiding walks (self-avoiding walks of length n). Since the position of the starting point of the walk is irrelevant for most problems, *we always assume that the initial point is* $0 = (0, \cdots, 0)$. A walk $w^{(n)}$ of length n is then a sequence of $(n+1)$ points $X_0 = 0, X_1, \cdots, X_n$ such that $|X_i - X_{i+1}| = 1$, $0 \leq i \leq n-1$ ($|Y|$ is the Euclidean distance between Y and 0), and $X_i \neq X_j$ if $i \neq j$. $X_i(w)$ will denote the ith point of the walk w.

An interesting subclass of the self-avoiding walks is the class of self-avoiding polygons. A self-avoiding polygon of length n is a self-avoiding walk w^n with $|X_n(w^n)| = 1$. $g(g^n)$ will denote a generic self-avoiding polygon (of length n). (The name self-avoiding polygon is of course used because the path $X_0, X_1, \cdots, X_n, X_0$ is closed and without double points.)

Our methods make much use of the papers of Hammersley[4,5] and Hammersley and Welsh[6] and concern mostly χ_n, the number of self avoiding walks of length n; and γ_n, the number of self-avoiding polygons of length n. (Clearly $\gamma_n = 0$ if n is even.) It was proved by Hammersley and Welsh[6] that there exist constants β and α_1 (depending on d) such that

$$\beta^n \leq \chi_n \leq e^{\alpha_1 n^{\frac{1}{2}}} \beta^n, \tag{1.1}$$

and for the same β, Hammersley[4] proved

$$\gamma_n \leq 2\, dn\, \beta^n, \quad \text{and} \quad \lim_{n\to\infty} \gamma_{2n-1}^{1/2n} = \beta. \tag{1.2}$$

Numerical evidence[3] suggests that for suitable constants $\alpha_2, \alpha_3, A_1, A_2$,

$$\chi_n \approx A_1 n^{\alpha_2} \beta^n, \quad \text{and} \quad \gamma_{2n-1} \approx A_2 n^{\alpha_3} \beta^{2n},$$

so that (1.1) and (1.2) still seem quite far from the true results. The only result in this direction proved here is

$$e^{-\alpha_4 n^{\frac{1}{2}}} \beta^n \leq \gamma_n, \tag{1.3}$$

for a certain constant α_4.

It has been conjectured that (cf reference 4 for the history of the conjecture)

$$\lim_{n\to\infty} \frac{\chi_{n+1}}{\chi_n} = \beta, \tag{1.4}$$

and

$$\lim_{n\to\infty} \frac{\gamma_{2n+3}}{\gamma_{2n+1}} = \beta^2. \tag{1.5}$$

(1.4) still seems to be open. We do prove that there exist positive constants $A_3 - A_5$ for which

$$|\chi_{n+2}/\chi_n - \beta^2| \leq A_3 n^{-\frac{1}{2}}, \tag{1.6}$$

and

$$-A_4 n^{-\frac{1}{2}} \leq \gamma_{2n+3}/\gamma_{2n+1} - \beta^2 \leq A_5 n^{-\frac{1}{2}} \tag{1.7}$$

[1] F. T. Wall and L. A. Miller, Jr., Ann. Rev. Phys. Chem. **5**, 267 (1954).
[2] S. R. Broadbent and J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 629 (1957).
[3] M. E. Fisher and M. F. Sykes, Phys. Rev. **114**, 45 (1959).
[4] J. M. Hammersley, Proc. Cambridge Phil. Soc. **57**, 516 (1961).
[5] J. M. Hammersley, Quart. J. Math, Oxford, Ser. 2 **12**, 250 (1961).
[6] J. M. Hammersley and D. J. A. Welsh, Quart. J. Math., Oxford, Ser. 2 **13**, 108 (1962).

hold. (1.7) of course implies (1.5), and (1.6) comes close to (1.4). As another result in the direction of (1.4), we prove

$$\lim_{n\to\infty} \frac{\nu_{n+1}}{\nu_n} = \beta, \tag{1.8}$$

where

ν_n = number of self avoiding walks w^n with

$$0 = X_0^{(1)}(w^n) < X_i^{(1)}(w^n) \leq X_n^{(1)}(w^n), \quad 1 \leq i \leq n. \tag{1.9}$$

An immediate consequence of reference 6 is, for a suitable α_5,

$$e^{-\alpha_5 n^{\frac{1}{2}}}\beta^n \leq \nu_n \leq \beta^n. \tag{1.10}$$

In Sec. 4 it is shown that $1/\beta$ is the unique positive solution of

$$\sum_{n=1}^{\infty} \lambda_n x^n = 1,$$

where λ_n is the number of w^n in a certain class, defined more precisely in Sec. 4. This fact can be used to find lower bounds for β.

It is possible by our methods to prove (1.4) for the plane triangular lattice and to show

$$\lim_{n\to\infty} \frac{\chi_{n+2}}{\chi_n} = \beta^2$$

for the tetrahedral lattice, where χ_n is the number of self-avoiding walks of length n on the appropriate lattice.

2. A BASIC INEQUALITY

The main tool for proving (1.6) and (1.7) will be Theorem 2 below. For the proof we need Theorem 1 which is analagous to exponential estimates for the strong law of large numbers of probability theory. From this point of view, Theorem 1 is quite intuitive. Its proof is rather involved, though, and deferred to the last section. Basically, Theorem 1 says that any configuration of k steps which can occur more than once in a self avoiding walk of n steps has to occur at least $a \cdot n$ times with the exception of "very few" w^n.

To state the theorems, we need the following definitions. A cube D is a set of lattice points of of form $D = \{X : c_i \leq X^{(i)} \leq c_i + b\ 1 \leq i \leq d\}$ for some integers $c_1, \cdots, c_d$. The meaning of an edge or vertex of a cube is self evident. If $C = \{X_i(C),\ i = 0, \cdots, k\}$ is a self-avoiding walk with $X_0(C) = 0$ we say that C *occurs at the* r*th step of* w^n *if*

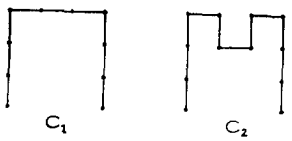

FIG. 1. The paths C_1 and C_2.

$$X_{r+j}(w^n) - X_r(w^n) = X_j(C) \quad j = 0, \cdots, k \tag{2.1}$$

[of course $Y_1 \pm Y_2$ is the point

$$(Y_1^{(1)} \pm Y_2^{(1)}, \cdots, Y_1^{(d)} \pm Y_2^{(d)})$$

if $Y_i = (Y_i^{(1)}, \cdots, Y_i^{(d)})$.] In other words, C occurs at the rth step of w^n if the piece of w^n between its rth and $(r + k)$th step is simply a translation of C. If there exists a cube D which has $0 = X_0(C)$ and $X_k(C)$ as two of its vertices and contains C [i.e., $X_i(C) \in D$, $0 \leq i \leq k$], then we say that (C, D) *occurs at the* r*th step of* w^n if C occurs at the rth step and, if in addition, w^n occupies no points of $X_r(w^n) + (D - C)$. In other words, if $X_{r+j}(w^r) - X_r(w^n) = X_j(C)$, $j = 0, \cdots, k$, and if $X_s(w^n) - X_r(w^n) \in D$ then $s = r + j$ for some $0 \leq j \leq k$ and $X_s(w^n) - X_r(w^n) = X_{s-r}(C)$. We put

$\chi_n(j, C)[\chi_n(j, (C, D))]$ = the number of w^n for which $C[(C, D,)]$ occurs at most at j steps. (2.2)

Theorem 1. Let $C = [X_i(C),\ 0 \leq i \leq k]$ be a self-avoiding path with $X_0(C) = 0$ and D a cube such that D has 0 and $X_k(C)$ as two of its vertices and contains C. Then there exists an $a > 0$ such that

$$\limsup_{n\to\infty} \left[\frac{\chi_n(an, (C, D))}{\chi_n}\right]^{1/n} < 1. \tag{2.3}$$

By (1.1), (2.3) is equivalent to

$$\limsup_{n\to\infty} \chi_n(an, (C, D))^{1/n} < \beta. \tag{2.4}$$

The theorem is proved in Sec. 5 in a slightly more general form.

Even though the choice of C for the next proof is rather arbitrary, we apply it to the paths C_1 and C_2 whose points have $X_i^{(j)}(C_1) = X_i^{(j)}(C_2) = 0$ for $j \geq 3$, and have length 9 and 11, respectively. The first two coordinates for the points of C_1 are in their proper order (0, 0), (0, 1), (0, 2), (0, 3), (1, 3), (2, 3), (3, 3), (3, 2), (3, 1), and (3, 0) (see Fig. 1). For C_2 they are (0, 0), (0,1), (0, 2), (0, 3), (1, 3), (1, 2), (2, 2), (2, 3), (3, 3), (3, 2), (3, 1), and (3, 0) (see Fig. 1). D will be the cube

$$D = [X : 0 \leq X^{(i)} \leq 3,\ i = 1, \cdots, d]$$

for both C_1 and C_2. Moreover, we use the abbreviations

$$\phi_n = \chi_{n+2}/\chi_n, \qquad \psi_n = \gamma_{2n+3}/\gamma_{2n+1}.$$

The basic inequalities for the proof of (1.6) and (1.7) are given in

Theorem 2. There exist constants B_1 and B_2 such that for sufficiently large n

$$\phi_{n+2} \geq \phi_n - B_1/n \tag{2.5}$$

and

$$\psi_{n+1} \geq \psi_n - B_2/n. \tag{2.6}$$

Proof: It will suffice to prove (2.5), the proof of (2.6) being quite similar. For any w, put $m_i(w) =$ number of steps at which (C, D) occurs in $w (i = 1, 2)$. Take an arbitrary w^n, w_0^n, say, with $m_1(w_0^n) > 0$, and choose an r such that (C_1, D) occurs at the rth step of w_0^n. We can then change w_0^n slightly and obtain a w^{n+2} such that (C_2, D) occurs at its rth step. In fact, if the $(r + 4)$th step of w_0 goes from $X_{r+4}(w_0)$ to $X_{r+5}(w_0) = X_{r+4}(w_0) + (1, 0, 0, \cdots, 0)$, we change it so as to go from $X_{r+4}(w_0)$ to $X_{r+5}(w_0)$ in three steps, namely from X_{r+4} to $Y = X_{r+4} + (0, -1, 0, \cdots, 0)$, from Y to $Z = X_{r+4} + (1, -1, 0, \cdots, 0)$, and from Z to X_{r+5}. Otherwise we leave the path unaltered. Clearly the new path has length $(n + 2)$. It is self avoiding because (C_1, D) occurred at the rth step of w_0^n so that w_0^n did not occupy Y or Z. Moreover, for the walk obtained, w_0^{n+2}, say, one has

$$m_1(w_0^{n+2}) = m_1(w_0^n) - 1, \tag{2.7}$$

$$m_2(w_0^{n+2}) = m_2(w_0^n) + 1. \tag{2.8}$$

For any w^n we can make this change on $m_1(w^n)$ different places, and if we consider all possible ways of making this change in all possible w^n, we obtain

$$\sum_{w^n} m_1(w^n) \tag{2.9}$$

walks of length $(n + 2)$. It is clear, however, that these are not all different. To be precise, if (C_2, D) occurs at the rth step of w_1^{n+2}, w_1^{n+2} could have arisen by making the above change in the walk w_1^n which has (C_1, D) at its rth step and is otherwise identical with w_1^{n+2}. If (C_2, D) occurs $m_2(w_1^{n+2})$ times in w_1^{n+2}, then w_1^{n+2} can arise from exactly $m_2(w_1^{n+2})$ walks of length n. We will therefore, count each walk w^{n+2}, obtained by changing w_0^n, once in total, if we give it a weight of

$$1/m_2(w^{n+2}) = 1/m_2(w_0^n) + 1 \quad \text{[cf. (2.8)]}$$

in the sum (2.9); i.e.,

$$\chi'_{n+2} = \sum_{w^n} \frac{m_1(w^n)}{m_2(w^n) + 1}, \tag{2.10}$$

where χ'_{n+2} is the number of w_{n+2} which can be obtained by changing a w^n as described above. The argument above shows that we will obtain any w^{n+2} with $m_2(w^{n+2}) > 0$. Thus,

$$\chi'_{n+2} = \chi_{n+2} - \chi_{n+2}(0, (C_2, D)) = \chi_{n+2}[1 - O(1 - \epsilon_2)^n]$$

for some $\epsilon_2 > 0$, since by Theorem 1,

$$\chi_n[0, (C_i, D)] \leq \chi_n[a_i n, (C_i, D)] \leq (1 - \epsilon_i)^n \chi_n \tag{2.11}$$

for some $a_i > 0$, $\epsilon_i > 0$, $i = 1, 2$, for large n. Consequently,

$$\phi_n = \frac{\chi_{n+2}}{\chi_n} = \frac{1}{\chi_n} \sum_{w^n} \frac{m_1(w^n)}{m_2(w^n) + 1} + O[(1 - \epsilon_2)^n]. \tag{2.12}$$

In essentially the same manner we obtain all but an exponentially small class of w^{n+4} by changing two occurrences of (C_1, D) in a w^n into (C_2, D). Using (2.7) and (2.8) we obtain

$$\phi_n \cdot \phi_{n+2} = \frac{\chi_{n+4}}{\chi_n} = \frac{1}{\chi_n} \sum_{w^n} \frac{m_1(w^n)(m_1(w^n) - 1)}{(m_2(w^n) + 1)(m_2(w^n) + 2)} + O(1 - \epsilon_2)^n. \tag{2.13}$$

By Schwarz's inequality and the obvious inequality $m_i(w^n) \leq n$, we finally obtain from (2.12), (2.13), and (2.11),

$$\begin{aligned}
&\phi_n \phi_{n+2} - \phi_n^2 \\
&= \frac{1}{\chi_n} \sum \left(\frac{m_1}{m_2 + 1}\right)^2 - \left(\frac{1}{\chi_n} \sum \frac{m_1}{m_2 + 1}\right)^2 \\
&\quad - \frac{1}{\chi_n} \sum \frac{m_1(m_1 + m_2 + 1)}{(m_2 + 1)^2(m_2 + 2)} + O(1 - \epsilon_2)^n \\
&\geq -\frac{n(2n + 1)}{(a_2 n)^3} - \frac{n(2n + 1)}{\chi_n} \chi_n(a_2 n, (C_2, D)) \\
&\quad + O(1 - \epsilon_2)^n \geq -\frac{B_3}{n},
\end{aligned} \tag{2.14}$$

for sufficiently large n. Since (2.12) and (2.11) show that

$$\phi_n \geq a_1[1 - (1 - \epsilon_1)^n] + O[(1 - \epsilon_2)^n],$$

the theorem follows from (2.14).

3. PROOF OF EQS. (1.3), (1.6), AND (1.7)

First a few words about (1.10) which is essentially in reference 6. The right-hand inequality appears explicitly in reference 5, formula (14). For the left-hand inequality we combine Eqs. (7) and (9) of reference 6, which, in our notation, state

$$\chi_n \leq \sum_{p=0}^{n} P_D(p)\nu_p P_D(n+p-1)\nu_{n-p+1}, \qquad (3.1)$$

where $P_D(p)$ is the number of partitions of p into unequal parts. Since $\nu_p\nu_{n-p+1} \leq \nu_{n+1}$ [Eq. (12) of reference 5], and $\log P_D(p) \sim \pi(\frac{1}{3}p)^{\frac{1}{2}}(p \to \infty)$,[7] (1.10) follows from (3.1) and the left-hand inequality in (1.1).

The next lemma is basically due to Hammersley.[4] We can order the lattice points lexicographically according to their coordinates. We say then that

$$Y = (Y^{(1)}, \cdots, Y^{(d)}) \prec Z = (Z^{(1)}, \cdots, Z^{(d)}),$$

if for some $j \leq d$, $Y^{(i)} = Z^{(i)}$ $1 \leq i \leq j$ and $Y^{(j)} < Z^{(j)}$. Let η_n = number of g^{n-1} with $X_0(g^{n-1}) = 0 \prec X_i(g^{n-1})$, $1 \leq i \leq n-1$. Our η_n corresponds to $2h_n$ of Hammersley. The difference by a factor 2 is a result of counting $X_0, X_1, \cdots, X_{n-1}$ and $X_0, X_{n-1}, \cdots, X_1$ as different objects (this is not done for h_n in reference 4). It is proved in reference 4, Eq. 11, that

$$\eta_n = \gamma_{n-1}/n. \qquad (3.2)$$

Lemma 1. For positive integers m and n,

$$\gamma_{m+n} \geq (1/2d)\gamma_m \cdot \eta_n, \qquad (3.3)$$

and

$$\chi_{m+n} \geq (1/2d)\chi_m \cdot \eta_n. \qquad (3.4)$$

Proof: Formula (7) of reference 4 states that

$$\eta_{m+n} \geq [1/2(d-1)]\eta_m \cdot \eta_n,$$

which is stronger than (3.3). The proof of (3.4) differs only in a minute detail from Sec. 8 in reference 4, and is given mainly for completeness sake. Let w^m be an arbitrary self-avoiding path of length m and let k_0 be such that $X_i(w^m) \prec X_{k_0}(w^n)$ $0 \leq i \leq n$, $i \neq k_0$. In particular, $X_{k_0+1}(w^m) \prec X_{k_0}(w^m)$. Assume first $k_0 \geq 1$, $X^{(t)}_{k_0-1}(w^m) = X^{(t)}_{k_0}(w^m)$ for $t \neq j_0$, $X^{(j_0)}_{k_0-1}(w^m) = X^{(j_0)}_{k_0}(w^m) - 1$, *and* $j_0 > 1$. Let h^n be a self-avoiding polygon of length $(n-1)$ such that $X_0(h^n) = 0 \prec X_i(h^n)$ $1 \leq i \leq n-1$, and such that $X^{(t)}_{n-1}(h^n) = 0$ for $t \neq j_0 X^{(j_0)}_{n-1}(h^n) = 1$. By symmetry there are at least $\eta_n/2d$ possible choices for h^n.[4] Consider now the path of length $m+n$ whose points in their proper order are: $X_i(w^m)$ $0 \leq i \leq k_0 - 1, X_{k_0-1}(w^m) + (1, 0 \cdots, 0) + X_r(h^n)$, $0 \leq r \leq n-1$, $X_i(w^m)$, $k_0 \leq i \leq m$. This path is self avoiding because the second piece $[X_{k_0-1}(w^m) + (1, 0, \cdots, 0) + X_r(h^n)]$ has only points with first coordinate at least $X^{(1)}_{k_0-1}(w^m) + 1 + X^{(1)}_0(h^n) = X^{(1)}_{k_0}(w^m) + 1$, whereas all the other points have a first coordinate at most $X^{(1)}_{k_0}(w^m)$. If $k_0 < n$ and $X^{(1)}_{k_0+1}(w^m) = X^{(1)}_{k_0}(w^m)$, a similar construction gives a path w^{m+n}. Moreover, if $0 < k_0 < n$, one of these cases must occur since not both X_{k_0-1} and X_{k_0+1} can be equal to $X_{k_0} + (-1, 0 \cdots 0)$. If, however, $k_0 = 0$ and $X_1(w^m) = (-1, 0, \cdots, 0)$, we choose h_n such that $X^{(1)}_{n-1}(h^n) = 0$ [and again $X_0(h^n) = 0 \prec X_i(h^n)$ $1 \leq i \leq n-1$], which can be done in at least $\frac{1}{2}\eta_n$ ways. We now obtain a w^{m+n} by taking the points $X_r(h^n)$, $0 \leq r \leq n-1$, $X_{n-1}(h^n) + (-1, 0 \cdots 0) + X_i(w^m)$ $0 \leq i \leq m$. A similar construction will always work if $k_0 = n$. For each w^m we can choose h^n in at least $\eta_n/2d$ ways, and we therefore obtain at least $\chi_m(\eta_n/2d)$ self-avoiding walks of length $(m+n)$ if we go through the above construction for each w^m. It is not hard to see that all the walks w^{m+n} obtained in this manner are different since there always exists a hyperplane $X^{(1)} = \alpha$ such that w^{m+n} has exactly n points with $X^{(1)}_i(w^{m+n}) > \alpha$. These n points form h^n except for a translation, and the remaining m points form w^m except for a translation. Thus,

$$\chi_{m+n} \geq \chi_m \cdot \eta_n/2d.$$

Theorem 3. There exists a constant $0 < \alpha_4 < \infty$, depending on d, such that for odd n,

$$e^{-\alpha_4 n^{\frac{1}{2}}}\beta^n \leq \gamma_n \leq 2(d-1)(n+1)\beta^n. \qquad (3.5)$$

Proof: The right-hand inequality is actually proved in reference 4 since $\gamma_n = (n+1)\eta_{n+1}$ and (see Eq. 7 in reference 4)

$$\frac{\eta_{2m+2n}}{2(d-1)} \geq \frac{\eta_{2m}}{2(d-1)} \frac{\eta_{2n}}{2(d-1)}. \qquad (3.6)$$

It is well known (see also reference 5) that (3.6) implies

$$\left[\frac{\eta_{2n}}{2(d-1)}\right]^{1/2n} \leq \lim_{m\to\infty}\left[\frac{\eta_{2m}}{2(d-1)}\right]^{1/2m} = \lim_{m\to\infty}\left[\frac{\gamma_{2m-1}}{4m(d-1)}\right]^{1/2m} = \beta.$$

For the left-hand inequality we construct a g^{4n+1} from 4 w^n's. This construction is a slight refinement of reference 4. Let N_n be the class of w^n with $0 = X^{(1)}_0(w^n) < X^{(1)}_i(w^n) \leq X^{(1)}_n(w^n)$ $1 \leq i \leq n$. N_n has ν_n members. Since there are at most $(2n+1)^d$ possible values for $X_n(w^n)$, there exists a $Y = (Y^{(1)}, \cdots, Y^{(d)})$ such that there are at least $(2n+1)^{-d}\nu_n$ members of N_n with $X_n(w^n) = Y$ [and, by symmetry, also with $X_n(w^n) = \tilde{Y} = (Y^{(1)}, -Y^{(2)}, \cdots, -Y^{(d)})$]. If $u, \tilde{u} \in N_n$ and $X_n(u^n) = Y$, $X_n(\tilde{u}^n) = \tilde{Y}$, then the walk w

[7] G. H. Hardy and S. Ramanujan, Proc. London Math. Soc. Ser 2 **16**, 112 (1917).

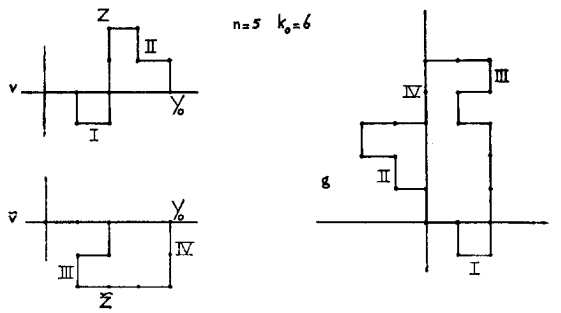

FIG. 2. Construction of the polygon g by regrouping the pieces I–IV of v and $\bar{v}$.

whose points are $X_i(u)$, $0 \leq i \leq n$, $X_n(u) + X_i(\tilde{u})$, $1 \leq i \leq n$, belongs to N_{2n} and $X_{2n}(w) = Y_0 = (2Y^{(1)}, 0, \cdots, 0)$. Thus, N_{2n} has at least $(2n+1)^{-2d}\nu_n^2$ members whose end point is Y_0. Let w be such a path and let k_0 be the smallest integer for which $X^{(2)}(w)$ is maximal, i.e., $X_i^{(2)}(w) \leq X_{k_0}^{(2)}(w)$, $0 \leq i \leq 2n$. There are at most $(2n+1)$ possible values for k_0 and once k_0 is fixed at most $(2n+1)^d$ values for $X_{k_0}(w)$. We can therefore choose k_0 and $Z = (Z^{(1)}, \cdots, Z^{(d)})$ such that V_{2n}, the class of $w \in N_{2n}$ with $X_{2n}(w) = Y_0$, $X_i^{(2)}(w) \leq X_{k_0}^{(2)}(w)$ $0 \leq i \leq n$, and $X_{k_0}(w) = Z$, has at least $(2n+1)^{-3d-1}\nu_n^2$ members. By symmetry, the same holds for the paths for which $X_{k_0}^{(2)}$ is minimal, $X_{k_0} = \tilde{Z} = (Z^{(1)}, -Z^{(2)}, Z^{(3)}, \cdots, Z^{(d)})$. That is, if $\tilde{V}_{2n}$ is the class of $w \in N_{2n}$ with $X_{2n}(w) = Y_0$, $X_i^{(2)}(w) \geq X_{k_0}^{(2)}(w)$, $0 \leq i \leq 2n$, $X_{k_0}(w) = \tilde{Z}$, then $\tilde{V}_{2n}$ also has at least $(2n+1)^{-3d-1}\nu_n^2$ members. Let now v and $\bar{v}$ be arbitrary members of V_{2n} and $\tilde{V}_{2n}$, respectively, and let Z_0 be the point $(0, 2Z^{(2)}, 0, \cdots)$. Consider the path g built out of four pieces and whose points in proper order are:

$$X_i(v) \quad 0 \leq i \leq k_0, \quad Z + (0, 1, 0, \cdots, 0) + X_i(\bar{v}) - X_{k_0}(\bar{v}),$$

$$i = k_0, \quad k_0 - 1, \cdots, 0, \quad Z_0 + (0, 1, 0, \cdots, 0) + X_i(\bar{v}) - X_{2n}(\bar{v});$$

$$i = 2n-1, \quad 2n-2, \cdots, k_0, \quad Z_0 + \tilde{Z} - Y_0 + X_i(v) - X_{k_0}(v) \quad k_0 \leq i \leq 2n-1.$$

Figure 2, demonstrating a typical two-dimensional example, should be of help to the reader. Actually one cuts v at its k_0th point and translates the endpoint of the final part to the origin and performs a similar operation on $\bar{v}$. g is then built by connecting the two paths, so obtained from v and $\bar{v}$. It is easy to see that each step from one point to the next (including from the last point back to 0) has length one if one takes into account such relations as $X_{k_0}(v) = Z, Z + (0, 1, 0, \cdots, 0) + X_0(\bar{v}) - X_{k_0}(\bar{v}) = Z + (0, 1, 0, \cdots, 0) - \tilde{Z} = Z_0 + (0, 1, 0, \cdots, 0)$, etc. To see that g is self avoiding, one only has to look at the first two coordinates of its points. The first piece always has $0 \leq X^{(1)}(g)$ and $X^{(2)}(g) \leq Z^{(2)}$. The second piece has $0 \leq X^{(1)}(g)$ and $Z^{(2)} + 1 \leq X^{(2)}(g)$, the third piece $X^{(1)}(g) < 0$, $Z^{(2)} + 1 \leq X^{(2)}(g)$, and finally, in the last piece, $X^{(1)}(g) < 0$, $X^{(2)}(g) \leq Z^{(2)}$. Thus g is a self-avoiding polygon of length $(4n+1)$. There are at least $(2n+1)^{-6d-2}\nu_n^4$ possible choices for v and $\bar{v}$ and all the polygons g obtained in this manner are different. After all, it is unambiguous which part of g comes from v or $\bar{v}$. Therefore,

$$\gamma_{4n+1} \geq (2n+1)^{-6d-2}\nu_n^4, \tag{3.7}$$

and (3.5) follows from (1.10) if n is of the form $4m+1$. If $4m+1 \leq n < 4(m+1)+1$, we use [cf. (3.3)]

$$\gamma_n \geq \gamma_{4m-3}(\eta_{n-4m+3}/2d).$$

The remaining part of the proof of (1.6) and (1.7) is purely mechanical and quite easy now.

Theorem 4. There exist constants $A_3 - A_5$ such that

$$|\chi_{n+2}/\chi_n - \beta^2| \leq A_3 n^{-\frac{1}{3}}, \tag{3.8}$$

and

$$-A_4 n^{-\frac{1}{3}} \leq \gamma_{2n+3}/\gamma_{2n+1} - \beta^2 \leq A_5 n^{-\frac{1}{3}}. \tag{3.9}$$

Proof: Define ρ_n by

$$\phi_n = \chi_{n+2}/\chi_n = \beta^2 + \rho_n n^{-\frac{1}{3}}.$$

Assume that $\rho_n > 0$. We shall show that it is necessarily bounded above. By (2.5) we have for $k = 0, 1, \cdots, m = [(\rho_n/2B_1)n^{\frac{2}{3}}]$

$$\phi \geq \beta_2 + \rho_n n^{-\frac{1}{3}} - \frac{mB_1}{n} \geq \beta^2 + \tfrac{1}{2}\rho_n n^{-\frac{1}{3}} \geq \beta^2\left(1 + \frac{\rho_n}{2\beta^2} n^{-\frac{1}{3}}\right).$$

Consequently,

$$\prod_{k=0}^{m-1} \phi_{n+2k} = \frac{\chi_{n+2m}}{\chi_n} \geq \beta^{2m}\left(1 + \frac{\rho_n}{2\beta^2} n^{-\frac{1}{3}}\right)^m.$$

On the other hand, it is clear that

$$\chi_{n+2m} \leq \chi_n \cdot \chi_{2m} \tag{3.10}$$

[cf. Eq. (9) in reference 5], and hence, by (1.1),

$$\beta^{2m}(1 + (\rho_n/2\beta^2)n^{-\frac{1}{3}})^m \leq \chi_{2m} \leq \beta^{2m} e^{\alpha_1 (2m)^{\frac{1}{3}}}$$

and, for large n,

$$\rho_n^{\frac{1}{3}} \log (1 + (\rho_n/2\beta^2)n^{-\frac{1}{3}}) \leq 4\alpha_1 B_1^{\frac{1}{3}} n^{-\frac{1}{3}}.$$

Thus, ρ_n is bounded from above. In order to show that ρ_n is bounded from below if $\rho_n < 0$, one uses, for large n,

$$\rho' = \min(|\rho_n|, \log n) \quad \text{and} \quad m' = [(\rho'/4B_1)n^{\frac{2}{3}}]$$

$$\phi_{n-2k} \leq \beta^2 - \rho' n^{-\frac{1}{3}} + 2m'B_1/n \leq \beta^2 - \tfrac{1}{2}\rho' n^{-\frac{1}{3}}, \qquad 0 \leq k \leq m';$$

and [cf. (3.4), (3.2), and (3.5)]

$$\prod_{k=0}^{m'-1} \phi_{n-2k} = \frac{\chi_n}{\chi_{n-2m'}} \geq \frac{\eta_{2m'}}{2d} \geq \frac{1}{4dm'} e^{-\alpha_4(2m'-1)^{\frac{1}{3}}} \beta^{2m'-1}.$$

In the same way one proves (3.9). A small difficulty arises in the proof of the right-hand inequality of (3.9) because we do not have an analogue of (3.10) for γ. Instead, we use the consequence of (3.5)

$$\gamma_{2n+2m+1}/\gamma_{2n+1} \leq \beta^{2m} \cdot 2(d-1) \times (2n + 2m + 1) e^{\alpha_4(2n+1)^{\frac{1}{3}}}.$$

In the exponent we have $(2n + 1)^{\frac{1}{3}}$ rather than $(2m)^{\frac{1}{3}}$, and as a result, we can only prove (3.9) with $A_5 n^{-\frac{1}{3}}$ instead of $A_5 n^{-\frac{1}{3}}$ in the right-hand side.

Remark: The class of w^n with $X_n(w^n) = Y$ for some fixed Y, can be treated in essentially the same way as the polygons. If we denote their number by ${}_n(Y)$, one proves as in reference 4, Sec. 8, or Lemma 1, that

$$\gamma_{n+m}(Y) \geq (1/4d)\gamma_m(Y) \cdot \eta_n,$$

and consequently, for some $0 < A_6(Y), A_7(Y) < \infty$, and large n of the same parity as $\sum_{i=1}^d Y^{(i)}$,

$$(A_6/n)e^{-\alpha_4 n^{\frac{1}{3}}} \beta^n \leq \gamma_n(Y) \leq \chi_n \leq e^{\alpha_1 n^{\frac{1}{3}}} \beta^n.$$

Since the change in the walks used in the proof of Theorem 2 does not change the end point of a walk, it is easy to obtain (2.6) and (3.9) with γ_{2n+1} replaced by ${}_{2n+1}(Y)$ [or $\gamma_{2n}(Y)$, depending on the parity of $\sum_{i=1}^d Y^{(i)}$]. Similarly, taking into account $\nu_n \leq \beta^n$ [formula (14) in reference 5], there is no difficulty in proving

$$\frac{\nu_{n+4}}{\nu_{n+2}} \geq \frac{\nu_{n+2}}{\nu_n} - \frac{B_3}{n}, \tag{3.11}$$

and

$$\frac{\nu_{n+2}}{\nu_n} \leq \beta^2 + A_8 n^{-\frac{1}{3}}, \tag{3.12}$$

for some $0 < B_3, A_8 < \infty$. We shall use (3.12) in the next section.

4. AN EQUATION FOR β

Let L_n be the class of w^n with $0 = X_0^{(1)}(w^n) < X_i^{(1)}(w^n) \leq X_n^{(1)}(w^n)$, $1 \leq i \leq n$, but such that no $k < n$ exists for which $0 < X_i^{(1)}(w^n) \leq X_k^{(1)}(w^n) < X_j^{(1)}(w^n) \leq X_n^{(1)}(w^n)$ whenever $0 < i \leq k < j \leq n$. The number of elements in L_n we shall denote by λ_n. Clearly $L_n \subseteq N_n$ (cf. proof of Theorem 3 for definition) which has ν_n members. We now prove an important relation between λ_n and ν_n.

Lemma 2.

$$\nu_n = \sum_{k=1}^n \lambda_k \nu_{n-k}, \qquad n = 1, 2, \cdots, \tag{4.1}$$

and consequently, for $|x| < \beta^{-1}$,

$$N(x) = \sum_{n=0}^\infty \nu_n x^n = \frac{1}{1 - \Lambda(x)}, \tag{4.2}$$

where $\nu_0 = 1$ and $\Lambda(x) = \sum_{k=1}^\infty \lambda_k x^k$.

Proof: For any $u^n \in N^n$, $0 < X_i^{(1)}(u^n) \leq X_n^{(1)}(u^n)$, and there exists therefore a *smallest* k, $0 < k \leq n$ such that $0 < X_i^{(1)}(u^n) \leq X_k^{(1)}(u^n)$, if $0 < i \leq k$ and $X_k^{(1)}(u^n) < X_j^{(1)}(u^n) \leq X_n^{(1)}(u^n)$ if $k < j \leq n$. It is clear that the walk $[X_i(u^n), 0 \leq i \leq k]$ belongs to L_k while $[X_j(u^n) - X_k(u^n), k < j \leq n]$ belongs to N_{n-k}. Moreover, k can take the values $1, 2, \cdots, n$ and is uniquely determined. Thus,

$$\nu_n \leq \sum_{k=1}^n \lambda_k \nu_{n-k}. \tag{4.3}$$

On the other hand, for any $w^k \in L_k$ and $u^{n-k} \in N_{n-k}$ the path $X_i(w^k)$ $0 \leq i \leq k$, $X_k(w^k) + X_j(u^{n-k})$, $1 \leq j \leq n - k$ belongs to N_n, proving that the inequality sign in (4.3) may be reversed. Thus (4.1) holds and (4.2) follows by taking generating functions, and the fact that $\lambda_n \leq \nu_n \leq \beta^n$ (formula (14) in reference 5).

Theorem 5. β^{-1} is the unique positive root of

$$\Lambda(x) = 1. \tag{4.4}$$

Proof: It is convenient to introduce $\lambda_n^* = \lambda_n \beta^{-n}$ and $\nu_n^* = \nu_n \beta^{-n}$ and their generating functions $\Lambda^*(x) = \sum_{n=1}^\infty \lambda_n^* x^n$, $N^*(x) = \sum_{n=0}^\infty \nu_n^* x^n$. Changing x into $x\beta^{-1}$ in (4.2) gives

$$N^*(x) = 1/[1 - \Lambda^*(x)], \qquad |x| < 1. \tag{4.5}$$

Moreover,

$$\lim_{n\to\infty} \nu_n^{*1/n} = 1. \tag{4.6}$$

It is known,[8] however, that

$$\lim_{n\to\infty} \nu_n^{*1/n} = \text{L.U.B.}\ \{x: x \geq 0, \Lambda^*(x) \leq 1\}. \tag{4.7}$$

[8] N. G. de Bruijn and P. Erdös, Indag. Math. **13**, 374 (1951).

(4.6) and (4.7) together imply

$$\Lambda^*(1) \leq 1. \tag{4.8}$$

To show that equality holds in (4.8) we remark that $\Lambda^*(1) < 1$ would imply

$$N^*(1) = \sum_{n=0}^{\infty} \nu_n \beta^{-n} < \infty, \tag{4.9}$$

which is impossible as we shall prove. Write $N_{n,k}$ for the class of $w^n \in N_n$ with $X_n^{(1)}(w^n) = k$ and denote the number of elements of $N_{n,k}$ by $\nu_{n,k}$. It is shown in reference 6 that each w^n can be built up uniquely from a finite sequence $u_1^{n_1}, v_1^{m_1}, u_1^{n_2}, v_2^{m_2}, \cdots, u_i^{n_i} \in N_{n_i,k_i}, v_i^{m_i} \in N_{m_i,l_i}$ where $k_1 > k_2 > \cdots$ and $l_1 > l_2 > \cdots$. Since for each value of k at most one $u \in \bigcup_n N_{n,k}$ occurs, and for each l at most one $v \in \bigcup_n N_{n,l}$, we find for the generating functions

$$\sum_{n=0}^{\infty} \chi_n x^n \leq \prod_{k=1}^{\infty} \left(1 + \sum_{n=1}^{\infty} \nu_{n,k} x^n\right)^2$$

$$\leq \exp 2\left(\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \nu_{n,k} x^n\right) = \exp 2[N(x) - 1]$$

if $x \geq 0$ (compare this for instance with the determination of the generating function of the number of partitions into different parts.[9]) Since $\chi_n \geq \beta^n$, we have $\sum_{n=0}^{\infty} \chi_n \beta^{-n} = \infty$. One must have equality in (4.8) therefore, which proves the theorem.

Remark: If β_N is the positive solution of

$$\sum_{n=1}^{N} \lambda_n \beta_N^{-n} = 1,$$

then $\beta_N \leq \beta$ and Theorem 4 can therefore be used to find lower bounds for β. These lower bounds are closely related to the lower bounds for β derived in reference 3.

Corollary.

$$\lim_{n\to\infty} \frac{\nu_{n+1}}{\nu_n} = \beta. \tag{4.10}$$

Proof: (4.10) is equivalent to

$$\lim_{n\to\infty} \frac{\nu_{n+1}^*}{\nu_n^*} = 1, \tag{4.11}$$

which follows from theorem 3.1[10] by means of (3.12). In fact, (3.12) in terms of ν^* states

$$\limsup_{n\to\infty} \frac{\nu_{n+2}^*}{\nu_n^*} \leq 1. \tag{4.12}$$

In particular, ν_{n+2}^*/ν_n^* is bounded. Since $\nu_{n+2} \geq \lambda_1 \nu_{n+1} = \nu_{n+1}$, also

$$\limsup_{n\to\infty} \frac{\nu_{n+1}^*}{\nu_n^*} < \infty. \tag{4.13}$$

(4.12) and (4.13) make Theorem 3.1 in reference 10 applicable.

5. PROOF OF THEOREM 1.

The proof of theorem 1 is broken down into several lemmas:

Let D be the cube defined by

$$D = \{X: 0 \leq X^{(i)} \leq b, 1 \leq i \leq d\}, \tag{5.1}$$

and D', the sides of cube enclosing D, i.e.,

$$D' = \{X: -1 \leq X^{(i)} \leq b + 1, \quad 1 \leq i \leq d, X \notin D\}. \tag{5.2}$$

Lemma 3. There exists a $w = w^k$, $k = (b+1)^d - 1$, such that $0 = X_0(w)$ and $X_k(w)$ are vertices of D, and such that w exactly fills D (i.e., $X_i(w) \in D$ $0 \leq i \leq k$ and for each $x \in D$, $X = X_i(w)$ for some $0 \leq i \leq k$). Moreover, for any $Y, Z \in D'$, $Y \neq Z$, there exists a self-avoiding path $w_1 = w_1^{k_1}$ such that $X_0(w_1) = Y$, $X_{k_1}(w_1) = Z$, $-1 \leq X_i^{(j)}(w_1) \leq b + 1$, $0 \leq i \leq k_1$, $1 \leq j \leq d$, and such that w_1 contains w, i.e., $X_{r+i}(w_1) = X_i(w)$ $0 \leq i \leq k$ for some r. One has $k_1 \leq (b+3)^d$. In particular, w_1 fills D.

Proof: The existence of w as required is proved by induction. For $d = 1$, it is trivial to fill up the interval $\{0, 1, \cdots, b\}$ in b steps. If it has been proved for $d - 1$, we construct w by first filling the $(d-1)$-dimensional cube

$$D^{(0)} = \{X^{(1)} = 0, 0 \leq X^{(i)} \leq b, 2 \leq i \leq d\}$$

by means of a self-avoiding path from 0 to the vertex

$$P = (0, P^{(2)}, \cdots, P^{(d)}), \quad P^{(i)} = 0 \text{ or } b, \text{ of } D^{(0)}.$$

Connect P to

$$Q = (1, P^{(2)}, \cdots, P^{(d)}).$$

Q is a vertex of

$$D^{(1)} = \{X^{(1)} = 1, 0 \leq X^{(i)} \leq b, 2 \leq i \leq d\}.$$

Again, by the induction hypothesis, we can fill $D^{(1)}$, ending at a vertex of $D^{(1)}$. Continuing in this manner one obtains a w as required. Since D has $(b+1)^d$ points, w must have length $(b+1)^d - 1$. Once the existence of w is established, it is not hard to show the existence of w_1. Without loss of generality, we assume $Y^{(1)} = -1$, $Y \prec Z$.

[9] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford University Press, London, 1954), 3rd Ed., Secs. 19.3 and 19.4.

[10] A. M. Garsia, S. Orey, and E. Rodemich, Ill. J. Math. 6, 620 (1962).

There exists a self-avoiding walk $v_1 = v_2^{k_2}$ from Y to $R = (-1, 0, \cdots, 0)$, with $X_i(v_1) \in D'$, $X_i^{(1)}(v_1) = -1$ [and $X_0(v_1) = Y$, $X_{k_2}(v_1) = R$]. We take v_1 as the first piece of w_1. After v_1 we continue with w which exactly fills D (this is possible since $|R - 0| = 1$). Finally we choose an $S \in D'$ such that $|S - X_k(w)| = 1$, and take as the last piece of w_1, a self-avoiding path $v_2 = v_2^{k_2}$ from S to Z with $X_i(v_2) \in D'$ [and $X_0(v_2) = S, X_{k_2}(v_2) = Z$]. A little reflection shows that we always can choose v_1, S, and v_2 in such a manner that v_1 and v_2 have no point in common and since $v_1, v_2 \in D'$, $w \in D$, they also have no point in common with w. Thus, the walk w_1 consisting of the pieces v_1, w, v_2, is self-avoiding. Obviously the length of w_1 does not exceed the number of points in D' which is $(b + 3)^d$. This proves the lemma.

Theorem 1 states that certain configurations occur "quite often" except in a small set of walks. We start with showing that "quite often" a self-avoiding walk fills a whole cube. The special feature of a filled cube is that, if the cube D is filled between the rth and sth step of w, then w can never occupy a point of D before its rth step or after its sth step, because w is self-avoiding. This will allow us to make changes in w between its rth and sth step without any interference of the other parts. That is, we shall not create double points by making changes only inside D, between the rth and sth step of w. We now give the details and show what kind of changes to make. First we need a few generalizations of the definitions in Sec. 2. Let b be a fixed positive integer. We say that E occurs at the rth step of w if for some $c = (c_1, c_2, \cdots, c_d)$ with $-b \leq c_i \leq 0$ $i = 1, \cdots, d$, all points of the cube

$$D(r, c) = \{X : c_i \leq X^{(i)} - X_r^{(i)}(w) \leq c_i + b, 1 \leq i \leq d\}$$

are occupied by w [i.e., if $X \in D(r, c)$, then $X = X_s(w)$ for some s]. For any $D(r, c)$, $D''(r, c)$ will be the cube just enclosing D and $D'(r, c)$ will be the sides of D''. Formally,

$$D''(r, c) = \{X : c_i - 1 \leq X^{(i)} - X_r^{(i)}(w) \leq c_i + b + 1, 1 \leq i \leq d\},$$

$$D'(r, c) = D''(r, c) - D(r, c).$$

Notice that always $X_r(w) \in D(r, c)$. The event E_k is said to occur at the rth step of w, if for some c, w occupies at least k points in $D''(r, c)$. If F is an event such as E or E_k, it makes sense to say that F occurs at the rth step of w through its $(r - n_1)$th until $(r + n_1)$th step. By this we mean that F occurs at the n_1th step of the walk $\bar{w}$ which consists of the piece of w between its $(r - n_1)$th and $(r + n_1)$th step (except for a translation). In that case we say that $F(n_1)$ occurs. In other words, $F(n_1)$ occurs at the rth step of w if F occurs at the n_1th step of $\bar{w}$ where $\bar{w} = \{X_i(\bar{w}), 0 \leq i \leq 2n_1\}$ with $X_i(\bar{w}) = X_{r-n_1+i}(w)$ (an obvious truncation is to be used if $r - n_1 < 0$ or $r + n_1 > n$). We shall only use this notation for such events F where occurrence of F in a certain piece of w implies that C also occurs in w itself (e.g., E and E_k are such events). $\chi_n(j, F)$ is defined as in (2.2) as the number of w^n for which F occurs at most at j steps [similarly for $\chi_n(j, F(n_1))$]. We then have

Lemma 4. If

$$\liminf_{n\to\infty} \left[\frac{\chi_n(0, F)}{\chi_n}\right]^{1/n} < 1,$$

then there exist an $a_1 > 0$ and integer n_1 such that

$$\limsup_{n\to\infty} \left[\frac{\chi_n(a_1 n, F(n_1))}{\chi_n}\right]^{1/n} < 1.$$

Proof: Since $\chi_n(0, F(n)) = \chi_n(0, F)$ (no path of length greater than n can occur before the nth step), we can find an $\epsilon > 0$ and an integer n_1 such that

$$\chi_{n_1}(0, F(n_1)) \leq \beta^{n_1}(1 - 2\epsilon)^{n_1},$$

and

$$\chi_{n_1} \leq \beta^{n_1}(1 + \epsilon)^{n_1}$$

[cf. (1.1)]. Any path w^{sn_1} consists of s self-avoiding walks of length n_1 put together in an obvious fashion. For $F(n_1)$ to occur at most $a_2 s$ times in w^{sn_1}, $F(n_1)$ may occur in at most $a_2 s$ of these pieces of length n_1 and not at all in the other pieces. There are $\chi_{n_1}(0, F(n_1))$ possible choices for a w^{n_1} without $F(n_1)$ occuring, and at most χ_{n_1} choices for an arbitrary w^{n_1}. If $F(n_1)$ is to occur in j of the s pieces of length n_1, their places can be chosen in $\binom{s}{j}$ ways. Thus

$$\chi_{sn_1}(a_2 s, F(n_1)) \leq \sum_{j=0}^{a_2 s} \binom{s}{j} \chi_{n_1}^j \chi_{n_1}^{s-j}(0, F(n_1)) \leq \beta^{sn_1} \sum_{j=0}^{a_2 s} \binom{s}{j} (1 + \epsilon)^j (1 - 2\epsilon)^{s-j}. \quad (5.3)$$

It is easy to see from (5.3) that for sufficiently small a_2,

$$\limsup_{s\to\infty} [\chi_{sn_1}(a_2 s, F(n_1))]^{1/sn_1} < \beta = \lim_{s\to\infty} (\chi_{sn_1})^{1/sn_1}. \quad (5.4)$$

For $sn_1 \leq n < (s+1)n_1$, we have

$$\chi_n(a_2 s, F(n_1)) \leq (2d)^{n_1} \chi_{sn_1}(a_2 s, F(n_1)),$$

and the lemma follows if we take $a_1 = a_2/2n_1$.

Lemma 5.

$$\liminf_{n\to\infty} \left[\frac{\chi_n(0, E)}{\chi_n}\right]^{1/n} < 1. \tag{5.5}$$

Proof: We assume that (5.5) is false and derive a contradiction. Assume therefore

$$\lim_{n\to\infty} \left[\frac{\chi_n(0, E)}{\chi_n}\right]^{1/n} = 1. \tag{5.6}$$

Let $S_{n,k}$ be the class of w^n for which neither E nor E_k ever occur and denote the number of its elements by $|S_{n,k}|$. k_0 will be the largest integer k for which

$$\liminf_{n\to\infty} \left(\frac{|S_{n,k}|}{\chi_n}\right)^{1/n} < 1. \tag{5.7}$$

One can always choose c such that at least two of the three points $X_{r-1}(w)$, $X_r(w)$, $X_{r+1}(w)$ belong to $D''(r, c)$. Consequently, E_2 must always occur and $|S_{n,2}| = 0$ so that $k_0 > 2$. On the other hand, if all points of $D''(r, c)$ would be occupied, so would be all points of $D(r, c)$ and thus by (5.6) we do not have (5.7) if k equals the number of points of D''. Therefore $k_0 \leq (b+3)^d - 1$. If we say that $\tilde{E}$ occurs when either E or E_{k_0} occurs, then the definition of k_0 implies

$$\liminf_{n\to\infty} \left[\frac{\chi_n(0, \tilde{E})}{\chi_n}\right]^{1/n} < 1$$

and thus, by lemma 4, there exists an $a_3 > 0$ and an integer n_1 such that

$$\limsup_{n\to\infty} \left[\frac{\chi_n(a_3 n, \tilde{E}(n_1))}{\chi_n}\right]^{1/n} < 1. \tag{5.8}$$

On the other hand, again by the definition of k_0,

$$\lim_{n\to\infty} \left(\frac{|S_{n,k_0+1}|}{\chi_n}\right)^{1/n} = 1. \tag{5.9}$$

If, finally, T_n is the class of those w^n in which neither E, nor E_{k_0+1} occur but in which $E_{k_0}(n_1)$ occurs at least $a_3 n$ times, then $|T_n|$ the number of members of T_n must satisfy

$$|T_n| \geq |S_{n,k_0+1}| - \chi_n(a_3 n, \tilde{E}(n_1)),$$

and thus, by (5.8) and (5.9),

$$\lim_{n\to\infty} \left(\frac{|T_n|}{\chi_n}\right)^{1/n} = 1. \tag{5.10}$$

For each $w^n \in T_n$, there are at least $a_3 n$ values of r such that w^n occupies k_0 points of $D''(r, c)$ for suitable c, between its $(r-n_1)$th and $(r+n_1)$th step, but never occupies a point of $D''(r, c)$ before the $(r-n_1)$th or after the $(r+n_1)$th step. If $D^*(r, c)$ denotes the union of $D''(r, c)$ and the points $X_i(w^n)$ with $r - n_1 \leq i \leq r + n_1$, then we can even find $a_4 n = [a_3/(8b + 8n_1)^d]n$ such values of r and c such that the corresponding $D^*(r, c)$ are pairwise disjoint and do not contain $X_0(w^n)$ or $X_n(w^n)$. In fact, $D^*(r, c)$ is contained in a cube with edgelength $2b + 2n_1 + 2$ with center at $X_r(w^n)$ and this cube intersects a similar cube around $X_s(w^n)$ for at most $(4b + 4n_1 + 5)^d$ values of s. $X_0(w^n)$, or $X_n(w^n)$ belong to $D^*(r, c)$ for at most $2(2b + 2n_1 + 3)^d$ values of r. Thus, for large n, there are at least $[a_3/(4b + 4n_1 + 5)^d]n - 2(2b + 2n_1 + 3)^d \geq a_4 n D^*(r, c)$ as desired. Assume that $a_4 n$ such values have been chosen, that r_0 is such a value, and D_0'' its corresponding $D''(r, c)$. Denote by $X_{r'}(w^n)$ the first point of w^n in D_0'', and by $X_{r''}(w^n)$ the last point in D_0''. $X_{r'}$ and $X_{r''}$ belong to $D'(r, c)$. By Lemma 3, there exists a self-avoiding walk w_1, of length not exceeding $n_2 = (b+3)^d$, going from $X_{r'}(w^n)$ to $X_{r''}(w^n)$, staying in D_0'' and occupying all points of $D(r_0, c)$. If we replace the piece of w^n between its r'th and r''th step by w_1, we obtain a self-avoiding walk of length at most $n + n_2$. If we make such a change for $a_5 n$ values of r, for which the corresponding $D^*(r, c)$ are disjoint, then we obtain a self-avoiding walk of length at most $n + a_5 n_2 n$. Since we can choose $a_5 n$ values of r out of $a_4 n$ possibilities in $\binom{a_4 n}{a_5 n}$ ways, we obtain at least $\binom{a_4 n}{a_5 n}$ $|T_n|$ (not necessarily different) self-avoiding walks by making the above change in each $w^n \in T_n$ at $a_5 n$ values of r with disjoint sets $D^*(r, c)$. To complete the proof we need an upper bound on the number of times a given w can be obtained in this manner. Since for $w_0^n \in T_n$ no cube $D(r, c)$ is completely filled, any $\bar{w}$ obtained from it by the above change has at most $(4b+3)^d a_5 n$ values of s for which $D(s, c)$ is filled. After all, E can occur at the sth step of $\bar{w}$ only if $D(s, c)$ intersects a D_0'' in which w_0^n was changed. w_0^n can be obtained by changing back some of those pieces of $\bar{w}$ where E occurs. If E occurs at the sth step of $\bar{w}$ let $\bar{c}$ be such that $\bar{w}$ fills $\bar{D} = D(s, \bar{c})$ and let $\bar{X}'(\bar{X}'')$ be the first (last) point of $\bar{w}$ in $\bar{D}'' = D''(s, \bar{c})$. Since there are at most χ_{2n_1} self-avoiding paths of length $2n_1$, there are no more than χ_{2n_1} ways of changing $\bar{w}$ back between $\bar{X}'$ and $\bar{X}''$.[11] Taking into account

[11] *Note added in proof:* Here, and in the remainder of this proof, χ_{2n_1} should be replaced by $\sum_{k=0}^{2n_1} \chi_k$, since we only knew $r - n_1 \leq r' \leq r \leq r'' \leq r + n_1$, and hence $r'' - r' \leq 2n_1$ and not $r'' - r' = 2n_1$.

that, given s and $X_s(\bar{w})$ there are at most $(b+1)^d$ $(b+3)^{2d}$ choices for $\bar{c}$, $\bar{X}'$, and $\bar{X}''$, we find that $\bar{w}$ can be changed back in at most

$$[1 + (b+1)^{d-1}(b+3)^{2d}\chi_{2n_1}]^{(4b+3)^d a_5 n} \tag{5.11}$$

ways. Equation (5.11) is an upper bound for the number of times any path can be obtained, and therefore, at least

$$\binom{a_4 n}{a_5 n} |T_n| \tau^{-a_5 n}$$

different paths are obtained, where

$$\tau = [1 + (b+1)(b+3)^{2d}\chi_{2n_1}]^{(4b+3)^d}.$$

Since all these paths have length at most $n + a_5 n_2 n$ one must have

$$\binom{a_4 n}{a_5 n} |T_n| \tau^{-a_5 n} \leq \sum_{i=1}^{n+a_5 n_2 n} \chi_i,$$

or, taking nth roots and letting n tend to ∞ [cf (5.10)],

$$\lim_{n\to\infty} \binom{a_4 n}{a_5 n}^{1/n} \leq \tau^{a_5}\beta^{a_5 n_2}. \tag{5.12}$$

It is easy to see from Stirling's formula that (5.12) is false for sufficiently small $a_5 > 0$. We have therefore derived a contradiction from (5.6) and proved the lemma.

We are now ready to prove slightly generalized version of Theorem 1.

Theorem 1. Let C' be a self-avoiding walk such that there exists a self avoiding walk $C = \{X_i(C), 0 \leq i \leq k\}$ in which C' occurs, and such that there exists a cube D which has $X_0(C) = 0$ and $X_k(C)$ as two of its vertices and contains C. Then [cf. definition (2.2)]

$$\limsup_{n\to\infty} \left[\frac{\chi_n(a_6 n, (C, D))}{\chi_n}\right]^{1/n} < 1 \tag{5.13}$$

for some $a_6 > 0$. *A fortiori*

$$\limsup_{n\to\infty} \left[\frac{\chi_n(a_6 n, C')}{\chi_n}\right]^{1/n} < 1.$$

Proof: Assume that (5.13) is false, i.e., that for each $a_6 > 0$,

$$\limsup_{n\to\infty} \left[\frac{\chi_n(a_6 n, (C, D))}{\chi_n}\right]^{1/n} = 1. \tag{5.14}$$

We may and shall assume that D is the same cube as in (5.1). D'' and D^* will have the same meaning as in Lemma 5. We shall say that E'' occurs at the rth step of w if for some c, w occupies all points of $D''(r, c)$. Let W_n be the class of these w^n in which (C, D) occurs at most $a_6 n$ times but $E''(n_1)$ occurs at least $a_7 n$ times. If $|W_n|$ is the number of elements in W_n, then

$$|W_n| \geq \chi_n(a_6 n, (C, D)) - \chi_n[a_7 n, E''(n_1)]. \tag{5.15}$$

By Lemmas 5 and 4 we can choose $a_7 > 0$ and n_1 such that

$$\limsup_{n\to\infty} \left[\frac{\chi_n(a_7 n, E''(n_1))}{\chi_n}\right]^{1/n} < 1,$$

and thus, by (5.14) and (5.15),

$$\limsup_{n\to\infty} \left(\frac{|W_n|}{\chi_n}\right)^{1/n} = 1. \tag{5.16}$$

We now go through a construction very similar to the one in Lemma 5. For each $w^n \in W_n$, we can find at least $a_8 n = [a_7/(8b + 8n)^d]n$ values of r such that w^n fills a cube $D''(r, c)$ between its $(r - n_1)$th and $(r + n_1)$th step and such that the corresponding $D^*(r, c)$ are disjoint. Let $w_0 \in W_n$, and let r be such a value, and let $X_{r'}(w_0)(X_{r''}(w_0))$ be the first (last) point of w_0 in $D''(r, c)$. Just as w_1 was constructed in Lemma 3, one can construct a self-avoiding walk from $X_{r'}$ to $X_{r''}$, which is entirely contained in $D''(r, c)$ and in which (C, D) occurs. One only has to replace D by $D(r, c)$ and w by C in the second part of the proof Lemma 3. We replace the piece between $X_{r'}$ and $X_{r''}$ by such a path in which (C, D) occurs. We make the same change for other values of r, in total in $a_9 n$ places. Entirely analogous to Lemma 5, we obtain

$$\binom{a_8 n}{a_9 n} |W_n|$$

walks whose length does not exceed $n + a_9 n(b+3)^d$. These walks are self avoiding because the original walks filled the cubes $D''(r, c)$ between $X_{r'}$ and $X_{r''}$ and hence never occupied a point of $D''(r, c)$ before the r'th or after the r''th step. In analogy to the situation of Lemma 5, each new path is obtained at most $\tau^{(a_6+a_9)n}$ times for suitable τ, and thus,

$$\binom{a_8 n}{a_9 n} |W_n| \tau^{-(a_6+a_9)n} \leq \sum_{i=1}^{n+a_9 n_2 n} \chi_i,$$

where $n_2 = (b+3)^d$. Taking nth roots, we obtain for $n \to \infty$, from (5.16),

$$\lim_{n\to\infty} \binom{a_8 n}{a_9 n}^{1/n} \leq \tau^{a_6+a_9}\beta^{a_9 n_2},$$

which again is impossible for sufficiently small a_6 and $a_9 > 0$. This proves that (5.14) cannot be true for all $a_6 > 0$, and hence completes the proof of the theorem.

New Method for Solving Eigenvalue Problems*

TATUYA SASAKAWA†
Department of Physics, University of California, Los Angeles, California
(Received 10 December 1962)

A new method of solving eigenvalue problems is proposed. The secular equation is represented in the form of a dispersion relation. Three methods of solving the dispersion relation are given. Several examples including both nondegenerate cases and degenerate cases are presented. All these examples demonstrate that the present treatment is better than other conventional perturbation treatments. The transformation coefficients between the unperturbed wavefunctions and the perturbed ones are given in terms of the eigenvalues thus obtained. The wavefunctions of the perturbed states form a complete orthonormal set if those of the unperturbed one do.

I. INTRODUCTION

TO solve an eigenvalue problem, that is, to diagonalize the secular equation

$$\det (E - H) = 0, \tag{1.1}$$

in terms of solutions of another eigenvalue problem, has been discussed for a long time. In the usual perturbation formula[1-3] a matrix element may occur more than once in a given perturbation order. As the result, the perturbation formula assumes an infinite series even if the system can be described by a finite number of unperturbed states. To avoid this difficulty, Feenberg[4] and Feshbach[5] have shown that it is possible to obtain a perturbation formula in which such repetition of *matrix elements* does not occur. Therefore, their formula ends with finite number of terms. However, because their formula is expressed in a form of a repeated fraction, it is difficult to apply when the system under consideration contains infinitely many states.

Now let us consider the problem from a slightly different point of view. It is well known that in the usual perturbation formula the *energy denominator* may occur more than once in a given perturbation order. However as we shall point out, it is also possible to obtain a perturbation formula in which repetition of energy denominator does not occur in a given perturbation order.

It is clear that the secular equation (1.1) can be written in the form

$$\prod_a (E - \omega_a) + \sum_a k_a \prod_{b\neq a}{}' (E - \omega_b) + \sum_a \sum_{b>a} k_{ab} \prod_{c\neq ab}{}'' (E - \omega_c) + \cdots = 0. \tag{1.2}$$

Here ω_a stands for an unperturbed energy ϵ_a plus the diagonal matrix element V_{aa} with respect to the unperturbed state $|a)$;

$$\omega_a = \epsilon_a + V_{aa}; \tag{1.3}$$

k_a, k_{ab} etc. are constants.

In (1.2), no repetition of a given factor $(E - \omega_a)$ occurs in a given term if there is no degeneracy with respect to ω_a. As the result, after dividing each term of Eq. (1.2) by $\Pi_a(E - \omega_a)$ and applying the partial fraction method, one can transform this equation into the dispersion formula

$$1 = \sum_i \frac{C_i}{E - \omega_i}. \tag{1.4}$$

Here C_i's consist of finite number of terms if number of unperturbed state is finite. This form is much simpler and handier than any of the other perturbation formulas.[6]

Now, however, in practice, if we follow the above procedure in obtaining the dispersion relation, we are forced to make a very tedious calculation from which it is almost impossible for us to reduce the general formula of C_i, especially when the number of states is very large.[7] Therefore, in the present

* This work was supported by the Army Research Office, and was done while on leave of absence from Kyoto University as a Fulbright Visiting Research Scholar.
† Now at Department of Physics, Kyoto University, Kyoto, Japan.
[1] E. Schrödinger, Ann. Physik 80, 437 (1926).
[2] L. Brillouin, J. Phys. Radium 3, 373 (1932).
[3] E. P. Wigner, Anz. Ungr. Akad. Wiss. Math. Naturw. 53, 475 (1935).
[4] E. Feenberg, Phys. Rev. 74, 206 (1948). Professor Feenberg called the author's attention to the improved Brillouin–Wigner-type formalism: P. Goldhammer and E. Feenberg, Phys. Rev. 101, 1233 (1956), and B. A. Lippman, Phys. Rev. 103, 1149 (1956). Also he noticed that the algebraic identities of P. I. Richard, Phys. Rev. 74, 835 (1948), can be used to transform the Feenberg–Feshbach method to the Brillouin–Wigner method. The author would express his thanks to Professor Feenberg for his communication and for sending his lecture notes. Also Dr. C. Bloch called the author's attention to his perturbation method which appeared in Nucl. Phys. 6, 329 (1958).
[5] H. Feshbach, Phys. Rev. 74, 1548 (1948).
[6] Along the same idea, M. Baker [Ann. Phys. (N. Y.) 4, 271 (1958)] treated the energy shift and the phase shift. The present work is more ready for use. The author thanks Dr. K. T. Mahanthappa for this comment.
[7] However, see note added, page 991.

paper, a procedure, which is rather similar to other perturbation theories in so far as the starting point is concerned, will be applied to derive C_i.

In Sec. II, we shall give the explicit form of C_i together with its derivation. There, also the transformation coefficients between unperturbed wavefunctions and the perturbed ones will be given. By using these transformation coefficients, we can get the complete orthonormal set of the wavefunction of the system in terms of those of the unperturbed system, and vice versa.

We can solve the dispersion relation (1.4) in many ways. Also, we may consider a special method suited to a given problem. In Sec. III, we shall present three methods which may be successfully applied to many problems. The first method is suited for the weak coupling case, the second one for the strong coupling case, and the last one is called the "nearest-neighbor-states" approximation.

In Sec. IV, to demonstrate the use of these methods, we shall give some examples together with the comparison with other perturbation methods, which are reproduced in Appendix for reader's convenience. The examples are (1) the case of two states, (2) the one-dimensional harmonic oscillator, (3) the Mathieu equation and (4) the energy levels in a deformed harmonic-oscillator potential with the spin-orbit and the l^2 forces. In all of these examples, the states are nondegenerate with respect to ω_a. The first two examples are especially interesting, because we already know the exact solution in an analytical form. In the first and the third example, our weak coupling formula will be compared with other perturbation formulas. In the second example, the nearest-neighbor-states method will be applied, while in the last example, the strong-coupling method will be used. All these examples demonstrate that the method is better than other conventional perturbation treatments.

When the system is degenerate with respect to some of ω_i's, the dispersion relation takes, in general, the form

$$\sum_i \sum_{m_i \geq 1} \frac{C_{m_i}}{(E-\omega_i)^{m_i}} = 1. \tag{1.5}$$

No general method for solving this equation is known. However, the method for the nondegenerate case is also applicable, if we can remove the degeneracy "tentatively". In Sec. V, we shall suggest a method of removing degeneracy for simple cases. Examples will be given for (1) the pairing correlation but without diagonal matrix elements, (2) the one-dimensional coupled oscillation, and (3) the level splitting of the spherical harmonic oscillator due to the spin-orbit force.

The present treatment is, of course, directly applicable to many-body problems. Also, we can apply this method to the eigenvalue problems of classical mechanics, e.g. to the problems of the disordered lattice. However, in the present paper, we consider only bound states of a quantum-mechanical system.

II. COEFFICIENTS OF THE DISPERSION RELATION AND TRANSFORMATION COEFFICIENTS

A. General Discussion

The Green's function $(E-H)^{-1}$ is expanded either by the complete orthonormal set $|\lambda)$ of the system or by another complete orthonormal set $|a)$ as

$$\frac{1}{E-H} = \sum_\lambda \frac{|\lambda)(\lambda|}{E-E_\lambda} = \sum_{ab} |a)(a| \frac{1}{E-H} |b)(b|.$$

Hence we have the relations

$$\left.\begin{aligned} \sum_\lambda \frac{(a|\lambda)(\lambda|a)}{E-E_\lambda} &= (a| \frac{1}{E-H} |a) \\ \sum_\lambda \frac{(b|\lambda)(\lambda|a)}{E-E_\lambda} &= (b| \frac{1}{E-H} |a) \end{aligned}\right\}. \tag{2.1}$$

Accordingly, the eigenvalues E_λ are given as the poles of the function $(a|\ (E-H)^{-1}\ |a)$ or $(b|\ (E-H)^{-1}\ |a)$. Here we may choose *any* state $|a)$ in deriving the equation for giving the position of the poles. The simple perturbation iteration procedure in the way of calculating the function $(a|\ (E-H)^{-1}\ |a)$ leads to the Brillouin–Wigner perturbation formula. The elimination of repetition of the matrix elements in a given order leads to the Feenberg–Feshbach perturbation theory, while elimination of the repetition of the energy denominator in a given order leads to the dispersion relation as will be displayed in this section.

In Sec. IIB. it will be shown that the function $(a|\ (E-H)^{-1}\ |a)$ takes the form

$$\left(1 - \sum_i \frac{C_i}{E-\omega_i}\right)(a| \frac{1}{E-H} |a) = \frac{1}{E-\omega_a} \mathfrak{U}_a(E), \tag{2.2}$$

where Σ_i denotes a summation over all the unperturbed states including the chosen state $|a)$. The explicit form of C_i and $\mathfrak{U}_a(E)$ will be given in Sec. IIB3. [C_i and $\mathfrak{U}_a(E)$ will be given by Eqs. (2.34) and (2.35), respectively.] The poles of the right-hand side are at ω_a's. Therefore the poles of

the function $(a|\ (E-H)^{-1}\ |a)$ must be given by the zeros of the first factor in the left-hand side of Eq. (2.2), namely by the dispersion relation (1.4).

The first equation of (2.1) shows that the square of the transformation coefficient $(a|\lambda)$ is the residue at $E = E_\lambda$ of the function $(a|\ (E-H)^{-1}\ |a)$. Since the first factor of Eq. (2.2) can be expressed in the form

$$1 - \sum_i \frac{C_i}{E-\omega_i} = \frac{\prod_\lambda (E-E_\lambda)}{\prod_i (E-\omega_i)}, \tag{2.3}$$

we get the expression for the square of the transformation coefficient as

$$|(a|\lambda)|^2 = \mathfrak{U}_a(E_\lambda) \frac{\prod_{a' \neq a} (E_\lambda - \omega_{a'})}{\prod_{\lambda' \neq \lambda} (E_\lambda - E_{\lambda'})}. \tag{2.4}$$

In Sec. IIC, it will be shown that the function $(b|\ (E-H)^{-1}\ |a)$ takes the form

$$\mathfrak{U}_a(E)(b|\frac{1}{E-H}|a) = \mathfrak{V}_{ba}(E)(a|\frac{1}{E-H}|a) \tag{2.5}$$

[$\mathfrak{V}_{ba}(E)$ will be given in Eq. (2.44)]. The second equation of (2.1) shows that the quantity $(b|\lambda)(\lambda|a)$ is the residue at $E = E_\lambda$ of the function $(b|\ (E-H)^{-1}\ |a)$. Therefore, the phase relation between $(a|\lambda)$ and $(b|\lambda)$ is given by the relation

$$\frac{(a|\lambda)}{(b|\lambda)} = \frac{\text{the residue of } (a|\ (E-H)^{-1}\ |a) \text{ at } E = E_\lambda}{\text{the residue of } (b|\ (E-H)^{-1}\ |a) \text{ at } E = E_\lambda}, \tag{2.6}$$

that is,

$$(b|\lambda) = [\mathfrak{V}_{ba}(E_\lambda)/\mathfrak{U}_a(E_\lambda)](a|\lambda). \tag{2.7}$$

Thus what we have to do is to calculate $(a|\ (E-H)^{-1}\ |a)$ and $(b|\ (E-H)^{-1}\ |a)$. The calculations up to the third order will be performed in Secs. IIB and IIC, respectively. Essential features of manipulation are already revealed in the way of deriving the second- and third-order expansions.

For simplicity, throughout these sections we shall use the notations

$$A \equiv E - H, \qquad A_a \equiv E - \omega_a, \qquad \text{and} \quad A_{aa'} \equiv -V_{aa'}. \tag{2.8}$$

The fundamental relation which will be used is the orthogonality relation

$$(a'|a) = \delta_{aa'}, \tag{2.9}$$

or

$$(a|\frac{1}{A}|a) = \frac{1}{A_a} - \sum_{a' \neq a} \frac{A_{aa'}}{A_a} (a'|\frac{1}{A}|a) \tag{2.10}$$

and

$$(a'|\frac{1}{A}|a) = -\sum_{a'' \neq a'} \frac{A_{a'a''}}{A_{a'}} (a''|\frac{1}{A}|a), \qquad (a' \neq a). \tag{2.11}$$

Later, we shall use the notations

$$\begin{aligned} S_{ik}f_{ik} &= f_{ik} + f_{ki}, \\ S_{ikl}f_{ikl} &= f_{ikl} + f_{kli} + f_{lik}, \\ S_{iklm}f_{iklm} &= f_{iklm} + f_{klmi} + f_{lmik} + f_{mikl}, \text{ etc.} \end{aligned} \tag{2.12}$$

Here f_{ik}, f_{ikl}, etc. are any functions, which have jk, jkl, etc. as suffixes respectively.

B. Calculation of $(a\ |\ (E\text{-}H)^{-1}\ |\ a)$

The function $(a|\ (E-H)^{-1}\ |a)$ will be calculated up to the second and the third order in Secs. IIB1 and IIB2, respectively. More general cases will be summarized in Sec. IIB3.

1. Second-Order Expansion

The iteration of the relation (2.11) into (2.10) leads to

$$\begin{aligned} (a|\frac{1}{A}|a) &= \frac{1}{A_a} - \sum_{a' \neq a} \frac{A_{aa'}}{A_a} (a'|\frac{1}{A}|a) \\ &= \frac{1}{A_a} + \sum_{a' \neq a} \sum_{a'' \neq a'} \frac{A_{aa'}A_{a'a''}}{A_aA_{a'}} (a''|\frac{1}{A}|a) \\ &= \frac{1}{A_a} + \sum_{a' \neq a} \frac{|A_{aa'}|^2}{A_aA_{a'}} (a|\frac{1}{A}|a) \\ &\quad + \sum_{a' \neq a} \Big(\sum_{\substack{a'' > a' \\ \neq a}} + \sum_{\substack{a' > a'' \\ \neq a}} \Big) \frac{A_{aa'}A_{a'a''}}{A_aA_{a'}} (a''|\frac{1}{A}|a) \\ &= \frac{1}{A_a} + \sum_{a' \neq a} \frac{|A_{aa'}|^2}{A_aA_{a'}} (a|\frac{1}{A}|a) \\ &\quad + \sum_{a' \neq a} \sum_{\substack{a'' > a' \\ \neq a}} S_{a'a''} \Big\{ \frac{A_{aa'}A_{a'a''}}{A_aA_{a'}} (a''|\frac{1}{A}|a) \Big\}. \end{aligned} \tag{2.13}$$

Here $S_{a'a''}$ was defined by (2.12). After using Eq. (2.11), the last term of the above equation is transformed as

$$\begin{aligned} &= -\sum_{a' \neq a} \sum_{\substack{a'' > a' \\ \neq a}} S_{a'a''} \\ &\quad \times \Big[\frac{A_{aa'}A_{a'a''}}{A_aA_{a'}} \Big\{ \sum_{a''' \neq a'a''} \frac{A_{a''a'''}}{A_{a''}} \\ &\quad \times (a'''|\frac{1}{A}|a) + \frac{A_{a''a'}}{A_{a''}} (a'|\frac{1}{A}|a) \Big\} \Big]. \end{aligned} \tag{2.14}$$

The last term of Eq. (2.14) is further transformed, using Eq. (2.10), as

$$= -\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}} \frac{|A_{a'a''}|^2}{A_a A_{a'} A_{a''}} \Big\{ 1 - A_a (a| \frac{1}{A} |a) - \sum_{a'''\neq aa'a''} A_{aa'''} (a'''| \frac{1}{A} |a) \Big\}. \quad (2.15)$$

Since we can easily show that

$$\sum_{a'\neq a} \frac{|A_{aa'}|^2}{A_a A_{a'}} + \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}} \frac{|A_{a'a''}|^2}{A_{a'} A_{a''}} = \sum_i \sum_{j>i} \frac{|A_{ij}|^2}{A_i A_j}, \quad (2.16)$$

we get the second-order expansion as

$$\Big(1 - \sum_i \sum_{j>i} \frac{|A_{ij}|^2}{A_i A_j}\Big)(a| \frac{1}{A} |a) = \frac{1}{A_a}\Big(1 - \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}} \frac{|A_{a'a''}|^2}{A_{a'} A_{a''}}\Big) - \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}} \frac{i}{A_a A_{a'} A_{a''}} \times \Big\{ \sum_{a'''\neq a'a''} S_{a'a''}(A_{aa'} A_{a'a''} A_{a''a'''}) - \sum_{a'''\neq aa'a''} |A_{a'a''}|^2 A_{aa'''} \Big\} (a'''| \frac{1}{A} |a), \quad (2.17)$$

after collecting Eqs. (2.13), (2.14), and (2.15). Therefore when the third- and higher-order interactions can be neglected, the zeros of the function $(a|\ (E - H)^{-1}\ |a)$ are determined by the equation

$$1 = \sum_i \sum_{j>i} \frac{|A_{ij}|^2}{A_i A_j}. \quad (2.18)$$

This equation is transformed into the dispersion relation (1.4) with

$$C_i = \sum_{j\neq i} \frac{|V_{ij}|^2}{\omega_i - \omega_j}. \quad (2.19)$$

Equation (1.4) with (2.19) shows that the perturbation causes an effect of repulsion to the eigenvalues, because the set of C_i once changes its sign in proceeding from low energy to high energy, the C_i corresponding to the lowest (highest) state being negative (positive).

When the third- and higher-order terms are neglected in Eq. (2.17), it takes the form (2.2) with

$$\mathfrak{U}_a(E) = 1 - \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}} \frac{|V_{a'a''}|^2}{(E - \omega_{a'})(E - \omega_{a''})}. \quad (2.20)$$

2. Third-Order Expansion

In proceeding to the third-order expansion of $(a|\ (E - H)^{-1}\ |a)$, we have to manipulate the last term of Eq. (2.17) as follows:

$$\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}} \frac{S_{a'a''}(V_{aa''} V_{a''a'} V_{a'a})}{A_a A_{a'} A_{a''}} (a| \frac{1}{A} |a) - \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{a'''\neq aa'a''} \{S_{a'a''}(A_{aa'} A_{a'a''} A_{a''a'''}) - |A_{a'a''}|^2 A_{aa'''}\} \frac{1}{A_a A_{a'} A_{a''}} (a'''| \frac{1}{A} |a). \quad (2.21)$$

If we define $\theta_{a'a''a'''}$ by

$$\theta_{a'a''a'''} = \{S_{a'a''}(A_{aa'} A_{a'a''} A_{a''a'''}) - |A_{a'a''}|^2 A_{aa'''}\} \frac{1}{A_a A_{a'} A_{a''}} (a'''| \frac{1}{A} |a), \quad (2.22)$$

the last term of Eq. (2.21) takes the form

$$\begin{aligned}
&\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{a'''\neq aa'a''} \theta_{a'a''a'''} \\
&= \Big(\sum_{a'''\neq a}\sum_{\substack{a'>a'''\\ \neq a}}\sum_{\substack{a''>a'\\ \neq a}} + \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a'\\ \neq a}} + \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}} \Big) \theta_{a'a''a'''} \\
&= \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a'\\ \neq a}} (\theta_{a'a''a'''} + \theta_{a''a'''a'} + \theta_{a'a'''a''}) \\
&= \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}} S_{a'a''a'''} \theta_{a''a'''a'}, \quad (2.23)
\end{aligned}$$

because

$$\theta_{a'a'''a''} = \theta_{a'''a'a''}$$

as can be shown from (2.22). Using Eq. (2.11), the last term of Eq. (2.21) is further transformed as

$$\begin{aligned}
&= \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}} \frac{1}{A_a A_{a'} A_{a''} A_{a'''}} S_{a'a''a'''} \\
&\times \Big[\sum_{a^{IV}\neq a'} \{S_{a''a'''}(A_{aa'''} A_{a'''a''} A_{a''a'}) - |A_{a''a'''}|^2 A_{aa'}\} A_{a'a^{IV}} \Big](a^{IV}| \frac{1}{A} |a) \\
&= \sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}}\sum_{a^{IV}\neq a'a''a'''} \frac{1}{A_a A_{a'} A_{a''} A_{a'''}} \\
&\times S_{a'a''a'''} [\{S_{a''a'''}(A_{aa'''} A_{a'''a''} A_{a''a'}) - |A_{a''a'''}|^2 A_{aa'}\} A_{a'a^{IV}}](a^{IV}| \frac{1}{A} |a)
\end{aligned}$$

$$+\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}}\frac{1}{A_aA_{a'}A_{a''}A_{a'''}}S_{a'a''a'''}$$
$$\times\Big[\{S_{a''a'''}(A_{aa'''}A_{a'''a''}A_{a''a'})$$
$$-|A_{a''a'''}|^2A_{aa'}\}\Big\{A_{a'a''}(a''|\frac{1}{A}|a)$$
$$+A_{a'a'''}(a'''|\frac{1}{A}|a)\Big\}\Big]. \qquad (2.24)$$

Now putting

$$\varphi_{a'a''a'''}=S_{a''a'''}(A_{aa'''}A_{a'''a''}A_{a''a'})-|A_{a''a'''}|^2A_{aa'}$$

we can calculate the last term as

$$S_{a'a''a'''}\Big\{\varphi_{a'a''a'''}A_{a'a''}(a''|\frac{1}{A}|a)+\varphi_{a'a''a'''}A_{a'a'''}(a'''|\frac{1}{A}|a)\Big\}$$
$$=S_{a'a''a'''}\Big\{(\varphi_{a'''a'a''}A_{a''a'}+\varphi_{a''a'''a'}A_{a'''a'})(a'|\frac{1}{A}|a)\Big\}$$
$$=S_{a'a''a'''}\Big\{S_{a''a'''}(A_{a'a'''}\times A_{a'''a''}A_{a''a'})A_{aa'}(a'|\frac{1}{A}|a)\Big\}$$
$$=S_{a''a'''}(A_{a'a'''}A_{a'''a''}A_{a''a'})\times\Big\{A_{aa'}(a'|\frac{1}{A}|a)+A_{aa''}(a''|\frac{1}{A}|a)+A_{aa'''}(a'''|\frac{1}{A}|a)\Big\},$$

since

$$S_{a''a'''}(A_{a'a'''}A_{a'''a''}A_{a''a'})=S_{a''a'}(A_{a''a}A_{aa'''}A_{a'''a''})=S_{a'a''}(A_{a'''a''}A_{a''a'}A_{a'a'''}).$$

Therefore, the last term of Eq. (2.24) is

$$\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}}\frac{1}{A_aA_{a'}A_{a''}A_{a'''}}S_{a''a'''}\times(A_{a'a'''}A_{a'''a''}A_{a''a'})\times\Big(1-A_a(a|\frac{1}{A}|a)-\sum_{a^{IV}\neq aa'a''a'''}A_{aa^{IV}}(a^{IV}|\frac{1}{A}|a)\Big). \qquad (2.25)$$

Now collecting Eqs. (2.17), (2.21), (2.24), and (2.25), we get the third-order equation

$$\Big(1-\sum_i\sum_{j>i}\frac{|V_{ij}|^2}{A_iA_j}-\sum_i\sum_{j>i}\sum_{k>j}\frac{S_{ik}(V_{ik}V_{kj}V_{ji})}{A_iA_jA_k}\Big)\times(a|\frac{1}{A}|a)=\frac{1}{A_a}\Big(1-\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\frac{|V_{a'a''}|^2}{A_{a'}A_{a''}}-\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}}\frac{S_{a''a'''}(V_{a'a'''}V_{a'''a''}V_{a''a'})}{A_{a'}A_{a''}A_{a'''}}\Big)$$
$$+\sum_{a'\neq a}\sum_{\substack{a''>a'\\ \neq a}}\sum_{\substack{a'''>a''\\ \neq a}}\frac{1}{A_aA_{a'}A_{a''}A_{a'''}}\times[\sum_{a^{IV}\neq a'a''a'''}S_{a'a''a'''}[\{S_{a''a'''}(A_{aa'''}A_{a'''a''}\times A_{a''a'})-|A_{a''a'''}|^2A_{aa'}\}A_{a'a^{IV}}]$$
$$-\sum_{a^{IV}\neq aa''a'''}S_{a''a'''}(A_{a'a'''}\times A_{a'''a''}A_{a''a'})A_{aa^{IV}}](a^{IV}|\frac{1}{A}|a). \qquad (2.26)$$

Therefore, when the fourth- and higher-order terms can be neglected, the zeros of the function $(a|(E-H)^{-1}|a)$ are determined by the equation

$$1=\sum_i\sum_{j>i}\frac{|V_{ij}|^2}{A_iA_j}+\sum_i\sum_{j>i}\sum_{k>j}\frac{S_{ik}(V_{ik}V_{kj}V_{ji})}{A_iA_jA_k}, \qquad (2.27)$$

or, if there is no degeneracy with respect to ω_i's, they are given by the dispersion relation (1.4) with

$$C_i=\sum_{j\neq i}\frac{|V_{ij}|^2}{\omega_i-\omega_j}+\sum_{j\neq i}\sum_{\substack{k>j\\ \neq i}}\frac{S_{jk}(V_{ik}V_{kj}V_{ji})}{(\omega_i-\omega_j)(\omega_i-\omega_k)}. \qquad (2.28)$$

Also when the fourth- and higher-order terms are neglected in Eq. (2.26), we get Eq. (2.2) with

$$\mathfrak{U}_a(E)=1-\sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}|V_{ij}|^2/(E-\omega_i)(E-\omega_j)-\sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\sum_{\substack{k>j\\ \neq a}}S_{jk}(V_{ik}V_{kj}V_{ji})/(E-\omega_i)\times(E-\omega_j)(E-\omega_k). \qquad (2.29)$$

3. *General Formulas*

A more general form of Eq. (2.18) or Eq. (2.27) is given by

$$1=\sum_i\sum_{j>i}\frac{g_{ij}}{A_iA_j}+\sum_i\sum_{j>i}\sum_{k>j}\frac{g_{ijk}}{A_iA_jA_k}+\sum_i\sum_{j>i}\sum_{k>j}\sum_{l>k}\frac{g_{ijkl}}{A_iA_jA_kA_l}+\cdots, \qquad (2.30)$$

where g_{ij}, g_{ijk}, etc. are given by

$$
\begin{aligned}
g_{ij} &= |V_{ij}|^2, \\
g_{ijk} &= S_{jk}(V_{ik}V_{kj}V_{ji}), \\
g_{ijkl} &= S_{jkl}\{S_{kl}(V_{il}V_{lk}V_{ki})V_{ji} - |V_{kl}|^2\,|V_{ij}|^2\}, \\
g_{ijklm} &= S_{jklm}\{S_{klm}(S_{lm}(V_{im}V_{ml}V_{lk})V_{kj}V_{ji} \\
&\quad - |V_{ml}|^2\,V_{ik}V_{kj}V_{ji}) - S_{lm}(V_{km}V_{ml}V_{lk})\,|V_{ij}|^2\}, \\
g_{ijklmn} &= S_{jklmn}[S_{klmn}\{S_{lmn} \\
&\quad \times (S_{mn}(V_{in}V_{nm}V_{ml})V_{lk}V_{kj}V_{ji} \\
&\quad\quad - |V_{nm}|^2\,V_{il}V_{lk}V_{kj}V_{ji}) \\
&\quad\quad - S_{mn}(V_{ln}V_{nm}V_{ml})V_{ik}V_{kj}V_{ji}\} \\
&\quad\quad - S_{lmn}(S_{mn}(V_{kn}V_{nm}V_{ml})V_{lk}\,|V_{ji}|^2 \\
&\quad\quad - |V_{mn}|^2\,|V_{kl}|^2\,|V_{ij}|^2)].
\end{aligned}
\tag{2.31}
$$

etc. S_{jk}, S_{jkl}, etc. were defined by (2.12).

g_{ij}, g_{ijk}, etc. take on simple forms; each of them consists of irreducible clusters, in which no repetition of a given element V_{ij} occurs. g_{ij}, g_{ijk}, etc. assume the form

$$
\begin{aligned}
g_{ij} &= (2), \qquad g_{ijk} = (3), \\
g_{ijkl} &= (4) - (2) \times (2), \\
g_{ijklm} &= (5) - \{(2) \times (3) + (3) \times (2)\}, \\
g_{ijklmn} &= (6) - \{(2) \times (4) + (3) \times (3) \\
&\quad + (4) \times (2)\} + (2) \times (2) \times (2), \quad \text{etc.}
\end{aligned}
\tag{2.32}
$$

where (n) means the irreducible linked cluster of order n.

g_{ij}, g_{ijk}, etc. are all real. Let us explain this statement, taking g_{ijklm}, as an example. Because of the properties of the S symbol defined by (2.12), the first term which takes the form

$$S_{jklm}S_{klm}(S_{lm}f_{ijklm})$$

represents the sum of the cyclic permutation with respect to $(ijklm)$, the total number being $(5-1)\,! = 24$. Thus one can always find the term f_{iabcd} which is the complex conjugate of f_{idcba}. The second term takes the form $S_{jklm}(S_{klm}f_{ijk}f'_{lm})$. Expansion of this term leads to

$$
\begin{aligned}
S_{jklm}(S_{klm}f_{ijk}f'_{lm}) &= (f_{ijk} + f_{ikj})f'_{lm} + (f_{ijl} + f_{ilj})f'_{km} \\
&\quad + (f_{ijm} + f_{imj})f'_{kl} + (f_{ikl} + f_{ilk})f'_{jm} \\
&\quad + (f_{ikm} + f_{imk})f'_{jl} + (f_{ilm} + f_{iml})f'_{jk}.
\end{aligned}
\tag{2.33}
$$

Thus $S_{jklm}(S_{klm}f_{ijk}f'_{lm})$ is real. The total number of terms is $4 \times 3 = 12$, as the definition of the S symbol shows. The right-hand side of Eq. (2.33) shows that $S_{jklm}(S_{klm}f_{ijk}f'_{lm})$ is the sum over terms which is obtained by the procedure that after dividing four numbers $(jklm)$ into two different groups of the form f_{iab} and f'_{cd}, we practice a permutation of a and b in the first group. Thus the total number of terms is ${}_4C_2 \times 2 \times 2\,! = 12$. Finally, the last term, $S_{jlkm}[S_{klm}\{S_{lm}(V_{km}V_{ml}V_{lk})\}|V_{ij}|^2]$, consists of the terms which we obtain after the permutation of four numbers $(jklm)$, so one can always find out the term $V_{ab}V_{bc}V_{ca}\,|V_{id}|^2$, which is the complex conjugate of $V_{ac}V_{cb}V_{ba}\,|V_{id}|^2$.

When there is no degeneracy with respect to ω_i's, Eq. (2.30) is transformed into the dispersion relation (1.4). Here, as in the generalization of Eq. (2.19) and Eq. (2.28), the coefficient C_i is expressed in terms of g_{ij}, g_{ijk}, etc. given by (2.31) as

$$
\begin{aligned}
C_i &= \sum_{j\neq i}\frac{g_{ij}}{\omega_i - \omega_j} + \sum_{j\neq i}\sum_{\substack{k>j\\ \neq i}}\frac{g_{ijk}}{(\omega_i - \omega_j)(\omega_i - \omega_k)} \\
&\quad + \sum_{j\neq i}\sum_{\substack{k>j\\ \neq i}}\sum_{\substack{l>k\\ \neq i}}\frac{g_{ijkl}}{(\omega_i - \omega_j)(\omega_i - \omega_k)(\omega_i - \omega_l)} + \cdots .
\end{aligned}
\tag{2.34}
$$

When there are degeneracies with respect to ω_i's, we must return to Eq. (2.30) which is quite general. In degenerate cases, the dispersion relation takes a form (1.5). In some cases, however, it seems more convenient to work with the polynomial which is obtained directly from Eq. (2.30), rather than to work with the dispersion relation (1.5). The degenerate cases will be treated in Sec. V.

As the extension of Eq. (2.20) or Eq. (2.29), the general form of the function $U_a(E)$ in Eq. (2.2) is given by

$$
\begin{aligned}
\mathfrak{U}_a(E) &= 1 - \sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\frac{g_{ij}}{(E - \omega_i)(E - \omega_j)} \\
&\quad - \sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\sum_{\substack{k>j\\ \neq a}}\frac{g_{ijk}}{(E - \omega_i)(E - \omega_j)(E - \omega_k)} \\
&\quad - \sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\sum_{\substack{k>j\\ \neq a}}\sum_{\substack{l>k\\ \neq a}} \\
&\quad \times \frac{g_{ijkl}}{(E - \omega_i)(E - \omega_j)(E - \omega_k)(E - \omega_l)} \\
&\quad - \cdots .
\end{aligned}
\tag{2.35}
$$

C. The Calculation of $(b \mid (E - H)^{-1} \mid a)$

The function $(b|\ (E - H)^{-1}\ |a)$ will be calculated up to the second and the third order in Secs. IIC1 and IIC2, respectively. More general cases will be given in Sec. IIC3.

1. Second-Order Expansion

The calculation of the function $(b|\ (E-H)^{-1}\ |a)$ is similar to that of the function $(a|\ (E-H)^{-1}\ |a)$. The relations (2.9), (2.10), and (2.11) will be used throughout:

$$
\begin{aligned}
(b|\,\frac{1}{A}\,|a) &= -\sum_{a'\neq b}\frac{A_{ba'}}{A_b}\,(a'|\,\frac{1}{A}\,|a)\\
&= -\frac{A_{ba}}{A_b}\,(a|\,\frac{1}{A}\,|a)\\
&\quad+\sum_{a'\neq ab}\sum_{a''\neq a'}\frac{A_{ba'}A_{a'a''}}{A_bA_{a'}}\,(a''|\,\frac{1}{A}\,|a)\\
&= \Big(-\frac{A_{ba}}{A_b}+\sum_{a'\neq ab}\frac{A_{ba'}A_{a'a}}{A_bA_{a'}}\Big)(a|\,\frac{1}{A}\,|a)\\
&\quad+\sum_{a'\neq ab}\frac{|A_{ba'}|^2}{A_bA_{a'}}\,(b|\,\frac{1}{A}\,|a)+\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{1}{A_b}\\
&\quad\times S_{a'a''}\Big(\frac{A_{ba'}A_{a'a''}}{A_{a'}}\,(a''|\,\frac{1}{A}\,|a)\Big).
\end{aligned}
\tag{2.36}
$$

Using Eq. (2.11), we have the last term of Eq. (2.36) equal to

$$
\begin{aligned}
&-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{1}{A_b}\,S_{a'a''}\Big(\frac{A_{ba'}A_{a'a''}}{A_{a'}}\\
&\quad\times\sum_{a'''\neq a''}\frac{A_{a''a'''}}{A_{a''}}\,(a'''|\,\frac{1}{A}\,|a)\Big)\\
&=-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{a'''\neq a'a''}\frac{1}{A_bA_{a'}A_{a''}}\,S_{a'a''}\\
&\quad\times(A_{ba'}A_{a'a''}A_{a''a'''})(a'''|\,\frac{1}{A}\,|a)\\
&\quad-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{|A_{a'a''}|^2}{A_bA_{a'}A_{a''}}\,S_{a'a''}\\
&\quad\times\Big(A_{ba'}(a'|\,\frac{1}{A}\,|a)\Big).
\end{aligned}
\tag{2.37}
$$

Using Eq. (2.9), we can further transform the last term of Eq. (2.37) as

$$
\begin{aligned}
&\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{|A_{a'a''}|^2}{A_bA_{a'}A_{a''}}\Big\{A_{ba}(a|\,\frac{1}{A}\,|a)+A_b(b|\,\frac{1}{A}\,|a)\\
&\qquad+\sum_{a'''\neq aba'a''}A_{ba'''}(a'''|\,\frac{1}{A}\,|a)\Big\}.
\end{aligned}
\tag{2.38}
$$

Finally, collecting (2.36), (2.37), and (2.38), we obtain the second-order expansion $(b\,|(E-H)^{-1}|\,a)$ as

$$
\begin{aligned}
&\Big(1-\sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\frac{|A_{ij}|^2}{A_iA_j}\Big)(b|\,\frac{1}{A}\,|a)\\
&=\Big\{-\frac{A_{ba}}{A_b}\Big(1-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{|A_{a'a''}|^2}{A_{a'}A_{a''}}\Big)\\
&\quad+\sum_{a'\neq ab}\frac{A_{ba'}A_{a'a}}{A_bA_{a'}}\Big\}(a|\,\frac{1}{A}\,|a)\\
&\quad+\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\Big\{\sum_{a'''\neq aba'a''}\frac{|A_{a'a''}|^2}{A_bA_{a'}A_{a''}}\,A_{ba'''}\\
&\quad-\sum_{a'''\neq a'a''}\frac{S_{a'a''}(A_{ba''}A_{a''a'}A_{a'a'''})}{A_bA_{a'}A_{a''}}\Big\}(a'''|\,\frac{1}{A}\,|a).
\end{aligned}
\tag{2.39}
$$

2. Third-Order Expansion

Two terms in the last curly brackets of Eq. (2.39) are now transformed to obtain the third-order expansion.

$$
\begin{aligned}
&\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\Big\{\sum_{a'''\neq aba'a''}\frac{|A_{a'a''}|^2}{A_bA_{a'}A_{a''}}\,A_{ba'''}\\
&\quad-\sum_{a'''\neq a'a''}\frac{S_{a'a''}(A_{ba''}A_{a''a'}A_{a'a'''})}{A_bA_{a'}A_{a''}}\Big\}(a'''|\,\frac{1}{A}\,|a)\\
&=-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{S_{a'a''}(A_{ba'}A_{a'a''}A_{a''a})}{A_bA_{a'}A_{a''}}\,(a|\,\frac{1}{A}\,|a)\\
&\quad-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{S_{a'a''}(A_{ba'}A_{a'a''}A_{a''b})}{A_bA_{a'}A_{a''}}\,(b|\,\frac{1}{A}\,|a)\\
&\quad+\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{a'''\neq aba'a''}\Big\{\frac{|A_{a'a''}|^2\,A_{ba'''}}{A_bA_{a'}A_{a''}}\\
&\quad-\frac{S_{a'a''}(A_{ba'}A_{a'a''}A_{a''a'''})}{A_bA_{a'}A_{a''}}\Big\}(a'''|\,\frac{1}{A}\,|a).
\end{aligned}
\tag{2.40}
$$

Analogous to (2.22), we define $\theta'_{ba'a''a'''}$ by

$$
\begin{aligned}
\theta'_{ba'a''a'''} &= \{S_{a'a''}(A_{ba'}A_{a'a''}A_{a''a'''})\\
&\quad-|A_{a'a''}|^2\,A_{ba'''}\}\,\frac{1}{A_bA_{a'}A_{a''}}\,(a'''|\,\frac{1}{A}\,|a).
\end{aligned}
$$

The last term can be transformed as

$$
\begin{aligned}
&-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}}S_{a'a''a'''}\theta'_{ba'a''a'''}\\
&=-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}}\sum_{a^{IV}\neq a'a''a'''}S_{a'a''a'''}\\
&\quad\times\Big[\Big\{\frac{|A_{a'a''}|^2\,A_{ba'''}}{A_bA_{a'}A_{a''}}-\frac{S_{a'a''}(A_{ba'}A_{a'a''}A_{a''a'''})}{A_bA_{a'}A_{a''}}\Big\}\\
&\quad\times\frac{A_{a'''a^{IV}}}{A_{a'''}}\,(a^{IV}|\,\frac{1}{A}\,|a)\Big]\\
&\quad-\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}}S_{a'a''a'''}[\{|A_{a'a''}|^2\,A_{ba'''}
\end{aligned}
$$

$$- S_{a'a'''}(A_{ba'}A_{a'a''}A_{a''a'''})\}$$
$$\times \left\{A_{a'''a'}(a'|\frac{1}{A}|a) + A_{a'''a''}\right.$$
$$\left.\times (a''|\frac{1}{A}|a)\right\}\Big] \frac{1}{A_bA_{a'}A_{a''}A_{a'''}}. \quad (2.41)$$

The last term is further transformed as

$$\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}}\Big[S_{a''a'''}(A_{a'a'''}A_{a'''a''}A_{a''a'})$$
$$\times \left\{A_{ba'}(a'|\frac{1}{A}|a) + A_{ba''}(a''|\frac{1}{A}|a)\right.$$
$$\left.+ A_{ba'''}(a'''|\frac{1}{A}|a)\right\}\Big] \frac{1}{A_bA_{a'}A_{a''}A_{a'''}}$$
$$= -\sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}} S_{a''a'''}(A_{a'a'''}A_{a'''a''}A_{a''a'})$$
$$\times \left\{A_{ba}(a|\frac{1}{A}|a) + A_b(b|\frac{1}{A}|a)\right.$$
$$\left.+ \sum_{a^{IV}\neq aba'a''a'''} A_{ba^{IV}}(a^{IV}|\frac{1}{A}|a)\right\} \frac{1}{A_bA_{a'}A_{a''}A_{a'''}}. \quad (2.42)$$

Thus, collecting Eqs. (2.39), (2.40), (2.41), and (2.42), we obtain

$$\Big\{1 - \sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\frac{g_{ij}}{A_iA_j}$$
$$- \sum_{i\neq a}\sum_{\substack{j>i\\ \neq a}}\sum_{\substack{k>j\\ \neq a}}\frac{g_{ijk}}{A_iA_jA_k}\Big\}(b|\frac{1}{A}|a)$$
$$= \Big\{\frac{V_{ba}}{A_b}\Big(1 - \sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{g_{a'a''}}{A_{a'}A_{a''}}$$
$$- \sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}}\frac{g_{a'a''a'''}}{A_{a'}A_{a''}A_{a'''}}\Big)$$
$$+ \sum_{a'\neq ab}\frac{V_{ba'}V_{a'a}}{A_bA_{a'}}$$
$$+ \sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\frac{S_{a'a''}(V_{ba''}V_{a''a'}V_{a'a})}{A_bA_{a'}A_{a''}}\Big\}(a|\frac{1}{A}|a)$$
$$- \sum_{a'\neq ab}\sum_{\substack{a''>a'\\ \neq ab}}\sum_{\substack{a'''>a''\\ \neq ab}}\frac{1}{A_bA_{a'}A_{a''}A_{a'''}}$$
$$\times [\sum_{a^{IV}\neq aba'a''a'''} A_{ba^{IV}}S_{a''a'''}(A_{a'a'''}A_{a'''a''}A_{a''a'})$$
$$+ \sum_{a^{IV}\neq a'a''a'''} S_{a'a''a'''}[\{|A_{a'a''}|^2 A_{ba'''}$$
$$- S_{a'a''}(A_{ba'}A_{a'a''}A_{a''a'''})\}$$
$$\times (a'''|A|a^{IV})]](a^{IV}|\frac{1}{A}|a). \quad (2.43)$$

3. General Formula

The extension of this procedure to higher-order terms leads to the relation (2.5) with U_a given by Eq. (2.35), and $\mathfrak{V}_{ba}$ given by

$$\mathfrak{V}_{ba}(E) = \frac{V_{ba}}{E-\omega_b}\Big\{1 - \sum_{i\neq ab}\sum_{\substack{j>i\\ \neq ab}}\frac{g_{ij}}{(E-\omega_i)(E-\omega_j)}$$
$$- \sum_{i\neq ab}\sum_{\substack{j>i\\ \neq ab}}\sum_{\substack{k>j\\ \neq ab}}\frac{g_{ijk}}{(E-\omega_i)(E-\omega_j)(E-\omega_k)} - \cdots\Big\}$$
$$+ \sum_{i\neq ab}\frac{h_{bia}}{(E-\omega_b)(E-\omega_i)}$$
$$+ \sum_{i\neq ab}\sum_{\substack{j>i\\ \neq ab}}\frac{h_{bija}}{(E-\omega_b)(E-\omega_i)(E-\omega_j)}$$
$$+ \sum_{i\neq ab}\sum_{\substack{j>i\\ \neq ab}}\sum_{\substack{k>j\\ \neq ab}}\frac{h_{bijka}}{(E-\omega_b)(E-\omega_i)(E-\omega_j)(E-\omega_k)}$$
$$+ \cdots. \quad (2.44)$$

Here h_{bia}, h_{bija}, h_{bijka}, etc. are related to g_{ai}, g_{aij}, g_{aijk}, etc. in such a manner that the "final" state of h_{bia} is $|b)$ whereas that of g_{ai} is $|a)$. Thus,

$$h_{bia} = V_{bi}V_{ia},$$
$$h_{bija} = S_{ij}(V_{bj}V_{ji}V_{ia}),$$
$$h_{bijka} = S_{ijk}\{S_{jk}(V_{bk}V_{kj}V_{ji})V_{ia} - |V_{jk}|^2 V_{bi}V_{ia}\},$$
$$h_{bijkla} = S_{ijkl}\{S_{jkl}(S_{kl}(V_{bl}V_{lk}V_{kj})V_{ji}V_{ia}$$
$$- |V_{kl}|^2 V_{bj}V_{ji}V_{ia})$$
$$- S_{kl}(V_{jl}V_{lk}V_{kj})V_{bi}V_{ia}\}, \quad \text{etc.} \quad (2.45)$$

The set of the wavefunctions $|\lambda)$ is related to another set of the wavefunctions $|a)$ by

$$|\lambda) = \sum_a |a)\cdot(a|\lambda). \quad (2.46)$$

When we use the transformation coefficient $(a|\lambda)$ obtained by (2.4) and (2.7), the wavefunctions $|\lambda)$ form a complete orthonormal set, if the wavefunctions $|a)$ form another complete orthonormal set. This is evident from the general discussion developed in Sec. IIa.

The simplest example is the case of only two states. Writing the unperturbed wavefunctions as $|1)$ and $|2)$ and the wavefunction of the system as $|+)$ and $|-)$, we get from (2.4)

$$|(1|\pm)|^2 = \frac{(E_\pm - \omega_1)}{(E_\pm - E_\mp)}, \qquad |(2|\pm)|^2 = \frac{(E_\pm - \omega_2)}{(E_\pm - E_\mp)}.$$

The orthonormality of $|1)$ and $|2)$ leads to

$$(\pm|\pm) = |(1|\pm)|^2 + |(2|\pm)|^2 = \frac{2E_\pm - (\omega_1 + \omega_2)}{E_\pm - E_\mp} = 1, \quad (2.47)$$

because

$$E_\pm = \tfrac{1}{2}[(\omega_1 + \omega_2) \pm \{(\omega_1 - \omega_2)^2 + 4|V_{12}|^2\}^{\frac{1}{2}}]. \quad (2.48)$$

Also using (2.7), we get

$$(2|-) = \frac{V_{21}}{E_- - \omega_2}(1|-), \qquad (+|2) = \frac{V_{12}}{E_+ - \omega_2}(+|1). \quad (2.49)$$

Thus we can show the orthogonality of $|+1)$ to $|-)$ as

$$(+|-) = (+|1)(1|-) + (+|2)(2|-) = (+|1)(1|-)\times\left\{1 + \frac{|V_{21}|^2}{(E_- - \omega_2)(E_+ - \omega_2)}\right\} = 0. \quad (2.50)$$

III. METHODS OF SOLVING THE DISPERSION RELATION

It is a special feature of the dispersion relation that it can be solved in many ways. Here we shall present three methods, each of which takes a very simple form and has a wide range of applicability. Some particular problems may lie within a common region of the convergence radius of these formulas.

In Secs. IIIA and IIIB, we shall give formulas for the weak coupling case and the strong coupling case, respectively. In Sec. IIIC, we shall give the "nearest-neighbor-states" approximation successfully applied to the case, where the coefficient C_i of the dispersion relation hardly converges.

A. Weak Coupling Case

Let us consider that all $|C_i|$'s are almost an equal order of magnitude smaller than $|E - \omega_i|$. Meanwhile, we shall multiply C_i by a parameter λ. Then the dispersion relation (1.4) can be expressed as

$$\prod_i (E - \omega_i) = \lambda \sum_i C_i \prod_{j\neq i}{}' (E - \omega_j). \quad (3.1)$$

The perturbed energy eigenvalue E is written

$$E = \sum_{n=0}^{\infty} \lambda^n E^{(n)} \quad (3.2)$$

We equate the coefficient of equal power of λ on both sides of Eq. (3.1). First, for $n = 0$,

$$\prod_i (E^{(0)} - \omega_i) = 0.$$

Accordingly, $E^{(0)}$ is any one of ω_i's. Hereafter, we suppose

$$E^{(0)} = \omega_i. \quad (3.3)$$

Then, for $n = 1$,

$$(E^{(1)} - C_i) \prod_{j\neq i} (E^{(0)} - \omega_j) = 0,$$

so

$$E^{(1)} = C_i. \quad (3.4)$$

For $n = 2$,

$$E^{(2)} \prod_{j\neq i} (\omega_i - \omega_j) = C_i \sum_{j\neq i} C_j \prod_{k\neq i,j} (\omega_i - \omega_k).$$

Thus,

$$E^{(2)} = \sum_{j\neq i} \frac{C_i C_j}{\omega_i - \omega_j}. \quad (3.5)$$

For $n = 3$,

$$E_i^{(3)} = C_i\left\{\left(\sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j}\right)^2 - \sum_{j\neq i} \frac{C_i C_j}{(\omega_i - \omega_j)^2}\right\}. \quad (3.6)$$

Hence, up to the third order,

$$E_i \approx \omega_i + C_i + \sum_{j\neq i} \frac{C_i C_j}{\omega_i - \omega_j} + C_i\left\{\left(\sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j}\right)^2 - \sum_{j\neq i} \frac{C_i C_j}{(\omega_i - \omega_j)^2}\right\}. \quad (3.7)$$

We can get this expression also from the third-order expansion with respect to C_i of

$$E = \omega_i + \frac{C_i}{1 - \sum_{j\neq i} \frac{C_j}{\omega_i + C_i - \omega_j}}. \quad (3.8)$$

Up to the fourth order, we get

$$E_i \approx \omega_i + C_i\left\{1 + \sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j} + \left(\sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j}\right)^2 + \left(\sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j}\right)^3\right\} - C_i^2 C_j \sum_{j\neq i} \frac{1}{(\omega_i - \omega_j)^2}\left\{1 + \sum_{k\neq i} \frac{C_k}{\omega_i - \omega_k}\right\} - C_i^3 C_j \sum_{j\neq i} \frac{1}{(\omega_i - \omega_j)^3} - 2\left(\sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j}\right)^2 \frac{C_i^2}{\omega_i - \omega_j}. \quad (3.9)$$

Again, we can get this expression also by the fourth-

order expansion of

$$E = \omega_i + \cfrac{C_i}{1 - \sum_{j\neq i} \cfrac{C_j}{\omega_i + \cfrac{C_i}{1 - \sum_{k\neq i} \cfrac{C_k}{\omega_i - \omega_k}} - \omega_j}}. \tag{3.10}$$

Equations (3.8) and (3.10) suggest that in the higher-order limit, we may have

$$E = \omega_i + \cfrac{C_i}{1 - \sum_{j\neq i} \cfrac{C_j}{E - \omega_j}}.$$

This is nothing but a transformation of the dispersion relation (1.4). This consideration leads to the iteration formula

$$E^{(n+1)} = \omega_i + \cfrac{C_i}{1 - \sum_{j\neq i} \cfrac{C_j}{E^{(n)} - \omega_j}}, \tag{3.11}$$

with the starting value

$$E^{(0)} = \omega_i.$$

Now, (n) in the formula (3.11) means the order of iteration. This is different from (n) in the expansion (3.2).

This formula is especially useful, when all C_i's are of equal order of magnitude, subject to the condition $1 > C_i/(\omega_{i+1} - \omega_i)$, because in such a case, E_i is not appreciably shifted from ω_i towards ω_{i+1} or ω_{i-1}. This formula is closely related to the conventional perturbation formulas.

The iteration of this formula converges uniformly:

$$\omega_i < E^{(1)} < E^{(2)} < \cdots < E < \omega_{i+1} \qquad (C_i > 0),$$

or

$$\omega_i > E^{(1)} > E^{(2)} > \cdots > E > \omega_{i-1} \qquad (C_i < 0)$$

when the condition

$$1 - \frac{C_i}{\omega_{i+1} - \omega_i} > \sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(1)} - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(2)} - \omega_j} > \cdots \begin{bmatrix} \omega_{i+1} \text{ for } C_i > 0 \\ \omega_{i-1} \text{ for } C_i < 0 \end{bmatrix} \tag{3.12}$$

is satisfied. The iteration formula (3.11) converges oscillating about the true value:

$$\omega_i < E^{(2)} < E^{(4)} < \cdots < E < \cdots < E^{(3)} < E^{(1)} < \omega_{i+1} (C_i > 0),$$

or

$$\omega_i > E^{(2)} > E^{(4)} > \cdots > E > \cdots > E^{(3)} > E^{(1)} > \omega_{i-1} (C_i < 0),$$

when the condition

$$1 - \frac{C_i}{\omega_{i+1} - \omega_i} > \sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(2)} - \omega_j} > \cdots > \sum_{j\neq i} \frac{C_j}{E^{(3)} - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(1)} - \omega_j} \tag{3.13}$$

is satisfied.

This formula is especially useful when applied for the maximum eigenvalue $E_M (> \omega_M)$ and the minimum eigenvalue $E_0 (< \omega_0)$ because for these eigenvalues, the condition for convergence is not so strict as for the intermediate eigenvalues. The condition for the uniform convergence of maximum or minimum eigenvalue is given by

$$1 > \sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(1)} - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(2)} - \omega_j} > \cdots, \tag{3.14}$$

and the condition for the oscillatory convergence is given by

$$1 > \sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(2)} - \omega_j} > \cdots > \sum_{j\neq i} \frac{C_j}{E^{(3)} - \omega_j} > \sum_{j\neq i} \frac{C_j}{E^{(1)} - \omega_j}. \tag{3.15}$$

B. Strong Coupling Case

Although the formula (3.11) is very simple, it fails when $C_i/(\omega_{i+1} - \omega_i) \gg 1$ (ω_{i+1} for $C_i > 0$, and ω_{i-1} for $C_i < 0$), or when $C_{i+1}/(\omega_i - \omega_{i+1}) \gg 1$ [(C_{i-1}, ω_{i-1}), or (C_{i+1}, ω_{i+1}), according as $(C_i > 0, C_{i-1} > 0)$, or $(C_i < 0, C_{i+1} < 0)$]. In such case, the eigenvalue is shifted either strongly to ω_{i+1} (according as $C_i \gtrless 0$) or around half the way of ω_{i+1} from ω_i. See Fig. 1. If we apply the formula (3.11) to these cases, already in the first or in the second-order iteration, the calculated eigenvalue, which must be expected to fall between ω_i and ω_{i+1}, goes beyond ω_{i+1}. Thus any further iteration becomes meaningless. In such cases, the following method is very powerful.

We shall consider the case that $C_i > 0$, $C_{i+1} > 0$, for example. Suppose that first we solved the equation

$$C_i/(E - \omega_i) + C_{i+1}/(E - \omega_{i+1}) = 1 \tag{3.16}$$

exactly. The lower solution E_- lies, of course, in between ω_i and ω_{i+1}. Now taking E_- as the starting value $E^{(1)}$, we shall solve the following equation

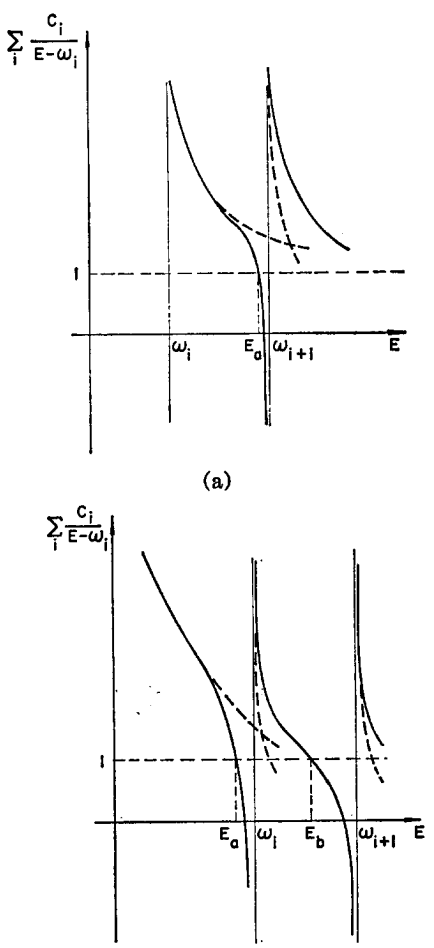

FIG. 1. (a) $c_i/(w_{i+1} - w_i) \gg 1$ and the eigenvalue E_a is very close to w_{i+1}. Here $c_i > 0, c_{i+1} > 0$; (b) $c_{i+1}/(w_i - w_{i-1})/(w_i - w_{i-1}) \gg 1$ and the eigenvalue E_b is around half the way from w_i to w_{i+1}. Here $c_{i-1} > 0$, $c_i > 0$, $c_{i+1} > 0$.

for $E^{(2)}$:

$$E^{(2)} = E_- + \frac{(E^{(1)} - \omega_i)(E^{(1)} - \omega_{i+1})}{E^{(1)} - E_+} \times \left(\sum_{j \neq i, i+1} \frac{C_j}{E^{(1)} - \omega_{i+1}} \right), \quad (3.17)$$

$$E^{(1)} = E_-,$$

where

$$E_\pm = \tfrac{1}{2}[(\omega_i + \omega_{i+1} + C_i + C_{i+1}) \pm \{(\omega_i + \omega_{i+1} + C_i + C_{i+1})^2 - 4(\omega_i\omega_{i+1} + C_i\omega_{i+1} + C_{i+1}\omega_i)\}^{\frac{1}{2}}]. \quad (3.18)$$

We shall repeat the iteration according to the equation

$$E^{(n+1)} = E_- + \frac{(E^{(n)} - \omega_i)(E^{(n)} - \omega_{i+1})}{E^{(n)} - E_+} \times \left(\sum_{j \neq i, i+1} \frac{C_j}{E^{(n)} - \omega_{i+1}} \right). \quad (3.19)$$

These equations, when $E^{(n)}$ converges to some limiting value E, converge to the original dispersion relation. The condition for the unifrom convergence of $E^{(n)}$ is given by

$$\frac{E_+ - E_-}{E_- - \omega_i} > \sum_{j \neq i, i+1} \frac{C_j}{E_- - \omega_j} > 0,$$

and

$$\frac{(\omega_{i+1} - E_-)(E^{(n)} - E_+)}{(E^{(n)} - \omega_i)(E^{(n)} - \omega_{i+1})} > \sum_{j \neq i, i+1} \frac{C_j}{E^{(n)} - \omega_j} > \frac{(E^{(n)} - E_-)(E^{(n)} - E_+)}{(E^{(n)} - \omega_i)(E^{(n)} - \omega_{i+1})}. \quad (3.20)$$

The condition for the oscillatory convergence of $E^{(n)}$ is

$$0 > \sum_{j \neq i, i+1} \frac{C_j}{E_- - \omega_j} > -\frac{E_+ - E_-}{\omega_{i+1} - E_-},$$

and

$$0 > \sum_{j \neq i, i+1} \frac{C_j}{E^{(1)} - \omega_j} > \frac{(E^{(1)} - E_-)(E_- - E_+)}{(E^{(1)} - \omega_i)(E^{(1)} - \omega_{i+1})} > \cdots. \quad (3.21)$$

The formula (3.19) is also applicable for the minimum eigenvalue E_0, if $C_i < 0, C_{i+1} > 0$. The formula for other cases; $\omega_i < E < \omega_{i+1}$ for $C_i > 0, C_{i+1} < 0$ and $\omega_{i-1} < E < \omega_i$ for $C_i < 0, C_{i-1} < 0$, or, for $C_i < 0, C_{i-1} > 0$ are similar to (3.19) and we shall not write them out.

C. "Nearest-Neighbor-States" Approximation

We can consider various methods according to a given problem. This is one of the characteristic features of the present treatment. Let us give a special method.

In some problems, as the second example in Sec. IV, the coefficient C_i may converge poorly. Even in this case, it may be possible to obtain quite accurate results by the following method:

Let $C_i^{(1)}$ include only couplings between the state i and the nearest neighbors $i + 1$ and $i - 1$, and $C_{i\pm1}^{(0)}$ does coupling only between the state $i \pm 1$ and the state i. First we calculate $E^{(1)}$ by the equation

$$\frac{C_i^{(1)}}{E^{(1)} - \omega_i} + \frac{C_{i+1}^{(0)}}{E^{(0)} - \omega_{i+1}} + \frac{C_{i-1}^{(0)}}{E^{(0)} - \omega_{i-1}} = 1, \quad (3.22)$$

where

$$E^{(0)} = \omega_i.$$

Next we calculate $E^{(2)}$ using $E^{(1)}$ thus found:

$$\frac{C_i^{(2)}}{E^{(2)}-\omega_i}+\frac{C_{i+1}^{(1)}}{E^{(1)}-\omega_{i+1}}+\frac{C_{i-1}^{(1)}}{E^{(1)}-\omega_{i-1}}+\frac{C_{i+2}^{(0)}}{E^{(0)}-\omega_{i+2}}+\frac{C_{i-2}^{(0)}}{E^{(0)}-\omega_{i-2}}=1. \quad (3.23)$$

Here, $C_i^{(2)}$ includes the couplings between the state i and $i \pm 1$, $i \pm 2$, $C_{i+1}^{(1)}$ ($C_{i-1}^{(1)}$) the couplings between i, $i+1$, and $i+2$ (i, $i-1$ and $i-2$), $C_{i+2}^{(0)}$ ($C_{i-2}^{(2)}$) the couplings between $i+2$ and $i+1$ ($i-2$ and $i-1$).

The general equation is given by

$$E^{(n+1)}=\omega_i+\frac{C_i^{(n+1)}}{1-\sum_{s=0}^{n}\left(\frac{C_{i+s+1}^{(n-s)}}{E^{(n-s)}-\omega_{i+s+1}}+\frac{C_{i-(s+1)}^{(n-s)}}{E^{(n-s)}-\omega_{i-(s+1)}}\right)}. \quad (3.24)$$

IV. EXAMPLES; COMPARISON WITH OTHER PERTURBATION METHODS

In this section, four examples will be given. In the first two examples we know the exact solution, while in the other two, only numerical solutions have been known. In the Appendix, Rayleigh–Schrödinger (R–S) formula, the Brillouin–Wigner (B–W) formula, and the Feenberg–Feshbach (F–F) formula are reproduced. We shall compare these formulas with the present one.

A. Case of Two States

When only two states $|1\rangle$ and $|2\rangle$ are the unperturbed states, the exact solutions are well known:

$$E_\pm=\tfrac{1}{2}\{(\omega_1+\omega_2)\pm[(\omega_1-\omega_2)^2+4g]^{\frac{1}{2}}\}, \quad (4.1)$$

where

$$\omega_i=\epsilon_i+V_{ii} \qquad (i=1 \text{ and } 2),$$

and

$$g=|V_{12}|^2.$$

The method described in Sec. IIIB also gives the exact solution.

We shall compare various perturbation theories with the expansion of (4.1) in powers of the perturbation interaction V. Putting

$$a=V_{22}-V_{11},$$

$$E_+=\omega_1+\frac{g}{\omega_1-\omega_2}-\frac{g^2}{(\omega_1-\omega_2)^3}+\frac{2g^3}{(\omega_1-\omega_2)^5}-\frac{5g^4}{(\omega_1-\omega_2)^7}+\frac{14g^5}{(\omega_1-\omega_2)^9}-\frac{42g^6}{(\omega_1-\omega_2)^{11}}+\cdots$$

$$=\omega_1+\frac{g}{\epsilon_1-\epsilon_2}+\frac{ga}{(\epsilon_1-\epsilon_2)^2}+\frac{ga^2-g^2}{(\epsilon_1-\epsilon_2)^3}+\frac{ga^3-3g^2a}{(\epsilon_1-\epsilon_2)^4}+\frac{ga^4-6g^2a^2+2g^3}{(\epsilon_1-\epsilon_2)^5}+\frac{ga^5-10g^2a^3+10g^3a}{(\epsilon_1-\epsilon_2)^6}+\frac{ga^6-15g^2a^4+30g^3a^2-5g^4}{(\epsilon_1-\epsilon_2)^7}+\frac{ga^7-21g^2a^5+70g^3a^3-35g^4a}{(\epsilon_1-\epsilon_2)^8}+\frac{ga^8-28g^2a^6+140g^3a^4-140g^4a^2+14g^5}{(\epsilon_1-\epsilon_2)^9}+\cdots. \quad (4.2)$$

As shown in the Appendix, (R–S), (B–W), and (F–F) agree up to the second order;

$$E^{(2)}=\omega_1+\frac{g}{\epsilon_1-\epsilon_2}.$$

The third-order and the fourth-order terms are

(R–S): $$E^{(3)}=\omega_1+\frac{g}{\epsilon_1-\epsilon_2}+\frac{ga}{(\epsilon_1-\epsilon_2)^2},$$

$$E^{(4)}=\omega_1+\frac{g}{\epsilon_1-\epsilon_2}+\frac{ga}{(\epsilon_1-\epsilon_2)^2}+\frac{ga^2-g^2}{(\epsilon_1-\epsilon_2)^3};$$

(B–W): $$E^{(3)}=\omega_1+\frac{g}{\omega_1-\epsilon_2}+\frac{gV_{22}}{(\epsilon_1-\epsilon_2)^2},$$

$$=\omega_1+\frac{g}{\epsilon_1-\epsilon_2}+\frac{ga}{(\epsilon_1-\epsilon_2)^2}+\frac{gV_{11}^2}{(\epsilon_1-\epsilon_2)^3}\left\{1-\frac{V_{11}}{\epsilon_1-\epsilon_2}+\cdots\right\},$$

$$E^{(4)}=\omega_1+\frac{g}{\omega_1+\frac{g}{\epsilon_1-\epsilon_2}-\epsilon_2}+\frac{gV_{22}}{(\omega_1-\epsilon_2)^2}+\frac{gV_{22}^2}{(\epsilon_1-\epsilon_2)^3}$$

$$=\omega_1+\frac{g}{\epsilon_1-\epsilon_2}+\frac{ga}{(\epsilon_1-\epsilon_2)^2}+\frac{ga^2-g^2}{(\epsilon_1-\epsilon_2)^3}+\frac{gV_{11}(2g+3V_{22}V_{11}-V_{11}^2)}{(\epsilon_1-\epsilon_2)^4}+\cdots,$$

$$\text{(F–F):}\quad E^{(3)} = \omega_1 + \frac{g}{\omega_1 - \omega_2}$$

$$= \omega_1 + \frac{g}{\epsilon_1 - \epsilon_2} + \frac{ga}{(\epsilon_1 - \epsilon_2)^2} + \frac{ga^2}{(\epsilon_1 - \epsilon_2)^3} + \cdots,$$

$$E^{(4)} = \omega_1 + \frac{g}{\left(\omega_1 + \frac{g}{\epsilon_1 - \epsilon_2}\right) - \omega_2}$$

$$= \omega_1 + \frac{g}{\epsilon_1 - \epsilon_2} + \frac{ga}{(\epsilon_1 - \epsilon_2)^2} + \frac{ga^2 - g^2}{(\epsilon_1 - \epsilon_2)^3} + \frac{ga^3 - 2g^2a}{(\epsilon_1 - \epsilon_2)^4} + \cdots.$$

Thus, in all of these perturbation methods, $E^{(n)}$ is correct up to the nth power of V. By the way, it is interesting to see that (B–W) gives better results than (R–S), and (F–F) gives much better results than (R–S) and (B–W).

(The present method):

$$C_1 = \frac{g}{\omega_1 - \omega_2} \qquad C_2 = -C_1,$$

$$E^{(0)} = \omega_1,$$

$$E^{(1)} = \omega_1 + \frac{C_1}{1 + \frac{C_1}{\omega_1 - \omega_2}}$$

$$= \omega_1 + \frac{g}{(\omega_1 - \omega_2)\left\{1 + \frac{g}{(\omega_1 - \omega_2)^2}\right\}}$$

$$= \omega_1 + \frac{g}{\omega_1 - \omega_2}\left\{1 - \frac{g}{(\omega_1 - \omega_2)^2} + \frac{g^2}{(\omega_1 - \omega_2)^4} - \cdots\right\}$$

$$= \omega_1 + \frac{g}{\epsilon_1 - \epsilon_2} + \frac{ga}{(\epsilon_1 - \epsilon_2)^2} + \frac{ga^2 - g^2}{(\epsilon_1 - \epsilon_2)^3} + \frac{ga^3 - 3g^2a}{(\epsilon_1 - \epsilon_2)^4} + \frac{ga^4 - 6g^2a^2 + g^3}{(\epsilon_1 - \epsilon_2)^5} + \cdots.$$

Thus, the first iteration of the present method gives a result which is correct up to the fifth power and almost correct up to the sixth power with respect to V. The second iteration is

$$E^{(2)} = \omega_1 + \cfrac{C_1}{1 + \cfrac{C_1}{\omega_1 + \cfrac{C_1}{1 + \cfrac{C_1}{\omega_1 - \omega_2}} - \omega_2}}$$

$$= \omega_1 + \frac{g}{\omega_1 - \omega_2} - \frac{g^2}{(\omega_1 - \omega_2)^3} + \frac{2g^3}{(\omega_1 - \omega_2)^5} - \frac{5g^4}{(\omega_1 - \omega_2)^7} + \frac{12g^5}{(\omega_1 - \omega_2)^9} - \cdots.$$

Further expansion in powers of a gives a result which is correct up to the ninth power with respect to V. The tenth power of V is

$$(ga^8 - 28g^2a^6 + 140g^3a^4 - 140g^4a^2 + 12g^5) \times (\epsilon_1 - \epsilon_2)^{-9},$$

so the difference in the tenth power between this term and the true one as given in (4.2) is very small.

B. The One-Dimensional Harmonic Oscillator

In this section we shall consider the equation

$$-\frac{1}{2m}\frac{d^2U}{dx^2} + \tfrac{1}{2}m(\omega_0^2 + \lambda^2)x^2U = EU, \quad (\hbar = 1). \tag{4.3}$$

The exact solution of this equation is

$$E = (n + \tfrac{1}{2})(\omega_0^2 + \lambda^2)^{\frac{1}{2}}. \tag{4.4}$$

We shall compare the various perturbation methods with the expansion of (4.4) into the power of $g \equiv (\lambda/\omega_0)^2$:

$$E = (n + \tfrac{1}{2})\omega_0(1 + \tfrac{1}{2}g - \tfrac{1}{8}g^2 + \tfrac{1}{16}g^3 - \tfrac{5}{128}g^4 + \tfrac{7}{256}g^5 - \cdots).$$

Up to the second order, (R–S), (B–W), and (F–F) give

$$E^{(0)} = (n + \tfrac{1}{2})\omega_0,$$

$$E^{(1)} = (n + \tfrac{1}{2})\omega_0(1 + \tfrac{1}{2}g),$$

$$E^{(2)} = (n + \tfrac{1}{2})\omega_0(1 + \tfrac{1}{2}g - \tfrac{1}{8}g^2).$$

The third and the fourth order are

$$\text{(R–S):}\quad E^{(3)} = (n + \tfrac{1}{2})\omega_0(1 + \tfrac{1}{2}g - \tfrac{1}{8}g^2 + \tfrac{1}{16}g^3),$$

$$E^{(4)} = (n + \tfrac{1}{2})\omega_0(1 + \tfrac{1}{2}g - \tfrac{1}{8}g^2 + \tfrac{1}{16}g^3 - \tfrac{5}{128}g^4);$$

$$\text{(B–W):}\quad E^{(3)} = (n + \tfrac{1}{2})\omega_0\left\{1 + \tfrac{1}{2}g - \tfrac{1}{8}g^2\frac{1 + \frac{1}{8}g(n^2 + n + 1)}{1 - \frac{1}{16}g^2(n + \frac{1}{2})^2} + \tfrac{1}{64}g^3(n^2 + n + 5)\right\}$$

$$= (n + \tfrac{1}{2})\omega_0\left\{1 + \tfrac{1}{2}g - \tfrac{1}{8}g^2 + \tfrac{1}{16}g^3 - \frac{g^4}{128}(n + \tfrac{1}{2})^2 + \cdots\right\},$$

$$E^{(4)} = (n+\tfrac{1}{2})\omega_0\Big[1+\tfrac{1}{2}g+\tfrac{1}{2}g^2 \times \frac{1+\frac{1}{8}g(n^2+n+1)(1-\frac{1}{4}g)}{-4+(n+\frac{1}{2})^2(\frac{1}{2}g-\frac{1}{8}g^2)^2} + \tfrac{1}{4}g^3\frac{(n^2+n+5)(1+(n+\frac{1}{2})^2\frac{1}{16}g^2)+2(n^2+n+\frac{5}{8})g}{(4-(n+\frac{1}{2})^2\frac{1}{4}g^2)^2} - \frac{1}{16^2}g^4\Big\{7(n^2+n)+\frac{31}{2}\Big\}\Big]$$

$$= (n+\tfrac{1}{2})\omega_0\Big\{1+\tfrac{1}{2}g-\tfrac{1}{8}g^2+\tfrac{1}{16}g^3 - \frac{5}{128}g^4 - \frac{(n+\frac{1}{2})^2}{512}(n^2+n+3)+\cdots\Big\};$$

(F–F): $$E^{(3)} = (n+\tfrac{1}{2})\omega_0\Big\{1+\tfrac{1}{2}g-\tfrac{1}{8}g^2\frac{1}{1+\frac{1}{2}g}\Big\}$$

$$= (n+\tfrac{1}{2})\omega_0(1+\tfrac{1}{2}g-\tfrac{1}{8}g^2 + \tfrac{1}{16}g^3-\tfrac{1}{32}g^4+\cdots),$$

$$E^{(4)} = (n+\tfrac{1}{2})\omega_0[(1+\tfrac{1}{2}g) + \tfrac{1}{16}g^2\Big\{\frac{(n+1)(n+2)}{-2(1+\frac{1}{2}g)-\frac{1}{8}g^2((n+\frac{1}{2})-\frac{1}{8}(n+3)(n+4))} + \frac{n(n-1)}{2(1+\frac{1}{2}g)-\frac{1}{8}g^2((n+\frac{1}{2})+\frac{1}{8}(n-2)(n-3))}\Big\}]$$

$$= (n+\tfrac{1}{2})\omega_0(1+\tfrac{1}{2}g-\tfrac{1}{8}g^2+\tfrac{1}{16}g^3 - \tfrac{5}{128}g^4+\tfrac{1}{128}g^5-\cdots).$$

Thus, all of these perturbation methods are correct up to g^3 in the third order, and up to g^4 in the fourth order.

(The present method): The method described in Sec. IIIC will be applied. Taking

$$E^{(0)} = \omega_n = (n+\tfrac{1}{2})\omega_0 a, \quad a = 1+\tfrac{1}{2}g \tag{4.5}$$

$$C_n^{(1)} = \frac{|V_{nn+2}|^2}{\omega_n-\omega_{n+2}} + \frac{|V_{nn-2}|^2}{\omega_n-\omega_{n-2}} = -\frac{\omega_0 g^2}{8a} \times (n+\tfrac{1}{2}), \tag{4.6}$$

$$C_{n+2}^{(0)} = \frac{|V_{n+2\,n}|^2}{\omega_{n+2}-\omega_n} = \frac{\omega_0 g^2(n+1)(n+2)}{32a},$$

$$C_{n-2}^{(0)} = \frac{|V_{n-2\,n}|^2}{\omega_{n-2}-\omega_n} = -\frac{\omega_0 g^2 n(n-1)}{32a}, \tag{4.7}$$

we calculate the first iteration as

$$E^{(1)} = \omega_n + \frac{C_n^{(1)}}{1-\left(\frac{C_{n+2}^{(0)}}{\omega_n-\omega_{n+2}}+\frac{C_{n-2}^{(0)}}{\omega_n-\omega_{n-2}}\right)} = (n+\tfrac{1}{2})\omega_0 a\left[1-\frac{g^2}{8a^2}\frac{1}{1+\frac{g^2}{32a^2}(n^2+n+1)}\right]$$

$$= (n+\tfrac{1}{2})\omega_0(1+\tfrac{1}{2}g-\tfrac{1}{8}g^2+\tfrac{1}{16}g^3-\cdots). \tag{4.8}$$

In the second iteration, we shall take

$$C_n^{(3)} = \frac{|V_{nn+2}|^2}{\omega_n-\omega_{n+2}} + \frac{|V_{nn-2}|^2}{\omega_n-\omega_{n-2}} - \frac{|V_{n-4n-2}|^2\,|V_{nn+2}|^2}{(\omega_n-\omega_{n-4})(\omega_n-\omega_{n-2})(\omega_n-\omega_{n+2})} - \frac{|V_{nn-2}|^2\,|V_{n+2n+4}|^2}{(\omega_n-\omega_{n-2})(\omega_n-\omega_{n+2})(\omega_n-\omega_{n+4})}. \tag{4.9}$$

In deriving $C_n^{(3)}$, we used Eq. (2.31). $C_{n\pm2}^{(1)}$ and $C_{n\pm4}^{(0)}$ are obtained by putting $n\pm2$ in place of n in (4.6) and $n\pm4$ in place of $n\pm2$ in (4.7), respectively.

We calculate the second iteration as

$$E^{(2)} = \omega_n + \frac{C_n^{(2)}}{1-\left(\frac{C_{n+2}^{(1)}}{\omega_n-\omega_{n+2}}+\frac{C_{n-2}^{(1)}}{\omega_n-\omega_{n-2}}+\frac{C_{n+4}^{(0)}}{\omega_n-\omega_{n+4}}+\frac{C_{n-4}^{(0)}}{\omega_n-\omega_{n-4}}\right)}$$

$$= (n+\tfrac{1}{2})\omega_0 a\left[1-\frac{g^2}{8a^2}\frac{\left[1+\frac{g^2}{64a^2}(n^2+n-3)\right]}{1+\frac{g^2}{4a^2}\left\{\frac{4-(n+\frac{1}{2})^2\frac{g^2}{8a^2}\frac{1}{\left[1+\frac{g^2}{32a^2}(n^2+n+1)\right]}}{4-(n+\frac{1}{2})^2\left[\frac{g^2}{8a^2}\frac{1}{\left[1+\frac{g^2}{32a^2}(n^2+n+1)\right]}\right]^2}+\frac{n^2+n+9}{16}\right\}}\right]$$

$$= (n+\tfrac{1}{2})\omega_0(1+\tfrac{1}{2}g-\tfrac{1}{8}g^2+\tfrac{1}{16}g^3-\tfrac{5}{128}g^4+\tfrac{7}{256}g^5-\cdots). \tag{4.10}$$

TABLE I. List of C_i's for the Mathieu equation (4.11). In calculating C_i, the cutoff was made at $n = 100$, so that the result is correct up to five significant figures.

$$
\begin{aligned}
C_0 &= -0.03125\,S^2\,(1 - 0.000155724\,S^2 + 0.000000003\,S^4)\\
C_2 &= 0.0260417\,S^2 - 0.0000065\,S^4\\
C_4 &= 0.00208333\,S^2 + 0.00000062\,S^4\\
C_6 &= 0.000892857\,S^2 + 0.000000045\,S^4\\
C_8 &= 0.000496032\,S^2 + 0.000000075\,S^4\\
C_{10} &= 0.000315659\,S^2 + 0.000000028\,S^4\\
C_{12} &= 0.000218532\,S^2 + 0.000000006\,S^4\\
C_{14} &= 0.000160256\,S^2 + 0.0000000003\,S^4\\
C_{16} &= 0.000122549\,S^2 + 0.0000000001\,S^4\\
C_{18} &= 0.0000967492\,S^2\\
C_{20} &= 0.0000783208\,S^2\\
C_{22} &= 0.0000646998\,S^2\\
C_{24} &= 0.0000543478\,S^2\\
C_{26} &= 0.0000462963\,S^2\\
C_{28} &= 0.0000399106\,S^2\\
C_{30} &= 0.0000347609\,S^2\\
C_{32} &= 0.0000305474\,S^2\\
C_{34} &= 0.0000270563\,S^2\\
C_{36} &= 0.0000241313\,S^2\\
C_{38} &= 0.0000216563\,S^2\\
C_{40} &= 0.0000195435\,S^2
\end{aligned}
$$

Thus the present perturbation is expected to be correct up to g^{2n+1} in the nth order iteration.

C. The Mathieu Equation

The matrix elements between the solutions for $S = 0$ of the Mathieu equation

$$(d^2\psi/d\theta^2) + [b - S\cos^2\theta]\psi = 0, \qquad \psi(0) = \psi(2\pi) \tag{4.11}$$

are given by

$$
\begin{aligned}
U_{mn} &= \tfrac{1}{2}\delta_{mn} + \tfrac{1}{4}(\delta_{m,n-2} + \delta_{m,n+2}), \quad m \text{ and } n \neq 0,\\
U_{m0} &= U_{0m} = (8)^{-\frac{1}{2}}\delta_{m,2}, \qquad m \neq 0,\\
U_{00} &= \tfrac{1}{2}.
\end{aligned} \tag{4.12}
$$

The unperturbed eigenvalues are n^2. C_i's defined by (2.31) are given, up to $i = 40$, in Table I. Using these C_i's we applied the formula (3.11) for the weak coupling to obtain the lowest eigenvalue E_0 for $S = 0.2$, 1.0, 2.0, and 4.0; and E_6 for $S = 0.5$, 2.0, and 30.0. The results are shown in Table II and IV, respectively. The results on the basis of the Brillouin–Wigner perturbation theory are quoted from the book by Morse and Feshbach,[8] and given as Tables III and V for E_0 and E_6, respectively. Compare, for example, the case $S = 4.0$ in Tables II and III. Our result is correct up to three figures even in the first iteration, whereas the Brillouin–Wigner method gives a rather poor result even in the fourth order.

TABLE II. Lowest eigenvalues E_0 of the Mathieu equation (4.11) for $S = 0.2$, 1.0, 2.0, and 4.0. In obtaining the results, the formula (3.11) was used. Broken line indicates convergence.

S	$E_0^{(0)}$	$E_0^{(1)}$	$E_0^{(2)}$	$E_0^{(3)}$	$E_0^{(4)}$
0.2	0.1000000	0.0987503			
1.0	0.5000000	0.468962	0.468959		
2.0	1.0000000	0.878326	0.878235	0.878234	
4.0	2.0000000	1.549077	1.544742	1.544704	1.544703

The improved (B–W) method in reference 4, yields faster convergence than the original one and better values for (at least) the ground state eigenvalues. Let us quote the value E_0 for $s = 4.0$ from his lecture note[4]:

Approximation	1	2	3	4	5
The improved (B–W)	$2 > E$	1.268	$1.55051 > E$	1.54429	$1.54487 > E$

Also the method by C. Bloch[4] yields, for $s = 2$,

Approximation	2	3	4
(A/B)	0.8750	0.8710	0.8785
α	0.8750	0.8750	0.8784

and for $s = 4$,

Approximation	2	3	4
(A/B)	1.5	1.4286	1.5625
α	1.5	1.5	1.5547

[8] P. M. Morse and H. Feshbach, *Method of Theoretical Physics* (McGraw-Hill Book Company, Inc., N. Y., 1953), Vol. II, p. 1001.

TABLE III. Lowest eigenvalues E_0 of the Mathieu equation (4.11) for $S = 0.2$, 1.0, 2.0 and 4.0, on the basis of the Brillouin–Wigner perturbation theory. These values are quoted from reference 8, Table 9.1.

S	$b_0^{(0)}$	$b_0^{(1)}$	$b_0^{(2)}$	$b_0^{(3)}$	$b_0^{(4)}$
0.2	0	0.10000	0.09875		
1.0	0	0.50000	0.46430	0.46892	0.46896
2.0	0	1.00000	0.83333	0.89197	0.87367
4.0	0	2.00000	1.00000	1.77778	1.15407

Here the ground state energy E_0 is given by (A/B) or by $\mathfrak{A}$.

At this point, the author adds a remark that as one of the specific features of the present method, the iteration formula keeps the same form as we continue the iteration successively. The length of the calculation is thus unchanged for every step. This makes actual calculation very simple once the C_i's are obtained. Indeed, Table II is equivalent to the solution of 50×50 secular determinant, because, the states up to $n = 100$ have been taken into account. Yet the calculation was performed by the desk computor. For the Mathieu equation, C_i consists only of terms (2) × (2), (2) × (2) × (2), $\cdots$, because of its selection rule. Thus, the calculation of Table I is also a simple work.

D. Nilsson's Problem: Energy Levels in a Deformed Harmonic-Oscillator Potential

In this section we shall calculate some eigenvalues of the Schrödinger equation

$$(H - E)\psi = 0, \qquad H = \dot{H}_0 + \kappa\hbar\dot{\omega}_0 R; \tag{4.13}$$

$$\dot{H}_0 = \hbar\omega_0(\tfrac{1}{2})(-\Delta + r^2),$$

$$R = \eta U - 2\bar{l}\bar{s} - \mu\bar{l}^2, \tag{4.14}$$

$$U = -\tfrac{4}{3}(\pi/5)^{\frac{1}{2}} r^2 Y_{20}.$$

As for notations see the original paper by Nilsson.[9]

TABLE IV. Eigenvalues E_6 of the Mathieu equation (4.11) for $S = 0.5$, 2.0, and 30.0. In obtaining the results, the formula (3.11) was used. Broken line indicates convergence.

S	$E^{(0)}$	$E^{(1)}$	$E^{(2)}$
0.5	36.25	36.250223218	
2.0	37.00	37.00357247	
30.0	51.00	51.75061	51.75116

[9] S. G. Nilsson, Dan. Mat. Fys. Medd. **29**, No. 16 (1955).

TABLE V. Eigenvalues E_6 of the Mathieu equation (4.11) for $S = 0.5$, 2.0, and 30.0 on the basis of the Brillouin–Wigner perturbation theory. These values are quoted from reference 8, Table 9.2.

S	$b_6^{(0)}$	$b_6^{(1)}$	$b_6^{(2)}$	$b_6^{(3)}$
0.5	36	36.25	36.25022	
2.0	36	37.00	37.00356	
30.0	36	51.00	48.27998	53.28531

In the following calculations, we neglect, as did Nilsson, the coupling between states with different N. The matrix element of the operator R were also given by Nilsson.

Here, as an example, we shall calculate eigenvalues of the states with $N = 3$, $\Omega = \frac{1}{2}$ (N: the total number of oscillator quanta; Ω: the third component of the total angular momentum). Four states belonging to this set of quantum numbers are numbered as shown in Table VI. The matrix eliments for this set are

$$R_{11} = -(0.8\eta + 12\mu), \quad R_{22} = -(0.6\eta - 1 + 12\mu),$$

$$R_{33} = -(1.2\eta + 2\mu), \quad R_{44} = -(-0.6\eta - 1 + 2\mu),$$

$$R_{12} = R_{21} = -2\sqrt{3} \quad R_{13} = R_{31} = -(\tfrac{2}{5}\sqrt{6})\eta,$$

$$R_{24} = R_{42} = -0.8\eta, \quad R_{34} = R_{43} = -\sqrt{2},$$

$$R_{14} = R_{41} = R_{32} = R_{23} = 0. \tag{4.15}$$

Writing $\{E^{N\Omega} - (N + \frac{3}{2})\hbar\omega_0\}/k\hbar\dot{\omega}_0$ and $C_i/k\hbar\dot{\omega}_0$ as E and a_i, respectively, the dispersion relation of this problem is

$$\frac{a_1}{E + (0.8\eta + 12\mu)} + \frac{a_2}{E + (0.6\eta - 1 + 12\mu)} + \frac{a_3}{E + (1.2\eta + 2\mu)} + \frac{a_4}{E + (-0.6\eta - 1 + 2\mu)} = 1, \tag{4.16}$$

TABLE VI. Four states considered in example IVD.

State No.	N	l	Λ	Σ
1	3	3	0	$\frac{1}{2}$
2	3	3	1	$-\frac{1}{2}$
3	3	1	0	$\frac{1}{2}$
4	3	1	1	$-\frac{1}{2}$

TABLE VII. Energy level of the states with $N = 3$, $\Omega = \frac{1}{2}$ in a deformed harmonic-oscillator potential with the spin-orbit force and the l^2 force ($\eta = 2$, $\mu = 0.35$).

Energy level	E_1		E_2	E_3	E_4
Formula used	(3.19)	(3.11)	(3.19)	(3.19)	(3.11)
Order of iteration					
0	−9.84929	−5.8	−3.29366	−2.14377	1.5
1	−9.11036	−7.18011	−3.40244	−1.58113	2.61230
2	−9.20785	−8.14681	−3.45724	−1.64632	2.50489
3	−9.19415	−8.69187	−3.47224	−1.63353	2.51225
4	−9.19605	−8.96366	−3.47755	−1.63600	2.51173
5	−9.19578	−9.09100	−3.47939	−1.63552	2.51177
6	−9.19582	−9.14892	−3.48002	−1.63561	
7		−9.17492	−3.48024	−1.63558	
8		−9.18652	−3.48031	−1.63560	
9		−9.19169	−3.48034	−1.63559	
10		−9.19398	−3.48035		
11		−9.19500			
12		−9.19545			
13		−9.19566			
14		−9.19575			
15		−9.19579			
16		−9.19580			
17		−9.19581			
Exact value	−9.19582		−3.48035	−1.63559	2.51177
Nilsson [10]	−9.196		−3.480	−1.636	2.512

where

$$a_1 = \frac{0.96\eta^2}{0.4\eta - 10\mu} - \frac{12}{0.2\eta + 1} + \frac{7.0656\eta^2 - 24}{(0.4\eta - 10\mu)(0.2\eta + 1)(1.4\eta + 10\mu + 1)},$$

$$a_2 = \frac{12}{0.2\eta + 1} - \frac{0.64\eta^2}{1.2\eta + 10\mu} - \frac{7.0656\eta^2 - 24}{(0.2\eta + 1)(0.6\eta - 10\mu + 1)(1.2\eta + 10\mu)},$$

$$a_3 = -\left\{\frac{0.96\eta^2}{0.4\eta - 10\mu} + \frac{2}{1.8\eta + 1} + \frac{7.0656\eta^2 - 24}{(0.4\eta - 10\mu)(0.6\eta - 10\mu + 1)(1.8\eta + 1)}\right\},$$

$$a_4 = \frac{2}{1.8\eta + 1} + \frac{0.64\eta^2}{1.2\eta + 10\mu} + \frac{7.0656\eta^2 - 24}{(1.4\eta + 10\mu + 1)(1.8\eta + 1)(1.2\eta + 10\mu)}.$$

TABLE VIII. Transformation coefficients between the wavefunctions of the system having eigenvalues shown in Table VII and those of the unperturbed states. The second column shows $|(a\,|\lambda = 1)|^2$ obtained from (2.4). The third and the fourth columns show $(a\,|\lambda = 1)$ and $(a\,|\lambda = 2)$, respectively, obtained from (2.7).

a	$\lvert(a\,\lvert\lambda = 1)\rvert^2$	$(a\,\lvert\lambda = 1)$	$(a\,\lvert\lambda = 2)$
1	0.568287	0.753848	0.117783
2	0.343045	0.585701	−0.546625
3	0.073428	0.270974	0.826925
4	0.015238	0.123441	0.059195
Total	0.999998		

Hereafter, we shall give a numerical example for $\eta = 2$ and $\mu = 0.35$. In this case, the dispersion relation is

$$\frac{-9.88093}{E + 5.8} + \frac{7.84787}{E + 4.4} + \frac{1.18008}{E + 3.1} + \frac{0.85298}{E - 1.5} = 1. \qquad (4.17)$$

The results of numerical calculation are given in Table VII.[10] In this problem, the unperturbed state is degenerate, so we cannot apply the Rayleigh–Schrödinger perturbation method, nor the iteration formulas of the Brillouin–Wigner and the Feenberg–Feshbach method.

We shall calculate the transformation coefficients between the wavefunctions of the system and the unperturbed states. First, we shall calculate $|(a|\lambda = 1)|^2$, using Eq. (2.4) with Eq. (2.35);

$$|(a = 1\,|\,\lambda = 1)|^2 = \mathfrak{A}_1(E_{\lambda=1}) \times \frac{(E_1 - \omega_2)(E_1 - \omega_3)(E_1 - \omega_4)}{(E_1 - E_2)(E_1 - E_3)(E_1 - E_4)},$$

[10] It may be interesting to note that the values in Tables VII and VIII were calculated by the desk computor. The use of the present iteration formula for this problem needs *only one page* of a note book for four iterations. Nilsson gave his result by the electronic computor.

$$\mathfrak{U}_1(E_{\lambda=1}) = 1 - \frac{|V_{24}|^2}{(E_1 - \omega_2)(E_1 - \omega_4)} - \frac{|V_{34}|^2}{(E_1 - \omega_3)(E_1 - \omega_4)}, \quad \text{etc.}$$

The results for $\eta = 2$ and $\mu = 0.35$ are given in the second column of Table VIII. The total sum of $|(a|\lambda = 1)|^2$ is shown in the lowest row. This is 0.999998, thus showing the wavefunction $|\lambda = 1)$ is normalized to 1.

Next, we shall calculate $(a|\lambda = 1)$, using Eq. (2.7) with $\mathfrak{V}_{ba}(E)$ given by Eq. (2.44). This formula determines the relative phase between the transformation coefficients $(a|\lambda = 1)$. We take the phase of $(a = 1|\lambda = 1)$ as positive, namely, $(a = 1|\lambda = 1) = (0.568278)^{\frac{1}{2}} = 0.753848$. The numerical results are given in the third column of Table VIII. The square of these values are, of course, in agreement with the values in the second column.

Similar calculations are performed for other wavefunctions. The fourth column of Table VIII shows $(a|\lambda = 2)$, derived by using the formula (2.7) from the value of $(a = 1|\lambda = 2)$, whose square was obtained by Eq. (2.4). One can easily check the orthogonality of the wavefunctions $|\lambda = 1)$ and $|\lambda = 2)$. Using the values in the third and the fourth columns, we obtain $\Sigma_a(\lambda = 2\,|a)\,(a|\lambda = 1) \simeq 0.000001$.

V. DEGENERATE CASE

When some of ω_i's take the same value, we have to go back to Eq. (2.30), from which the dispersion relation originated. In general, the dispersion relation takes the form (1.5). No general method for solving this equation is known. Nevertheless, Eq. (2.30) is very useful.

Example 1: Solve the eigenvalue problem where $V_{nn} = 0$ and $V_{nm} = -a$. (The pairing correlation but without diagonal matrix elements.)In this case,

$$g_{ij} = a^2, \quad g_{ijk} = -2a^3, \quad g_{ijkl} = 3a^4 \quad \text{etc.},$$

$$\sum_i \sum_{j>i} = {}_nC_2, \quad \sum_i \sum_{j>i} \sum_{k>j} = {}_nC_3,$$

$$\sum_i \sum_{j>i} \sum_{k>j} \sum_{l>k} = {}_nC_4, \quad \text{etc.}$$

Thus putting $x = E/a$, from Eq. (2.30) we get the polynomial

$$x^n - {}_nC_2 x^{n-2} + 2{}_nC_3 x^{n-3} - 3{}_nC_4 x^{n-4} + \cdots + (-)^{s-1}(s-1){}_nC_s x^{n-s} + \cdots + (-)^n(n-1) = 0. \tag{5.1}$$

Now multiplying both sides of the equation

$$(x-1)^{n-1} = \sum_{s=1}^{n} (-)^{s-1}{}_{n-1}C_{s-1} x^{n-s} \tag{5.2}$$

by $x + (n-1)$, we get Eq. (5.1) because

$$(-)^s{}_{n-1}C_s + (-)^{s-1}(n-1)_{n-1}C_{s-1} = (-)^{s-1}(s-1)_nC_s. \tag{5.3}$$

As the result, the eigenvalues of this problem are $E = a$, which is $(n-1)$-fold degenerate, and $E = -(n-1)a$.

Example 2: Find the energy equation of the eigenvalue problem, where $V_{n,n+1} = V_{n,n-1} = -a$ and all other matrix elements are zero. (The one-dimensional coupled oscillation.)

If we only take two states n and $n+1$, putting $i = n$ and $j = n+1$ in Eq. (2.30), we get

$$1 - |V_{nn+1}|^2/E^2 = 0,$$

or

$$E^2 - a^2 = 0. \tag{5.4}$$

If we take four states n, $n+1$, $n+2$, and $n+3$, then putting $i = n$, $j = n+1$, $k = n+2$, and $l = n+3$ in Eq. (2.30), we get

$$1 - \frac{|V_{nn+1}|^2 + |V_{n+1,n+2}|^2 + |V_{n+2,n+3}|^2}{E^2} + \frac{|V_{nn+1}|^2|V_{n+2n+3}|^2}{E^4} = 0;$$

that is,

$$E^4 - 3a^2E^2 + a^4 = 0. \tag{5.5}$$

and so on. In general, for N states,

$$\sum_i \sum_{j>i} g_{ij} = (N-1)a^2$$

$$\sum_i \sum_{j>i} \sum_{k>j} \sum_{l>k} g_{ijkl} = -\frac{(N-2)(N-3)}{2!} a^4$$

$$\sum_i \sum_{j>i} \sum_{k>j} \sum_{l>k} \sum_{m>l} \sum_{n>m} g_{ijklmn} = \frac{(N-3)(N-4)(N-5)}{3!} a^6.$$

Hence writing $x = E/a$, we get the polynomial

$$x^N - \frac{(N-1)}{1!} x^{N-2} + \frac{(N-2)(N-3)}{2!} x^{N-4} - \frac{(N-3)(N-4)(N-5)}{3!} x^{N-6} + \cdots + (-)^s \frac{(N-s)(N-s-1)\cdots(N-(2s-1))}{s!} \times x^{N-2s} + \cdots + (-)^{\frac{1}{2}N} = 0. \tag{5.6}$$

In some cases, we can further transform the polynomial obtained from Eq. (2.30) into a dispersion relation of the form (1.4). Then we may apply the methods described in Sect. III.

The following formulas may be used to transform the polynomial to the dispersion relation:

(1) $E^3 - bE - c = 0, \quad (b > 0)$

is transformed into

$$\frac{a_1}{E} + \frac{a_2}{E - b^{\frac{1}{2}}} + \frac{a_3}{E + b^{\frac{1}{2}}} = 1, \tag{5.7}$$

where $a_1 = -c/b$, $a_2 = a_3 = c/2b$.

(2) $E^4 - bE^2 - cE + d = 0, \quad (b > 0, d > 0)$

is transformed into

$$\frac{a_1}{E - b^{\frac{1}{2}}} + \frac{a_2}{E + b^{\frac{1}{2}}} + \frac{a_3}{E - (d/b)^{\frac{1}{2}}} + \frac{a_4}{E + (d/b)^{\frac{1}{2}}} = 1, \tag{5.8}$$

where

$$a_1 = \frac{b^{\frac{1}{2}}(d - cb^{\frac{1}{2}})}{2(d - b^2)}, \qquad a_2 = -\frac{b^{\frac{1}{2}}(d + cb^{\frac{1}{2}})}{2(d - b^2)},$$

$$a_3 = -\frac{b[(d/b)^{\frac{3}{2}} - c]}{2(d - b^2)}, \qquad a_4 = \frac{b[(d/b)^{\frac{3}{2}} + c]}{2(d - b^2)}.$$

(3) $E^5 - bE^3 - cE^2 + dE + e = 0, \quad (b > 0, d > 0)$

is transformed into

$$\frac{a_1}{E} + \frac{a_2}{E - b^{\frac{1}{2}}} + \frac{a_3}{E + b^{\frac{1}{2}}} + \frac{a_4}{E - (d/b)^{\frac{1}{2}}} + \frac{a_5}{E + (d/b)^{\frac{1}{2}}} = 1, \tag{5.9}$$

where

$$a_1 = -\frac{e}{d}, \qquad a_2 = \frac{db^{\frac{1}{2}} - bc + e}{2(d - b^2)},$$

$$a_3 = -\frac{(db^{\frac{1}{2}} + bc - e)}{2(d - b^2)},$$

$$a_4 = -\frac{(d^{5/2}b^{-\frac{1}{2}} - bcd + eb^2)}{2(d - b^2)d},$$

$$a_5 = \frac{d^{5/2}b^{-\frac{1}{2}} + bcd - eb^2}{2(d - b^2)d}.$$

(4) $E^6 - bE^4 - cE^3 + dE^2 + eE - f = 0, \quad (b > 0, d > 0, f > 0)$

is transformed into

$$\frac{a_1}{E - b^{\frac{1}{2}}} + \frac{a_2}{E + b^{\frac{1}{2}}} + \frac{a_3}{E - (d/b)^{\frac{1}{2}}} + \frac{a_4}{E + (d/b)^{\frac{1}{2}}} + \frac{a_5}{E - (f/d)^{\frac{1}{2}}} + \frac{a_6}{E + (f/d)^{\frac{1}{2}}} = 1, \tag{5.10}$$

where

$$a_1 = \frac{b^{\frac{1}{2}}\{(bd - f) + b^{\frac{1}{2}}(e - cb)\}d}{2(d - b^2)(bd - f)},$$

$$a_2 = -\frac{b^{\frac{1}{2}}\{(bd - f) - b^{\frac{1}{2}}(e - cb)\}d}{2(d - b^2)(bd - f)},$$

$$a_3 = \frac{d^{\frac{1}{2}}b^{-\frac{1}{2}}\{(d^3 - fb^3) + b^{\frac{3}{2}}d^{\frac{1}{2}}(eb - cd)\}}{2(d - b^2)(bf - d^2)},$$

$$a_4 = -\frac{d^{\frac{1}{2}}b^{-\frac{1}{2}}\{(d^3 - fb^3) - b^{\frac{3}{2}}d^{\frac{1}{2}}(eb - cd)\}}{2(d - b^2)(bf - d^2)},$$

$$a_5 = \frac{bd^{-\frac{1}{2}}f^{-\frac{1}{2}}\{f^2(f - bd) + f^{\frac{1}{2}}d^{\frac{3}{2}}(ed - cf)\}}{2(bd - f)(bf - d^2)},$$

$$a_6 = -\frac{bd^{-\frac{1}{2}}f^{-\frac{1}{2}}\{f^2(f - bd) - f^{\frac{1}{2}}d^{\frac{3}{2}}(ed - cf)\}}{2(bd - f)(bf - d^2)}.$$

Example 3: Find the solution of *Example 2* for $N = 6$ and compare it with the exact solutions $2\cos s\pi/(N + 1)$, $(N = 6, s = 1, 2, \cdots, 6)$.

In this case, Eq. (5.6) is

$$x^6 - 5x^4 + 6x^2 - 1 = 0. \tag{5.11}$$

Now putting $x^2 = \lambda + \frac{5}{3}$, Eq. (5.11) is transformed as

$$\lambda^3 - (7/3)\lambda - 7/27 = 0. \tag{5.12}$$

By the aid of Eq. (5.7), this equation can be expressed in the form of the dispersion relation

$$-\frac{1}{9\lambda} + \frac{1}{18[\lambda - (7/3)^{\frac{1}{2}}]} + \frac{1}{18[\lambda + (7/3)^{\frac{1}{2}}]} = 1. \tag{5.13}$$

This equation is solved by using the weak coupling formula (3.11).

(1) The solution near $\lambda = (\frac{7}{3})^{\frac{1}{2}}$:

$$\lambda_1^{(0)} = (7/3)^{\frac{1}{2}}, \qquad \lambda_1^{(1)} = (7/3)^{\frac{1}{2}}\left\{1 + \frac{2}{36 + 3(3/7)^{\frac{1}{2}}}\right\},$$

which leads to

$$x_1^{(0)} = \pm 1.78723, \qquad x_1^{(1)} = \pm 1.80191.$$

(2) The solution near $\lambda = 0$:

$$\lambda_2^{(0)} = -\frac{1}{9}, \qquad \lambda_2^{(1)} = -\frac{1}{9}\frac{188}{187},$$

which leads to

$$x_2^{(0)} = \pm 1.24720, \qquad x_2^{(1)} = \pm 1.24698.$$

(3) The solution near $\lambda = -(7/3)^{\frac{1}{2}}$:

$$\lambda_3^{(0)} = -(7/3)^{\frac{1}{2}}$$

$$\lambda_3^{(1)} = -(7/3)^{\frac{1}{2}}\left\{\frac{2}{36 - 3(3/7)^{\frac{1}{2}}}\right\},$$

which leads to

$$x_3^{(0)} = \pm 0.37569, \qquad x_3^{(1)} = \pm 0.44486.$$

These values should be compared with the exact solutions

$$x_1 = \begin{cases} 2\cos(\pi/7) = 1.80192 \\ 2\cos(6\pi/7) = -1.80192, \end{cases}$$

$$x_2 = \begin{cases} 2\cos(2\pi/7) = 1.24695 \\ 2\cos(\pi/7) = -1.24695, \end{cases}$$

$$x_3 = \begin{cases} 2\cos(3\pi/7) = 0.44490 \\ 2\cos(4\pi/7) = -0.44490. \end{cases}$$

Thus even the first iteration gives a very good result.

Example 4: Calculate Nilsson's problem (Sec. IVD.) with $\eta = \mu = 0$, for $N = 5$, $\Omega = \frac{1}{2}$. (The level splitting of the spherical harmonic oscillator due to the spin-orbit force.)

Six states belonging to this set of quantum numbers are shown in Table IX. According to Sec. IVD, the matrix elements are

$$V_{44} = V_{55} = V_{66} = -1, \qquad V_{14} = V_{41} = -(30)^{\frac{1}{2}},$$
$$V_{25} = -2\sqrt{3} \qquad V_{36} = V_{63} = -\sqrt{2}.$$

All other matrix elements are zero. Thus Eq. (2.30) applied to this problem is

$$\frac{44}{E(E+1)} - \frac{444}{E^2(E+1)^2} + \frac{720}{E^3(E+1)^3} = 1,$$

or

$$E^6 + 3E^5 - 41E^4 - 87E^3 + 400E^2 + 444E - 720 = 0.$$

Putting $E = \epsilon - 3/6$, this equation is transformed as

$$\epsilon^6 - 44.75\epsilon^4 + 466.1875\epsilon^2 - 833.76562 = 0,$$

or further, putting $\epsilon^2 = \eta + 44.75/3$,

$$\eta^3 - 201.3333\eta - 517.9264 = 0.$$

This equation is transformed into the dispersion relation

$$\frac{-2.572482}{\eta} + \frac{1.286241}{\eta - 14.1892} + \frac{1.286241}{\eta + 14.1892} = 1.$$

This equation can be solved by the weak coupling formula (3.11). Finally we obtain $E = 5.0$, 3.0, 1.0, -2.0, -4.0 and -6.0. Thus the energy levels split as shown in Fig. 2.

TABLE IX. Six states appearing in Example 4 in Sec. V.

state no.	N	l	Λ	Σ
1	5	5	0	+
2	5	3	0	+
3	5	1	0	+
4	5	5	1	−
5	5	3	1	−
6	5	1	1	−

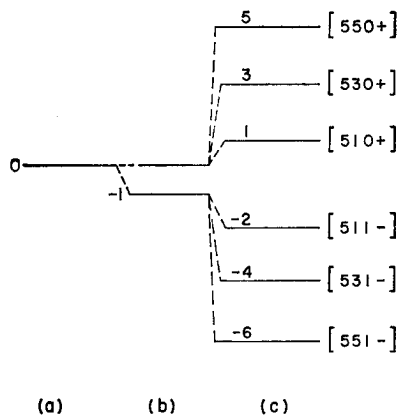

FIG. 2. The spin-orbit splitting of the single-particle states with $N = 5$, $\Omega = \frac{1}{2}$ in a harmonic-oscillator potential. (a) without the spin-orbit force; (b) the partial removing of degeneracy due to the diagonal matrix elements of the spin-orbit correction; (c) the final result.

ACKNOWLEDGMENT

The author would like to express his gratitude to Professor S. A. Moszkowski for hospitality extended to him during the course of his stay at UCLA, and for the interest and many suggestions shown in this work.

APPENDIX. VARIOUS PERTURBATION FORMULAS

As an appendix, we shall write the various perturbation formulas for comparison.[11]

(I) Rayleigh–Schrödinger Formula

$$E = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{|V_{np}|^2}{\epsilon_n - \epsilon_p} + \sum_{pq\neq n} \frac{V_{np}V_{pq}V_{qn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)} - V_{nn} \sum_{p\neq n} \frac{|V_{np}|^2}{(\epsilon_n - \epsilon_p)^2} + \sum_{pqr\neq n} \frac{V_{np}V_{pq}V_{qr}V_{rn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)(\epsilon_n - \epsilon_r)} - V_{nn} \sum_{pq\neq n} \left\{ \frac{V_{np}V_{pq}V_{qn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)^2} + \frac{V_{np}V_{pq}V_{qn}}{(\epsilon_n - \epsilon_p)^2(\epsilon_n - \epsilon_q)} \right\} + |V_{nn}|^2 \sum_{p\neq n} \frac{|V_{np}|^2}{(\epsilon_n - \epsilon_p)^3} - \sum_{p\neq n} \frac{|V_{np}|^2}{(\epsilon_n - \epsilon_p)} \sum_{q\neq n} \frac{|V_{nq}|^2}{(\epsilon_n - \epsilon_q)^2}. \tag{A1}$$

[11] Since the primary purpose of this Appendix is to aid a comparison of practical iteration procedures, in (II) and (III) we give only the iteration formulas found in reference 8. Their respective original formulas, which have the extremum property, are obtained simply by replacing $E^{(a)}$, $E^{(a-2)}$, $E^{(a-3)}$, etc., by E. The similar note is also applied to (IV), the present method.

(II) Brillouin–Wigner Formula

$$E^{(a)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{E^{(a-2)} - \epsilon_p} + \sum_{\substack{p\neq n \\ q\neq n}} \frac{V_{np}V_{pq}V_{qn}}{(E^{(a-3)} - \epsilon_p)(E^{(a-3)} - \epsilon_q)} + \sum_{pq\cdots\neq n} \frac{V_{np}V_{pq}V_{qr}V_{rs}\cdots V_{zn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)(\epsilon_n - \epsilon_r)\cdots(\epsilon_n - \epsilon_z)}. \tag{A2}$$

First iteration:

$$E^{(1)} = \epsilon_n + V_{nn}.$$

Second iteration:

$$E^{(2)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{\epsilon_n - \epsilon_p}.$$

Third iteration:

$$E^{(3)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{\epsilon_n + V_{nn} - \epsilon_p} + \sum_{pq\neq n} \frac{V_{np}V_{pq}V_{qn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)}.$$

Fourth iteration:

$$E^{(4)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{\epsilon_n + V_{nn} + \sum_{q\neq n} \frac{V_{nq}V_{qn}}{\epsilon_n - \epsilon_q} - \epsilon_p} + \sum_{pq\neq n} \frac{V_{np}V_{pq}V_{qn}}{(\epsilon_n + V_{nn} - \epsilon_p)(\epsilon_n + V_{nn} - \epsilon_q)} + \sum_{pqr\neq n} \frac{V_{np}V_{pq}V_{qr}V_{rn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)(\epsilon_n - \epsilon_r)}.$$

(III) Feenberg–Feshbach formula

$$E^{(a)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{E^{(a-2)} - \epsilon_{np}^{(a-2)}} + \sum_{\substack{p\neq n \\ q\neq np}} \frac{V_{np}V_{pq}V_{qn}}{(E^{(a-3)} - \epsilon_{np}^{(a-3)})(E^{(a-3)} - \epsilon_{npq}^{(a-3)})} + \cdots + \sum_{\substack{p\neq n \\ q\neq np \\ r\neq npq}} \frac{V_{np}V_{pq}V_{qr}V_{rs}\cdots V_{zn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)(\epsilon_n - \epsilon_r)\cdots(\epsilon_n - \epsilon_z)}. \tag{A3}$$

Here

$$\epsilon_{np}^{(a)} = \epsilon_p + V_{pp} + \sum_{q\neq np} \frac{V_{pq}V_{qp}}{E^{(a-2)} - \epsilon_{npq}^{(a-2)}} + \sum_{\substack{q\neq np \\ r\neq npq}} \frac{V_{pq}V_{qr}V_{rp}}{(E^{(a-3)} - \epsilon_{npq}^{(a-3)})(E^{(a-3)} - \epsilon_{npqr}^{(a-3)})} + \cdots$$

$$\epsilon_{npq}^{(a)} = \epsilon_q + V_{qq} + \sum_{r\neq npq} \frac{V_{qr}V_{rq}}{E^{(a-2)} - \epsilon_{npqr}^{(a-2)}} + \sum_{\substack{r\neq npq \\ s\neq npqr}} \frac{V_{qr}V_{rs}V_{sq}}{(E^{(a-3)} - \epsilon_{npqr}^{(a-3)})(E^{(a-3)} - \epsilon_{npqrs}^{(a-3)})} + \cdots.$$

First iteration:

$$E^{(1)} = \epsilon_n + V_{nn}.$$

Second iteration:

$$E^{(2)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{\epsilon_n - \epsilon_p}.$$

Third iteration:

$$E^{(3)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{(\epsilon_n + V_{nn}) - (\epsilon_p + V_{pp})} + \sum_{\substack{p\neq n \\ q\neq np}} \frac{V_{np}V_{pq}V_{qn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)}.$$

Fourth iteration:

$$E^{(4)} = \epsilon_n + V_{nn} + \sum_{p\neq n} \frac{V_{np}V_{pn}}{\left(\epsilon_n + V_{nn} + \sum_{q\neq n} \frac{V_{nq}V_{qn}}{\epsilon_n - \epsilon_q}\right) - \left(\epsilon_p + V_{pp} + \sum_{q\neq p} \frac{V_{pq}V_{qp}}{\epsilon_n - \epsilon_q}\right)} + \sum_{\substack{p\neq n \\ q\neq np}} \frac{V_{np}V_{pq}V_{qn}}{\{(\epsilon_n + V_{nn}) - (\epsilon_p + V_{pp})\}\{(\epsilon_n + V_{nn}) - (\epsilon_q + V_{qq})\}} + \sum_{\substack{p\neq n \\ q\neq np \\ r\neq npq}} \frac{V_{np}V_{pq}V_{qr}V_{rn}}{(\epsilon_n - \epsilon_p)(\epsilon_n - \epsilon_q)(\epsilon_n - \epsilon_r)}.$$

These three perturbation formulas agree up to the second order.

(IV) Present Method

(IVa) weak coupling case

$$E^{(n+1)} = \omega_i + \frac{C_i}{1 - \sum_{j\neq i} \frac{C_j}{E^{(n)} - \omega_j}}. \tag{3.11}$$

C_i's are given by

$$C_i = \sum_{j\neq i} \frac{|V_{ij}|^2}{\omega_i - \omega_j} + \sum_{j\neq i} \sum_{\substack{k>j \\ \neq i}} \frac{S_{jk}(V_{ik}V_{kj}V_{ji})}{(\omega_i - \omega_j)(\omega_i - \omega_k)}$$

$$+ \sum_{j\neq i} \sum_{\substack{k>j \\ \neq i}} \sum_{\substack{l>k \\ \neq i}} \frac{S_{jkl}\{S_{kl}(V_{il}V_{lk}V_{kj})V_{ji} - |V_{kl}|^2 |V_{ij}|^2\}}{(\omega_i - \omega_j)(\omega_i - \omega_k)(\omega_i - \omega_l)}$$

$$+ \cdots, \qquad (2.34)$$

where S_{jk}, S_{jkl}, etc. mean the cyclic sum over the suffixes indicated [as defined by Eq. (2.12)].

Starting value:

$$E^{(0)} = \omega_i = \epsilon_i + V_{ii}.$$

First iteration:

$$E^{(1)} = \omega_i + \frac{C_i}{1 - \sum_{j\neq i} \frac{C_j}{\omega_i - \omega_j}}.$$

Second iteration:

$$E^{(2)} = \omega_i + \frac{C_i}{1 - \sum_{j\neq i} \frac{C_j}{\omega_i + \frac{C_i}{1 - \sum_{k\neq i} \frac{C_k}{\omega_i - \omega_k}} - \omega_j}}.$$

(IVb) Strong Coupling Case

$$E^{(n+1)} = E_- + \frac{(E^{(n)} - \omega_i)(E^{(n)} - \omega_{i+1})}{(E^{(n)} - E_+)} \times \left(\sum_{j\neq i,i+1} \frac{C_j}{E^{(n)} - \omega_{i+1}} \right), \qquad (3.19)$$

with

$$E^{(1)} = E_-.$$

C_i's are given by (2.34). E_- is

$$E_- = \tfrac{1}{2}[(\omega_i + \omega_{i+1} + C_i + C_{i+1}) \pm \{(\omega_i + \omega_{i+1} + C_i + C_{i+1})^2 - 4(\omega_i\omega_{i+1} + C_i\omega_{i+1} + C_{i+1}\omega_i)\}^{\frac{1}{2}}]. \qquad (3.18)$$

This is the formula applied for $C_i > 0$ and $C_{i+1} > 0$. The other cases are similar to this formula.

(IVc) Nearest-Neighbor-States Approximation

$$E^{(n+1)} = \omega_i + \frac{C_i^{(n+1)}}{1 - \sum_{s=0}^{n} \left(\frac{C_{i+s+1}^{(n-s)}}{E^{(n-s)} - \omega_{i+s+1}} + \frac{C_{i-(s+1)}^{(n-s)}}{E^{(n-s)} - \omega_{i-(s+1)}} \right)}. \qquad (3.24)$$

Starting value:

$$E^{(0)} = \omega_i = \epsilon_i + V_{ii}.$$

First iteration:

$$E^{(1)} = \omega_i + \frac{C_i^{(1)}}{1 - \left(\frac{C_{i+1}^{(0)}}{\omega_i - \omega_{i+1}} + \frac{C_{i-1}^{(0)}}{\omega_i - \omega_{i-1}} \right)}, \qquad (3.22)$$

where

$$C_i^{(1)} = \frac{|V_{ii+1}|^2}{\omega_i - \omega_{i+1}} + \frac{|V_{ii-1}|^2}{\omega_i - \omega_{i-1}} \quad \text{[see (4.6)]},$$

$$C_{i\mp1}^{(0)} = \frac{|V_{i\mp1,i}|^2}{\omega_{i\mp1} - \omega_i} \quad \text{[see (4.7)]}.$$

Second iteration:

$$E^{(2)} = \omega_i + \frac{C_i^{(2)}}{1 - \left(\frac{C_{i+1}^{(1)}}{\omega_i - \omega_{i+1}} + \frac{C_{i-1}^{(1)}}{\omega_i - \omega_{i-1}} + \frac{C_{i+2}^{(0)}}{\omega_i - \omega_{i+2}} + \frac{C_{i-2}^{(0)}}{\omega_i - \omega_{i-2}} \right)}, \qquad (3.23)$$

where

$$C_i^{(2)} = \frac{|V_{ii+1}|^2}{\omega_i - \omega_{i+1}} + \frac{|V_{ii-1}|^2}{\omega_i - \omega_{i-1}} - \frac{|V_{i-2i-1}|^2 |V_{ii+1}|^2}{(\omega_i - \omega_{i-2})(\omega_i - \omega_{i-1})(\omega_i - \omega_{i+1})} - \frac{|V_{ii-1}|^2 |V_{i+1i+2}|^2}{(\omega_i - \omega_{i-1})(\omega_i - \omega_{i+1})(\omega_i - \omega_{i+2})} \quad \text{[see (4.9)]},$$

$$C_{i\mp1}^{(1)} = \frac{|V_{i\mp1i\mp2}|^2}{\omega_{i\mp1} - \omega_{i\mp2}} + \frac{|V_{i\mp1i}|^2}{\omega_{i\mp1} - \omega_i},$$

$$C_{i\mp2}^{(0)} = \frac{|V_{i\mp2i\mp1}|^2}{\omega_{i\mp2} - \omega_{i\mp1}}.$$

Note added: The general relation (2.30), i.e., the dispersion formula (1.4), is also derived from the expansion (1.2), directly. Equation (1.2) assumes the form

$$\prod_{i=1}^{n} (E - \omega_i) - \sum_i \sum_{j>i} V_{ij}V_{ji} \prod_{a\neq ij}'' (E - \omega_a) - \sum_i \sum_{j>i} \sum_{k>j} (V_{ij}V_{jk}V_{ki} + V_{ik}V_{kj}V_{ji}) \times \prod_{a\neq ijk}''' (E - \omega_a) - \cdots = 0.$$

Here the coefficient of $\prod_{a\neq ijk\ldots}^{(n)} (E - \omega_a)$ takes

on the form

$$\begin{vmatrix} 0 & -V_{ij} & -V_{ik} & -V_{il} & \cdots \\ -V_{ji} & 0 & -V_{jk} & -V_{jl} & \cdots \\ -V_{ki} & -V_{kj} & 0 & -V_{kl} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{vmatrix}.$$

The author thanks Dr. K. Maki for this comment. However, the author would keep the original Green's function method in the text for the following reasons. Firstly, the Green's function method in the present form itself seems interesting in connection with previous perturbation methods. In the previous perturbation methods, the function $(a|\ 1/A\ |a)$ is extracted out only at the step described in Eq. (2.13), and not at Eq. (2.15). (The seeming complexity of the method developed in Sec. IIb comes from this two-step extraction of $(a|\ 1/A\ |a)$ and from the ordering of states.) Secondly, by the Green's function method one can calculate $(b|\ 1/A\ |a)$, which is necessary for calculation of the transformation coefficients, as shown in the text. Thirdly, it seems that the use of the functions $(a|\ 1/A\ |a)$ and $(b|\ 1/A\ |a)$ in the present form seems advantageous also in the time-dependent perturbation theory now under investigation.

X-Ray Line Broadening in Isotactic Polystyrene

D. R. BUCHANAN AND ROBERT L. MILLER
Cemstrand Research Center, Inc., Durham, North Carolina

Wide-angle x-ray line-broadening data from several samples of isotactic polystyrene with different histories of pretreatment have been obtained by diffractometer techniques. These data have been analyzed according to several theories of the separation of the line-broadening effects due to small crystallite size and to lattice distortion; the results obtained from several integral breadth methods and from full line-profile analysis via Fourier transformation are compared. In each instance, annealing produces larger crystallite sizes and smaller degrees of lattice distortion. However, the actual numerical results obtained vary with the method of separation. Some of these variations arise from questionable assumptions in particular theories of size-distortion separation and others occur because different physical quantities are determined. It is possible, for example, to obtain information concerning number-average *and* weight-average crystallite sizes. For one isotactic polystyrene sample, good correlation of crystallite size from wide-angle diffraction, long period from low-angle diffraction, and x-ray crystallinity is demonstrated. Experimental problems encountered in the collection of line-profile data for polymer systems are also discussed, with particular emphasis on methods of correcting data for the effects of instrumental broadening. It is shown that a simple method based upon a Gaussian approximation to the profile shapes is often adequate to carry out the instrumental broadening correction.

INTRODUCTION

THE wide-angle x-ray diffraction pattern obtained from most polymer systems consists of diffraction maxima that are quite broad compared to those obtained from more perfectly crystalline materials. This same effect is observed for a cold-worked metal; as a result most of the procedures for analyzing diffraction broadening have come from work in this field where it has repeatedly been demonstrated that either small crystallite size or distortions of the crystallographic lattice (or both) lead to a broadening of the diffraction profiles.[1] Since both the distribution and magnitude of crystallite sizes and the degree of lattice distortion affect the mechanical properties of a material, it can be quite useful to be able to determine these parameters. Polymer systems have not generally been quantitatively investigated by x-ray line-broadening techniques, due principally to the somewhat difficult experimental problems involved. Diffraction patterns from polymers usually contain few orders of any one reflection; interpretation of such a diffraction pattern is obstructed by the presence of a relatively high, structured background and the unfortunate circumstance that, due to generally low crystallographic symmetry and broad diffraction maxima, most of the diffraction maxima are partially overlapped by at least one (and often more) of their neighbors. Previous studies (on polyethylene, for example[2–4]) have each concentrated on a single method of analysis of line broadening without establishing the validity of the results.

This investigation was undertaken with the object of determining the feasibility of applying metallurgical techniques of analyzing x-ray line broadening to polymer systems. In particular we wished to determine the nature and the severity of modifications to existing theories of line broadening which would allow them to be used for polymer systems. We chose a system, which, from the point of view of the diffraction phenomenon, is probably as similar to a metal system as possible: isotactic polystyrene. It crystallizes in a hexagonal lattice[5] and may exhibit many, quite sharp, diffraction maxima (including up to six orders of 110). For favorable conditions of pretreatment, many of the diffraction maxima are reasonably well-resolved, a condition aided by the hexagonal symmetry of the unit cell. This symmetry should also make possible the use of $hk0$ reflections as "pseudo-orders of reflection" of 110 without the introduction of an exorbitant error into the analysis.

Extraction of size and/or distortion information from experimental x-ray line-profile data occurs in three steps: (a) correction of the experimental data for geometrical factors and for background scattering, (b) correction of these data for instrumental broadening, and (c) separation of the final curve into size and distortion components. The first step is straightforward, although frequently tedious, and is considered part of the experimental procedure. The latter two steps each involve assumptions that are not always recognized. The problem of obtaining x-ray data adequately corrected for instrumental broadening is considered first. This is followed by a discussion of different methods that have been proposed to effect the size-distortion separation. The different methods are then applied to isotactic polystyrene samples differing in their thermal and/or mechanical histories and the results compared.

EXPERIMENTAL

Sample Description

Approximately 0.5-mm-thick films of an isotactic polystyrene were molded in a heated press at 240°C,

[1] B. E. Warren, Progr. Metal Phys. 8, 147 (1959).
[2] K. Katayama, J. Phys. Soc. Japan 16, 462 (1961).
[3] S. Kuribayashi, K. Tanaka, and A. Nagai, Seni-i Gakkaishi 19, 386 (1963).
[4] H. G. Thielke and F. W. Billmeyer, Jr., J. Polymer Sci. A2, 2947 (1964).
[5] G. Natta, P. Corradini, and I. W. Bassi, Nuovo Cimento Suppl. 15, 68 (1960).

then quenched rapidly in cold water. Three samples were prepared from strips of the films:

1. Sample 0-165 was the material as molded, annealed for 20 min (in 5-min increments) at 165°C. After 20 min (total) annealing, little further crystallization could be detected from x-ray flat-plate photographs.

2. Sample 4-165 was drawn approximately four times (300% elongation) at 105°C, quenched, and annealed at constant length for 5 min at 165°C.

3. Sample 4-120 was drawn approximately four times at 105°C, quenched, and annealed at constant length for 5 min at 120°C.

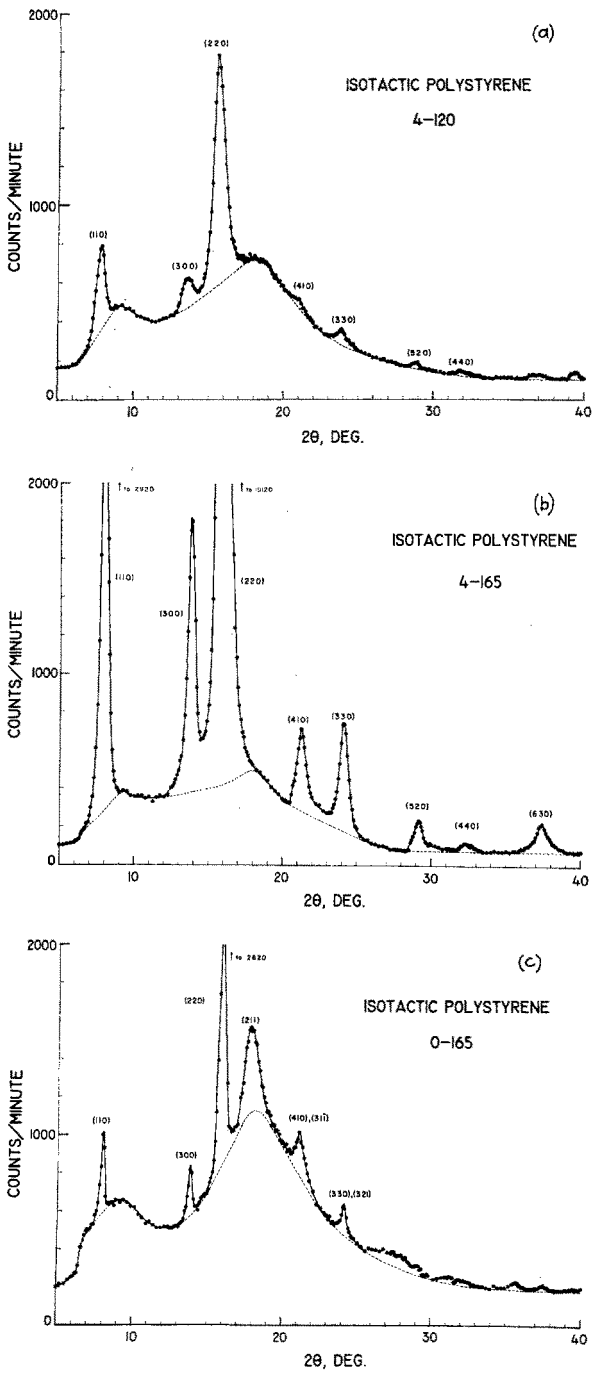

FIG. 1. Uncorrected data for isotactic polystyrene samples. Dashed line represents "amorphous" background. (a) 4-120, (b) 4-165, (c) 0-165.

Instrumental Standard

The instrumental standard (i.e., a sample for which it is assumed that no line broadening occurs other than that due to the instrument) in cold-worked metal studies is commonly taken as the cold-worked sample which had been thoroughly annealed. In polymeric systems, this procedure is obviously not applicable. The ideal instrumental standard in this case is composed of large ($\sim 1\,\mu$), perfect crystals and absorbs the x-ray beam to approximately the same extent as does the polymer sample. Hexamethylenetetramine, $(CH_2)_6N_4$, satisfies these requirements better than other substances we have investigated. We used a disk of this standard $\sim$0.36 mm thick compacted at approximately 85°C.

X-Ray Measurements

All of the diffraction profiles were measured with Cu$K\alpha$ radiation generated at 35 kV and 20 mA, a scintillation counter and pulse-height analyzer, and point-count techniques. Intensities were counted at 0.1° intervals in Bragg angle (2θ) in the peak and, in some cases, at 0.25° intervals in the vicinity of the tails. The counting period was generally 10 000 counts or 30 min, whichever occurred first. The $hk0$ profiles were obtained in the reflection mode; a curved-crystal monochromator in the diffracted beam passed predominantly $K\alpha_1$ radiation. A few hkl profiles from sample 4-165 were obtained in the transmission mode (a Ni filter was used for monochromatization). In the transmission mode, the plane of the sample was maintained at the angle $(90°-\theta)$ with respect to the incident and diffracted beams.

Uncorrected data for the three samples studied are shown in Fig. 1. Line breadths are noticeably different for the three samples; those from sample 4-120 are broadest whereas those from sample 0-165 are clearly the narrowest. Figure 1 also illustrates the major experimental difficulties encountered in determining line broadening in polymers. Most of the profiles are incompletely resolved from neighboring profiles. Those at the lower diffraction angles are superimposed upon a structured, "amorphous" background whose exact shape is difficult to determine.

Low-angle x-ray measurements were made with a Kratky camera set for photographic techniques. Two photographs were taken from each of the drawn samples: one with the machine direction parallel, and one with the machine direction perpendicular to the length of the slit-shaped incident beam. The Cu x-ray tube was operated at 30 kV and 20 mA; exposure times, with a 100-μ entrance slit, were 2–3 h. The incident

beam was monochromatized by a Ni filter. Although microdensitometer traces of the resulting photographs were approximately corrected for background scattering, no other corrections were applied.

Reduction of Data

The data were corrected for background and for polarization and absorption by means of a digital computer (ASI 210A). The background correction for curves obtained in reflection was estimated from the scattering curve from a quenched, isotactic polystyrene sample, obtained under the same instrumental conditions. For each sample, a certain amount of guesswork was necessary in order to draw a reasonable background curve. The background correction for the profiles obtained in transmission was determined more directly; ratemeter scans from 5° to 55° 2θ were made with the draw direction in the sample tilted, with respect to the diffraction plane, about the normal to the sample surface. The background curve constructed from these curves consisted of the line drawn through the set of intensity minima.

In many cases, it was necessary to resolve a corrected diffraction profile from its overlapping neighbors. This was done graphically by trial and error. Each profile was assumed to be symmetrical about its maximum, an assumption which is completely valid if the deconvolution process has been applied correctly, and approximately true even for our data before deconvolution. An iterative process was used to construct resolved maxima whose intensities at any given point in 2θ space could be summed to give the correct intensity. While the accuracy of this procedure may be doubtful, the precision with which one can carry out this process seems quite good. We applied this method of resolution both before and after correction for instrumental broadening by Fourier deconvolution. Integral breadths computed in the two cases generally agreed within 3%–5%.

Experimental Problems

The successful distinction between crystallite size broadening and lattice distortion broadening depends to a large extent on the accuracy of the experimental data. With the assumption that the usual experimental requirements are met (e.g., monochromatization of the x-radiation, use of point-counting methods), the problems of uncertain background and of unresolved peaks remain. It is usually not sufficient to assume that the background scattering is given simply by the scattering from a quenched sample, although this information is of some use in constructing a background curve. If the diffraction profiles are obtained in the transmission mode, a background curve may be constructed from data taken with the unique axis of the sample at different angles to the diffraction plane. Unfortunately, the overlapping diffraction maxima are best resolved if the experiment is conducted in the reflection mode, thus utilizing the focusing geometry of the diffractometer. In this case, the background curve must be constructed by trial-and-error methods based, perhaps, on the scattering curve of a quenched sample.

Resolution of partially overlapping peaks may be accomplished with varying degrees of accuracy dependent upon the effort expended. Graphical resolution by trial-and-error methods can yield quite precise results, especially if the background level has been well-defined, but a large number of iterations is required to achieve the desired accuracy. Analytical schemes of resolution are possible if the functionality of the individual profiles is assumed.

The matching of both the thicknesses and the linear absorption coefficients of the polymer samples and the instrumental standard is an experimental problem of considerable difficulty. Our instrumental standard, the thinnest ($t\sim0.36$ mm) HMT disk that could support its own weight in a sample holder, was only slightly too thick to match the undrawn polystyrene sample ($t\sim0.5$ mm) in total absorption. However, the absorption difference was much greater with respect to the drawn polystyrene samples ($t\sim0.25$ mm). Thus it is possible that some broadening has been neglected for these samples, or, alternatively, that the use of a too thick standard has overcorrected for the instrumental broadening contribution. We have, however, assumed that this error is negligible for two reasons: Both the HMT standard and all of the polystyrene samples were thin enough to exhibit apparently symmetrical diffraction profiles; the profiles from slightly thicker HMT disks (up to $t\sim0.5$ mm) showed an integral breadth increase of only a few percent, indicating that, for these thicknesses, the contribution of the experimental geometry and slit systems to the instrumental broadening is much greater than that of the thickness of the standard.

Correction for Instrumental Broadening

The line profile obtained after correction for geometrical factors and for background scattering is a broadened profile produced by the convolution product of the true line profile and an instrumental broadening profile. Extraction of the true profile from the data corresponds to a deconvolution (or unfolding) operation which is commonly carried out by means of the Fourier transform method of Stokes.[6] Both the corrected profile and its Fourier transform are thereby obtained.

In many of the line-broadening-analysis methods which have been proposed, only the integral breadth of the true diffraction profile is required. This may be approximated without the use of the Fourier deconvolution process if an assumption is made regarding the shapes of the observed, instrumental, and corrected

[6] A. R. Stokes, Proc. Phys. Soc. (London) **A61**, 382 (1948).

TABLE I. Observed, instrumental, and corrected integral breadths.

Sample	$hk0$	B(°2θ)	b(°2θ)	β(Å⁻¹)[a]	β(Å⁻¹)[b]	β(Å⁻¹)[c]	β(Å⁻¹)[d]
4-120	110	0.751	0.235	0.0079	0.0076	0.0080	0.0090
	300	0.946	0.235	0.0100	0.0091	0.0103	0.0124
	220	1.028	0.235	0.0111	0.0109	0.0112	0.0138
	401	0.806	0.227	0.0081	0.0078	0.0086	0.0101
	330	0.773	0.227	0.0078	0.0075	0.0082	0.0095
4-165	110	0.624	0.235	0.0061	0.0059	0.0065	0.0044
	300	0.726	0.235	0.0080	0.0078	0.0077	0.0055
	220	0.756	0.235	0.0070	0.0069	0.0080	0.0058
	410	0.804	0.227	0.0087	0.0085	0.0085	0.0064
	330	0.853	0.227	0.0084	0.0082	0.0091	0.0069
0-165	110	0.374	0.235	0.0035	···	0.0033	0.0024
	300	0.435	0.235	0.0053	0.0051	0.0041	0.0035
	220	0.436	0.235	0.0044	0.0043	0.0041	0.0035
	330	0.863	0.227	0.0055	0.0052	0.0057	0.0063

[a] Determined directly from corrected curve.
[b] Equation (3).
[c] Equation (1).
[d] Equation (2).

profiles. Possible assumptions are discussed in detail by Klug and Alexander.[7] For this work, we investigated the assumptions that all three profiles were Gaussian,[8] in which case

$$\beta^2 = B^2 - b^2, \tag{1}$$

where β, B, and b are integral breadths of the corrected, observed (but corrected for geometrical factors), and instrumental profiles, respectively, and that all three profiles were Cauchy (after Scherrer, cf. Ref. 7),

$$\beta = B - b. \tag{2}$$

If the Stokes deconvolution operation is performed, the integral breadth of the true diffraction profile (β) may be obtained from the Fourier coefficients (A_n) of the true profile,

$$\beta = 1/\sum_{-\infty}^{\infty} A_n, \tag{3}$$

or may be determined directly from the corrected profile itself. Table I contains integral breadths for the more accurately determined profiles from the three polystyrene samples (column 3) and for the hexamethylenetetramine standard (column 4). The remaining columns (5–8) contain corrected integral breadths for the polystyrene profiles obtained in the manner indicated.

If we assume that the deconvolution process yields the most reliable values of β, we find (Table I) that the Gaussian approximation [Eq. (1)] yields values of β which differ, on the average, by 10% or less from those computed directly via the deconvolution process. Consequently, for those cases in which the utmost accuracy is not required, Eq. (1) may serve as a useful, rapid means to obtain approximate integral breadths. On the other hand, the results of the Cauchy assumption [Eq. (2)] clearly are inconsistent with the other results and are not considered further.

SEPARATION OF SIZE AND DISTORTION BROADENING

Fourier Transform Methods

The observed diffraction profile (after correction for instrumental broadening) is assumed to be given by the convolution product of a profile due entirely to crystallite size and of a profile due only to lattice distortion. The Fourier transforms of these profiles are related by

$$A(t,s) = A^S(t) \cdot A^D(t,s), \tag{4}$$

where $A(t,s)$, $A^S(t)$, and $A^D(t,s)$ are the Fourier transforms of the corrected profile, the crystallite size profile, and the distortion profile, respectively. The variable t is a length along the normal to (hkl) and s $[=(2/\lambda)\ \sin\theta]$ is the diffraction space variable. Several authors[9–11] have discussed methods of extracting information about the "number-average" crystallite size and the mean-square lattice distortion, as a function of t, from Fourier transforms of a set of line profiles. These authors have shown that, for small distortions and for planes of low index (small s), the distortion transform becomes

$$A^D(t,s) = \exp[-2\pi^2 s^2 t^2 \langle \epsilon_t^2 \rangle], \tag{5}$$

where $\epsilon_t (= \delta d/d)$ is the fractional difference between the length of a column of unit cells parallel to s and the length it would have in the undistorted lattice. The angle brackets $\langle\ \rangle$ indicate the average over-all unit cells of length t.

From Eqs. (4) and (5) it follows that

$$\ln A(t,s) = \ln A^S(t) - 2\pi^2 s^2 t^2 \langle \epsilon_t^2 \rangle. \tag{6}$$

When several orders of reflection (i.e., several values of s) are available, one may plot $\ln A(t,s)$ vs s^2 with t as a parameter. In general, the resulting curves will not be linear but will approach linearity for small s. The intercept for each t curve gives $\ln A^S(t)$ and the initial slope gives $\langle \epsilon_t^2 \rangle$. At least three orders of reflection are necessary to make an accurate extrapolation to $s=0$. The "number-average" crystallite size, $\bar{N}$, is obtained from the initial slope of a plot of $A^S(t)$ vs t. A frequently observed negative curvature near the origin of these curves (the "hook" effect) has been ascribed[12] to an effect of the finite summation of the Fourier series and is commonly caused by underestimation of the extent of the tails of the profile. To a first approximation, this error can be corrected if one considers the initial straight portion of the $A^S(t)$ vs t curve to represent the

[7] H. P. Klug and L. E. Alexander, *X-Ray Diffraction Procedures* (John Wiley & Sons, New York, 1954), Chap. 9.
[8] B. E. Warren, J. Appl. Phys. **12**, 375 (1941).
[9] B. E. Warren and B. L. Averbach, J. Appl. Phys. **21**, 595 (1950).
[10] B. E. Warren, Acta Cryst. **8**, 483 (1955).
[11] E. F. Bertaut, Compt. Rend. **228**, 492 (1949).
[12] E. F. Bertaut, Acta Cryst. **5**, 117 (1952).

initial slope, extends this line to $t=0$, and renormalizes the entire curve.

Integral Breadth Methods

To avoid the computational labor necessary in the Fourier transform method, various schemes to obtain size-distortion separation from integral breadths alone have been suggested. The integral breadth of the lattice distortion profile as given by Wilson[13] is

$$\beta_D = 4e(\sin\theta/\lambda) = 2e/d_{hkl}, \tag{7}$$

where $e(\propto \delta d_{hkl}/d_{hkl})$ is a measure of the maximum lattice distortion. The integral breadth of the crystallite size profile is taken as the Scherrer equation (Ref. 7):

$$\beta_S = K/L_{hkl}, \tag{8}$$

where K is a constant related both to crystallite shape and to the way in which β_S and L_{hkl} are defined. We take β_S(in Å^{-1}), L_{hkl} (in Å), and K to be an integral breadth, the dimension of the crystal normal to the (hkl) plane and unity, respectively.

Since size broadening [Eq. (8)] is independent of the order of reflection, while distortion broadening [Eq. (7)] is not, it is possible to distinguish between the two types if only one of them is operative (provided data for at least two orders of reflection are available).

[Equations (7) and (8) are most commonly applied in an alternative form:

$$\beta_D(\text{rad}) = \beta_D(\text{Å}^{-1})\lambda/\cos\theta = 4e\tan\theta, \tag{7a}$$

and

$$\beta_S(\text{rad}) = \beta_S(\text{Å}^{-1})\lambda/\cos\theta = K\lambda/L_{hkl}\cos\theta. \tag{8a}$$

Equations (7a) and (8a) indicate that, *for a constant angle* θ, size broadening depends on the radiation wavelength used, whereas distortion broadening does not. It is not possible, however, to generate extra data from measurements of one reflection at several wavelengths. Equations (7) and (8) indicate that for the same reflection, hkl, neither β_D nor β_S is affected by a change in wavelength; therefore, measurements of one reflection at several x-ray wavelengths can never distinguish between pure size and pure distortion broadening.]

If it is assumed that both β_S and β_D arise from Cauchy profiles, the integral breadth (β) of the resulting convolution product is[14]

$$\beta = \beta_S + \beta_D. \tag{9}$$

If, on the other hand, β_S and β_D are assumed to result from Gaussian profiles[15]:

$$\beta^2 = \beta_S^2 + \beta_D^2. \tag{10}$$

It has been observed[16,17] that the cold-worked line profiles from pure copper, pure aluminum, and several aluminum–copper alloys are neither Gaussian nor Cauchy, but have an intermediate character. One possible explanation is that the size profile follows one law, while the distortion profile follows another. For the assumption that the size profile is Cauchy, while the distortion profile is Gaussian, Schoening[18] has published parameters suitable for a graphical solution for e and L_{hkl} from integral breadth data for the first two orders of reflection.

The lattice distortions represented by e are microstrains, i.e., random, local lattice distortions. Bonart, Hosemann, and McCullough[19] have discussed a method of separating size from distortion broadening in which paracrystalline lattice distortions are responsible for a portion of the observed line broadening. In this case, both the size and the distortion profiles are assumed to be Gaussian, but the integral breadth of the distortion profile shows a squared dependence on the order of reflection, instead of the first-power dependence shown for microstrain broadening. The total integral breadth is given by

$$\beta^2 = (1/L_{hkl})^2 + \{(1/2d_{hkl})[1-\exp(-2\pi^2 g^2 h_0^2)]\}^2, \tag{11}$$

where $h_0 = 1, 2, \cdots$ is the order of reflection, $g(=\Delta d_{hkl}/d_{hkl})$ is the standard deviation of the (Gaussian) distribution $H(x)$, which, in turn, is the probability function describing the paracrystalline lattice in the direction normal to the planes (hkl).

It will be noted that, for all of the integral breadth methods discussed, the Scherrer equation (8) is incorporated as the size-broadening component. It can be shown[20] that the use of the Scherrer equation (with integral breadths) for the determination of crystallite size results in a "weight-average" size, in contrast to a "number-average" size obtained from the Fourier transform method. Furthermore, this result is independent of the assumed shape of the diffraction profile.

RESULTS

Basal Plane

Fourier Transform Method

The results obtained by the Warren–Averbach Fourier transform method are shown in Figs. 2–4 and in the first row of Table II. For example, the extrapolation procedure [Eq. (6)] is shown in Fig. 2 for sample 4-165. The upward concavity of the curves implies that the strain distribution falls off more slowly than does a Gaussian.[21] This feature was observed for all three polystyrene samples. Figure 3 contains $A^S(t)$ data as a function of t for all three samples. In each case, re-

[13] A. J. C. Wilson, *X-Ray Optics* (Methuen and Co., London, 1949), p. 5.
[14] W. H. Hall, Proc. Phys. Soc. (London) **A62**, 741 (1949).
[15] H. Romeis and P. Royen, Z. Anorg. Allgem. Chem. **318**, 126 (1962).
[16] G. B. Mitra, Acta Cryst. **16**, 429 (1963).
[17] G. B. Mitra, Brit. J. Appl. Phys. **16**, 77 (1965).
[18] F. R. L. Schoening, Acta Cryst. **18**, 975 (1965).
[19] R. Bonart, R. Hosemann, and R. L. McCullough, Polymer **4**, 199 (1963).
[20] D. R. Buchanan, R. L. McCullough, and R. L. Miller, Acta Cryst. **20**, 922 (1966).
[21] B. E. Warren, Acta Met. **11**, 995 (1963).

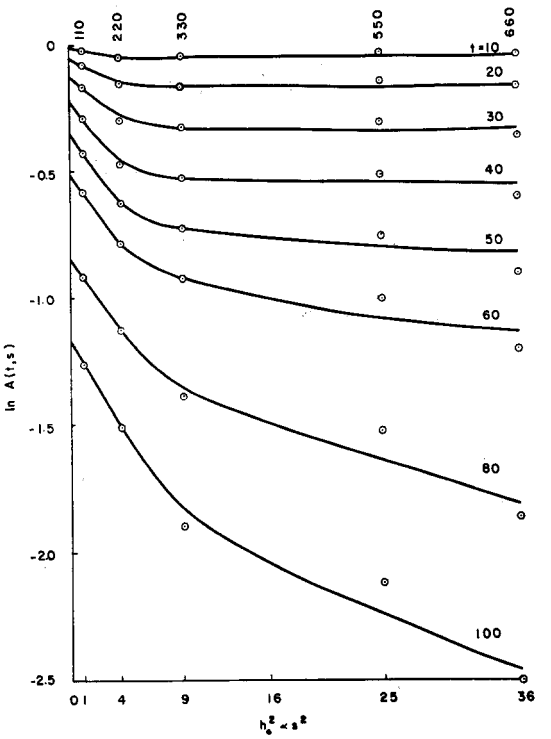

FIG. 2. Extrapolation curves for Warren–Averbach Fourier transform method [Eq. (6)]. Sample 4-165.

normalization of the $A^S(t)$ was necessary, indicating the difficulty encountered in a correct determination of profile tails. Figure 4 contains the variation of $\langle \epsilon_t^2 \rangle^{\frac{1}{2}}$ with t: For each sample, $\langle \epsilon_t^2 \rangle^{\frac{1}{2}}$ decreases rapidly with t and then levels off to a constant value. Warren[1] has suggested that this behavior arises from inhomogeneous strains and from the fact that, with increasing t, there is no longer a significant contribution from the smaller domains where the strains are probably larger.

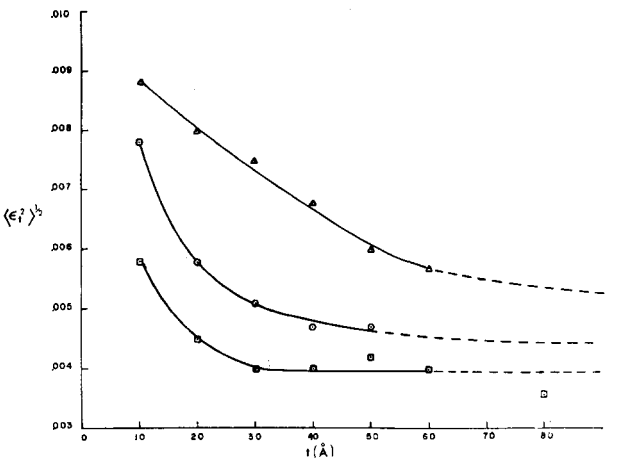

FIG. 4. $\langle \epsilon_t^2 \rangle^{\frac{1}{2}}$ vs t. △: 4-120, ⊙: 4-165, ⊡: 0-165.

Integral Breadth Methods

Because observing even two true orders of reflection is difficult in many polymer systems, we investigated the possibility of using data from other reflections (i.e., pseudo-orders) in the broadening analysis. Since the polystyrene unit cell has a hexagonal cross section it is reasonable to anticipate that crystallite sizes and lattice distortions might be nearly isotropic in this

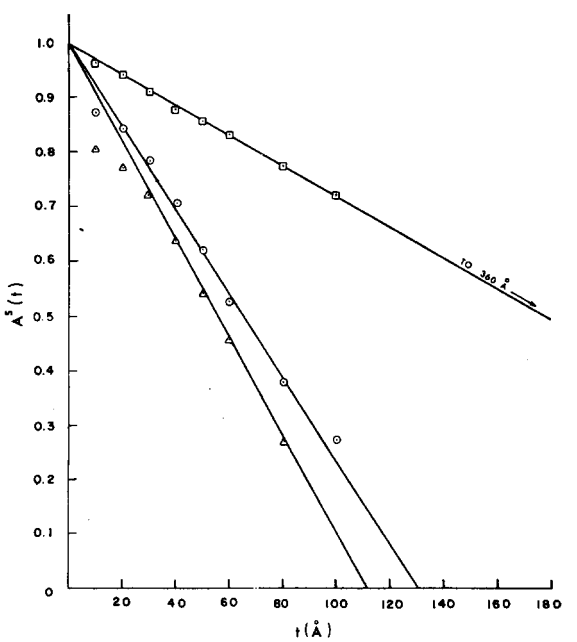

FIG. 3. $A^S(t)$ vs t. △: 4-120, ⊙: 4-165, ⊡: 0-165.

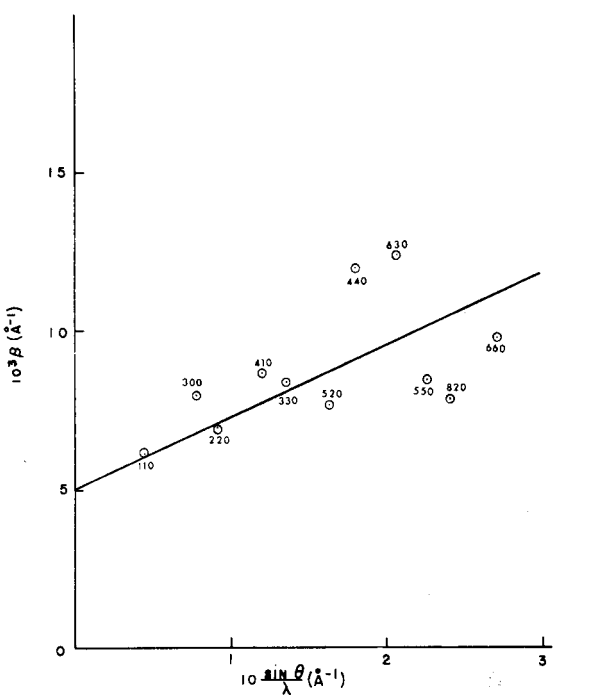

FIG. 5. Integral breadths (corrected for instrumental broadening by Fourier deconvolution) for sample 4-165 as a function of $\sin\theta/\lambda$ [Eq. (9)].

TABLE II. Crystallite size and lattice distortion parameters—basal plane.

Separation method	Instrumental correction	4-120 Size (Å)	4-120 Distortion (%)	4-165 Size (Å)	4-165 Distortion (%)	0-165 Size (Å)	0-165 Distortion (%)
1. Fourier transform[a]	F[f]	111	0.66	130	0.51	360	0.41
2. Cauchy[b]	F	217	1.78	208	0.63	400	0.54
	G[g]	213	1.78	192	0.72	400	0.45
3. Gaussian[c]	F	150	0.61	177	0.28	305	0.13
	G	150	0.61	160	0.31	338	0.11
4. Cauchy–Gaussian[d]	F	172	1.92	176	0.70	330	0.56
	G	162	1.95	176	1.02	355	0.55
5. Paracrystal[e]	F	131	4.95	162±6	2.80±0.30	275±19	2.21±0.60
	G	129	4.95	156	3.76	310	2.92

[a] Equation (6), number-average size; distortion shown is root-mean $\langle \epsilon_\ell^2 \rangle$ over the range $\ell = 10$–80 Å.
[b] Equation (9). [c] Equation (10).
[d] See Ref. 19; weight-average size; distortion shown is e, the "maximum" distortion.
[e] Equation (11); mean values from 110, 220, 330. Otherwise solution for 110, 220 only.
[f] Fourier deconvolution (see Ref. 6). [g] Equation (1).

plane. In Fig. 5 are shown the corrected integral breadths of usable *hk*0 reflections from sample 4-165 (Table I) as a function of $\sin\theta/\lambda$. The solid line represents the best fit to the 110, 220, and 330 datum points. The *hk*0 datum points appear to fit this line as well as do the *hh*0 points, especially for the higher-order reflections. Evidently, any anisotropy in the basal plane is masked by experimental inaccuracies.

Integral breadths for sample 4-165 are not all equally accurate. For this reason, the data of Fig. 5 were fit to Eq. (9) by a weighted least-squares technique based on the usual statistical methods for counting techniques and utilized, for each peak, the peak-to-background ratio and the number of counts at the peak. The background error was arbitrarily taken as 1% (probably an optimistic estimate). The results of the analysis of the *hk*0 reflections were: (a) from *hh*0 reflections only, $L = (181 \pm 28)$ Å and $e = (0.46 \pm 0.10)\%$, and (b) from *hh*0 and *hk*0 reflections, $L = (172 \pm 25)$ Å and $e = (0.47 \pm 0.10)\%$. Clearly, in this polymeric system the *hk*0 reflections may be used to augment the *hh*0 data without appreciably affecting the results of the analysis.

Weight-average crystallite sizes, L, and percent "maximum" lattice distortions, e, were obtained from the integral breadth data of Table I via Eqs. (9) and (10). Two types of instrument correction were used: Fourier deconvolution and Gaussian approximation [Eq. (1)].

The Cauchy analysis indicated in Fig. 5, when applied to the other two samples, indicated agreement of the *hk*0 and *hh*0 datum points equally as good as that obtained with sample 4-165. The resulting size and strain parameters are shown as the second entry in Table II. Application of the Gaussian approximation [Eq. (10)] to the integral breadth data is shown in Fig. 6 and as the third entry in Table II. The fourth and fifth entires in Table II correspond to the results obtained upon analysis of the integral breadth data of Table I by the Schoening and paracrystalline methods, respectively.

It is necessary to decide whether the integral breadth data are fit better by a microstrain or by a paracrystalline theory of distortion broadening. For the relatively small amounts of distortion deduced here, paracrystalline distortions should broaden a line according to the square of the order of reflection, and microstrain distortions should broaden a line according to the first power of the order of reflection, even in the presence of crystallite size broadening. The integral breadth data for sample 4-165 were plotted accordingly as shown in Fig. 7, where h_0 is the order of reflection. It would appear that the microstrain theory is fit slightly better by the data than is the paracrystalline theory, but inaccuracies of the data make a categorical distinction impossible. Note that a minimum of *three* true orders of reflection are required in order to distinguish between

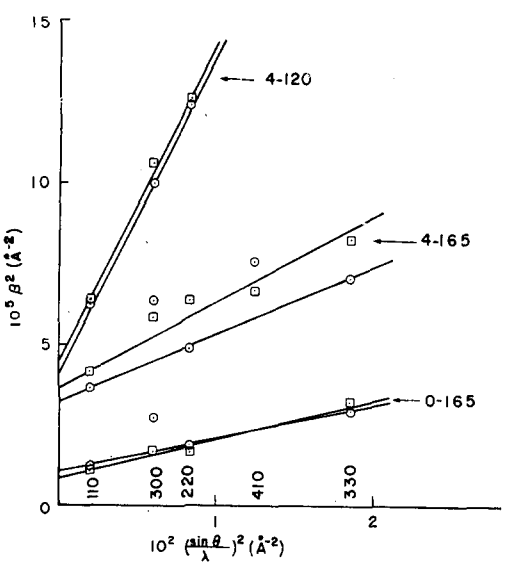

FIG. 6. Squared integral breadths as a function of $(\sin\theta/\lambda)^2$ [Eq. (10)]. ⊙: instrumental correction by Fourier deconvolution; ⊡: instrumental correction according to Eq. (1).

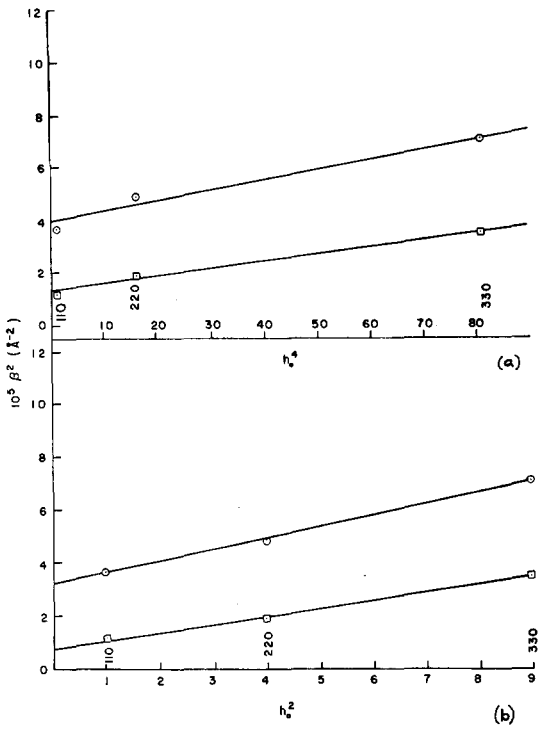

FIG. 7. Squared integral breadths (corrected for instrumental broadening by Fourier deconvolution) as a function of (a) h_0^4 and (b) h_0^2, where h_0 is the order of reflection. ⊙ : 4-165; ⊡ : 0-165.

the two theories. This is the same number of orders required for the full Fourier transform analysis.

Comparison of Results

Crystallite size and lattice distortion parameters obtained by each of the methods discussed are collected in Table II. For each of the separation methods, except the Cauchy approximation (second entry), the crystallite size increases from left to right, i.e., with annealing, and the degree of lattice distortion decreases. Another general feature is that the weight-average size derived from each of the integral breadth methods is larger than the corresponding number-average size derived from the Fourier transform technique, as must be expected. An inconsistency arises for sample 0-165 (unstretched, highly annealed) in which the number-average size is larger than most of the corresponding weight-average sizes. We attribute this anomaly to the fundamental difficulty of deconvoluting an observed peak with an instrumental peak of nearly the same breadth. Thus the number-average size for sample 0-165 and also the corresponding lattice-distortion value are probably in error. Fortunately, this particular situation is not likely to arise often in analyses of polymer systems.

The process required to apply the Fourier transform technique of Warren and Averbach yields, in principle, the most complete information from line profile data, even though the accuracy required of these data is higher than that required for integral breadth interpretations. With sufficiently accurate data, this process allows the determination of the number- and weight-average crystallite sizes and the distributions, as a function of distance in the sample, of lattice distortions. Unfortunately, it appears that the Warren–Averbach technique cannot be used for any but the most favorable polymer systems. Most importantly, to obtain accurate results requires that at least three true orders of reflection be observed in order to determine the curvature of the extrapolation curve [Eq. (6)]. Warren[21] has shown that a straight-line extrapolation from only two datum points (i.e., two orders of reflection) leads to consistently low values for both the "number-average" crystallite size and the mean square lattice distortion $\langle\epsilon^2_t\rangle$. Since only two orders of $hh0$ reflection were available for sample 4-120, the number-average crystallite size and the lattice distortions computed for this sample by the Warren–Averbach technique are somewhat low. Even from our best (and most complete) set of data, that from sample 4-165, we were not able to determine the size transform $A^S(t)$ well enough for any rational computation of the crystallite size distribution.

While integral breadth methods of separating size and distortion broadening are perhaps less susceptible to the errors introduced by overlapping peaks and uncertain background determinations, and, in favorable cases, can utilize pseudo-orders of reflection, the results are biased by the choice of the assumed functionality of the size and distortion profiles. The suggestion that both profiles be approximated by Cauchy functions (Table II, second entry) seems to be unsuited for very broad profiles; the use of this assumption results in a larger weight-average size for sample 4-120 than for sample 4-165, contrary to the general trend. On the other hand, the distortions computed on this basis for samples 4-165 and 0-165 are very nearly the mean values found using the Warren–Averbach technique. If both the size and the distortion profiles are assumed Gaussian (third entry), the weight-average crystallite sizes increase with annealing in the expected manner, but the distortion computed for the most highly annealed sample (0-165) is much lower than that obtained by any other method. It appears quite possible that if one intends to assume the same functionality for both the size and distortion profiles, a set of broad profiles ought to be assumed Gaussian, while a set of narrow profiles ought to be assumed Cauchy. In actual practice, the Gaussian assumption may be more useful due to the generally broad nature of the diffraction profiles from polymer systems.

A better approximation is to allow the size and distortion profiles to follow different functionalities. Solution of an equation of the type, $\beta^n=\beta_S{}^n+\beta_D{}^n$, is no longer possible, but Schoening[18] has shown that, through the use of Fourier transforms, one may use first- and second-order reflection data to solve graphically for the case in which the size profile is Cauchy and the distortion profile is Gaussian (Table II, fourth entry). This

method yields results generally in accord with those expected, with the exception that the distortion value for sample 4-120 seems somewhat high. Extending the method to make use of all of the available data would improve its utility.

The crystallite sizes and paracrystalline distortions computed from Eq. (10) are shown as the fifth entry of Table II. Again crystallite sizes increases and lattice distortions decrease with annealing. The crystallite sizes tend to be somewhat smaller, and the lattice distortion parameters somewhat larger, than those obtained from other integral breadth methods. Our observation is that this procedure is not very sensitive for values of the lattice distortion g below about 5%. That is, given an integral breadth for 110, one computes the same crystallite size for any value of g less than about 5%.

Molecular Axis

In sample 4-165, two nearly meridional reflections, 102 and 113, are accessible. No 00l data are available. These two reflections make angles of 9° and 11°, respectively, with the fiber axis (c) and we have taken them to be pseudo-orders. Separating size and distortion broadening by the Gaussian and paracrystalline approximations results in crystallite sizes of 89 and 77 Å and lattice distortions of 0.31% and 3.70%, respectively. The values of the lattice distortion are essentially identical to those found from the $hk0$ data (Table II): 0.31% and 3.76%, respectively. Although such a coincidence appears unlikely, it is equally difficult to rationalize isotropic lattice distortions in this sample. The crystallite size results indicate a "pill" shape in which the thickness is approximately one-half the diameter. A low-angle x-ray meridional maximum occurs at 145 Å in this sample; a regular alternation of crystalline and amorphous regions of approximately equal size is thereby indicated, in reasonable agreement with maximum polystyrene crystallinities detected by Challa, Hermans, and Weidinger.[22] Tsvankin[23] has constructed parametric curves based upon calculations of meridional small-angle scattering expected from a fiber consisting of alternating crystalline and amorphous regions. From his curves and the measured half-width of the 145-Å maximum, the ratio of crystalline to total length of the repeating segment is estimated to be 0.45. Thus, the wide-angle crystallite size, the low-angle maximum, the low-angle half-width, and crystallinity estimates are mutually consistent in predicting a structure containing reasonably large quantities of amorphous material (more than half).

DISCUSSION

Once experimental requirements have been satisfied, correction of data for instrumental broadening is fairly straightforward. If a digital computer is available, the Stokes[6] Fourier transform deconvolution procedure is preferable to more approximate methods, particularly since the Fourier transform of the corrected profile is obtained as part of the process. Without a computer this process is rather tedious. A reasonably good (within 10%) approximation to the instrumental correction is obtained under the assumption that the profiles are Gaussian. Of course, this latter procedure is restrictive in that only integral breadth analyses of size and distortion broadening are possible and only one measure of the crystallite size distribution (the "weight-average" size) is available.

When the data are available, the Fourier transform method of separating size from distortion broadening[9] gives the most complete information of any procedure considered here. Manipulation of the Fourier coefficients of the true profile as per Warren and Averbach yields a "number-average" crystallite size and an rms strain parameter. Integral breadth methods yield a "weight-average" crystallite size and a different strain parammeter; the magnitudes of the size and the strain parameters obtained depend somewhat upon which assumption regarding the functionality of the size and distortion profile is made. Our results for isotactic polystyrene suggest that a Gaussian approximation (or the Cauchy–Gaussian approximation of Schoening[18]) is most suitable for polymer systems. A Cauchy approximation leads to results which are quite inconsistent with the rest of our work. Our results also indicate that paracrystalline lattice distortions are not of significance in this system; considerably greater residual lattice distortion magnitudes would be required before the paracrystalline model would be a sensitive basis for consideration. In any case, data for at least three true orders of reflection are required to distinguish categorically between microstrain and paracrystalline lattice distortions.

In isotactic polystyrene we find that the observed diffraction line broadening is caused by both small crystals and distorted lattices. Furthermore, annealing produces a simultaneous increase in crystallite size and decrease in lattice distortions, irrespective of the method employed for size-distortion separation. Agreement between crystallite sizes determined by the different acceptable integral breadth measurements is reasonable. Comparison between Fourier transform and integral breadth analyses indicated a moderately broad crystallite size distribution with weight-to-number-average size ratios of the order of 1.2:1. Since the different strain parameters have different definitions, comparison of them is inappropriate.

Crystallite sizes range from 100 to 350–400 Å in the direction normal to the molecular axis but are only 80–100 Å in a direction parallel to the molecular axis. Chain folding must occur in this direction. The small amount of low-angle x-ray-scattering evidence available suggests that there exists along the molecular axis

[22] G. Challa, P. H. Hermans, and A. Weidinger, Makromol. Chem. **56**, 169 (1962).
[23] D. Ya. Tsvanskin, Polymer Sci. USSR **6**, 2310 (1964).

alternating crystalline and amorphous regions of approximately equal size. This conclusion is in agreement with independent measurements of crystallinities in isotactic polystyrene.[22]

Lattice distortions parallel to the basal plane in all three samples are observed to fall off rapidly to a constant value as the domain over which they are averaged is increased. This behavior is consistent with a dislocation theory of lattice distortions, although explaining large amounts of noncrystalline material along the molecular axis is difficult in terms of dislocations. We have computed the approximate density of edge dislocations which would correspond to the observed values of lattice distortion parallel to the basal plane. An edge dislocation produces a maximum displacement of an atom from an equilibrium lattice site of $\mathbf{b}/2$, where $\mathbf{b}$ is the Burgers vector, and a minimum displacement of 0; therefore an average displacement per edge dislocation is approximately $\mathbf{b}/4$. Table II contains values of lattice distortions $e=\delta d/d$, (percent), over the entire sample. If D is the number of edge dislocations (per angstrom) in the direction of $\mathbf{b}$, the total lattice distortion observed is

$$e=\delta d/d=Db/4,$$

from which

$$D=4e/b. \tag{12}$$

An approximate measure of the dislocation density per unit area is thus available. We take $|b|$ to be equivalent to d_{110} (10.95 Å). For over-all lattice distortions of 0.002, 0.005, and 0.010, Eq. (12) leads to dislocation densities of 5×10^{9}, 3×10^{10}, and 1×10^{11} dislocations/cm^2, respectively. These values are slightly less than those observed for highly strained metals.[24]

SUMMARY AND CONCLUSIONS

A number of the different methods proposed to separate the effects of crystallite size and of lattice distortion on x-ray line broadening were investigated for samples of isotactic polystyrene. Analysis of the x-ray patterns from three samples with different thermal and/or mechanical histories indicated a concomitant increase of crystallite size and decrease of lattice distortion on annealing. Observed diffraction line broadenings, then, cannot be interpreted solely in terms of one or the other possible effect.

Our results indicate clearly that a meaningful separation of size and distortion effects on the x-ray pattern is possible for polymeric systems; this had not been established before. Instrumental requirements are very severe: (a) Precise line-profile data must be collected over large regions of reciprocal space regardless of the size-distortion separation method to be used. (b) An accurate correction for background scattering is necessary. (c) A reliable method to separate (resolve) overlapping peaks must be found. Once these criteria are met, any of the size-distortion analyses may be made.

We found, in particular, that the integral breadth methods qualitatively are all equally valid, with the sole exception of the assumption of Cauchy line profiles for the size and the distortion effects. "Weight-average" crystallite sizes of the order of a few hundred angstroms or less are thereby obtained. In addition, a complete Fourier transform process supplies alternative measures of size and distortion to those obtained by integral breadth methods: "*number*-average" crystallite size and rms lattice strain. Having two measures of each distribution furnishes a better characterization of the sample than does a single one. Of significance, then, is the fact that true separation of size from strain effects requires exactly the same number of resolved reflection orders for the integral breadth methods as for the Fourier transform method, whereas the quality of data required for integral breadth methods is only slightly inferior to that required by the Fourier transform method. Therefore, in order to obtain the maximum available amount of information, the full Fourier transform method should be applied whenever possible; the integral breadth methods become of importance whenever data adequate for the Fourier transform method are not available.

Finally, our crystallite size results for the molecular axis direction, coupled with positions and half-widths of low-angle x-ray maxima, indicate that a structural model containing regularly alternating crystalline and noncrystalline regions of approximately the same size is appropriate for the isotactic polystyrene system.

[24] C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, New York, 1957), pp. 554–557.

Solution of the Cauchy Problem for Laplace's Equation in Axially Symmetric Systems*

KENNETH J. HARKER
Microwave Laboratory, W. W. Hansen Laboratories of Physics, Stanford University, Stanford, California
(Received 27 March 1963)

The determination of the electrodes for an electron gun reduces to solving the Cauchy problem for Laplace's equation. Using Green's Theorem, an analytic solution for the case of axial symmetry is obtained in terms of a line integral involving the product of the analytically continued values of the voltage and normal electric fields at the beam edge by an appropriate weighting function. Since an explicit solution is given in terms of line integrals, numerical results should be obtained far more accurately and quickly than those obtained by integrating Laplace's equation directly by finite-difference methods.

I. INTRODUCTION

In recent papers[1,2] the author has described a general method for obtaining the electrodes for axially symmetric electron guns by integrating a hyperbolic partial differential equation by finite-difference methods. Daykin[3] and Radley[4] have given analytic solutions for the electrodes of a cylindrical beam, and the latter[4] has also given an analytic solution for the conical beam.

In this paper we shall give an analytic and completely general solution to the electrode problem for axially symmetric beams in terms of line integrals according to a method developed recently by Garabedian.[5] Since the solution given involves only the evaluation of line integrals, it should be far more accurate and faster than the method cited above[1,2] involving the integration of a partial differential equation by finite-difference methods. Also, being completely general, it should be applicable to all types of beams, including those which are curvilinear.

In the next section of this paper, a derivation of this solution will be given in detail. An analytic continuation of Laplace's equation is made in order to transform it into hyperbolic form. This, then, allows us to use Green's Theorem[6] to obtain a solution in terms of a line integral involving the product of an appropriate weighting function by the analytically continued values of the voltage and normal field at the beam edge.

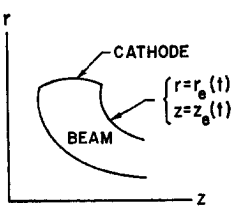

FIG. 1. Gun geometry.

In the following two sections these results are applied to the determination of the electrode shapes for cylindrical and conical beams. The paper concludes with a brief discussion of the effects of singularities on the solution.

II. GENERAL SOLUTION OF LAPLACE'S EQUATION

Consider an axially symmetric flow with cylindrical coordinates r and z, shown in the meridian plane in Fig. 1. The beam edge is described in terms of some parameter t by the equations

$$r = r_e(t), \tag{1}$$

$$z = z_e(t), \tag{2}$$

and the corresponding values of the voltage and its normal derivative at this edge by

$$V = V_e(t), \tag{3}$$

$$V_n = V_{ne}(t). \tag{4}$$

The parameter t is arbitrary, and could be either time, arc length, or some function thereof. The variables r_e, z_e, V_e, and V_{ne} must be analytic functions of t. This hypothesis is necessary whenever an elliptic equation such as Laplace's is solved for open boundary conditions. The means by which these functions are determined is discussed in previous papers.[1,2] We will also find it desirable at this point to define the additional variables α and β by the equations

$$\alpha(t) = dr_e(t)/dt, \tag{5}$$

* The research reported in this paper was sponsored by the Lewis Research Center, NASA, Cleveland, Ohio, under Contract NAS 8-1509.

[1] K. J. Harker, J. Appl. Phys. **31**, 2165 (1960).
[2] K. J. Harker, J. Appl. Phys. **33**, 1861 (1962).
[3] P. N. Daykin, Brit. J. Appl. Phys. **6**, 248 (1955).
[4] D. E. Radley, J. Electron. Control **4**, 125 (1958).
[5] P. R. Garabedian, *Studies in Mathematics and Physics Presented to Richard von Mises*, (Academic Press Inc., New York, 1954), p. 149.
[6] D. V. Widder, *Advanced Calculus* (Prentice-Hall, Inc., New York, 1947), p. 191.

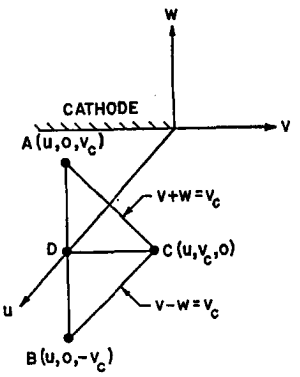

FIG. 2. Conformal coordinate system.

$$\beta(t) = dz_e(t)/dt. \tag{6}$$

In order to obtain the focusing electrodes, we must find a solution of Laplace's equation,

$$\frac{\partial^2 V}{\partial r^2} + \frac{\partial^2 V}{\partial z^2} + \frac{1}{r}\frac{\partial V}{\partial r} = 0, \tag{7}$$

outside the beam, which satisfies the boundary conditions given by Eqs. (3) and (4).

Our first step is to introduce a new dependent variable W related to V by the relation

$$W = r^{\frac{1}{2}}V. \tag{8}$$

When this is introduced into Eq. (7), we obtain the new equation

$$\frac{\partial^2 W}{\partial r^2} + \frac{\partial^2 W}{\partial z^2} + \frac{W}{4r^2} = 0. \tag{9}$$

Next it will be necessary to introduce the conformal transformation

$$r + iz = r_e(t + iv) + iz_e(t + iv), \tag{10}$$

where r_e and z_e are given by Eqs. (1) and (2). This transformation maps the beam boundary into the curve $v = 0$. Equation (9) transforms under this transformation into the equation

$$\frac{\partial^2 W}{\partial t^2} + \frac{\partial^2 W}{\partial v^2} + \frac{1}{4r^2}\left[\left(\frac{\partial r}{\partial v}\right)^2 + \left(\frac{\partial r}{\partial t}\right)^2\right]W = 0. \tag{11}$$

In order to solve this equation, it will be necessary to reduce it to hyperbolic form by analytically continuing t into the complex domain. This is done by setting

$$t = u + iw. \tag{12}$$

Since

$$\partial/\partial t = (1/i)(\partial/\partial w), \tag{13}$$

Equation (11) is reduced to the hyperbolic form

$$\frac{\partial^2 W}{\partial v^2} - \frac{\partial^2 W}{\partial w^2} + \frac{1}{4r^2}\left[\left(\frac{\partial r}{\partial v}\right)^2 - \left(\frac{\partial r}{\partial w}\right)^2\right]W = 0. \tag{14}$$

By substituting Eq. (12) into Eq. (10) and its complex conjugate, we find that the analytically continued values of r and z are determined by the equations

$$r + iz = r_e[u + i(w + v)] + iz_e[u + i(w + v)], \tag{15}$$

$$r - iz = r_e[u + i(w - v)] - iz_e[u + i(w - v)]. \tag{16}$$

Figure 2 shows the conformal coordinate system (u, v, w) to be used for solving Laplace's equation. We shall determine the voltage at a point C located at the point (u, v_C, o) in terms of a line integral from D to A. From Eqs. (15) and (16), the cylindrical coordinates of C are given by

$$r_c = \text{Re}\,[r_e(u + iv_c)] - \text{Im}\,[z_e(u + iv_c)], \tag{17}$$

$$z_c = \text{Im}\,[r_e(u + iv_c)] + \text{Re}\,[z_e(u + iv_c)]. \tag{18}$$

Points A and B are located at the points (u, o, v_C) and (u, o, $-v_C$), respectively, in order that points along the lines AC and BC shall satisfy the relations

$$v + w = v_c, \tag{19}$$

$$v - w = v_c, \tag{20}$$

respectively. The data given by Eqs. (1–4) can be continued analytically so that initial data are known over the surface $v = 0$ and in particular along the line AB.

An important property which holds along lines AC and BC and which we will need later can be obtained as follows. Equations (15) and (19) imply that $r + iz$ is constant and that, in particular,

$$(r + iz) - (r_c + iz_c) = 0 \tag{21}$$

along AC. Multiplying both sides of this equation by $[(r - iz) - (r_c - iz_c)]$, we find that

$$(r - r_c)^2 + (z - z_c)^2 = 0 \tag{22}$$

along AC. A similar argument for Eqs. (16) and (20) leads to the same result along BC.

The voltage at point C is obtained by the application of Green's Theorem[6] over the surface ABC shown in Fig. 2. First, we introduce the Riemann function, R, for Eq. (14).[7] This function is defined to take the value unity along AC and BC and satisfies the same equation as W:

$$\frac{\partial^2 R}{\partial v^2} - \frac{\partial^2 R}{\partial w^2} + \frac{1}{4r^2}\left[\left(\frac{\partial r}{\partial v}\right)^2 - \left(\frac{\partial r}{\partial w}\right)^2\right]R = 0. \tag{23}$$

If we define the functions P and Q with the relations

[7] B. Epstein, *Partial Differential Equations*, (McGraw-Hill Book Company, Inc., New York, 1962), p. 55.

$$P = R(\partial W/\partial w) - W(\partial R/\partial w), \tag{24}$$

$$Q = R(\partial W/\partial v) - W(\partial R/\partial v), \tag{25}$$

then it is clear from Eqs. (14) and (23) that

$$(\partial Q/\partial v) - (\partial P/\partial w) = 0. \tag{26}$$

It then follows from Green's Theorem that

$$\oint_{\text{ACBA}} (P\,dv + Q\,dw) = 0. \tag{27}$$

Evaluation of this integral along the contour ACBA leads to the expression

$$W_c = \text{Re}\left\{W_a + \int_0^{w_c} \left(R\frac{\partial W}{\partial v} - W\frac{\partial R}{\partial v}\right) dw\right\}, \tag{28}$$

where W_c and W_a are the values of W at C and A, respectively.

Before proceeding with the further reduction of this integral, it will be necessary to obtain an explicit form for R. By noting that R satisfies the same differential equation as W, it is clear that

$$\frac{\partial^2 R}{\partial r^2} + \frac{\partial^2 R}{\partial z^2} + \frac{R}{4r^2} = 0. \tag{29}$$

If we introduce the new independent variable

$$k = i\{[(r - r_c)^2 + (z - z_c)^2]/4rr_c\}^{\frac{1}{2}}, \tag{30}$$

Eq. (29) reduces to the ordinary differential equation

$$k(1 - k^2)\frac{d^2R}{dk^2} + (1 - 3k^2)\frac{dR}{dk} - kR = 0. \tag{31}$$

This equation is satisfied by the complete elliptic integral of the first kind, $K(k)$, and therefore the solution for R is given by

$$R = (2/\pi)K(k). \tag{32}$$

According to Eq. (22), k assumes the value zero along the lines AC and BC, and consequently, since $K(0) = \frac{1}{2}\pi$, R assumes the correct value of unity along these same two lines. Noting that[8]

$$K[i\gamma/(1 - \gamma^2)^{\frac{1}{2}}] = (1 - \gamma^2)^{\frac{1}{2}}K(\gamma), \tag{33}$$

it is possible for us to write Eq. (32) in the more suitable form

$$R = (2/\pi)\{4rr_c/[(r + r_c)^2 + (z - z_c)^2]\}^{\frac{1}{2}}K(\sigma), \tag{34}$$

where

$$\sigma = \left[\frac{(r - r_c)^2 + (z - z_c)^2}{(r + r_c)^2 + (z - z_c)^2}\right]^{\frac{1}{2}}. \tag{35}$$

[8] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists*, (Springer-Verlag, Berlin, 1954), p. 40.

We are now in a position to derive the final formula for the voltage. In carrying this out, we will need the relations

$$\partial r/\partial v = -\beta(u + iw), \tag{36}$$

$$\partial z/\partial v = \alpha(u + iw), \tag{37}$$

which hold on the surface $v = 0$, and follow from Eqs. (15) and (16). In addition we will need the relation[9]

$$\frac{dK}{d\sigma} = \frac{E - (1 - \sigma^2)K}{\sigma(1 - \sigma^2)}. \tag{38}$$

By substituting Eqs. (8), (34), and (36–38) into Eq. (28) and collecting terms, we obtain the final formula for the voltage,

$$V(r_c, z_c) = \text{Re}\left\{\left(\frac{r_a}{r_c}\right)^{\frac{1}{2}} V_a + \frac{2}{\pi}\int_0^{w_c} \frac{dw}{[(r_e + r_c)^2 + (z_e - z_c)^2]^{\frac{1}{2}}} \times \left[2r_eK(\sigma)(\alpha^2 + \beta^2)^{\frac{1}{2}}V_{ne} + 2r_eV_e[K(\sigma) - E(\sigma)] \times \frac{\alpha(z_e - z_c) - \beta(r_e - r_c)}{(r_e - r_c)^2 + (z_e - z_c)^2} - \beta V_eE(\sigma)\right]\right\}, \tag{39}$$

where r_a, V_a, r_e, z_e, V_e, V_{ne}, α, and β are abbreviated symbols representing $r_e(u + iv_c)$, $V_e(u + iv_c)$, $r_e(u + iw)$, $z_e(u + iw)$, $V_e(u + iw)$, $V_{ne}(u + iw)$, $\alpha(u + iw)$, and $\beta(u + iw)$, respectively. The unit normal vector by which the sign of V_{ne} is determined has been taken to point in the direction of increasing v. Functions K and E are complete elliptic integrals of the first and second kind of the argument σ defined in Eq. (35) and can be obtained by numerical integration of Eq. (38) and the equation[9]

$$dE/d\sigma = (E - K)/\sigma. \tag{40}$$

In integrating Eqs. (38) and (40) it should be borne in mind that σ is, in general, complex.

III. ELECTRODES FOR A CYLINDRICAL BEAM

In order to clarify the application of Eq. (39), we shall apply it to the determination of the voltages, and thereby the focusing electrodes, of a parallel-flow cylindrical beam. Results in a different form for this case have been derived by Daykin[3] and Radley.[4]

By using properly normalized variables, the boundary equations corresponding to Eqs. (1)–(4) can be written as

[9] Reference 8, p. 282.

$$r_e = 1, \tag{41}$$

$$z_e = t, \tag{42}$$

$$V_e = t^{4/3}, \tag{43}$$

$$V_{ne} = 0, \tag{44}$$

where t is the arc length along the beam edge. By replacing t by $u + iw$ in Eqs. (41)–(43), we find that the values of r, z, and V along the path of integration DA in Fig. 2 are given by

$$r_e = 1, \tag{45}$$

$$z_e = u + iw, \tag{46}$$

$$V_e = (u + iw)^{4/3}. \tag{47}$$

From Eqs. (17) and (18), the coordinates of point C are given by

$$r_c = 1 - v_c, \tag{48}$$

$$z = u. \tag{49}$$

We shall replace w as our variable of integration in Eq. (39) by the radial coordinate

$$r = 1 - w, \tag{50}$$

which ranges in value from 1 to r_c between D and A. By substituting into Eqs. (46) and (47) we obtain the relations

$$z_e = z + i(1 - r), \tag{51}$$

$$V_e = [z + i(1 - r)]^{4/3} \tag{52}$$

along the path DA.

By substituting Eqs. (45), (51), and (52) into Eq. (39), and simplifying, we obtain for the voltage the relation

$$V(r_c, z) = \mathrm{Re}\,[z + i(r_c - 1)]^{4/3} r_c^{-\frac{1}{3}} + \frac{2}{\pi}\int_1^{r_c} \frac{\mathrm{Re}[z + i(r - 1)]^{4/3}\, dr}{[(r_c + r)(2 + r_c - r)]^{\frac{1}{2}}} \times \left\{E(\sigma) - \frac{2(r_c - 1)}{(r_c - r)(r_c + r - 2)}\,[K(\sigma) - E(\sigma)]\right\}, \tag{53}$$

where

$$\sigma^2 = \frac{(r_c - r)(r_c + r - 2)}{(r_c + r)(2 + r_c - r)}. \tag{54}$$

IV. ELECTRODES FOR A CONICAL BEAM

In this section we will solve Laplace's equation for the conical beam. This result can be compared with Radley's,[4] which is given in terms of Legendre functions.

We will let ρ and θ be the usual spherical coordinates, and θ_0 the cone half-angle. Also, the parameter t in Eqs. (1)–(4) will be defined by the relation

$$\rho = e^t, \tag{55}$$

in which case these equations assume the form

$$r_e = e^t \sin\theta_0, \tag{56}$$

$$z_e = e^t \cos\theta_0, \tag{57}$$

$$V_e = V_e(t), \tag{58}$$

$$V_{ne} = 0, \tag{59}$$

where V_e is the potential function for the spherical diode. By combining Poisson's equation and the equation of continuity and then normalizing the voltage properly, it can be shown that V_e satisfies the equation

$$(d^2V_e/dt^2) + (dV_e/dt) - V_e^{-\frac{1}{3}} = 0. \tag{60}$$

By replacing t by $u + iw$, we find that the values of r, z, and V along the path of integration DA in Fig. 2 are given by

$$r_e = e^{u+iw} \sin\theta_0, \tag{61}$$

$$z_e = e^{u+iw} \cos\theta_0, \tag{62}$$

$$V_e = V_e(u + iw). \tag{63}$$

From Eqs. (17) and (18), the coordinates of C are given by

$$r_c = e^u \sin\theta_c, \tag{64}$$

$$z_c = e^u \cos\theta_c, \tag{65}$$

where

$$\theta_c = \theta_0 - v_c. \tag{66}$$

Instead of w, we shall use as our variable of integration the angle coordinate

$$\theta = \theta_0 - w. \tag{67}$$

In terms of θ, Eqs. (61)–(63) take the form

$$r_e = e^u e^{i(\theta_0 - \theta)} \sin\theta_0, \tag{68}$$

$$z_e = e^u e^{i(\theta_0 - \theta)} \cos\theta_0, \tag{69}$$

$$V_e = V_e[u + i(\theta_0 - \theta)]. \tag{70}$$

By substituting into Eq. (39) and simplifying, we obtain for the voltage the formula

$$V(\rho, \theta) = \mathrm{Re}\left\{V_e[\ln\rho + i(\theta_0 - \theta_c)]e^{i(\theta_0 - \theta_c)/2}\left[\frac{\sin\theta_0}{\sin\theta_c}\right]^{\frac{1}{2}} - \frac{(2)^{\frac{1}{2}}}{\pi}\int_{\theta_0}^{\theta_c} \frac{V_e[\ln\rho + i(\theta_0 - \theta)]e^{i(\theta_0 - \theta)/2}\, d\theta}{[\cos(\theta - \theta_0) - \cos(\theta_c + \theta_0)]^{\frac{1}{2}}} \times \left[\frac{\sin\theta_0 \sin(\theta_c - \theta_0)}{\cos(\theta - \theta_0) - \cos(\theta_c - \theta_0)} \times (K(\sigma) - E(\sigma)) - \cos\theta_0 E(\sigma)\right]\right\}, \tag{71}$$

where

$$\sigma^2 = \frac{\cos(\theta - \theta_0) - \cos(\theta_c - \theta_0)}{\cos(\theta - \theta_0) - \cos(\theta_c + \theta_0)}, \quad (72)$$

and

$$\rho = e^u. \quad (73)$$

The function $V_e[\ln \rho + i(\theta_0 - \theta)]$ in this equation may be obtained numerically by integrating from θ_0 to θ the differential equation

$$\frac{d^2V_e}{d\theta^2} + \frac{1}{i}\frac{dV_e}{d\theta} + V_e^{-\frac{1}{2}} = 0, \quad (74)$$

obtained from Eq. (60) by substitution of Eqs. (12) and (67).

V. EFFECT OF SINGULARITIES

The functions r_e, z_e, V_e, and V_{ne} are generally continued analytically by the solution of a series of ordinary differential equations with t as the independent variable. By integrating these equations along contours through the complex domain of t which bypass any singularities present, it is possible to ascertain the values of the dependent variables for all values of t for which they are analytic. Isolated singularities occurring in V_e or V_{ne} at some point $t = a + ib$ in the plane $v = 0$ will be reflected in the physical solution as an axially symmetric distribution of charge and/or multipoles on the circle passing through the point with coordinates

$$r_c = \text{Re}\,[r_e(a + ib)] - \text{Im}\,[z_e(a + ib)], \quad (75)$$

$$z_c = \text{Im}\,[r_e(a + ib)] + \text{Re}\,[z_e(a + ib)], \quad (76)$$

obtained from Eqs. (17–18) by substitution of $a + ib$ for $u + iv_c$.

If either V_e or V_{ne} is multivalued, then a branch cut must be made along some line $u = u_0(w)$ in the plane $v = 0$. This is reflected in the physical solution as a discontinuity in the solution along the line $u = u_0(v)$ in the plane $w = 0$. Physically, the discontinuity represents the combination of an axially symmetric free charge and dipole-layer distribution on the surface of revolution obtained from the line represented parametrically as a function of v by the equations

$$r_c = \text{Re}\,[r_e(u_0(v) + iv)] - \text{Im}\,[z_e(u_0(v) + iv)], \quad (77)$$

$$z_c = \text{Im}\,[r_e(u_0(v) + iv)] + \text{Re}\,[z_e(u_0(v) + iv)]. \quad (78)$$

The charge and dipole densities are determined by the corresponding discontinuities in the voltage and normal field, respectively.

Another difficulty arises when $d(r + iz)/dt$ vanishes somewhere in the plane $v = 0$. This causes t, and thus possibly V_e and V_{ne}, to become multivalued functions of $(r_e + iz_e)$. This in turn may allow V to become of a multivalued function of r_c and z_c. The result is again a physical discontinuity consisting of a free charge and dipole-layer distribution on a surface of revolution, just as in the case of the branch cut discussed above.

The detailed effect of singularities is quite involved and has been carefully investigated by Bergman.[10]

[10] S. Bergman, *Integral Operators in the Theory of Linear Partial Differential Equations* (Springer-Verlag, Berlin, 1961).

Errata: An Approach to Gravitational Radiation by a Method of Spin Coefficients

E. NEWMAN
University of Pittsburgh, Pittsburgh, Pennsylvania
AND
R. PENROSE
Kings College, University of London, London, England
[J. Math. Phys. 3, 566 (1962)]

Page 568: The next-to-the-last term in Eq. (3.14) should read $\Lambda\epsilon_{a'b'}(\epsilon_{cd}\epsilon_{af} + \epsilon_{ad}\epsilon_{cf})$;

Page 569: The last term in the first equation of (3.17) should read $2\Lambda\xi_{(A}\epsilon_{B)C}$;

Page 570: The third term from the end of Eq. (4.2g) should read $-\nu\bar{\kappa}$;
The third term from the end of Eq. (4.2k) should read $+(\mu - \bar{\mu})\kappa$;
The first term after the equals sign in Eq. (4.2m) should read $+(\rho - \bar{\rho})\nu$;
The third term from the end of Eq. (4.2r) should read $+(\bar{\gamma} - \bar{\mu})\alpha$;
The third equation of (4.3a) should read $\Psi_2 = -\frac{1}{2}(C_{1212} - C_{1234}) = -\frac{1}{2}C_{\alpha\beta\gamma\delta} \times (l^\alpha n^\beta l^\gamma n^\delta - l^\alpha n^\beta m^\gamma \bar{m}^\delta) = \Psi_{0011}$.

Page 572: The second sentence after Eq. 6.3 should read "Sachs uses a "luminosity" parameter $\tilde{r}$, satisfying $D\tilde{r} = -\tilde{r}\rho$, which $\cdots$," etc.

Page 574: In footnote 26, the second sentence should read, " $\cdots$ affine parameter $\tilde{r}$ and tangent vector $\tilde{l}_\mu$ $\cdots$ ".

Page 575: The lemma should read, "Let the $(n \times n)$ matrix A be independent of r and have no eigenvalue with positive real part. Suppose, also that any eigenvalue of A with vanishing real part is regular (i.e. its multiplicity is equal to the number of linearly independent eigenvectors corresponding to it). Then all the solutions $\cdots$ ".

Page 578: In Eqs. (A3) of the Appendix, the differentiated terms on the right-hand side all appear with the wrong sign. Also the δ's on the right-hand sides of the third, fifth, and seventh equations should be $\bar{\delta}$'s. The equality symbol on the left in the seventh equation should, of course, be a minus sign. As an example this seventh equation should be correctly written

$$(D - 3\rho)\Psi_2 + 2\kappa\Psi_3 - (\bar{\delta} + 2\pi - 2\alpha)\Psi_1 + \lambda\Psi_0 = -\bar{\Phi}_0\Delta\Phi_0 + \bar{\Phi}_1\bar{\delta}\Phi_0 + 2(\bar{\Phi}_0\Phi_0\gamma - \bar{\Phi}_1\Phi_0\alpha - \bar{\Phi}_0\Phi_1\tau + \bar{\Phi}_1\Phi_1\rho).$$

Errata: Cluster Sums for the Ising Model

G. S. RUSHBROOKE* AND H. I. SCOINS
University of Durham, King's College, Newcastle upon Tyne, England
[J. Math. Phys. 3, 176 (1962)]

The following misprints should be noted, on pp. 183–4:

Simple Cubic Lattice β_5: $871\frac{1}{2}f^5$ should read $871\frac{1}{5}f^5$;

Body-centered Lattice β_5: $16f^2$ should read $16f^3$;

$\bar{b}_5$: $+9072\eta^{34}$ should read $-9072\eta^{34}$;

Equation (29) should read

$$\ln \Lambda(1, \eta) = 3 \ln (1 + \eta) - 2 \ln 2 + 3u^4 + 22u^6 + 187\tfrac{1}{2}u^8 + \cdots ;$$

and in the expression for b_l in terms of β_k, $l!$ should read l^2.

* At present, Visiting Professor, Chemistry Department, University of Oregon, Eugene, Oregon.