DEPARTMENT OF PHYSICS

VOLUME 1 | 1949 | NUMBER 6

THE APPLICATION OF GEOMETRICAL CONSTRUCTIONS TO THE THEORY OF OPTICAL INTERFERENCE AND DIFFRACTION PHENOMENA

By

A. L. REIMANN, PH.D., D.SC., F.INST.P.

(Published as an original paper by the University of Queensland.)

DATE OF PUBLICATION:
4th April, 1949

Printed by WATSON, FERGUSON & Co.
Stanley Street, South Brisbane

THE APPLICATION OF GEOMETRICAL CONSTRUCTIONS TO THE THEORY OF OPTICAL INTERFERENCE AND DIFFRACTION PHENOMENA.*

By A. L. REIMANN, PhD., D.Sc., F.Inst.P.

Introduction—It is well known that the sum of two simple harmonic disturbances,

$$s_1 = a_1 \sin (\omega t + \epsilon_1)$$

and

$$s_2 = a_2 \sin (\omega t + \epsilon_2),$$

in the same direction and having the same frequency is itself a simple harmonic disturbance,

$$s = a \sin (\omega t + \epsilon),$$

whose frequency and direction are those of its components, and that the amplitude, a, of this resultant disturbance may be obtained by the simple vector construction of Fig. 1. If more than the two disturbances are to be compounded the process is extended, and instead of a vector triangle we then have a vector polygon, as shown in Fig. 2.

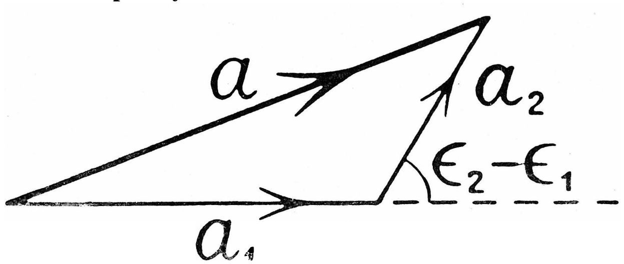

Fig. 1

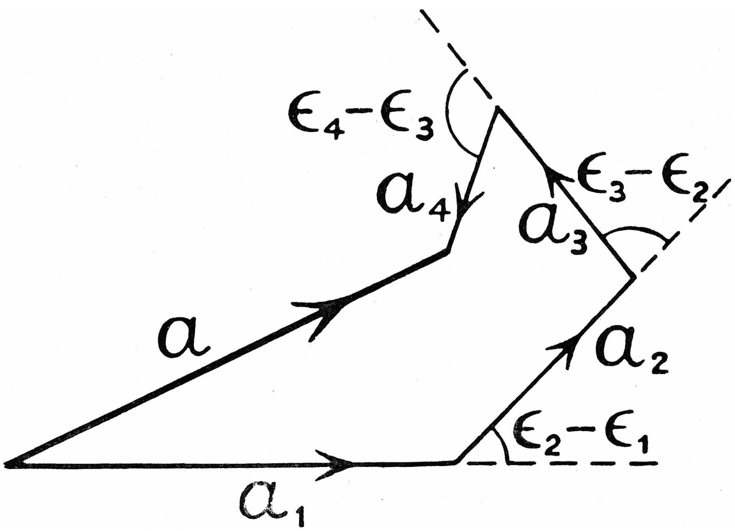

Fig. 2

* Paper read at the Perth meeting of the Australian and New Zealand Association for the Advancement of Science, August, 1947.

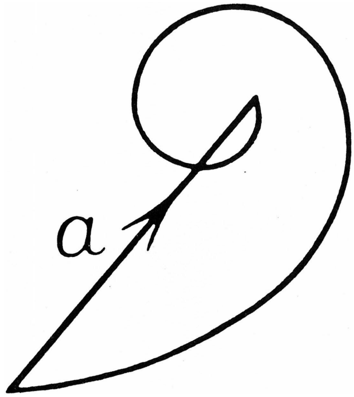

Fig. 3

It frequently happens that we wish to compound an infinite number of infinitesimal disturbances. In such a case our vector polygon becomes a smooth curve, as shown in Fig. 3. Perhaps the best-known example of such a curve is Cornu's spiral.

Now Cornu's spiral is the particular form of construction appropriate to one very special class of Fresnel diffraction phenomena. While corresponding constructions are, on occasion, employed for other cases, it is still somewhat unusual to do so. Yet the method is very versatile, and lends itself quite simply to the solution of a great variety of problems in physical optics. It seems that it would surely be profitable to make much more widespread use of these constructions than is at present customary. In the present paper it is proposed to apply this method to a few selected problems in optics which do not appear to have been treated in this manner elsewhere.

Actual Light Sources—Light sources are never simple, but always contain innumerable radiating centres having both a random spatial distribution and a random phase distribution. This situation applies to all optical interference and diffraction phenomena. Thus if, in Fig. 4, QM and QN represents two rays

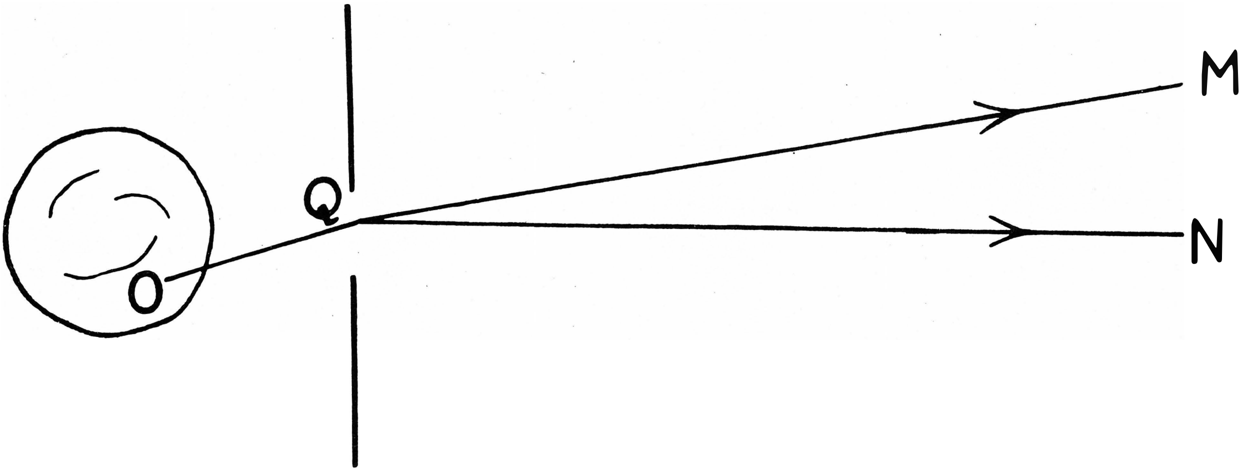

Fig. 4

emanating from some point Q in an " illuminated slit," which ultimately interfere, it is obvious that they can do so only by virtue of diffraction at Q of light derived from one and the same centre in the source, e.g., that marked O. Contributing to the two rays QM and QN we must have disturbances coming from all centres such as O in the light source, the light being diffracted at Q, i.e., points in the plane of the slit being the seat of Huygens' secondary wavelets. What will be the most probable amplitude of the sum of these secondary wavelets ?

The existence of this problem was recognized long ago by Rayleigh,* though it is often forgotten by modern writers. Rayleigh treated the problem algebraically. There is, however, an interesting alternative vector treatment. If we combine all the components in any particular plane of polarization of the Huygens secondary wavelets emanating from Q which originate from all centres such as O by the geometrical method we shall obtain a very peculiar sort of vector polygon, some idea of which is given in Fig. 5. The question resolves itself into this : What will

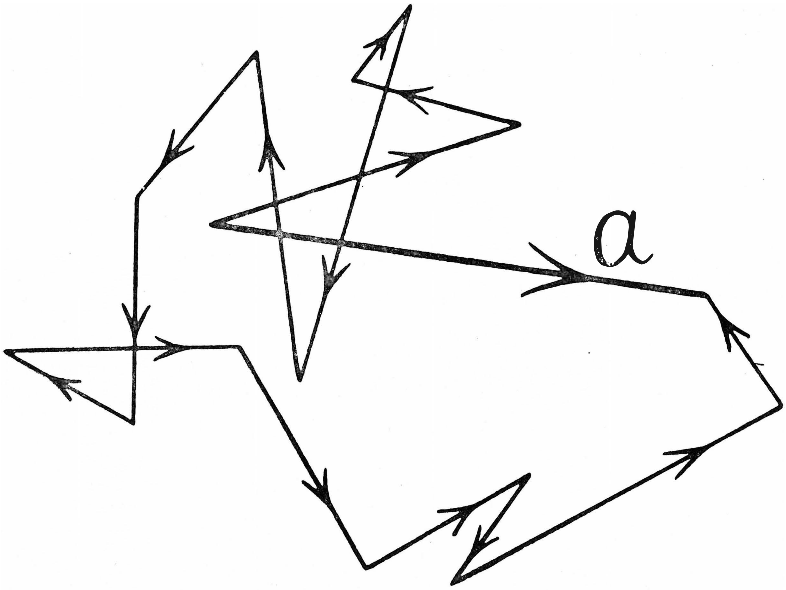

Fig. 5

be the most probable magnitude of the resultant of a very large number of vectors in the same plane, having some well-defined distribution of magnitudes, but whose directions are distributed entirely at random ?

Now a well-known theorem in probability states that the most probable magnitude of the algebraic sum of a number of positive and negative quantities, equal in magnitude but having signs distributed at random, is proportional to the square root of the number of quantities added.

This theorem may be extended in two respects. First, the quantities need not all be of the same magnitude, provided they conform to some definite law of distribution. Secondly, they can be vectors instead of scalars, the components of the vectors in some chosen direction then corresponding to the positive and negative scalar quantities. The statement of the extended theorem is that the most probable magnitude of the resultant of a large number of vectors whose magnitudes are distributed according to some definite statistical law and whose directions are distributed at random is proportional to the square root of the number of vectors compounded. Thus in our optical problem the most probable amplitude of the resultant of all the Huygens secondary wavelets sent out simultaneously from Q

* Lord Rayleigh, "Wave Theory of Light," Encyl. Britt., XXIV., 1888.

must be proportional to the square root of the number of radiating centres. The resultant intensity is therefore proportional to the number itself, which is as it should be according to energy considerations. We may note, incidentally, as a matter of interest, that from the same mathematical theorem as that on which this conclusion is based we may also infer that in Brownian movement the most probable distance a particle will be found from its starting point is proportional to the square root of the time.

The above reasoning concerned itself only with the light diffracted at Q. But the observed interference or diffraction fringes are due to the light diffracted at all such points as Q. It is easily seen that, as regards intensity, the combined effect due to light originating from all centres such as O and diffracted at all points in the slit such as Q will again be proportional to the number of centres.

Similarly we see that if, instead of having an illuminated slit as light source, we used a short, straight and narrow incandescent wire, or a section of capillary tube containing a gas through which a discharge is passed, the intensity at any point in the interference or diffraction pattern will be proportional to the projected area of the source.

Of course in no actual case do the centres from which the light originates continue to radiate indefinitely; they all have finite radiation "lives." Also in most cases the centres are not stationary, in consequence of which the radiations received from them are not all of exactly the same frequency (Doppler effect). We may, however, consider any interval of time which is so short that during it no important number of centres has begun or ceased to radiate, and no appreciable phase shift between the waves received from different centres has occurred. Then the most probable value of the intensity due to all the centres during this interval is proportional to their number n, and consequently the mean intensity over a long period of time is likewise proportional to n.

It is interesting to note in this connection that even if all the centres were stationary and radiated with strictly the same frequency there would, owing to the finiteness of their radiation lives, necessarily be a fluctuation about a mean, of both the intensity and the frequency of the resultant Huygens secondary wavelets emanating from Q. The fluctuation in frequency follows from the fact that at instants of time whose separation is comparable with, or larger than, the mean radiation life of a centre, the phases of the resultant disturbances at Q are related to one another in an entirely random manner.

Subsidiary Maxima Produced by a Diffraction Grating—The vector diagram appropriate to the problem of a diffraction grating consists of a number of equal elements, each inclined at the same angle to its predecessor, the number of elements being equal to the number of lines. When the number of lines is large this approximates to an arc of a circle of constant length, which curls up more and more as the angle of viewing is increased. Main maxima correspond to the special cases where this arc is a straight line (radius infinite), i.e., where the angle between adjacent elements of the diagram is a whole number of complete revolutions, while subsidiary maxima and minima correspond to a curling up of the diagram into an odd and even number of half-circles respectively. By consideration of these diagrams we easily obtain the angles at which the main and subsidiary maxima and the minima between them are produced, and also the relative amplitudes ($\pi : \frac{2}{3} : \frac{2}{5} : \frac{2}{7} : \ldots\ldots$) and intensities of the maxima. All this has been done elsewhere, and it is not proposed to repeat it here. However, one problem in this field which does not appear to have been dealt with by the geometrical method, or even, for that matter, very satisfactorily by the algebraical method, is that of the *number* of subsidiary maxima between two consecutive main maxima. This will therefore now be considered.

The easiest way to find the number of subsidiary maxima is the indirect one of finding the number of minima.

Let the number of lines in the grating be n. Then a necessary condition that a vector diagram having n equal elements each inclined at the same angle to its predecessor shall close is that a fictitious (n + 1)th element shall be in line with the first, as indicated in Fig. 6. We shall have this state of affairs when the angle turned through between the first and the (n + 1)th element is a whole-number multiple of 2π. This condition, though necessary, is not sufficient, however, for if the number in question is a multiple of n (the number of lines), say mn, so that the angle between the first and the (n + 1)th element is mn.2π, then the angle ϕ between *successive* elements is m.2π, and the elements will all be strung out along a straight line, giving a main maximum—that of order m. However, whenever the number of complete revolutions between the first and (n + 1)th element is *not* a multiple of n, the vector diagram will close, giving a minimum. Thus when the number is mn we have the main maximum of order m. When it is (m + 1)n the main maximum of order (m + 1)n is formed. At all whole-number values *between* mn and mn + n we have minima. There are, therefore, n − 1 occasions on which the condition for a minimum will be fulfilled between two successive main maxima. And since between each adjacent pair of these minima there is a subsidiary maximum, the number of these must be n − 2.

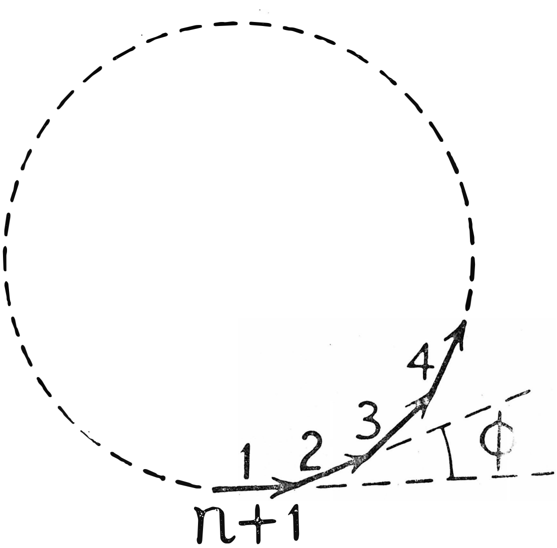

Fig. 6

Diffraction by a Circular Aperture—This is of importance in connection with optical instruments. Let us consider first Fraunhofer diffraction, which is involved in the theory of the resolving power of a telescope.

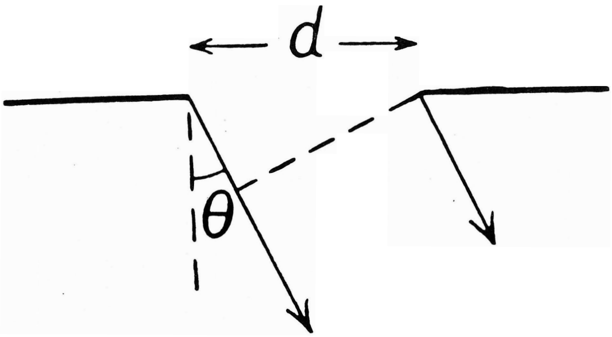

Fig. 7

In the case of a *rectangular* aperture or slit of width d (see Fig. 7) the appropriate vector diagram for light diffracted in a direction making an angle θ with the incident light is an arc of a circle of constant length, and the first minimum will occur when this just closes, forming one complete circle, d sin θ then being equal to the wave-length, λ.

Now it can be shown, by a rather lengthy and tedious process of integration, first carried out by Airy,* that if, instead of a rectangular aperture, we employ a circular one of the same diameter, d, the first minimum occurs at a somewhat greater angle ϕ, such that

$$\phi = 1.22\ \theta\ .$$

The value of the correcting factor, 1.22, which it is necessary to apply for the case of a circular aperture can be obtained quite simply, with an error not greater than about 1 per cent., geometrically. The principle of the method is as follows :

In the case of the rectangle, equal elements of the vector diagram correspond to strips of equal width (and therefore area) of the rectangle, as indicated in Fig. 8.

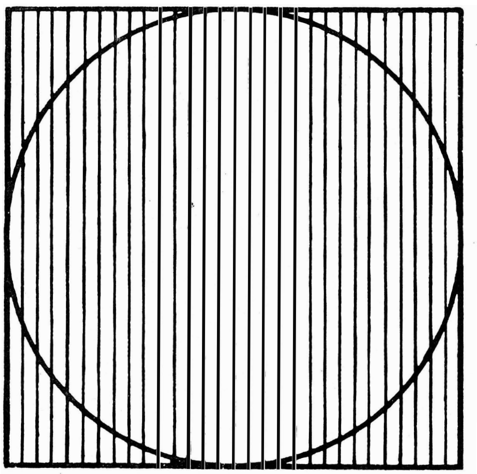

Fig. 8

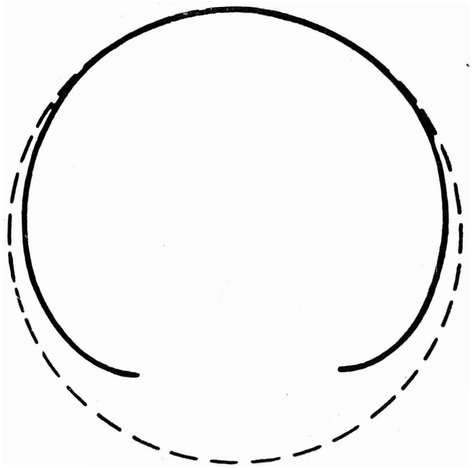

Fig. 9

Now if, instead, we consider only those parts of the rectangle contained within the circle, we have the same inclination of successive elements of the vector diagram to one another as before, but the lengths become progressively less as we approach the sides of the circle, so that, for the angle θ that produces closure for the rectangle, we have, instead of the closed-circle vector diagram (shown dashed in Fig. 9), one

* G. B. Airy, Cambr. Phil. Trans., 1834, p. 283.

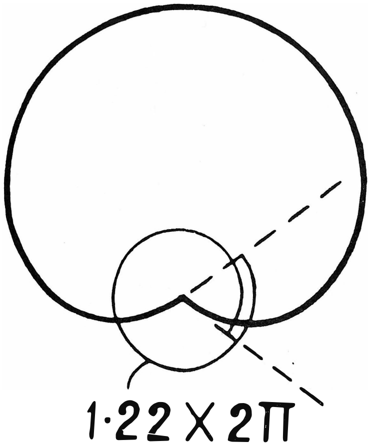

Fig. 10

that fails to close. To make it close we find we have to increase the angle from θ to ϕ, where ϕ is equal to 1.22 θ, when we obtain the diagram shown in Fig. 10.

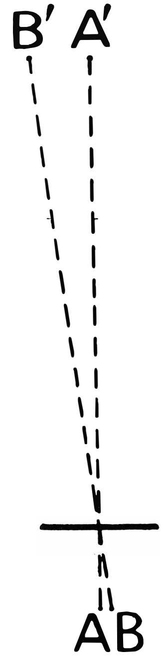

Fig. 11

The case of the microscope is much less simple, and only the main features of the problem will be sketched here.

Let us consider two neighbouring points A and B in the object viewed (Fig. 11). It is required to find under what circumstances it will be possible to see these as separate. Let the geometrical images of A and B be A′ and B′ respectively. Then all the optical paths from A, through the lens system, to A′ will be the same, and likewise those from B to B′. A′ will be the central maximum of the diffraction pattern due to light from A, and B′ that due to light from B. We shall have to find, then, the condition that A′ shall lie on the first diffraction minimum of the light from B and B′ on that of the light from A.

To study this condition, let us consider rays from B converging, via Huygens' secondary wavelets generated at the surface of the first lens of the objective, to A′ (Fig. 12). Let O be the point on this surface where the axis intersects it, while P is some other point. Then the problem is to find the condition that the resultant amplitude at A′ due to light from B passing through all points such as P is zero.

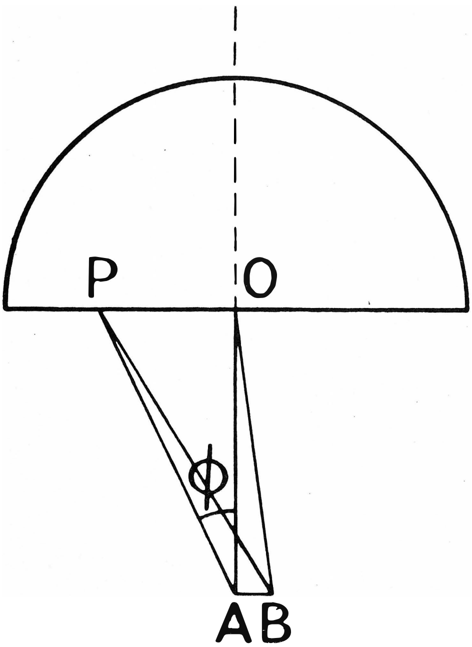

Fig. 12

We see in the first place that since all rays from A to A′ have the same optical path, and since to a sufficiently close approximation AO and BO are equal, the optical path difference between rays from B passing through P and O and ultimately converging in A′ is equal to BP—AP, which in turn is equal to AB sin ϕ. Since this is a function of AB and ϕ only, we may disregard the lens itself except only in so far as it sets a limit to ϕ. Let this limit be denoted by θ.

The problem, being three-dimensional, is correspondingly somewhat intractable. As a first step in its simplification, let us consider a very special case, where θ is small, and also let us admit light only through a narrow central strip of the segment of a sphere subtended by the lens, as indicated in Fig. 13, in a plane containing AB. Then if we cut this strip into equal rectangular elements, each of these will correspond to an equal increment in ϕ ($= \sin \phi$), and so to an equal increment in phase. Under these circumstances the appropriate vector diagram becomes the arc of a circle, which will just close, producing a minimum at A′, if the path difference between the two extreme rays is λ, i.e., if

$$2 \text{ d} \sin \theta = \lambda,$$

where d is written for AB.

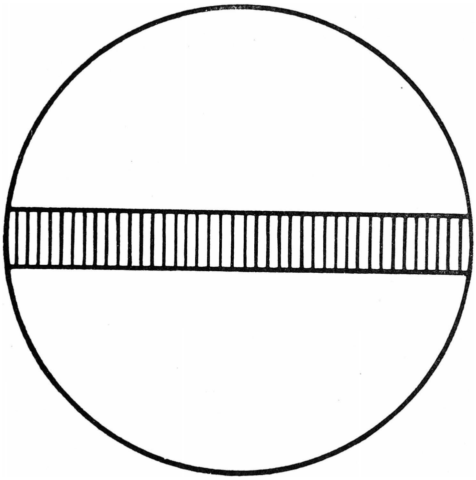

Fig. 13

From this stage it is not difficult to extend the reasoning to the actual case of a circular aperture, by expanding the elements laterally to the boundaries of this aperture. Thus for a circular aperture and θ small we shall have as the condition to be fulfilled :

$$2 \text{ d} \sin \theta = 1.22 \lambda.$$

The factor 1.22 arises, as before, from the fact that the contributions from parts for which ϕ is small relative to θ (ϕ being still the same angle in the plane of the original strip) are now more important than from parts where it is comparable with θ.

Now in an actual microscope θ is never small, and we must take account of the fact that $\sin \phi$ increases at a progressively less rapid rate as ϕ increases, so that on this score the areas contributing to equal steps in phase become progressively larger as we go outwards, and this tendency opposes that which we had in passing from a rectangular to a circular aperture, so that for *some* value of θ lying within the practical range the two opposing effects will just cancel each other out, and the condition to be fulfilled will be the same as for a narrow strip and small θ, viz.,

$$2 \text{ d} \sin \theta = \lambda.$$

This is the equation which is usually taken as valid for all values of θ. It has, however, been sufficiently demonstrated that in fact it can apply for only one particular value. For others it should be replaced by

$$2 \text{ d} \sin \theta = \text{f} \cdot \lambda,$$

where the factor f has the value 1.22 when θ is small, and has values less than 1 for θ sufficiently large. The variation of f with θ can only be found by means of a much fuller and more systematic investigation than has been attempted here. This problem lends itself well to solution by the geometrical method, the condition for closure being found for a series of values of θ. Such an investigation, which it is hoped to undertake shortly, will show that f is not only a function of θ but also of the state of polarization of the light used relative to the direction AB.