GEOMETRIC THEORY OF FRESNEL DIFFRACTION PATTERNS

Part I. Basic Ideas

BY Y. V. KATHAVATE

(From the Department of Physics, Indian Institute of Science, Bangalore)

Received April 2, 1945

(Communicated by Sir C. V. Raman, KT., F.R.S., N.L.)

1. INTRODUCTION

LIGHT diverging from a point source falls upon a distant screen, while between them is interposed an opaque plane obstacle of limited extension, or alternatively, an opening of determinate shape in a opaque sheet of unlimited extension. Diffraction effects (known as Fresnel patterns) are then observed on the illuminated screen. The usual procedure for a theoretical determination of the nature of these patterns is to express the light disturbance at any point on the screen as a summation of elementary disturbances having their origins distributed over the unobscured part of the wave-front. The numerical evaluation of the integrals which arise and the computation of the intensities at a sufficient number of points on the screen is, in general, a laborious process. Apart from this, it may be remarked that the method followed for the computation of intensities affords no insight into the relationship between the form of the obstacle or aperture on the one hand and the geometric characters of the diffraction pattern on the other. From the physical point of view, therefore, the usual approach to the theory of Fresnel diffraction patterns leaves much to be desired.

It is proposed in this paper to indicate a different approach to the subject which enables the nature of the Fresnel pattern due to an obstacle or an aperture of arbitrary form to be very readily derived in such manner as to admit of direct comparison with the experimental facts. In the succeeding papers of the series, various cases of special interest are worked out and the theoretically derived patterns are exhibited alongside of those observed experimentally. The remarkable power of the method and its essential simplicity become evident in the light of these examples.

2. DIFFRACTION BY A STRAIGHT EDGE

The present approach to the general theory of Fresnel patterns will be most readily understood by reference to the most familiar case of all, namely, that of the diffraction of light by an opaque obstacle bounded by a straight edge. Here, as is well known, instead of the discontinuity between light and darkness indicated by geometrical optics, the pattern exhibits an intensity distribution which is continuous in passing through the boundary, but is different on the two sides of it, oscillating in the region of light, and progressively diminishing to zero in the region of shadow. The evaluation of the Fresnel integrals enables this distribution to be quantitatively derived. But, as has often been remarked, it is easier to understand the observed results physically, if, taking the hint suggested by the actual facts of observation (Gouy, 1886), we assume the edge of the obstacle to be a source of cylindrical waves radiating from it. These waves give rise to the observed illumination in the region of shadow, while, in the region of light, their interference with the primary incident waves gives rise to the observed fluctuations of intensity. Since the discontinuity of illumination at the geometric boundary of the shadow disappears and is replaced by a continuous distribution, it is evident that the cylindrical waves exhibit a sudden reversal of phase at the boundary, and also that their amplitude becomes very large and comparable with that of the primary waves in the same direction. These features are a natural consequence of the manner in which the cylindrical waves arise. The genesis of the cylindrical waves, and their relation to the physical properties of the obstacle has been discussed in considerable detail by Raman and Krishnan (1925). These authors have successfully explained the observations of Gouy on the colour and polarisation of the light diffracted through large angles by metallic edges. The basic ideas of the Raman-Krishnan theory have been amply confirmed by the recent work of Savornin (1934).

3. Diffraction by Boundaries of Arbitrary Form

The ideas put forward in the preceding section regarding the Fresnel diffraction pattern of an infinite straight edge, *viz.*, that the illumination in the region of shadow is produced only by the boundary radiations, while the region of light, one has also to consider the effect of the direct light, should evidently be capable of being extended to boundaries of arbitrary form. Exactly as in the case of a straight edge, the fact that the light in the shadow has its origin in the boundary is readily observed in the general case also. For example, if we keep a circular disc in the path of a beam of light diverging from a pinhole, and look at it, keeping the eye at the centre of the geometric shadow, then the whole of the circular boundary appears luminous. But, if the eye is moved away, the bright portions reduce to two narrow regions on the boundary at the ends of a diameter, parallel to the line joining the eye to the centre. These become more and more different in intensity as the eye is moved farther away from the centre. Clearly therefore, in this case, the diffraction rings around the centre can be represented as the interferences of the radiations from those poles.

The formal recognition of the fact that the Fresnel pattern, even with apertures and obstacles of arbitrary form, arises from the boundary radiation enables us to explain the effects observed in special cases, such as the elliptical aperture, which was studied by Sir C. V. Raman (1919). He found that there is a concentration of luminosity on and within the evolute of the elliptic boundary between light and shadow. Further observations by Mitra (1919) have confirmed such a relationship between the form of the diffraction pattern and that of the boundary.

It is evident, however, that there are some special features requiring consideration in the general case. In the case of the straight edge, it is possible to consider the edge as a whole radiating a cylindrical wave-front. This, however, is not possible with a boundary of arbitrary form. Here, one has to consider each element of the boundary as a luminous source, which radiates spherical waves, and the effect at any point is the sum-total of the effects of all of them. The basic idea is, however, the same; namely that, in the region of shadow, the illumination is due only to the radiations from the various parts of the boundary, while, in the region of light, the effects are due both to the geometrically transmitted wave-front and to the boundary radiations. It is obvious that the phase of the wave diffracted to a particular point of observation will be determined by the total path it has to traverse from the source to this point, *via*, a point on the boundary. This will vary as the point on the boundary makes a circuit round it, so that the different diffracted waves reaching the point of observation will have varying phases. We have then to consider what the summation of the effects of all the portions of the boundary would be.

Suppose that the boundary is a smooth curve, without any discontinuities. Then, the rate of change of phase of the radiations received at the point of observation from the various points on the boundary would be continuous. Since, after a complete circuit round the boundary, the phase would have returned to its original value, there should be regions on the boundary for which the optical path would be either a maximum or a minimum, and the phase consequently stationary. Considering those regions where the phase varies continuously, as the emitting centre moves through a short distance along the boundary, the phase would have varied by a complete circle, and the resultant disturbance at the point of observation would be zero. Such a cancellation of the effects of neighbouring regions takes place as long as the rate of variation of phase is not too small. But near about those points where the phase is stationary, the radiations from the points nearby actually interfere constructively, and give rise to a large amplitude. After passing on the other side of this region, the neighbouring portions of the boundary once again begin to cancel one another until the next region of stationary phase is reached. Thus, the resultant disturbance at the point of observation is contributed only by a small number of regions of the boundary. These may be called "poles" on the boundary with respect to the point of observation. The positions of these poles may be geometrically defined to be those points on the boundary, whose projections on the plane of observation are the feet of the normals drawn to the projection of the boundary from the point of observation. This condition thus provides one with an easy method of determining the poles of a point of observation.

In general, these considerations do not give one the amplitude of the resultant radiation from the poles ; but this is not a serious difficulty, for the attempt is only to find the positions of the maxima and minima in the interference field, and not their absolute intensities. For the former, an idea of the phases alone will suffice, which can be obtained quite rigorously (see Section 5).

In special cases, however, considerations of amplitude do assume a certain importance, *e.g.*, when certain regions of the boundary are of constant curvature. In such a case, if the point of observation O happens to be at the centre of curvature of the projection of this region in the plane of observation, then the waves reaching O from the whole of this region would be in phase, so that the disturbance at the point, which may be called the "focus" of the region of constant curvature becomes very great. The bright central spot in the diffraction pattern of a circular obstacle arises from this cause. But, if the point of observation is removed away from the focus, the exact equality of phase of the radiations from a large portion of the boundary is no longer maintained, and the boundary radiation can be treated exactly as in the general case.

In fact, for every boundary, each region has a centre of curvature, so that it is natural to expect a small concentration of intensity along the locus of the centre of curvature of the geometric projection of the boundary. As is well known, this locus is the evolute of the projection, so that the evolute must appear bright, as is in fact observed (see above).

Another special case requiring consideration is that of boundaries having discontinuities or angular points. At such a point, the rate of change of phase of the boundary radiation undergoes abrupt changes, and the mutual cancellation of the disturbances produced by contiguous regions does not take place completely, so that the discontinuities act also as effective sources on the boundary in addition to the poles. However, while the poles shift about on the boundary as the point of observation is altered, the discontinuities always act as sources, irrespective of the position of the observation point.

The above is the case in the region of shadow, where the illumination is due only to the boundary radiation. But, in the region of light, one more complicating factor arises in that the incident light is also present. Consequently, the diffraction pattern has to be divided into two classes. The first, which may be called the " primary ", arises as a result of the interferences between the incident light, and the light diffracted by the boundary (which may be replaced by the light radiated by a small number of poles situated on it). The " secondary " pattern is produced by the interference of the radiations diffracted by the various poles on the boundary with themselves. It may be noted that this secondary pattern is of the same nature as the pattern in the region of shadow, being produced solely by the boundary radiation.

4. Relation between the Diffraction Patterns of Apertures and Obstacles of the Same Form

Consider two cases, which are complementary to each other, namely (A) an opaque obstacle of specified form, and (B) an opening of the same form in an opaque screen of great extension. In the region of shadow of the former, the disturbance is due only to the boundary radiations, which, as described in the previous section, produces the diffraction pattern. In the region of light of the latter (which obviously is the same as the region of shadow in the former case), we have both the primary and the secondary patterns. As explained above, the secondary pattern arises solely from the interference of the radiations from the boundary, so that one should expect this to be identical with the pattern in the shadow of the obstacle. *Vice versa*, the secondary pattern in the region of light of the obstacle should be identical with the pattern in the region of shadow of the aperture.

Another interesting result is obtained from a study of the relation between the diffraction patterns of obstacles and apertures of identical forms. Since the part of the incident light wave which is cut off in case (A) is that which is allowed to pass through in case (B), if the disturbances at any point in the two cases are imagined to be superposed, then one should get everywhere merely the undisturbed effect of the complete light wave, Consequently, if E_I is the light field due to the original source, E_{SA} and E_{SB} are the derived disturbances in the region of shadow in the cases (A) and (B) respectively, and E_{LA} and E_{LB} the corresponding disturbances in the region of light due to the boundary alone, the following relations hold :

$$E_I + E_{SA} + E_{LB} = E_I; \quad E_I + E_{SB} + E_{LA} = E_I,$$

so that $$E_{LB} = - E_{SA} \text{ and } E_{SB} = - E_{LA}.$$

From this, we see that the boundary radiation reaching the same point in the two cases are in opposite phases. Consequently, to the statement made above that the secondary pattern in the region of light of one is the same as the pattern in the region of shadow of the other has to be added the statement that the disturbances in the two cases are in opposite phases.

5. Mathematical Theory of Boundary Radiation

Formally, Rubinowicz (1917, 1923) has shown that it is possible to convert the surface integral, which has to be evaluated to determine the diffracted intensity due to an aperture, into a line integral over the boundary of the aperture. However, the final expression that is obtained is far too complicated, and its evaluation in any particular case (as for example, for an elliptic aperture) is perhaps as difficult as the evaluation of the corresponding surface integral. If one is concerned only with the effects at small angles, the so-called rigour in the treatment is superfluous. In fact, as Sir C. V. Raman has recently shown,* the transformation of the surface integral into a line integral can be performed in a very simple and elegant manner under these conditions. This method has been used by G. N. Ramachandran (1945) to obtain a mathematical theory of boundary radiation in a paper appearing in these *Proceedings*. We shall give here, for convenient reference, the main results obtained by him.

It is shown that with either an aperture or an obstacle, the illumination in the region of shadow can be completely represented as the effect of radiations arising from the boundary. On the other hand, in the region of light, the disturbance at any point is the sum of the disturbance due to the direct light, and that due to the boundary. The boundary radiation in the region of light has the opposite phase to that of the incident light, and has the same phase in the region of shadow. An important result is that the phase is determined by the region (of light or shadow) to which the ray towards the point of observation proceeds immediately from the boundary. Thus if a ray to a point in the region of shadow has to proceed through the region of

* Sayaji Memorial Lectures under publication.

light, it will have a phase opposite to the incident light, and *vice versa.* It is however shown that this leads to no discontinuity in the illumination as the point of observation passes from the region of light into the region of shadow.

As has been shown earlier the resultant effect of the radiation from the boundary may be considered as being due to the radiations arising from a finite number of point sources called poles, situated on the boundary. The question now arises whether the phase of the resultant radiation reaching the point of observation is the same as the phase of the light wave radiated by the pole itself. This has been considered by G. N. Ramachandran, and he finds that the resultant phase lags behind the radiations from the pole by $\pi/4$ if the pole is one of minimum path, and leads by $\pi/4$ if the pole is one of maximum path.

We shall outline in the next section a method for geometrically mapping out the diffraction pattern of an aperture or obstacle of arbitrary form.

6. Interference Pattern in the Region of Shadow

As we have seen in Section 3, the poles of a point of observation P can be geometrically defined as the feet of the normals to the diffracting boundary drawn from the point of intersection of the line joining the source and the point of observation P, with the plane of the diffracting screen. It is obvious from this that the lines joining P with the projection of these poles on the plane of observation are normal to the projection of the boundary on the same plane. This then provides one with a convenient method of determining the poles. It is also clear that for all points lying along a particular normal to the geometric shadow of the boundary on the plane of observation, the pole is the same, namely, that point on the boundary, whose projection is the foot of the normal. But the phase of the waves arriving from the pole at the different points on the normal will vary. Reversing the idea, one may attempt to draw curves, for points on which the phase is a constant ; but it would be found that the resulting geometric construction is too complicated. However, by a simple modification, namely, by drawing curves, not of equal phase, but of what may be termed *equal phase difference*, the difficulty can be overcome. The phase difference here conceived of is the difference in phase between the ray reaching the point *via* the pole, and the one that would reach it straight from the source if the obstacle were not present. It is obvious that, for points in the field of observation lying on the projection of the boundary, the two phases would be identical, and the phase difference would be zero. As one goes farther and farther away, either into the region of light or the region of shadow, the phase of the ray *via* the pole increases with respect to the direct ray, and the phase difference steadily increases. Since for points on the same normal, the pole is the same, one can mark out points on each normal for which the phase difference is π, 2π, 3π, etc. Through points having the same phase difference, one can draw smooth curves, and thus obtain the curves of equal phase difference.

In the case of a straight edge, the phase difference for a point of observation distant s from the geometric shadow of the edge is easily shown to be

$$\phi = \frac{\pi}{\lambda} \cdot \frac{as^2}{b(a+b)},$$

where a and b respectively are the distances from the plane of the diffracting boundary of the source and the plane of observation. This expression also holds good for a boundary of arbitrary form, if we mean by s the distance of the point from the foot of the normal through it to the projection of the boundary ; for the shape of the boundary for an infinitesimal distance on either side of the pole can be supposed to be a straight line. Consequently, the distance s_n corresponding to a phase difference of $n\pi$ is

$$s_n = \sqrt{n\lambda b(a+b)/a}.$$

Therefore, the lines of equal phase difference would be curves which are parallel to the form of the boundary, *i.e.*, the distances measured along any normal between two curves would be constant. If one numbers these curves sequentially as 1, 2, 3,.... corresponding to the values π, 2π, 3π,... of the phase difference, then the distance of the nth curve from the projection of the boundary would be s_n.

Having drawn these curves, it is a very simple matter to determine the form of the interference curves in the region of shadow. The bright lines in the pattern would be given by the curve drawn through those curves of equal phase diffeıence, the difference between whose sequential numbers is an even number. This is so because the difference between the values of the phase difference of two curves *at a point of intersection* is the difference between their actual phases, which is here $2m\pi$, where m is an integer, so that the two rays reaching the point from the corresponding poles would re-inforce each other, and produce a maximum intensity. On the other hand, the points of intersection of two sets of curves, whose sequential numbers differ by an odd number would evidently lie on an interference fringe of minimum intensity.

It may happen that not two, but more, sets of curves of equal phase difference intersect. In such a case, one has to take them in pairs, and draw the interference fringes for each pair. In this way, the whole interference pattern can be mapped.

The pattern becomes more complicated, if there are points of discontinuity on the boundary. In such a case, these points also act as centres of radiation, waves emitted by which will interfere with those radiated by the poles and produce an additional interference pattern. The lines of equal phase difference for these points are easily seen to be circles with the projection of the point as centre and radii given by the formula derived above.

The above statements are made on the supposition that the phase of the resultant radiation from a pole is the same as that of the light falling on it. This is not actually the case, as was remarked in the preceding section. In that case, the lines of equal phase difference have to be shifted suitably. This fact will be taken into account while considering particular cases in the succeeding papers.

7. Interference Pattern in the Region of Light

In the region of light, as already mentioned, there are both the direct radiation from the source as well as the diffracted radiation from the boundary, and two types of pattern, namely the primary and secondary, are produced. The primary pattern is due to the interference of the direct light with the radiations from the poles. Here, it is obvious that the most intense effects would be produced by those poles which are nearest to the point of observation. Also, since the lines of equal phase difference represent the difference in phase between the direct and the diffracted rays, these lines must coincide with the primary pattern in the region of light, those with odd numbers being dark fringes, and those with even numbers being bright. Further, these fringes will run parallel to the boundary of the diffracting figure in the region of light, since the lines of equal path difference do the same. Of course, this is true only if the shape of the boundary is continuous, without any singularities.

The secondary interferences are of the same nature as the pattern within the region of shadow, and can be mapped out in precisely the same manner. There is also a third, though minor, feature that can be observed in the region of light. This is the interference pattern due to points of discontinuity acting as centres of radiation, which can also be drawn in a manner similar to that in the region of shadow.

8. Transition from the Fresnel to the Fraunhofer Pattern

The application of the boundary idea to the interpretation of Fraunhofer patterns has already been pointed out by several writers.

Mitra has worked out the case of the semi-circular aperture in detail. Other applications will be found in the lectures by Sir C. V. Raman quoted above. A rigorous theory has been given by Laue (1936). But the new point which has been developed in this paper is that this technique can equally well be applied to Fresnel patterns. This procedure enables one to appreciate the relation between the two, and also to understand the transition from the Fresnel to the Fraunhofer class.

In Fraunhofer diffraction, the light from the source is focussed by means of a lens on the observation screen, so that the geometric shadow of the diffracting screen is reduced to a point, and the entire pattern is due to the boundary waves only. In the Fresnel pattern, on the other hand, the geometric shadow of the aperture is finite, and within this region, we have to consider not only the interferences of the boundary radiations with themselves, but also with the incident light. There is no strict analogue of this in the Fraunhofer pattern. In the region of shadow of the Fresnel pattern no direct light falls and the effects are produced only by the boundary radiation : but even here the relationship of phase is not exactly the same as in the Fraunhofer pattern, because the angle of diffraction is not the same for all points on the boundary, and also because variations of phase due to the finite distances are involved.

It is fairly obvious that the differences between the two types of patterns can be diminished by one of three procedures, *viz.*, increasing the angle of diffraction, increasing the distance of the source and the screen from the aperture, and diminishing the size of the aperture. If we consider the diffraction at large angles, it is seen that the small variations in phase caused by the finite distances of the source and the point of observation are overwhelmed by the large variations caused by the angle of diffraction, so that the pattern at these angles resembles the Fraunhofer pattern very closely. The same thing can be accomplished by reducing the size of the aperture, without altering its shape. Since the variations in phase due to the first cause are of the order of the square of the dimensions of the aperture, these decrease very much, so that the resemblance to the Fraunhofer pattern can be observed at smaller angles. So also, by increasing the distances of the source and the observation point from the diffracting screen, the variations in phase due to these finite distances can be decreased, and the resemblance to the Fraunhofer pattern brought still closer to the centre of the field. However, very close to the centre, particularly in the region of light, the pattern in the Fresnel case will be different from that in the other.

It is a well known fact that the Fraunhofer pattern always exhibits a centre of symmetry, irrespective of the shape of the aperture. Consequently, the above phenomena can also be described as follows. Starting with a large size for the aperture, if one goes on reducing the size, the outer regions of the Fresnel diffraction pattern progressively develop more and more of symmetry, the asymmetric portion contracting towards the centre.

The author wishes to express his indebtedness to Prof. Sir C. V. Raman, for suggesting the problem and for kind interest in the course of this investigation. His thanks are also due to Dr. R. S. Krishnan for helpful suggestions, and to Mr. G. N. Ramachandran with whom he had many useful discussions.

SUMMARY

The paper deals with a new approach to the subject of diffraction which enables the nature of the Fresnel pattern due to an obstacle or aperture of arbitrary form to be easily derived. Taking the hint suggested by observation, it is assumed that in the region of shadow the pattern is produced by the interference of radiations having their origin in the boundary of the aperture or the obstacle. In the region of light, the boundary radiation interferes with the primary incident waves and produces the fluctuations of intensity. The radiation from the boundary can again be effectively replaced by spherical waves originating from a finite number of point-sources on the boundary, called ' poles ', at which the path to the point of observation from the boundary is a maximum or a minimum. This immediately gives a geometric definition of the poles as those points on the boundary whose projection on the plane of observation are the feet of the normals from the observation point to the projection of the boundary. Using this result, it is possible to geometrically map out the positions of maximum and minimum intensity in the diffraction pattern of an arbitrary aperture or obstacle. The relation between the Fresnel and Fraunhofer patterns of an aperture is discussed and it is shown that the Fresnel pattern approaches more and more to the other type as the size of the aperture is diminished or the distance to the observation screen is increased.

REFERENCES

1. Gouy .. *Ann. d. Phys. et de Chem.*, 1886 **(6)**, **8**, 145.
2. Laue, M. von .. *Sitzungsber. Preuss. Akad.*, 1936, 89.
3. Mitra, S. K. .. *Phil. Mag.*. 1919, **38**, 289.
4. Ramachandran, G. N. .. *Proc. Ind. Acad. Sci.*, 1945, this issue.
5. Raman, Sir C. V. .. *Phys. Rev.*, 1919, **13**, 259.
6. Raman, Sir C. V., and Krishnan, K. S. .. *Proc. Roy. Soc.*, A, 1927, **116**, 254.
7. Rubinowicz, A. .. *Ann. d. Phys.*, 1917, **57**, 257.
8. ———— .. *Ibid.*, 1923, **73**, 339.
9. Savornin .. *Compt. Rend.*, 1934. **198**, 647, 941.