Geometrical theory of diffraction—A historical perspective

P. B. Sunil Kumar and G. S. Ranganath

The geometrical theory of diffraction is a very convenient and easy method of calculating diffraction patterns, and an elegant approach to the problems of Fresnel diffraction at apertures and obstacles. In spite of its long and chequered history, it has not found sufficient emphasis in the standard literature on optical diffraction. Between Young (1802) and Keller (1962), the Indian school (1917–45) led by Raman was active in the field.

In recent years, there has been a revival of interest in the so-called geometrical theory of the diffraction of light[1–3], consequent on the systematic work of Keller[4], who in the late 1950s developed the detailed methodology for this approach. This concept of treating optical diffraction using geometrical ideas was first introduced by Thomas Young in 1802. In the intervening years, much progress was made and many salient features of the recent ideas in geometric theory were anticipated during the three decades ending in 1945 by the active school led by C. V. Raman. Unfortunately, this work has gone unnoticed in the subsequent, modern literature. It is our intention in this article to give a connected account of the development of the geometric theory bridging the gap between Young and Keller by presenting the work of Gouy, Sommerfeld, Rubinowicz and Raman.

The Helmholtz–Kirchhoff theory of scalar waves presents many computational difficulties in the theoretical calculation of a general diffraction pattern. The procedures become inordinately complex even in the cases of apertures and obstacles having standard geometrical shapes. Over the years, attempts at improving this technique have met with very limited success. It is in this context that the oldest and perhaps the simplest theory, viz. the geometrical theory becomes relevant.

Diffraction at edges and apertures

Thomas Young[5] was the first to propose that when light falls on a straight edge, the edge 'reflects' the light into space and the associated interference between the 'edge wave' and the geometrically transmitted wave gives rise to the observed diffraction effect. Gouy[6] in 1886 gave reality to Young's edge waves when he observed that the sharp metallic edge held in a pencil of light appears luminous and the strongly polarized light is diffracted through large angles. Maggi[7] later elaborated Young's model and showed mathematically that the diffraction integral over an aperture can be reduced to a line integral on the boundary of the aperture and a contribution due to the geometrically transmitted light. Sommerfeld[8], who was apparently unaware of this work, independently solved the problem exactly for a straight edge. This theory of electromagnetic diffraction at a straight edge made of perfectly conducting material leads to an interesting result. The field at any point can be looked upon as a sum of the transmitted wave and the wave that appears to emanate from the edge. This edge wave is given by the asymptotic formula

$$u(r, \phi) = v(r, \phi - \phi_o) \pm v(r, \phi + \phi_o), \tag{1}$$

where

$$v(r, \theta) = \left[\frac{1+i}{4\sqrt{k}}\right]\left[\frac{e^{ikr}}{\sqrt{r}}\right]\left[\frac{1}{\cos\theta/2}\right].$$

The + or − sign is taken according as the electric vector is parallel or perpendicular to the edge and ϕ_o is the angle of the incident ray and ϕ that of the diffracted ray as measured from the plane of the diffracting screen (see Figure 1). Along the shadow boundary v diverges since $\theta = \pi$. (The so-called uniform geometrical theory of diffraction overcomes this lacuna[3], but we do not discuss it further in this article.)

The geometrical theory resolved, for the first time, the apparent puzzle associated with the concept of edge

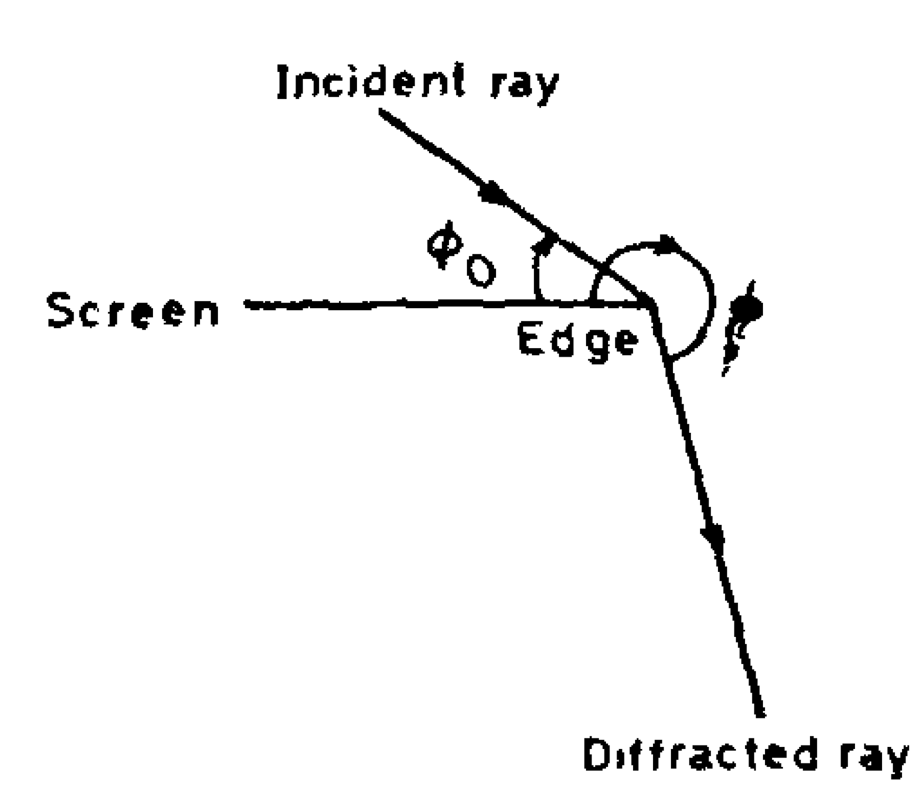

Figure 1. Diffraction geometry: parallel rays incident on a straight edge at a glancing angle of ϕ_o.

The authors are in the Raman Research Institute, Bangalore 560 080.

diffraction, that the intensity of the bright fringe remains almost constant as the observational plane recedes from the edge. Although the edge wave is cylindrical, its amplitude is dependent on the angle of diffraction. The amplitude increases as the shadow boundary is approached (by a factor sec $\theta/2$) and the $1/\sqrt{r}$ law for amplitude diminution exactly compensates this.

Sommerfeld's theory also agrees with the Fresnel scalar theory of diffraction for a straight edge. But, it fails to account for the observed diffraction pattern by metallic edges. Raman and Krishnan[9] pointed out that this failure is due to the assumption that the material is perfectly conducting. Instead, by incorporating the complex metallic reflection coefficient in the second term of eq. (1) these authors neatly accounted for the experimentally observed features[9,10].

Rubinowicz[11] many years later rediscovered Maggi's result that the Kirchhoff's surface integral over the diffracting aperture in the limit of short wavelengths for an incident spherical wave could be reduced to a line integral over the aperture. He also obtained the result that the diffracted field at any point is made up of two components—(i) the familiar geometrical optical field, and (ii) a wave emitted by the boundary of the aperture. Laue[12] showed that in Fraunhofer diffraction also, a similar transformation from surface integral to line integral along the boundary is possible. Raman[13] showed the integral transformation to be a far simpler procedure if one makes the justifiable approximation of ignoring the obliquity factor. It must be remarked that all these procedures are valid only when the size of the diffracting object is large compared to the wavelength of light.

Another important aspect of Rubinowicz's work is that the contour integral can be reduced by the stationary phase method to contributions from a *finite number of points* on the boundary whose locations depend on the point of observation in the diffraction field. Further, these special points for normal incidence can be easily obtained by drawing perpendiculars to the diffracting boundary from the point of observation. At these points, the incident light ray and the diffracted ray reaching the point of observation satisfy a *reflection condition*. According to this, when the incident rays are parallel and normal to the diffracting edge, the diffracted light rays reaching any given point of interest are also normal to the edge. Clearly, this answer yields the cylindrical boundary waves for straight edges as shown in Figure 2,*b*. The feet of the perpendiculars mentioned earlier are the special points which seem to be the source of radiation. Based on Raman's model, Ramachandran[14,15] also obtained the same result.

From a given point of observation only these points—poles—should be visible and this was experimentally demonstrated by Raman[13,16], who showed that only a finite number of luminous points are visible on the boundary when viewed from the shadow region. For these special points the total optical path from the source to the point of observation via these points is an extremum. This principle is very reminiscent of the well known Fermat's principle in geometrical optics. For this reason, it has been referred to as Fermat's principle for edge diffraction by Keller. When the incident light rays are parallel but are incident on the edge at an angle then Fermat's principle of diffraction will result in diffracted rays travelling on a cone symmetrical about the local tangent to the edge. Thus, one gets diffraction wavefronts to be parallel cones with the edge as their common axis. This has been depicted in Figure 2,*a*.

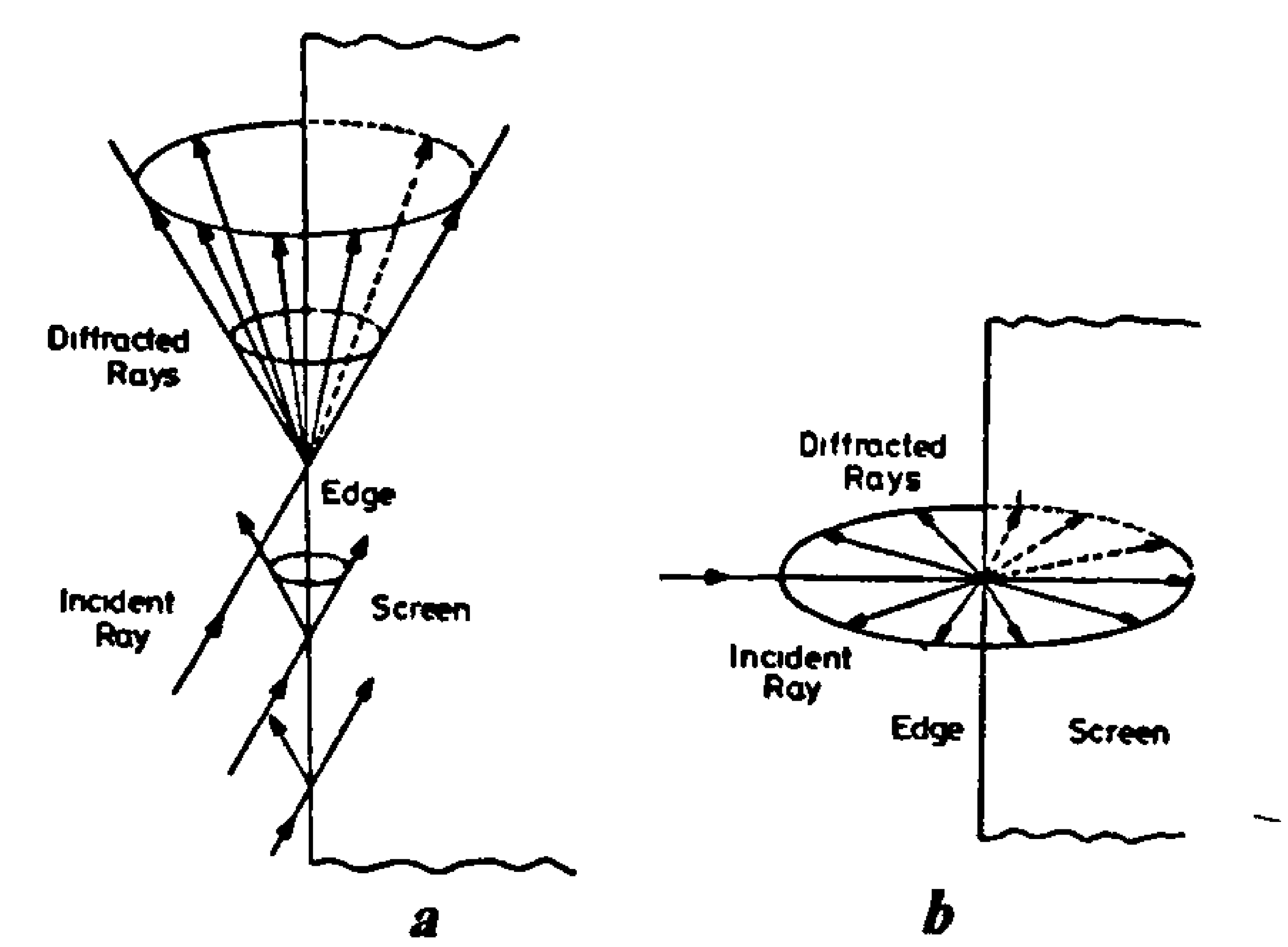

Figure 2. Diffracted rays can be obtained from the Rubinowicz 'reflection' condition. *a*, Diffracted rays will be on a cone symmetric about the edge for an oblique incidence. *b*, Diffracted rays will be on a disc perpendicular to the edge for normal incidence.

Kathavate[17] stated that, when dealing with sharp corners of apertures and obstacles, the sharp corners should be taken as additional point sources of light emitting spherical waves. These are in addition to the poles already considered. A decade later Keller[4] also suggested the same procedure. However, these workers did not work out the diffraction coefficient. Independently, around the same year, Miyamoto and Wolf[18] not only came to the same conclusion but also worked out the corner-diffraction coefficient.

Diffraction within the shadow

The geometrical theory of diffraction clearly indicates that the shadow region of an obstacle gets light only from the edge wave. These edge waves will have to be added at any point within the shadow to get the net optical field there. From what has been said in the previous section it follows that we need to take only two types of contributions: (i) from the poles obtained with respect to the point of observation, and (ii) from the corners. The question of the phase of the radiation from the pole was considered by Ramachandran[14]. He showed using the Cornu spiral method that the radiation received at the point of observation from the regions neighbouring the special points (poles) resulted in a phase advance or a phase lag of $\pi/4$ depending upon whether it is one of maximum or minimum optical path from the source to the point of observation via the edge. The same result was obtained many years later by Miyamoto and Wolf[18].

Surface diffraction

So far we have considered diffraction only at sharp edges. But in reality edges are never perfectly sharp, but are rounded. An extreme example of this was considered by Raman and Krishnan[19]—the Fresnel diffraction by a spherical object. One might naively regard this as equivalent to diffraction by a circular disc. However, in the case of diffraction around metallic spheres, they noticed that the intensity of the central bright spot is always lower than what one observes for a circular disc of equal diameter. Also, the intensity of the central spot is found to be a very sensitive function of the distance from the centre of the sphere. In fact, it exponentially decays below the intensity of the disc spot as the point of observation approaches the sphere. They accounted for this by suggesting that light actually creeps over the spherical surface and the light reaching any point of observation emanates from the circular boundary along the tangent cone drawn to the sphere from the point of observation. They used the exponential law derived by Riemann–Weber[20] for electromagnetic wave propagation around the earth and got a beautiful fit with experimentally observed data.

The more recent work[3] on the geometrical theory of diffraction at smooth surfaces is based on essentially the same mechanism. The detailed theory gives a series expansion for the attenuation coefficient which turns out to be different for the electric vector parallel or perpendicular to the surface. Raman and Krishnan's theory has only the leading term of this series. This is sufficient to account for the experimental data. But one feature which is important in this process of creeping is that the attenuation coefficient is a complex number. Hence when the waves interfere after creeping they have additional phase differences over and above that due to the actual path travelled by light. Keller invokes the generalized Fermat's principle, whereby the actual path from the source to the observer via the surface should be an extremum. This is only possible when light 'creeps' on the surface, travelling along a geodesic on the surface. In fact, for oblique incidence on a cylindrical surface the light creeps along a helix.

Implications of the theory

Diffraction caustics

An important implication of the geometrical theory of diffraction was stressed by Raman[16] as early as 1919. He argued that for normal incidence the diffracted 'rays' will be predominantly proceeding in the direction of the local normals to the edge of the aperture. Hence, there will be a concentration of light along the evolute (the envelope of the normals to a given curve is defined as its evolute) to the diffracting boundary. Raman also demonstrated this experimentally and called these the diffraction caustics (Figure 3). Of course, the diffraction caustic degenerates into a point in the case of a circular disc, leading to the familiar Poisson spot. A few years later, Coulson and Becknell[21] (for a disc) and Nienhuis[22] (for an aperture) did similar experiments and obtained the same results.

Slits and gratings

In the case of multiple straight edges as in a slit or an array of slits, the standard procedure is to employ the Fresnel integral or Cornu spiral to work out the diffraction pattern. In the geometrical theory of diffraction, as Raman[13] showed, the diffraction pattern can be obtained by adding the various cylindrical edge waves. This model leads to the well-known answers for a slit or a grating in the Fraunhofer diffraction limit. Keller's recipe to deal with these situations is also essentially the same.

Semitrasparent edge

Anathanarayanan[23] studied the diffraction at straight edges of thin films of metals coated on glass. When the metallic coatings were thin enough, he saw fringes of high visibility in the shadow region behind the metallic film. But when the film was thick, he observed in this region, the familiar gradual decay in intensity. He explained this fringe system in the shadow region as a consequence of the interference between the cylindrical edge wave and the wave weakly transmitted by the thin metallic film. When both these waves are of nearly comparable amplitudes the fringe system had a high visibility or contrast.

Figure 3. Diffraction caustic of an elliptical aperture. (After Raman[16])

Fraunhofer diffraction

A special mention may be made of Raman's[13] studies on Fraunhofer diffraction by triangular and semi-circular apertures. Here, the boundary of the diffracting object is replaced by a set of points. Fraunhofer diffraction of an equilateral triangular aperture has a six-fold symmetry and is obtainable from an interference of radiation from three point sources placed at the vertices of the triangle. In the case of a semicircular aperture, Raman[13] argued that in effect we can replace the boundary by three points. One lies on the curved edge its position given by the foot of the perpendicular from the point of observation to the curved edge, and two more respectively at the two corners. This leads to the observed higher symmetry in the Fraunhofer pattern than that of the object.

Again in all their studies on apertures, Raman's school made a special experimental study of pattern transformation as one went from the Fresnel diffraction limit to the Fraunhofer diffraction limit. This is important in view of the fact that Fraunhofer diffraction is centrosymmetric while Fresnel diffraction is not. In Figure 4 we have shown this phenomenon.

Shadow patterns

On the experimental side, Kathavate[17], using objects of a few millimetres and 10 to 50 hours of exposure, got beautiful and intricate diffraction patterns in the case of apertures and discs of various shapes. He came up with a simple and an elegant geometrical procedure, based on the geometrical theory of diffraction, to work out the positions of diffraction maxima (or minima). The whole geometrical construction is carried out on the plane of observation on to which we project the obstacle and the rays from the special points. It is easy to convince oneself that to get the projection of the poles, we just draw perpendiculars to the boundary of the shadow from the point observation. Light from the feet of these perpendiculars must be considered while calculating the positions of maxima or minima. In Figure 5, we show his theoretical calculation along with the observed diffraction pattern for the square disc. This work appears to have gone unnoticed in the literature. Recently English and George[24] have reported the same result. The shadow patterns from elliptic discs of different eccentricities are shown in Figure 6.

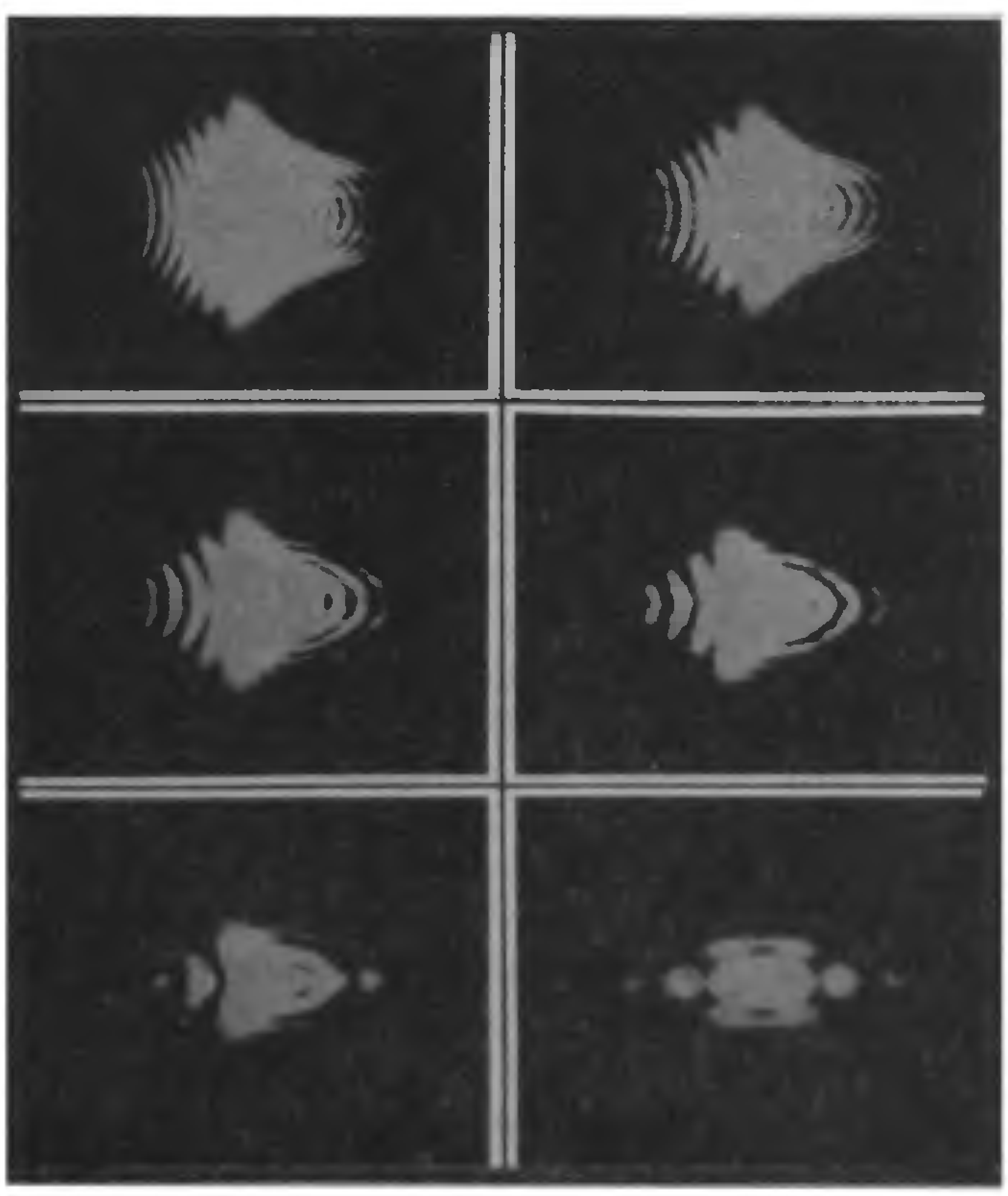

Figure 4. Transition from Fresnel to Fraunhofer diffraction pattern in the case of a semicircular aperture. (After Kathavate[17])

Some new results

The Poisson spot

The bright central Poisson spot in the case of a circular disc is the brightest region of the diffraction pattern in the shadow and it is due to the constructive interference of radiation from the entire boundary. At other points of observation we have only two boundary waves emanating from diametrically opposite poles, leading to periodic weak maxima (and minima). Even in the case of other obstacles like elliptic, square, triangular and rectangular discs, we get such a central spot. It is easy to work out the features associated with this Poisson spot in the language of the geometrical theory of diffraction. For example, in the case of an elliptic disc, the centre of the pattern gets light from four poles—two poles at the ends of the major axis and two poles at the ends of the minor axis of the ellipse. Radiations from the poles of the major axis (or minor axis) are in phase at the centre. But radiations from a pole associated with the major axis may not always be in phase with the radiations from a pole associated with the minor axis. In fact, as we recede from the diffracting plane along the central axis these pole radiations will be successively in and out of phase, giving rise to a brightness fluctuation in the Poisson spot. For a square obstacle, the fluctuations are due to the pole and the corner radiation being in and out of phase. Since corner radiations are weaker as their intensity falls as $1/r^2$, these flutuations will not be prominent.

The Poisson spot associated with an elliptic disc is, in many ways, different from the one associated with a rectangular obstacle whose length and breadth are respectively equal to the major and minor diameters. If we ignore the corner radiation, then we get four poles

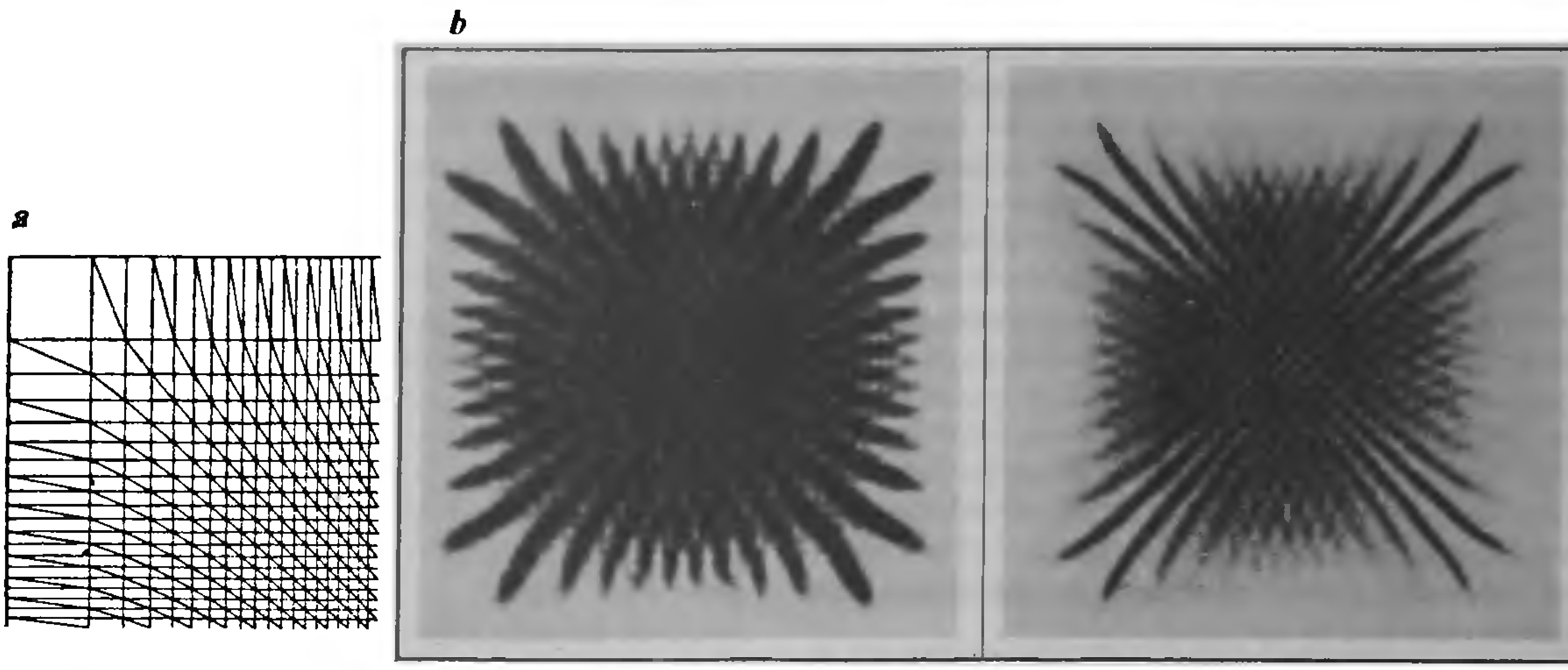

Figure 5. ***a***, Positions of intensity maxima for a 90° sector as obtained from geometrical theory of diffraction. ***b***, Diffraction pattern for a square disc: monochromatic light (left) and white light (right). (After Kathavate[17])

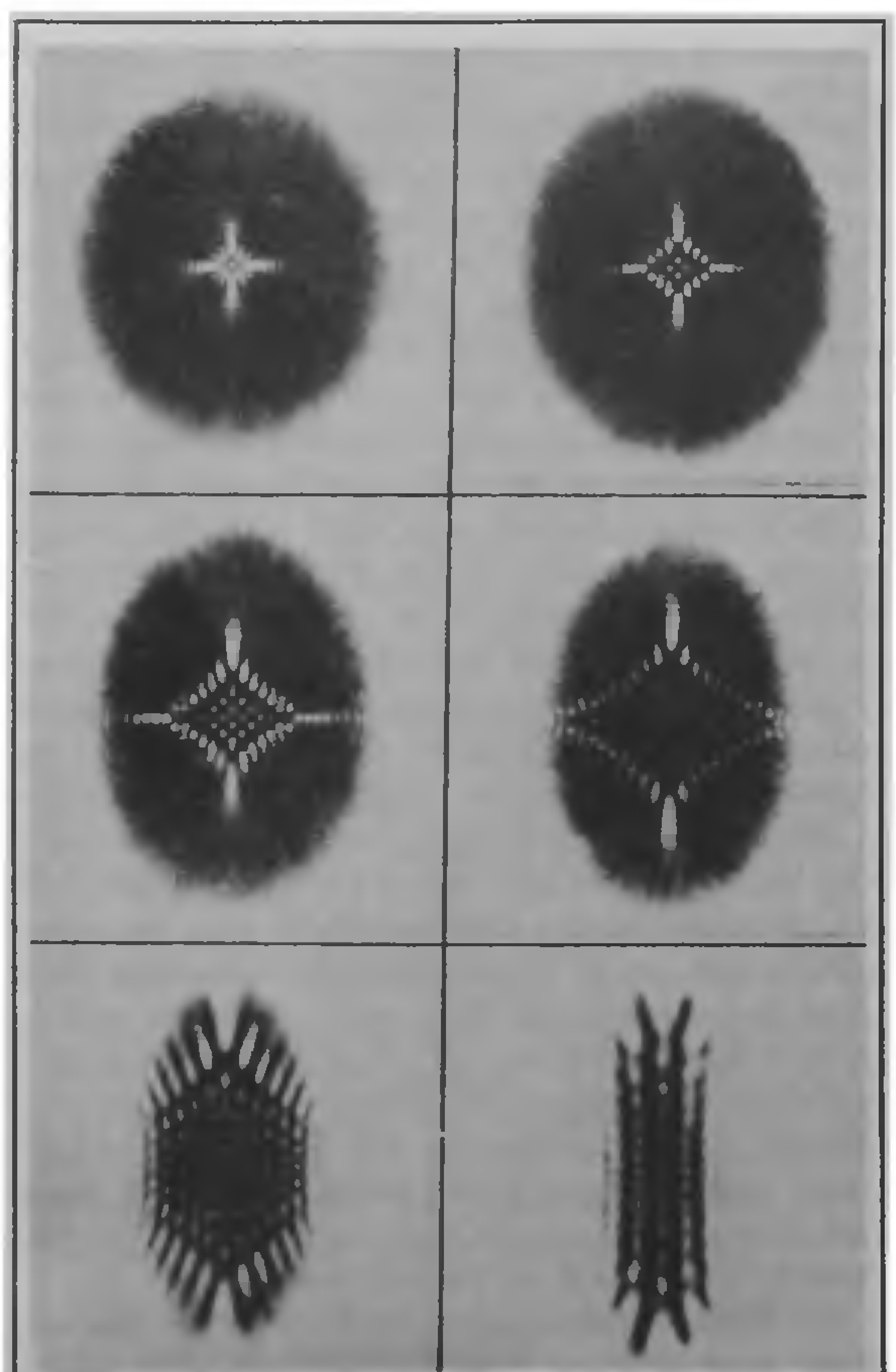

Figure 6. Diffraction pattern within the shadow of an elliptic disc. The picture shows the effect of increasing eccentricity. (After Kathavate[17])

as in an elliptic disc. Yet the net intensity at the Poisson spot will be different for the elliptic disc due to the curvature at the boundary. This arises owing to the focusing effect of a curved wavefront emitted by a curved boundary. In fact, Keller[4] shows from the geometrical theory of diffraction that a curved boundary contributes more than a straight boundary when the point of observation is towards the centre of curvature. Hence poles of the rectangle make a weaker contribution to the Poisson spot than the poles of an ellipse of 'equal' size. To our knowledge these interesting consequences of geometrical theory of diffraction have not been emphasized in the literature.

Diffraction at a strip and at a cylinder

In the shadow region at a finite distance from an opaque strip or a cylinder, experimentally one observes a fringe system. Careful investigations show that this fringe system is strictly not a set of equidistant bright and dark fringes. Nor is the visibility of the fringe system the same all over. In the language of the geometrical theory of diffraction the fringe pattern in the two situations is due to entirely different processes. For a strip it is the interference between the two cylindrical waves from the two straight edges. On the other hand, to reach any point in the shadow of a cylinder, light will have to creep from both sides along the boundary. Thus the interference pattern in general will be different in the two cases. The same arguments are valid even in the case of diffraction at a circular disc and a sphere. However, calculations of the diffraction pattern are easy for the case of a strip and cylinder. At extremely large distances there is very little creeping and the two patterns can be expected to be nearly the

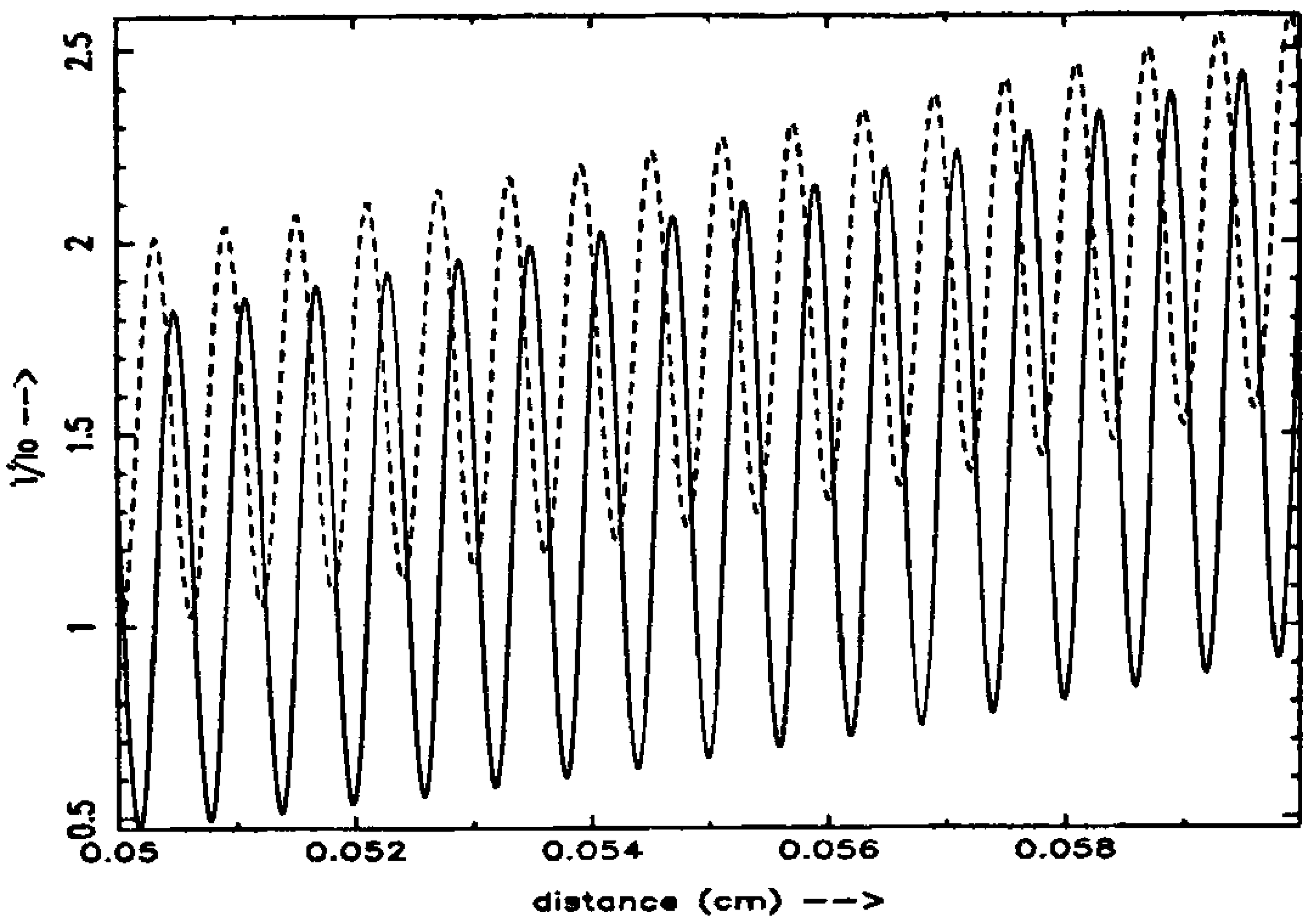

Figure 7. Calculated intensity pattern in the shadow region of a strip (full line) and a cylinder of equal width (dashed line).

same. In Figure 7 we have compared the calculated diffraction pattern due to a strip and that due to a cylinder of diameter equal to the width of the strip.

Diffraction symmetry

We have already touched upon the symmetry of a diffraction pattern in relation to aperture symmetry. Diffraction symmetries are also strongly influenced by polarization. Implications of the geometrical theory of diffraction in this regard will be briefly considered here. In the case of scalar wave diffraction, a circle or a sphere gives rise to a pattern with circular symmetry while a square yields a pattern with a four-fold symmetry. In polarized light, however, we can come to some interesting conclusions concerning these symmetries. Let us say that a linearly polarized wave is incident normally on a square aperture [or an obstacle] with its electric vector parallel to one pair of edges. Then from eq. (1) we conclude that the cylindrical waves emitted by this pair of edges are not identical to the ones emitted by the other pair, since for this second pair the electric vector is normal to the diffracting edge. Thus the diffraction pattern will not have four-fold symmetry but a two-fold symmetry. However, when the electric vector is parallel to the diagonal of the square the pattern will have a four-fold symmetry. Only when the incident vibration is circularly polarized do we have the symmetry obtained for scalar waves. Calculations based on the Keller's theory indicate that these features are noticeable only at a short distance from the screen. At larger distances, i.e. in the paraxial approximation, this symmetry is lost owing to smallness of the second term in eq. (1).

Multiple-edge radiations

In another respect Keller improved the geometrical theory of diffraction. We shall illustrate Keller's correction with the example of single-slit diffraction. It was argued earlier that in this case we have two cylindrical waves diverging from the two edges. A wave from one such edge will reach the other edge and will result in a second cylindrical wave. This process could go on endlessly, indicating that each edge gives rise to a multiplicity of edge rays. These have been termed by Keller as second, third, etc. diffracted edge rays. In principle, a complete solution must include the effect of these multiply diffracted rays. However, in practice, these extra effects do not appear to be all that significant, since the strength of the diffracted ray decreases considerably with increasing order.

Acknowledgements. It is pleasure to acknowledge Prof. S. Ramaseshan for introducing us to the geometrical theory of diffraction and for many of his suggestions and hints in our discussions with him. We also thank Prof. Rajaram Nityananda for useful comments.

1. Jones, D. S., *Methods in Electromagnetic Wave Propagation*, Oxford, 1979.
2. James, G. L., *Geometrical Theory of Diffraction for Electromagnetic Waves*, Peter Peregrinus, 1980.
3. McNamara, D. A., Pistorius C. W. and Malherbe, J. A. G., *Introduction to the Uniform Geometrical Theory of Diffraction*, Artech House, Boston, London, 1990.
4. Keller, J. B., *J. Opt. Soc. Am.*, 1962, **52**, 116. and references therein.
5. T. Young, *Philos. Trans. R. Soc. (London)*, 1802, **20**, 26.
6. Gouy, L. G., *Ann. Chim. Phys.*, 1886, **8**, 145.
7. Maggi, G. A., *Annali di Matem IIA*, 1888, **16**, 21.
8. Sommerfeld, A., *Math. Ann.*, 1986, **47**, 314.
9. Raman, C. V. and Krishnan, K. S., *Proc. R. Soc. (London)*. 1927, A116, 254.
10. Savornin, J., *Ann. Phys.*, 1939, **11**, 129.
11. Rubinowicz, A., *Ann. Phys.*, 1917, **53**, 257; 1924, **73**, 339; see also Born, M., and Wolf, E., *Principles of Optics*, Pergamon, 1964.
12. Laue, Von, M., *Berliner Sitzungsberichte*, 1936, 89; see also Jame, R. W., *The Optical Principles of the Diffraction of X-rays*, G. Bell and Sons Ltd, 1967.
13. Raman, C. V., *Sayaji Rao Gaekwar Foundation Lectures on Physical Optics*; The Indian Academy Sciences, Bangalore, 1959
14. Ramachandran, G. N., *Proc. Indian Acad. Sci.*, 1944 **A19**, 123.
15. Ramachandran, G. N., *Proc. Indian Acad. Sci.*, 1945, **A21**, 165.
16. Raman, C. V., *Phys. Rev.*, 1919, **13**, 259.
17. Kathavate, Y. V., *Proc. Indian Acad. Sci.*, 1945, **A21**, 177 (in six parts).
18. Miyamoto, K. and Wolf, E., *J. Opt. Soc. Am.*, 1962, **52**, 615 and 626.
19. Raman, C. V. and Krishnan, K. S., *Proc. Phys. Soc. (London)*, 1926, **38**, 350.
20. Riemann Weber, *Differential-Gleichung Phys.*, 1927, **2**, 594.
21. Coulson, J. and Becknell, C. G., *Phys. Rev.*, 1922, **20**, 594.
22. Nienhuis, K., Thesis, Groningen, 1948.
23. Ananthanarayanan, N. *Proc. Indian Acad Sci.*, 1940, **A10**, 477.
24. English, R. E. and George, N. *Appl. Opt.*, 1988, **27**, 1581.