Phase Distribution near Focus in an Aberration-free Diffraction Image

By E. H. LINFOOT† and E. WOLF‡

†The Observatories, University of Cambridge

‡The Physical Laboratories University of Manchester

MS. received 12*th October* 1955, *and in revised form* 24*th February* 1956

Abstract. A satisfactory picture of the three-dimensional intensity distribution near focus in the aberration-free diffraction image of a monochromatic point object was first given by Zernike and Nijboer in 1949. No corresponding general picture of the phase relations between different parts of the image seems to have been worked out, although Gouy's discovery in 1890 of the so-called phase anomaly at focus aroused an interest in this topic which, as a succession of publications shows, has not decreased in the intervening 66 years.

In the present paper a sharpened version of Lommel's classical analysis is applied to obtain a general picture of the phase distribution near focus and to examine in detail, in the special case of an F/3·5 pencil, its peculiarities near the geometrical focal point, the Airy dark rings and the axial nodes (points of zero intensity) of the diffraction image. The 'phase anomaly' near focus and the singular behaviour of the phase along the axis of the pencil become more readily intelligible when they are considered against the background of this general picture.

§ 1. Introduction

A central problem in the theory of image formation in optical instruments is the determination of the three-dimensional light distribution which represents the aberration-free image of a point source by an axially symmetric optical system. Analytical formulae for the intensity distribution in such an image were first given by von Lommel (1885) and almost simultaneously by Struve (1886). The first satisfactory diagram of the intensity distribution was published by Zernike and Nijboer (1949) more than sixty years later,† and similar diagrams for images affected by selected amounts of spherical aberration have been worked out by them (1949) and by A. Maréchal and his co-workers (1948).

Less progress seems to have been made in the complementary part of the problem, namely the study of the phase distribution in the three-dimensional diffraction image, even though modern developments in microwave optics have greatly enhanced its practical interest. Theoretical interest was never lacking; as early as 1890 Gouy discovered the so-called phase anomaly near focus, and the subject has since been treated by many authors‡, notably by Joubin (1892), Fabry (1893), Julius (1895), Zeeman (1897, 1900, 1901), Sagnac (1903, 1904 a, b), Debye (1909), Reiche (1909 a, b), Ignatowsky (1919), Fokker (1923, 1924), Picht (1930), Rubinowicz (1938), Breuninger (1938, 1939), Bouwkamp (1940) and Toraldo di Francia (1942).

† An inaccurate diagram, widely reproduced in optical textbooks, was given by Berek (1926). The isophotes in figure 2 were constructed from tables given by von Lommel (1885).

‡ For a review of the early papers, see Reiche (1909 a).

All these investigations were concerned with the aberration-free image and their joint outcome was, broadly speaking, to confirm Gouy's results, to determine the detailed behaviour of the phase distribution along particular rays through the geometrical focus, and to establish that the properties of the phase distribution along the axial ray are qualitatively different from those elsewhere. The last result seems to have been first established by Picht (1930) However, a satisfactory overall picture of the three-dimensional phase distribution, as distinct from the phase distribution along isolated rays, did not emerge.

Such a picture seems nowadays hardly less essential to a proper understanding of the diffraction image than that obtained by Zernike and Nijboer for the three-dimensional intensity distribution. In the present paper we go some way towards obtaining one by an application of Lommel's classical formulae which, for pencils of not too large numerical aperture, allow the calculation of the phases at points not too far from the geometrical focus. From the computed values of the phases at a sufficiently dense set of sample points, the phase distribution near the geometrical focus has been obtained by graphical methods.

It turns out that in the geometrical pencil of rays near focus the co-phasal surfaces are very nearly plane; the corresponding phenomenon for sound waves can be seen in schlieren photographs given by Pohl (1948). The singular behaviour of the phase along the axis becomes readily intelligible when its connection with the existence of points of zero intensity is made visible to the eye (see figures 5 and 6). The singular behaviour of the phase at those points of the geometrical focal plane which corresponds to the Airy dark rings becomes intelligible in the same way (figures 3 and 4).

§ 2. Phase near Focus

The notation and approximations are similar to those of our paper (1952) on telescopic star images. In figure 1, ABA'B' represents a circular aperture

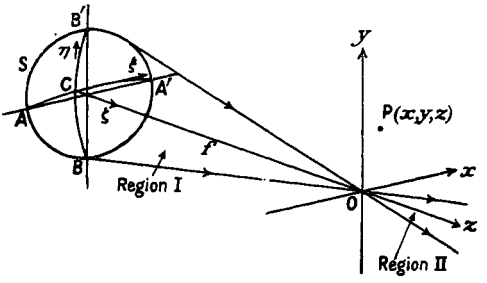

Figure 1.

through which issues a train of converging spherical waves of wavelength λ. The diameter AA' of this aperture is of length $2a$; C is the pole of the wave surface S which momentarily fills it. O is the centre of curvature of S, CO$=f$ its radius of curvature. It is assumed throughout that $a^2 \ll f^2$; for example, $a^2/f^2 = 1/49$ in an F/3·5 pencil. P (x, y, z) is an arbitrary point in the space near O.

We define the variables u, v by the equations

$$u = ka^2z/f^2, \qquad v = kar/f, \qquad \ldots\ldots (2.1)$$

where $k = 2\pi/\lambda$ and r stands for $\sqrt{(x^2+y^2)}$. In physical terms, $u/4\pi$ is the

number of fringes of defocusing and v/π the number of fringes of lateral displacement of P relative to O. We note that $|v/u| \lesssim 1$ according as P lies in the geometrical cone of rays or in the geometrical shadow.

The complex displacement† at P(x, y, z) in the space near O which results from waves of unit amplitude in S is given by the approximate formula

$$D_\lambda(\mathrm{P}) = \frac{2\pi i}{\lambda} a^2 \exp[ik(f-R')] \int_0^1 \exp(\tfrac{1}{2}iu\rho^2) \mathrm{J}_0(v\rho)\rho \, d\rho, \qquad \ldots\ldots (2.2)$$

where R' denotes the distance CP and J_0 is the Bessel function of order zero.

The equation (2.2) is valid, broadly speaking, at points P for which the number of fringes of defocusing and of lateral displacement are both O(1), that is to say do not exceed about 5 or 10. To define its domain of validity more precisely, we should need to specify the amount of inaccuracy which is regarded as tolerable and to examine the approximation errors in the same way as was done, for an F/15 pencil, in our paper on the intensity distribution already referred to.

The integral on the right of (2.2) can be evaluated in terms of the functions

$$U_n(u, v) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{u}{v}\right)^{n+2m} \mathrm{J}_{n+2m}(v) \qquad \ldots\ldots (2.3)$$

introduced by von Lommel for the purpose; in fact‡

$$\int_0^1 \mathrm{J}_0(v\rho) \exp(\tfrac{1}{2}iu\rho^2)\rho \, d\rho = -iu^{-1} \exp(\tfrac{1}{2}iu)(U_2 + iU_1) \qquad \ldots\ldots (2.4)$$

and (2.2) therefore gives

$$D_\lambda(\mathrm{P}) = \frac{2\pi a^2}{\lambda f u} \exp\{i[k(f-R') + \omega + \tfrac{1}{2}u]\} \sqrt{(U_1^2 + U_2^2)}, \qquad \ldots\ldots (2.5)$$

where

$$\cos\omega = \frac{U_2}{\sqrt{(U_1^2 + U_2^2)}}, \quad \sin\omega = \frac{U_1}{\sqrt{(U_1^2 + U_2^2)}} \qquad \ldots\ldots (2.6)$$

and $\sqrt{}$ denotes the positive square root here and elsewhere.

In equation (2.5)

$$\begin{aligned} R' = \mathrm{CP} &= \sqrt{[x^2 + y^2 + (z+f)^2]} \\ &= (z+f)\left[1 + \frac{r^2}{(z+f)^2}\right]^{1/2} \\ &= (z+f) + \frac{1}{2}\frac{r^2}{z+f} - \frac{1}{8}\frac{r^4}{(z+f)^3} - \ldots \end{aligned} \qquad \ldots\ldots (2.7)$$

Hence, by (2.1),

$$\begin{aligned} k(f-R') &= -\left(\frac{f}{a}\right)^2 u - \mathrm{O}\left(\frac{kr^2}{2f}\right) \\ &= -\left(\frac{f}{a}\right)^2 u - \frac{\lambda}{4\pi f}\left(\frac{f}{a}\right)^2 \mathrm{O}(v^2). \end{aligned} \qquad \ldots\ldots (2.8)$$

In the region of space where

$$u = \mathrm{O}(1), \qquad v = \mathrm{O}(1), \qquad \ldots\ldots (2.9)$$

the error term on the right of (2.8) is negligible in comparison with unity in a practical case; for example, if $\lambda = 2 \times 10^{-5}$ inch the value of $\lambda f/4\pi a^2$ is

† Following Rayleigh, we give this name to the time-independent part $D_\lambda(\mathrm{P})$ of the wave displacement D (P) exp ($ikct$).

‡ See Watson 1922, p. 541.

approximately $F^2/157\,000$, where F is the focal ratio and f is the distance measured in inches. Thus we may safely use the approximation

$$k(f-R')=-\left(\frac{f}{a}\right)^2 u. \qquad \ldots\ldots (2.10)$$

It follows from (2.5) that, in the region where (2.9) holds, the amplitude $M(u, v)$ and the phase $\phi(u, v)$ of the complex displacement $D_\lambda(\mathrm{P})$ are given by the equations

$$M(u,v)=\frac{2\pi a^2}{\lambda f|u|}\sqrt{(U_1^2+U_2^2)}, \qquad \ldots\ldots (2.11)$$

$$\phi(u,v)=\omega+\tfrac{1}{2}u-(f/a)^2u\,(\mathrm{mod}\,2\pi) \qquad \ldots\ldots (2.12)$$

respectively.†

From (2.12) and (2.6) it appears that $\phi(u, v)$, unlike $M(u, v)$, cannot be expressed in terms of the parameters u, v alone but has a structure which changes as the focal ratio varies. The multivalued function $\phi(u, v)$ has a branch point at each zero of the intensity $M^2(u, v)$ and is continuous elsewhere. At the focal point $u=v=0$ (which is not a branch point) one of its values is $\pi/2$.

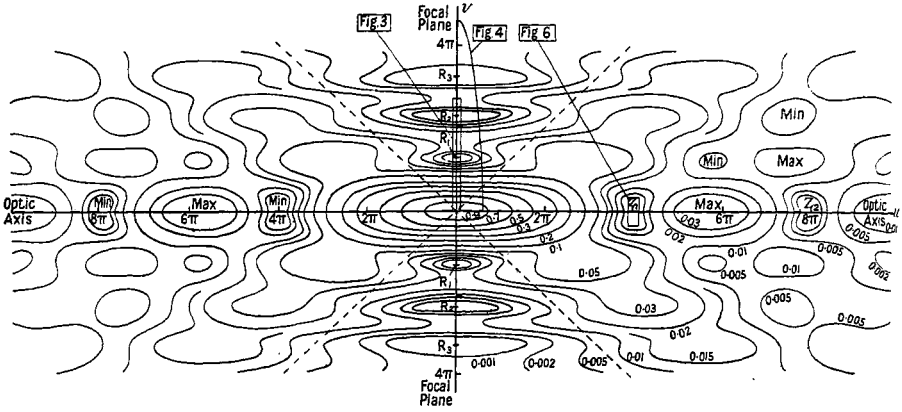

Figure 2. Key diagram showing the relation of figures 3, 4 and 6 to the intensity distribution in the diffraction image. The dotted lines represent the boundary of the geometrical cone of rays and the principal ray lies along the horizontal axis (u-axis). The contour lines are isophotes (lines of equal intensity) and the minima on the v-axis generate the Airy dark rings when the figure is rotated about the u-axis. The region covered by figure 5 is too large to be included in the diagram. The boundary of the quadrant labelled ' Fig. 4 ' appears elliptical because of the scale-distortion.

The co-phasal surfaces $\phi(u, v)=$ constant are surfaces of revolution about the axis $v=0$. By (2.12), (2.6), (2.3) they possess the further symmetry expressed by the equation

$$\phi(-u,v)+\phi(u,v)=\pi\ (\mathrm{mod}\,2\pi). \qquad \ldots\ldots (2.13)$$

The form of the equiphase surfaces can be computed from (2.12) and (2.6). We first consider the region, very close to the geometrical focal plane, represented in figure 2 by the rectangle labelled ' Fig. 3 '. This region is a thin, disc-shaped volume extending a little beyond the second Airy dark ring R_2. Close to the

† The notation (mod 2π) means that the two sides of the equation differ either by zero or by a multiple of 2π. ϕ itself is indeterminate to the extent of an additive multiple of 2π.

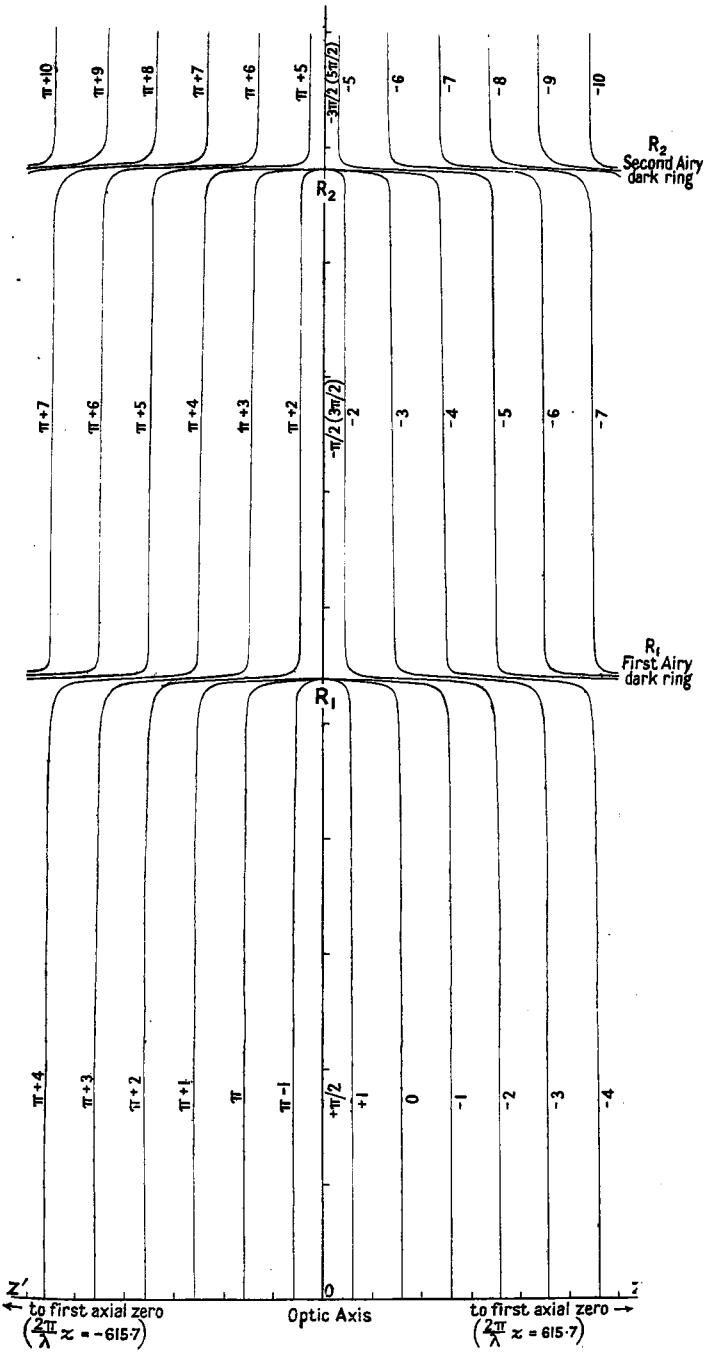

Figure 3. Co-phasal surfaces in the neighbourhood of the geometrical focal planeofan F/3·5 pencil. O is the geometrical focus, ZZ' the axis of the pencil, OR_1 and OR_2 the radii of the first and second Airy dark rings.

geometrical focus O, the equiphase surfaces are found to be substantially plane except (see figure 3) at points whose distances from the optic axis are nearly equal to the radii OR_1, OR_2, ... of the successive Airy dark rings. However, in the annular regions where they are nearly plane the equiphase surfaces are spaced closer together, by a factor $1-a^2/4f^2$, than would be those in a parallel beam of light of the same frequency; moreover (see figure 2) the intensity is far from uniform over each equiphase surface, being greatest on the optic axis.

The co-phasal surfaces which near O represent a phase-range $-\pi/2$ to $+3\pi/2$ are almost exactly plane until they begin to draw level with the first Airy dark ring R_1; at this distance from the optic axis ZZ′ (see figure 3) they make a sudden swerve inward to unite at the points of zero intensity which constitute the first Airy dark ring. The co-phasal surfaces on either side of this set all make a similar inward swerve opposite R_1 which brings them closer to the focal plane by a distance (a little less than $\frac{1}{2}\lambda$) which corresponds to a phase change of π. At a distance from the axis equal to the radius of the second Airy dark ring R_2, the co-phasal surfaces corresponding to the phase-ranges $-3\pi/2$ to $-\pi/2$ and $3\pi/2$ to $5\pi/2$ of total length 2π suddenly swerve inward to meet in R_2, while the remainder again swerve towards the focal plane through a distance corresponding to a phase change π. This swerving of the equiphase surfaces where they face the Airy dark rings becomes gradually less and less sudden as their distance from the focal plane increases.

From (2.12), in which ω is a function of u, v alone, we see that as the focal ratio varies the form of the equiphase surfaces undergoes a transformation which is more complicated than a simple scale distortion. To display the phase distribu-

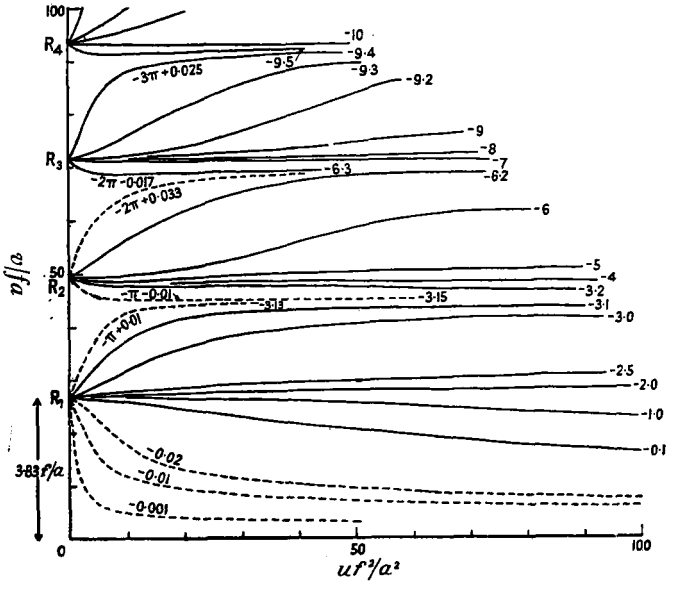

Figure 4. Difference $\phi-\tilde{\phi}$ between actual phase ϕ and phase $\tilde{\phi}=\frac{1}{2}u-(f^2/a^2)u$ of the fiducial plane wave. The numerical scales refer to an F/3·5 pencil; in this case unit distance on the graph represents an actual distance $\lambda/2\pi$, the radius OR_1 of the first Airy dark ring is $3{\cdot}83f/a$ units $=0{\cdot}61\lambda f/a$, and the distance OZ_1 to the first axial node Z_1 (not shown) is $4\pi f^2/a^2=606$ units, nearly.

tion in the neighbourhood of the origin we therefore introduce a fiducial plane wave,† of wavelength $\lambda(1-a^2/4f^2)$, whose phase $\tilde{\phi}=\frac{1}{2}\pi+\frac{1}{4}u-(f/a)^2u$ agrees (see (2.12) and (3.8) below) with that of the actual wave along the part of the u-axis near the origin. Figure 4 shows the difference $\phi-\tilde{\phi}=\omega+\frac{1}{4}u-\frac{1}{2}\pi$ over the interior of a region which, for an F/3·5 pencil, is a sphere of radius 16λ centred on the geometrical focus. As will be seen from figure 2, this sphere is about 3·5 times as wide, and one-sixth as long, as the central bright nucleus of the diffraction image. A change in the value of the focal ratio $f/2a$ leaves the curves in figure 4 unaltered, but changes the numerical scales along the two axes in accordance with equations (2.1).

§ 3. The Phase Anomaly

In the approximation of geometrical optics, the aberration-free image of a point object is formed by a pencil of concurrent rays issuing from the exit pupil of an optical system. The orthogonal surfaces of this pencil, usually called the geometrical wave fronts, are spherical with the focal point O as common centre of curvature (see figure 1). In the region I, light is converging towards the focus O; in the region II the rays have passed through O and are diverging again. In this simplified picture, the geometrical wave fronts are regarded as surfaces of constant phase, and the phase difference between any two wave fronts is taken as $2\pi/\lambda$ times the optical distance between them, measured along the geometrical rays.

If the wave amplitude is taken as 1 over the suface of the geometrical wave front filling the exit pupil, the disturbance at the point $P(x,y,z)$ in the image space is represented in the simplified geometrical picture by the complex displacement function

$$\left.\begin{aligned} D_\lambda^*(P) &= f\frac{e^{ikR}}{R} \text{ in region I} \\ &= f\frac{e^{-ikR}}{R} \text{ in region II} \\ &= 0 \text{ elsewhere in the image space} \end{aligned}\right\}, \quad \ldots\ldots(3.1)$$

where $f=OC$ and $k=2\pi/\lambda$ as before, while

$$R=\sqrt{(x^2+y^2+z^2)}=\frac{f}{ka}\sqrt{\left\{\left(\frac{f}{a}\right)^2u^2+v^2\right\}}$$

is the distance of P from the focus O. In regions I and II, the phase attributed to the displacement at P is thus represented by the quantity

$$\phi^*(u,v)=-kR\,\mathrm{sgn}\,u, \quad \ldots\ldots(3.2)$$

where $\mathrm{sgn}\,u$ is defined as ± 1 according as $u\gtrless 0$. In the remainder of the image space, there is complete darkness according to the present approximation, and the phase $\phi^*(u,v)$ is not defined.

The difference

$$\delta(u,v)=\phi(u,v)-\phi^*(u,v) \quad \ldots\ldots(3.3)$$

between the phases predicted by (2.5) and (3.1) respectively is called the *phase anomaly*. It is defined only in the regions I and II, where by (3.2)

$$\delta(u,v)=\phi(u,v)+kR\,\mathrm{sgn}\,u \quad \ldots\ldots(3.4)$$

† This expedient was suggested by a referee.

and by (2.13) and (3.4)

$$\delta(-u, v)+\delta(u, v)=\pi \pmod{2\pi}. \qquad \ldots\ldots(3.5)$$

When P tends to O, the geometrical phase $\phi^*(0, 0)$ tends to zero, while the 'principal value' of $\phi(0, 0)$ is $\frac{1}{2}\pi$ by (2.2); (3.4) then gives the well-known result

$$\delta(0, 0)=\tfrac{1}{2}\pi. \qquad \ldots\ldots(3.6)$$

In the parts of the regions I and II which lie within 6λ of the geometrical focus, each surface of phase ϕ is substantially coincident with the 'best' plane approximation to the corresponding geometrical wave cap of phase $\phi-\frac{1}{2}\pi$, namely the plane which deviates from the geometrical wave cap by equal and opposite amounts at centre and edge. This has the consequence, already noted above, that consecutive surfaces of the same phase (mod 2π) in this region are spaced a distance $(1-a^2/4f^2)\lambda$ apart, in agreement with (2.12) and (3.9), instead of a distance λ apart.

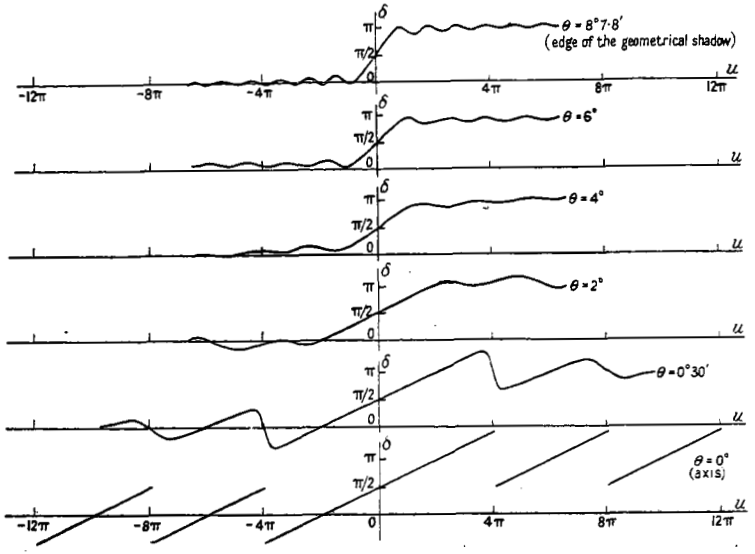

Figure 5. Phase anomaly δ along geometrical rays through the focal point of an F/3·5 pencil. θ is the angle between the ray and the axis of the pencil.

Further from the origin, the phase anomaly exhibits a different type of behaviour, shown for the special case of an F/3·5 pencil in figures 5 and 6. When $v=0$ (that is to say, on the axis of the pencil) the Lommel functions U_1, U_2 in (2.6) take the special values

$$\left.\begin{aligned} U_1(u, 0)&=\sin\tfrac{1}{2}u=2\sin\tfrac{1}{4}u\cos\tfrac{1}{4}u \\ U_2(u, 0)&=1-\cos\tfrac{1}{2}u=2\sin^2\tfrac{1}{4}u \end{aligned}\right\} \qquad \ldots\ldots(3.7)$$

and we obtain from (2.6)

$$\left.\begin{aligned} \omega&=\tfrac{1}{2}\pi-\tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad 2s\pi<\tfrac{1}{4}u<(2s+1)\pi \\ &=\tfrac{3}{2}\pi-\tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad (2s+1)\pi<\tfrac{1}{4}u<2s\pi \end{aligned}\right\} \qquad \ldots\ldots(3.8)$$

where $s=0, \pm 1, \pm 2, \ldots$. Hence by (2.12), (3.1), (3.3)

$$\begin{aligned} \delta(u, 0)&=\tfrac{1}{2}\pi+\tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad 8s\pi\leqslant|u|<8(s+\tfrac{1}{2})\pi \\ &=-\tfrac{1}{2}\pi+\tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad 8(s+\tfrac{1}{2})\pi\leqslant|u|<8(s+1)\pi. \end{aligned} \qquad \ldots\ldots(3.9)$$

There is a phase jump of π at each of the axial 'nodes' (points of zero intensity) $u=4s\pi$; $s=\pm1,\ \pm2,\ldots$. Between two consecutive nodes the phase anomaly $\delta(u,0)$ varies linearly, as shown in the bottom curve of figure 5, until $|u|$ becomes so large that the approximation (2.2) begins to break down.

The computation of the phase distribution on the edge $v=|u|$ of the geometrical shadow is simplified by observing that the Lommel functions U_1, U_2 take for $v=u$ the simple forms†

$$U_1(u,u)=\tfrac{1}{2}\sin u$$
$$U_2(u,u)=\tfrac{1}{2}[J_0(u)-\cos u]. \qquad \ldots\ldots(3.10)$$

The resulting phase anomaly is shown, for the case of an F/3·5 pencil, in the top curve of figure 5.

Figure 5 illustrates the difference between the behaviour of the phase anomaly along the axial ray and along the non-axial rays of the pencil. Along each non-axial ray, the phase anomaly approaches π as the distance from the focus increases (Debye 1909). The different behaviour, described by (3.9), of the phase anomaly along the axial ray is by no means surprising, since this ray passes through the axial nodes. In figure 6 is shown the behaviour of the co-phasal surfaces near the first axial node.

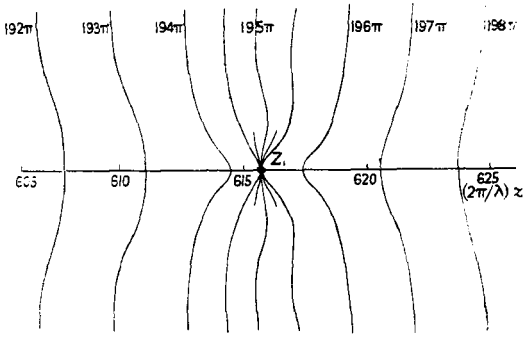

Figure 6. Co-phasal surfaces in the neighbourhood of the first axial node (point of zero intensity) Z_1 in an F/3·5 pencil.

ACKNOWLEDGMENTS

We should like to thank Mrs. I. Berrer for the care and skill with which she has carried out the heavy computations. One of us (E.W.) wishes to acknowledge gratefully financial assistance received from the Carnegie Trust for the Universities of Scotland towards computing expenses. He is also indebted to the Universities of Edinburgh and Manchester for awards of I.C.I. Research Fellowships.

REFERENCES

BEREK, M., 1926, *Z. Phys.*, **40**, 421.
BOUWKAMP, C. J., 1940, *Physica*, **7**, 485.
BREUNINGER, H. W., 1938, *Thesis*, University of Jena; 1939, *Ann. Phys., Lpz.*, **35**, 238.
DEBYE, P., 1909, *Ann. Phys., Lpz.*, **30**, 755.
FABRY, C., 1893, *J. Phys. Theor. Appl.* (3), **2**, 22.

† These follow from the Jacobi identities (Watson 1922, p. 22).

FOKKER, A. D., 1923, *Physica*, **3**, 334; 1924, *Ibid.*, **4**, 166.
GOUY, L. G., 1890, *C. R. Acad. Sci., Paris*, **110**, 1251; 1891, *Ann. Chim. (Phys.)*, **24**, 14
IGNATOWSKY, B. C., 1919, *Trans. State Optical Institute, Petrograd*, **1**, III.
JOUBIN, P., 1892, *C.R. Acad. Sci., Paris*, **115**, 932.
JULIUS, W. H., 1895, *Archives Néerl. Sci.*, **28**, 221, 226.
LINFOOT, E. H., and WOLF, E., 1952, *Mon. Not. Roy. Astr. Soc.*, **112**, 452.
VON LOMMEL, E., 1885, *Abh. Bayer Akad. Wiss.*, **53**, 233.
MARÉCHAL, A., 1948, *Rev. Opt. (Theor. Instrum.)*, **27**, 73.
PICHT, J., 1930, *Z. Phys.*, **65**, 14.
POHL, R. W., 1948, *Optik*, 7th and 8th Ed. (Berlin : Springer), p. 353.
REICHE, F., 1909 a, *Ann. Phys., Lpz.*, **29**, 65; 1909 b, *Ibid.*, **29**, 401.
RUBINOWICZ, A., 1938, *Phys. Rev.*, **54**, 931.
SAGNAC, G., 1903, *J. Phys. (Théor. Instrum.)*, **2**, 721; 1904 a, *C.R. Acad. Sci., Paris*, 1
479, 619, 678; 1904 b, *Boltzmann Festschrift*, 528.
STRUVE, H., 1886, *Mém. de l'Acad. de St. Petersb.* (7), **34**, No. 5, 1.
TORALDO DI FRANCIA, G., 1942, *Ottica*, **7**, N.2.
WATSON, G. N., 1922, *Bessel Functions* (Cambridge : University Press).
ZEEMAN, P., 1897/98, *Versl. Gewone Vergad. Akad. Amst.*, **6**, 11; 1900, *Phys. Z.*, 1,
1901, *Archives Néerl. Sci.*, **4**, 314, 318.
ZERNIKE, F., and NIJBOER, B. R. N., 1949, Contribution to *Théorie des Images Optiq*
(Paris : Éditions de la Revue d'Optique).