Optics Communications 283 (2010) 3371–3375

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

The origin of the Gouy phase anomaly and its generalization to astigmatic wavefields

T.D. Visser [a,b,*], E. Wolf [c,d]

[a] Department of Electrical Engineering, Delft University of Technology Delft, The Netherlands
[b] Department of Physics and Astronomy, Free University Amsterdam, The Netherlands
[c] Department of Physics and Astronomy, University of Rochester, River Campus Station, Rochester, NY 14627, USA
[d] The Institute of Optics, University of Rochester, River Campus Station, Rochester, NY 14627, USA

ARTICLE INFO

Article history:
Received 5 March 2010
Received in revised form 22 April 2010
Accepted 27 April 2010

ABSTRACT

One of the most poorly understood subjects in physical optics is the origin of the Gouy phase (sometimes called "the phase anomaly near focus"). This is evident from the large number of publications on the subject, many of which attribute it to quite different causes. In this paper we show that the Gouy phase anomaly can be clearly understood from elementary properties of normal congruences of light rays and from the relationship between geometrical optics and physical optics. We also show that the Gouy phase anomaly may be regarded as a degenerate case of a rapid $\pi/2$ phase change that is found to occur at each focal line of an astigmatic pencil of rays. The intensity distribution in the region of the phase changes is also presented. Furthermore, symmetry relations for both the phase anomaly and the intensity distribution are derived.

1. Introduction

In two seminal papers L.G. Gouy [1,2] described an anomalous behavior of the phase of a converging diffracted spherical wave as it passes through a focus. He wrote (translated from French [1]):

> *If one considers a converging wave that has passed through a focus and has then become divergent, a simple calculation shows that the vibration of that wave has advanced half a period compared to what it should be according to the distance traveled and the speed of light.*

Gouy confirmed his theoretical analysis experimentally. By letting the light from a point source impinge onto two mirrors, one concave, the other plane, two beams were generated. The mirrors were positioned so that the beams were nearly parallel to each other. In any transverse plane of observation, their superposition yielded a circular interference pattern, with ring-shaped fringes. The central disk was found to change from dark to bright, or vice versa, when the observation plane was moved through the focus of the converging beam. This transition demonstrated the predicted 180° phase change. Since Gouy's original work, many observations of such a phase anomaly have been reported. (See, for example, Refs. [3–13].)

The origin of the phase anomaly continues to be a matter of debate, with different authors attributing it to widely differing causes. For example, it has been associated with Heisenberg's uncertainty relations [14,15], with Berry's geometric phase [16,17], and with geometrical properties of Gaussian beams [18]. It has even been "explained, one might say somewhat esoterically," by relating it to the i/λ factor in front of the Kirchhoff diffraction integral [19]!

In the present paper we show that the origin of the Gouy phase anomaly can be clearly understood from elementary properties of normal congruences of light rays ([20], Sec. 3.2) and from the well-known relationship between geometrical optics and physical optics that is brought out by the principle of stationary phase (see ([20], App. III) or ([21], Ch. 9)). We present curves that show the behavior of the phase and of the intensity in the region of the anomaly.

One of the earliest treatments of the Gouy phase is due to Walker [22], who used a forerunner of the principle of stationary phase to demonstrate that when a ray associated with an astigmatic wavefront (i.e., a wavefront with two unequal principle radii of curvature) passes through the two centers of curvature, there is a phase discontinuity of an amount of $\pi/2$ at each of them. Walker's analysis is in agreement with the following remark of Gouy (again in translation [2]):

> *A similar calculation shows that when a wave passes through a focal line, the phase advance is exactly half of that in the above case; non-spherical converging waves pass through two successive focal lines, the total phase advance is thus the same as for spherical waves.*

A detailed analysis of the behavior of the phase and the intensity in the focal region was made by Linfoot and Wolf [23]. They considered a monochromatic, diffracted spherical converging wave (with time-dependence $exp[-i\omega t]$) that emerges from a circular aperture with radius a (see Fig. 1). Using the Huygens-Fresnel Principle, they expressed the space-dependent part of the field as (see also Ref. ([20], Sec. 8.8))

$$U(u,v) = -\frac{2\pi i a^2 C}{\lambda f^2} e^{if^2 u/a^2} \int_0^1 J_0(v\rho) e^{-iu\rho^2/2} \rho \, d\rho, \tag{1}$$

* Corresponding author. Department of Electrical Engineering, Delft University of Technology Delft, The Netherlands.
E-mail address: tvisser@nat.vu.nl (T.D. Visser).

0030-4018/$ – see front matter
doi:10.1016/j.optcom.2010.04.099

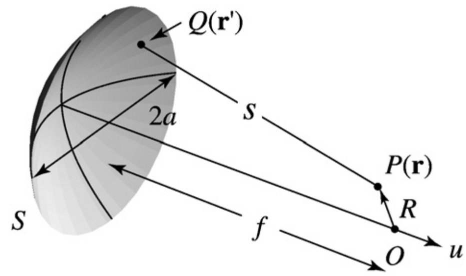

Fig. 1. Illustrating the notation. The origin O of a Cartesian coordinate system is taken at the geometrical focus of the converging spherical wave.

with λ denoting the wavelength and f the radius of the wave front. Furthermore, C is a positive constant, and J_0 denotes the Bessel function of the first kind and zero order. The parameters u and v are the so-called Lommel variables, representing the position of an observation point in the focal region. They are defined as

$$u = \frac{2\pi}{\lambda}\left(\frac{a}{f}\right)^2 z, \quad v = \frac{2\pi}{\lambda}\left(\frac{a}{f}\right)(x^2 + y^2)^{1/2}. \tag{2}$$

The conditions under which Eq. (1) is valid are discussed in [24] and in ([21], Sec 12.1.2).

The phase anomaly δ is defined as the difference between the actual phase of the wave, $\arg[U(u, v)]$, and that of a (non-diffracted) spherical wave converging to the focus in the half-space $u<0$ and diverging from it in the halfspace $u>0$, (cf. ([20], Sec. 8.8, Eq. (48))), i.e.,

$$\delta(u, v) = \arg[U(u, v)] - \text{sign}\,(u)kR, \tag{3}$$

where $k = 2\pi/\lambda$ and

$$R = \left(x^2 + y^2 + z^2\right)^{1/2} = \left\{(f/ak)^2\left[v^2 + (fu/a)^2\right]\right\}^{1/2} \geq 0, \tag{4}$$

is the distance from the observation point to the geometrical focus (see Fig. 1). The function $\delta(u, v)$ represents the difference between the actual phase of the wave and of the phase given by geometrical optics. In Ref. [23] the phase anomaly was calculated along geometrical rays that pass through the focus (see Fig. 2). As the angle of inclination between the ray and the axis of rotational symmetry (the u-axis) decreases, the π phase change becomes more gradual. Also, in agreement with the predictions of Gouy, it was found that half of the advance takes place in front of the geometrical focus and the other half behind it. The discontinuous behavior of the phase anomaly along the u-axis ($\alpha = 0°$) occurs at zeros of intensity. Since at these zeros the field has also zero value, therefore its phase is undetermined. The behavior of the phase in the vicinity of such points has been intensively studied in recent years in the field of singular optics (see ([20], Sec. 8.8.4) and [25,26]).

2. The diffraction integral in the presence of aberrations

When an aberrated converging, monochromatic wave is focused, the field in the focal region is given by the expression ([20], Sec. 9.1.1)

$$U(P) = -\frac{\mathrm{i}}{\lambda}\frac{Ce^{-\mathrm{i}kf}}{f}\iint_S \frac{e^{\mathrm{i}k[\Phi + s]}}{f}\,\mathrm{d}S, \tag{5}$$

where Φ denotes the wave aberration function (see Fig. 3), and s is the distance from a point of integration Q to the observation point P (see Fig. 1). The integration extends over the Gaussian reference sphere S that approximately fills the aperture. For a wavefront suffering from

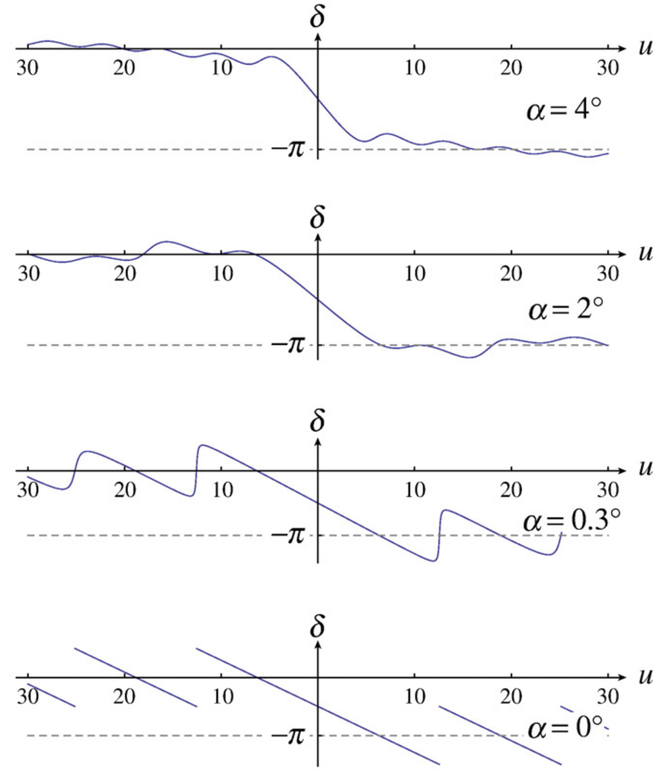

Fig. 2. The phase anomaly δ of a diffracted converging spherical wave along selected geometrical rays through focus. The angle α denotes the inclination of the ray to the u-axis. The dashed line indicates the value $\delta = -\pi$. In this example $f = 35$ cm, $a = 5$ cm, and $\lambda = 1$ µm. After Ref. [23].

astigmatism, the aberration function, expressed in spherical polar coordinates, is ([20], Sec. 9.3)

$$\Phi(\rho, \theta) = A_0\rho^2 cos^2\theta, \qquad (0 \leq \rho \leq 1, 0 \leq \theta \leq 2\pi), \tag{6}$$

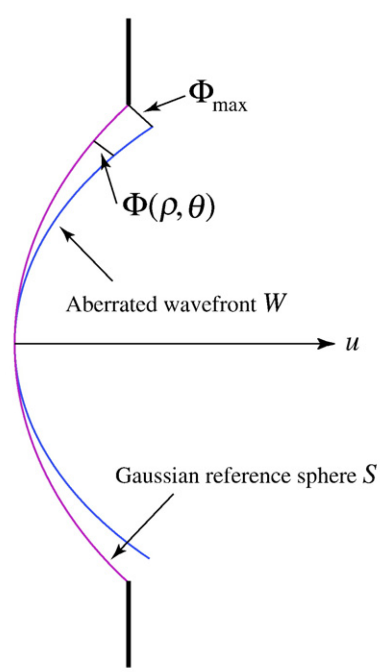

Fig. 3. An aberrated wavefront W, the Gaussian reference sphere S and the aberration function $\Phi(\rho, \theta)$.

where A_0 is a real number. It is to be noticed that the maximum value that the aberration function can attain on the reference sphere, $\Phi_{max} = A_0$. For observation points on the axis of symmetry (the u-axis), we find from Eqs. (5) and (6) that

$$U(u,0) = -\frac{iCa^2}{\lambda f^2} e^{if^2u/a^2} \int_0^1 \int_0^{2\pi} \rho e^{i(kA_0\rho^2\cos^2\theta - u\rho^2/2)} d\rho d\theta, \quad (7)$$

$$= -\frac{i2\pi Ca^2}{\lambda f^2} e^{if^2u/a^2} \int_0^1 e^{i[(kA_0-u)\rho^2]/2} \rho J_0\left(kA_0\rho^2/2\right) d\rho. \quad (8)$$

Let us define the normalized on-axis intensity by the formula

$$I(u,0) = |U(u,0)|^2 / I_0, \quad (9)$$

with

$$I_0 = \left(\frac{\pi Ca^2}{\lambda f^2}\right)^2 \quad (10)$$

being the intensity at the geometrical focus of the aberration-free wave. When the astigmatism is sufficiently large (i.e., when the coefficient A_0 is large compared to the wavelength λ), one might expect that the on-axis intensity distribution will exhibit a peak at each of the two focal lines. The intersections of these focal lines with the u-axis are given by the equations (cf. ([20], Sec. 9.3, Eq. (17)))

$$u_s = 0, \qquad u_t = 2kA_0, \quad (11)$$

where the indices s and t denote the sagittal and tangential focus, respectively. It follows from Eqs. (8), (9) and (11) that the on-axis intensity is symmetric about the midpoint of the two focal lines, that is to say the point labeled by the Lommel variable $u = \overline{u} \equiv (u_s + u_t)/2 = kA_0$, i.e.,

$$I(\overline{u} + \Delta, 0) = I(\overline{u} - \Delta, 0). \quad (12)$$

The on-axis phase anomaly $\delta(u, 0)$ may again be defined as in Eq. (3), but with the axial field $U(u, 0)$ now given by Eq. (8) and with $\mathrm{sign}(u)kR = f^2u/a^2$. Hence

$$\delta(u,0) = \arg\left[\int_0^1 e^{i[(kA_0-u)\rho^2]/2} \rho J_0\left(kA_0\rho^2/2\right) d\rho\right] - \frac{\pi}{2}. \quad (13)$$

It follows from Eq. (11) that the integral on the right-hand side of Eq. (13) is real-valued when $u = \overline{u}$. We verified numerically for wave aberrations for which $|\Phi_{max}| < 100\lambda$, that the integral is always positive. Hence we conclude that the on-axis phase anomaly halfway between the two focal lines ($u = \overline{u}$) equals $-\pi/2$, i.e.,

$$\delta[\overline{u}, 0] = -\pi/2, \qquad (|\Phi_{max}| < 100\lambda). \quad (14)$$

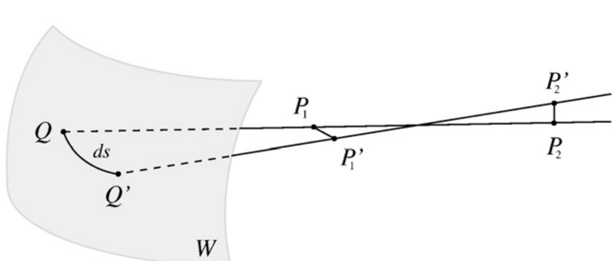

Fig. 4. A wavefront W is orthogonal to two rays at points Q and Q'. The arc length QQ' is denoted by ds. The distances between the astigmatic foci P_1, P'_1 and P_2, P'_2 are of the order $(ds)^2$. For details see ([20], Sec. 3.2).

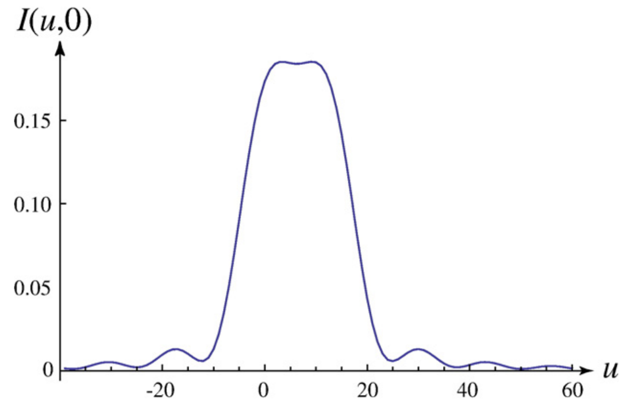

Fig. 5. The on-axis intensity of a diffracted, converging spherical wave in the presence of astigmatic aberration with $\Phi_{max} = 1\lambda$, as given by Eq. (9). Furthermore, $f = 2$ m, $a = 3$ cm, $\lambda = 1$ μm. The two focal lines are at $u_s = 0$ and $u_t = 4\pi$.

Eq. (13) also brings into evidence the "symmetry relation"

$$\delta\left[\overline{u} + \Delta, 0\right] + \delta\left[\overline{u} - \Delta, 0\right] = -\pi. \quad (15)$$

In the next two sections Eqs. (8) and (13) will be used and their implications will be compared with the predictions of geometrical optics.

3. Behavior of the field in the focal region

3.1. Geometrical optics

We will now discuss the phase anomaly on the basis of geometrical optics. It will be useful to begin by recalling some basic results relating to families of geometrical wavefronts and of the associated light rays.

A system of curves that fills a portion of space in such a way that a single curve passes through each point in the region, is called a *congruence* (see ([20], Sec. 3.2) or ([27], Ch. 10)). If there exists a family of surfaces which intersects each of the curves orthogonally, the congruence is said to be *normal*. If each of the curves is a straight line, the congruence is said to be *rectilinear*. In the context of optics, a congruence of curves is a bundle of light rays, and the surfaces which intersect them orthogonally represent the wavefronts.

Let us now consider a wavefront W in free space, and let us consider two rays passing through two closely-separated points Q and Q' on W, with a small distance ds between them (see Fig. 4). If we choose a point on each of these two rays, say P and P', the distance

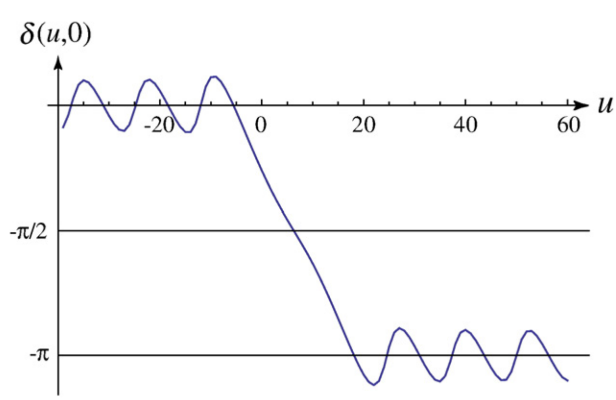

Fig. 6. The on-axis phase anomaly $\delta(u, 0)$ of a diffracted, converging spherical wave in the presence of astigmatic aberration with $\Phi_{max} = 1\lambda$, as given by Eq. (13). The parameters f, a and λ are the same as in Fig. 5. The levels $\delta(u, 0) = -\pi/2$ and $-\pi$ are also indicated.

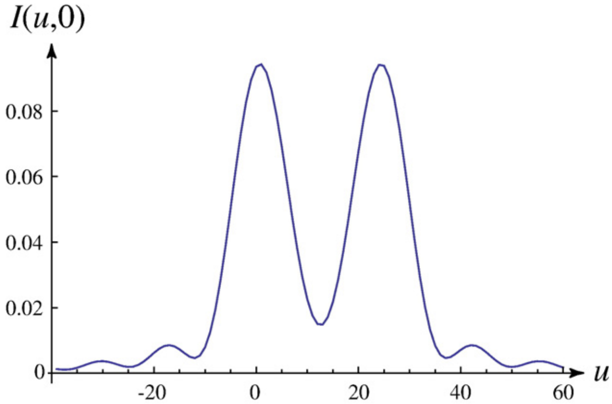

Fig. 7. The on-axis intensity of a diffracted, converging spherical wave in the presence of astigmatic aberration with $\Phi_{max} = 2\lambda$. The parameters f, a and λ are the same as in Fig. 5. The two focal lines are at $u_s = 0$ and $u_t = 8\pi$.

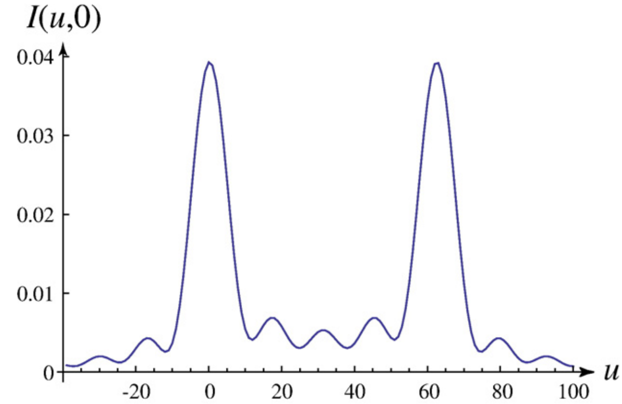

Fig. 9. The on-axis intensity of a diffracted, converging spherical wave in the presence of astigmatic aberration with $\Phi_{max} = 5\lambda$. The parameters f, a and λ are the same as in Fig. 5. The two focal lines are at $u_s = 0$ and $u_t = 20\pi$.

between them will be of order ds or greater. However, one can show ([20], Sec. 3.2) that there are two points P_1 and P_2 on one ray and P_1' and P_2' on the other ray, whose separation is of order $(ds)^2$. These special points are known as *astigmatic foci*, and their locus when the points Q and Q' are varied, are two line elements known as *astigmatic focal lines*. They may be shown to be mutually orthogonal.

3.2. Physical optics

As is well known, geometrical optics may be regarded as the asymptotic limit of physical optics as the wavenumber $k = 2\pi/\lambda$ tends to infinity ([20], Sec. 3.1). It can be shown by the use of the two-dimensional principle of stationary phase that in this limit the field associated with astigmatism exhibits a phase discontinuity of $\pi/2$ at each focal line [28]. Now the actual time-independent field obeys the Helmholtz equation rather than the eikonal equation that governs geometrical optics. The Helmholtz equation belongs to the class of elliptic differential equations, and it is well known that solutions of such equations are boundary values of on the real axis of functions that are analytic and regular in any closed domain, see, for example ([29], pp. 42–43). Hence the actual field, in contrast to the geometrical optics field, is necessarily continuous, and, consequently, so is its phase except at points where the field has zero value and consequently the phase is singular there. The actual field in the region of the astigmatic focal lines may be expected to exhibit the same features as the geometrical optics field, but the discontinuities at the focal lines will be 'smoothed out'. Put differently, the actual physical optics field will have sharp variations in the region of the focal lines, but will have no discontinuities. This situation is illustrated in Figs. 5–10 in which the intensity and the phase anomaly are shown for different amounts of astigmatism. It can be seen how for a relatively small amount of astigmatism, viz. $\Phi_{max} = 1\lambda$, the on-axis intensity is still single-peaked, and the phase changes continuously by an amount of π at the peak. When the aberration is increased, say to $\Phi_{max} = 2\lambda$, an intensity peak is present at each focal line. The phase change now takes place in two $\pi/2$ steps. For an even larger amount of astigmatic aberration, $\Phi_{max} = 5\lambda$ say, the two intensity peaks are separated by a greater distance, in accordance with Eq. (11). Also, the phase changes by an amount of $\pi/2$ around each focal line. We note that the two symmetry properties expressed by Eqs. (12) and (15) are clearly visible in Figs. 5–10.

We summarize our analysis by saying that the geometrical optics behavior of focused fields and the relation between geometrical optics and physical optics, make it clear that the field at each astigmatic focal line undergoes a rapid $\pi/2$ phase change. In the limiting case when the astigmatic wave aberration tends to zero, i.e., when the field in the aperture becomes a converging spherical wave, the two foci coincide and the sharp change in the focal region is the Gouy phase change by an amount π.

Thus we have shown that the Gouy phase is due to 1) the asymptotic behavior of the wavefield in the focal region as the wavenumber $k \to \infty$, and 2) the well-understood relation between geometrical optics and physical optics.

Acknowledgments

The research of TDV is supported by the Foundation for Fundamental Research on Matter (FOM) and by the Dutch Technology

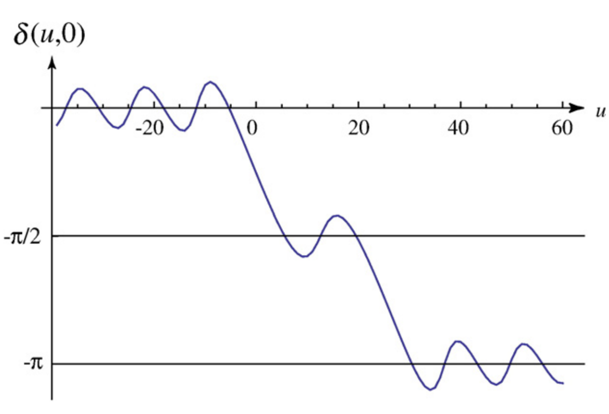

Fig. 8. The on-axis phase anomaly $\delta(u, 0)$ of a diffracted, converging spherical wave in the presence of astigmatic aberration with $\Phi_{max} = 2\lambda$. The parameters f, a and λ are the same as in Fig. 5.

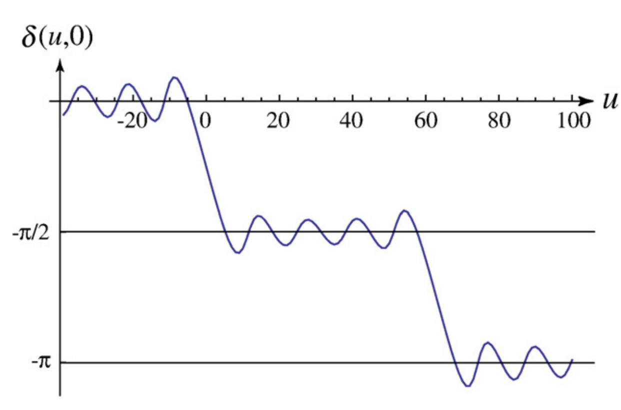

Fig. 10. The on-axis phase anomaly $\delta(u, 0)$ of a diffracted, converging spherical wave in the presence of astigmatic aberration with $\Phi_{max} = 5\lambda$. The parameters f, a and λ are the same as in Fig. 5.

Foundation (STW). That of EW is supported by the U.S. Air Force Office of Scientific Research under grant FA 9550-08-1-0417 and by the Air Force Research Laboratory (AFRL) under contract FA 9451-04-C-0296.

References

[1] L.G. Gouy, Sur une propriété nouvelle des ondes lumineuses, CR Hebdomadaires Séances Acad. Sci. 110 (1890) 1251.
[2] L.G. Gouy, Sur la propagation anomale des ondes, Annales des Chimie et de Physique 24 (1891) 145 6e séries.
[3] G.W. Farnell, Measured phase distribution in the image space of a microwave lens, Canadian J. Phys. 36 (1958) 935.
[4] L. Mertz, A new demonstration of the anomalous phase change of light at a focus, J. Opt. Soc. Am. 49 (1959) AD 10-iv.
[5] A.B. Ruffin, J.V. Rudd, J.F. Whitaker, S. Feng, H.G. Winful, Direct observation of the Gouy phase shift with single-cycle Terahertz pulses, Phys. Rev. Lett. 83 (1999) 3410.
[6] R.W. McGowan, R.A. Cheville, D. Grischkowskya, Direct observation of the Gouy phase shift in THz impulse ranging, Appl. Phys. Lett. 76 (2000) 670.
[7] T. Feurer, N.S. Stoyanov, D.W. Ward, K.A. Nelson, Direct visualization of the Gouy phase by focusing phonon polaritons, Phys. Rev. Lett. 88 (2002) 257402.
[8] J.H. Chow, G. de Vine, M.B. Gray, D.E. McClelland, Measurement of Gouy phase evolution by use of spatial mode interference, Opt. Lett. 29 (2004) 2339.
[9] T. Klaassen, A. Hoogeboom, M.P. van Exter, J.P. Woerdman, Gouy phase of nonparaxial eigenmodes in a folded resonator, J. Opt. Soc. Am. A 21 (2004) 1689.
[10] G. Lamouche, M.L. Dufour, B. Gauthier, J.-P. Monchalin, Gouy phase anomaly in optical coherence tomography, Opt. Commun. 239 (2004) 297.
[11] F. Lindner, G.G. Paulus, H. Walther, A. Baltuska, E. Goulielmakis, M. Lezius, F. Krausz, Gouy phase shift for few-cycle laser pulses, Phys. Rev. Lett. 92 (2004) 113001.
[12] O. Steuernagel, E. Yao, K. O'Holleran, M. Padgett, Observation of Gouy-phase-induced transversal intensity changes in focused beams, J. Mod. Optics 52 (2005) 2713–2721.
[13] H.C. Kandpal, S. Raman, R. Mehrotra, Observation of Gouy phase anomaly with an interferometer, Opt. Lasers Eng. 45 (2007) 249.
[14] P. Hariharan, P.A. Robinson, The Gouy phase shift as a geometrical quantum effect, J. Mod. Optics 43 (1996) 219.
[15] S. Feng, H.G. Winful, Physical origin of the Gouy phase shift, Opt. Lett. 26 (2001) 485.
[16] D. Subbarao, Topological phase in Gaussian beam optics, Opt. Lett. 20 (1995) 2162.
[17] R. Simon, N. Mukunda, Bargmann invariant and the geometry of the Güoy effect, Phys. Rev. Lett. 70 (1993) 880.
[18] R.W. Boyd, Intuitive explanation of the phase anomaly of focused light beams, J. Opt. Soc. Am. 70 (1980) 877.
[19] A.E. Siegman, *Lasers* (University Science Books, Mill Hill, 1986), Sec. 17.4. In this work Gouy's name is consistently misspelled as “Guoy”.
[20] M. Born, E. Wolf, Principles of Opticsseventh (expanded) ed, Cambridge University Press, Cambridge, 1999.
[21] J.J. Stamnes, Waves in Focal Regions, Adam Hilger, Bristol, 1986.
[22] J. Walker, The Analytical Theory of Light, Cambridge University Press, Cambridge, 1904 pp. 91.
[23] E.H. Linfoot, E. Wolf, Phase distribution near focus in an aberration-free diffraction image, Proc. Phys. Soc. B 69 (1956) 823.
[24] E. Wolf, Y. Li, Conditions for the validity of the Debye integral representation of focused fields, Opt. Commun. 39 (1981) 205.
[25] J.F. Nye, Natural Focusing and Fine Structure of Light, IOP Publishing, Bristol, 1999.
[26] M.S. Soskin, M.V. Vasnetsov, in: E. Wolf (Ed.), Progress in Optics, Vol. 42, Elsevier, Amsterdam, 2001.
[27] C.E. Weatherburn, Differential Geometry of Three Dimensions, Cambridge University Press, Cambridge, 1931.
[28] N.G. van Kampen, An asymptotic treatment of diffraction problems, Physica XIV (1949) 575.
[29] A. Sommerfeld, Partial Differential Equations in Physics, Academic Press, New York, 1949.