

# Axially symmetric focusing as a cuspid diffraction catastrophe: Scalar and vector cases and comparison with the theory of Mie

Johannes Kofler\*

*Institute of Experimental Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria*

Nikita Arnold†

*Institute for Applied Physics, Johannes Kepler University Linz, Altenbergerstrasse 69, 4040 Linz, Austria*

(Received 2 January 2006; revised manuscript received 23 March 2006; published 2 June 2006)

An analytical description of arbitrary strongly aberrated axially symmetric focusing is developed. This is done by matching the solution of geometrical optics with a wave pattern which is universal for the underlying ray structure. The corresponding canonical integral is the Bessoid integral, which is a three-dimensional generalization of the Pearcey integral that approximates the field near an arbitrary two-dimensional cusp. We first develop the description for scalar fields and then generalize it to the vector case. As a practical example the formalism is applied to the focusing of light by transparent dielectric spheres with a few wavelengths in diameter. The results demonstrate good agreement with the Mie theory down to Mie parameters of about 30. Compact analytical expressions are derived for the intensity on the axis and the position of the diffraction focus both for the general case and for the focusing by microspheres. The high intensity region is narrower than for an ideal lens of the same aperture at the expense of longitudinal localization and has a polarization dependent fine structure, which can be explained quantitatively. The results are relevant for aerosol and colloid science where natural light focusing occurs and can be used in laser micro- and nano-processing of materials.

DOI: [10.1103/PhysRevB.73.235401](https://doi.org/10.1103/PhysRevB.73.235401)

PACS number(s): 42.15.Dp, 41.20.Jb, 42.25.Fx, 81.16.-c

## I. INTRODUCTION

Axially symmetric focusing of wave fields occurs in various areas of science, since physical systems often possess an intrinsic rotational symmetry. In particular, the electromagnetic field enhancement by small spherical particles is important in many situations. Spheres have minimal surface energy for a given volume and thus are naturally formed as a result of phase separation, for example, as aerosols or colloids. Applications of colloidal microspheres in photonic crystals and photonic crystal slabs led to an explosion of the experimental and theoretical studies of their optical properties.<sup>1,2</sup> The majority of these investigations concentrate on their collective properties in a periodic arrangement. Single microspheres are used as high quality optical resonators and as agents that allow controlled and highly localized wavelength-dependent field enhancement for nonlinear optical studies and in resonance spectroscopy.<sup>3,4</sup> Here, the emphasis is placed on the eigenmode analysis and the distribution of the field within the sphere or in the immediate vicinity of its surface.

Lately, it was demonstrated that self-assembling arrays of transparent colloidal microspheres can be employed for high-throughput laser-assisted micro- and nano-structuring of materials.<sup>5-7</sup> Similar effects were observed in experiments on dry laser cleaning, where such particles are used as controlled contaminants.<sup>8-10</sup> This necessitates better understanding of the focusing of light by microspheres with diameters of several wavelengths. Only a few rigorous results are available for the intermediate range of sphere sizes and distances from the particle. The majority of analytical approaches either deal with the properties of the eigenmodes, or refer to the integral characteristics and/or to the far field behavior.<sup>11</sup> Mie resonances were analyzed on the basis of advanced geo-

metrical optics<sup>12</sup> and detailed numerical calculations for transparent spheres of several wavelengths in size were performed in connection with the use of laser tweezers in biology<sup>13</sup> and for needs of aerosol science.<sup>14</sup>

In this work we develop a theoretical description for an arbitrary nonparaxial strongly aberrated axially symmetric focusing and apply it to the case of dielectric microspheres. Our emphasis is on the fine structure of the field distribution in the exterior of the sphere up to the focal region, which can be used to control and improve the concentration of energy.

Strong spherical aberration makes the focusing nontrivial. Usually, the exact solution is obtained using the Mie theory,<sup>15</sup> which does not give much of a physical insight as it requires the summation of a large number of terms in a multiple expansion even for moderate sphere sizes. At the same time, the main focusing properties of transparent dielectric microspheres originate rather from the picture of geometrical optics.

One might think that in the lowest approximation a small sphere acts as an ideal lens. However, in the range of sizes we are interested in, this picture does not even provide a description which is qualitatively correct. Also classical formulas for weak spherical aberration<sup>16</sup> do not yield useful results for the field behind a sphere: They predict that the maximum intensity is kept unchanged and its position does not depend on the wavelength.

Our approach, following the method of uniform caustic asymptotics,<sup>17</sup> is based on the canonical integral for the cuspid ray topology of strong spherical aberration. Though this Bessoid integral—a member of the hierarchy of diffraction catastrophes<sup>18,19</sup>—appears naturally in the paraxial approximation, it can be used to describe arbitrary axially symmetric strong spherical aberration by appropriate coordinate and amplitude transformations. For angularly dependent vec-

torial amplitudes the formalism uses higher-order Bessoid integrals.

The Bessoid integral is the axially symmetric generalization of the Pearcey integral,<sup>20</sup> which plays an important role in many short wavelength phenomena.<sup>21</sup> Therefore, the present approach can be applied in various areas of physics where axially symmetric focusing is of importance, e.g., acoustics, semiclassical quantum mechanics, radio wave propagation and scattering theory.

## II. THE BESSOID INTEGRAL

### A. Definition

We first consider the diffraction of a *scalar* spherically aberrated wave on a circular aperture with radius  $a$  in the plane  $z=-f$  around the  $z$  axis, where  $f$  is the focal distance. The origin of the coordinate system is put into the focus  $F$ . In cylindrical coordinates  $(\rho, z)$ , the paraxial Fresnel-Kirchhoff diffraction integral<sup>16</sup> yields the field amplitude

$$U(\rho, z) = -\frac{ikU_0}{f} e^{ikz} \int_0^a J_0\left(k \frac{\rho \tilde{\rho}_1}{f}\right) e^{-ikz\tilde{\rho}_1^2/2f^2 - ikB\tilde{\rho}_1^4} \tilde{\rho}_1 d\tilde{\rho}_1. \quad (1)$$

Here  $U_0$  is the amplitude of the incident wave in the center of the aperture,  $k$  is the wave number ( $k=2\pi/\lambda$ , where  $\lambda$  is the wavelength) and  $\tilde{\rho}_1$  is the distance from the axis on the aperture. The Bessel function  $J_0$  comes from the integration over the polar angle  $\varphi$ . The parameter  $B$  in the exponent determines the strength of the spherical aberration. For  $B>0$  the diffraction focus shifts towards the aperture, while  $B=0$  corresponds to ideal focusing.<sup>16</sup>

We introduce the dimensionless coordinates  $\rho_1 \equiv \sqrt[4]{4kB}\tilde{\rho}_1$ ,  $R \equiv \sqrt[4]{k^3/4B}\rho/f$  and  $Z \equiv \sqrt{k/4B}z/f^2$  and consider an infinitely large aperture. Then the field (1) becomes proportional to the *Bessoid integral*<sup>22</sup>

$$I(R, Z) = \int_0^\infty \rho_1 J_0(R\rho_1) e^{-i(Z\rho_1^2/2 + \rho_1^4/4)} d\rho_1 \quad (2)$$

$$= \frac{1}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty e^{i\phi} dx_1 dy_1, \quad (3)$$

where

$$\phi \equiv -Rx_1 - Z\frac{x_1^2 + y_1^2}{2} - \frac{(x_1^2 + y_1^2)^2}{4}. \quad (4)$$

Its absolute square is shown in Fig. 1. In the Cartesian representation  $x_1 = \rho_1 \cos \varphi$  and  $y_1 = \rho_1 \sin \varphi$  are dimensionless coordinates in the plane of integration. Expression (3) is the axially symmetric generalization of the Pearcey integral<sup>20</sup>

$$I_P(X, Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-i(Xx_1 + Zx_1^2/2 + x_1^4/4)} dx_1, \quad (5)$$

which is also shown in Fig. 1.

Both integrals correspond to so-called *diffraction catastrophes*.<sup>17-19</sup> Their field distribution contains caustic zones

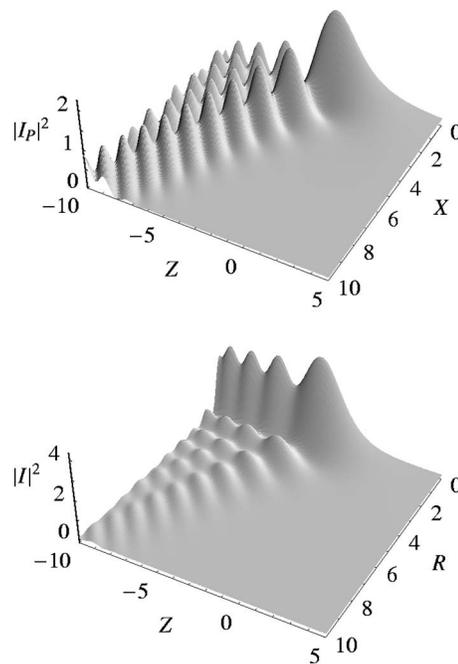


FIG. 1. Absolute square of the Pearcey integral  $I_P$  (top) and the Bessoid integral  $I$  (bottom). The latter is proportional to the field of a spherically aberrated wave within small angles approximation.

where the intensity predicted by geometrical optics goes to infinity. The Pearcey integral corresponds to a *cuspid caustic*, i.e., a single one-dimensional curve in a two-dimensional space, and does not reveal a high intensity along the axis, while the Bessoid integral corresponds to a *cuspid caustic*, i.e., to a surface of revolution of the cusp in three dimensions, as well as the caustic line up to the focus  $F$  at  $z=Z=0$ . The equation of the cusp is given by the semicubic parabola

$$27R^2 + 4Z^3 = 0. \quad (6)$$

Henceforth we will apply the term cusp also for the whole cuspid. A caustic is denoted as stable, if it does not change its topology under small perturbations. This is the case for the Pearcey integral. The Bessoid integral corresponds to a structurally unstable caustic, because an infinitely small perturbation will destroy the radial symmetry and the axis will not be a caustic zone any longer. It is, however, stable on the class of axially symmetric wave fronts.

The cusp is the envelope of the family of rays. The latter correspond to the *points of stationary phase* in the Bessoid integral, i.e., those points where the two first partial derivatives with respect to  $R$  and  $Z$  of the phase  $\phi$  in Eq. (3) vanish. Inside the cusp, for  $27R^2 + 4Z^3 < 0$ , three rays (tangents to the cusp) arrive at each point of observation  $P \equiv (\rho, z)$ , and outside, for  $27R^2 + 4Z^3 > 0$ , there is only one real ray (Fig. 2). Thus, the cusp forms the border between the lit region and the (partial) geometrical shadow, where two rays merge.

Without loss of generality, we assume that all rays lie in the meridional plane  $\varphi=0$  ( $y_1=0$ ) and hence correspond to the roots  $x_{1,j}$  ( $j=1,2,3$ ) of the cubic equation

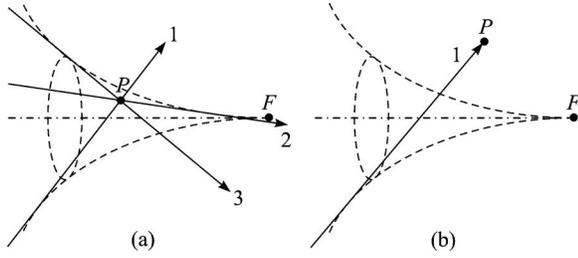


FIG. 2. (a) Three-ray region inside the cuspid (dashed line). (b) One-ray region outside. The  $z$  axis is represented by a dashed-dotted line.

$$R + Zx_1 + x_1^3 = 0, \quad (7)$$

which are given by Cardan's formulas.<sup>23</sup> On the axis,  $R=0$ , a cone formed by an infinite number of rays converges. These rays originate from the circle  $x_1^2 + y_1^2 = -Z$  on the aperture. They are all in phase and produce a high intensity along the axis (compare the two pictures in Fig. 1). The oscillations occur due to interference with the ray propagating along the  $z$  axis. One can also directly observe that the Bessoid integral has the topology of spherical aberration as the maximum of intensity does not lie in the geometrical focus  $Z=0$  but is spherically aberrated to a negative value of  $Z$ , i.e., towards the aperture.

### B. Asymptotic expressions

Off the caustic—away from the cusp and the focal line—the Bessoid integral (3) can be approximated by the method of stationary phase. As the integrand in Eq. (3) is highly oscillatory, the only significant contributions to the integral come from those regions where the phase is stationary<sup>24</sup>

$$I(R, Z) \approx \sum_{j=1}^m \frac{e^{i\phi_j + i\pi/4 \text{ sign} \mathbf{H}_j}}{\sqrt{|\det \mathbf{H}_j|}}, \quad (8)$$

where the summation runs over all real rays, i.e.,  $m=1$  (lit region) or  $m=3$  (shadow). The phase  $\phi_j$  is obtained by inserting the  $j$ th stationary point  $(x_{1,j}, y_{1,j}=0)$  into Eq. (4). The determinant and signature of the Hessian are given by

$$\det \mathbf{H}_j = Z^2 + 4x_{1,j}^2 Z + 3x_{1,j}^4, \quad (9)$$

$$\text{sign} \mathbf{H}_j = \text{sgn}(-Z - 3x_{1,j}^2) + \text{sgn}(-Z - x_{1,j}^2). \quad (10)$$

Near the cusp the Bessoid integral shows an Airy-type behavior typical for caustics where two real rays disappear and become complex.

One can also derive a different approximation valid on and near the caustic axis, i.e., for  $Z \leq 0$  and small  $R$  (Appendix A):

$$I(R, Z) \approx \frac{\sqrt{\pi}}{2} J_0(R\sqrt{-Z}) e^{i(Z^2 - \pi)/4} \text{erfc}\left(\frac{Z}{2} e^{i\pi/4}\right). \quad (11)$$

Here  $\text{erfc}$  is the complementary error function,<sup>25</sup> which can also be written in terms of Fresnel sine and cosine functions.<sup>22</sup> Expression (11) becomes exact at the axis  $R=0$ ,

where  $J_0(0)=1$ . It shows that near the axis the Bessoid integral is virtually a Bessel beam<sup>26</sup> with a variable cross section.

### C. Numerical evaluation

As the Bessoid integrand is highly oscillatory, its evaluation for the whole range of coordinates  $R$  and  $Z$  is nontrivial and of large practical importance. Direct numerical integration along the real axis and the method of steepest descent in the complex plane both have their disadvantages. By far the fastest technique is based on the numerical solution of the ordinary differential equation (derivation in Appendix B)<sup>27,28</sup>

$$L_R - ZI_R + iRI = 0. \quad (12)$$

Indices denote (partial) derivatives and  $L \equiv I_{RR} + I_R/R$  is an abbreviation for the radial Laplacian applied onto  $I$ . The three initial conditions at  $R=0$  are

$$I(0, Z) = \frac{\sqrt{\pi}}{2} e^{i(Z^2 - \pi)/4} \text{erfc}\left(\frac{Z}{2} e^{i\pi/4}\right), \quad (13)$$

$$I_R(0, Z) = 0, \quad (14)$$

$$L(0, Z) = ZI(0, Z) + i. \quad (15)$$

$I(0, Z)$  was taken from Eq. (11),  $I_R(0, Z)$  vanishes due to symmetry, and the last condition arises from the fact that the Bessoid integral satisfies the paraxial Helmholtz equation  $2iI_Z + L = 0$ , where  $I_Z$  is calculated from Eq. (13).

In the literature the Pearcey integral was calculated by solving differential equations,<sup>29</sup> by a series representation<sup>30</sup> and by the first terms of its asymptotic expansion.<sup>31</sup> The Bessoid integral was expressed in terms of parabolic cylinder functions<sup>32</sup> and as a series.<sup>22</sup> The latter work gives reference to an unpublished work of Pearcey,<sup>33</sup> stating that differential equations for the Bessoid integral were employed there.

### D. Geometrical optics for the cuspid

In geometrical optics, the rays carry the information of amplitude and phase. The total field in a point  $P$  is given by the sum of all ray fields there. A ray's field at  $P$  is determined by<sup>34</sup>

$$U(P) = U_0 \frac{e^{ik\psi}}{\sqrt{J}}, \quad (16)$$

where  $U_0$  is the amplitude at some initial wave front,  $\psi$  is the eikonal, and  $J$  is the generalized geometrical divergence, which can be calculated from flux conservation along the ray. For a homogeneous medium with constant refractive index<sup>34</sup>

$$J = \frac{R_m R_s}{R_{m0} R_{s0}}. \quad (17)$$

$R_m, R_s$  are the main radii of curvature at the point  $P$  and  $R_{m0}, R_{s0}$  are the radii on the initial wave front, where  $U=U_0$ .

When a ray touches a caustic, its radius of curvature (the geometrical divergence in the general case) changes the sign

and the ray undergoes a *caustic phase delay*<sup>17,34</sup> of  $-\pi/2$ , which is taken into account by the proper choice of the square root in Eq. (16). When a ray touches several caustics, these delays must be added. The total caustic phase shift, denoted as  $\Delta\varphi$ , can be explicitly written in the phase. For the cuspid topology and ray numbering ( $j=1,2,3$ ) according to Fig. 2, we obtain

$$U(P) = U_0 \frac{e^{ik\psi}}{\sqrt{J}} = U_0 \frac{e^{ik\psi+i\Delta\varphi}}{\sqrt{|J|}} \quad (18)$$

with

$$\Delta\varphi_j = \begin{cases} -\pi & \text{for } j=1, \\ 0 & \text{for } j=2, \\ -\pi/2 & \text{for } j=3. \end{cases} \quad (19)$$

Ray 1 touched the cuspid and the focal line, ray 2 is not shifted, and ray 3 touched the cuspid.

### III. RELATION BETWEEN GEOMETRICAL AND WAVE OPTICS

#### A. Matching with the Bessoid integral

If we have found the phases  $\varphi \equiv k\psi$  and divergences  $J$  of the rays, the (scalar) geometrical optics solution with an axially symmetric 3-ray cuspid topology can be written as

$$U(\mathbf{r}) = \sum_{j=1}^3 \frac{U_{0,j} e^{i\varphi_j(\mathbf{r})}}{\sqrt{J_j(\mathbf{r})}}. \quad (20)$$

Here  $\mathbf{r} \equiv (\rho, z)$  are the real-space coordinates and we have allowed for different initial amplitudes  $U_{0,j}$  of the rays. This field shows singularities at the caustic, especially on the axis, which is the most interesting region for applications.

We want to describe *arbitrary axially symmetric focusing* by matching the solution of geometrical optics (where it is correct) with a wave field constructed from the Bessoid integral (3), which naturally appears in the paraxial approximation and is finite everywhere, and its partial derivatives  $I_R$  and  $I_Z$  (*method of uniform caustic asymptotics*). We make the *ansatz*<sup>17</sup>

$$U = \left( AI + \frac{1}{i} A_R I_R + \frac{1}{i} A_Z I_Z \right) e^{i\chi}. \quad (21)$$

The yet unknown arguments of the Bessoid integral and its derivatives are  $\mathbf{R} \equiv (R(\mathbf{r}), Z(\mathbf{r}))$ .  $A(\mathbf{r})$ ,  $A_R(\mathbf{r})$  and  $A_Z(\mathbf{r})$  are three amplitude factors and  $\chi(\mathbf{r})$  is a phase function. (The indices  $R$  and  $Z$  in the amplitudes do not indicate derivatives.) Now the geometrical optics solution (20) is matched with the stationary phase approximation of Eq. (21) by equating the amplitudes and phases:<sup>17,35</sup>

$$\frac{U_{0,j}}{\sqrt{J_j}} = \frac{A(\mathbf{r}) + A_R(\mathbf{r})\phi_R(\mathbf{R}, \mathbf{t}_j) + A_Z(\mathbf{r})\phi_Z(\mathbf{R}, \mathbf{t}_j)}{\sqrt{H_j}}, \quad (22)$$

$$\varphi_j(\mathbf{r}) = \chi(\mathbf{r}) + \phi(\mathbf{R}, \mathbf{t}_j). \quad (23)$$

The terms  $\phi_R$  and  $\phi_Z$  are the partial derivatives of Eq. (4) and

$$\frac{1}{\sqrt{H_j}} \equiv \frac{e^{i\pi/4 \operatorname{sign} H_j}}{\sqrt{|\det \mathbf{H}_j|}}, \quad (24)$$

where the determinant and signature of the Hessian are written in Eqs. (9) and (10), respectively. Outside the cusp, the rays 2 and 3 are complex and the general definition of  $H_j$  is more subtle, namely

$$\frac{1}{\sqrt{H_j}} \equiv \sqrt{\frac{i}{\phi_{20,j}}} \sqrt{\frac{i}{\phi_{02,j}}}, \quad (25)$$

with  $\phi_{20} \equiv \partial^2 \phi / \partial x_1^2$ ,  $\phi_{02} \equiv \partial^2 \phi / \partial y_1^2$  (the index  $j$  denotes substitution of the  $j$ th point of stationary phase as argument).

The three points of stationary phase were denoted as  $\mathbf{t}_j \equiv (t_j, 0)$ , where the  $t_j$  are given by the (correctly ordered) Cardan's solutions of Eq. (7), i.e., of

$$R + Zt + t^3 = 0. \quad (26)$$

Note that they are functions of the Bessoid coordinates,  $\mathbf{t}_j = \mathbf{t}_j(\mathbf{R})$ , and the latter depend on the real space coordinates:  $\mathbf{R} = \mathbf{R}(\mathbf{r})$ . The partial derivatives with respect to  $R$  and  $Z$  in Eq. (22) must be evaluated in such a way as the  $t_j$  were held constant, although they are functions of  $\mathbf{R}$  themselves. The conditions (22) and (23) give six equations for the six unknowns  $R$ ,  $Z$ ,  $\chi$ ,  $A$ ,  $A_R$ , and  $A_Z$ .

It is convenient to solve Eq. (23), that is

$$\varphi_j = \chi - Rt_j - \frac{1}{2} Zt_j^2 - \frac{1}{4} t_j^4, \quad (27)$$

using quantities that are permutationally invariant with respect to the roots  $t_j$ .<sup>21,35</sup> This yields

$$R = \sqrt{\frac{Z^3}{54} - \frac{4b_2}{9Z}},$$

$$Z = \pm \sqrt[4]{\frac{2}{3}} \sqrt{-2 \operatorname{sgn}(b_3) \sqrt{b_2 + q} + 2\sqrt{D}},$$

$$D \equiv 2b_2 - q + 2\sqrt{b_2^2 - b_2q + q^2},$$

$$\chi = b_1 - \frac{1}{6} Z^2, \quad (28)$$

where  $\operatorname{sgn}(Z) = \operatorname{sgn}(Z^4 - 24b_2)$ . The  $b_l$  ( $l=1,2,3$ ) are given by  $b_1 \equiv (1/3)\sum_{j=1}^3 \varphi_j$ ,  $b_2 \equiv \sum_{j=1}^3 (\varphi_j - b_1)^2$  and  $b_3 \equiv \sum_{j=1}^3 (\varphi_j - b_1)^3$ . The quantity  $q$  (sometimes called discriminant) can be expressed in different ways:

$$\begin{aligned} q^3 &\equiv 6b_3^2 - b_2^3 = \frac{1}{2_{11}} R^2 (27R^2 + 4Z^3)^3 \\ &= -2(\varphi_1 - \varphi_2)^2 (\varphi_2 - \varphi_3)^2 (\varphi_3 - \varphi_1)^2. \end{aligned} \quad (29)$$

Hence, it vanishes exactly at the caustic where two phases are equal. At the cuspid  $\varphi_2 = \varphi_3$  ( $27R^2 + 4Z^3 = 0$ ) and on the axis  $\varphi_1 = \varphi_3$  ( $R=0$ ).

The solutions of Eq. (22), that is

$$\frac{U_{0,j}}{\sqrt{J_j}} = \frac{A - t_j A_R - \frac{1}{2} t_j^2 A_Z}{\sqrt{H_j}}, \quad (30)$$

are<sup>35</sup>

$$\begin{aligned} A &= -U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} \frac{t_2 t_3}{(t_3 - t_1)(t_1 - t_2)} - \dots - \dots, \\ A_R &= U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} \frac{t_1}{(t_3 - t_1)(t_1 - t_2)} + \dots + \dots, \\ A_Z &= 2U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} \frac{1}{(t_3 - t_1)(t_1 - t_2)} + \dots + \dots, \end{aligned} \quad (31)$$

where the cyclic terms permute the numbering of rays:  $(1, 2, 3) \rightarrow (2, 3, 1) \rightarrow (3, 1, 2)$ . The Bessoid matching solution (21) does not show the divergences of geometrical optics.

Note that this method utilizes also the so-called complex rays which have less apparent physical meaning. It turns out that both real and complex rays provide the *geometrical skeleton for the wave flesh*.<sup>17</sup>

### B. Expressions on and near the axis in the general case

All formulas can be strongly simplified on and near the axis inside the cuspid (small  $\rho$ ,  $z < 0$ ). The Bessoid coordinates have the simple form (Appendix C)

$$R \approx \frac{(\varphi_1 - \varphi_3)/2}{\sqrt{2^4} \sqrt{(\varphi_1 + \varphi_3)/2 - \varphi_2}} \approx \frac{k\rho \sin \beta}{\sqrt{-Z}}, \quad (32)$$

$$Z \approx -2 \sqrt{\frac{\varphi_1 + \varphi_3}{2} - \varphi_2} \approx -2 \sqrt{\varphi_{np} - \varphi_p}, \quad (33)$$

where  $\beta > 0$  is the local angle of the nonparaxial cone of rays with the axis and  $\varphi_{np}$  and  $\varphi_p$  denote the phases of the nonparaxial rays and the (par)axial ray, respectively (see Fig. 9 in Appendix C). The simple natural combination

$$R\sqrt{-Z} \approx \frac{\varphi_1 - \varphi_3}{2} \approx k\rho \sin \beta \quad (34)$$

also appears in the near axis approximation for the Bessoid integral (11). On the axis ( $\rho=0$ ,  $\varphi_1=\varphi_3$ ) we obtain  $R=0$  and  $Z=-2\sqrt{\varphi_1-\varphi_2}$ .

The results (32)–(34) have transparent physical meaning. Indeed, near the axis the largest contribution to the field comes from the converging cone of nonparaxial rays (similar to ray 1) that intersect the axis at an angle  $\beta$ . If the angle  $\beta$  is constant and all rays have the same intensity, the result is the Bessel beam.<sup>26</sup> Such beams have a propagation constant along the  $z$  direction equal to the  $z$  component of the wave vector of the plane waves which form them and correspondingly the argument of the Bessel function (cylindrical analog of a plane wave) is equal to  $k\rho \sin \beta$ . As the angle  $\beta$  gradually changes for the spherically aberrated wave, so does the argument of the Bessel function.

Additionally, there exists the axial ray 2, which is not present in the canonical Bessel beam (though it often appears in real experimental situations). The interference of this beam with the converging ray cone results in the intensity oscillations along the axis (Fig. 1, bottom). Clearly, these oscillations are largely due to the phase difference  $\varphi_{np} - \varphi_p$ . At large negative  $Z$  in Eq. (11)

$$\operatorname{erfc} \left[ \frac{Z}{2} \exp \left( i \frac{\pi}{4} \right) \right] \rightarrow 2 \left[ 1 - \frac{1}{\sqrt{\pi Z}} \exp \left( i \frac{3\pi - Z^2}{4} \right) \right],$$

and the oscillating behavior is governed by the phase of the exponent, which is equal to  $3\pi/4 - (\varphi_{np} - \varphi_p)$ . This clarifies the origin of expression (33), as it is  $Z^2$  entering the final formulas.

In particular, the global maximum is expected on axis at the first constructive interference of the axial and the nonparaxial rays. Because  $Z < 0$  in this region, the two terms of the erfc expansion are first in phase when the phase difference is  $\varphi_1 - \varphi_2 = 3\pi/4$ . The geometrical meaning of this result is that rays 1 and 3 are shifted by  $-\pi/2$  as they touch the cusp. In addition, they acquire a further shift of  $-\pi/2$  when crossing the focal line. But exactly *on the axis* only half of this delay has occurred yet, which yields the  $3\pi/4$  difference. The numerical maximum of the Bessoid intensity (absolute square) occurs at  $Z_m \approx -3.051$  and hence this yields the condition

$$\varphi_1 - \varphi_2 = \frac{Z_m^2}{4} \approx 2.327, \quad (35)$$

which is close to  $3\pi/4 \approx 2.356$ .

The width of the focal line caustic,  $\rho_w$ , is defined by the first zero  $w_0 \approx 2.405$  of the Bessel function in Eq. (11). Hence, with Eq. (34)

$$\rho_w \equiv \frac{w_0}{k \sin \beta} \approx 0.383 \frac{\lambda}{\sin \beta}. \quad (36)$$

In the geometrical optics picture the first minimum occurs when rays 1 and 3 interfere destructively, i.e., when their phase difference becomes  $\pi$ . This results in  $\varphi_1 - \varphi_3 = \pi + \pi/2$ , where the term  $\pi/2$  takes into account the caustic phase shift of ray 1:  $\rho_w \approx (\varphi_1 - \varphi_3)/2k \sin \beta = 0.375\lambda/\sin \beta$ . Note, that this is smaller than the Airy spot for the same aperture angle<sup>16</sup> and large angles  $\beta$  are indeed realized, e.g., in the case of the sphere studied below.

Finally, we present an expression for the field (21) on the axis. The equations for the amplitudes (31) simplify tremendously (Appendix D) and result in

$$U = \left[ \frac{U_{0,1} \sqrt{2k\rho \sin \beta}}{\sqrt{J_1}} \left( iI - \frac{1}{Z} \right) + \frac{U_{0,2}}{\sqrt{J_2}} \right] e^{i\varphi_2}. \quad (37)$$

The structure of expression (37) helps to understand its physical meaning. It details the contribution of the cone of nonparaxial rays, represented by ray 1 (first term), and the axial ray 2 (second term) to the overall structure of the field. Note that in the general case not only the angle  $\beta$ , but also the amplitude of the converging cone may vary along  $z$  ( $Z$ ), thus slowly modifying the properties of the Bessel beam in

the axial region. This enters Eq. (37) via amplitude transformations and is manifested by the presence of the initial ray amplitudes in both terms. Inside the cusp on the axis ( $z, Z < 0$  and  $\rho, R \rightarrow 0$ ) both  $1/\sqrt{J_2}$  and the ratio  $\sqrt{\rho}/\sqrt{J_1}$  remain finite as the divergence of the paraxial ray 2 is nonsingular, while the sagittal divergence  $J_1$  of the cone of nonparaxial rays 1 is proportional to  $\rho$ . Due to the Bessoid matching procedure the singularity of the converging cone is removed by the compensating factor  $\sqrt{\rho}$ . Along the axis the last term in Eq. (37) partly cancels with the second term in the parentheses of the first term. As a result, the on axis field behavior up to the focus is dominated by a single term proportional to the Bessoid integral  $I$ , which justifies the maximum condition (35) discussed above.

### C. Angular dependencies and vectorial problems: Higher-order Bessoid matching

Often—especially in *vectorial* problems—there exists axial symmetry with respect to the wave fronts, ray phases and generalized divergences, but not with respect to the amplitudes. In this case, new functions are required to represent arbitrary angular dependence of the field. The natural generalizations of Eq. (3) are the *higher-order Bessoid integrals*<sup>32</sup> with the non-negative integer  $m$ :

$$I_m(R, Z) = \int_0^\infty \rho_1^{m+1} J_m(R\rho_1) e^{-i(Z\rho_1^2/2 + \rho_1^4/4)} d\rho_1, \quad (38)$$

where  $I_0 \equiv I$  and  $J_m$  are higher-order Bessel functions. The higher-order Bessoid integrals obey the recurrence relation

$$I_{m+1} = -I_{m,R} + m \frac{I_m}{R}. \quad (39)$$

The integral  $I_m$  is canonical for angular dependent geometrical field components  $U^{(m)}(\rho, z) \sin m\varphi$  or  $U^{(m)}(\rho, z) \cos m\varphi$ . In matching similar to Eq. (21),

$$U^{(m)} = \left( A_m I_m + \frac{1}{i} A_{m,R} I_{m,R_m} + \frac{1}{i} A_{m,Z} I_{m,Z_m} \right) e^{i\chi_m}, \quad (40)$$

the angular dependence cancels. Here  $A_m$ ,  $A_{m,R}$ , and  $A_{m,Z}$  are the higher-order amplitude factors, whereas  $I_{m,R_m}$  and  $I_{m,Z_m}$  are partial derivatives of the higher-order Bessoid integrals  $I_m$ . Since the latter can be written in terms of  $I_0$ , it can be shown that the points of stationary phase, the matching of phases and thus the higher-order coordinates ( $R_m$ ,  $Z_m$ ) and phases ( $\chi_m$ ) are identical with the original ones:

$$R_m = R, \quad Z_m = Z, \quad \chi_m = \chi. \quad (41)$$

From the physical point of view, this reflects the *conservation of the wavefront* and thus the ray phases and divergences.

The equations for the amplitudes have to be generalized. The higher-order amplitudes  $A_m$ ,  $A_{m,R}$ , and  $A_{m,Z}$  have the same form as Eq. (31), but with an additional factor  $(it_j)^m$  in each denominator, i.e.,

$$A_m = -U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} \frac{t_2 t_3}{(it_1)^m (t_3 - t_1)(t_1 - t_2)} - \dots - \dots,$$

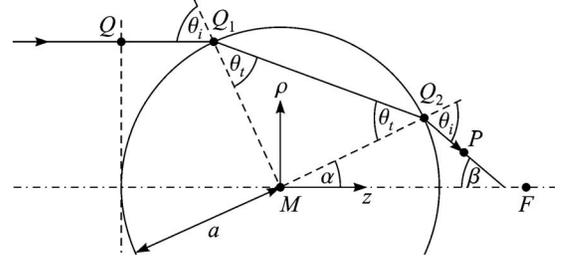


FIG. 3. Refraction of a ray—propagating from  $Q$  to  $P$ —by a sphere with radius  $a$  and refractive index  $n$ . The picture is drawn in the meridional plane and all indicated angles are positive.

$$A_{mR} = U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} \frac{t_1}{(it_1)^m (t_3 - t_1)(t_1 - t_2)} + \dots + \dots,$$

$$A_{mZ} = 2U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} \frac{1}{(it_1)^m (t_3 - t_1)(t_1 - t_2)} + \dots + \dots. \quad (42)$$

A more detailed description of the higher-order Bessoid integrals as well as the derivation of the recurrence relation and the amplitude equations can be found in Appendix E.

## IV. THE SPHERE

### A. Geometrical optics solution

Consider a plane wave falling on a transparent sphere in vacuum. Figure 3 illustrates the refraction of a single ray in the meridional plane, containing the point of observation  $P$  and the axis. Within the frame of geometrical optics the cuspid is formed behind the sphere in analogy to Fig. 2.

Let  $a$  be the sphere radius and  $n > 1$  its refractive index. In contrast to the previous sections, we choose the origin of the axially symmetric cylindrical coordinate system  $(\rho, z)$  differently now, namely as the center  $M$  of the sphere. The incident plane wave propagates parallel to the  $z$  axis. The geometrical optics focus, formed by the paraxial rays, is located at  $F \equiv (0, f)$  with<sup>36</sup>

$$f \equiv \frac{a}{2} \frac{n}{n-1}. \quad (43)$$

A ray passes the point  $Q$ , is first refracted at  $Q_1$ , a second time at  $Q_2$  and propagates to  $P$ . The incident and transmitted angle,  $\theta_i$  and  $\theta_t$ , are related by Snell's law,  $\sin \theta_i = n \sin \theta_t$ . Writing the position of  $P \equiv (\rho, z)$  in polar coordinates,  $\rho = l \sin \theta$  and  $z = l \cos \theta$ , one can find the following expression, determining the three rays that arrive at  $P$ :

$$l \sin(\theta + 2\theta_i - 2\theta_t) = a \sin \theta_i, \quad (44)$$

where one has to substitute  $\theta_t = \arcsin[(\sin \theta_i)/n]$ . This is a transcendental cubic-like equation which has three roots, either all real or one real and two complex conjugate. (For  $n \geq \sqrt{2}$  this is true for  $z \geq a$ ; if  $n < \sqrt{2}$ , the three-ray region does not start until some distance behind the sphere.) We denote them as  $\theta_{i,j}$  ( $j=1, 2, 3$ ) and choose their order consistently with the previous notations. Therefore,  $\theta_{i,1}$  is always real and

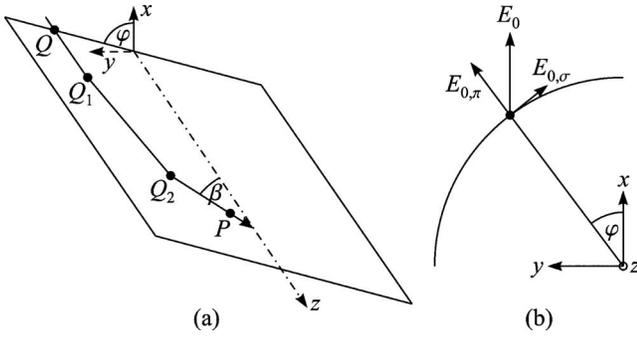


FIG. 4. (a) A ray propagates from  $Q$  to  $P$  in the meridional plane (plane of incidence). (b) Decomposition of the initial electric field vector with length  $E_0$  into its  $\pi$ - and  $\sigma$ -component parallel and perpendicular to the meridional plane.

negative, whereas  $\theta_{i,2}$  and  $\theta_{i,3}$  are either real and positive (lit region) with  $\theta_{i,2} < \theta_{i,3}$  or complex conjugate (geometrical shadow).

When the  $\theta_{i,j}$  are known, we find the  $\theta_{i,j}$  from Snell's law and the  $\alpha_j$  and  $\beta_j$  from

$$\alpha = 2\theta_i - \theta_i, \quad \beta = 2\theta_i - 2\theta_i. \quad (45)$$

Omitting the index  $j$ , the three ray coordinates can be written as

$$s \equiv \overline{Q_2P} = \frac{l \cos \theta - a \cos \alpha}{\cos \beta}. \quad (46)$$

The eikonal is the optical path accumulated from  $Q$  to  $P$  (on the dashed vertical line in Fig. 3 all rays are still in phase):

$$\psi = \overline{QQ_1} + n\overline{Q_1Q_2} + \overline{Q_2P} - a = a(2n \cos \theta_i - \cos \theta_i) + s. \quad (47)$$

The sphere radius  $a$  was subtracted from the path contributions to make the eikonal zero in the center  $M$ , if there were no sphere.

Next we calculate the geometrical optics amplitudes by determining the meridional and sagittal radii of curvature,  $R_m$  and  $R_s$ , and their changes due to refraction. Formulas for the refraction on an arbitrary surface with arbitrary orientation of the main radii exist in the literature.<sup>34,37</sup> A simple derivation for the sphere can be found in Appendix F. It yields the dependence of the actual radii of curvature  $R_m$  and  $R_s$  (right after the refraction) on the initial radii  $R_{m0}$  and  $R_{s0}$  (just before the refraction):

$$R_m = \frac{naR_{m0} \cos^2 \theta_i}{a \cos^2 \theta_i + R_{m0}(\cos \theta_i - n \cos \theta_t)}, \quad (48)$$

$$R_s = \frac{naR_{s0}}{a + R_{s0}(\cos \theta_i - n \cos \theta_t)}. \quad (49)$$

For a plane wave,  $R_{m0}, R_{s0} \rightarrow \infty$ , the radii of curvature in the points  $Q_1$  (inside the sphere) and  $Q_2$  (outside the sphere) have the compact form

$$R_{m,Q_1} = -a \frac{\sin \theta_i \cos^2 \theta_t}{\sin(\theta_i - \theta_t)}, \quad (50)$$

$$R_{m,Q_2} = -a \frac{\cos \theta_t}{2} \left( \frac{\cos \theta_t \sin \theta_t}{\sin(\theta_i - \theta_t)} - 1 \right), \quad (51)$$

$$R_{s,Q_1} = -a \frac{\sin \theta_i}{\sin(\theta_i - \theta_t)}, \quad (52)$$

$$R_{s,Q_2} = -a \frac{\sin(2\theta_t - \theta_t)}{\sin(2\theta_t - 2\theta_t)}. \quad (53)$$

The overall geometrical generalized divergence after both refractions reads (index  $j$  omitted)

$$\frac{1}{\sqrt{J}} = \frac{\sqrt{R_{m,Q_1}R_{s,Q_1}}}{\sqrt{(R_{m,Q_1} + d)(R_{s,Q_1} + d)}} \frac{\sqrt{R_{m,Q_2}R_{s,Q_2}}}{\sqrt{(R_{m,Q_2} + s)(R_{s,Q_2} + s)}}, \quad (54)$$

where  $d \equiv 2a \cos \theta_t$  is the distance of propagation within the sphere. Note that ray 1 has a negative angle  $\theta_i$ . Besides, a double caustic phase shift should be added (manually) to the phase of this ray (minus sign) as in Eq. (18). The caustic shifts of the rays 2 and 3 are taken into account automatically if the branch cut for the square roots in Eq. (54) is along the negative real axis from  $-\infty$  to 0 and the branch with  $\sqrt{-1} = +i$  is used. In this procedure it is not allowed to multiply the radicands and write them under one common square root. Also the case of complex rays 2 and 3 is covered correctly by this convention.

Finally, the geometrical optics solution for the sphere is given by Eq. (20), where the eikonal  $\psi$  and divergence  $J$  are given by Eqs. (47) and (54). The equation determining the three rays is Eq. (44). In the geometrical shadow the sum (20) becomes only the term with  $j=1$ .

To incorporate Fresnel transmission coefficients, we assume that the incident light is linearly polarized in  $x$  direction, i.e., the incident electric field vector is

$$\mathbf{E}_0 = E_0 \mathbf{e}_x, \quad (55)$$

with  $\mathbf{e}_x$  the unit vector in  $x$  direction and  $E_0 \equiv U_0$ . Since axial symmetry is broken, we introduce the polar angle  $\varphi$  which is measured from  $x$  to  $y$ . The point of observation  $P \equiv (\rho, \varphi, z)$  will be reached by three rays (two may be complex) and their angles  $\theta_{i,j}$  are still determined by Eq. (44), for all three rays lie in the meridional plane, containing  $P$  and the  $z$  axis [Fig. 4(a)]. The initial  $\pi$ - and  $\sigma$ -polarized components depend on  $\varphi$  [Fig. 4(b)]:

$$E_{0,\pi} = E_0 \cos \varphi, \quad (56)$$

$$E_{0,\sigma} = E_0 \sin \varphi. \quad (57)$$

We define the overall transmission coefficients

$$T_\pi \equiv t_{12,\pi} t_{21,\pi} = 1 - r_{12,\pi}^2, \quad (58)$$

$$T_\sigma \equiv t_{12,\sigma} t_{21,\sigma} = 1 - r_{12,\sigma}^2. \quad (59)$$

Here the  $t_{12}$  ( $r_{12}$ ) are the standard Fresnel transmission (reflection) coefficients<sup>16</sup> from the medium 1, i.e., vacuum, into the medium 2, i.e., the sphere.

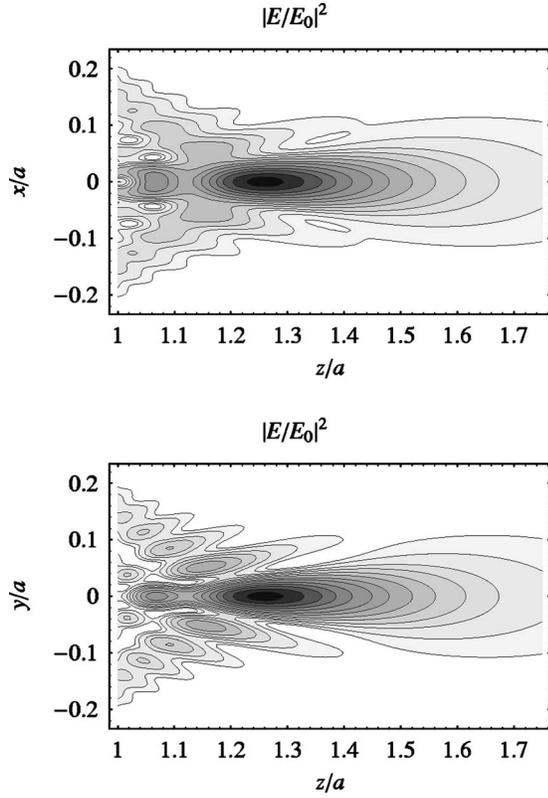


FIG. 5. Normalized intensity  $|E/E_0|^2$  in the normalized  $x,z$  plane (top) and in the  $y,z$  plane (bottom). Contour shadings go from white (zero) to black ( $\approx 700$ ). Parameters: refractive index  $n=1.5$ , dimensionless wave number  $ka=100$ . The initial electric field vector is  $\mathbf{E}_0=E_0\mathbf{e}_x$ . The sphere with radius  $a$  is positioned at the origin of the coordinate system. The focus of geometrical optics is located at  $z=f=1.5a$ , whereas the diffraction focus (the point of maximum intensity) is significantly shifted towards the sphere:  $f_d \approx 1.25a$ . In dimensional units, for a wavelength of  $\lambda=0.248 \mu\text{m}$  the sphere radius is  $a \approx 4 \mu\text{m}$ .

The ray field behind the sphere is found by the projection onto the original Cartesian system  $(x,y,z)$ . We write the components of the transmission vector  $\mathbf{T} \equiv (T_x, T_y, T_z)$  and show the ray index  $j=1, 2, 3$  explicitly. The  $\varphi$  dependence is indicated with the superscript  $(m)$ :

$$\begin{aligned} T_{x,j} &= T_j^{(0)} + T_j^{(2)} \cos 2\varphi, & T_j^{(0)} &\equiv \frac{T_{\pi,j} \cos \beta_j + T_{\sigma,j}}{2}, \\ T_{y,j} &= T_j^{(2)} \sin 2\varphi, & T_j^{(1)} &\equiv T_{\pi,j} \sin \beta_j, \\ T_{z,j} &= T_j^{(1)} \cos \varphi, & T_j^{(2)} &\equiv \frac{T_{\pi,j} \cos \beta_j - T_{\sigma,j}}{2}. \end{aligned} \quad (60)$$

Hence, the geometrical optics solution for the electric field  $\mathbf{E} \equiv (E_x, E_y, E_z)$ —including the eikonal  $\psi$  (47) and divergence  $J$  (54)—reads

$$\begin{aligned} E_{x,j} &= E_j^{(0)} + E_j^{(2)} \cos 2\varphi, \\ E_{y,j} &= E_j^{(2)} \sin 2\varphi, \\ E_{z,j} &= E_j^{(1)} \cos \varphi, \end{aligned} \quad E_j^{(m)} \equiv E_0 \frac{T_j^{(m)} e^{ik\psi_j}}{\sqrt{J_j}}. \quad (61)$$

## B. Bessoid matching solution

Matching each term  $E^{(m)} = \sum_{j=1}^3 E_j^{(m)}$  by the ansatz (21) in its higher-order formulation (40) with the corresponding integral  $I_m$ , we obtain the vectorial electric field  $\mathbf{E} \equiv (E_x, E_y, E_z)$  in the form

$$\mathbf{E} = E^{(0)} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + E^{(1)} \begin{pmatrix} 0 \\ 0 \\ \cos \varphi \end{pmatrix} + E^{(2)} \begin{pmatrix} \cos 2\varphi \\ \sin 2\varphi \\ 0 \end{pmatrix}. \quad (62)$$

Figure 5 illustrates the intensity, i.e., the absolute square of the electric field  $|E|^2 \equiv \mathbf{E}\mathbf{E}^*$ , for  $\varphi=0$  ( $x,z$  plane) and  $\varphi=\pi/2$  ( $y,z$  plane).

The magnetic field  $\mathbf{H}$  can be calculated similarly (incident magnetic field  $\mathbf{H}_0 = H_0\mathbf{e}_y$ ,  $H_0 = E_0$ ) and the (normalized) Poynting vector is given by  $\mathbf{S} \equiv \text{Re}(\mathbf{E} \times \mathbf{H}^*)$ .

## C. On the axis

On the axis the electric field is given by its  $x$  component only (direction of polarization) due to averaging over  $\varphi$  in Eq. (62). For  $z < f$  (inside the cusp) it is given by the analytical expression (37). After several simplifications<sup>27</sup> it can be written as

$$E = E_0 \left[ T_1 D_1 \left( iI - \frac{1}{Z} \right) + \frac{T_2}{1 - z/f} \right] e^{i\varphi_2}, \quad (63)$$

where the transmission factors  $T_j \equiv T_j^{(0)}$  are given in Eq. (60) and for dielectric spheres have the form

$$T_1 = \frac{n(1 + 3 \cos \beta_1) \cos \theta_{i,1} \cos \theta_{t,1}}{(n \cos \theta_{i,1} + \cos \theta_{t,1})^2}, \quad (64)$$

$$T_2 = \frac{4n}{1 + n^2}. \quad (65)$$

The phases in the coordinate  $Z = -2\sqrt{\varphi_1 - \varphi_2}$  are

$$\varphi_1 = \varphi_3 = ka \left( 2n \cos \theta_{i,1} - \cos \theta_{i,1} + \frac{\sin \alpha_1}{\sin \beta_1} \right), \quad (66)$$

$$\varphi_2 = 2ka(n-1) + kz \quad (67)$$

and  $D_1 \equiv \sqrt{\varphi_1 - \varphi_3} / \sqrt{J_1}$  is the first ray's *compensated sagittal divergence*:

$$\begin{aligned} D_1 &= - \frac{\sqrt{(R_{m,Q_1})_1 (R_{s,Q_1})_1}}{\sqrt{[(R_{m,Q_1})_1 + d_1][ (R_{s,Q_1})_1 + d_1]}} \\ &\times \frac{\sqrt{(R_{m,Q_2})_1 (R_{s,Q_2})_1}}{\sqrt{(R_{m,Q_2})_1 + s_1}} \sqrt{2k \sin \beta} \\ &= -2\sqrt{2ka \cos(\beta_1/2)} \\ &\times \sqrt{\frac{\cot \theta_{i,1} \cos \theta_{t,1} \sin(\beta_1/2)}{1 + 1/n^2 - 3 \sin^2(\beta_1/2) / \sin^2 \theta_{i,1}}}, \end{aligned} \quad (68)$$

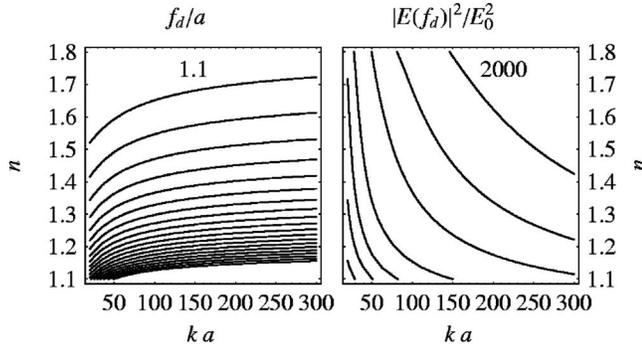


FIG. 6. Left: Diffraction focus in units of the sphere radius as a function of  $n$  and  $ka$  (contour lines from top to bottom go from 1.1 to 3.0 in steps of 0.1). Right: Intensity enhancement at  $f_d$  (contour lines from bottom left to top right are 20, 50, 100, 200, 500, 1000 and 2000).

which manifestly has no singularity until the geometrical focus where  $(R_{m,Q_2})_1 + s_1 \rightarrow 0$ . The minus sign comes from the manually inserted phase shift of the first ray. (All aforementioned quantities should be expressed in terms of the angles  $\theta_{i,1}$  and  $\theta_{i,2} \equiv 0$  as described in detail in Sec. IV D below.) The structure of the first two lines in Eq. (68) is general and is valid for arbitrary axially symmetric systems.  $D_1$  is always finite on the axis, since both the sagittal radius of curvature and the phase difference  $\varphi_1 - \varphi_3$  are proportional to the distance  $\rho$ .

Equation (63) is valid even near the focus, since the diverging terms  $D_1/Z$  and  $(1-z/f)^{-1}$  almost cancel. For  $z \rightarrow f$ , however, the divergence of  $D_1$  itself becomes important, as the nonparaxial ray 1 becomes axial.

In Fig. 6 we show the position and the value of the maximum of  $|E|^2$  as a function of the refractive index and the dimensionless product  $ka$ , calculated from Eq. (63). The  $z$  coordinate of this global maximum is denoted with  $f_d$  (diffraction focus) and the intensity there is  $|E(f_d)|^2$ . Contrary to the square dependence for the case of an ideal lens,<sup>16</sup> even for macroscopic spheres the maximum intensity turns out to be about proportional to  $ka$ , in agreement with the general theory.<sup>17</sup>

The main contribution in Eq. (63) stems from the Bessoid integral, that is from the term  $\propto T_1 D_1 I$ . Thus, the position of the maximum can be estimated from condition (35), i.e.,  $\varphi_1 - \varphi_2 \approx 3\pi/4$ . If the phase difference  $\varphi_1 - \varphi_2$  from Eqs. (66) and (67) is expressed as a function of  $\theta_{i,1}$ , Taylor expanded and equated to  $3\pi/4$ , then we get in the lowest nontrivial order of the inverse product  $ka$

$$f_d \approx \frac{a}{2n-1} \frac{n}{n-1} \left( 1 - \sqrt{\frac{3\pi n(3-n)-1}{4ka n(n-1)}} \right). \quad (69)$$

Hence, in the limit of small wavelengths or large spheres the relative difference between the diffraction and the geometrical focus decreases proportionally to the inverse square root of  $ka$ . The factor  $3\pi/4 \approx 2.356$  can be replaced by the more exact Bessoid value 2.327 from Eq. (35). Expression (69) approximates the position of the maximum within an error of

$<5\%$  for  $ka > 100$  and values of the refractive index in the range  $1.4 < n < 1.6$ . The transcendental phase difference condition (35), which holds for large angles, naturally has a wider range of applicability. With very good accuracy the diffraction focus also provides the maximum for the absolute square of the magnetic field,  $|H|^2 \equiv \mathbf{H}\mathbf{H}^*$ , as well as for the  $z$  component of the Poynting vector  $\mathbf{S}$ . Note that on the axis  $\mathbf{H} = H\mathbf{e}_y$  and  $\mathbf{S} = S\mathbf{e}_z$ .

#### D. Protocol for the electromagnetic field calculation behind the sphere

For convenience we summarize the sequence of steps that should be used for the calculation of the field behind a sphere irradiated with linearly polarized light on the basis of the formulas developed above:

(1) *Finding the rays*: The origin of the coordinate system is in the center of the sphere. Choose a point  $P \equiv (\rho, \varphi, z)$  behind the sphere and numerically calculate the three rays arriving at  $P$ . These rays are characterized by the three incident angles  $\theta_{i,j}$  ( $j=1,2,3$ ), found numerically from Eq. (44), and numbered according to Fig. 2. All other angles follow from Snell's law and from Eq. (45), respectively. Outside the cuspid the rays 2 and 3 are complex.

(2) *The geometrical optics solution*: Compute the geometrical optics solution for the electric field (61), which is the sum of contributions from three rays. The eikonals  $\psi_j$  are calculated from Eq. (47) with Eq. (46). The geometrical optics amplitudes are given by the Fresnel transmission components  $T_j^{(m)}$  (60) and the generalized divergence factors  $1/\sqrt{J_j}$  (54), which follow from the radii of curvature (50)–(53) and the distances  $s_j$  from the sphere to  $P$  (46). The conventions for the complex roots shall correctly add up all individual caustic phase shifts. Ignore the fact that the geometrical field diverges near caustic regions.

(3) *Bessoid matching*: Starting from the eikonals  $\psi_j$ , determine the Bessoid coordinates, i.e., first  $Z$ , and then  $R$  and  $\chi$  (28). Next, compute and correctly order the points of stationary phase  $t_j$  (26), most conveniently using trigonometric formulas.<sup>23</sup> With the generalized divergences (54) and the Hessians (25) the Bessoid amplitudes  $A_m$ ,  $A_{mR}$  and  $A_{mZ}$  can be computed from Eq. (42) for all orders  $m=0,1,2$ . The electric field component  $E^{(m)}$  associated with the  $m$ th order results from the ansatz (40). The Bessoid-matched field  $\mathbf{E}$  is finally given by Eq. (62). In the case of a scalar plane wave (or exactly on the axis) only the order  $m=0$  contributes. Proceed accordingly for the magnetic field  $\mathbf{H}$ , employing different transmission coefficients in step 2.

(4) *The Bessoid integral*: The final solution (62) contains the Bessoid integral and higher-order Bessoid integrals as well as their partial derivatives. The Bessoid integral  $I(R, Z)$  can be efficiently computed numerically via the differential Eq. (12). The higher-order Bessoid integrals follow from a recursive relation (39).

(5) *Remarks*: The speed limiting bottleneck of this procedure is the finding of the rays  $\theta_{i,j}$  in step 1. The numerical evaluation of the Bessoid integral is very efficient and in all other steps analytical expressions are applied. When

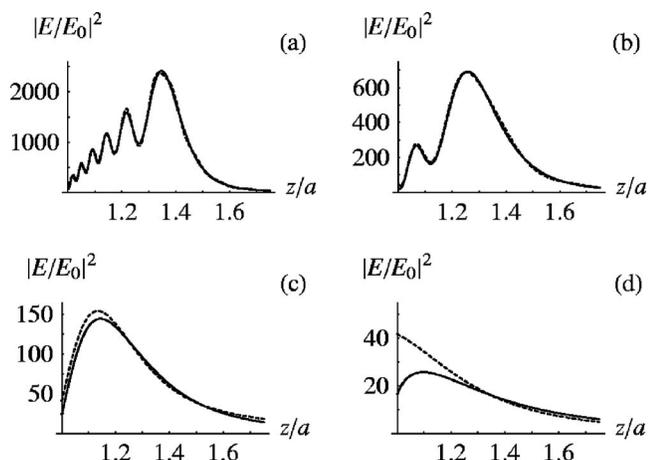


FIG. 7.  $|E/E_0|^2$  on the axis. Dashed lines represent the Mie theory, solid lines are the results of Bessoid matching. The parameters are as in Fig. 5 and the cases (a), (b), (c), and (d) correspond to  $q=300, 100, 30$ , and  $10$ , respectively. In dimensional units, for  $\lambda=0.248 \mu\text{m}$  this corresponds to sphere radii of  $a \approx 12, 4, 1.2$ , and  $0.4 \mu\text{m}$ .

approaching the caustic (axis, cuspid), individual quantities—the geometrical optics amplitudes—diverge but their combinations remain finite. In our calculations we observed perfect numerical stability up to distances from the caustic of the order of  $10^{-5}$  times the sphere radius, which is more than enough for any practical purposes.

### E. Comparison with the theory of Mie

We presented a general way to match geometrical optics solutions with the Bessoid integrals. It can be applied to any axially symmetric system with the cuspid topology of spherical aberration.

For the sphere we can compare our approximate results with the theory of Mie.<sup>15</sup> A main quantity characterizing the sphere is the dimensionless Mie parameter  $q \equiv ka$ . Figure 7 compares the intensity on the axis obtained from the Mie theory with the Bessoid approximation. The parameters are as in Fig. 5 and the Mie parameter is  $q=300, 100, 30$  and  $10$ .

We see very good agreement down to  $q \approx 30$  ( $a/\lambda \approx 4.8$ ). For  $q=10$  ( $a/\lambda \approx 1.6$ ) the asymptotic behavior far from the sphere is still correct. However, for small  $q$  the characteristic scale  $a$  is no longer large compared to the wavelength  $\lambda$  and geometrical optics becomes invalid.

Next, we compare the off-axis electric and magnetic field as well as the  $z$  component of the Poynting vector,  $S_z$  (Fig. 8). Right behind the sphere ( $z=a$ ) the agreement is not perfect (see Fig. 7), though all qualitative features are preserved. Sections at  $z=1.02a$  already show good agreement (Fig. 8) and for  $z \approx 1.05a$  the pictures become visually almost indistinguishable. Discussing the quality of these results, one has to differentiate between the accuracy of the method and the influence of those factors which can be taken into account, but were not included into the current consideration.

The accuracy of the Bessoid matching procedure itself was studied separately for the case of a spherically aberrated wave incident onto an aperture. The deviation—defined as

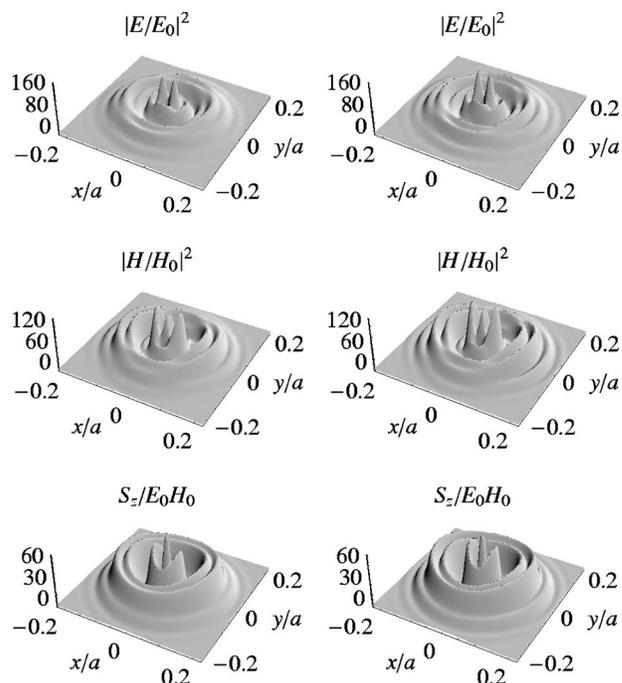


FIG. 8. Normalized  $|E|^2$ ,  $|H|^2$  and  $S_z$  in the normalized  $x, y$  plane for  $z=1.02a$  calculated with Bessoid matching (left) and with the Mie theory (right). The parameters are the same as in Fig. 5.

the maximal relative error of the intensity ( $(|E_{\text{Bessoid}}|^2 - |E_{\text{exact}}|^2)/|E_{\text{exact}}|^2$ )—between the Bessoid matched geometrical optics solution and the corresponding (exact) Rayleigh-Sommerfeld diffraction integral<sup>16</sup> decreases as the aperture increases. If the aperture is large enough, the deviation is below  $10^{-3}$  for spherical aberration strengths and wave vectors approximately corresponding to the focusing by spheres studied in Figs. 7 and 8.

For the sphere, for all investigated Mie parameters  $30 \leq q \leq 300$ , the deviations from the exact Mie solution are about  $\pm 5\%$  in those regions where the intensity is not very small (including the caustic axis and cuspid). A detailed analysis indicates that this deviation originates from several factors:

(i) Influence of the finite size of the sphere. There exist diffractive contributions from creeping rays<sup>38</sup> propagating along the sphere surface. They can in principle be accounted for at the expense of the simplicity of the procedure, essentially by considering the interference of the Bessoid field with these additional rays. This results in oscillations which can actually be seen in the Mie curve on the right side of the Bessoid tail in Figs. 7(a) and 7(b).

(ii) Rays entering the sphere undergo multiple interior reflections. Some of them satisfy resonance conditions, accumulate significant energy inside the sphere and refract outside. This produces an additional field behind the sphere, in particular the intensity becomes nonzero in the regions of geometrical shadow for the directly transmitted rays used in the Bessoid matching. Here again, one can (in principle) study such multiply reflected rays separately and add them to the Bessoid field, but their contribution to the focusing properties of the microspheres is of secondary importance.

(iii) Finally, very close to the sphere surface at distances of the order of a fraction of  $\lambda$ , there exist evanescent contributions, which are taken into account in the Mie theory, but are obviously absent in the Bessoid matching procedure.

Thus, the quality of the Bessoid matching in the most interesting regions near caustic surfaces is quite satisfactory. It rectifies the divergences of geometrical optics, which is asymptotically correct for large  $q$  in nonsingular regions of space. Clearly, the procedure has to be extended whenever contributions from rays other than those three used for matching become significant.

The field distribution behind the sphere has a rich fine structure (Fig. 8) which our geometrical approach helps to clarify. It is known that the ring-type field enhancement corresponds to the cuspid caustic, having the approximate radial distance<sup>10</sup>

$$\rho_c = a \frac{(4 - n^2)^{3/2}}{3\sqrt{3}n^2}. \quad (70)$$

Our approach explains the double-peak structure of  $|E|^2$  along the direction of polarization. It is related to the axial field component  $E_z$  and can be understood in terms of geometrical optics. On the axis, the  $E_z$  components from the rays 1 and 3 point into opposite directions and cancel, having an effective phase difference of  $\pi$ . Off the axis, ray 1 underwent a caustic phase shift of  $-\pi/2$  when crossing the axis, which makes the condition for constructive interference:  $\varphi_1 - \varphi_3 = 3\pi/2$ . Then, according to Eq. (34), the peak occurs at the radial distance

$$\rho_p \approx \frac{\varphi_1 - \varphi_3}{2k \sin \beta} = \frac{3}{8} \frac{\lambda}{\sin \beta}. \quad (71)$$

More details on the derivation are given elsewhere<sup>27</sup> together with the refined coefficient 0.293 (instead of  $3/8$ ) obtained from the Bessoid asymptotic (E8).

Double-peak structures have been observed in nanopatterning experiments<sup>8,39,40</sup> and were semiquantitatively explained on the basis of the Mie solution.<sup>9</sup> In an actual experiment it may depend on the laser pulse parameters and the properties of the patterned material whether the Poynting vector or the electric field is responsible for the patterning process. For small spheres this double peak effect can be understood using the near field pattern for a scattering dipole.<sup>9,10</sup> The present explanation (for sphere diameters of a few wavelengths and larger) results in the same orientation of the maxima and thus these two limiting cases cover almost all range of sphere sizes. Similar polarization dependence of the field distribution in focal regions can be used to improve the resolution.<sup>41</sup>

## V. CONCLUSIONS

We described theoretically arbitrary axially symmetric aberrated focusing and studied light focusing by microspheres as an example. Following the method of uniform caustic asymptotics,<sup>17</sup> we introduced a canonical integral describing the wave field for the given cuspid ray topology. This Bessoid integral appears naturally in the paraxial approxima-

tion. In some regions (off the caustic or exactly on the axis) it reduces to simple analytical expressions. In other regions we efficiently computed this highly oscillatory integral via a single ordinary differential equation.

For arbitrary axially symmetric focusing, coordinate and amplitude transformations match the Bessoid wave field and the solution of geometrical optics. The caustic divergences of the latter are removed thereby. For vectorial problems with angularly dependent field components, higher-order Bessoid integrals are used for the matching procedure. The formulas significantly simplify on and near the axis. An approximate universal condition for the diffraction focus can be given in terms of phase differences. Here, the concept of caustic phase shifts is of main importance.

The central part of the Bessoid integral is essentially a Bessel beam<sup>26</sup> with a variable cross section due to the variable angle of the nonparaxial rays. Its local diameter is always smaller than in the focus of an ideal lens with the same numerical aperture. Besides, the largest possible apertures can be physically realized, which is hardly possible with lenses. All this is achieved at the expense of longitudinal confinement.

As an example the focusing of a linearly polarized plane wave by a transparent sphere is studied in detail. We calculate the geometrical optics eikonals and divergences, incorporate Fresnel transmission coefficients and perform Bessoid matching. Using the general theory, simple expressions for the light field on the axis and for the diffraction focus are derived. The two strong maxima in the intensity observed immediately behind the sphere can be explained as well.

Finally, the results of the Bessoid matching procedure are compared with the Mie theory. The agreement is good for Mie parameters  $ka > 30$ . Near the sphere the correspondence is worse due to unaccounted evanescent contributions.

The developed formulas can be directly applied in other areas of physics where nonparaxial axially symmetric focusing is of importance, e.g., acoustics, semiclassical quantum mechanics,<sup>42</sup> flat superlenses based on left-handed materials,<sup>43</sup> radio wave propagation, scattering theory,<sup>21</sup> chiral conical diffraction,<sup>44</sup> etc.

Concluding, let us briefly enumerate several possibilities to extend and refine the developed formalism. Weak absorption can be incorporated easily, for it just changes the amplitudes along the rays and the transmission coefficients. Strong absorption additionally modifies Snell's law of refraction, still preserving the axial symmetry. One can consider incoming radially or azimuthally polarized beams, which are known to produce better resolution than linear polarization.<sup>41</sup> The diffraction of light from regions beyond the sphere radius can be incorporated by considering the interference of the Bessoid field with creeping rays.<sup>38</sup> For other geometries, in particular finite apertures with sharp boundaries, edge rays, the Rubinowitz representation,<sup>16</sup> or an approach based on catastrophe theory<sup>45</sup> have to be used. Such corrections become relevant, for example, for the ray structure and the field distribution immediately behind spheres with a refractive index  $n < \sqrt{2}$ . Finally, one can calculate the interference of the diffracted light with the original incident wave or the interference of the light refracted by several spheres or arrays of spheres. The latter yields interesting secondary patterns<sup>46</sup> related to the so-called Talbot effect.<sup>47</sup>

## ACKNOWLEDGMENTS

The authors thank D. Bäuerle (Johannes Kepler University, Linz) for many stimulating discussions on microsphere patterning experiments, which initiated this study, and for his continuous support of this work. The authors also thank B. Luk'yanchuk and Z. B. Wang (both at the Data Storage Institute, Singapore) for their Mie program and discussions on Mie calculations. J.K. appreciates helpful conversations with G. Langer (Johannes Kepler University, Linz). N.A. thanks V. Palamodov (Tel Aviv University) for illuminating mathematical suggestions. Financial support was provided by the FWF (Austrian Science Fund) under Contract No. P16133-N08. N.A. also thanks the Christian Doppler Laboratory of Surface Optics (Johannes Kepler University, Linz). This work was done at the Institute for Applied Physics, Johannes Kepler University Linz, Austria.

## APPENDIX A: NEAR AXIS APPROXIMATION FOR THE BESSOID INTEGRAL

We make the substitution  $w \equiv \rho_1^2$  in Eq. (2)

$$I(R, Z) = \frac{1}{2} \int_0^\infty J_0(R\sqrt{w}) e^{-i(Zw/2 + w^2/4)} dw. \quad (\text{A1})$$

Near the axis (small  $R$ ) the Bessel function is slowly varying compared with the exponent. The integral will have significant contribution only from the region in which the exponent's phase is stationary, i.e., regions near  $w = -Z$ . We consider the most interesting caustic part of the axis for which  $Z \leq 0$ . In a lowest order approximation the Bessel function is considered as constant near the stationary point  $-Z$  and can be pulled out of the integral. The phase can be written as a complete quadratic form. With the full square of  $v \equiv (w + Z)/2$

$$I(R, Z) \approx J_0(R\sqrt{-Z}) e^{iZ^2/4} \int_{Z/2}^\infty e^{-iv^2} dv. \quad (\text{A2})$$

The remaining integral can be expressed in terms of the complementary error function<sup>25</sup>  $\text{erfc}$  (of complex argument) and hence we arrive at Eq. (11).

For  $Z > 0$  the point  $w = 0$  should be taken as a stationary edge point of the integration.<sup>24</sup> And the near axis approximation (11) remains valid as long as the Bessel function is set to  $J_0(0) = 1$ .

## APPENDIX B: AN ORDINARY DIFFERENTIAL EQUATION FOR THE BESSOID INTEGRAL

We derive the paraxial Helmholtz equation

$$I_{RR} + \frac{1}{R} I_R + 2iI_Z = 0, \quad (\text{B1})$$

as well as the following *ordinary differential equation* for the Bessoid integral

$$I_{RRR} + \frac{1}{R} I_{RR} - \left( \frac{1}{R^2} + Z \right) I_R + iRI = 0. \quad (\text{B2})$$

Indices denote partial derivatives. Both equations can be rewritten in the compact form

$$L + 2iI_Z = 0, \quad (\text{B3})$$

$$L_R - ZI_R + iRI = 0, \quad (\text{B4})$$

where  $L$  is the radial Laplacian

$$L \equiv I_{RR} + \frac{1}{R} I_R. \quad (\text{B5})$$

We begin with the proof of Eq. (B1) and state that we may differentiate under the integral sign, since the partial derivatives of the integrand exist and are continuous functions. Starting from the Bessoid integral in the polar representation (2), its integrand can be written as

$$G \equiv \rho_1 J_0(R\rho_1) E \quad (\text{B6})$$

with the abbreviation

$$E \equiv e^{-i(Z\rho_1^2/2 + \rho_1^4/4)}. \quad (\text{B7})$$

The (multiple) partial derivatives are

$$G_R = -\rho_1^2 J_1(R\rho_1) E, \quad (\text{B8})$$

$$G_{RR} = -\frac{\rho_1^3}{2} [J_0(R\rho_1) - J_2(R\rho_1)] E, \quad (\text{B9})$$

$$G_Z = -i\frac{\rho_1^3}{2} J_0(R\rho_1) E. \quad (\text{B10})$$

Here we used the derivative formula for Bessel functions<sup>25</sup>

$$\frac{d}{dt} J_m(t) = \frac{J_{m-1}(t) - J_{m+1}(t)}{2} \quad (\text{B11})$$

with  $m=0$  to obtain Eq. (B8) and  $m=1$  for Eq. (B9). Note that  $J_{-1}(t) = -J_1(t)$ . Applying the recurrence relation for Bessel functions<sup>25</sup>

$$J_{m+1}(t) = -J_{m-1}(t) + \frac{2m}{t} J_m(t), \quad (\text{B12})$$

one can eliminate  $J_2$  from Eq. (B9). And then it is enough to notice and verify that

$$G_{RR} + \frac{1}{R} G_R + 2iG_Z = 0. \quad (\text{B13})$$

This proves Eq. (B1).

For the proof of (B2), we need to note that its left hand side can be expressed as the integral of a partial derivative

$$H \equiv \int_0^\infty \frac{\partial}{\partial \rho_1} [i\rho_1 J_1(R\rho_1) e^{-i(Z\rho_1^2/2 + \rho_1^4/4)}] d\rho_1. \quad (\text{B14})$$

With the help of Eqs. (B11) and (B12) both the left hand side of Eq. (B2) and  $H$  become

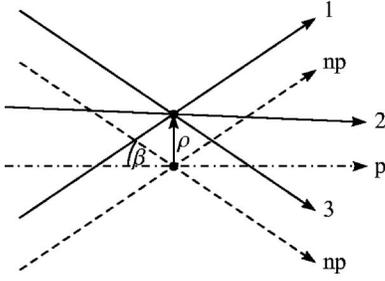


FIG. 9. Near the axis, the phases of the rays 1 and 3 differ from the phase of the nonparaxial rays np (crossing the axis at the same  $z$ ) by  $\pm k\rho \sin \beta$ , whereas the phases of ray 2 and the (par)axial ray are the same in first order. The nonparaxial ray and ray 3 cross the axis at an angle  $\beta > 0$ .

$$\int_0^\infty [iR\rho_1 J_0(R\rho_1) + \rho_1^2(Z + \rho_1^2)J_1(R\rho_1)] E d\rho_1. \quad (\text{B15})$$

Thus, in order to prove Eq. (B2), it is enough to show that  $H=0$ . This follows from the Newton-Leibniz formula applied to the (definite) integral (B14):

$$H = [i\rho_1 J_1(R\rho_1) e^{-i(Z\rho_1^2 + \rho_1^4)}]_0^\infty = 0. \quad (\text{B16})$$

The lower bound at 0 vanishes for obvious reasons. For the upper bound at  $\infty$  one assumes an *infinitely small* imaginary part in front of the fourth order term in the exponent:  $\rho_1^4 \rightarrow (1-i\varepsilon)\rho_1^4$  with  $\varepsilon > 0$ . This completes the proof of the differential Eqs. (B1) and (B2) for the Bessoid integral.

### APPENDIX C: THE NEAR AXIS BESSOID COORDINATES

Near the axis the phases of the rays can be Taylor expanded. From Fig. 9 one infers that up to the first order in  $\rho$  the phases can be written as

$$\varphi_1 \approx \varphi_{np} + k\rho \sin \beta, \quad \varphi_2 \approx \varphi_p, \quad \varphi_3 \approx \varphi_{np} - k\rho \sin \beta. \quad (\text{C1})$$

Here  $\varphi_{np}$  and  $\varphi_p$  denote the phases of the nonparaxial rays and the (par)axial ray (with  $\rho=0$ ) and  $\beta > 0$  is the angle of ray 3 with the axis.

We insert these phases into the exact expressions for  $R$  and  $Z$  in Eq. (28), Taylor expand the result with respect to  $\rho$  and resubstitute  $\varphi_{np} \approx (\varphi_1 + \varphi_3)/2$ ,  $\varphi_p \approx \varphi_2$  and  $k\rho \sin \beta \approx (\varphi_1 - \varphi_3)/2$  from Eq. (C1). This yields Eqs. (32) and (33).

### APPENDIX D: THE ON AXIS FIELD

Here we derive a simple on axis expression for the Bessoid-matched field  $U$  (21), namely Eq. (37).

On the axis and inside the cusp  $\rho=0$  ( $R=0$ ) and  $z < 0$  ( $Z < 0$ ). The stationary points, given by Eq. (26), are

$$t_1 = -\sqrt{-Z}, \quad t_2 = 0, \quad t_3 = -t_1. \quad (\text{D1})$$

Then, the amplitude  $A$  in Eq. (31) simplifies to

$$A = U_{0,2} \frac{\sqrt{H_2}}{\sqrt{J_2}}, \quad (\text{D2})$$

because on the axis the ratios  $\sqrt{H_{1,3}}/\sqrt{J_{1,3}}$  are both finite and the corresponding other terms disappear upon multiplication with  $t_2=0$ . Due to the restriction to the lit region ( $Z < 0$ ), all rays are real and Eq. (24) holds. By virtue of Eq. (9)  $\det \mathbf{H}_2 = Z^2$ , and due to Eq. (10)  $\text{sign } \mathbf{H}_2 = 2$ , one finds

$$\sqrt{H_2} = iZ \quad (\text{D3})$$

and thus

$$A = i \frac{U_{0,2} Z}{\sqrt{J_2}}. \quad (\text{D4})$$

This approximation for the amplitude  $A$  is valid up to the focus ( $Z=0$ ). As ray 2 converges like the inverse distance from the focus,  $\sqrt{J_2}$  is proportional to  $Z$ .

The amplitude  $A_R$  in Eq. (32) vanishes due to  $t_3 = -t_1$  and the fact that

$$\frac{\sqrt{H_1}}{\sqrt{J_1}} = \frac{\sqrt{H_3}}{\sqrt{J_3}}, \quad (\text{D5})$$

which means that rays 1 and 3 have equal amplitudes and that the caustic phase shifts are in accordance with the signature of the Hessian. Consequently,

$$A_R = 0. \quad (\text{D6})$$

With Eqs. (D1) and (D5) the amplitude  $A_Z$  reads

$$A_Z = \frac{2}{Z} \left( U_{0,1} \frac{\sqrt{H_1}}{\sqrt{J_1}} - U_{0,2} \frac{\sqrt{H_2}}{\sqrt{J_2}} \right). \quad (\text{D7})$$

The first term is nontrivial. Both  $\sqrt{H_1}$  and  $\sqrt{J_1}$  are zero on the axis, but their ratio is finite and well defined. Indeed, the Taylor expansion of Cardan's solution  $t_1$  in its trigonometric representation<sup>23</sup> yields in the first order in  $R$

$$t_1 = -\sqrt{-Z} + \frac{R}{2Z}. \quad (\text{D8})$$

Therefore, again in first order in  $R$ :  $\det \mathbf{H}_1 = 2R\sqrt{-Z}$ . Due to  $\text{sign } \mathbf{H}_1 = -2$  we obtain

$$\sqrt{H_1} = i\sqrt{2R\sqrt{-Z}} \quad (\text{D9})$$

and with Eq. (D3)

$$A_Z = 2i \frac{U_{0,1} \sqrt{2R\sqrt{-Z}}}{Z\sqrt{J_1}} - 2i \frac{U_{0,2}}{\sqrt{J_2}}. \quad (\text{D10})$$

This approximation for  $A_Z$  holds for small values of  $R$ . It is finite, since  $\sqrt{J_1}$  approaches zero as  $\sqrt{R}$  for  $R \rightarrow 0$ .

For the final representation of the field  $U$ , we can substitute the near axis expression for  $R\sqrt{-Z}$  (34) into  $A_Z$ . On the axis the phase coordinate becomes  $\chi = \varphi_2$ , which results from substituting  $\varphi_1 = \varphi_3$  and  $Z = -2\sqrt{\varphi_1 - \varphi_2}$  into the corresponding expression in Eq. (28). This leads to

$$\begin{aligned}
U &= \left( AI + \frac{1}{i} A_R I_R + \frac{1}{i} A_Z I_Z \right) e^{i\chi} \\
&= \left( \frac{U_{0,2}}{\sqrt{J_2}} (iZI - 2I_Z) + \frac{2U_{0,1} \sqrt{2k\rho \sin \beta}}{Z\sqrt{J_1}} I_Z \right) e^{i\varphi_2}.
\end{aligned} \tag{D11}$$

Using the linear relationship between the Bessoid integral and its  $Z$ -derivative Eq. (15),

$$iZI - 2I_Z = 1, \tag{D12}$$

we end up with Eq. (37).

#### APPENDIX E: HIGHER-ORDER BESSOID INTEGRALS

*Higher-order Bessoid integrals* (38) appear naturally, if one expands an arbitrary initial field amplitude on the aperture in a Fourier series

$$U_0(\rho_1, \varphi_1) = \sum_{m=0}^{\infty} [a_m(\rho_1) \cos(m\varphi_1) + b_m(\rho_1) \sin(m\varphi_1)]. \tag{E1}$$

The form of the coefficients  $a_m$  and  $b_m$  can be seen from a two-dimensional Taylor expansion in Cartesian coordinates around the point  $(0, 0)$ , rewritten into polar coordinates

$$U_0(\rho_1, \varphi_1) = \sum_{m=0}^{\infty} \sum_{n=0}^m c_{mn} \rho_1^m \cos^{m-n} \varphi_1 \sin^n \varphi_1 \tag{E2}$$

with

$$c_{mn} \equiv \frac{1}{m!} \binom{m}{n} \left. \frac{\partial^m U_0(x'_1, y'_1)}{\partial x_1^{m-n} \partial y_1^n} \right|_{x'_1=0, y'_1=0}. \tag{E3}$$

Thus,  $\rho_1^m$  is the lowest possible power of  $\rho_1$  which can be found in the term with  $\exp(im\varphi_1)$ . An additional  $\rho_1$  comes from the transformation from Cartesian to polar coordinates.

If we define the functions

$$\tilde{I}_m \equiv I_m e^{im\varphi}, \tag{E4}$$

we find that they satisfy the paraxial Helmholtz equation

$$2i\tilde{I}_{m,Z} + \tilde{I}_{m,RR} + \frac{1}{R}\tilde{I}_{m,R} + \frac{1}{R^2}\tilde{I}_{m,\varphi\varphi} = 0, \tag{E5}$$

where  $\tilde{I}_{m,\varphi\varphi} = -m^2\tilde{I}_m$ .

Due to Eqs. (B11) and (B12), one can write the identity

$$\rho_1^{m+2} J_{m+1}(R\rho_1) = -\frac{\partial}{\partial R} [\rho_1^{m+1} J_m(R\rho_1)] + \frac{m}{R} \rho_1^{m+1} J_m(R\rho_1). \tag{E6}$$

Hence, the recursive relation for the Bessoid integrals, Eq. (39), follows:

$$I_{m+1} = -I_{m,R} + \frac{m}{R} I_m, \tag{E7}$$

i.e.,  $I_1 = -I_{0,R} \equiv -I_R$ ,  $I_2 = I_{RR} - I_R/R$ , etc. Using Eqs. (E7) and (B12) as well as expression (11) for  $I$ , one obtains

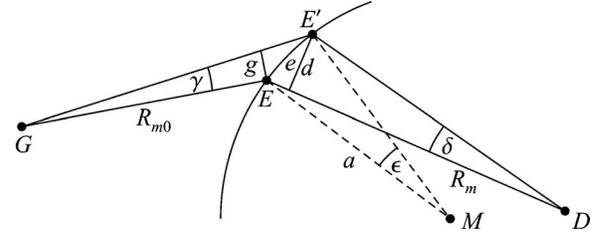


FIG. 10. Meridional cross section for the determination of the meridional radius of curvature,  $R_m$ .  $M$  is the center of the sphere.

$$I_m(R, Z) \approx \frac{\sqrt{\pi}(-Z)^{m/2}}{2} J_m(R\sqrt{-Z}) e^{i(Z^2-\pi)/4} \operatorname{erfc}\left(\frac{Z}{2} e^{i\pi/4}\right), \tag{E8}$$

which is the analytic near axis expression for the higher-order Bessoid integrals.

While the coordinates and phases  $(R, Z, \chi)$  remain unchanged, the derivation of the higher-order amplitudes  $(A_m, A_{mR}, A_{mZ})$  requires some insight for  $m \geq 2$ . Let us briefly consider the case  $m=2$ . For the matching procedure we need the asymptotic behavior of  $I_2 = I_{RR} - I_R/R$  far from the caustic regions where  $R \gg 1$  and where it is dominated by the term  $I_{RR}$ . Note that for the matching procedure we need exactly this asymptotic representation and in the noncaustic regions only. Thus—although we need the second-order Bessoid integral  $I_2$  on and near the axis, where it vanishes—we shall use its asymptotic stationary phase expressions far from the axis for the derivation of the amplitudes. In this region it is equivalent to the asymptotic of  $I_{RR}$ .

In fact, we may generalize this statement to arbitrary order. Due to Eq. (E7) the leading term in the stationary phase calculation is always

$$I_m \rightarrow \left(-\frac{\partial}{\partial R}\right)^m I. \tag{E9}$$

Therefore, the equations for the amplitudes (30) become

$$\frac{U_{0,j}^{(m)}}{\sqrt{J_j}} = (it_j)^m \frac{A_m - it_j A_{mR} - \frac{1}{2} t_j^2 A_{mZ}}{\sqrt{H_j}}, \tag{E10}$$

which can be seen from the Bessoid integral's Cartesian representation with the phase (4):

$$\left(-\frac{\partial}{\partial R}\right)^m I \rightarrow \left(-i\frac{\partial\phi}{\partial R}\right)^m I = (ix_1)^m I. \tag{E11}$$

Equations (E10) have the same form as Eq. (30) except for an additional factor  $(it_j)^m$  on the left hand side, proving Eq. (42).

#### APPENDIX F: WAVE FRONT RADII OF CURVATURE FOR THE REFRACTION ON A SPHERE

Let us consider a point source  $G$  and start with the derivation of the meridional radius of curvature (Fig. 10).

The initial radius of curvature is  $R_{m0} \equiv GE$ , the one after refraction is  $R_m \equiv ED$ . The infinitesimally neighbored beam

( $\gamma \ll 1$ ) which is refracted in  $E'$  (the angles of incidence and transmission in  $E$  are  $\theta_i$  and  $\theta_t$ , in  $E'$  they are denoted  $\theta'_i$  and  $\theta'_t$ ) also propagates to  $D$ . The normals onto  $R_{m0}$  through  $E$  and onto  $R_m$  through  $E'$  are  $g$  and  $d$ , respectively. In the necessary order the length of the arc  $EE'$  can be approximated by the distance  $e \approx EE'$ . As all angles are small:  $g = \gamma R_{m0}$ ,  $d = \delta R_m$ , and  $e = \epsilon a$ . On the other hand, we find from the infinitesimal triangles:  $g = e \cos \theta_i$ ,  $d = e \cos \theta_t$ . This leads to

$$R_m = -\frac{\gamma R_{m0} \cos \theta_t}{\delta \cos \theta_i}, \quad (\text{F1})$$

where we have introduced a minus sign because the wave is converging after the refraction. The remaining problem is the angle  $\delta$  in the denominator. To find  $\delta$  we write the relations between angles and primed angles

$$\theta'_i = \theta_i + \gamma + \epsilon, \quad \theta'_t = \theta_t - \delta + \epsilon. \quad (\text{F2})$$

With Snell's law

$$\theta_t - \theta'_t = \arcsin \frac{\sin \theta_i}{n} - \arcsin \frac{\sin(\theta_i + \gamma + \epsilon)}{n}, \quad (\text{F3})$$

a first-order Taylor expansion in  $(\gamma + \epsilon)$  yields

$$\theta_t - \theta'_t = -\frac{(\gamma + \epsilon) \cos \theta_i}{n \cos \theta_t}. \quad (\text{F4})$$

We express  $\delta$  from Eq. (F2), substitute it into Eq. (F1) and finally obtain the meridional radius of curvature (48)

$$R_m = \frac{naR_{m0} \cos^2 \theta_t}{a \cos^2 \theta_i + R_{m0}(\cos \theta_i - n \cos \theta_t)}. \quad (\text{F5})$$

For the sagittal radius of curvature we consider Fig. 11.

A ray which emerged from  $G$  is refracted in  $E$  (incident angle  $\theta_i$  and transmitted angle  $\theta_t$ ). The distance from  $E$  to the

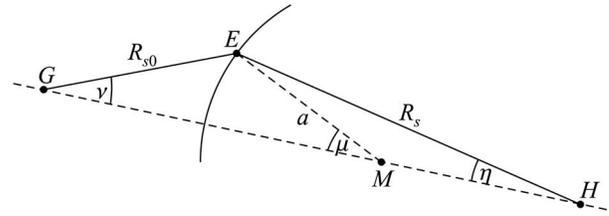


FIG. 11. Meridional cross section for the determination of the sagittal radius of curvature,  $R_s$ .

intersection  $H$  of the ray with the line passing through  $G$  and the sphere center  $M$  is the sagittal radius of curvature, as a neighbored ray, emerging from  $G$  and hitting the sphere not in  $E$  but infinitesimally shifted perpendicular to the meridional plane, will also propagate to  $H$  due to symmetry around the line  $GM$ . We have  $R_{s0} \equiv GE$  and  $R_s \equiv EH$ . The tangent theorem states (all angles are in general large now)

$$\tan \frac{\nu - \mu}{2} = \frac{a - R_{s0}}{a + R_{s0}} \cot \frac{\pi - \theta_i}{2}, \quad (\text{F6})$$

where  $\pi - \theta_i$  is just the third angle in the triangle  $GME$ . Due to  $\nu + \mu = \theta_t$  we find

$$\mu = \frac{\theta_t}{2} - \arctan \left( \frac{a - R_{s0}}{a + R_{s0}} \cot \frac{\pi - \theta_i}{2} \right). \quad (\text{F7})$$

In the triangle  $MHE$  the sine theorem reads

$$-\frac{R_s}{a} = \frac{\sin(\pi - \mu)}{\sin \eta}, \quad (\text{F8})$$

where we have again introduced a minus sign due to the convergence of the refracted wave. Trigonometric transformations finally give the sagittal radius of curvature (49)

$$R_s = \frac{naR_{s0}}{a + R_{s0}(\cos \theta_i - n \cos \theta_t)}. \quad (\text{F9})$$

\*Electronic address: johannes.kofler@univie.ac.at

†Electronic address: nikita.arnold@jku.at

<sup>1</sup>J. E. G. J. Wijnhoven and W. L. Vos, *Science* **281**, 802 (1998).

<sup>2</sup>Y. A. Vlasov, X.-Z. Bo, J. C. Sturm, and D. J. Norris, *Nature (London)* **414**, 289 (2001).

<sup>3</sup>M. L. Gorodetsky, A. A. Savchenkov, and V. S. Ilchenko, *Opt. Lett.* **21**, 453 (1996).

<sup>4</sup>K. J. Vahala, *Nature (London)* **424**, 839 (2003).

<sup>5</sup>F. Burmeister, W. Badowsky, T. Braun, S. Wieprich, J. Boneberg, and P. Leiderer, *Appl. Surf. Sci.* **144**, 461 (1999).

<sup>6</sup>D. Bäuerle, *Laser Processing and Chemistry*, 3rd ed. (Springer-Verlag, Berlin, 2000).

<sup>7</sup>D. Bäuerle, L. Landström, J. Kofler, N. Arnold, and K. Piglmayer, *Proc. SPIE* **5339**, 20 (2004).

<sup>8</sup>H.-J. Münzer, M. Mosbacher, M. Bertsch, J. Zimmermann, P. Leiderer, and J. Boneberg, *J. Microsc.* **202**, 129 (2001).

<sup>9</sup>*Laser Cleaning*, edited by B. S. Luk'yanchuk (World Scientific, Singapore, 2002).

<sup>10</sup>B. S. Luk'yanchuk, N. Arnold, S. M. Huang, Z. B. Wang, and M. H. Hong, *Appl. Phys. A* **77**, 209 (2003).

<sup>11</sup>C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).

<sup>12</sup>G. Roll and G. Schweiger, *J. Opt. Soc. Am. A* **17**, 1301 (2000).

<sup>13</sup>A. Rohrbach, *Phys. Rev. Lett.* **95**, 168102 (2005).

<sup>14</sup>W. Zimmer, Ph.D. thesis, Freie Universität, Berlin, Germany, 2001.

<sup>15</sup>G. Mie, *Ann. Phys.* **25**, 377 (1908).

<sup>16</sup>M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge, 2002).

<sup>17</sup>Y. A. Kravtsov and Y. I. Orlov, *Caustics, Catastrophes and Wave Fields*, 2nd ed., Springer Series on Wave Phenomena Vol. 15 (Springer-Verlag, Berlin, 1999).

<sup>18</sup>M. V. Berry, *J. Phys. A* **13**, 149 (1980).

<sup>19</sup>M. V. Berry, *Phys. Today* **54**(4), 11 (2001).

<sup>20</sup>T. Pearcey, *Philos. Mag.* **37**, 311 (1946).

<sup>21</sup>J. N. Connor and D. Farrelly, *J. Chem. Phys.* **75**, 2831 (1981).

- <sup>22</sup>N. P. Kirk, J. N. Connor, P. R. Curtis, and C. A. Hobbs, *J. Phys. A* **33**, 4797 (2000).
- <sup>23</sup>I. N. Bronstein and K. A. Semendjajew, *Handbook of Mathematics*, 4th ed. (Springer-Verlag, Berlin, 2004).
- <sup>24</sup>J. Focke, *Opt. Acta* **3**, 110 (1956).
- <sup>25</sup>*Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (Wiley, New York, 1993).
- <sup>26</sup>D. McGloin and K. Dholakia, *Contemp. Phys.* **46**, 15 (2005).
- <sup>27</sup>J. Kofler, Master thesis, Johannes Kepler Universität, Linz, Austria, 2004.
- <sup>28</sup>In the present work all plots of the Bessoid integral (and its derivatives) contain  $101 \times 101$  mesh points, computed with the software package MATHEMATICA 5 (Wolfram Research). Direct numerical integration of Eq. (2) takes more than one hour on a modern personal computer. Integration along a line in the complex plane decreases the amount of time by approximately a factor of 3. Solving the ordinary differential equation (12), however, lasts only a few seconds.
- <sup>29</sup>J. N. Connor and P. R. Curtis, *J. Math. Phys.* **25**, 2895 (1984). In this work the term *cuspid* stands for the whole family of canonical catastrophe integrals with corank 1, beginning with fold (co-dimension 1), cusp (2), swallowtail (3), and butterfly (4), but does not mean the surface of revolution of a cusp with corank 2 and co-dimension 2 as we use it.
- <sup>30</sup>J. N. Connor, *Mol. Phys.* **26**, 1217 (1973).
- <sup>31</sup>J. J. Stamnes and B. Spjelkavik, *Opt. Acta* **30**, 1331 (1983).
- <sup>32</sup>A. J. Janssen, *J. Phys. A* **25**, L823 (1992).
- <sup>33</sup>T. Pearcey and G. W. Hill, *Spherical Aberrations of Second Order: The Effect of Aberrations Upon the Optical Focus* (Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia, 1963), p. 71.
- <sup>34</sup>Y. A. Kravtsov and Y. I. Orlov, *Geometrical Optics of Inhomogeneous Media*, Springer Series on Wave Phenomena Vol. 6 (Springer-Verlag, Berlin, 1990).
- <sup>35</sup>L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II*, Springer Series on Wave Phenomena Vol. 10 (Springer-Verlag, Berlin, 1992).
- <sup>36</sup>L. Bergmann and C. Schäfer, *Lehrbuch der Experimentalphysik* (Band 3, Optik), edited by H. Gobrecht (Walter de Gruyter, 7. Auflage, 1978).
- <sup>37</sup>V. Červený, *Seismic Ray Theory* (Cambridge University Press, Cambridge, 2001).
- <sup>38</sup>J. B. Keller, *J. Opt. Soc. Am.* **52**, 116 (1962).
- <sup>39</sup>D. Bäuerle, G. Wysocki, L. Landström, J. Klimstein, K. Piglmayer, and J. Heitz, *Proc. SPIE* **5063**, 8 (2003).
- <sup>40</sup>G. Langer, Master thesis, Johannes Kepler Universität, Linz, Austria, 2005.
- <sup>41</sup>R. Dorn, S. Quabis, and G. Leuchs, *Phys. Rev. Lett.* **91**, 233901 (2003).
- <sup>42</sup>A. D. Peters, C. Jaffé, J. Gao, and J. B. Delos, *Phys. Rev. A* **56**, 345 (1997).
- <sup>43</sup>P. V. Parimi, W. T. Lu, P. Vodo, and S. Sridhar, *Nature (London)* **426**, 404 (2003).
- <sup>44</sup>M. V. Berry and M. R. Jeffrey, *J. Opt. A: Pure Appl. Opt.* **8**, 363 (2006).
- <sup>45</sup>J. F. Nye, *J. Opt. A: Pure Appl. Opt.* **7**, 95 (2005).
- <sup>46</sup>D. Bäuerle, K. Piglmayer, R. Denk, and N. Arnold, *Lambda Phys.* **60**, 1 (2002).
- <sup>47</sup>M. V. Berry, I. Marzoli, and W. Schleich, *Phys. World* **14**(6), 39 (2001).