

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Modeling physical optics phenomena by complex ray tracing

James E. Harvey
Ryan G. Irvin
Richard N. Pfisterer

SPIE.

Modeling physical optics phenomena by complex ray tracing

James E. Harvey,* Ryan G. Irvin, and Richard N. Pfisterer

Photon Engineering, LLC, 310 South Williams Boulevard, Suite 222, Tucson, Arizona 85711, United States

Abstract. Physical optics modeling requires propagating optical wave fields from a specific radiometric source through complex systems of apertures and reflective or refractive optical components, or even complete instruments or devices, usually to a focal plane or sensor. The model must accurately include the interference and diffraction effects allowed by the polarization and coherence characteristics of both the initial optical wave field and the components and media through which it passes. Like a spherical wave and a plane wave, a Gaussian spherical wave (or Gaussian beam) is also a solution to the paraxial wave equation and does not change its fundamental form during propagation. The propagation of a Gaussian beam is well understood and easily characterized by a few simple parameters. Furthermore, a paraxial Gaussian beam can be propagated through optical systems using geometrical ray-trace methods. The decomposition of arbitrary propagating wave fields into a superposition of Gaussian beamlets is, thus, an alternative to the classical methods of propagating optical wave fields. This decomposition into Gaussian beamlets has been exploited to significant advantage in the modeling of a wide range of physical optics phenomena. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.54.3.035105](https://doi.org/10.1117/1.OE.54.3.035105)]

Keywords: physical optics modeling; complex ray tracing; Gaussian beam decomposition of arbitrary wave fields.

Paper 141706 received Nov. 3, 2014; accepted for publication Mar. 2, 2015; published online Mar. 23, 2015.

1 Introduction

Over the last three decades, there has been a quiet revolution occurring in the computer modeling capability of both fundamental physical optics phenomena and performance predictions of sophisticated and advanced optical systems. This revolution is based upon the practice of decomposing an arbitrary optical wave field into a superposition of (coherent) Gaussian beams, propagating those beams by Arnaud's method of complex ray tracing¹ and then coherently recombining the resultant optical fields of each beam at the analysis plane or sensor.

Every physics and optical engineering student learns that an arbitrary optical wave field can be decomposed into a superposition of (Huygens') spherical wavelets, i.e., the Rayleigh–Sommerfeld diffraction theory. They also learn that an arbitrary optical wave field can be decomposed into a superposition of plane wave components, i.e., the angular spectrum approach of Fourier optics. Meanwhile, the alternative method of decomposing an arbitrary optical wave field into a superposition of Gaussian beamlets (this terminology, in analogy to the well-known Huygens' spherical wavelets, was introduced by Al Greynolds in Ref. 2) has been implemented by software engineers in several commercially available software packages. These software packages are being extensively used by industry and government agencies to model the physical optics performance of increasingly advanced optical systems, including the polarization and coherence characteristics of those systems. The resulting software is fast, accurate, user-friendly, provides impressive graphical output, and can potentially be used as a great tool

in education for illustrating a wide variety of physical optics phenomena.

After three decades, this powerful modeling technique has not yet been published in the peer-reviewed literature, or in physics or optics textbooks, nor is it generally being taught to physics or optics students even in our academic institutions specializing in optical sciences or optical engineering.

Our motivation in publishing this paper is to introduce and demonstrate this powerful concept to optical engineers and educators with the hope that they will incorporate it into their toolbox of techniques for modeling and analyzing optical systems exhibiting polarization, interference, diffraction, and coherence phenomena.

2 Historical Background

The zero-width ray has been used for centuries as the basic tool for designing and analyzing optical systems.³ Geometrical optics has been a remarkably effective model for the design and analysis of both imaging and nonimaging optical systems. Bundles of rays not only determine the phase fronts of the associated wave field, but also effectively describe the flow of optical power. Also, both the phase and amplitude of the optical field can be predicted from the rays. However, there are two prominent weaknesses with this approach. First, when a ray bundle collapses (i.e., at a caustic or focus), the model wrongly predicts an infinite field amplitude. Second, when the field has encountered an aperture, simple ray-based predictions are entirely inconsistent with the familiar diffraction patterns of the wave solutions.⁴

For over a century, the decomposition of an arbitrary wave field into the superposition of (Huygens') spherical wavelets has been effectively used to model the diffractive effects of truncating apertures (the Rayleigh–Sommerfeld diffraction theory). For the past few decades, the decomposition into a superposition of plane waves (the angular

*Address all correspondence to: James E. Harvey, E-mail: jimh@photonengr.com

spectrum approach of Fourier optics) combined with extensive ray tracing has been successful in modeling physical optical effects, such as interference, diffraction, and wavefront aberrations.⁵

However, this angular spectrum approach to incorporating wave optics in a ray-optics model is limited to the determination of the diffraction image on a plane close to the focus and nearly perpendicular to the axis of the imaging light cone. Also, the system must have a well-defined exit pupil. These requirements exclude calculating physical optics effects at arbitrary points in both imaging and nonimaging systems for arbitrary coherence and polarization conditions.⁴ The standard work-around is to reconstruct the field on a fictional surface (the exit pupil) that is well removed from the detector. When the image space is homogeneous, it is straightforward to apply wave theory for the last step of propagating this field to the detector in order to determine the image quality. This allows the ray model to be used where a rigorous wave solution would have been impractical, and wave theory is then called upon only where it seems to be essential. Yet there exists serious problems as there is not always a well-behaved exit pupil (for example, in the presence of astigmatism).⁴ Such issues provided motivation for the developments presented in Refs. 6–8.

3 Gaussian Beam Characteristics

The electric field amplitude of a Gaussian spherical wave that satisfies the paraxial wave equation is expressed in Eq. (1).⁹

$$E(r, z) = E_0 \frac{w_0}{w(z)} \exp \left\{ - \left[\frac{r}{w(z)} \right]^2 - j \left[kz + k \frac{r^2}{2R(z)} - \varphi(z) \right] \right\}. \quad (1)$$

It has a Gaussian amplitude and both a linear and a quadratic phase term.

The ABCD matrix method of propagating Gaussian laser beams was discussed by Kogelnik and Li in 1966.¹⁰ Over the next couple of decades, the free-space propagation behavior of Gaussian spherical waves, or Gaussian (laser) beams, as illustrated in Fig. 1 and described by the associated parametric equations, became quite well understood by laser physicists and optical engineers.^{11,12}

As it propagates through space, diffraction effects cause the Gaussian beam to broaden and diverge. Every Gaussian spherical wave field can be traced backward (or forward) to a unique real or virtual waist at a unique axial position (usually designated as $z = 0$). The principle parameters associated

with the beam are the beam radius at the waist, w_0 , and the Rayleigh range

$$Z_R = \frac{\pi w_0^2}{\lambda}, \quad (2)$$

where λ is the wavelength; the beam radius at an arbitrary distance from the waist is

$$w(z) = w_0 \sqrt{1 + (z/z_R)^2}, \quad (3)$$

and the radius of curvature of the spherical wavefront at an arbitrary distance from the waist is

$$R(z) = z[1 + (z_R/z)^2]. \quad (4)$$

The beam radius is a minimum at the waist, and the wavefront is flat, i.e., $R(z) = \infty$. The Rayleigh range, z_R , is the distance at which $w(z)$ is $\sqrt{2}$ times w_0 , i.e., the beam area doubles. The distance over which a Gaussian beam can be considered to be collimated is, thus, nominally $2z_R$. The beam divergence near the waist is very small; however, the asymptotic divergence angle, θ , can be quite large for small beam waist sizes.

$$\tan \theta = \frac{dw}{dz} = w_0 \frac{(z/z_R)}{\sqrt{Z_R^2 + z^2}}. \quad (5)$$

In the far field, ($z \gg z_R$).

$$\tan \theta_{\text{lim } z \gg z_R} \approx \frac{w_0}{Z_R} = \frac{\lambda}{\pi w_0}. \quad (6)$$

The irradiance distribution, $I(r, z)$, of the Gaussian beam produced by the electric field amplitude expressed in Eq. (1) is given by

$$I(r, z) = \frac{c\mathcal{E}}{2} |E(r, z)|^2 = I_0 \left(\frac{w_0}{w(z)} \right)^2 \exp \left[-2 \left(\frac{r}{w(z)} \right)^2 \right]. \quad (7)$$

The total radiant power in the Gaussian beam is obtained by integrating the irradiance over a plane perpendicular to the propagation direction.

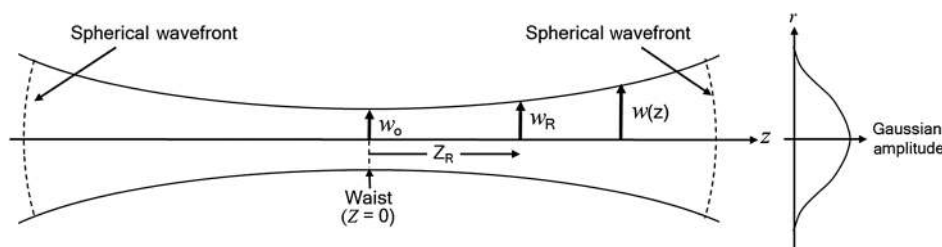


Fig. 1 Illustration of the Gaussian spherical wave propagation behavior.

$$\begin{aligned}
 \text{Power} &= \int \int \frac{c\epsilon}{2} |E|^2 dA \\
 &= \left(\frac{w_0}{w}\right)^2 I_0 \int_0^{2\pi} \int_0^\infty \exp\left[-2\left(\frac{r}{w}\right)^2\right] r dr d\varphi \\
 &= \frac{\pi w_0^2}{2} I_0.
 \end{aligned} \tag{8}$$

The quantity can, thus, be thought of as an effective area of the beam waist. From Eqs. (7) and (8), the general equation for the irradiance in terms of the beam power, P , for arbitrary z and r is

$$I(r, z) = \frac{2P}{\pi w^2(z)} \exp\left\{-2\left(\frac{r}{w(z)}\right)^2\right\}, \tag{9}$$

and the on-axis irradiance can be written as

$$I(0, z) = I_0(z) = \frac{2P}{\pi w^2(z)}. \tag{10}$$

In the far field ($z \gg z_R$), the on-axis irradiance is given by

$$I(0, z) \approx \frac{2\pi w_0^2 P}{\lambda^2 z^2}. \tag{11}$$

Thus, $w(z)$ is the $1/e$ half-width of the field amplitude, E , and the $1/e^2$ half-width of the irradiance distribution, I . The quadratic phase part of the Gaussian spherical wave of Eq. (1) is of key importance in understanding the limitations of accurately modeling physical optical phenomena by complex ray tracing.

4 Propagating Gaussian Beams by Complex Ray Tracing

In 1969, Arnaud and Kogelnik published a paper yielding equations for the propagation of “generally astigmatic” Gaussian beams through a general optical system.¹³ A generally astigmatic beam can be produced by sending a rotationally symmetric Gaussian beam through two crossed cylinder lenses.

Arnaud also showed that a Gaussian beam can be fully characterized in any space by representing the imaginary and real components with rays, the combination of which describes a “complex ray” that “formally obeys the laws of geometrical optics.”^{14,15} This complex ray is a representation of a skew ray that generates the Gaussian beam when rotated about the optical axis. Thus, the complex ray representation provides a method of propagating the beam by ordinary geometrical ray tracing methods. Arnaud discussed, first in an unpublished internal Bell Labs technical memorandum (1968) and then, finally, in a peer-reviewed journal article 16 years later,¹ a graphic method for determining the beam parameters by projecting the two rays onto a plane perpendicular to the optical axis.

The similarities of this graphical method to the Delano $y\bar{y}$ diagrams^{16,17} led Kessler and Shack¹⁸ to discuss this two-ray representation of Gaussian beams and show that a Gaussian beam with a wavelength (in air) of λ is defined by any two paraxial rays such that

$$y\bar{u} - \bar{y}u = \lambda/\pi, \tag{12}$$

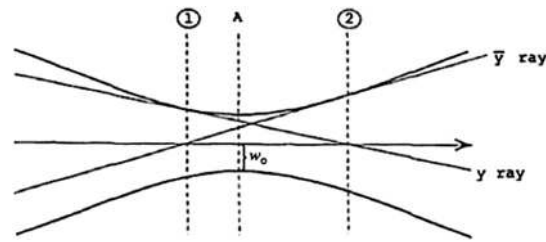


Fig. 2 Representation of a Gaussian beam by two arbitrary paraxial rays.

$$w = \sqrt{y^2 + \bar{y}^2}, \tag{13}$$

where y and \bar{y} are the ray heights at an arbitrary plane, and u and \bar{u} are the (reduced) paraxial angles of these rays. As illustrated in Fig. 2, the Gaussian beam (as described by its spot size) is tangent to each of these two rays at planes one and two, where the other ray intersects the axis. The use of the notation $y\bar{y}$ should not be taken to imply that these two rays are the marginal and chief rays of an imaging system.

One particular choice of rays is shown in Fig. 3, where the two tangent points are at the waist (plane A) and at infinity. Herloski et al.¹⁹ used the terms “waist ray” and “divergence ray” for these rays and used them in the Code V optical design program to optimize lens systems for desired Gaussian beam properties.

The waist ray, traveling parallel to the optical axis shown in Fig. 3, corresponds to the imaginary part of a complex ray. And the divergence ray, with zero height at the beam waist and a ray angle equal to the far-field divergence of the beam, corresponds to the real part of a complex ray. Thus, we actually trace two geometrical rays to model the propagation of a complex ray, i.e., a simple Gaussian beam propagating through a homogeneous media.

By complex ray tracing we mean Arnaud’s ray-equivalent method for propagating generally astigmatic (twisted) Gaussian beams along a skew path through a nonsymmetric optical system. For the general astigmatic case, a base (or chief) ray, four secondary divergence rays, and four secondary parallel waist rays must be traced. Two of these pairs of rays are shown in Fig. 4 (there are another two pairs of rays in the plane perpendicular to the paper and containing the base ray). Thus, nine rays total per complex ray must be traced in order to represent the generally astigmatic Gaussian beam.

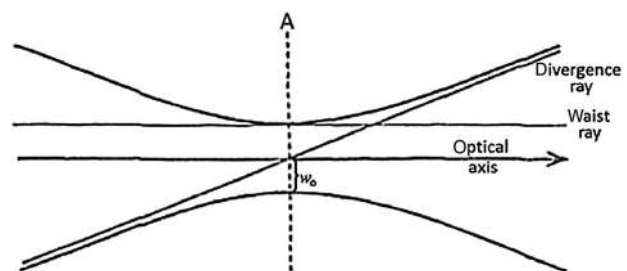


Fig. 3 Representation of a Gaussian beam by a divergence ray and a waist ray.

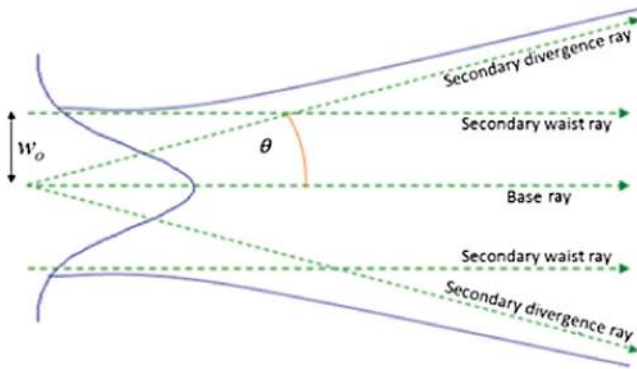


Fig. 4 Representation of a general Gaussian beam by a complex ray consisting of a central base ray, four secondary waist rays, and four secondary divergence rays (there are another two pairs of rays in the plane perpendicular to the paper and containing the base ray).

5 Arbitrary Wave Fields as a Superposition of Gaussian Beamlets

Greynolds published a vector formulation of the ray-equivalent method for general Gaussian beam propagation in 1986.²⁰ However, his real interest was not motivated by the need to analyze optical systems that steer, focus, and/or shape only Gaussian beams. The primary thrust of his research was to show that the Gaussian beam can be used as a general tool in the diffraction analysis of arbitrary wavefronts in any optical system.²¹ Implementation of this vector formulation in software is made possible because any arbitrary wave field (of finite spatial frequency bandwidth) can be decomposed into a collection of paraxial Gaussian beams, and Gaussian beams can be easily propagated by complex ray tracing.

Greynolds proposed complex ray tracing of Gaussian beamlets as a powerful means of performing system diffraction calculations for a wide variety of applications. There are three fundamental steps in performing the calculations: (1) decomposition of the incident arbitrary wave field into a superposition of equally spaced, mutually coherent Gaussian beamlets, (2) propagation of the individual generally astigmatic Gaussian beamlets through the diffracting aperture or optical system using the complex ray tracing technique, and (3) at the desired point “downstream,” the total field is found by the coherent recombination of new values of the elementary complex fields. Note that this recombination can be done in any space throughout the optical system.

This concept of decomposition, propagation, and recombination is not new since that is exactly what is being done in the angular spectrum method of Fourier optics, where the elementary field is a plane wave. However, difficulties can arise when the angular spectrum method is being used to propagate plane waves (of infinite extent) through general optical systems with bounded, nonplanar refracting (or reflecting) surfaces.²¹

Strictly speaking, Gaussian beams are also of infinite extent. However, the following desirable features of Gaussian beams make them ideal elementary fields for the decomposition process: (1) Gaussian beams are easy to propagate through optical systems, (2) Gaussian beams are fundamental solutions to the paraxial wave equation, (3) the Gaussian function is perfectly smooth (all derivatives

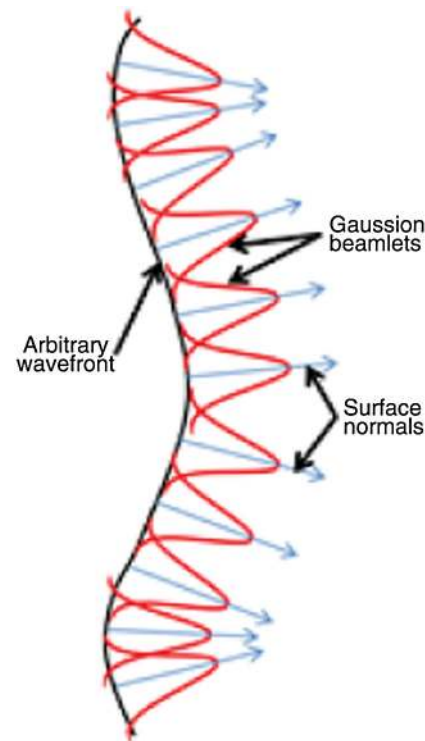


Fig. 5 Illustration of the rules for the spatial decomposition of arbitrary wavefronts into Gaussian beamlets.

are continuous), and (4) Gaussian beams are relatively compact. Indeed, for practical purposes, they can be considered to have a finite width.²¹

There are several rules for the decomposition of arbitrary wavefronts into Gaussian beamlets: (1) the base rays corresponding to individual Gaussian beamlets must be directed perpendicular to the local wavefront in accordance with the theorem of Malus and Dupin from geometrical optics, (2) the principal curvatures of the individual (astigmatic) Gaussian spherical waves must match the local principal curvatures of the arbitrary optical wave field, (3) the Gaussian beamlets must have an appropriate ratio of beamlet diameter to adjacent beamlet separation referred to as the overlap factor (this quantity is variable, with a default value of 1.5), and (4) the Gaussian beamlet density must be sufficient to adequately sample the aperture or wavefront deformations in the application of interest. These rules are illustrated schematically in Fig. 5. Note that for the case illustrated, some of the Gaussian beamlets are converging and some of them are diverging. Figure 6 illustrates the irradiance resulting from recombination of a uniform-amplitude plane wavefront truncated by a square diffracting aperture and decomposed into a superposition of Gaussian beamlets. Note the slight inaccuracies in the irradiance (ripple on the top and finite slope of the sides) of this representation of the truncated uniform-amplitude plane wavefront.

Figure 7 shows that an overlap factor of 1.5 between the individual Gaussian beamlets provides a compromise between this ripple in the irradiance and the finite slope of the sides of this representation of a truncated uniform-amplitude plane wavefront. This overlap factor can, and should, be varied according to the requirements of different applications.

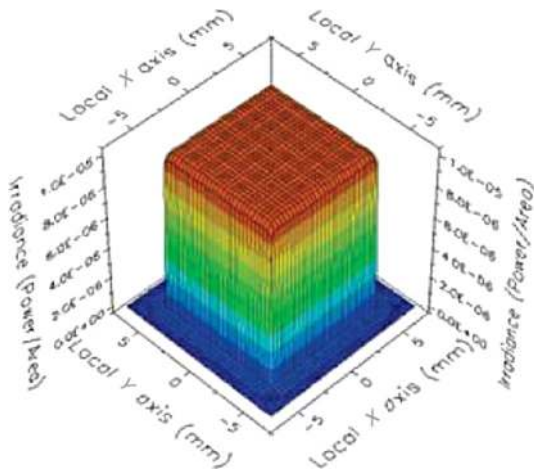


Fig. 6 Illustration of a uniform-amplitude plane wavefront truncated by a square aperture spatially decomposed into a 21×21 array of Gaussian beamlets.

The circles in Fig. 7(a) represent the $1/e^2$ radii of the 11×11 array of Gaussian beamlets into which a truncated uniform-amplitude plane wavefront has been decomposed. Overlap factors of 1.0, 1.5, and 2.0 are illustrated.

Figure 7(b) shows a false-color map of the irradiance distributions for this representation of the uniform-amplitude truncated plane wave for overlap factors of 1.0, 1.5, and 2.0. The overlap factor of 1.0 results in very substantial irradiance variations as indicated by the presence of local maxima located at the center of the individual Gaussian beamlets, while the distributions shown for overlap factors of 1.5 and 2.0 show very little variation by comparison. The broad width of the band around the edges of the truncated irradiance distribution for overlap factors of 1.5 and 2.0 qualitatively indicates the steepness of the slope of the sides of this representation of a truncated uniform-amplitude plane wavefront.

Finally, Fig. 7(c) shows the profile across the center of the irradiance distribution. An overlap factor of 1.0 results in irradiance variations or a ripple of $\pm 16\%$ but very steep edge slopes. An overlap factor of 2.0 virtually eliminates the irradiance ripple but exhibits a quite gentle, or broad, edge roll-off. And an overlap factor of 1.5 reduces the irradiance ripple to $\pm 0.35\%$ with a moderate edge roll-off.

Although Greynolds was not the first author to suggest the use of Gaussian beams as an elementary field for decomposition applications,^{22–25} to the authors' knowledge, his 1985 article²¹ was the first to suggest and demonstrate this powerful technique as a routine tool for the detailed optical analysis of diffraction effects in not only standard imaging systems, but also nonimaging concentrators, multimode fibers, interferometers, and synthetic aperture systems. Greynolds was also the chief architect of the first popular commercially available optical analysis ray-trace code that uses the decomposition of arbitrary optical wave fields into a superposition of Gaussian beamlets to accurately model physical optics phenomena.²⁶ Today, there are at least three such optical analysis codes in the market that graduating physicists and optical engineers should be aware of, even if not trained to be proficient users.^{26–28}

Although it is not clear from their literature whether, in fact, the Code V Beam Synthesis Propagation algorithm uses Gaussian beam decomposition, they have reported in detail upon simulating beam propagation with optical design software,²⁹ including modeling interferometers³⁰ and describing tests for assessing beam propagation algorithms.³¹

6 Modeling Physical Optics Phenomena by Complex Ray Tracing

We will now demonstrate a few of the basic physical optics phenomena that can be modeled by this powerful technique

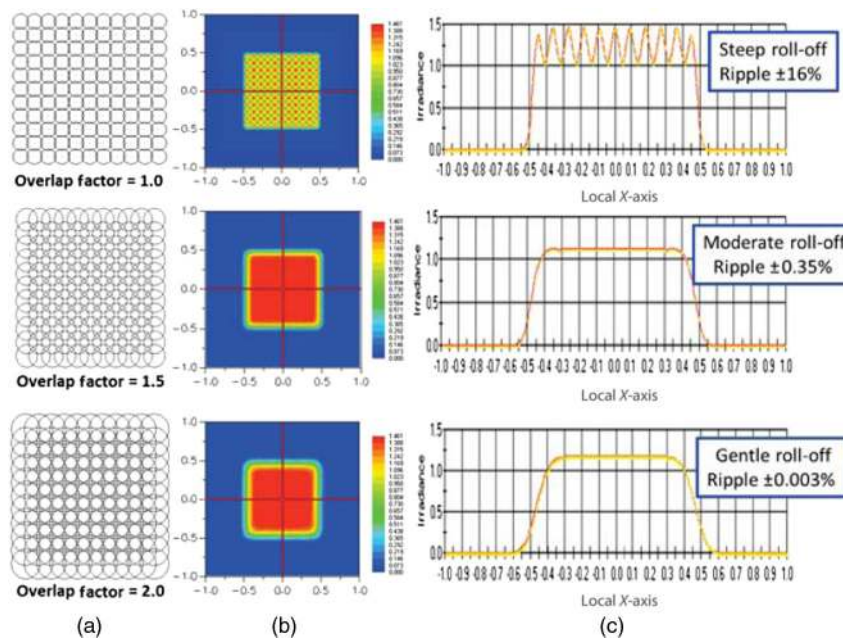


Fig. 7 (a) $1/e^2$ radii of the 11×11 array of Gaussian beamlets for different overlap factors, (b) False-color map of the irradiance distribution for this representation of the uniform-amplitude truncated plane wave for different overlap factors, and (c) illustration of the trade-off between the steep slopes and the amount of ripple upon the irradiance distribution for different overlap factors.

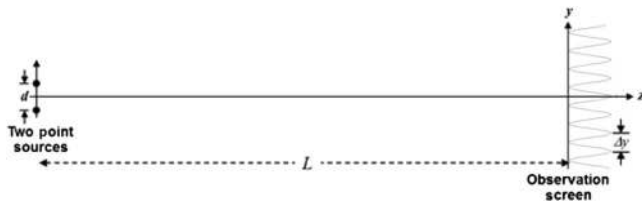


Fig. 8 Optical layout for producing straight-line interference fringes from two mutually coherent point sources.

of decomposition into Gaussian beamlets, propagation by complex ray tracing, and coherent re-combination of the resultant fields at any arbitrary point in space.

Consider first the interference produced by two mutually coherent point sources as shown in Fig. 8. Let the wavelength $\lambda = 0.5 \mu\text{m}$, the separation of the two point sources $d = 0.5 \text{ mm}$, and the distance to the observation plane (or analysis plane for our numerical computations) is parallel to the line connecting the two point sources as illustrated. We will observe a time-independent (or stationary) sinusoidal interference fringe pattern described by³

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\varphi_1 - \varphi_2). \quad (14)$$

The fringe pattern will exhibit a bright fringe if the phase difference, $\varphi_1 - \varphi_2$, is an integer multiple of 2π and a dark fringe if the phase difference is a half-integer multiple of 2π . Furthermore, if the two point sources are of equal strength (each yielding an irradiance of I at the analysis plane),

$$I = 4I_0 \cos^2[(\varphi_1 - \varphi_2)/2], \quad (15)$$

from which it follows that $I_{\min} = 0$ and $I_{\max} = 4I_0$. Since $L \gg d$, we know that the fringe period at the observation plane is given by

$$\Delta y = L\lambda/d = 1.0 \text{ mm}. \quad (16)$$

By launching an array of 51×101 Gaussian beamlets from each point source over the subtended angle of a

$4 \text{ mm} \times 8 \text{ mm}$ rectangular analysis region of interest, as shown in Fig. 9(a), and coherently recombining the complex amplitudes of all beamlets, as shown in Fig. 9(b), at each of 121×241 points on the rectangular analysis plane,

$$I_{j,k} = \sum_n E_n^2 = \left| \sum_n [U_{n_1}(j,k) + U_{n_2}(j,k)] \right|^2, \quad (17)$$

we can accurately calculate the interference pattern produced by these two mutually coherent point sources as illustrated in Fig. 9(c). Note that only the base rays of each Gaussian beamlet are being drawn, but realize that there are eight secondary rays associated with each base ray that complete the complex ray representation for each Gaussian beamlet.

Figure 9 provides an example of the graphics obtained from the commercially available software that utilizes the technique of decomposition of optical wave fields into a superposition of Gaussian beamlets, propagation of those beamlets by complex ray tracing, and coherent recombination of the resultant fields for modeling physical optical phenomena. These graphics provide not only qualitative visual interpretation of quantitative numerical results, but also real-time diagnostic graphical information to the user during the setup and preliminary analysis of models for sophisticated and advanced optical systems.

Figure 10 shows the direct comparison of this physical optics model to theoretical calculations. Note that by setting the radiant power on the analysis plane from each source equal to 32 (unit irradiance on the $4 \text{ mm} \times 8 \text{ mm}$ analysis plane), the resulting average irradiance of the interference pattern profile is 2 with $I_{\min} = 0$ and $I_{\max} = 4$ as expected.

One more basic interference phenomenon that we will demonstrate is the Newton's rings that one observes when a long-radius spherical optical surface is placed on an optically flat test plate in an optical shop. Figure 11(a) is a schematic drawing of a Newton interferometer.³² Figure 11(b) is the software model of such an instrument, and Fig. 11(c) shows the resulting interference pattern and the irradiance profile across the center of the pattern. We have not included the beam divider in the software model because the complex rays can pass right through the source to reach the analysis surface. Note that the center of the interference pattern is

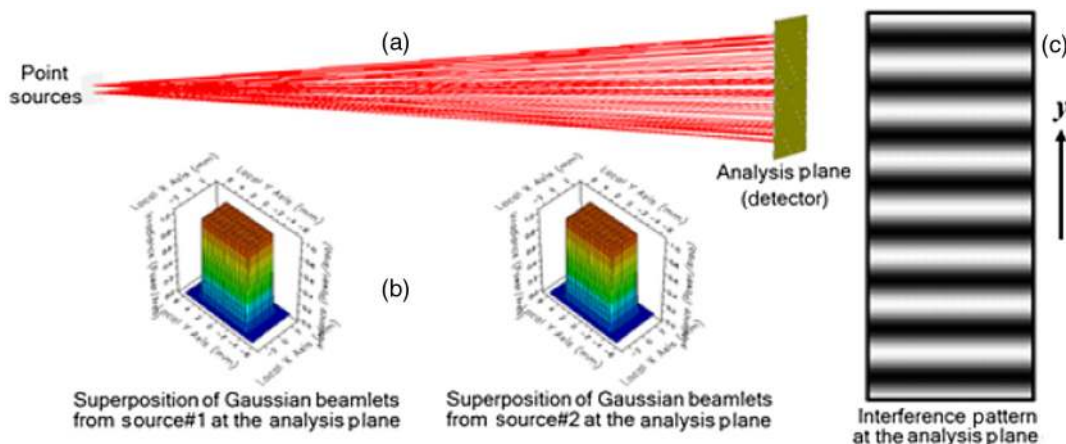


Fig. 9 (a) Illustration of the base rays traced from two mutually coherent point sources to a rectangular analysis plane, (b) graphical illustration of the superposition of Gaussian beamlets at the analysis plane, and (c) the interference pattern produced when all beamlets are coherently recombined on the analysis plane.

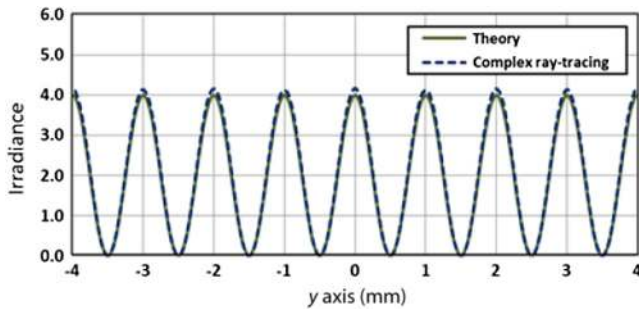


Fig. 10 Excellent agreement of numerical predictions of interference phenomena by complex ray tracing from the software package from Ref. 27 with theoretical calculations.

dark rather than bright because reflection from glass/air interface has a $\pi/2$ phase change associated with it, but the reflection from the subsequent air/glass interface does not.

The next physical optics phenomenon that we want to model by complex ray tracing is Fraunhofer diffraction from binary-amplitude apertures, i.e., transparent apertures in an otherwise opaque screen. The classical method would be to use the angular spectrum approach from Fourier optics,^{5,12} where the Fraunhofer diffraction pattern produced by a normally incident plane wave striking an aberration-free lens followed immediately by the diffracting aperture can be expressed as

$$E_2(x_2, y_2) = \frac{E_0}{\lambda^2 f^2} \left| \mathcal{F}\{t_A(x_1, y_1)\} \right|_{\substack{\xi = x_2/\lambda f \\ \eta = y_2/\lambda f}}^2. \quad (18)$$

Here, E_0 is the incident irradiance, λ is the wavelength of the incident light, f is the focal length of the lens, and $t_A(x_1, y_1)$ is the complex amplitude transmittance of the diffracting aperture. The script \mathcal{F} operator denotes the Fourier transform operation.

We will now calculate the Fraunhofer diffraction pattern of a square diffracting aperture by complex ray tracing and then compare the results to the analytical solution described by Eq. (18).

Following the method laid out in the previous section of decomposition into a superposition of Gaussian beamlets, propagation by complex ray tracing, and coherent recombination of the complex amplitudes of all beamlets at each point on the analysis plane, we calculate the Fraunhofer diffraction pattern of a square aperture shown in Fig. 12. The two-dimensional \sin^2 Fraunhofer diffraction pattern predicted by the software package from Ref. 27 is illustrated as a log irradiance plot (floor at 10^{-4}) for qualitative observation, and the predicted irradiance profile along the x axis is indistinguishable from theory down to values of 10^{-6} at the \sin^2 minima. The accuracy of the numerical calculations depends upon the sampling of both the source (density of beamlets) and the analysis plane (density of analysis points). For the results illustrated in Fig. 12, we decomposed the optical field emerging from the square aperture into a 71×71 square array of Gaussian beamlets and performed the recombination at an array of 101×101 points on the analysis plane.

Figure 13 illustrates the Fraunhofer diffraction patterns of a semicircular aperture, an equilateral triangular aperture, and a hexagonal aperture as calculated by the methods of complex ray tracing by the software package of Ref. 27. Note the superb detail in these numerically calculated diffraction patterns that could not be easily calculated analytically.

The optical field incident on each aperture was decomposed into a 51×51 grid of Gaussian beamlets whose grid extent is slightly oversized relative to the aperture. The base rays (i.e., Gaussian beamlets) were directed to a distant point producing a perfect spherical wavefront that was clipped by the aperture. The analysis plane was sampled at $\sim 241 \times 241$ points.

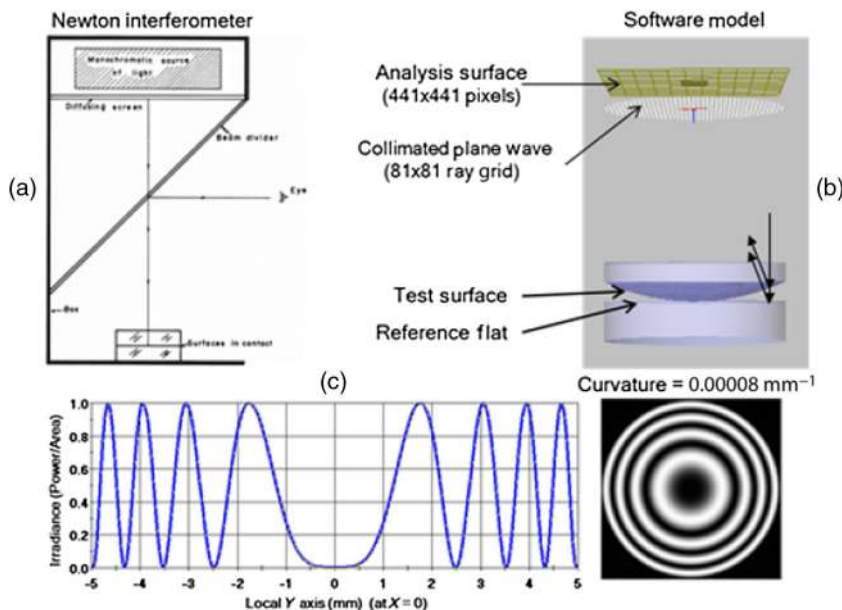


Fig. 11 (a) Schematic diagram of a Newton interferometer, (b) software model of the Newton interferometer (curvature of test surface exaggerated for clarity), and (c) the bull's-eye Newton's rings interference pattern with irradiance profile through the center.

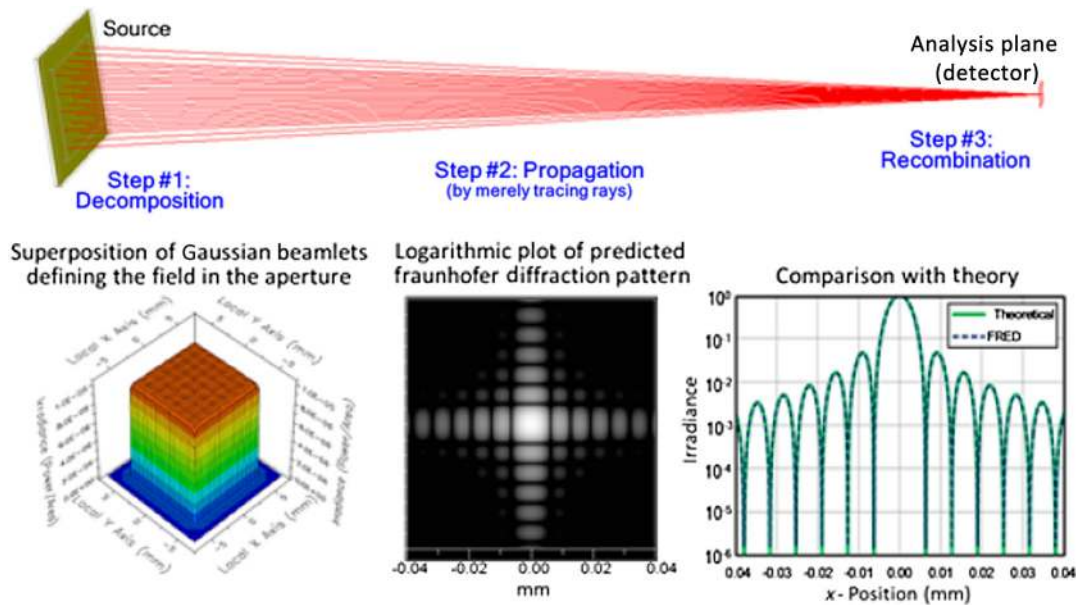


Fig. 12 The method for predicting Fraunhofer diffraction patterns by complex ray tracing is illustrated. The predicted two-dimensional log irradiance diffraction pattern is displayed and the predicted irradiance profile along the x axis is superposed on the theoretical profile for comparison.

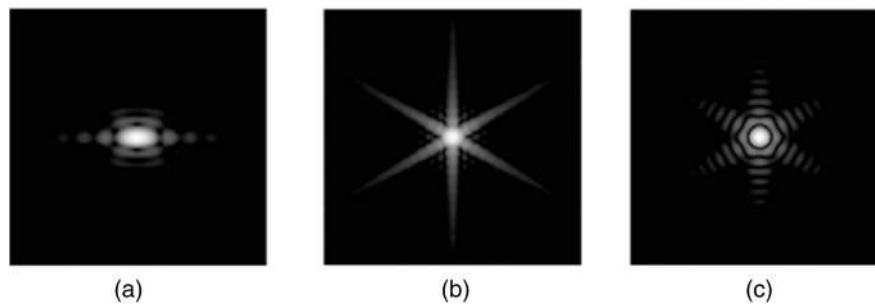


Fig. 13 Fraunhofer diffraction patterns of (a) semicircular, (b) equilateral triangular, and (c) hexagonal apertures predicted by complex ray tracing with the software package of Ref. 27 exhibit superb detail.

Fresnel diffraction patterns are merely defocused Fraunhofer diffraction patterns.^{33,34} Figure 14 shows defocused point spread functions (PSFs) of an unaberrated imaging system with an unobscured circular aperture as calculated by tracing complex rays and coherently combining the resulting complex amplitudes at each point on the analysis plane. Note the symmetry in the image as we go out of focus on either side of the paraxial image plane, with the expected dark spot appearing at the center of the pattern when we are an integral number of wavelengths out of focus. These predicted PSFs are virtually identical to the corresponding photographs of experimentally produced PSFs in the classical reference.³⁵

To obtain further insight into the structure of the optical image, we must study the irradiance distribution not only in selected defocused planes perpendicular to the optical axis, but also as a function of z (parallel to the optical axis) in the neighborhood of a focal plane. We can calculate the irradiance (or energy density) at any point throughout an optical system, even very near a focus or within a geometrical caustic region, by simply orienting and positioning the analysis grid in the proper position. Figure 15 shows isophotes

(contour lines of energy density), predicted by complex ray tracing, in a meridional plane near the focus of a converging spherical wave diffracted by a circular aperture. In this region, and in the absence of aberrations, this irradiance distribution is symmetrical about the paraxial focal plane as illustrated.

The isophote plot shown in Fig. 15(d) is virtually identical to Fig. 8.41 in Ref. 36, and it was calculated with absolutely no knowledge of Lommel functions. Clearly, using the technique of complex ray tracing, one could produce similar plots of the axial energy density in the presence of arbitrary aberrations, which would be very difficult to calculate analytically.

It should also be pointed out that this process of decomposition of an arbitrary optical wave field into a superposition of Gaussian beamlets, propagation by Arnaud's method of complex ray tracing, and, finally, the coherent recombination of optical wave fields on the desired observation plane provides either the resultant optical field (amplitude and phase) or the irradiance or energy density (by taking the squared modulus of the field).

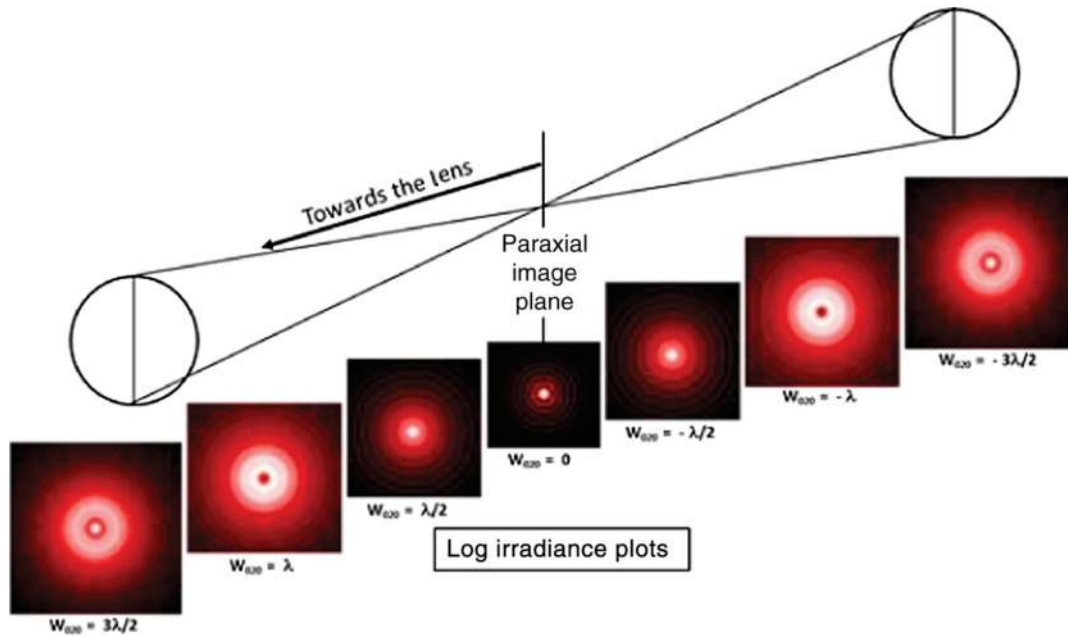


Fig. 14 Fresnel diffraction patterns (defocused Airy pattern) calculated by complex ray tracing.

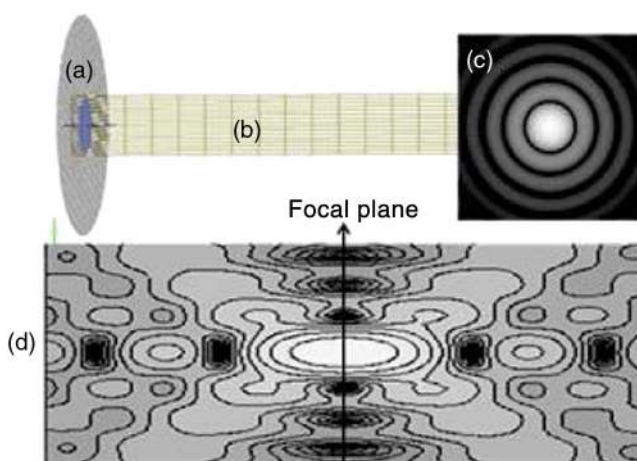


Fig. 15 (a) Source and diffracting aperture, (b) axial analysis plane parallel to the optical axis, (c) irradiance distribution in the transverse focal plane, and (d) isophote contours of the energy density through the focus in the meridional plane.

Siegman discusses Fresnel diffraction due to a uniform-amplitude plane wave incident upon a circular aperture.¹¹ Figure 16(a) shows the radial profile of two such Fresnel diffraction patterns, one at a Fresnel number $N = 5$ (bright spot in the center) and one at a Fresnel number $N = 10$ (dark spot in the center).

Note that the left side of these profiles (taken from Ref. 11) is due to careful experimental measurements and the right side is theoretically calculated from the Huygens' integral. Figure 16(b) shows the corresponding Fresnel diffraction profiles calculated by the complex ray tracing technique discussed in this paper. The sampling density of the analysis plane for both Figs. 16(b) and 16(c) was set at 241×241 . Decomposition with a higher beamlet sampling [indicated in Fig. 16(b)] in the $N = 10$ case was empirically determined to be necessary due to the finer structure

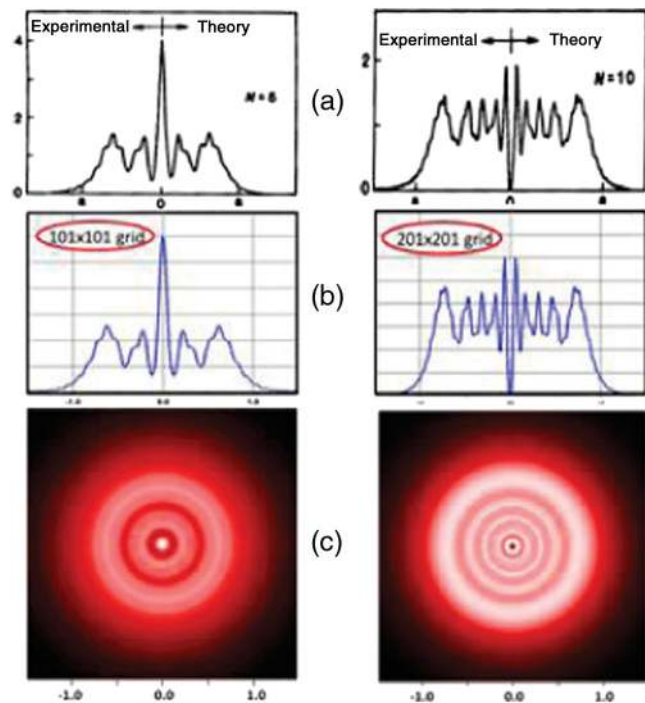


Fig. 16 (a) Experimental and theoretical predictions of Fresnel diffraction patterns by a uniformly illuminated circular aperture, (b) virtually identical irradiance profiles calculated by complex ray tracing (aperture sampling indicated), and (c) two-dimensional Fresnel diffraction patterns calculated by complex ray tracing.

being resolved in the diffraction pattern. Reducing the aperture sampling by a factor of two significantly reduced the accuracy of the predicted profile for both Fresnel numbers. Figure 16(c) displays the full two-dimensional Fresnel diffraction pattern calculated by complex ray tracing. Once again, the technique of decomposing into Gaussian beamlets, propagation by complex ray tracing, and then coherently

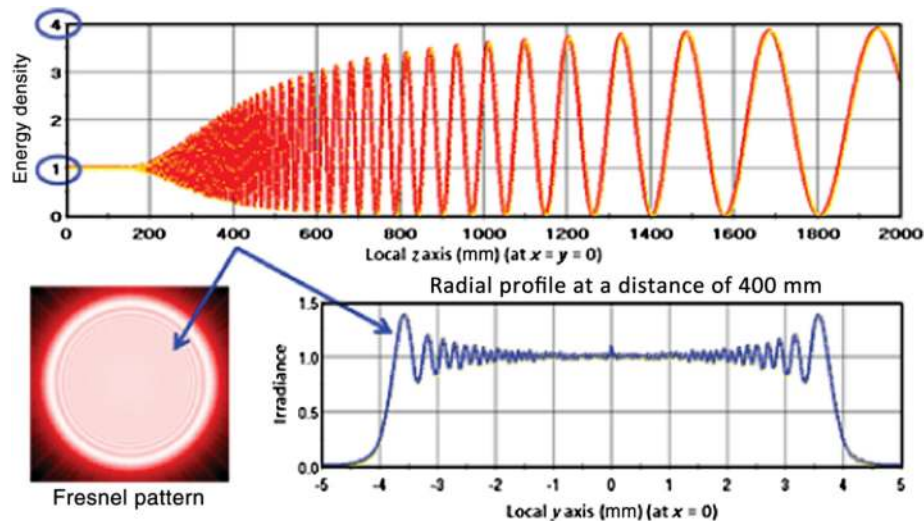


Fig. 17 Illustration of the oscillatory behavior of the on-axis irradiance in the near-field behind an 8-mm-diameter circular aperture uniformly illuminated with a unit-amplitude, normally incident collimated beam of wavelength $0.6328 \mu\text{m}$. Also shown are the two-dimensional near-field diffraction pattern and a radial profile, at a distance of 400 mm behind the circular aperture.

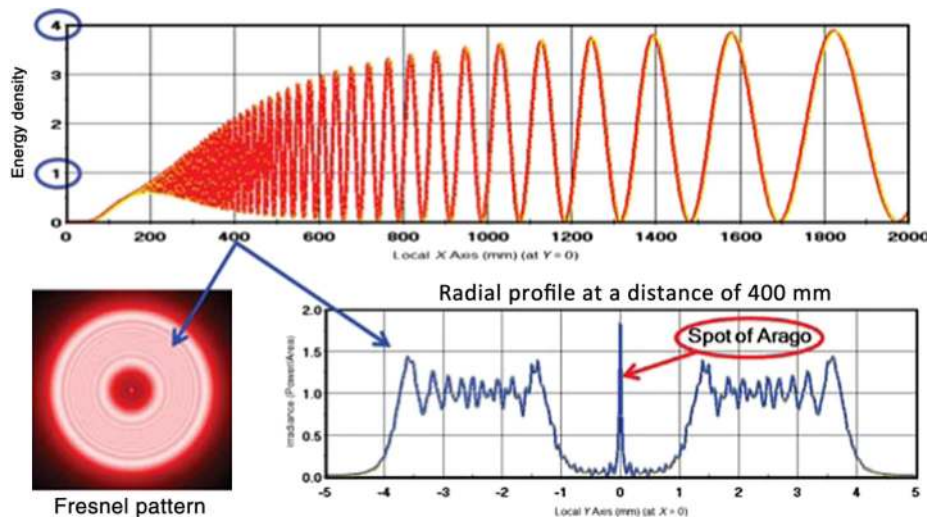


Fig. 18 Illustration of the on-axis irradiance in the near-field behind an annular aperture (8 mm outer diameter, 2 mm inner diameter) illuminated as in Fig. 17. Note that the irradiance is zero immediately behind the aperture, increasing rapidly and oscillating as before. Also note Poisson's bright spot at the center of this diffraction pattern at a distance of 400 mm.

re-combining the resulting complex amplitudes provides excellent results.

Very near-field diffraction phenomena can also be readily modeled by complex ray tracing. Consider the on-axis irradiance behind a circular aperture illuminated by a unit-irradiance plane wave. Most physics or optics textbooks illustrate this on-axis irradiance throughout the Fresnel and Fraunhofer regions, but few show how these oscillations diminish in the near-field and asymptotically approach unity immediately behind the circular aperture.³⁷ Figure 17 illustrates this on-axis oscillatory behavior in the near-field for an 8-mm-diameter circular aperture for an incident wavelength of $0.6328 \mu\text{m}$.

Recall that for unit incident irradiance, these oscillations range from 0 to 4 (with a mean of 2) throughout the Fresnel

region. Also shown are the two-dimensional near-field diffraction pattern and a radial profile at a distance of 400 mm behind the circular aperture.

If we add a 2-mm-diameter circular obscuration (i.e., model an annular aperture), we observe similar diminishing on-axis oscillatory behavior asymptotically approach zero immediately behind the circular aperture as shown in Fig. 18. However, at a distance of 400 mm behind the aperture, we now observe the spot of Arago (or Poisson's bright spot) at the center of the geometrical shadow of the central obscuration.³⁸

It is also a very simple matter to produce an aberrated diffraction PSF tree as illustrated in Fig. 19 by applying various Zernike deformations to an optical surface. This configuration is very useful in studying the nature and symmetry of

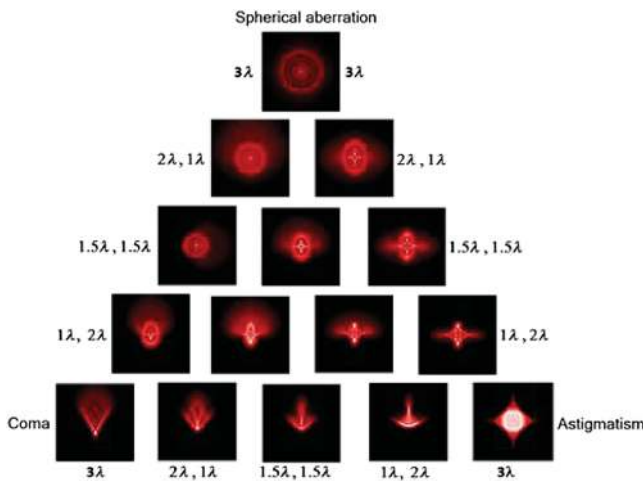


Fig. 19 This diffraction point spread function tree for low-order aberrations is readily produced by placing Zernike deformations on a focusing mirror and utilizing the complex ray tracing technique, which is the subject of this paper.

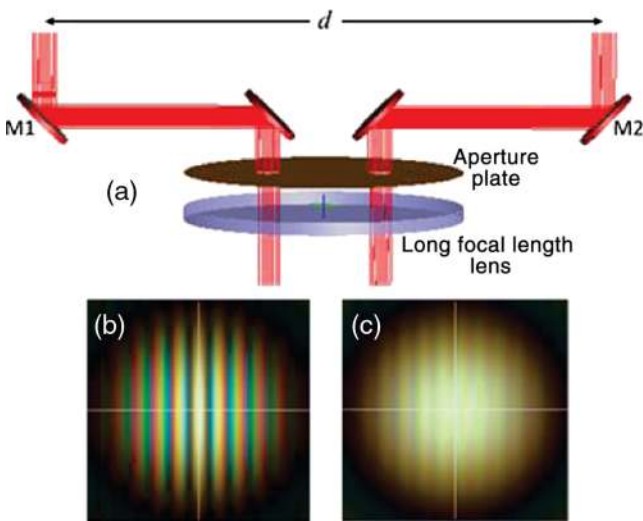


Fig. 20 (a) The optical layout of a Michelson stellar interferometer as modeled by the commercially available software package of Ref. 27, and (b) and (c) illustrations of high-visibility and low-visibility interferograms calculated by complex ray tracing.

PSFs degraded by various combinations of primary aberrations. Spherical aberration exhibits rotational symmetry, astigmatism exhibits bilateral symmetry, and coma exhibits lateral symmetry. Hence, only the aberrated PSFs going down the right side of the tree retain at least bilateral symmetry, with the remaining aberrated PSFs (containing some coma) retaining only lateral symmetry.

A wide variety of additional physical optics phenomena can be accurately modeled by complex ray tracing. Examples include Babinet’s principle, the Talbot effect, frustrated total internal reflection, speckle phenomena, pulse propagation, optical fiber coupling efficiency, and polarization effects, such as Brewster’s law, birefringent materials, and the Maltese cross.³⁹

Even partial coherence effects can be accurately modeled by complex ray tracing. Figure 20 illustrates the optical layout of a Michelson stellar interferometer for determining the small angular diameter of remote astronomical bodies. Note the out-rigger mirrors M1 and M2 of variable spacing d and the aperture plate with two holes that will produce Young’s interference fringes in the focal plane of the long focal length telescope objective lens. Also shown are examples of high-visibility and low-visibility white-light interferograms calculated by complex ray tracing for two different values of the mirror separation d . By adjusting the mirror separation until the visibility goes to zero, the star angular diameter can be determined.³⁶

Reference 39 provides a detailed discussion of how the software package of Ref. 27 models this Michelson stellar interferometer by creating an extended incoherent source of a specific size with a collection of randomly positioned, mutually incoherent point sources for each of the discrete wavelengths making up a desired spectrum (there is a feature in the code for creating any desired spectrum by specifying a mean wavelength, spectral bandwidth, amplitude function, and number of discrete wavelengths). The resulting interference pattern is produced by Gaussian beam decomposition, propagation by complex ray tracing, and then recombination by summing equal wavelengths coherently and different wavelengths incoherently. The resulting degree of partial coherence compares very well with predictions of the van Cittert–Zernike theorem.³⁶

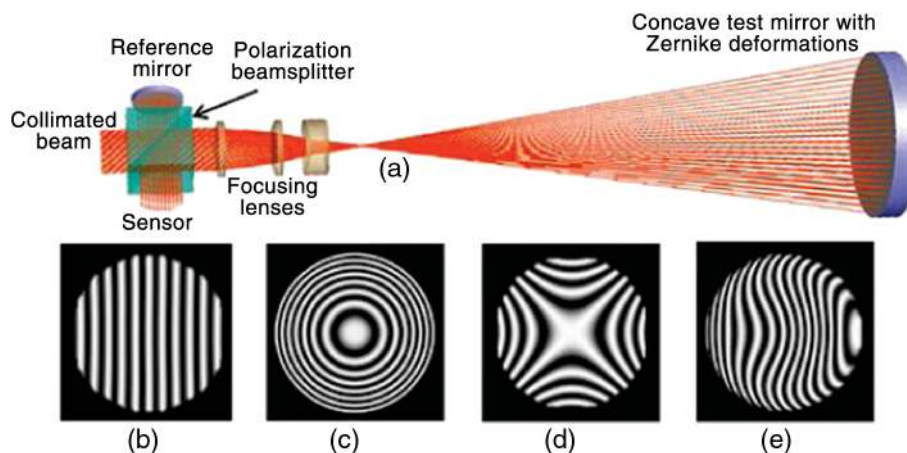


Fig. 21 (a) Software model illustrating the optical layout of a laser unequal path interferometer; (b) through (e) are examples of interferograms obtained by putting different Zernike deformations on the large test mirror.

Finally, Fig. 21 is an example of a complete precision optical instrument that can be modeled by complex ray tracing for accurate performance analysis. Detailed optical prescriptions and optomechanical details can be readily downloaded directly into the software package of Ref. 27 from conventional optical design codes and CAD programs. A laser unequal path interferometer (LUPI) has, thus, been modeled. The LUPI consists of a collimated laser beam incident upon a polarization beamsplitter, a small flat reference mirror, and several lenses producing an aberration-free point image that then expands to fill a large concave test mirror. The returning beam is then combined with the plane reference beam by the beamsplitter to form the interferogram. Several resulting example interferograms are illustrated.

7 Summary and Conclusions

A very versatile method of accurately modeling a wide variety of physical optics phenomena has been discussed and the results demonstrated. The method involves three steps: (1) decomposition of an arbitrary optical wave field into a superposition of Gaussian beamlets, (2) propagation of the individual beamlets by complex ray tracing, and (3) the coherent recombination of the individual Gaussian beamlets at the observation plane or sensor. This technique has been extensively used in specialized optical analysis software packages by the aerospace industry and in research laboratories for over three decades. However, even the concept of describing a diffracted wave field as a superposition of Gaussian beamlets as an alternative to the classical approaches (superposition of spherical wavelets and the superposition of plane waves) does not seem to be generally included in current physics or optics textbooks or course materials. The authors' hope is that this article might raise the awareness of this powerful optical analysis method within the optical engineering community and in our educational institutions.

References

1. J. A. Arnaud, "Representation of Gaussian beams by complex rays," *Appl. Opt.* **24**, 538–543 (1985).
2. A. W. Greynolds, "Fat rays: a synthesis of physical and geometrical optics with Gaussian beamlets," in *1990 Int. Optical Design Conf.*, Vol. 10, pp. 3–4, Optical Society of America, Washington, DC (1990).
3. E. Hecht, *Optics*, 2nd ed., Addison-Wesley Publishing Co., Reading, MA, (1987).
4. G. W. Forbes and M. A. Alonzo, "What on earth is a ray and how can we best use them?," *Proc. SPIE* **3482**, 22–31 (1998).
5. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996).
6. G. W. Forbes and M. A. Alonzo, "Using rays better. I. Theory for smoothly varying media," *J. Opt. Soc. Am. A* **18**, 1132–1145 (2001).
7. G. W. Forbes and M. A. Alonzo, "Using rays better. II. Ray families for simple wave fields," *J. Opt. Soc. Am. A* **18**, 1146–1159 (2001).
8. G. W. Forbes and M. A. Alonzo, "Using rays better. III. Error estimates and illustrative applications in smooth media," *J. Opt. Soc. Am. A* **18**, 1357–1370 (2001).
9. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, 2nd ed., John Wiley & Sons Inc., Hoboken, New Jersey (2007).
10. H. Kogelnik and T. Li, "Laser beams and resonators," *Appl. Opt.* **5**, 1550–1567 (1966).
11. A. E. Seigman, *Lasers*, University Science Books, Mill Valley, CA (1986).
12. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York (1978).
13. J. A. Arnaud and H. Kogelnik, "Gaussian light beams with general astigmatism," *Appl. Opt.* **8**(8), 1687–1693 (1969).
14. J. A. Arnaud, "Degenerate optical cavities," *Appl. Opt.* **8**, 189–196 (1969).
15. J. A. Arnaud, "Hamiltonian theory of beam mode propagation," in *Progress in Optics*, E. Wolf, Ed., Vol. 11, pp. 249–304, North-Holland Publishing Company, Amsterdam (1973).
16. E. Delano, "First order design and the diagram," *Appl. Opt.* **2**, 1251–1256 (1963).
17. R. V. Shack, "Analytic system design with pencil and ruler—the advantages of the diagram," *Proc. SPIE* **39**, 127–140 (1973).
18. D. Kessler and R. V. Shack, "Diagram, a powerful optical design method for laser systems," *Appl. Opt.* **31**(15), 2692–2707 (1992).
19. R. Herloski, S. Marshall, and R. Antos, "Gaussian beam ray-equivalent modeling and optical design," *Appl. Opt.* **22**(8), 1168–1174 (1983).
20. A. W. Greynolds, "Vector formulation of the ray-equivalent method for general Gaussian beam propagation," *Proc. SPIE* **679**, 129–133 (1986).
21. A. W. Greynolds, "Propagation of generally astigmatic Gaussian beams along skew ray paths," *Proc. SPIE* **560**, 33–50 (1985).
22. D. Gabor, "Theory of communications," *J. Inst. Electr. Eng. (Lond.)* **93**(3), 429–457 (1946).
23. M. J. Bastiaans, "Gabor's expansion of an optical signal into a discrete set of Gaussian beams," *Optik* **57**, 95–102 (1980).
24. M. M. Popov, "A new method of computation of wavefields using Gaussian beams," *Wave Motion* **4**, 85–97 (1982).
25. P. D. Einziger, S. Raz, and M. Shapira, "Gabor representation and aperture theory," *J. Opt. Soc. Am. A* **3**(4), 508–522 (1986).
26. www.breault.com/software/asap.
27. www.photonengr.com/software.
28. www.synopsys.com/codev.
29. B. D. Stone and T. J. Bruegge, "Practical considerations for simulating beam propagation: a comparison of three approaches," *Proc. SPIE* **4832**, 359 (2002).
30. B. D. Stone and K. P. Thompson, "Modeling interferometers with lens design software: beyond ray-based approaches," *Proc. SPIE* **7427**, 74270A (2009).
31. B. D. Stone, "Tests for assessing beam propagation algorithms," *Proc. SPIE* **8171**, 81710E (2011).
32. D. Malacara, *Optical Shop Testing*, pp. 3–4, John Wiley & Sons, New York (1978).
33. J. E. Harvey and R. V. Shack, "Aberrations of diffracted wave fields," *Appl. Opt.* **17**(18), 3003–3009 (1978).
34. J. E. Harvey, "A Fourier treatment of near-field scalar diffraction theory," *Am. J. Phys.* **47**(11), 974–980 (1979).
35. C. Michel, M. Françon, and J. C. Thrierr, *Atlas of Optical Phenomena*, Springer-Verlag, Berlin (1962).
36. M. Born and E. Wolf, *Principles of Optics*, 6th ed., pp. 271, 440, 508, Oxford, England (1980).
37. J. E. Harvey and A. Krywonos, "Axial irradiance distribution throughout the whole space behind an annular aperture," *Appl. Opt.* **41**(19), 3790–3795 (2002).
38. J. E. Harvey and J. L. Forgham, "The spot of Arago: new relevance for an old phenomenon," *Am. J. Phys.* **52**(3), 243–247 (1984).
39. www.photonengr.com/wp-content/uploads/2011/04/appNoteCoherence.pdf.

James E. Harvey is a retired associate professor from CREOL: the College of Optics and Photonics at the University of Central Florida, and currently a senior optical engineer with Photon Engineering, LLC in Tucson AZ. He has a PhD in optical sciences from the University of Arizona and is credited with over 200 publications and conference presentations in diverse areas of applied optics. He is a member of OSA and a Fellow and past board member of SPIE.

Ryan G. Irvin received a bachelor's degree in physics from Arizona State University in 2005, and a master's degree in optical sciences from the University of Arizona in 2009. He is currently a senior optical engineer with Photon Engineering, LLC, in Tucson, AZ.

Richard N. Pfisterer is co-founder and president of Photon Engineering, LLC. He received his bachelor's and master's degrees in optical engineering from the Institute of Optics at the University of Rochester in 1979 and 1980, respectively. Previously he served as head of optical design at TRW (now Northrup-Grumman) and senior optical engineer at Breault Research Organization. He is credited with 20 articles and conference presentations in the areas of optical design, stray light analysis, and phenomenology. He is a member of OSA and SPIE.