Ray-based method for simulating cascaded diffraction in high-numerical-aperture systems

Mout, Marco; Flesch, Andreas; Wick, Michael; Bociort, Florian; Petschulat, Joerg; Urbach, Paul

DOI
10.1364/JOSAA.35.001356

Publication date
2018

Document Version
Accepted author manuscript

Published in
Journal of the Optical Society of America A: Optics and Image Science, and Vision

Ray-based method for simulating cascaded diffraction in high NA systems

Marco Mout[1,2,*], Andreas Flesch[1], Michael Wick[3], Florian Bociort[2], Joerg Petschulat[1], and Paul Urbach[2]

[1]Carl Zeiss AG, Carl-Zeiss-Straße 22, 73447 Oberkochen, Germany
[2]Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
[3]Coburg University of Applied Sciences, Friedrich-Streib-Straße 2,96450 Coburg, Germany
[*]Corresponding author: marco.mout@zeiss.com

17 July 2019

Abstract

The electric field at the output of an optical system is in general affected by both aberrations and diffraction. Many simulation techniques treat the two phenomena separately, using a geometrical propagator to calculate the effects of aberrations and a wave-optical propagator to simulate the effects of diffraction. We present a ray-based simulation method that accounts for both the effects of aberrations and of diffraction within a single framework. The method is based on the Huygens-Fresnel principle, is entirely performed using Monte Carlo ray tracing, and, in contrast to our previously published work, is able to calculate the full electromagnetic field. The method can simulate the effects of multiple diffraction in systems with a high numerical aperture (NA).

1 Introduction

In the development of optical systems it can be extremely useful to predict the optical performance of a system by computer simulations. For devices such as head-mounted displays (HMDs), this requires the consideration of both (multiple) diffraction and refraction.

Frequently, macroscopic optical systems are simulated using geometrical optics. Although it is possible to keep track of the polarization of the electric field during ray tracing [4, 26] (an option in, e.g., the commercial ray tracers OSLO [18, p. 60–65], ZEMAX [28, p. 216–222], and CODE V [6]), and one can propagate an electric field using geometrical optical ray tracing techniques (see, e.g., Wyrowski *et al.* [27]), geometrical optics as such does not account for diffraction effects and breaks down in the focal region of a system. Since the Point Spread Function (PSF), which is the field of a point object in the focal region, is often very important, several methods have been developed to simulate this field. Because the wavelength of light is very small compared to the size of the optical components (e.g., lenses) this problem is far from trivial.

One method that includes diffraction effects is the exit pupil diffraction method. This method uses ray tracing to calculate the field in the exit pupil, which is the image of the aperture of the optical system in the image space, and then uses a diffraction integral to perform the last propagation step from the exit pupil to the focal plane. Many commercial ray tracers use this method to compute scalar PSFs, and when combined with polarization ray tracing the exit pupil diffraction method can also provide the vectorial PSF. The ray tracing program OSLO has this option [18, p. 365-368]. A limitation of the exit pupil diffraction method is that it assumes all diffraction to occur at the exit pupil. For systems with multiple diffracting surfaces (e.g., an HMD with a diffractive element and an aperture) the exit pupil diffraction method is insufficient.

One way to deal with such cascaded diffraction is to sequentially apply diffraction integrals (e.g., the Fresnel integral), which can propagate the field through homogeneous media, and geometrical optical propagators, which are better suited for propagating fields through optical interfaces. Several commercial simulation packages (e.g., VirtualLab by

LightTrans and ZEMAX) provide this option to the user. To alternate between the propagation methods, one needs to resample the field which may require a lot of memory [28, p. 628], especially when the local wavefront is strongly curved.

Another approach for cascaded diffraction simulations is the decomposition of the field into subfields that can be propagated using geometrical optical techniques. The Gaussian Beam Decomposition (see e.g., Greynolds [11]), which is for instance used in ASAP, or the Beam Synthesis Propagation of CODE V [5] are two commercially available examples. They depend on resampling whenever diffraction occurs and a subsequent propagation of the subfields using geometrical optics would cause inaccuracies.

There exists a range of other ray-based diffraction simulations that initiate additional rays at diffracting surfaces. The geometrical theory of diffraction proposed by Keller [15] initiates 'diffracted rays' from the edges of apertures (or other diffracting surfaces) (see [13] for an overview of the method). Heisenberg uncertainty ray bending initiates secondary rays near the edge of an aperture, but a distribution of the secondary rays over the radiation cones that consistently gives correct results could not be determined conclusively [10, 22, 9]. A method called stable aggregate of flexible elements [8] provides accurate results, but is limited to two-dimensional configurations.

Initially ray-based diffraction methods using the Huygens-Fresnel principle [7, 19, 20, 21] were also limited to two spatial dimension or to problems which were invariant in one of the three spatial dimensions. In [17] we proposed a ray-based diffraction method based on the Huygens-Fresnel principle and derived from the Rayleigh-Sommerfeld diffraction integral, that was capable of simulating the diffraction in a three-dimensional optical system. We showed results for the propagation of a coherent scalar field through an optical system with aberrations and multiple diffracting surfaces. Mahan *et al.* [16] recently presented a similar method for the simulation of diffraction at a single aperture in free-space and convincingly validated the model using experimental data.

The current paper is an extension of our previously published method [17], known as Huygens-Fresnel path integration (HFPI), to vectorial optics. As such, it inherits many of the properties of HFPI. It can be used to simulate macroscopic optical systems, and accounts for all geometrical aberrations and for the diffraction effects (e.g., introduced by multiple apertures in the system). Like scalar HFPI, vectorial HFPI is robust but computationally intensive. It bears some conceptual resemblance to the vectorial ray-based diffraction integral (VRBDI) introduced by Andreas *et al.* [2], but its implementation and intended domain of application differ. While VRBDI was developed to simulate the propagation of a diffracting beam (i.e., a laser beam) through a non-diffracting optical system for interferometry, HFPI is developed for the simulation of multiple diffraction in optical systems like the before mentioned HMDs. To simplify the simulation of multiple diffraction, the Monte Carlo approach of HFPI (in contrast to the grid sampling of VRBDI) does not require an explicit resampling at diffracting surfaces.

The working principle of HFPI is to use geometrical optics in the subparts of the system where it is valid. At surfaces where the geometrical optical approximations break down, secondary sources are initiated according to the Huygens-Fresnel principle. These secondary sources can again be propagated using geometrical optics. Multiple diffraction is accounted for by applying the Huygens-Fresnel principle in multiple planes. The electromagnetic field is calculated by coherently summing the subfields of the secondary sources.

The main contribution of this paper is the extension from scalar HFPI to vectorial HFPI. Since light is a vectorial phenomenon, this extension also offers the possibility of a more fundamental derivation of the method. In Section 2 this derivation is provided. In Section 3 the implementation is described and in Section 4 the results of HFPI simulations are presented. These results also demonstrate the importance of taking cascaded diffraction into account. In Section 5 we evaluate the properties of HFPI before drawing the final conclusions in Section 6.

2 Theory

The core idea of our method is to successively propagate the kernel of the m-theory diffraction integrals using geometrical optics. We first present the m-theory diffraction integrals in Section 2.1. In the following Section 2.2 it will be shown that these diffraction integrals can be interpreted as a decomposition of the electric field into magnetic dipoles. These elementary fields are of such a form that they can be propagated using geometrical optics. In Section 2.3 we will apply this diffraction integral multiple times, to arrive at a cascaded diffraction integral for multiple diffraction in an optical system. The topic of Section 2.4 is the propagation of the electric field by a Monte Carlo ray tracer, since this will be the propagator of choice for HFPI. Note that throughout this paper we

use the negative time convention, $\vec{\tilde{E}}(\vec{r},t) = \vec{E}(\vec{r})\exp(-\mathrm{i}\omega t)$, with $\vec{\tilde{E}}(\vec{r},t)$ the vectorial electric field in the point $\vec{r}$ at time t, and ω the frequency of the light. Since we are interested in steady state solutions, the factor $\exp(-\mathrm{i}\omega t)$ will be omitted.

2.1 The m-theory diffraction integral

A diffraction integral expresses the electric field, $\vec{E}$, in a point of interest, $\vec{r_1}$, as an integral over field quantities over a surface and/or in a volume. In this paper we will apply the m-theory diffraction integral, derived by, e.g., Severin [23], which expresses $\vec{E}(\vec{r_1})$ as a function of the electric field in a plane

$$\vec{E}(\vec{r_1}) = \frac{-\mathrm{i}k}{2\pi}\iint \frac{\exp(\mathrm{i}k\rho)}{\rho}\hat{\rho}\times\left(\hat{n}_0\times\vec{E}\right)\,\mathrm{d}S_0, \qquad (1)$$

where k is the wavenumber of the light, $\vec{\rho}$ is a vector (with length ρ and normalized direction $\hat{\rho}$) from the point of integration $\vec{r_0}$ to the point of interest $\vec{r_1}$, S_0 is the plane of integration and $\hat{n}_0$ is the normal of this plane that points towards the half space containing $\vec{r_1}$. All the vectors are indicated by an arrow and normalized vectors have a hat (e.g., $\hat{n}_0$). By taking the curl with respect to $\vec{r_1}$, one can obtain the diffraction integral for the magnetic field

$$\vec{H}(\vec{r_1}) = \frac{-\mathrm{i}k^2}{2\pi\omega\mu}\iint \frac{\exp(\mathrm{i}k\rho)}{\rho}\hat{\rho}\times\left[\hat{\rho}\times\left(\hat{n}_0\times\vec{E}\right)\right]\,\mathrm{d}S_0, \quad (2)$$

with μ the magnetic permeability of the medium, which we will assume to be equal to the permeability of free-space.

In deriving Eqs. (1, 2) one must assume that the surface of integration is an infinitely extending flat plane and that the field in the point of interest only depends on the field in this plane. Furthermore, one must assume that the point of interest is many wavelengths away from the plane of integration.

2.2 Geometrical propagation of the integral kernel

The integrals of the m-theory, Eqs. (1) and (2), are only valid within a homogeneous medium. Since optical systems are only piecewise homogeneous, these diffraction integrals cannot describe the propagation through an entire system. Applying separate diffraction integrals for every homogeneous region (e.g., a lens or free space) is numerically challenging and requires the solution of boundary value problems at every interface (e.g., between glass and air) to determine transmission losses. Furthermore, Eqs. (1) and (2) require a flat plane, whereas interfaces are frequently curved. To simulate optical system comprised of multiple optical media we therefore choose a different approach based on a physical interpretation of these diffraction integrals.

Equations (1, 2) are integrations over the following two quantities

$$\vec{E}^{(m)}(\vec{r_0},\vec{r_1},\vec{E}(\vec{r_0})) \equiv -2\mathrm{i}kG_H\hat{\rho}\times\left(\hat{n}_0\times\vec{E}(\vec{r_0})\right), \qquad (3)$$

$$\vec{H}^{(m)}(\vec{r_0},\vec{r_1},\vec{E}(\vec{r_0})) \equiv \frac{-2\mathrm{i}k^2}{\omega\mu}G_H\hat{\rho}\times \left[\hat{\rho}\times\left(\hat{n}_0\times\vec{E}(\vec{r_0})\right)\right], \qquad (4)$$

where

$$G_H(\rho) = \frac{\exp(\mathrm{i}k\rho)}{4\pi\rho} \qquad (5)$$

is a Green's function of the Helmholtz equation. These are equal to the magnetic and electric far field of a magnetic dipole $\vec{m}$ in $\vec{r_0}$ (which is why the name m-theory was introduced by Karczewski and Wolf [14])

$$\vec{E}^{(md)}(\vec{r_0},\vec{r_1},\vec{m}(\vec{r_0})) = -k^2\sqrt{\frac{\mu}{\epsilon}}G_H\hat{\rho}\times\vec{m}, \qquad (6)$$

$$\vec{H}^{(md)}(\vec{r_0},\vec{r_1},\vec{m}(\vec{r_0})) = -k^2G_H\hat{\rho}\times(\hat{\rho}\times\vec{m}) \qquad (7)$$

(see, e.g., Jackson [12]), with dipole moment

$$\vec{m}(\vec{r_0}) = \frac{2\mathrm{i}}{k}\sqrt{\frac{\epsilon}{\mu}}\hat{n}_0\times\vec{E}(\vec{r_0}). \qquad (8)$$

The field predicted by the m-theory, Eq. (1), is therefore the same as the field of infinitely many magnetic dipoles, located in the integration plane S_0

$$\vec{E}(\vec{r_1}) = \iint \vec{E}^{(md)}(\vec{r_0},\vec{r_1},\vec{m}(\vec{r_0}))\,dS_0 \qquad (9)$$

where the strength and direction of the dipoles are given by Eq. (8).

The diffraction integrals of the m-theory can be seen as a twofold process. First the decomposition of the field in S_0 into magnetic dipoles according to Eq. (8) and then the propagation of these dipole fields through a homogeneous medium to the point $\vec{r_1}$ using Eq. (6) or (7).

Since these subfields are equivalent to the fields of magnetic dipoles, they have the properties of physical electromagnetic fields (e.g., being divergence free). Although this can help in physically understanding the diffraction integral, the decomposition is primarily useful if the subfields are easier to propagate than the original field. It turns out that the elementary fields of the m-theory integrals can be propagated through a series of homogeneous media (i.e., an optical system) using geometrical optics. To show this, we must first briefly introduce geometrical optics.

In geometrical optics, fields are propagated along paths that are perpendicular to surfaces of equal phase. These paths, known as rays, are straight in homogeneous media and change direction at optical interfaces according to Snell's law

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2), \tag{10}$$

where n is the refractive index of the medium, θ the angle between the surface normal and the direction of propagation, and the indices indicate the quantities at the two sides of the interface. Geometrical optics can be derived by using the following ansatz for the electric field (see, e.g., Born and Wolf [3, p. 117–127])

$$\vec{E}(\vec{r}) = \vec{E}'(\vec{r}) \exp(\mathrm{i}k_0 s(\vec{r})), \tag{11}$$

where $\vec{E}'$ can be complex, k_0 is the wavenumber in free-space, and s is a real function that determines the phase propagation. The direction of the rays are related to Eq. (11) through

$$\hat{\rho} = \frac{\nabla s}{n}, \tag{12}$$

with $\hat{\rho}$ the ray directions. From the electric field carried by a ray, Eq. (11), one can derive the corresponding magnetic field by taking the curl

$$\vec{H} \approx \sqrt{\frac{\epsilon}{\mu}} \hat{\rho} \times \vec{E}. \tag{13}$$

Note that Eq. (13) also relates the electric to the magnetic field of a plane wave (see, e.g., Jackson [12, p. 299]), which is why geometrical optical rays are often said to carry local plane waves. To obtain Eq. (13) we make the approximation that $\vec{E}'$ is slowly varying with respect to the wavelength λ. This is the core approximation of geometrical optics, which is also needed to obtain the time-averaged Poynting's vector in a source-free homogeneous medium

$$\vec{S} \approx \sqrt{\frac{\epsilon}{4\mu}} \left|\vec{E}\right|^2 \hat{\rho}, \tag{14}$$

where $\left|\vec{E}\right|$ is the amplitude of the field. The Poynting's vector is the vector describing the energy transport. By integrating the Poynting's vector over a tube of rays, so that we have one surface (S_1) where all the rays enter the volume, one surface (S_2) where all the rays exit the volume and the other sides of the surfaces are along the paths followed by the rays (see Figure 1) we can derive the intensity law of geometrical

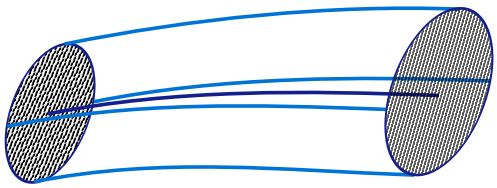

Figure 1: A ray tube.

optics

$$I_1 \, \mathrm{d}S_1 = I_2 \, \mathrm{d}S_2, \tag{15}$$

where the intensity is defined by

$$I \equiv |\hat{n} \cdot \vec{S}|, \tag{16}$$

with $\hat{n}$ the surface normal. The intensity of the field is therefore inversely proportional to the area spanned by the ray tube. By combining Eq. (16) with Eq. (14) and assuming the material to be non-magnetic, one can obtain

$$I = \hat{n} \cdot \hat{\rho} n \sqrt{\frac{\epsilon_0}{4\mu_0}} |\vec{E}|^2. \tag{17}$$

The amplitude of the electric field is therefore

$$|\vec{E}| = \sqrt{\frac{2I\sqrt{\mu_0}}{\hat{n} \cdot \hat{\rho} n \sqrt{\epsilon_0}}} \tag{18}$$

and through its dependence on I, it is a function of the area spanned by the ray tube. In a ray tracer, these areas can for instance be calculated by tracing several very close rays (i.e., neighbouring rays). At optical interfaces, part of the field can be reflected, leading to a decrease of the amplitude by a factor given by the Fresnel equations (see, e.g., Born and Wolf [3, p. 42]). The attenuation is different for the component of the electric field in the plane spanned by the surface normal and the propagation direction of the field (the p-component) and for the component in the direction perpendicular to this plane

(the s-component). The amplitude transmission through the interface is given by

$$T'_s = \sqrt{\frac{\cos(\theta_2)}{\cos(\theta_1)}} \frac{2n_1\cos(\theta_1)}{n_1\cos(\theta_1)+n_2\cos(\theta_2)}, \tag{19}$$

$$T'_p = \sqrt{\frac{\cos(\theta_2)}{\cos(\theta_1)}} \frac{2n_1\cos(\theta_1)}{n_2\cos(\theta_1)+n_1\cos(\theta_2)}. \tag{20}$$

These equations are slightly different from the familiar versions of the Fresnel equations to account for the fact that the change in surface of a ray tube due to the refraction is already affecting the amplitude of the field through the intensity law of geometrical optics.

In summary, to perform geometrical optical simulations using ray tracing, one starts by specifying the ray distribution at the initial surface and the starting power, phase, and polarization of the rays. The rays are then transported through the optical system, where their paths are determined by Snell's law, Eq. (10), their phases by the optical path lengths, and their amplitudes by the (adjusted) Fresnel equations, Eqs. (19) and (20), and the intensity law of geometrical optics, Eq. (15). This is an accurate way to propagate the complex, vectorial electromagnetic field, provided that the amplitude of the field does not vary significantly over the length scale of a wavelength.

Returning to the fields of a magnetic dipole, Eqs. (6, 7), we find that the their phases are determined by the phase of the Green's function ($\exp(\mathrm{i}k\rho)$). The surfaces of constant phase are therefore spheres around the point $\vec{r_0}$. The phase changes fastest along the gradient of the phase ϕ,

$$\nabla\phi = k\hat{\rho}, \tag{21}$$

which defines vectors perpendicular to the surfaces of constant phase and with a magnitude equal to the wave number. This is consistent with the concept of a ray in geometrical optics. The second useful property is that both $\vec{E}^{(md)}$ and $\vec{H}^{(md)}$ are perpendicular to $\hat{\rho}$, the direction of phase propagation. Furthermore, the local electric and magnetic field are related through Eq. (13). We thus conclude that the fields of a magnetic dipole, Eqs. (6, 7), and thereby those of the kernel of the m-theory, Eqs. (3, 4), can be calculated using a geometrical optics propagator $\tilde{G}$. Using this propagator, the electric field of Eq. (6) can be written as

$$\vec{E}^{(md)}(\vec{r_0},\vec{r_1},\vec{m}(\vec{r_0})) = \tilde{G}\left[\frac{-k^2}{4\pi}\sqrt{\frac{\mu}{\epsilon}}\vec{m}(\vec{r_0}),\vec{r_0},\vec{r_1}\right]. \tag{22}$$

In free-space this propagator has the simple form

$$\tilde{G}\left[\vec{A},\vec{r_0},\vec{r_1}\right] = \hat{\rho}\times\vec{A}\frac{\exp(\mathrm{i}k\rho)}{\rho}, \tag{23}$$

which reduces Eq. (22) to Eq. (6). The advantage of propagating the magnetic dipole field by a geometrical optical operator is that the propagation is no longer restricted to a single homogeneous medium. Where Eq. (6) was limited to free-space propagation, Eq. (22) can propagate the field through a series of optical components (e.g., lenses).

Since we are not so much interested in propagating dipole fields as we are in the m-theory diffraction integral, we use Eq. (22) to write the diffraction integrals of Eqs. (1, 2) as

$$\vec{E}(\vec{r_1}) = \iint_{S_0}\tilde{G}\left[\frac{-\mathrm{i}k}{2\pi}\hat{n}_0\times\vec{E}(\vec{r_0}),\vec{r_0},\vec{r_1}\right]\mathrm{d}S_0, \tag{24}$$

$$\vec{H}(\vec{r_1}) = \iint_{S_0}\sqrt{\frac{\epsilon}{\mu}}\hat{\rho}\times\tilde{G}\left[\frac{-\mathrm{i}k}{2\pi}\hat{n}_0\times\vec{E}(\vec{r_0}),\vec{r_0},\vec{r_1}\right]\mathrm{d}S_0. \tag{25}$$

Again, the advantage of using Eq. (24) over the integral of Eq. (1) is the ability to propagate the field through an optical system with multiple lenses. Since Eq. (24) is based on the free-space integral of Eq. (1), the implicit assumption is that the field in the medium after the plane of integration can be calculated as if the medium were to continue to infinity. At optical interfaces (i.e., lenses) we neglect the part of the field that is reflected and only consider the forward propagating field (with an amplitude adjusted by the Fresnel equations). Although the Fresnel equations provide the reflected fields and these could in principle be propagated using our method, their influence is typically small whereas the associated computational cost would be substantial. We therefore only consider the direct path, neglecting (multiple) reflections. Under this minor assumption, the geometrical propagation of the integral kernel allows us to propagate a polarized field at the input surface S_0, through a (sub)system.

2.3 Cascaded diffraction integral

By combining the diffraction integral with a geometrical optics propagator we can calculate the field after a set of lenses and account for the diffraction caused by the initial field (e.g., if the initial field is a top hat or a diffracting laser beam). One can accurately describe the propagation through a series of elements, as long as the amplitude of the field is slowly varying. At for instance apertures (or lenses that are small

compared to the beam of light) this condition is violated, requiring another approach. In such a plane one needs to perform another decomposition into secondary sources and propagate them using geometrical optics. Labelling the initial plane by S_0 and the second diffracting plane by S_1, we can summarize this mathematically by

$$\vec{E}(\vec{r_2}) = \iint_{S_1}\iint_{S_0} \tilde{G}\left[\frac{-\mathrm{i}k}{2\pi}\hat{n}_1 \times \tilde{G}\left[\frac{-\mathrm{i}k}{2\pi}\hat{n}_0 \times \vec{E}(\vec{r_0}), \vec{r_0}, \vec{r_1}\right], \vec{r_1}, \vec{r_2}\right] \mathrm{d}S_0 \mathrm{d}S_1. \quad (26)$$

Any further planes where the geometrical optical propagator breaks down, can be treated in a similar manner. The magnetic field can be obtained by changing the final propagator to the one defined by Eq. (25)

$$\vec{H}(\vec{r_2}) = \iint_{S_1}\iint_{S_0} \sqrt{\frac{\epsilon_2}{\mu}}\hat{\rho} \times \tilde{G}\left[\frac{-\mathrm{i}k}{2\pi}\hat{n}_1 \times \tilde{G}\left[\frac{-\mathrm{i}k}{2\pi}\hat{n}_0 \times \vec{E}(\vec{r_0}), \vec{r_0}, \vec{r_1}\right], \vec{r_1}, \vec{r_2}\right] \mathrm{d}S_0 \mathrm{d}S_1. \quad (27)$$

The result is a cascaded diffraction integral that fully accounts for the diffraction introduced at multiple surfaces.

2.4 Monte Carlo ray tracing propagator

We have seen that a diffraction simulation can be stated as a decomposition of the field in magnetic dipoles followed by a geometrical optical propagation of the field of every magnetic dipole, and the superposition of the resulting fields. Since HFPI is implemented in a Monte Carlo ray tracer, this section will compare the Monte Carlo ray tracing propagator to the standard propagator of geometrical optics. We will see that the main difference lies in the method used to account for the conservation of energy.

A Monte Carlo ray tracer initiates random rays from a predefined source set and traces each ray independently through the optical system. Upon passing through a pixel, the ray weight is added to the pixel value.

Mathematically, a Monte Carlo ray tracer relies on the central limit theorem. This theorem states that the averaged sum of the function values $w(\vec{x}_i)$ evaluated at (two-dimensional) positions $\vec{x}_i$ sampled from a random variable $\boldsymbol{X}$ with probability distribution function $p(\vec{x})$,

$$w_N = \frac{1}{N}\sum_{i=1}^{N} w(\vec{x}_i), \quad (28)$$

converges in the limit of $N \to \infty$ to the analytical expectation value

$$\langle w(\boldsymbol{X})\rangle = \iint w(\vec{x})p(\vec{x})d^2\vec{x}. \quad (29)$$

For Monte Carlo ray tracing, the sample points $\vec{x}_i$ are the intersection points of rays and a final plane. The probability of sampling a certain ray is proportional to the local density of rays

$$p(\vec{x}) = \lim_{\Delta S\to 0,\ N\to\infty} \frac{\Delta N(\vec{x}, N)}{\Delta S(\vec{x})N}, \quad (30)$$

with ΔN the number of rays, ΔS a surface area, and N the total number of rays.

By choosing

$$w(\vec{x}) = \frac{I(\vec{x})}{p(\vec{x})}, \quad (31)$$

and limiting the domain of integration of Eq. (29) to the area of a pixel S_p, Eq. (29) evaluates the power incident on that pixel. Since generally the intensity distribution in the pixel plane $I(\vec{x})$ is not known, one can use the ray tracer to calculate it from the intensity distribution in another plane $I_0(\vec{x_0})$. Rewriting Eq. (15)

$$I(\vec{x}) = I_0(\vec{x_0})\frac{\Delta S_0(\vec{x_0})}{\Delta S(\vec{x})}, \quad (32)$$

where ΔS_0 and ΔS are infinitesimally small surface areas connected by a tube of rays and the mapping between $\vec{x}$ and $\vec{x_0}$ is defined by the same ray tube. Combining Eqs. (30), (31), and (32), we obtain

$$w(\vec{x}) = \lim_{\Delta S_0\to 0,\ N\to\infty} I_0(\vec{x_0})\frac{\Delta S_0(\vec{x_0})N}{\Delta N_0(\vec{x_0}, N)} = \frac{I_0(\vec{x_0})}{p_0(\vec{x_0})}, \quad (33)$$

where we have used that $\Delta N_0(\vec{x_0}, N) = \Delta N(\vec{x}, N)$ if $\vec{x}$ and $\vec{x_0}$ are connected by a ray (tube). The expectation value of a Monte Carlo ray tracer will thus be the power incident on a plane, if each ray carries a power defined by Eq. (33).

Since for HFPI we are interested in calculating the amplitude of the field, the weight of the rays must be adjusted accordingly. From Eq. (18) we see that by choosing

$$w'(\vec{x}) = \lim_{\Delta S\to 0,\ N\to\infty} \sqrt{\frac{2I_0\Delta S_0 N}{\Delta N\ \hat{n}\cdot\hat{\rho}n}}\left(\frac{\mu_0}{\epsilon_0}\right)^{1/4}\sqrt{\frac{\Delta SN}{\Delta N}}, \quad (34)$$

the expectation value of $w'(\vec{X})$ is

$$\begin{aligned}\left\langle w'(\vec{X})\right\rangle &= \iint_{S_p} \lim_{\Delta S_0 \to 0,\ N\to\infty} \sqrt{\frac{2I_0 \Delta S_0 N}{\Delta N\ \hat{n}\cdot\hat{\rho}n}} \\ &\quad \left(\frac{\mu_0}{\epsilon_0}\right)^{1/4} \sqrt{\frac{\Delta N}{\Delta S N}} \mathrm{d}^2\vec{x} \\ &= \iint_{S_p} \sqrt{\frac{2I(\vec{x})}{\hat{n}\cdot\hat{\rho}n}} \left(\frac{\mu_0}{\epsilon_0}\right)^{1/4} \mathrm{d}^2\vec{x} \\ &= \iint_{S_p} |\vec{E}(\vec{x})| \mathrm{d}^2\vec{x}. \end{aligned} \tag{35}$$

In a simulation, different factors of the ray weight defined by Eq. (34) are implemented at different parts of the simulation. At the starting surface, every ray is assigned a weight proportional to the square root of the intensity divided by the local ray density

$$w_s = \sqrt{\frac{2I_0 \Delta S_0 N}{\Delta N_0}} \left(\frac{\mu_0}{\epsilon_0}\right)^{1/4}. \tag{36}$$

For an HFPI simulation one has access to the electric field instead of the intensity, so one should rewrite this using Eq. (17) as

$$w_s = \sqrt{\frac{\Delta S_0 N}{\Delta N_0}} \sqrt{n_0} |\vec{E}_0|, \tag{37}$$

with n_0 the refractive index at S_0. At the final surface, upon incidence on the pixel, the weight is adjusted using the cosine of the angle between the ray direction and the surface normal, $\hat{n}\cdot\hat{\rho}$, and the local refractive index n

$$w_i = \sqrt{\frac{1}{\hat{n}\cdot\hat{\rho}n}}. \tag{38}$$

Furthermore, the reciprocal of the local ray density is needed

$$w_p = \sqrt{\frac{\Delta S N}{\Delta N}}. \tag{39}$$

As stated in Section 2.2, this can be obtained by tracing neighbouring rays. So far we have assumed that the power carried by a ray remains constant, but the effect of transmission losses can easily be incorporated into an additional weight w_f, depending on the adjusted Fresnel equations, Eqs. (19, 20).

Since we are not only interested in the amplitude, but also in the phase of the electric field, we must also include the phase function $\exp[\mathrm{i}k_0 s(x,y)]$ into the ray weight w'. If the local wavefront and the pixel are not parallel, the phase will vary quickly over the pixel. We are not interested in the complex integral over the pixel, but in approximating the field at the centre of the pixel. Therefore we will approximate the field of the ray by a local plane wave and calculate its

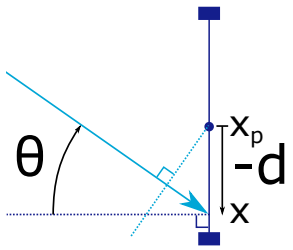

Figure 2: Local plane wave approximation of a ray at a pixel.

contribution to the field in the centre of the pixel, (x_p, y_p), using the adjusted phase function

$$\exp\left[\mathrm{i}\phi_p\right] \approx \exp\left[\mathrm{i}k_0\left(s \pm dn\sin(\theta)\right]\right. \tag{40}$$

where θ is the angle between the incident ray and the local surface normal (see Figure 2) and

$$d = \sqrt{(x-x_p)^2 + (y-y_p)^2} \tag{41}$$

is the distance between the centre of the pixel and the intersection point of the ray with the pixel, and the sign depends on the position of (x,y) with respect to (x_p, y_p). Using the final ray weight

$$w' = w_s w_i w_p w_f \exp\left[\mathrm{i}\phi_p\right], \tag{42}$$

we can approximate the complex electric field at the centre of a pixel by averaging the electric fields of all rays properly weighted, which arrive at the pixel and dividing by the area of the pixel.

Like the geometrical optical propagator, this adjusted Monte Carlo ray tracing propagator is thus able to propagate the complex electric field and thereby evaluate the cascaded diffraction integral of Eq. (26).

3 Implementation in a Monte Carlo ray tracer

In Section 2 we derived a method to compute a cascaded diffraction integral that simulates the effects of diffraction

and aberrations. In this section we will present the individual steps of a vectorial HFPI simulation with a Monte Carlo ray tracer.

As a pre-step to any simulation one must define the optical system and the initial electric field. Furthermore, one should define at which planes diffraction should be taken into account (e.g., at an aperture), and at which planes one wants to calculate the electric field.

The first step of the actual simulation consists of generating random rays at the initial surface. If the input field can be propagated using geometrical optics, the set of possible initial rays is defined by the type of field. For a plane wave for instance, the Monte Carlo algorithm chooses random starting positions for the rays, but the initial propagation and polarisation directions are determined by the corresponding directions of the plane wave. If the input field does not propagate according to the laws of geometrical optics (e.g., a Gaussian beam in its waist) the diffraction integral of Eq. (24) should be applied. The initiated rays have random starting positions and directions, and carry a field according to

$$\vec{E}_{ray} = \sqrt{\Delta S_0 N n_0}\frac{\mathrm{i}k}{2\pi}\hat{\rho} \times \left(\hat{n}_0 \times \vec{E}(\vec{r_0})\right) \qquad (43)$$

where $\vec{E}(\vec{r_0})$ is the field in the starting point of the ray, $\hat{n}_0$ the normal of the starting plane and $\hat{\rho}$ the normalized ray direction.

After initiating the rays, they are propagated through the system using standard ray tracing procedures. At optical interfaces, the transmission coefficients are determined using the adapted Fresnel equations for ray tracing, Eqs. (19) and (20), and the weight of the rays are adapted accordingly.

At a diffracting surface, a ray is terminated and its complex contribution to the local electric field is determined by adjusting the ray weight with the factors w_i (Eq. (38)) and w_p (Eq. (39)). Since w_p is an amplitude factor related to the intensity law of geometrical optics, one needs a procedure like tracing neighbouring rays to determine w_p. This is not implemented in our Monte Carlo ray tracer, so we have to neglect this term, leading to an error in the amplitude of the field. In Section 4 the effect of this error is quantified for a microscope with an NA of 0.7. After determining the local E-field, the diffraction integral of Eq. (24) is applied. Due to the linearity of optics, one does not first have to calculate the total electric field, but can directly apply the integral on the local field contribution of a single ray. In practice we do this by changing the direction of the ray randomly, but one could achieve faster convergence by initiating a set of new rays with random directions at the point of intersection of the ray and the diffracting surface. The weight of the new ray is adjusted according to Eq. (43), where $\vec{E}(\vec{r_0})$ is the previously calculated contribution of the terminated ray to the local electric field.

These secondary rays are propagated further through the optical system. Any additional diffracting surface is treated identical to the first one, terminating the current rays and generating new rays instead. Finally, at the last (or any intermediate) surface, the field is determined by having every ray contribute to the pixel it intersects. The weight of each ray is again adjusted using the factors w_i (Eq. (38)) and w_p (Eq. (39)). Furthermore, since we are now approximating the field at the centre of the pixel, the phase of each ray is shifted using the local plane wave approximation of Eq. (40).

Following this procedure, a Monte Carlo ray tracer can propagate an electric field through an optical system and include the aberrations introduced by the system and the diffraction introduced at specific planes of the system.

4 Results

We compare results obtained using vectorial HFPI with those of two other simulation methods. The first method is the vectorial ray-based diffraction integral (VRBDI) of Andreas *et al.* [2] for which the authors have published their MATLAB code online [1]. The second reference method is beam synthesis propagation (BSP) in the commercial software package CODE V [5]. The first simulated system is a low NA configuration, with diffraction of the input field, whereas the second and third system have higher NAs and demonstrate the effects of single and multiple diffraction respectively.

The first system is one of the example systems of the VRBDI MATLAB code [1] and consists of a singlet with in its front focal plane a ring aperture illuminated by an x-polarised plane wave with a wavelength of 632.8 nm. A schematic representation of the system is shown in Figure 3 and its parameters are listed in Table 1. This system can be modelled using HFPI by placing secondary sources in the ring aperture. The results of HFPI and the absolute difference between the HFPI results and those using the MATLAB code by Birk Andreas [1] are shown in Figure 4. Note that these results and the other results presented in this section are normalized with respect to the squared amplitude. The statistical nature of HFPI can be clearly seen in the difference

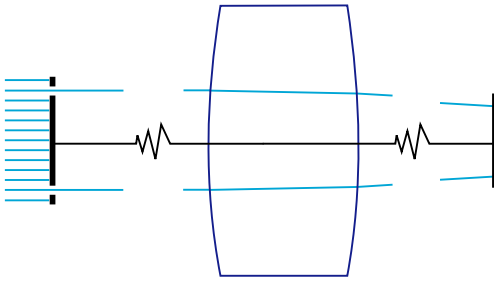

Figure 3: A schematic representation of the ring aperture and singlet. The parameters are listed in Table 1, the results of an HFPI simulation of the system are shown in Figure 4.

plots. The absolute difference between the methods is around 2% of the maximum amplitude for all three electrical components, even though the y-component of the field is around six orders of magnitude smaller than the x-component. The L_2-difference between the two methods is defined as

$$\text{L}_2\text{-dif} = \frac{\sqrt{\sum_{p=1}^{N_p} \sum_{i=x,y,z} \left| E_{i,p}^{(2)} - E_{i,p}^{(1)} \right|^2}}{\sqrt{\sum_{p=1}^{N_p} \sum_{i=x,y,z} \left| E_{i,p}^{(1)} \right|^2}} \tag{44}$$

where the superscript (1) indicates the results of the reference method, the superscript (2) those of HFPI, the subscript p runs over all N_p pixels, and the subscript i runs over the three Cartesian components of the field. An overall phase difference (i.e., a piston phase term) between the two methods is corrected before calculating Eq. (44). For this example, the L_2-difference between the results of HFPI and those of the MATLAB code by Birk Andreas is 2.4 %. The simulation times and the number of rays drawn for this simulation and the other simulations presented in this section are listed in Table 2.

In the previous example almost all the energy was concentrated in the main polarisation direction because the system had a very low effective numerical aperture of 0.004. We now consider the results for a microscope with an NA above 0.7 for which the effects of polarisation become more pronounced. The system corresponds roughly to the example microscope system of CODE V which is based on a patent by Tojyo [25]. It is depicted in Figure 5 and its parameters are listed in Table 3. The system is in reversed order compared to the original design and the normal usage of microscopes. Such a reversed configuration can be beneficial in the design process [24, p. 257–258] and, more important for the current paper, makes the polarisation effects clearly visible. As explained in Sections 2 and 3, an error is introduced by propagating the amplitude of the field using a Monte Carlo ray tracer designed for propagating intensities. This means that the factor w_p in Eq. (39) is assumed to be 1. For the first system w_p does not vary significantly for different rays and this error has little effect. For the current system, the neglected term does vary noticeably and as a result one can see some systematic differences between the results of HFPI and BSP in Figure 6. To verify that these difference are indeed mainly due to neglecting w_p, which is a function of the ray density, we adjust the system such that we can compensate for this error. This is done by letting the diffraction occur in the exit pupil, which is the image of the aperture in image space, instead of at the aperture itself. The two

Table 1: System parameters of the singlet with a ring aperture [1] (see Figure 3).

Quantity	**Value**
Ring aperture radius	1.25 mm
Ring aperture width	0.01 mm
Distance aperture to lens (front focal length)	300 mm
Thickness lens	3 mm
Refractive index lens	1.5155
Radius of curvature first surface	308.5 mm
Radius of curvature second surface	-308.5 mm
Distance lens to output plane	100 mm
Wavelength	632.8 nm
Numerical aperture	0.004

Table 2: The number of paths and simulation times for the HFPI simulations. The simulations were performed on multiple (6-8) computers using a total of 36 to 48 cores simultaneously. Every path was traced from object to image plane.

System name	**Number of paths**	**Core hours**
ring aperture	10^9	12
microscope	10^9	23
microscope corrected HFPI	$1.36 \cdot 10^9$	45
cascaded diffraction	10^{10}	306

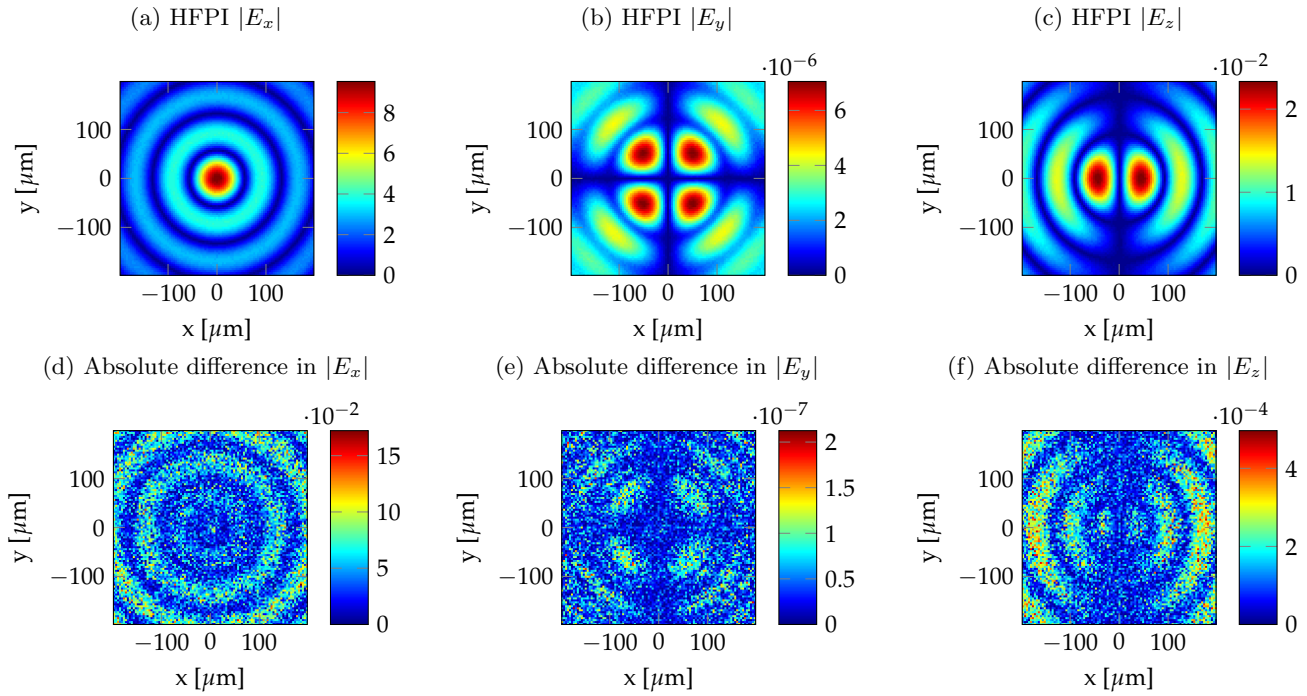

Figure 4: The results from HFPI for a plane wave illuminating a ring aperture, followed by a singlet (see Figure 3 and Table 1 for the system description). The amplitudes of the electric field according to the HFPI simulation are shown in the top row, and the difference between these results and those of the model of Andreas *et al.* in the bottom row. The L_2-difference (Eq. (44)) between the methods is 2.4 %.

w_p factors of a ray (one for the propagation to the diffracting surface and one for the following propagation step to the image plane) are now both functions of the position of the ray in the exit pupil only. In such cases, our implementation enables us to account for w_p, by adjusting the amplitudes of the rays that pass through the exit pupil using a thin element with a spatially varying absorption. The bottom row of Figure 6 shows that the differences between the BSP simulation and the HFPI simulation with correct w_p are significantly lower than for the HFPI simulation that assumes w_p to be 1. This may appear to suggest that assuming w_p to be constant is insufficient for simulating high NA systems but, as the cross plots in Figure 7 demonstrate, the errors due to this approximation are mainly in the secondary polarisation directions and might be acceptable for a wide range of applications. Furthermore, as mentioned before, the method can be adapted to account for w_p by tracing closely neighbouring rays and tracking the area spanned between these rays.

In the last example, diffraction is introduced in multiple planes. Moreover, the field in the second diffracting plane is affected by the diffraction in the first plane, i.e., cascaded diffraction occurs. The system consists of a pinhole placed in the focal plane of a singlet (see Table 4 and Figure 8) followed by the previously described reversed microscope. The light passing through the singlet generates a PSF in the plane of the pinhole. There, the semi-circular pinhole with a radius of 40 μm, completely blocks the lower half of the PSF. The pinhole is then imaged by the reversed microscope of the previous example (see Table 3). Both the object NA of the microscope and the image NA of the singlet are 0.018. The system has three diffracting planes: the aperture of the singlet, the semi-circular pinhole, and the physical aperture of the microscope. If we apply the cascaded diffraction integral presented in Section 2.3 to this problem, we need to perform

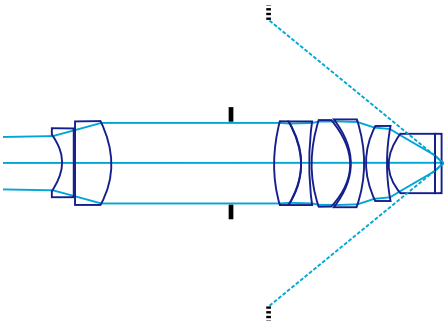

Figure 5: A schematic representation of the microscope. The exit pupil and the rays at the exit pupil are depicted using dashed lines.

three consecutive integrals. This means that the fields are first decomposed into secondary sources at the aperture of the singlet, then decomposed again at the pinhole, and for a finally time at the aperture of the microscope. This is implemented in the Monte-Carlo ray tracer by randomly changing the direction of every ray, and adjusting its field using Eq. (43), in each of the before mentioned diffracting planes. The results of HFPI and the difference with a simulation with BSP are depicted in Figure 9. Simulating the triple diffraction is computationally more demanding than the previous simulations which is why one can observe stronger statistical noise (due to the Monte Carlo nature of the method). This noise can be reduced at the cost of a longer simulation time by sampling more rays. The cascaded diffraction simulation shows similar differences between our implementation of HFPI and BSP as the simulation of the microscope did (Figure 6), which suggests that the remaining differences are at least partly due to taking w_p constant. A simulation method that cannot describe cascaded diffraction would produces results very different from the ones shown in Figure 9. For the exit pupil diffraction method for instance, the current system is nearly identical to the microscope of Figure 5 since the geometrical optical rays are only blocked by a single aperture. It would therefore produce the results shown in Figure 6 for both the microscope and the current system. A comparison between Figure 9 and 6 (which have an L_2-difference in absolute amplitude of 50%) thus shows the effect of (neglecting) cascaded diffraction.

Table 3: System parameters of the microscope. The aperture is indicated by $*$ and has a radius of 4.1 mm. The simulations were performed at a wavelength of 546.1 nm.

Surface number (j)	**Radius [mm]**	**Distance [mm] (j to j+1)**	**Refractive index (after j)**
0	∞	156.466582	1
1	-4.893054	1.235052	1.6164
2	331.759004	3.773815	1.7546
3	-7.999296	12.115041	1
4*	∞	0.000000	1
5	∞	4.439448	1
6	17.811462	2.602460	1.4984
7	-6.811396	1.000946	1.7918
8	65.868865	0.140215	1
9	11.474481	3.940087	1.4984
10	-6.105357	1.470398	1.5236
11	-14.404937	0.100095	1
12	7.120780	2.202081	1.4984
13	14.369366	0.100095	1
14	4.017020	4.660603	1.7919
15	2.431637	0.672514	1
16	∞	0.170000	1.5246
17	∞		

5 Discussion

By keeping track of the phase of optical paths and initiating secondary sources at diffracting surfaces (e.g., apertures) both diffraction and aberrations can be simulated using a Monte Carlo ray tracer. The method presented here is an extension to vectorial optics of our previously published work [17]. The scalar version of HFPI can be obtained from the method described here by using scalar ray-tracing and initiating secondary sources according to the kernel of the Rayleigh-Sommerfeld diffraction integral of the first kind. As a result the secondary sources in scalar HFPI are point sources (with a cosine-attenuation) instead of magnetic dipoles. A similar model for diffraction at an aperture in free-space was recently presented by Mahan *et al.* [16]. The excellent agreement between their Monte-Carlo model and experimental results demonstrates the potential of Monte-Carlo ray-based diffraction methods.

Huygens-Fresnel path integration is a Monte Carlo method and therefore does not require an explicit sampling grid.

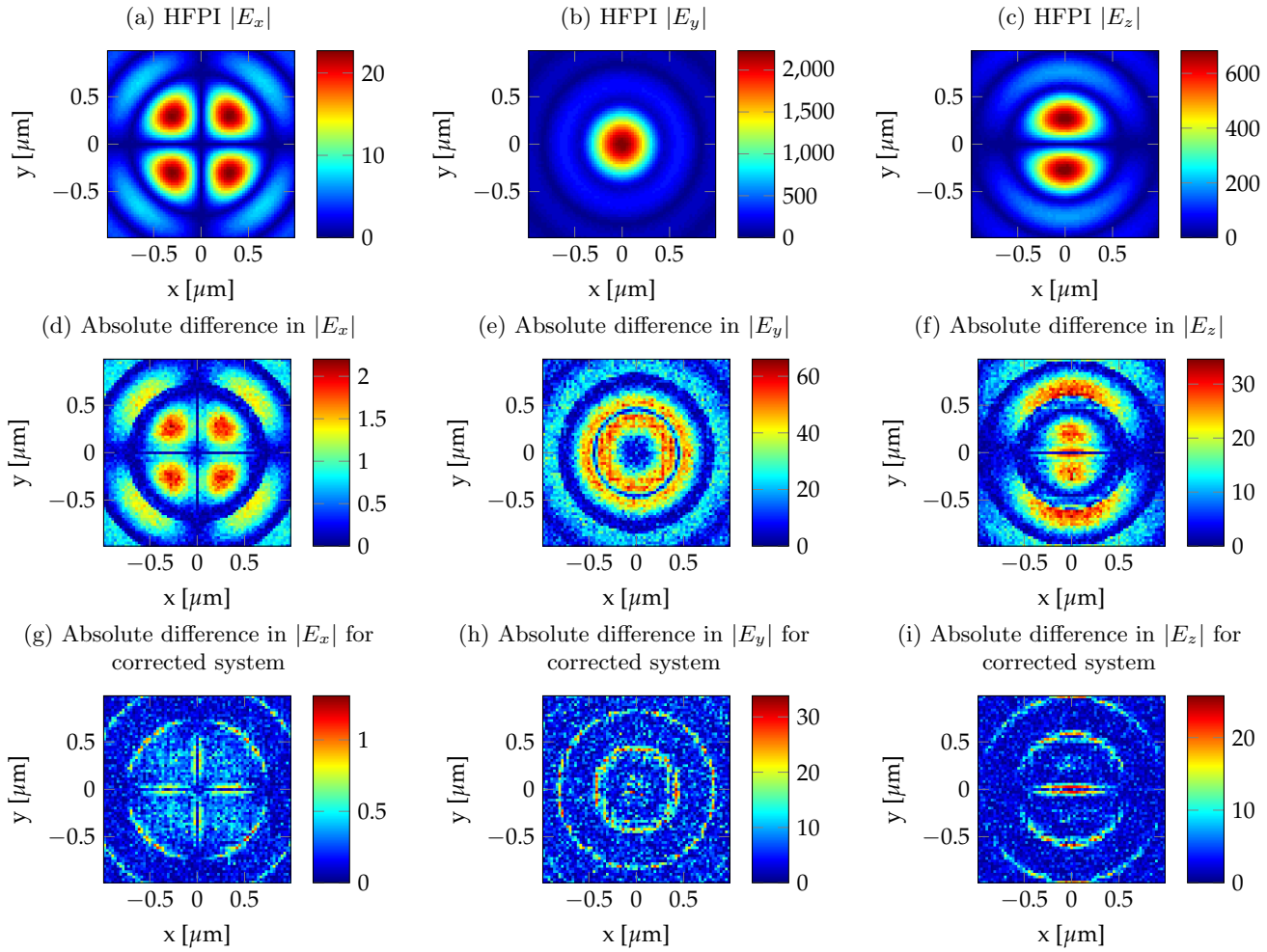

Figure 6: A magnetic dipole (see Eq. (6)) imaged by the microscope with the parameters listed in Table 3. The top row shows the amplitudes of the three components of the electric field in the focal plane, as predicted by HFPI. The middle row shows the absolute differences between these results and those of a BSP simulation. For the bottom row the amplitude factor w_p, which is usually neglected in our implementation of HFPI, was accounted for. The plots show the differences between the HFPI and BSP simulation. The L_2-differences between HFPI and BSP are 7.3% and 3.3% for respectively the uncorrected and the corrected HFPI.

Since every ray that arrives in an aperture generates a new source, one does not need to explicitly calculate the field in an aperture. As a result, HFPI can calculate the field in the focal plane of a system, even when the field in the aperture cannot be resolved because the curvature of the local wave front requires an unrealistically dense sampling. This is a considerable advantage over methods that rely on intermediate resampling.

The Monte Carlo implementation without explicit resampling has several other benefits. First, it aids the user by reducing the input required for the simulation. Apart from a system and source definition, the user only needs to de-

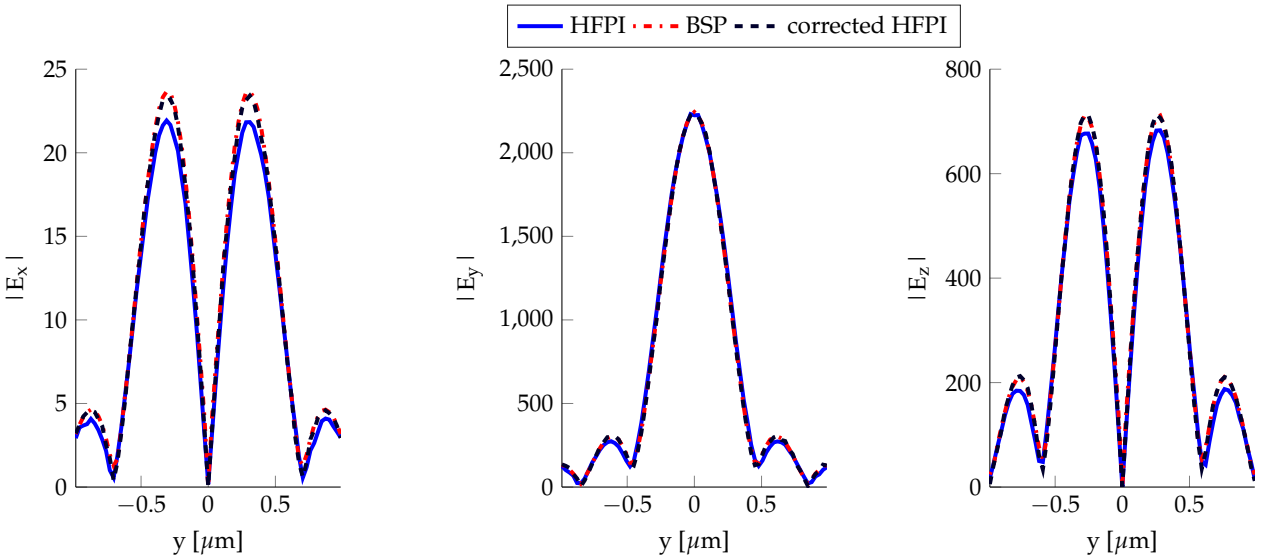

Figure 7: Cross sections of the amplitude of the field components at the focal plane of the microscope. In order to include the peak amplitudes, the cross section of the x-component is at $x = 2.5\mu m$ and the other components are shown at $x = 0$. The plots show the results from the BSP simulation as well as the results from the standard HFPI simulation, which neglects an amplitude factor w_p (Eq. (39)), and the results from the HFPI simulation that accounted for w_p (corrected HFPI). It can be seen that, in comparison to the other simulations, the peak amplitudes of HFPI are a bit lower for the x- and z-polarisation and the PSF is slightly wider.

fine the diffracting surfaces where new sources are generated. The second benefit of the Monte Carlo implementation is that it lends itself to parallel computing since every ray can be propagated completely independent of all other rays and only little memory is required. For the results shown here, the simulations were performed on six to eight computers using a total of 36 to 48 cores simultaneously. The perfect scalability of HFPI is particularly relevant in view of increasing availability of cloud computing services, which provide large number of cores at very low effort and cost and could significantly reduce real world computation times.

The simulation times also depend strongly on the numerical implementation. We expect that the computational load can be reduced by allowing every ray to initiate multiple (as opposed to a single) new rays in diffracting surfaces (i.e., by ray splitting), and by optimizing the code for the use of graphics processing units (GPUs). Other methods to increase the convergence are to trace a grid of rays and interpolate the field between them or to sum multiple subfields at an aperture before initiating the secondary sources. These last two methods may increase the convergence, but come at the cost of increased complexity in the implementation. This would compromise one of the main advantages of HFPI, the relative ease with which it can be implemented in an existing polarization ray tracer.

Using distributed or parallel computing is desired since the computational demand of HFPI is high (see Table 2). By running on multiple cores simultaneously, all the presented simulations were performed in under an hour, with the exception of the cascaded diffraction system. The simulation times can be further reduced by calculating the results on a coarser grid. A too large pixel size would likely reduce the accuracy of the results, as was reported for a related method [16], but this issue was not observed for the systems presented in this paper. Another effective way to decrease simulation times is by sampling fewer paths. In contrast to many other simulation methods, such a reduction of the sampling rate does not cause aliasing but only increases the noise. The error of HFPI is inversely proportional to the square root of the number of rays [17], so decreasing the number of rays (thereby decreas-

Table 4: System parameters for the first part of the system showing cascaded diffraction. The second part of the system is the microscope with the parameters listed in Table 3. The aperture is indicated by $*$ and has a radius of 4.1 mm. The simulation is performed at a wavelength of 546.1 nm.

Surface number (j)	**Radius [mm]**	**Distance [mm] (j to j+1)**	**Refractive index (after j)**
0	∞	0.000000	1
1	244.210307	5.000000	1.5187
2	-244.210307	0.500000	1
3*	∞	0.000000	1
4	∞	234.000000	1
5	∞	0.000000	1

ing computational load) by two or three orders of magnitude increases the L_2-error by roughly a factor 10 or 32. As an example we present such results for the ring aperture (Table 1) in Figure 10. The plots are noisy but may suffice for many practical applications.

6 Conclusion

We present the theory for a ray-based vectorial diffraction simulation method based on a physical interpretation of the m-theory diffraction integrals and compare its results to those of other methods. Huygens-Fresnel path integration is a robust method to simulate the effects of aberrations, (multiple) diffraction, and polarization.

In the derivation of HFPI several assumptions are made. The distances between optical interfaces are assumed to be much larger than the wavelength, light reflected at optical interfaces is not further propagated and all apertures are treated using the Kirchhoff boundary conditions, which means that the light outside of the aperture is assumed to be completely blocked. Note that most of the methods listed in Section 1 operate under the same assumptions, and many of those methods need additional assumptions. The exit pupil diffraction integral for instance only allows a single diffracting surface and the Fresnel integral assumes all lenses to be thin lenses. The assumptions of HFPI allow for the simulation of many practical systems, such as head-mounted displays, that are excluded by these additional assumptions.

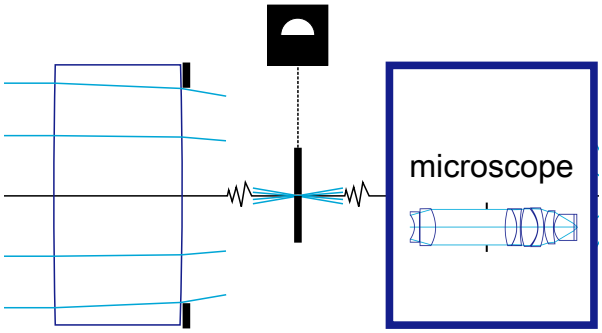

Figure 8: A schematic representation of the system with cascaded diffraction. The semi-circular aperture (shown at the top) is placed in the focal plane of the singlet and followed by the microscope (represented by the blue box) which is depicted in Figure 5. The parameters of the singlet and the microscope are listed in Tables 4 and 3.

HFPI can be used to simulate the propagation of coherent fields (e.g., laser beams) through optical systems for which aberrations and polarization effects cannot be neglected and where diffraction is introduced at one or more surfaces. The method is implemented by making some small adjustments to an existing Monte Carlo ray tracing engine, thereby hugely extending its domain of application with relatively little programming effort. Its main disadvantage is its computational demand, but this drawback is strongly reduced by distributed computing. The absence of explicit resampling in HFPI not only makes the method relatively easy to implement but also makes it robust and fairly easy to use.

Funding information

European Union (EU); People Programme (Marie Curie Actions) of the Seventh Framework Programme (FP7/2007-2013/ N-608082).

Aknowledgements

We thank Hans-Joachim Frasch and George Firanescu for support in implementing the method and Steffen Siegler for discussions on geometrical optics and for pointing our attention to the m-theory diffraction integrals.

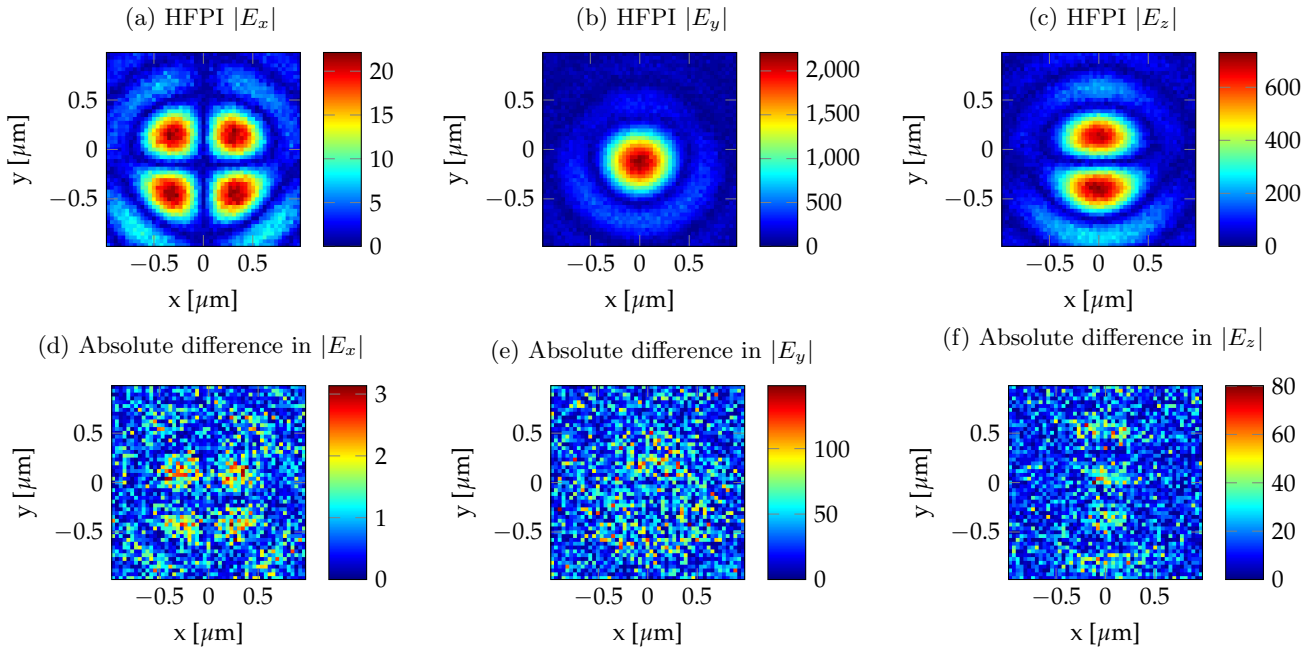

Figure 9: The amplitudes of the field components in the focal plane of the system demonstrating cascaded diffraction (top row) and the absolute differences with the results from the BSP simulations (bottom row). The L_2-difference between the methods is 13.2%.

References

[1] Birk Andreas. Vectorial ray-based diffraction integral (VRBDI), 2016. URL `http://nl.mathworks.com/matlabcentral/fileexchange/52210-vectorial-ray-based-diffraction-integral--vrbdi-`. version 1.7, MathWorks® File Exchange, accessed 6 May, 2017.

[2] Birk Andreas, Giovanni Mana, and Carlo Palmisano. Vectorial ray-based diffraction integral. *J. Opt. Soc. Am. A*, 32(8):1403–1424, 2015. doi: 10.1364/JOSAA.32.001403.

[3] Max Born and Emil Wolf. *Principles of optics*. Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, UK, seventh edition, 1999.

[4] Russell A. Chipman. Polarization ray tracing. *Proc. SPIE*, 0766:61–68, 1987. doi: 10.1117/12.940204.

[5] CODE V® is a trademark of Synopsys Inc. CODE V® beam synthesis propagation, 2017. URL `https://optics.synopsys.com/codev/pdfs/CV_BSP_datasheet.pdf`. Accessed June 10, 2018.

[6] CODE V® is a trademark of Synopsys Inc. CODE V features. https://optics.synopsys.com/codev/codev-features.html, 2017. Accessed July 29, 2017.

[7] David G. Fischer, Scott A. Prahl, and Donald D. Duncan. Monte Carlo modeling of spatial coherence: free-space diffraction. *J. Opt. Soc. Am. A*, 25(10):2571–2581, 2008. doi: 10.1364/JOSAA.25.002571.

[8] G. W. Forbes and Miguel A. Alonso. Using rays better. I. Theory for smoothly varying media. *J. Opt. Soc. Am. A*, 18(5):1132–1145, 2001. ISSN 1084-7529. doi: 10.1364/JOSAA.18.001132.

[9] Edward R. Freniere. Numerical experiments in modeling diffraction phenomena with Monte Carlo ray tracing. *Proc. SPIE*, 6289:62890N, 2006. doi: 10.1117/12.681797.

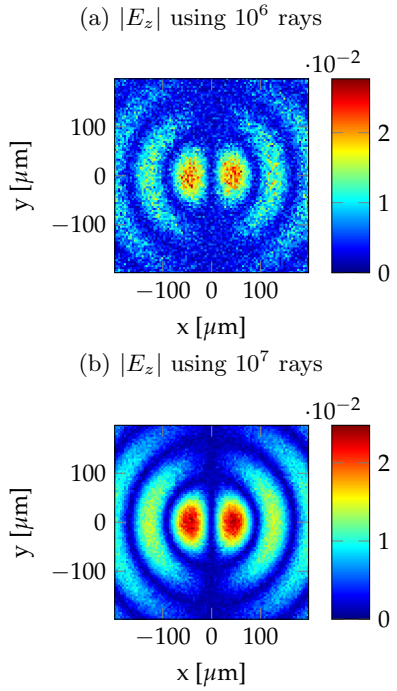

Figure 10: The amplitude of the z-component of the electric field in the output plane of the singlet with a ring aperture (see Table 1). The simulations are performed using two or three orders of magnitudes fewer rays than the simulations shown in Figure 4. Consequently the results are much noisier than the result shown in Figure 4, and the simulation times are reduced to roughly one and roughly six core minutes. This could offer a better compromise between accuracy and simulation time than the simulation for Figure 4.

[10] Edward R. Freniere, G. Groot Gregory, and Richard A. Hassler. Edge diffraction in Monte Carlo ray tracing. *Proc. SPIE*, 3780:151–157, 1999. ISSN 0277786X. doi: 10.1117/12.363773. URL `http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=997486`.

[11] Alan W. Greynolds. Propagation of generally astigmatic Gaussian beams along skew ray paths. *Proc. SPIE*, 0560: 33–51, 1986. doi: 10.1117/12.949614. URL `http://dx.doi.org/10.1117/12.949614`.

[12] John D. Jackson. *Classical electrodynamics*. John Wiley & Sons, Inc., New York, NY,, third edition, 1999.

[13] Graeme L. James. *Geometrical theory of diffraction for electromagnetic waves*. Institution of Engineering and Technology, UK, third edition, 1986.

[14] B. Karczewski and E. Wolf. Comparison of three theories of electromagnetic diffraction at an aperture.∗ part I: Coherence matrices. *J. Opt. Soc. Am.*, 56(9):1207–1214, Sep 1966. doi: 10.1364/JOSA.56.001207. URL `http://www.osapublishing.org/abstract.cfm?URI=josa-56-9-1207`.

[15] Joseph B. Keller. Geometrical theory of diffraction. *J. Opt. Soc. Am.*, 52:116–130, 1962. doi: 10.1364/JOSA.52.000116.

[16] J. R. Mahan, N. Q. Vinh, V. X. Ho, and N. B. Munir. Monte carlo ray-trace diffraction based on the Huygens-Fresnel principle. *Appl. Optics*, 57(18):D56–D62, Jun 2018. doi: 10.1364/AO.57.000D56.

[17] Marco Mout, Michael Wick, Florian Bociort, Joerg Petschulat, and Paul Urbach. Simulating multiple diffraction in imaging systems using a path integration method. *Appl. Optics*, 55(14):3847–3853, May 2016. doi: 10.1364/AO.55.003847. URL `http://ao.osa.org/abstract.cfm?URI=ao-55-14-3847`.

[18] *Optics Reference*. OSLO® is a trademark of Lambda Research Corporation, www.lambdares.com/oslo, user's manual, June 2005.

[19] Scott A. Prahl, Donald D. Duncan, and David G. Fischer. Monte Carlo propagation of spatial coherence. *Proc. SPIE*, 7187:71870G, 2009. doi: 10.1117/12.809603.

[20] Scott A. Prahl, David G. Fischer, and Donald D. Duncan. Monte Carlo Green's function formalism for the propagation of partially coherent light. *J. Opt. Soc. Am. A*, 26(7):1533–1543, 2009. doi: 10.1364/JOSAA.26.001533.

[21] Scott A. Prahl, Donald D. Duncan, and David G. Fischer. Stochastic Huygens and partial coherence propagation through thin tissue. *Proc. SPIE*, 7573:75730D, January 2010.

[22] Vladimir Serikov and Shinji Kawamoto. Numerical experiments in Monte Carlo modeling of polarization, diffraction, and interference phenomena. *Proc. SPIE*, 4436:80–88, 2001. doi: 10.1117/12.451286.

[23] Hans Severin. Zur Theorie der Beugung elektromagnetischer Wellen. *Z. Phys.*, 129(4):426–439, 1951. doi: 10.1007/BF01379593.

[24] Warren J. Smith. *Modern lens design*. McGraw-Hill, New York, first edition, 1992.

[25] T. Tojyo. Microscope objective lens system. US4208099, 1980.

[26] Eugene Waluschka. Polarization ray trace. *Optical Engineering*, 28(2):86–89, 1989. doi: 10.1117/12.7976913.

[27] Frank Wyrowski and M. Kuhn. Introduction to field tracing. *J. Mod. Optic.*, 58(5-6):449–466, 2011. doi: 10.1080/09500340.2010.532237.

[28] *Optical Design Program, User's Manual*. ZEMAX® is a trademark of Radiant Zemax LLC, www.zemax.com, July 2011.