

Vectorial phase retrieval in super-resolution polarization microscopy

Special Collection: 2024 Future Luminary Collection

R. Gutiérrez-Cuevas  ; L. A. Alemán-Castañeda  ; I. Herrera  ; S. Brasselet  ; M. A. Alonso 



APL Photonics 9, 026106 (2024)

<https://doi.org/10.1063/5.0179906>



Articles You May Be Interested In

Vectorial metasurface holography

Appl. Phys. Rev. (February 2022)

Vectorial sculpturing of spatiotemporal wavepackets

APL Photonics (September 2022)

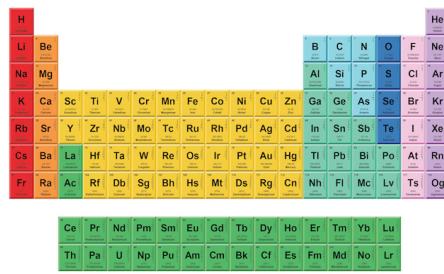
Demonstration of a vectorial optical field generator with adaptive close loop control

Rev. Sci. Instrum. (December 2017)



THE MATERIALS SCIENCE MANUFACTURER®

Now Invent.™



American Elements
Opens a World of Possibilities

...Now Invent!

www.americanelements.com

Vectorial phase retrieval in super-resolution polarization microscopy

Cite as: APL Photon. 9, 026106 (2024); doi: 10.1063/5.0179906

Submitted: 5 October 2023 • Accepted: 25 January 2024 •

Published Online: 27 February 2024



View Online



Export Citation



CrossMark

R. Gutiérrez-Cuevas,^{1,a)} L. A. Alemán-Castañeda,² I. Herrera,² S. Brasselet,² and M. A. Alonso^{2,3,b)}

AFFILIATIONS

¹ Institut Langevin, ESPCI Paris, CNRS, Université PSL, 75005 Paris, France

² Aix Marseille Univ, CNRS, Centrale Med, Institut Fresnel, Marseille 13013, France

³ The Institute of Optics, University of Rochester, Rochester, New York 14627, USA

^{a)} Author to whom correspondence should be addressed: rodrigo.gutierrez-cuevas@espci.fr

^{b)} Electronic mail: miguel.alonso@fresnel.fr

ABSTRACT

In single-molecule orientation localization microscopy, valuable information about the orientation and longitudinal position of each molecule is often encoded in the shape of the point spread function (PSF). Yet, this shape can be significantly affected by aberrations and other imperfections in the imaging system, leading to an erroneous estimation of the measured parameters. A basic solution is to model the aberrations as a scalar mask in the pupil plane that is characterized through phase retrieval algorithms. However, this approach is not suitable for cases involving polarization-dependent aberrations, introduced either through unintentional anisotropy in the elements or by using birefringent masks for PSF shaping. Here, this problem is addressed by introducing a fully vectorial model in which the polarization aberrations are represented via a spatially dependent Jones matrix, commonly used to describe polarization-dependent elements. It is then shown that these aberrations can be characterized by a set of PSF measurements at varying focal planes and for various polarization projections. This PZ-stack of PSFs, which contains diversity in both phase and polarization projection, is used in a phase retrieval algorithm based on nonlinear optimization to determine the aberrations. This methodology is demonstrated with numerical simulations and experimental measurements. The PYPSFSTACK software developed for modeling and characterization is made freely available.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0179906>

I. INTRODUCTION

Fluorescence microscopy is a widely used imaging modality in biological research¹ given its strong signal, selective labeling within complex systems,² and compatibility with super-resolution methods.³ Moreover, this technique also allows access to the sample's structural properties,^{4,5} making it very useful for studying biomechanics at the molecular level. For example, in single-molecule orientation localization microscopy (SMOLM), 3D spatial localization can reach a precision of a few nanometers while simultaneously allowing the characterization of the molecule's 3D orientational behavior (e.g., mean orientation and degree of wobble).⁵ Common SMOLM techniques include polarization channel splitting^{6–9} and point spread function (PSF) engineering,^{10–14} which can be used together or separately. The shape of the PSF can change considerably with the emitter's orientation and

longitudinal position, so it is crucial to take this into account to enable a full estimation of the parameters^{13,15,16} and to avoid localization biases.¹⁷ In that respect, PSF engineering techniques have the aim of enhancing these shape changes. Nevertheless, any optical aberration, polarization distortion, or misalignment in the imaging system can affect the final shape of the PSFs and, thus, lead to an inaccurate estimation of the parameters. In particular, polarization aberrations are delicate to correct as they require additional adaptive strategies that account for the vectorial nature of light propagation.^{18,19}

A common solution to this problem is to perform a set of calibration measurements^{20,21} and use them in a phase retrieval optimization algorithm to determine the aberrations present in the system. For this approach to work, it is important to have an accurate model for light's propagation from a known source to the camera and to incorporate phase diversity, e.g., through measurements at

varying focal planes, both to avoid falling into local minima and to accelerate convergence.²² For single-molecule localization microscopes, the standard approach is to measure the PSFs generated by fluorescent nanobeads in varying focal planes and use this Z-stack of images in a phase retrieval algorithm.²³ Initial approaches relied on scalar models, assuming a point source emitting a spherical wavefront along with a scalar pupil representing the aberrations. In this simplified case, Gerchberg–Saxton iterative algorithms^{24,25} can be used, thus reducing the complexity of the implementation. However, these algorithms are less flexible since the pupil model cannot be chosen freely, and they are not directly applicable to full vectorial models that consider the dipolar emission pattern and polarization distortions.^{20,21,26} A more flexible approach is offered by casting the phase retrieval problem as a nonlinear optimization routine where most parameters in the model can be found during the optimization process, although this requires providing analytical formulas for the gradients in these parameters, hence complicating implementation.²⁵ More accurate models that take into account the vector nature of the emitted light have also been proposed.^{27,28} They incorporate the effect of the interface between the medium embedding the fluorescent particles and the coverslip, which causes extra aberrations, polarization-dependent transmission, and supercritical angle fluorescence (SAF) radiation.^{28–30} However, these approaches assume a scalar mask to characterize all the remaining aberrations, thus preventing them from correcting any residual polarization-dependent effects. Moreover, the polarized components of the emitted fields are eventually summed in order to model an unpolarized measurement, which is not directly applicable to polarized PSFs in SMOLM.

Recently, new SMOLM techniques have been proposed that use birefringent elements either to efficiently encode the 3D orientation and 3D localization information of the emitting dipole into the shape of the two polarization components of the PSF^{10,12,13,31} or to understand the intensity and/or shape of their different polarization projections.^{8,9,32} For such approaches, it is essential to take into account the emission pattern of the dipole source, its interaction with the interface between the embedded medium and coverslip, and any polarization-dependent aberrations.³³ To address these issues, a model is used here where all vector aspects of the propagation of the light emitted by the source to the back focal plane (BFP) are taken into account and where the aberrations and polarization distortions are represented by a birefringent distribution at the pupil plane (BDPP) modeled with a spatially varying Jones matrix.³⁴ It is shown that in order to characterize this BDPP, it is necessary to

introduce polarization projection diversity, obtained by projecting the PSFs onto various polarization states, in addition to the phase diversity given by changing the location of the focal plane. The PSF images generated with these two diversities form a PZ-stack that is fed into a nonlinear optimization algorithm that allows retrieving the unknown BDPP. This approach allows for the inclusion of many parameters that are necessary for proper characterization, such as photo-bleaching amplitudes, background illumination, or diversity-dependent tilts. Accurate and computationally amenable models for the light produced by fluorescent beads (commonly used for calibration measurements) are also included. These models take into account the unpolarized nature of the emitted light and the blurring due to bead size.³⁵ The software package PYPSF-STACK used for the modeling of PZ-stacks and the phase retrieval process can be found in Ref. 36. The retrieval algorithm was implemented with the neural network framework PYTORCH,³⁷ which greatly simplifies its implementation and flexibility due to the automatic computation of all gradients and its integration with GPUs if available. While the emphasis of this work is on the characterization of BDPPs, both the theory and code are equally applicable to scalar-only aberration pupil distributions, as shown in the supplementary material.

II. POINT-SPREAD FUNCTION OF A DIPOLAR SOURCE

A. Field at the back-focal plane

In order to properly characterize a given system, one must first derive an accurate model. Here, the situation depicted in Fig. 1 is considered: the incoherent light emitted by a fluorescent bead is collected by an immersion microscope objective with a high numerical aperture (NA). The bead is assumed to be placed at an axial position z_0 from the interface between its embedding medium with refractive index n_i and the coverslip assumed to have the same refractive index n_f as that of the immersion liquid. (Because the bead is behind this interface, $z_0 < 0$.) The index mismatch between these media introduces extra aberrations (see the supplementary material, Secs. SI and SII, for more details) and polarization-dependent transmission following the Fresnel coefficients. It also allows the coupling of evanescent components emitted by the bead when $n_f > n_i$, leading to SAF radiation, which can make up a significant portion of the light detected by the camera.^{30,38–40} All of these effects can be encapsulated into the Green tensor for a dipolar source at the BFP of the microscope objective, which can be written as^{29,41–43}

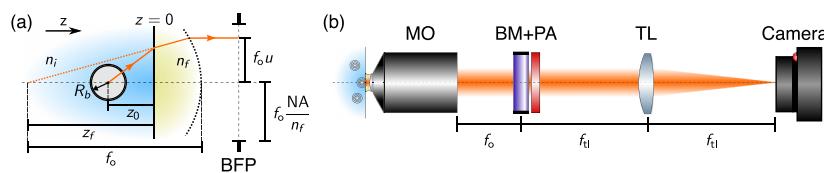


FIG. 1. Schematic of the experimental setup for the collection and shaping of the emission by a source. (a) Position of the fluorescent nanobead of radius R_b , embedded in a medium of refractive index n_i , with respect to the interface created by the coverslip and the immersion liquid of the microscope objective, with refractive index n_f , and the focal plane located at z_f , $z_f < 0$ ($z_f > 0$) if the focal plane is in the medium with refractive index n_i (n_f). (b) Schematic of the collection arm composed of a microscope objective (MO), followed by a birefringent mask (BM), and a polarization analyzer (PA) at the back focal plane (BFP). The light at the BFP is then focused onto the camera by the tube lens (TL) of focal length f_{tl} .

$$\mathbb{G}(\mathbf{u}) = D^{(\Delta)}(\mathbf{u})S(\mathbf{u})\mathbb{g}(\mathbf{u}), \quad (1)$$

where

$$S(\mathbf{u}) = \exp \left\{ ikn_f |z_0| \left[\frac{n_i}{n_f} \sqrt{1 - \left(\frac{n_f u}{n_i} \right)^2} - \alpha \sqrt{1 - u^2} \right] \right\}, \quad (2)$$

$$D^{(\Delta)}(\mathbf{u}) = \exp \left(ikn_f \Delta \sqrt{1 - u^2} \right),$$

and \mathbb{g} is a 2×3 matrix (see the supplementary material, Sec. SI, for the explicit form) that includes the effect of the Fresnel coefficients for the interface and depends only on the normalized pupil coordinates at the BFP, $\mathbf{u} = (u_x, u_y)$, whose maximum value is limited by the NA through $\|\mathbf{u}\|_{\max} = \text{NA}/n_f$, and $k = 2\pi/\lambda$, with λ being the wavelength of the emitted light. (Note that the definition of \mathbf{u} differs from that in Ref. 35 by a factor of n_f .) For simplicity, it was assumed that the source is centered on the optical axis. The position of the focal plane z_f shown in Fig. 1 requires specifying two parameters: $z_f = \Delta - \alpha|z_0|$, where α is a dimensionless parameter fixing the position chosen for the reference focal plane (RFP) and Δ is the position of the focal plane with respect to the RFP. The parameter α is generally taken to be the one producing the best focus when $\Delta = 0$, although its definition is not unique (one option being the one that minimizes the wavefront error⁴⁴). As detailed in the supplementary material, minimizing the root-mean-square (rms) spot size with fourth-order corrections for the wavefront difference between the SAF and defocus produces the simple expression,

$$\alpha = \frac{n_r^3 (32n_r^2 + 11)}{24n_r^4 + 16n_r^2 + 3} \quad (3)$$

with $n_r = n_f/n_i$, which gives satisfactory results for both water/oil ($n_r = 1.1391$) and air/oil ($n_r = 1.515$) interfaces. This value is taken as the default in this work, but it can be changed to match other experimental configurations. In the supplementary material, Sec. SII, we include other possible criteria for defining the best focus plane in our scenario: index matching between the objective immersion liquid and the coverslip. For situations in which this is not the case and the coverslip has to be considered as a parallel slab of a finite thickness, expressions for the best focal plane can be found in Ref. 44.

B. Birefringence distribution at the pupil plane

As mentioned in the introduction, the goal of this work is to characterize any residual aberrations and polarization-dependent effects due to stress and/or interfaces, the use of masks aimed at tailoring the PSF, or a combination thereof. These residual effects produce a birefringent distribution at the BFP, shown as a mask in Fig. 1, which can be represented by a 2×2 space-dependent Jones matrix,³⁴

$$\mathbb{J}_M(\mathbf{u}) = e^{i2\pi W(\mathbf{u})} \begin{pmatrix} q_0(\mathbf{u}) + iq_3(\mathbf{u}) & q_2(\mathbf{u}) + iq_1(\mathbf{u}) \\ -q_2(\mathbf{u}) + iq_1(\mathbf{u}) & q_0(\mathbf{u}) - iq_3(\mathbf{u}) \end{pmatrix}, \quad (4)$$

where W represents a scalar aberration function, and the scalar pupil functions q_j are real. This matrix can be made unitary by enforcing the condition $\sum_j q_j^2 = 1$, and otherwise, it can include

apodization effects. The birefringence distribution \mathbb{J}_M can be used to represent both a mask introduced to shape the PSFs, such as a stress-engineered optic (SEO)^{13,45,46} or a q-plate,^{12,47,48} and the scalar and polarization aberrations of the system. (The simplest case of a scalar mask corresponds to $q_0 = 1$, $q_1 = q_2 = q_3 = 0$.) This Jones matrix acts on the Green tensor of the dipolar source at the BFP, and the result is then propagated to the image plane via

$$\mathbb{G}_{\text{IP}}(\tilde{\mathbf{p}}) = \iint \mathbb{J}_M(\mathbf{u}) \cdot \mathbb{G}(\mathbf{u}) \exp \left(-ik \frac{n_f \tilde{\mathbf{p}}}{M} \cdot \mathbf{u} \right) d^2 u, \quad (5)$$

where $\tilde{\mathbf{p}}$ denotes the transverse position at the image plane, and M is the total magnification of the system. Note that for setups using a relay system, the coordinates of \mathbb{G} should be flipped, $\mathbf{u} \rightarrow -\mathbf{u}$.

For a fully polarized dipole oriented along the unit vector $\hat{\mathbf{p}} = (\mu_x, \mu_y, \mu_z)$, the electric field distribution at the image plane is given by

$$\mathbf{E}_{\text{IP}}(\mathbf{u}) = \mathbb{G}_{\text{IP}}(\tilde{\mathbf{p}}) \cdot \hat{\mathbf{p}}. \quad (6)$$

Therefore, the three columns of the Green tensor represent the field distribution at the pupil plane produced by a dipole along each of the three coordinate axes. Since the information about the orientation of the dipole is encoded into the components of the Green tensor, in order to retrieve the dipole's orientation from the shape of the PSF, it is necessary to spatially separate its projections into two appropriately chosen orthogonal polarization states, such as horizontal and vertical linear or left and right circular. These polarization projections can be represented by two matrices, \mathbb{P}_1 and \mathbb{P}_2 , thus for a fully polarized dipole, the PSF pair is given by

$$I_{\text{IP},j}(\tilde{\mathbf{p}}) = \|\mathbb{P}_j \cdot \mathbb{G}_{\text{IP}}(\tilde{\mathbf{p}}) \cdot \hat{\mathbf{p}}\|^2 \quad (7)$$

with $j = 1, 2$. For unpolarized emitters such as fluorescent beads used for characterization, the PSFs are given by the incoherent sum of the components of the final Green tensor, which amounts to the incoherent sum of the dipoles oriented along the three coordinate axes. In this case, the pair of PSFs is given by

$$I_{\text{IP},j}(\tilde{\mathbf{p}}) = \|\mathbb{P}_j \cdot \mathbb{G}_{\text{IP}}(\tilde{\mathbf{p}})\|^2. \quad (8)$$

Note that if no polarization projection is used, the PSF is given by the sum of the pair of PSFs $I_{\text{IP},1}$ and $I_{\text{IP},2}$.

III. FORWARD MODEL FOR CHARACTERIZING A BIREFRINGENT DISTRIBUTION AT THE PUPIL PLANE

A. Polarization aberrations

The goal of this work is to provide a method for determining from a set of calibration PSFs the system's BDPP, represented by a Jones matrix of the form in Eq. (4). The various functions in this expression must first be expanded in terms of a basis, whose expansion coefficients are then determined through nonlinear optimization. Common basis choices for optimization are the Zernike polynomials⁴⁹ and a direct pixel representation,²⁸ both of which result in comparable speeds since they employ the same number of discrete Fourier transforms, the most costly operation even when implemented through the fast Fourier transform (FFT) algorithm.

In what follows, a decomposition in the Zernike polynomial basis is used given that its elements are easy to interpret and allow an accurate description with fewer parameters (although examples using the pixel-based approach are shown in the supplementary material, Sec. VI). The components of the Jones matrix are then decomposed as

$$W(\mathbf{u}) = \sum_l c_l^{(W)} Z_l(\mathbf{u}/u_{\max}), \quad (9a)$$

$$q_l(\mathbf{u}) = \sum_l c_l^{(j)} Z_l(\mathbf{u}/u_{\max}), \quad (9b)$$

where $j = 0, \dots, 3$, and a single index notation was used for the basis elements Z_l (e.g., the Fringe notation⁵⁰). Note that \sum' in the expression for W indicates that the terms corresponding to piston and defocus are excluded. The piston term would simply add a global phase that is unimportant and cannot be determined from intensity measurements, while the defocus term would be redundant with the more accurate defocus parameter Δ in Eq. (1), which is also included as an optimization parameter. In addition, note that any misalignment of the PSFs with respect to the optical axis is corrected by the scalar tilts present in W . The number of Zernike polynomials included in the decomposition depends on the expected spatial dependence of the BDPP. In this work, we found that the first 15 polynomials are usually sufficient, even when retrieving discontinuous BDPPs such as a q-plate (see Fig. 3). It should also be mentioned that the smooth description provided by the Zernike model works despite fast variations due to SAF radiation since these are already included in the propagation model. This Zernike expansion is inspired by the Nijboer-Zernike theory, where a scalar mask would be separated into real and imaginary parts before decomposing them in terms of Zernike polynomials.⁵¹⁻⁵³

B. Phase and polarization diversity

It is common practice in phase retrieval algorithms for optical microscopes to assume access to a stack of intensity images for varying focal distances Δ_ζ from the location of the best focus (a Z-stack). The phase diversity provided by the varying focal distances is taken into account by the phase factor $D^{(\Delta_\zeta)}(\mathbf{u})$ defined in Eq. (2) within the Green tensor in Eq. (1). This additional information, referred to as phase diversity, helps the algorithm converge to an appropriate solution without falling into local minima, as well as discriminate between vortices with opposite topological charges.

While a Z-stack is sufficient to determine scalar masks and aberrations, it is not so for BDPPs since it does not discriminate between the true \mathbb{J} and its unitary transformations $\mathbb{J}_{\mathbb{U}} = \mathbb{U} \cdot \mathbb{J}$, where \mathbb{U} is a constant unitary matrix, as exemplified in Sec. V A. Therefore, it is necessary to include information about the polarization dependence of the PSFs used for the retrieval. This additional information is obtained by introducing a polarization analyzer after the birefringent mask (see Fig. 1), composed of a combination of waveplates and polarizers where at least one element rotates to generate various polarization projections of the output. This polarization diversity is modeled by a set of constant Jones matrices $\mathbb{P}^{(p)}$ that are applied to

the Green tensor at the BFP along with the defocus terms for the phase diversity in order to generate a PZ-stack of Green tensors,

$$\mathbb{G}_{\text{BFP}}^{(\zeta,p)}(\mathbf{u}) = D^{(\zeta)}(\mathbf{u}) \mathbb{P}^{(p)} \cdot \mathbb{J}_M(\mathbf{u}) \cdot \mathbb{G}_0(\mathbf{u}). \quad (10)$$

It is important to notice that the constant matrices $\mathbb{P}^{(p)}$ cannot be unitary since they would then have no effect on the shape of the PSFs. The simplest nonunitary matrix to implement is a projection matrix obtained by placing a linear polarizer at the end of the waveplate sequence.

This PZ-stack of Green tensors is then propagated to the image plane via

$$\mathbb{G}_{\text{IP}}^{(\zeta,p)}(\rho) = \iint \mathbb{G}_{\text{BFP}}^{(\zeta,p)}(\mathbf{u}) \exp\left(-ik \frac{n_f \rho}{M} \cdot \mathbf{u}\right) d^2 u \quad (11)$$

and its components are added incoherently (modeling an unpolarized source) to obtain a PZ-stack of PSFs,

$$I_{\text{IP}}^{(\zeta,p)}(\rho) = \left\| \mathbb{G}_{\text{IP}}^{(\zeta,p)} \right\|^2 = \sum_{i=x,y} \sum_{j=x,y,z} |G_{\text{IP},ij}^{(\zeta,p)}(\rho)|^2, \quad (12)$$

such as the one shown in Fig. 2. The polarization projections \mathbb{P}_1 and \mathbb{P}_2 used to extract the orientation information can also be used to define the polarization diversity, as discussed in Sec. V. It is worth noting that while experimentally the polarization diversity happens at the BFP, computationally it is better to perform it at the image plane in order to avoid the computation of unnecessary FFTs. However, as discussed in Sec. VI, there are some situations in which it must be implemented at the BFP.

C. Modeling the measured PSFs

Before comparing the PZ-stack computed by the model presented thus far with the one measured experimentally, it is necessary to account for other effects. First, depending on the size of the fluorescent bead, it might be necessary to include a blurring effect. As shown in Ref. 35, the exact three-dimensional blurring corresponds to a superposition of two-dimensional convolutions between the PSFs generated by point sources located along the longitudinal diameter of the bead with a kernel that weights more heavily the contributions near the center of the bead and vanishes for those at the poles. This exact blurring cannot be rewritten as a three-dimensional convolution due to the lack of translation invariance along the axis of propagation, but it can be approximated through a semi-analytic method based on a Taylor expansion. By keeping the first term of the expansion, we obtain a two-dimensional blurring model based on a two-dimensional convolution of the PSFs. If instead we keep two terms in the expansion, we get a model for the blurring along both the transverse and longitudinal directions, which is computed via two-dimensional convolutions of the PSFs and their second-derivatives with respect to the bead's axial position, thus giving a three-dimensional blurring model. Both of these approximate models are computationally more efficient than the exact one and have been implemented in PYPSFSTACK. Moreover, they can be used during the BDPP retrieval process, as shown in Sec. V B, albeit at a computational cost due to the need for supplementary Fourier transforms. Nonetheless, they allow for the analytic computation of the gradients, hence significantly limiting the computational slowdown; using readily available functions to blur images, on the other hand,

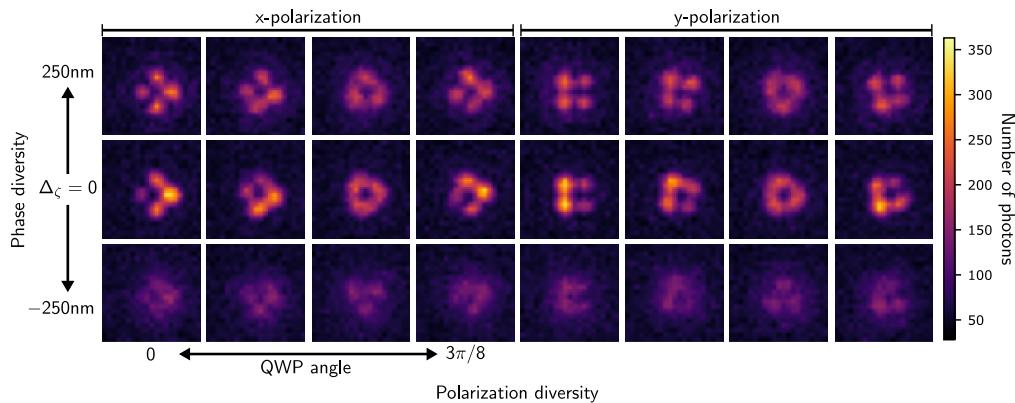


FIG. 2. PZ-stack for an unpolarized emitter shaped with an SEO element and modeled with PYPSFSTACK. Only the PSFs at the initial, middle, and final values of the defocus parameter Δ_ζ used for the phase diversity in the retrieval of the BDPP shown in Fig. 3 are shown. For each Δ_ζ , we show the PSFs for all polarization projections of diversity. This polarization diversity is generated by a rotating quarter wave plate followed by a projection onto linear horizontal or vertical polarization states.

such as the one used in Ref. 28 with a Gaussian kernel, would require the gradients to be computed via finite differences.

The other two effects that must be considered are the photobleaching of the fluorescent beads and background illumination. Photobleaching causes the number of photons emitted by the bead to diminish with time. Its effect can be taken into account by implementing an overall amplitude factor $a^{(p,\zeta)}$ that depends on both the phase and polarization diversities. Background illumination is then added incoherently to the photobleached PSF stack. The simplest model is to assume that the background illumination is determined by a constant term $b^{(p,\zeta)}$ that depends on the phase and polarization diversities. Extra terms for a spatially dependent background can be added if needed.⁵⁴ The modeled PZ-stack to be compared to the experimental measurements is then given by

$$I_{\text{tot}}^{(\zeta,p)}(\rho) = a^{(p,\zeta)} \mathcal{B}[I_{\text{IP}}^{(\zeta,p)}(\rho); R_b] + b^{(p,\zeta)}, \quad (13)$$

where \mathcal{B} denotes the blurring operation that depends on the bead radius. In the supplementary material, Sec. SIII, we include a short derivation of the blurring operation when beads are modeled as solid spheres, following Ref. 35.

D. Assessing the accuracy with a cost function

The last piece of the forward model to be considered is the choice of cost function used to tune the parameters so that the modeled PSFs, $I_{\text{tot}}^{(\zeta,p)}$, best fit the measured ones, $I_{\text{exp}}^{(\zeta,p)}$. In the absence of noise, any choice of a cost function that has a minimum when the two quantities are the same should provide the same result. However, noise is always present in experimental measurements and, thus, must be taken into account. In single molecule fluorescent microscopy, one is normally limited by shot noise following a Poisson distribution, in which case the log-likelihood cost function,⁵⁵

$$\mathcal{C} = -\sum_{\zeta,p} \iint w(\rho) \{ I_{\text{exp}}^{(\zeta,p)}(\rho) \log [I_{\text{tot}}^{(\zeta,p)}(\rho)] - I_{\text{tot}}^{(\zeta,p)}(\rho) \} d^2\rho \quad (14)$$

should be used. Here, w denotes a binary window function used to represent the region considered for optimization due to the smaller size of the experimental data and/or to exclude bad pixels from the camera. Another common option for the cost function is the sum of differences squared, which is appropriate when the noise follows a Gaussian distribution. Both of these options are implemented in PYPSFSTACK. Note that for the choice of the cost function to be consistent, the values of $I_{\text{exp}}^{(\zeta,p)}$ must actually follow the assumed distribution. This means that the images should not be denoised and that the offset of the camera should be removed.

IV. IMPLEMENTING THE NONLINEAR OPTIMIZATION

The goal of the nonlinear optimization routine is to find the set of optimization parameters in the forward model that minimizes the cost function, assessing the differences between the measured and modeled PZ-stacks. This is achieved by supplying an optimization algorithm, such as Adam⁵⁶ or L-BFGS,⁵⁷ with a function that uses all the current values of the parameters to compute the forward model all the way to the value of the cost function. In addition, one should supply the optimization algorithm with another function that performs a backward computation to obtain the gradients of the cost function with respect to all the optimization parameters. These gradients are used to change the parameter values until a minimum is reached. The gradient computation is straightforward but tedious and can be achieved by following the rules outlined in Ref. 58. However, an advantage of implementing nonlinear optimization with the neural network framework PYTORCH is that only the forward model must be implemented explicitly since the backward model for computing the gradients of the various parameters is computed automatically. This framework also offers the most common optimization algorithms.

The implementation using PYTORCH greatly simplifies the vectorial phase retrieval algorithm, as shown in Algorithm 1. First, the system parameters and the initial values for the optimization parameters are used to initialize the forward model used to compute the estimated PSFs and evaluate the current value of the cost function. For all the examples presented in this work, the BDPP was

ALGORITHM 1. Vectorial phase retrieval.

Input

θ_{sys}	system parameters ($n_i, n_f, \lambda, z_0, \alpha, R_b, NA, M, p_{\text{cam}}, N$)
$\theta_{\text{opt}}^{(0)}$	initial optimization parameters ($\Delta, c_l^{(W)}, c_l^{(j)}, R_b, a^{(p,\zeta)}, b^{(p,\zeta)}$)
$l_{\text{exp}}^{(\xi, p)}$	measured PSFs
$\Delta \zeta$	N_z defocuses for phase diversity
$\mathbb{P}^{(p)}$	N_p Jones matrices for polarization diversity
\mathcal{B}	blurring model
η	learning rate
N_{iter}	number of iterations

Output

J_M Jones matrix at the pupil plane

Procedure

```

1: Initialize model with  $\theta_{\text{sys}}$ ,  $\theta_{\text{opt}}^{(0)}$ , and  $\mathcal{B}$ 
2: for  $n \leftarrow 1$  to  $N_{\text{iter}}$  do
3:   Compute forward model            $\triangleright$  See Algorithm 2
4:   Compute gradients       $\triangleright$  Uses automatic differentiation
5:   Update optimization parameters  $\triangleright$  Takes a step size of  $\eta$  using the gradients
6: end for

```

initialized as the identity matrix, i.e., $c_0^{(0)} = 1$ and $c_l^{(W)} = c_l^{(j)}$ for all $j, l \neq 0$. However, this can be changed if there is *prior* knowledge about the BDPP. After initialization, the phase retrieval algorithm enters a for loop in which the forward model is used to compute

the cost function using the current values of the parameters. Then, automatic differentiation is called for to obtain the gradients. Finally, the hyperparameters of the chosen optimization algorithm and the values of the gradients are used to update the values for all the

ALGORITHM 2. Implementation of the forward model.

Input

θ_{sys}	system parameters ($n_i, n_f, \lambda, z_0, \alpha, NA, M, p_{\text{cam}}, N$)
$I_{\text{exp}}^{(\zeta, p)}$	measured PSFs
∂	Green tensor
Δ_ζ	N_z defocuses for phase diversity
$\mathbb{P}^{(p)}$	N_p Jones matrices for polarization diversity
\mathcal{B}	blurring model
θ_{opt}	optimization parameters ($\Delta, c_l^{(W)}, c_l^{(j)}, R_b, a^{(p, \zeta)}, b^{(p, \zeta)}$)

Procedure Forward model

- 1: $\mathbb{G}(\boldsymbol{\ell}) \leftarrow D^{(\Delta)}(\boldsymbol{\ell}) S(\boldsymbol{\ell}) \mathbb{g}(\boldsymbol{\ell})$ ▷ Compute Green tensor
- 2: $\mathbb{G}_J(\boldsymbol{\ell}) \leftarrow \mathbb{J}[\boldsymbol{\ell}; c_l^{(W)}, c_l^{(j)}] \cdot \mathbb{G}(\boldsymbol{\ell})$ ▷ Apply birefringent mask to Green tensor
- 3: $\mathbb{G}^{(\zeta)}(\boldsymbol{\ell}) \leftarrow D^{(\Delta_\zeta)}(\boldsymbol{\ell}) \mathbb{G}_J(\boldsymbol{\ell})$ ▷ Apply phase diversity
- 4: $\mathbb{G}_{\text{IP}}^{(\zeta)}(\boldsymbol{\ell}) \leftarrow \text{FFT}\{\mathbb{G}^{(\zeta)}(\boldsymbol{\ell})\}$ ▷ Propagate to image plane
- 5: $\mathbb{G}_{\text{IP}}^{(\zeta,p)}(\boldsymbol{\ell}) \leftarrow \mathbb{P}^{(p)} \cdot \mathbb{G}_{\text{IP}}^{(\zeta)}(\boldsymbol{\ell})$ ▷ Apply polarization diversity
- 6: $I^{(\zeta,p)}(\boldsymbol{\ell}) \leftarrow \sum_i \sum_j |G_{\text{IP},i,j}^{(\zeta,p)}(\boldsymbol{\ell})|^2$ ▷ Compute intensity
- 7: $I_{\text{blur}}^{(\zeta,p)}(\boldsymbol{\ell}) \leftarrow \mathcal{B}[I_{\text{IP}}^{(\zeta,p)}(\boldsymbol{\ell}), R_b]$ ▷ Apply blurring
- 8: $I_{\text{tot}}^{(\zeta,p)}(\boldsymbol{\ell}) \leftarrow a^{(p,\zeta)} I_{\text{blur}}^{(\zeta,p)}(\boldsymbol{\ell}) + b^{(p,\zeta)}$ ▷ Photobleaching and background illumination
- 9: $\mathcal{C} \leftarrow -\sum_{\boldsymbol{\ell}, \zeta, p} w(\boldsymbol{\ell}) \left\{ I_{\text{exp}}^{(\zeta,p)}(\boldsymbol{\ell}) \ln \left[I_{\text{tot}}^{(\zeta,p)}(\boldsymbol{\ell}) \right] - I_{\text{tot}}^{(\zeta,p)}(\boldsymbol{\ell}) \right\}$ ▷ Compute cost function

Return C

optimization parameters. This process is iterated N_{iter} times, after which the process is completed. Note that other criteria can be used to terminate the process; for example, the process could be terminated when the change in the cost function falls below a set limit. All the results presented in this work were obtained using the Adam optimizer since it is more robust than simple gradient descent and faster and less memory intensive than L-BFGS. While this optimization algorithm has several parameters that can be tuned, in practice, it was found that choosing the appropriate learning rate was sufficient to obtain satisfactory solutions. The convergence was assessed by analyzing the evolution of the cost function. A more detailed explanation can be found in the examples in the repository.³⁶

To implement the forward model, all calculations are performed numerically using FFTs, requiring all spatial quantities to be sampled consistently with the camera's pixel pitch p_{cam} , so that the size at the BFP is fixed to $L_{\text{pupil}} = \lambda M / (p_{\text{cam}} n_f)$. The total size at the BFP is divided into $N \times N$ points used for the computation, and the resulting two-dimensional sampling is labeled with the lexicographic index ℓ , which takes over all spatial dependence in terms of \mathbf{u} and ρ , for instance, $\mathbb{G}_0(\mathbf{u}) \rightarrow \mathbb{G}_0(\ell)$. All the parameters and steps necessary to compute the forward model are summarized in Algorithm 2.

V. NUMERICAL EXPERIMENTS

A. Polarization diversity vs more phase diversity

To exemplify the implementation of the phase retrieval algorithm and, in particular, the need for polarization diversity to properly characterize a BDPP, we consider the retrieval of two masks

used recently for estimating the position and orientation of single emitters: an SEO^{45,46} with

$$\mathbb{J}_M = \cos\left(\frac{cu}{2}\right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + i \sin\left(\frac{cu}{2}\right) \begin{pmatrix} \cos \phi & -\sin \phi \\ -\sin \phi & -\cos \phi \end{pmatrix} \quad (15)$$

and the parameter $c = 1.25\pi$, and a q-plate with unit topological charge,^{12,47,48} also known as a vortex waveplate, with

$$\mathbb{J}_M = \begin{pmatrix} i \cos 2\phi & i \sin 2\phi \\ i \sin 2\phi & -i \cos 2\phi \end{pmatrix}, \quad (16)$$

both of which are shown in Fig. 3. Following the strategy outlined thus far, PYPSFSTACK is used to model a PZ-stack for each BDPP, such as the one shown in Fig. 2 for the SEO. For the phase diversity, images are taken from -250 to 250 nm of the RFP with a step size of 50 nm. For the polarization diversity, a quarter wave plate (QWP) is rotated from 0 to $3\pi/8$ with a step of $\pi/8$ and is followed by a Wollaston prism that projects the output onto horizontal and vertical linear polarizations. This choice is inspired by the setup used in Ref. 13, where the SEO is followed by a QWP and a Wollaston prism to project the PSF into left and right circular polarizations. It is also assumed that $10\,000$ photons arrive on average at the camera to form the PSFs, to which an additional 50 photons per pixel are added as background. Noise following a Poisson distribution is incorporated. Note that the choice of diversity follows the results presented in the supplementary material, Sec. SV, where the accuracy of the retrieval is studied as a function of measured photons for

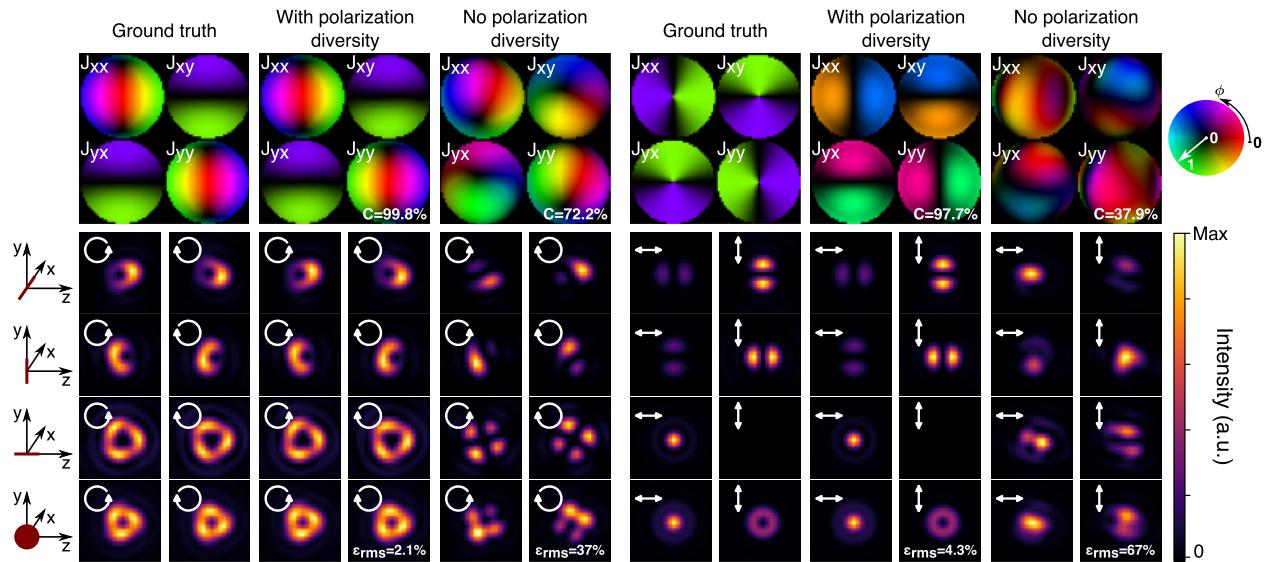


FIG. 3. BDPP retrieval with and without polarization diversity. (First row) Elements of the Jones matrix for the ground truth and the retrieved BDPP with and without polarization diversity for (left) an SEO and (right) a q-plate. In addition, shown are the PSFs generated by point dipoles (second to fourth row) oriented along each of the three Cartesian axes and (last row) an unpolarized dipole, all for each of the standard projectors used for each type of birefringent window. For the SEO, the PSFs are projected onto left- and right-circular polarizations, while for the q-plate they are projected onto linear horizontal and vertical polarizations. In addition, given are the correlation between the retrieved and true BDPP, as well as the rms error between the retrieved and true PSFs generated along each of the three Cartesian axes and an unpolarized dipole.

various combinations of phase and polarization diversity. For simplicity, we consider a small fluorescent bead with $R_b = 10$ nm so that spatial blurring is negligible. Nonetheless, the exact blurring model introduced in Ref. 35 was used to compute the PZ-stacks for testing the BDPP retrieval algorithm. Moreover, a random error of the order of 20 nm was introduced to the distance between the bead and the coverslip, and one of the order of 25 nm to the location of the RFP. The algorithm was run on a desktop computer with an AMD Ryzen 9 3900X 12-Core central processing unit (CPU) and a NVIDIA GeForce RTX 2080 graphics processing unit (GPU). Using a pupil sampling of 128×128 points, the full retrieval routine consisting of 200 iterations took only a few seconds (see Table I for more details).

The results of the procedure are shown in Fig. 3, where the Jones matrix for the ground truth is compared to the ones retrieved from the PZ-stacks. This figure also shows the PSFs formed by dipoles oriented along the three Cartesian axes and by an unpolarized dipole (constructed as an incoherent mixture of the three orthogonal dipoles). The PSFs shown are modeled using the standard projectors \mathbb{P}_1 and \mathbb{P}_2 [Eqs. (7) and (8)] for each birefringent mask, for the SEO, the output is projected onto the left and right circular polarizations, while for the q-plate, the output is projected onto the horizontal and vertical polarizations. To assess the accuracy of the retrieval, both the correlation between the retrieved and true BDPPs as well as the root-mean square (rms) error, ϵ_{rms} , between the retrieved and true PSFs are shown in Fig. 3. For the SEO, the retrieved BDPP and the corresponding PSFs are almost indistinguishable from those corresponding to the ground truth. However, for the q-plate, there are appreciable differences between the original and retrieved BDPPs. The deviation at the center is due to the chosen model, since the Zernike polynomials struggle to reproduce the singularity at the center of the q-plate. In general, this type of singularity does not arise as an aberration but rather by design of the mask, so that it can be incorporated as prior information. For instance, on top of the Zernike decomposition, a q-plate could be added to the model with its center and orientation as optimization parameters. Another noticeable difference in the case of the q-plate is that the phase between the two rows of the Jones matrix is not correct. This error happens only for birefringent masks that generate rotationally symmetric PSFs for unpolarized emitters for all polarization projections. For these masks, the algorithm cannot determine the global phase between the rows of the corresponding Jones matrix. This problem can be solved by placing the device that introduces the polarization

TABLE I. Runtime per iteration for a pupil sampling of 128×128 points, 11 phase diversities, and eight polarization diversities for different blurring models, with and without the diversity-dependent tilts mentioned in Sec. VI.

Model	Time per iteration	
	CPU (ms)	GPU (ms)
No blurring and no tilts	45	9.6
2D blurring and no tilts	110	13.7
3D blurring and no tilts	273	21.4
No blurring and with tilts	222	16.7
2D blurring and with tilts	287	20.6

diversity before the birefringent mask. Even when this solution is not used, the correlation (when neglecting the phase difference between the rows) is quite high, and the reproduced PSFs for any dipole orientation are again indistinguishable from the true ones. Therefore, these differences are inconsequential for the purposes of single molecule fluorescence microscopy. In addition, the error introduced for z_0 and α has no impact on the retrieval as long as the defocus Δ is included as an optimization parameter. Similar results using a pixel-based model are shown in the supplementary material, Sec. VI, as is a comparison of the convergence of the cost function between the two models.

For comparison, the retrieval of the same birefringent mask is also performed without using polarization diversity. To make this comparison consistent and fair, the number of photons reaching the camera and those in the background illumination is doubled since there is no polarization separation. Moreover, the phase diversity images are now taken from -250 to 250 nm of the RFP with a reduced step size of 5 nm, so that the total number of PSFs used is larger than in the previous case. The results are shown in Fig. 3, where the algorithm is seen to fail to retrieve the appropriate BDPP, and thus, it cannot generate the correct PSFs when they are projected onto a given polarization state. These results show the need for polarization diversity when characterizing a BDPP. A more detailed comparison between the scalar and vector models is provided in the supplementary material, Sec. IV. Note that the runtime is much longer due to the increase in the number of FFTs required. (This runtime should be comparable with that of a model that includes tilts but no blurring; see Table I.)

B. Incorporating blurring due to size to the birefringent distribution retrieval

As an additional test of the present implementation, the retrieval of a BDPP from a highly blurred PZ-stack such as the one shown in Fig. 4 is considered. Using the SEO as an example, a PZ-stack is constructed as in Sec. III B and III C with the significant difference of increasing the radius of the nanobead, which is chosen randomly from a uniform distribution between 150 and 170 nm. Here again, the distance to the coverslip is taken to be equal to the radius of the bead. As shown in Fig. 4, this bead size causes a significant blur in the resulting PSFs. The blurred PSFs were computed with the exact blurring presented in Ref. 35, but this approach turned out to be computationally expensive and not necessary for the retrieval procedure, as shown in what follows.

A first attempt to retrieve the BDPP is performed by completely neglecting the blurring effect. However, as shown in Fig. 4, this approach fails to retrieve the appropriate BDPP, showing that bead size effects cannot be neglected in this case. Next, as discussed in the supplementary material, Sec. SIII, the two approximate models presented in Ref. 35 were used for the retrieval, where the radius of the bead is used as an optimization parameter for the blurring with an initial value of 150 nm. The first of these approaches is a 2D convolution with a kernel given by a spherical Bessel function. This approach produces a satisfactory result, whose retrieved BDPP has a correlation of 98.6% with the ground truth. The second is the semi-analytic model described in Sec. III C for

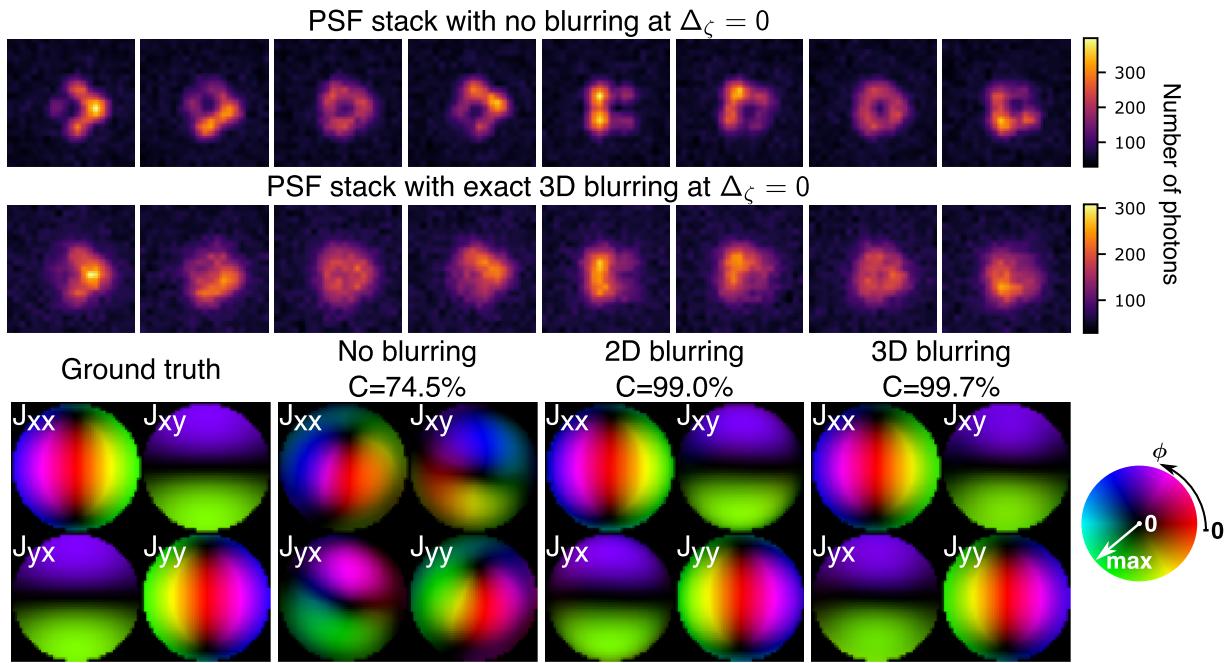


FIG. 4. BDPP retrieval from highly blurred PSFs. Comparison of PSFs for an SEO at the reference focal plane ($\Delta_\zeta = 0$) for the same polarization diversities as those shown in Fig. 2 for (first row) a point source and (second row) a fluorescent bead with a radius $R_b = 150$ nm. (Third row) Elements of the Jones matrix for the ground truth and the retrieved BDPPs without blurring, with a two-dimensional blurring model, and a three-dimensional semi-analytic blurring model.

reproducing the effects of the three-dimensional blurring, based on a Taylor expansion around the center of the bead. In this case, a BDPP indistinguishable from the true one is again retrieved with a slightly better correlation of 99.5%. As shown in Table I, both blurring models significantly increase the time per iteration due to the required supplementary Fourier transforms. This increase is particularly important for the 3D blurring model, which requires the computation of the first and second derivatives of the Green tensor with respect to the bead's axial position, hence tripling the number of Fourier transforms in addition to two convolutions at the image plane. Therefore, the small gain provided by the 3D model might not justify the extra computational resources needed for it. Nonetheless, both models are able to perform the full retrieval in well under a minute with the CPU and in a few seconds with the GPU.

VI. CHARACTERIZATION FROM EXPERIMENTAL DATA

After validating the proposed retrieval procedure on simulated data, we apply it to retrieve a BDPP distribution from a PZ-stack measured experimentally. We use a sparse sample of fluorescent nanobeads with a diameter of 20 nm (orange carboxylate-modified FluoSpheres), immobilized on the surface of a poly-L-lysine-coated coverslip, and embedded in water. The sample is mounted on a XYZ piezo stage (Physik Instrumente) and is excited by a continuous wave laser emitting at 561 nm (Oxxius L4Cc) in a wide-field illumination configuration using an oil immersion objective lens (APO TIRF $\times 100$, NA = 1.49, Nikon). The emitted

fluorescence is collected by the same objective lens and then passes through two multiband dichroics (Semrock, R405/488/561/635-t1-25x36) and a fluorescence filter (Semrock, 605/40). A telescopic relay system (composed of two achromatic doublets with $f = 250$ mm so that the magnification is unity) is used to access the BFP of the objective. The different polarization projections are taken by using a QWP (AQWP05M-600, Thorlabs) placed on a rotating mount (Newport, PR50CC), followed by a quartz Wollaston polarizing 2.2° beamsplitter (Edmunds, 68-820). The final images are measured using an ORCA Fusion-Digital CMOS C14440-20 UP (1024×1024 pixels, $6.5 \times 6.5 \mu\text{m}^2$ pixel size, Hamamatsu). PZ-stacks were acquired with a step size of 50 nm and a rotation step for the QWP of 30°. Figure 5 shows part of the experimental PZ-stack.

Before launching the retrieval on the experimental PZ-stack, it should be noted that there are several factors that lead to the introduction of a diversity-dependent phase tilt at the BFP. First, the use of a Wollaston prism to spatially separate the two polarization components onto different sections of the camera might make it difficult to have the same center for the PSFs for each polarization component. Second, any slight wedge on the rotating QWP introduces a tilt that rotates with it. Finally, any slight misalignment between the stage moving the sample and the optical axis defined by the microscope objective leads to a defocus-dependent tilt. Therefore, it is best to introduce in the forward model outlined in Sec. IV extra optimization parameters to independently adjust these tilts at the BFP for each combination of diversities. In particular, steps 4 and 5 in Algorithm 2 should be reversed in order to apply the polarization

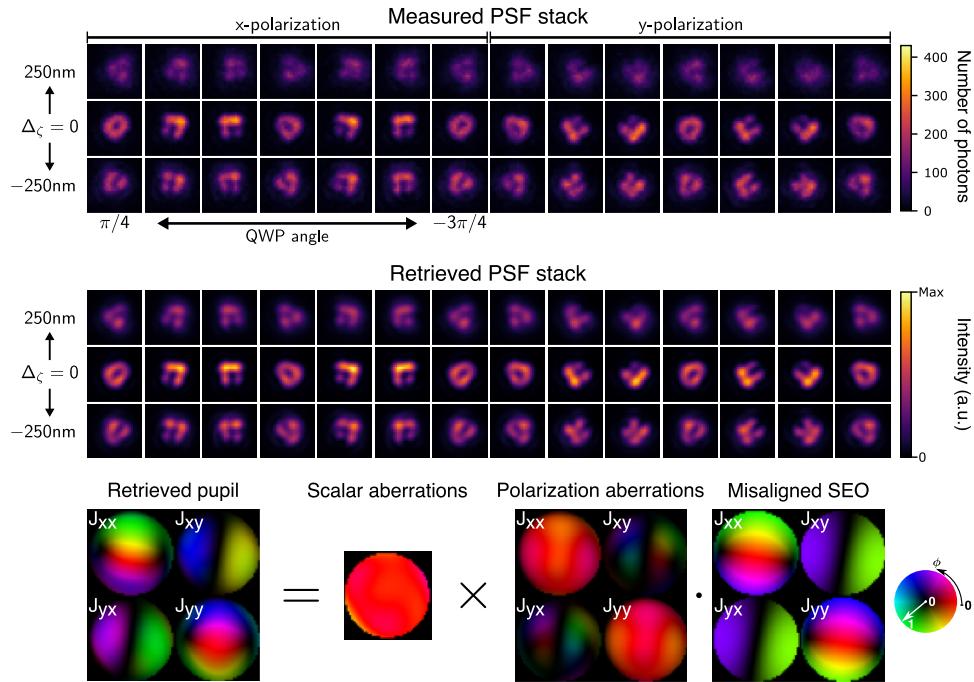


FIG. 5. BDPP retrieval from experimental data, taken from fluorescent nanobeads of diameter 20 nm. (First row) Measured and (second row) retrieved PZ-stacks, where only the PSFs at the initial, middle, and final values of the defocus parameter Δ used for the phase diversity are shown, for all polarization projections comprising the polarization diversity. (Last row) The retrieved elements of the Jones matrix for the BDPP and its decomposition into a misaligned SEO element, and scalar and polarization aberrations.

diversity before propagating to the image plane, and the following additional step should be added after applying the polarization diversity:

4b. Apply a phase tilt $T(\ell) = \exp(i2\pi t^{(\zeta,p)} \cdot \ell)$ to each diversity.

Optimization parameters: $t^{(\zeta,p)} = (t_x^{(\zeta,p)}, t_y^{(\zeta,p)})$.

The downside of including these extra parameters is that the number of FFTs needed for each iteration increases by a factor equal to the number of polarization projection steps (see Table I). However, the total run time remains quite manageable even if the 2D blurring model is used.

All the optimization parameters for the retrieval procedure were used, except for the bead radius (for the blurring), since it can be neglected due to the small bead size ($R_b = 10$ nm). The results of the retrieval process are shown in Fig. 5, where the strong agreement between the measured PZ-stack and the one modeled with the retrieved BDPP (see the supplementary material, Sec. SVIII, for a full comparison) can be appreciated. Moreover, the presence of the SEO is visible in the retrieved BDPP. In this case, since it is known that there is an SEO at the BFP, it is worth trying to separate the total retrieved BDPP into a misaligned ideal SEO and a BDPP containing the scalar and polarization aberrations of the system. It should be noted that scalar tilts and defocus are not shown as part of the scalar aberrations. From this decomposition, it can be seen that the largest birefringent contribution comes from the SEO, but the aberrations are not negligible and must be taken into account

(see the supplementary material, Sec. SVIII, for more details). Note also that the polarization aberrations show larger variations than the scalar ones, which are almost flat, showing the need to consider polarization aberrations when using polarization-dependent systems. Given that for retrieval from experimental data, the ground truth is not known *a priori*, the same retrieval was performed for other beads within the field of view to corroborate the results. The retrieved BDPPs are indeed highly correlated with the one shown in Fig. 5 (see the supplementary material, Sec. SVIII, for full details).

VII. CONCLUSIONS

A methodology and phase retrieval algorithm were presented for the characterization of birefringent distributions at the pupil plane (BDPP) from stacks of PSFs. In particular, it was shown that, for polarization-dependent systems, aberrations should be modeled as a BDPP encoding both scalar and polarization deviations from the ideal system, and that the use of polarization diversity is essential for its proper characterization. The software program PYPSFSTACK created and used for the modeling and retrieval presented in this work is freely available. In particular, the BDPP retrieval implementation based on PYTORCH makes this software flexible and customizable. This is shown in this work by incorporating several optimization parameters apart from those used to describe the BDPP, as well as the blurring models presented in Ref. 35. Even when considering these extra parameters and models, the runtime remains

manageable. In addition, this software can also be used for the retrieval of scalar BDPPs, as shown in the supplementary material in Sec. SVII for a mask known as the tetrapod.⁵⁹

While it was assumed that the sources used for the characterization were unpolarized, the retrieval model and algorithm presented here can be easily adapted to sources that are dipolar, partially polarized, or with a fixed orientation, such as molecules fixed on a surface¹⁴ or in DNA origami.^{60,61} Similarly, this model can also be used to find optimal scalar or BDPPs minimizing a particular combination of the Cramér–Rao bounds for the parameters to be extracted from the shape of the PSFs. Finally, note that there are some modifications that could be implemented in the retrieval algorithm that might allow faster convergence, such as the use of stochastic gradient descent, where only a given set of measurements is used during each iteration or spectral initialization.⁶²

SUPPLEMENTARY MATERIAL

See the supplementary material for the exact expression of the Green tensor, an analysis of the various criteria used to define the position for the best focus, the blurring models, a detailed comparison between the scalar and vector models, the effect of noise and the number of measurements composing the diversity in the performance of the retrieval, examples using a pixel-based model for the retrieval of birefringent masks, the retrieval of a scalar mask, and further corroboration of the retrieval from experimental data.

ACKNOWLEDGMENTS

R.G.-C. acknowledges S. W. Paine and S. M. Popoff for useful discussions. L.A.A.-C. acknowledges M. Sison for the experimental advice. The authors also thank T. G. Brown for supplying the SEO. R.G.-C. acknowledges funding from the Labex WIFI (Grant Nos. ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL*). L.A.A.-C., S.B., and M.A.A. acknowledge funding from Grant No. ANR-21-CE24-0014 and S.B. from Grant No. ANR-20-CE42-0003.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

R. Gutiérrez-Cuevas: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (lead); Validation (equal); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **L. A. Alemán-Castañeda:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (supporting); Supervision (equal); Validation (equal); Writing – review & editing (equal). **I. Herrera:** Conceptualization (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Validation (equal); Writing – review & editing (equal). **S. Brasselet:** Conceptualization (supporting); Funding

acquisition (equal); Supervision (equal); Writing – review & editing (equal). **M. A. Alonso:** Conceptualization (supporting); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Software (supporting); Supervision (equal); Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data underlying the results presented in this paper are available in Ref. 36.

REFERENCES

- Shashkova, S. and M. C. Leake, "Single-molecule fluorescence microscopy review: Shedding new light on old problems," *Biosci. Rep.* **37**, BSR20170031 (2017).
- K. M. Dean and A. E. Palmer, "Advances in fluorescence labeling strategies for dynamic cellular imaging," *Nat. Chem. Biol.* **10**, 512–523 (2014).
- B. Huang, M. Bates, and X. Zhuang, "Super-resolution fluorescence microscopy," *Annu. Rev. Biochem.* **78**, 993–1016 (2009).
- S. Brasselet, "Polarization-resolved nonlinear microscopy: Application to structural molecular and biological imaging," *Adv. Opt. Photonics* **3**, 205 (2011).
- S. Brasselet and M. A. Alonso, "Polarization microscopy: From ensemble structural imaging to single-molecule 3D orientation and localization microscopy," *Optica* **10**, 1486–1510 (2023).
- J. T. Fourkas, "Rapid determination of the three-dimensional orientation of single molecules," *Opt. Lett.* **26**, 211–213 (2001).
- C. A. Valades Cruz, H. A. Shaban, A. Kress, N. Bertaux, S. Monneret, M. Mavrakis, J. Savatier, and S. Brasselet, "Quantitative nanoscale imaging of orientational order in biological filaments by polarized superresolution microscopy," *Proc. Natl. Acad. Sci. U. S. A.* **113**, E820–E828 (2016).
- C. V. Rimoli, C. A. Valades-Cruz, V. Curcio, M. Mavrakis, and S. Brasselet, "4polar-STORM polarized super-resolution imaging of actin filament organization in cells," *Nat. Commun.* **13**, 301 (2022).
- M. Sison, C. A. Valades Cruz, C. S. Senthil Kumar, V. Curcio, L. A. Aleman-Castañeda, M. Mavrakis, and S. Brasselet, "4polar3D SMOLM: Single molecule orientation and localization microscopy using a simple pupil diaphragm and ratio metric polarization splitting" (unpublished).
- M. P. Backlund, M. D. Lew, A. S. Backer, S. J. Sahl, G. Grover, A. Agrawal, R. Piestun, and W. E. Moerner, "Simultaneous, accurate measurement of the 3D position and orientation of single molecules," *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19087–19092 (2012).
- A. S. Backer and W. E. Moerner, "Extending single-molecule microscopy using optical Fourier processing," *J. Phys. Chem. B* **118**, 8313–8329 (2014).
- O. Zhang, W. Zhou, J. Lu, T. Wu, and M. D. Lew, "Resolving the three-dimensional rotational and translational dynamics of single molecules using radially and azimuthally polarized fluorescence," *Nano Lett.* **22**, 1024–1031 (2022).
- V. Curcio, L. A. Aleman-Castañeda, T. G. Brown, S. Brasselet, and M. A. Alonso, "Birefringent Fourier filtering for single molecule coordinate and height super-resolution imaging with dithering and orientation," *Nat. Commun.* **11**, 5307 (2020).
- C. N. Hulleman, R. Ø. Thorsen, E. Kim, C. Dekker, S. Stallinga, and B. Rieger, "Simultaneous orientation and 3D localization microscopy with a vortex point spread function," *Nat. Commun.* **12**, 5934 (2021).
- T. Ding and M. D. Lew, "Single-molecule localization microscopy of 3D orientation and anisotropic wobble using a polarized vortex point spread function," *J. Phys. Chem. B* **125**, 12718–12729 (2021).
- T. Wu, J. Lu, and M. D. Lew, "Dipole-spread-function engineering for simultaneously measuring the 3D orientations and 3D positions of fluorescent molecules," *Optica* **9**, 505–511 (2022).
- J. Enderlein, E. Toprak, and P. R. Selvin, "Polarization effect on position accuracy of fluorophore localization," *Opt. Express* **14**, 8111–8120 (2006).

¹⁸Q. Hu *et al.*, “Arbitrary complex retarders using a sequence of spatial light modulators as the basis for adaptive polarisation compensation,” *J. Opt.* **23**, 065602 (2021).

¹⁹C. He, M. B. C. He, and M. J. Booth, “Enhancing polarisation imaging through novel polarimetry and adaptive optics,” *Proc. SPIE* **11963**, 1196302 (2022).

²⁰B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, “Phase retrieval for high-numerical-aperture optical systems,” *Opt. Lett.* **28**, 801 (2003).

²¹B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, “Phase-retrieved pupil functions in wide-field fluorescence microscopy,” *J. Microsc.* **216**, 32–48 (2004).

²²R. A. Gonsalves, “Phase diversity: Math, methods and prospects, including sequential diversity imaging,” *Proc. SPIE* **10677**, 106771S (2018).

²³P. N. Petrov, Y. Shechtman, and W. E. Moerner, “Measurement-based estimation of global pupil functions in 3D localization microscopy,” *Opt. Express* **25**, 7945–7959 (2017).

²⁴R. W. Gerchberg and W. O. Saxton, “A practical algorithm for the determination of plane from image and diffraction pictures,” *Optik* **35**, 237–246 (1972).

²⁵J. R. Fienup, “Phase retrieval algorithms: A comparison,” *Appl. Opt.* **21**, 2758 (1982).

²⁶N. A. Clark, “Microscope characterization using phase retrieval applied to determine the spatial distribution of membrane-associated proteins in hematocytes,” Ph.D. thesis, University of Rochester, 2012.

²⁷N. Hieu Thao, O. Soloviev, and M. Verhaegen, “Phase retrieval based on the vectorial model of point spread function,” *J. Opt. Soc. Am. A* **37**, 16 (2019).

²⁸B. Ferdman, E. Nehme, L. E. Weiss, R. Orange, O. Alalouf, and Y. Shechtman, “VIPR: Vectorial implementation of phase retrieval for fast and accurate microscopic pixel-wise pupil estimation,” *Opt. Express* **28**, 10179 (2020).

²⁹L. Novotny and B. Hecht, *Principles of Nano-Optics* (Cambridge University Press, 2006).

³⁰D. Axelrod, “Fluorescence excitation and imaging of single molecules near dielectric-coated and bare surfaces: A theoretical study,” *J. Microsc.* **247**, 147–160 (2012).

³¹O. Zhang, J. Lu, T. Ding, and M. D. Lew, “Imaging the three-dimensional orientation and rotational mobility of fluorescent emitters using the tri-spot point spread function,” *Appl. Phys. Lett.* **113**, 031103 (2018).

³²E. Bruggeman, O. Zhang, L.-M. Needham, M. Körbel, S. Daly, M. Cheetham, R. Peters, T. Wu, A. S. Klymchenko, S. J. Davis, E. K. Paluch, D. Klenerman, M. D. Lew, K. O’Holleran, and S. F. Lee, “POLCAM: Instant molecular orientation microscopy for the life sciences,” *bioRxiv* **207.527479**.

³³E. W. Hansen, “Overcoming polarization aberrations in microscopy,” *Proc. SPIE* **0891**, 1 (1988).

³⁴A. Vella and M. A. Alonso, “Poincaré sphere representation for spatially varying birefringence,” *Opt. Lett.* **43**, 379 (2018).

³⁵L. A. Alemán-Castañeda, S. Y.-T. Feng, R. Gutiérrez-Cuevas, I. Herrera, T. G. Brown, S. Brasselet, and M. A. Alonso, “Using fluorescent beads to emulate single fluorophores,” *J. Opt. Soc. Am. A* **39**, C167 (2022).

³⁶Dataset: R. Gutiérrez-Cuevas (2023). “pyPSFstack,” Github. <https://github.com/rodgut190/pyPSFstack>.

³⁷A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “PyTorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), Vol. 32.

³⁸E. H. Hellen and D. Axelrod, “Fluorescence emission at dielectric and metal-film interfaces,” *J. Opt. Soc. Am. B* **4**, 337 (1987).

³⁹D. Axelrod, “Total internal reflection fluorescence microscopy in cell biology,” *Traffic* **2**, 764–774 (2001).

⁴⁰D. Axelrod, “Evanescent excitation and emission in fluorescence microscopy,” *Biophys. J.* **104**, 1401–1409 (2013).

⁴¹B. Richards, E. Wolf, and D. Gabor, “Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system,” *Proc. R. Soc. A* **253**, 358–379 (1959).

⁴²E. Wolf and D. Gabor, “Electromagnetic diffraction in optical systems—I. An integral representation of the image field,” *Proc. R. Soc. A* **253**, 349–357 (1959).

⁴³M. A. Lieb, J. M. Zavislán, and L. Novotny, “Single-molecule orientations determined by direct emission pattern imaging,” *J. Opt. Soc. Am. B* **21**, 1210 (2004).

⁴⁴S. Stallinga, “Compact description of substrate-related aberrations in high numerical-aperture optical disk readout,” *Appl. Opt.* **44**, 849–858 (2005).

⁴⁵A. K. Spilman and T. G. Brown, “Stress birefringent, space-variant wave plates for vortex illumination,” *Appl. Opt.* **46**, 61 (2007).

⁴⁶A. K. Spilman and T. G. Brown, “Stress-induced focal splitting,” *Opt. Express* **15**, 8411 (2007).

⁴⁷L. Marrucci, C. Manzo, and D. Paparo, “Optical spin-to-orbital angular momentum conversion in inhomogeneous anisotropic media,” *Phys. Rev. Lett.* **96**, 163905 (2006).

⁴⁸A. Rubano, F. Cardano, B. Piccirillo, and L. Marrucci, “Q-plate technology: A progress review [invited],” *J. Opt. Soc. Am. B* **36**, D70 (2019).

⁴⁹N. Yamamoto, J. Kye, and H. J. Levinson, “Polarization aberration analysis using Pauli-Zernike representation,” *Proc. SPIE* **6520**, 65200Y (2007).

⁵⁰E. P. Goodwin and J. C. Wyant, *Field Guide to Interferometric Optical Testing* (SPIE, 2006).

⁵¹A. J. E. M. Janssen, “Extended Nijboer-Zernike approach for the computation of optical point-spread functions,” *J. Opt. Soc. Am. A* **19**, 849 (2002).

⁵²J. J. M. Braat, P. Dirksen, A. J. E. M. Janssen, and A. S. van de Nes, “Extended Nijboer-Zernike representation of the vector field in the focal region of an aberrated high-aperture optical system,” *J. Opt. Soc. Am. A* **20**, 2281 (2003).

⁵³J. J. Braat, P. Dirksen, A. J. Janssen, S. van Haver, and A. S. van de Nes, “Extended Nijboer-Zernike approach to aberration and birefringence retrieval in a high-numerical-aperture optical system,” *J. Opt. Soc. Am. A* **22**, 2635 (2005).

⁵⁴A. Aristov, B. Lelandais, E. Rensen, and C. Zimmer, “ZOLA-3D allows flexible 3D localization microscopy over an adjustable axial range,” *Nat. Commun.* **9**, 2409 (2018).

⁵⁵R. G. Paxman, T. J. Schulz, and J. R. Fienup, “Joint estimation of object and aberrations by using phase diversity,” *J. Opt. Soc. Am. A* **9**, 1072 (1992).

⁵⁶D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv:1412.6980* [cs.LG] (2017).

⁵⁷D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Math. Program.* **45**, 503–528 (1989).

⁵⁸A. S. Jurling and J. R. Fienup, “Applications of algorithmic differentiation to phase retrieval algorithms,” *J. Opt. Soc. Am. A* **31**, 1348 (2014).

⁵⁹Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, “Optimal point spread function design for 3D imaging,” *Phys. Rev. Lett.* **113**, 133902 (2014).

⁶⁰A. K. Adamczyk, T. A. P. M. Huijben, M. Sison, A. Di Luca, G. Chiarelli, S. Vanni, S. Brasselet, K. I. Mortensen, F. D. Stefani, M. Pilo-Pais, and G. P. Acuna, “DNA self-assembly of single molecules with deterministic position and orientation,” *ACS Nano* **16**, 16924–16931 (2022).

⁶¹K. Hübner, H. Joshi, A. Aksimentiev, F. D. Stefani, P. Tinnefeld, and G. P. Acuna, “Determining the in-plane orientation and binding mode of single fluorescent dyes in DNA origami structures,” *ACS Nano* **15**, 5109–5117 (2021).

⁶²E. J. Candes, X. Li, and M. Soltanolkotabi, “Phase retrieval via Wirtinger flow: Theory and algorithms,” *IEEE Trans. Inf. Theory* **61**, 1985–2007 (2015).