



HAL
open science

Optimal birefringence distributions for imaging polarimetry

Anthony Vella, Miguel Alonso

► **To cite this version:**

Anthony Vella, Miguel Alonso. Optimal birefringence distributions for imaging polarimetry. *Optics Express*, 2019, 27 (25), pp.36799. 10.1364/OE.27.036799 . hal-02395513

HAL Id: hal-02395513

<https://hal.science/hal-02395513>

Submitted on 5 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal birefringence distributions for imaging polarimetry

ANTHONY VELLA¹ AND MIGUEL A. ALONSO^{1,2,*}

¹The Institute of Optics, University of Rochester, Rochester NY 14627, USA

²Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, UMR 7249, 13397 Marseille, France

*miguel.alonso@fresnel.fr

Abstract: Star test polarimetry is an imaging polarimetry technique in which an element with spatially-varying birefringence is placed in the pupil plane to encode polarization information into the point-spread function (PSF) of an imaging system. In this work, a variational calculation is performed to find the optimal birefringence distribution that effectively encodes polarization information while producing the smallest possible PSF, thus maximizing the resolution for imaging polarimetry. This optimal solution is found to be nearly equivalent to the birefringence distribution that results from a glass window being subjected to three uniformly spaced stress points at its edges, which has been used in previous star test polarimetry setups.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Polarimetry is the measurement of the polarization state of light and/or the polarization properties of materials. Such measurements are usually characterized in terms of the Stokes parameters and the Mueller matrix, respectively, which are directly accessible from measurements of the intensity. Imaging polarimetry, in which the polarization is measured as a function of position, is particularly important in applications ranging from microscopy to remote sensing.

Conventional techniques for Stokes polarimetry require multiple intensity measurements, either through time-sequencing or by splitting different polarization components into several separate detection channels [1, 2]. For example, one common method uses a rotating quarter-wave plate (QWP) followed by a fixed linear polarizer, in which the Stokes parameters are deduced from successive intensity measurements with the QWP oriented at different angles [3]. While these techniques can produce highly accurate measurements, they can be relatively complicated and/or time-consuming, generally involving moving parts or multiple beam paths.

When a short acquisition time is desirable, rapid polarization measurements may be taken using *single-shot polarimetry*, in which the Stokes parameters are estimated from a single intensity measurement. A variety of methods exist for single-shot polarimetry, involving gradient index lenses [4], patterned nanoscale gratings [4], or a split aperture composed of multiple polarizers [5]. One particularly simple method, referred to as *star test polarimetry*, uses a spatially-varying birefringent mask (BM) followed by a uniform polarization analyzer in the pupil plane of an exit-telecentric imaging system. With an appropriately chosen birefringence distribution, the inserted elements can encode full polarization information into the shape of the PSF in the rear focal plane of the lens. This method has been demonstrated experimentally using a stress-engineered optic (SEO), which is a BK7 glass window subjected to stress with trigonal symmetry at its periphery, followed by a circular analyzer [6, 7]. Natural applications for this approach are those in which the object is a sparse set of discrete points, such as in astronomy and confocal microscopy. A recent specific application was the real-time monitoring of the output polarization states of each core within a multicore fiber bundle for applications in medical endoscopes [8]. An extension of this technique is being implemented within the context of superresolution microscopy for determining the 3D position, orientation, and vibration of independent fluorescent molecules [9]. Star test polarimetry can also be applied for imaging

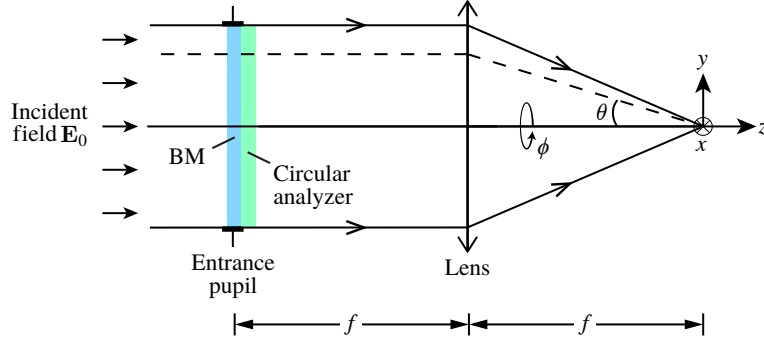


Fig. 1. System layout for a star test polarimetry measurement of an unknown input field \mathbf{E}_0 , illustrated for the case in which a single circular polarization component is imaged. The pupil plane is described by a radial coordinate $u = \sin \theta$ and an azimuthal angle ϕ .

continuous objects if these are discretized by placing a pinhole array in the object plane of an afocal $4f$ relay system with unit magnification [10, 11]. In all these applications, the polarization state of each object point can then be deduced from the corresponding PSF in the image plane, provided that the PSFs from different points do not overlap significantly. The spatial resolution of the measurement is thus limited by the size of the PSF. Encoding polarization information in the PSF necessarily implies increasing its size, since more information is being included in it.

The goal of this work is to find the birefringence distribution of the mask in the pupil plane that maximizes spatial resolution by producing the smallest possible PSF while imposing reasonable restrictions to ensure that the object's polarization is effectively encoded in the shape of the PSF. The optimal solution is found to give very similar results to those of the birefringence distribution of the SEO. The solution's statistical performance (i.e., the expected error in the retrieved Stokes parameters) is assessed by calculating the Fisher information matrix for the measurement.

2. System layout and notation

Consider the system shown in Fig. 1, in which a spatially-varying, transparent, thin birefringent mask with Jones matrix $\mathbb{J}(\mathbf{u})$ is placed at the pupil plane of an exit-telecentric imaging system, where $\mathbf{u} = (u \cos \phi, u \sin \phi)$ is the pupil coordinate normalized such that $u \leq \text{NA}$, with NA being the system's numerical aperture. Collimated light (e.g., from a distant localized source) with uniform polarization \mathbf{E}_0 is then incident on this BM in the plane of an aperture with pupil function $A(\mathbf{u})$ (binary or apodized) and is focused by a lens, producing a polarization-dependent PSF. The input polarization may then be deduced from the shape of the PSF by separating two orthogonal polarization components of the field, denoted by \mathbf{e}_1 and \mathbf{e}_2 , before forming an image of one (or both) of them. For applications where photons are not scarce, only one of the two images is often necessary, so the extraction of the component in question (say, \mathbf{e}_1) can be performed with a polarizer. As will be discussed in Section 9, however, there are applications in which it is best to use all photons, so instead the two components (\mathbf{e}_1 and \mathbf{e}_2) are separated by using a polarizing beamsplitter or a Nomarski or Wollaston prism and then they are imaged separately [9].

In what follows, all calculations are performed in the circular polarization basis, so that $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$, corresponding to right-hand circular (RHC) and left-hand circular (LHC) polarizations, respectively. This choice facilitates the comparison with the SEO (which is used with a circular analyzer) in Section 8. If other output polarization components were analyzed, the results would remain valid after minor modifications. It is assumed that the imaging lens is slow (around 0.05 NA) so that the PSFs are imaged onto several pixels. All calculations then assume the paraxial limit, and angle-dependent polarization effects at the lens can be neglected.

3. Birefringence distribution and point-spread function

We use here the notation introduced in [12] to describe spatially-varying birefringence, where the Jones matrix of the BM in the circular polarization basis may be written as

$$\mathbb{J}(\mathbf{u}) = e^{i\Gamma(\mathbf{u})} \begin{bmatrix} q_0(\mathbf{u}) + iq_3(\mathbf{u}) & -q_2(\mathbf{u}) + iq_1(\mathbf{u}) \\ q_2(\mathbf{u}) + iq_1(\mathbf{u}) & q_0(\mathbf{u}) - iq_3(\mathbf{u}) \end{bmatrix}, \quad (1)$$

with $\Gamma(\mathbf{u})$ being a global phase function, and

$$q_0(\mathbf{u}) = \cos \delta(\mathbf{u}), \quad (2a)$$

$$q_1(\mathbf{u}) = \sin \delta(\mathbf{u}) \cos \Theta(\mathbf{u}) \cos \Phi(\mathbf{u}), \quad (2b)$$

$$q_2(\mathbf{u}) = \sin \delta(\mathbf{u}) \cos \Theta(\mathbf{u}) \sin \Phi(\mathbf{u}), \quad (2c)$$

$$q_3(\mathbf{u}) = \sin \delta(\mathbf{u}) \sin \Theta(\mathbf{u}), \quad (2d)$$

where $\Theta(\mathbf{u}) \in [-\pi/2, \pi/2]$ and $\Phi(\mathbf{u}) \in [0, 2\pi)$ are the latitude and longitude on the Poincaré sphere of the “slow” eigenpolarization at the pupil point \mathbf{u} , and $\delta(\mathbf{u})$ represents half the retardance between the two eigenpolarizations. As discussed in [12], the four quantities in Eq. (2) can be used to construct a four-dimensional unit vector $\vec{q}(\mathbf{u}) = (q_0, q_1, q_2, q_3)$, and $\mathbb{J}(\mathbf{u})$ itself can be considered as a unit quaternion. A particularly simple visualization of the birefringence results from using the 3-vector $\mathbf{q} = (q_1, q_2, q_3)$, which points in the direction joining the two eigenpolarizations in the Poincaré sphere, and whose magnitude encodes the half-retardance δ as $|\mathbf{q}(\mathbf{u})| = \sin \delta(\mathbf{u})$. Because \vec{q} and $-\vec{q}$ correspond to the same effective birefringence, this vector can always be chosen such that $q_0 = (1 - |\mathbf{q}|^2)^{1/2} \geq 0$.

After transmission through the BM, the \mathbf{e}_1 and \mathbf{e}_2 polarization components of the field are given by $\mathbf{e}_1^\dagger \mathbb{J}(\mathbf{u}) \mathbf{E}_0$ and $\mathbf{e}_2^\dagger \mathbb{J}(\mathbf{u}) \mathbf{E}_0$, respectively, where the dagger denotes a conjugate transpose. The paraxial PSF $I^{(j)}(\mathbf{x})$ of each polarization component \mathbf{e}_j is the squared modulus of the Fourier transform of the pupil distribution:

$$I^{(1,2)}(\mathbf{x}) = \left\langle \left| \iint A(\mathbf{u}) \mathbf{e}_{1,2}^\dagger \mathbb{J}(\mathbf{u}) \mathbf{E}_0 \exp[ik(\mathbf{u} \cdot \mathbf{x})] d^2u \right|^2 \right\rangle_{\text{T}}, \quad (3)$$

where $k = 2\pi/\lambda$ is the wavenumber, \mathbf{x} is the two-dimensional spatial coordinate in the image plane, and $\langle \cdot \rangle_{\text{T}}$ denotes a time average in the case of partially polarized light. The phase function $\Gamma(\mathbf{u})$ causes the PSF to increase in size without helping to encode polarization information [12]. Its optimal value is then a constant, assumed here without loss of generality to be zero.

The PSF may be rewritten succinctly by introducing the pupil functions

$$g(\mathbf{u}) = A(\mathbf{u})[q_0(\mathbf{u}) + iq_3(\mathbf{u})], \quad h(\mathbf{u}) = A(\mathbf{u})[-q_2(\mathbf{u}) + iq_1(\mathbf{u})] \quad (4)$$

and their Fourier transforms

$$G(\mathbf{x}) = \iint g(\mathbf{u}) \exp[ik(\mathbf{u} \cdot \mathbf{x})] d^2u, \quad H(\mathbf{x}) = \iint h(\mathbf{u}) \exp[ik(\mathbf{u} \cdot \mathbf{x})] d^2u. \quad (5)$$

Note that $G(\mathbf{x})$ and $H(\mathbf{x})$ represent the output \mathbf{e}_1 field component that results from an RHC or LHC polarized incident beam, respectively. As shown in Appendix A, the PSF of each polarization component may then be expressed in terms of the Stokes parameters S_0, S_1, S_2, S_3 of the incident field \mathbf{E}_0 in the form

$$I^{(1,2)}(\mathbf{x}) = \frac{1}{2} \sum_{n=0}^3 S_n \mathcal{I}_n^{(1,2)}(\mathbf{x}), \quad (6)$$

where the normalized intensity contributions $\mathcal{I}_n^{(1,2)}(\mathbf{x})$ are given by

$$\mathcal{I}_0^{(1,2)}(\mathbf{x}) = |G(\pm\mathbf{x})|^2 + |H(\pm\mathbf{x})|^2, \quad (7a)$$

$$\mathcal{I}_1^{(1,2)}(\mathbf{x}) = \pm 2 \operatorname{Re}\{G^*(\pm\mathbf{x})H(\pm\mathbf{x})\}, \quad (7b)$$

$$\mathcal{I}_2^{(1,2)}(\mathbf{x}) = \pm 2 \operatorname{Im}\{G^*(\pm\mathbf{x})H(\pm\mathbf{x})\}, \quad (7c)$$

$$\mathcal{I}_3^{(1,2)}(\mathbf{x}) = \pm (|G(\pm\mathbf{x})|^2 - |H(\pm\mathbf{x})|^2). \quad (7d)$$

As shown in Appendix B, the Fisher information matrix for a measurement of the normalized Stokes vector $\mathbf{s} = (S_1, S_2, S_3)/S_0$ when \mathcal{N} photons are detected is given by

$$[\mathcal{N}\mathbb{F}(\mathbf{s})]_{mn} = \frac{\mathcal{N}}{1 + \bar{\boldsymbol{\mu}} \cdot \mathbf{s}} \left[\left(\frac{\overline{\mu_m \mu_n}}{1 + \bar{\boldsymbol{\mu}} \cdot \mathbf{s}} \right) - \frac{\bar{\mu}_m \bar{\mu}_n}{1 + \bar{\boldsymbol{\mu}} \cdot \mathbf{s}} \right], \quad (8)$$

where, in what follows, we assume that only the image for the first polarization component (\mathbf{e}_1) is being used, so that $\boldsymbol{\mu}(\mathbf{x})$ is defined by

$$\boldsymbol{\mu}(\mathbf{x}) = \frac{1}{\mathcal{I}_0^{(1)}(\mathbf{x})} \begin{bmatrix} \mathcal{I}_1^{(1)}(\mathbf{x}) \\ \mathcal{I}_2^{(1)}(\mathbf{x}) \\ \mathcal{I}_3^{(1)}(\mathbf{x}) \end{bmatrix} \quad (9)$$

and $\bar{f} = \iint f(\mathbf{x})w(\mathbf{x})d^2x$ represents a weighted integral of a function f with weight $w(\mathbf{x}) = \mathcal{I}_0^{(1)}(\mathbf{x})/\iint \mathcal{I}_0^{(1)}(\mathbf{x})d^2x$. Note that $\boldsymbol{\mu}(\mathbf{x})$ is a unit vector, and therefore $|\bar{\boldsymbol{\mu}}| \leq 1$. The case in which both images are used is discussed in Section 9.

4. Constraints on the BM distribution

As mentioned previously, the primary performance metric for a candidate BM distribution is its impact on the width of the resulting PSF. However, it is necessary to impose some constraints in order to avoid solutions that would be unsuitable for polarimetry. For practical purposes, it is useful for the PSFs corresponding to each polarization component to have the same total power $\Psi^{(j)} = \iint I^{(j)}(\mathbf{x})d^2x$. This makes the signal-to-noise ratio roughly independent of the input polarization, and it ensures that all polarization information encoded by the BM is fully contained within the shape of each intensity distribution rather than the overall power. Consequently, the input polarization can be deduced by analyzing the PSF of a single component, which is guaranteed to contain roughly half of the incident photons when the signal is sufficiently large.

The total power of each polarization component can be found by integrating Eq. (6) over \mathbf{x} and applying Parseval's theorem to each term, leading to

$$\begin{aligned} \Psi^{(1,2)} &= \frac{1}{2} \iint \left[S_0(|g|^2 + |h|^2) \pm 2S_1 \operatorname{Re}\{g^*h\} \pm 2S_2 \operatorname{Im}\{g^*h\} \pm S_3(|g|^2 - |h|^2) \right] d^2u \\ &= \frac{1}{2} S_0 \iint A^2 (1 \pm \boldsymbol{\beta} \cdot \mathbf{s}) d^2u, \end{aligned} \quad (10)$$

where $\boldsymbol{\beta}(\mathbf{u}) = (-2q_0q_2 + 2q_1q_3, 2q_0q_1 + 2q_2q_3, q_0^2 - q_1^2 - q_2^2 + q_3^2)$ is a unit vector. Then $\Psi^{(1)} = \Psi^{(2)} = \frac{1}{2} S_0 \int A^2 d^2u$ for any arbitrary input polarization if $\langle \boldsymbol{\beta} \rangle_A \equiv \iint A^2(\mathbf{u})\boldsymbol{\beta}(\mathbf{u})d^2u = \mathbf{0}$. Note that this condition is not satisfied when \vec{q} is constant, ruling out the possibility of a uniform BM. From Parseval's theorem, it is easy to see that $\bar{\boldsymbol{\mu}} = \langle \boldsymbol{\beta} \rangle_A / \Psi^{(1)}$. Therefore, the condition $\langle \boldsymbol{\beta} \rangle_A = \mathbf{0}$ implies that $\bar{\boldsymbol{\mu}} = \mathbf{0}$, which causes a significant simplification of the Fisher information matrix in Eq. (8), in particular making zero the second term inside the brackets that is guaranteed to be non-positive for the diagonal elements of the matrix.

5. Derivation of differential equation and boundary conditions

When the BM is introduced into the system, the PSF must increase in size since it now contains polarization information. Because the PSF would provide information about four values (the Stokes parameters) rather than one, the width of $I^{(j)}(\mathbf{x})$ can be expected to be roughly twice as large as that of a system with zero (or uniform) birefringence. Of course, there are many measures for the size of the PSF, each of which would give a slightly different result. The metric chosen here is the RMS irradiance width, which allows a simple variational calculation. It is expected that the optimal birefringence distribution for this measure will be nearly optimal for other measures, such as the full width at half maximum (FWHM) or the Strehl ratio.

Generally speaking, the width of the PSF does not strongly depend on the incident polarization state since the normalized intensity contributions $\mathcal{I}_n^{(1)}(\mathbf{x})$ ($n = 0, 1, 2, 3$) have similar compositions. Therefore, for simplicity, the performance may be evaluated in terms of the spread of the average PSF over all possible input polarizations, i.e., over all values of S_1 , S_2 , and S_3 within the range $[-S_0, S_0]$. This average produces $I^{(1)}(\mathbf{x}) = \frac{1}{2}S_0\mathcal{I}_0^{(1)}(\mathbf{x})$, which is the PSF for unpolarized light. (Since $\mathcal{I}_0^{(1)}(\mathbf{x}) = \mathcal{I}_0^{(2)}(-\mathbf{x})$, the analysis is also valid when the PSFs of both polarization components are used, as discussed in Section 9.) As shown in [12], the squared RMS width r^2 of the PSF for unpolarized light is

$$r^2 = \frac{\iint |\mathbf{x}|^2 I^{(1)}(\mathbf{x}) d^2x}{\iint I^{(1)}(\mathbf{x}) d^2x} = \frac{1}{\kappa} \iint \left(|\nabla A|^2 + A^2 \|\nabla \vec{q}\|^2 \right) d^2u, \quad (11)$$

where ∇ is the gradient with respect to \mathbf{u} , $\|\nabla \vec{q}\|$ denotes the Frobenius norm of the 2×4 matrix $\nabla \vec{q}$, and $\kappa = k^2 \iint A^2 d^2u$. (This result can also be derived from Eqs. (6) and (7a).) The term involving $|\nabla A|^2$ accounts for diffraction, while the second term is the increase in width due to the birefringence distribution of the BM. Notice that if A represents a hard aperture, the RMS width is not well-defined since ∇A diverges at the edge of the pupil. However, the increment

$$\Delta r^2 = \frac{1}{\kappa} \iint A^2 \|\nabla \vec{q}\|^2 d^2u = \frac{1}{\kappa} \sum_{n=0}^3 \iint A^2 \nabla q_n \cdot \nabla q_n d^2u \quad (12)$$

caused by the BM can be well-defined and finite even for a hard-aperture pupil; it is this increment that we seek to minimize. Even for hard apertures, this measure of increase of the PSF width is expected to produce meaningful results, which can be verified by evaluating the FWHM of the solution for comparison.

Since \vec{q} is a unit vector, it contains only three free components for optimization. A natural choice is to use the three-dimensional vector \mathbf{q} and let $q_0 = (1 - |\mathbf{q}|^2)^{1/2}$. To find the solutions we use a variational approach: we consider a linear combination of the PSF's width increase and the three constraints, namely $M = \kappa \Delta r^2 + \boldsymbol{\Lambda} \cdot \langle \boldsymbol{\beta} \rangle_A$, where $\boldsymbol{\Lambda} = (\Lambda_1, \Lambda_2, \Lambda_3)$ is a vector of Lagrange multipliers and the constant prefactor κ was introduced for future convenience. The variational procedure consists of finding the conditions under which M is stationary under infinitesimal changes in \mathbf{q} . For this purpose, consider a variation $\vec{q} \rightarrow \vec{q} + \vec{\epsilon}$, where $\vec{\epsilon} = (\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3)$ is an infinitesimal change. Since after this change the vector must remain a unit vector, $\vec{\epsilon}$ must be perpendicular to \vec{q} , so only three of its components are independent parameters, e.g., the zeroth component can be chosen to be a function of the remaining ones according to

$$\epsilon_0 = -\frac{\boldsymbol{\epsilon} \cdot \mathbf{q}}{q_0}. \quad (13)$$

Let us first analyze the variations of each of the different parts, starting with the PSF width

increase. It can be seen from Eq. (12) that $\vec{q} \rightarrow \vec{q} + \vec{\epsilon}$ causes the change

$$\begin{aligned} \Delta r^2 &\rightarrow \Delta r^2 + \frac{2}{\kappa} \sum_{n=0}^3 \iint A^2 \nabla q_n \cdot \nabla \epsilon_n \, d^2 u \\ &= \Delta r^2 - \frac{2}{\kappa} \sum_{n=0}^3 \iint \epsilon_n \nabla \cdot (A^2 \nabla q_n) \, d^2 u + \left[\frac{2}{\kappa} \int_{\text{edge}} A^2 \vec{\epsilon} \cdot \partial_{\perp} \vec{q} \, du \right], \end{aligned} \quad (14)$$

where we ignored terms quadratic in $\vec{\epsilon}$ and in the second step we removed the derivatives from ϵ_n by using integration by parts. The term in square brackets is an edge contribution that is present in the case of hard apertures, where ∂_{\perp} denotes the derivative normal to the edge, and it should be included only if edges are excluded from the region of integration within the second term (since otherwise this edge contribution is already included in this integral). The corresponding variations in the three constrained quantities $\langle \beta \rangle_A$ give

$$\langle \beta_1 \rangle_A \rightarrow \langle \beta_1 \rangle_A + 2 \iint A^2 (-\epsilon_0 q_2 + \epsilon_1 q_3 - \epsilon_2 q_0 + \epsilon_3 q_1) \, d^2 u, \quad (15a)$$

$$\langle \beta_2 \rangle_A \rightarrow \langle \beta_2 \rangle_A + 2 \iint A^2 (\epsilon_0 q_1 + \epsilon_1 q_0 + \epsilon_2 q_3 + \epsilon_3 q_2) \, d^2 u, \quad (15b)$$

$$\langle \beta_3 \rangle_A \rightarrow \langle \beta_3 \rangle_A + 2 \iint A^2 (\epsilon_0 q_0 - \epsilon_1 q_1 - \epsilon_2 q_2 + \epsilon_3 q_3) \, d^2 u. \quad (15c)$$

In Eqs. (14) and (15), ϵ_0 can be expressed in terms of the remaining ϵ_n by using Eq. (13). The substitution of the resulting variations into M causes a variation of the form $M \rightarrow M + \iint (M_1 \epsilon_1 + M_2 \epsilon_2 + M_3 \epsilon_3) \, d^2 u$. The total variation of M is then insensitive to the infinitesimal changes ϵ_n at any point inside the pupil only if $M_1 = M_2 = M_3 = 0$ everywhere inside the pupil, that is, if the following three constraints hold:

$$q_1 \nabla \cdot (A^2 \nabla q_0) - q_0 \nabla \cdot (A^2 \nabla q_1) + A^2 [\Lambda_1 (q_0 q_3 + q_1 q_2) + \Lambda_2 (q_0^2 - q_1^2) - 2\Lambda_3 q_0 q_1] = 0, \quad (16a)$$

$$q_2 \nabla \cdot (A^2 \nabla q_0) - q_0 \nabla \cdot (A^2 \nabla q_2) + A^2 [\Lambda_1 (q_2^2 - q_0^2) + \Lambda_2 (q_0 q_3 - q_1 q_2) - 2\Lambda_3 q_0 q_2] = 0, \quad (16b)$$

$$q_3 \nabla \cdot (A^2 \nabla q_0) - q_0 \nabla \cdot (A^2 \nabla q_3) + A^2 [\Lambda_1 (q_0 q_1 + q_2 q_3) + \Lambda_2 (q_0 q_2 - q_1 q_3)] = 0. \quad (16c)$$

Note that for hard apertures, these equations are dominated at the edges by the derivatives acting on A^2 , which is discontinuous. However, as mentioned earlier, it is perhaps more convenient to account for variations of M due to the edges by using the edge term in Eq. (14). After following similar steps, it is easy to find three edge constraints that can be written succinctly as

$$q_0 \partial_{\perp} \mathbf{q} \Big|_{\text{edge}} = \mathbf{q} \partial_{\perp} q_0 \Big|_{\text{edge}}. \quad (17)$$

This way, Eqs. (16) apply within the smooth regions of the aperture (even infinitesimally close to the edges) and constitute the differential equations for $\mathbf{q}(\mathbf{u})$ to be solved, while Eq. (17) applies at the edges, providing appropriate boundary conditions.

6. Solution ignoring the boundary conditions

In order to solve these equations, it is illustrative to first solve the simpler problem in which the constraints are ignored in the variational calculation (that is, the Lagrange multipliers are set to zero), as are the boundary conditions. In this case, Eqs. (16a) through (16c) can be written compactly as

$$q_0 \nabla \cdot (A^2 \nabla \mathbf{q}) = \mathbf{q} \nabla \cdot (A^2 \nabla q_0) \quad (18)$$

or as a set of several equalities:

$$\frac{\nabla \cdot (A^2 \nabla q_0)}{q_0} = \frac{\nabla \cdot (A^2 \nabla q_1)}{q_1} = \frac{\nabla \cdot (A^2 \nabla q_2)}{q_2} = \frac{\nabla \cdot (A^2 \nabla q_3)}{q_3}. \quad (19)$$

The latter representation highlights the underlying symmetry between the coordinates of $\vec{q}(\mathbf{u})$ on the Poincaré hypersphere they inhabit [12].

Consider the standard case of a circular hard aperture with pupil function

$$A(\mathbf{u}) = \begin{cases} 1, & u \leq \text{NA}, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

The rotational symmetry of the pupil combined with the symmetries of the differential equations mentioned earlier allow solving the problem through separation of the polar pupil variables u, ϕ . The solution becomes particularly simple if we assume that the BM is not birefringent at the pupil's center, i.e., that $\mathbf{q} = \mathbf{0}$ for $u = 0$. (Note that this choice causes no loss in generality, since the cascading of the resulting BM with plates with uniform birefringence before and/or after it gives access to other solutions.) It is easy to show that a set of solutions to Eq. (18) is given by

$$\mathbf{q}(u, \phi) = \frac{2b_m u^{|m|}}{1 + b_m^2 u^{2|m|}} [\cos(m\phi), \sin(m\phi), 0] \quad (21)$$

for integer m , where b_m is a constant. (A derivation of this result in terms of a stereographic projection of the hyperspherical coordinates \vec{q} is given in [13].) Note that we also made the choice of setting $q_3 = 0$. Again, there is no loss of generality in this choice, because other solutions can be found through cascading with uniform birefringent plates.

The solutions in Eq. (21) turn out to automatically satisfy the constraints $\langle \beta_1 \rangle_A = \langle \beta_2 \rangle_A = 0$. For the third constraint ($\langle \beta_3 \rangle_A = 0$), we calculate

$$\begin{aligned} \frac{\langle \beta_3 \rangle_A}{2\pi \text{NA}^2} &= \frac{1}{\text{NA}^2} \int_0^{\text{NA}} \frac{b_m^4 u^{4|m|} - 6b_m^2 u^{2|m|} + 1}{(1 + b_m^2 u^{2|m|})^2} u \, du \\ &= \int_0^1 \frac{(b_m \text{NA}^{|m|} v^{|m|})^4 - 6(b_m \text{NA}^{|m|} v^{|m|})^2 + 1}{[1 + (b_m \text{NA}^{|m|} v^{|m|})^2]^2} v \, dv, \end{aligned} \quad (22)$$

where we used the change of variables $v = u/\text{NA} \in [0, 1]$ in the last step. The value of this expression is plotted in Fig. 2 as a function of $b_m \text{NA}^{|m|}$ for $|m| = 1, 2, \dots, 10$. Observe that when $1 \leq |m| \leq 4$, there are exactly two values of b_m for which $\langle \beta_3 \rangle_A = 0$. For each of these eight cases, the values of the increment in the RMS width of the PSF can be calculated by using

$$\begin{aligned} \Delta r &= \frac{1}{2\pi} \frac{\lambda}{\text{NA}} \left[\frac{1}{\pi} \sum_{n=0}^3 \int_0^{2\pi} \int_0^1 (\nabla_{\mathbf{v}} q_n \cdot \nabla_{\mathbf{v}} q_n) v \, dv d\phi \right]^{1/2} \\ &= \frac{1}{2\pi} \frac{\lambda}{\text{NA}} \left(2 \int_0^1 \left\{ [\bar{\delta}'(v)]^2 + \frac{m^2}{v^2} \sin^2 \bar{\delta}(v) \right\} v \, dv \right)^{1/2}, \end{aligned} \quad (23)$$

where $\bar{\delta}(v) = \delta(v\text{NA})$. Note from this expression that solutions with higher $|m|$ tend to cause larger increases in the size of the PSF. For the solutions in Eq. (21), $\bar{\delta}(v) = \frac{1}{2} \arctan(b_m \text{NA}^{|m|} v^{|m|})$. The numerical results obtained for each solution are shown in the inset of Fig. 2, where we can see that the solution with $|m| = 1$, $b_1 = 0.625/\text{NA}$ produces the smallest increase in PSF width, which is $\Delta r = 0.239\lambda/\text{NA}$. The resulting half-retardance distribution is then $\bar{\delta}(v) = 2 \arctan(0.625v)$.

While the solutions found above satisfy the constraints $\langle \beta \rangle_A = \mathbf{0}$, the fact that these constraints were ignored at the outset in the variational calculation means that the solutions cannot simultaneously satisfy the constraints and the boundary condition in Eq. (17), which for the hard circular aperture considered here (where the normal derivative ∂_{\perp} reduces to a radial derivative) is given by

$$\bar{\delta}'(1) = 0. \quad (24)$$

These solutions are therefore not optimal, but they provide insights that are useful when considering the true optimal solution.

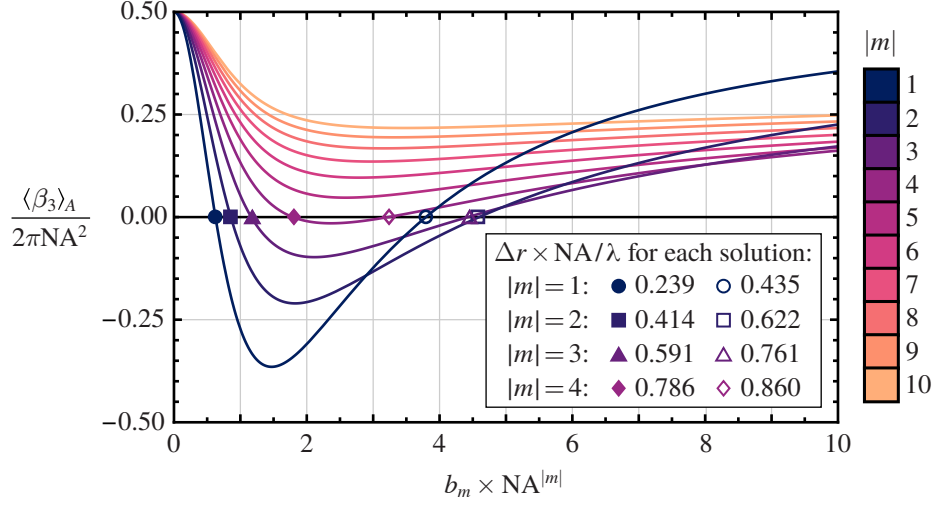


Fig. 2. Value of Eq. (22) as a function of b_m (times a scaling factor) for several values of $|m|$. Note that this quantity can vanish only for $1 \leq |m| \leq 4$. The inset shows the corresponding values of Δr for the eight values of $|m|$ and b_m for which the condition $\langle \beta_3 \rangle_A = 0$ is satisfied.

7. Optimal birefringence distribution

Let us now go back to the general equations (16) while again considering a hard circular pupil. Drawing inspiration from the previous case, we propose a separable solution constrained to the $q_1 q_2$ plane, namely $\mathbf{q}(u, \phi) = \sin \delta(u) [\cos(m\phi), \sin(m\phi), 0]$ with $m \neq 0$, which automatically satisfies the constraints $\langle \beta_1 \rangle_A = \langle \beta_2 \rangle_A = 0$. Notice from Eqs. (2) that this ansatz corresponds to $q_0(u) = \cos \delta(u)$, $\Theta = 0$, and $\Phi = m\phi$. The substitution of this ansatz into Eq. (16c) implies that $\Lambda_1 = \Lambda_2 = 0$, and the resulting form of both Eqs. (16a) and (16b) gives a differential equation for the half-retardance δ , which after changing variables to $v = u/NA$ becomes

$$\delta''(v) + \frac{\delta'(v)}{v} + \left(NA^2 \Lambda_3 - \frac{m^2}{2v^2} \right) \sin[2\delta(v)] = 0 \quad (25)$$

for $v \in [0, 1]$, where the assumption of no birefringence at the pupil center translates into the condition $\delta(0) = 0$, and the boundary condition in Eq. (24) must be satisfied. Additionally, the solution must satisfy the constraint $\langle \beta_3 \rangle_A = 0$, which implies $\int_0^1 \cos[2\delta(v)] v dv = 0$. Equation (25) must be solved numerically subject to these boundary conditions and constraints by choosing appropriately the initial slope and $NA^2 \Lambda_3$. Like for the case considered in the previous section, the optimal solution corresponds to $|m| = 1$. This solution is shown in Fig. 3, and it turns out to be very well approximated by the quadratic expression $1.856v - 0.922v^2$ (the R-squared of the fit being 0.9996). This solution achieves an increase in the PSF width of $\Delta r = 0.220\lambda/NA$, which is indeed slightly smaller than that found in the previous section.

8. Performance evaluation and comparison to SEO

As discussed in the introduction, star test polarimetry is implemented experimentally by using an SEO having three points of symmetric stress. The birefringence distribution near the center of an SEO turns out to be approximately given by the same general form as the previous two solutions, namely $\mathbf{q}(u, \phi) = \sin \delta(u) [\cos(m\phi), \sin(m\phi), 0]$, with $m = -1$ and where in this case the half-retardance is given by $\delta(u) = cu$, with c being a measure of the stress in the SEO [6, 7].

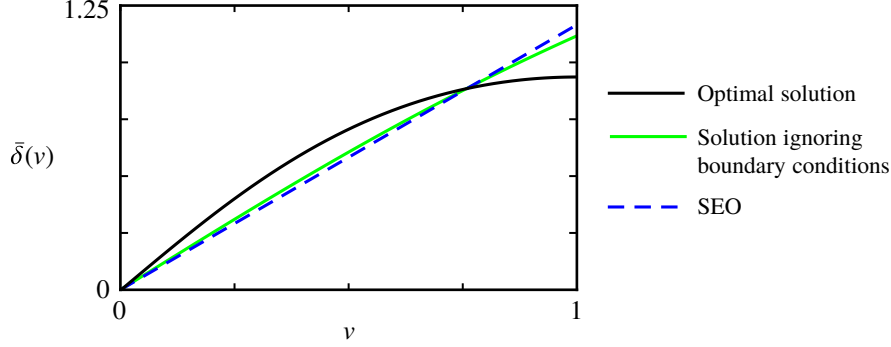


Fig. 3. Radial retardance distribution $\bar{\delta}(v)$ for the optimal BM solution ignoring the boundary conditions, the true optimum, and an SEO with stress coefficient $c = 1.166/\text{NA}$, plotted as functions of the normalized radial pupil coordinate v .

Table 1. Performance metrics for the PSFs resulting from using three BM distributions characterized by the half-retardance functions in the second column, as well as for a diffraction-limited system, when the incident field is unpolarized.

Solution	$\bar{\delta}(v)$	Δr	FWHM	Strehl ratio
Optimal	$\approx 1.856v - 0.922v^2$	$0.220 \lambda/\text{NA}$	$0.764 \lambda/\text{NA}$	0.490
Solution w/o BC's	$\frac{1}{2} \arctan(0.625v)$	$0.239 \lambda/\text{NA}$	$0.792 \lambda/\text{NA}$	0.474
SEO	$1.166v$	$0.249 \lambda/\text{NA}$	$0.798 \lambda/\text{NA}$	0.469
Diffraction limit	0	Not applicable	$0.514 \lambda/\text{NA}$	1.000

Given the similarity of this birefringence distribution to the optimal one, the constraints $\langle \beta_1 \rangle_A = 0$ and $\langle \beta_2 \rangle_A = 0$ are also automatically satisfied. The third constraint reduces to

$$\frac{\langle \beta_3 \rangle_A}{2\pi \text{NA}^2} = \frac{1}{\text{NA}^2} \int_0^{\text{NA}} \cos(2cu)u \, du = \left[\cos(c\text{NA}) - \frac{1}{2} \text{sinc}(c\text{NA}) \right] \text{sinc}(c\text{NA}) = 0, \quad (26)$$

where $\text{sinc}(x) = x^{-1} \sin x$. This equation has infinitely many solutions for c , but the one that leads to the smallest PSF width increase is the first one, $c = 1.166/\text{NA}$, shown in Fig. 3.

Table 1 provides a comparison of the performance metrics of the BM solutions found in the previous two sections and the SEO. The half-retardance distributions are also summarized. All three solutions correspond to $|m| = 1$. (The SEO corresponds specifically to $m = -1$, but the results are the same for $m = 1$, which corresponds to an SEO followed a uniform half-wave plate.) The optimal solution does outperform the two other solutions not only in RMS width increase Δr (calculated through Eq. (23)) for which it was optimized, but also for two other performance metrics: the FWHM and the Strehl ratio. However, the differences in these metrics between the optimal solution and the SEO are of only 5 and 10%. Also provided for comparison are corresponding metrics for a diffraction-limited imaging system where no BM is used (or equivalently where the BM is uniform). We can see that the cost of encoding polarization information into the PSF (on average over all possible polarizations) is an increase in its FWHM of about 60% and a reduction of the Strehl ratio by about 50%.

The functions $G(\mathbf{x})$ and $H(\mathbf{x})$ and the corresponding PSF contributions $\mathcal{I}_n^{(j)}(\mathbf{x})$ for the optimal BM are shown in Fig. 4. Note that the PSF contributions involve negative contributions for $n \geq 1$, but the total measured PSF, which corresponds to the superposition in Eq. (6), is always

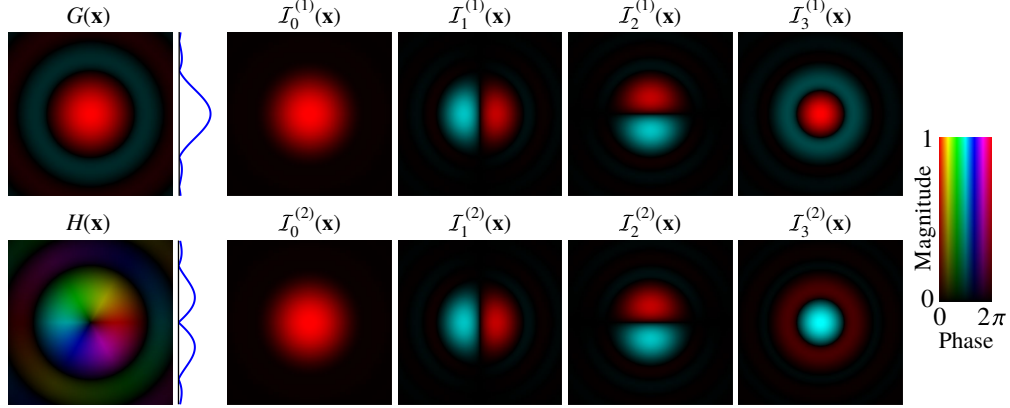


Fig. 4. For the optimal BM solution: complex fields G and H (left, with amplitude cross-sections through the center shown on the right), and PSF contributions $\mathcal{I}_n^{(j)}(\mathbf{x})$ to the PSFs of the \mathbf{e}_1 (top row) and \mathbf{e}_2 (bottom row) output polarization components. The plots are shown over a square region with half-width $1.25\lambda/\text{NA}$.

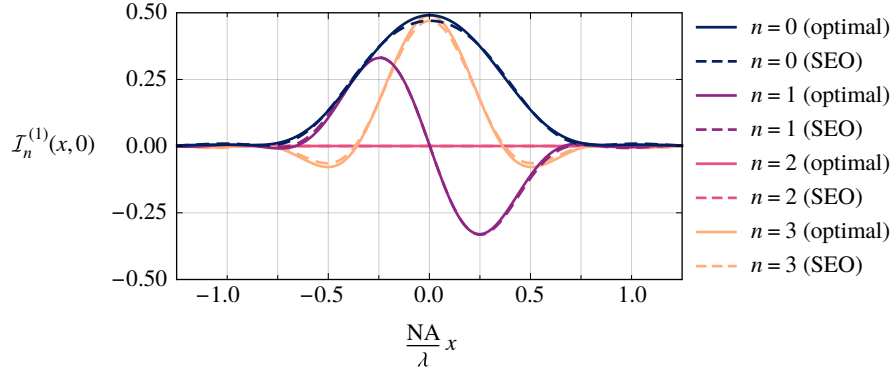


Fig. 5. Horizontal slices through $y = 0$ of the intensity contributions $\mathcal{I}_n^{(1)}$. The solid and dashed curves correspond to the optimal birefringence distribution and an SEO, respectively.

non-negative as a result of the constraint $S_0^2 \geq S_1^2 + S_2^2 + S_3^2$. For the PSF contributions, the top/bottom row corresponds to the case where the right/left circularly polarized component emerging from the BM is focused on the detector. Notice that, for these solutions, $\mathcal{I}_n^{(j)}(\mathbf{x})$ for $n = 0, 1, 2$ are independent of j (the circular component being analyzed) while $\mathcal{I}_3^{(j)}(\mathbf{x})$ changes by a global sign. That is, the symmetries of the functions $G(\mathbf{x})$ and $H(\mathbf{x})$ are such that the only sign change (\pm) that has an effect in Eqs. (7) is that in Eq. (7d). The equivalent plots for an SEO are virtually indistinguishable and are therefore not shown. Instead, to appreciate the very subtle differences, plots of the horizontal cross-sections $\mathcal{I}_n^{(1)}(x, 0)$ for both the optimal BM solution and the SEO are shown in Fig. 5. In practice, the effect of these differences is probably insignificant compared with variations that would arrive, for example, from the pixelization of the detector.

The accuracy of the measurement can be assessed by using the Fisher information matrix in Eq. (8), which given the constraints imposed on the solutions reduces to

$$[\mathcal{NF}(\mathbf{s})]_{mn} = \mathcal{N} \left(\frac{\mu_m \mu_n}{1 + \boldsymbol{\mu} \cdot \mathbf{s}} \right). \quad (27)$$

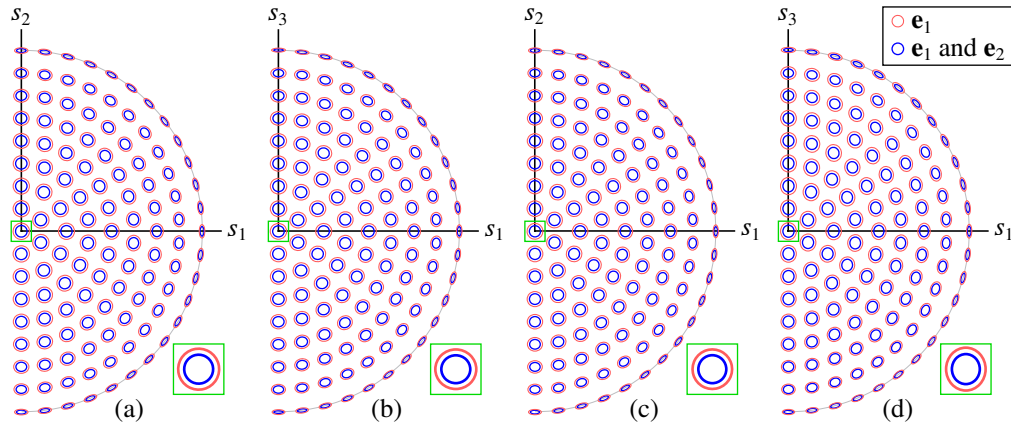


Fig. 6. Two-dimensional cross-sections of the error ellipsoids for incident polarization states in the Poincaré sphere's s_1 - s_2 and s_1 - s_3 cross-sections (shown only for $s_1 \geq 0$ due to symmetry) for (a-b) the optimal BM and (c-d) an SEO, when only the \mathbf{e}_1 output polarization component is measured with $\mathcal{N} = 1500$ detected photons (red) and when both polarization components are measured with $\mathcal{N} = 3000$ detected photons (blue). The error ellipses for unpolarized light are magnified in the inset of each plot.

The inverse of this matrix provides an estimate of the variance in the accuracy of the measurements. That is, for a given incident polarization $\hat{\mathbf{s}}$, the retrieved normalized Stokes vector \mathbf{s} is expected (within one standard deviation) to be within an ellipsoid given by $(\mathbf{s} - \hat{\mathbf{s}}) \cdot [\mathcal{N}\mathbb{F}(\mathbf{s})] \cdot (\mathbf{s} - \hat{\mathbf{s}}) = 1$. Figure 6 shows cross-sections (in red) of these ellipsoids for several incident polarization states, both for the optimal BM and for the SEO, for the case of $\mathcal{N} = 1500$ detected photons. The error ellipsoids are slightly more symmetric with respect to the s_1 - s_2 plane for the optimal BM, but overall the performance is essentially the same. Note that these expected deviations of the normalized Stokes vectors scale as $\mathcal{N}^{-1/2}$, and that the width of the ellipsoids in the plane normal to the cross-sections is comparable to that in the azimuthal direction.

The fact that all solutions described so far have the form $\mathbf{q}(u, \phi) = \sin \delta(u) [\cos(\phi), \pm \sin(\phi), 0]$, where $\delta(u)$ is linear near the center, can be understood in terms of the two circular components of the incident light. Recall that the BM is followed by an RHC polarizer and a Fourier-transforming lens before reaching the detector. Since the birefringence is smallest at the center of the BM and grows away from it, the BM/RHC polarizer combination acts as a rotationally-symmetric apodizer for the incident RHC-polarized light. The incident LHC-polarized component, on the other hand, gets converted into RHC light away from the BM's center, and acquires a phase vortex due to a geometric phase effect. After Fourier transformation by the lens, the field distributions at the detector plane for these components are precisely the distributions G and H (shown in Fig. 5 for the optimal BM), whose forms are a localized central lobe with uniform phase and a slightly larger “donut” with a phase vortex of unit charge at its center, respectively. The radial extent of the latter would increase with $|m|$, the magnitude of the vortex charge written by the BM on H , so $|m| = 1$ guarantees the most compact vortex distribution. The superposition of these two fields then gives access to combinations with all relative phases between them, as well as with all relative amplitudes due to their different radial dependence.

The accuracy of a polarimetric measurement is known to increase with the volume within the Poincaré sphere enclosed by a simplex whose corners correspond to the polarization components being measured [14]. The analysis above tells us that, in theory, each point of the PSF being measured corresponds to the measurement of a different polarization component, and that these cover the complete surface of the Poincaré sphere in a fairly uniform way, so that all the interior

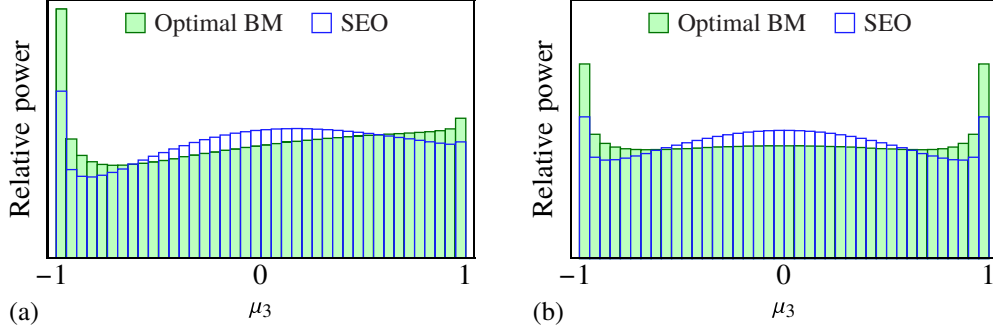


Fig. 7. Histograms of the power coverage of different values of s_3 (characterized by μ_3) over the extent of the PSFs, for the cases when (a) only the PSF for the \mathbf{e}_1 component is used, and (b) the PSFs for both components are used.

of the sphere is enclosed. The uniformity of the coverage of the Poincaré sphere in the s_3 direction following from the variation in the radial direction of the relative weights of the two field components is examined in Fig. 7(a) for both the optimal solution and the SEO. While the uniformity is not perfect, all points are represented in similar amounts and the centroid of the distribution is zero, as guaranteed by the constraint $\overline{\mu_3} = 0$. In practice, of course, the sampling of these different polarization components is compromised by detector pixelation, and even if the pixels were to be made very small, this would come at the cost of each detecting less photons. However, it is clear that this type of measurement provides a more “democratic” coverage of the sphere when compared, say, to a polarimeter where six polarization components are measured.

9. Results when both polarization components are measured

So far we have assumed that only one circular polarization component (\mathbf{e}_1) emerging from the BM is focused to form an image. However, as mentioned earlier, it is possible to instead separately image both circular components and analyze the two resulting PSFs jointly. While doing so requires a more complex system, there are applications in which it is advantageous [9]. One clear advantage is that the number of detected photons is essentially doubled, reducing the uncertainty in the estimation of the Stokes parameters by roughly a factor of $1/\sqrt{2}$. Note, however, that when both images are used, $\overline{\boldsymbol{\mu}}$ vanishes automatically (where the bar now indicates weighted integration over both images) because the total number of detected photons is independent of the polarization state. The Fisher information matrix is then given by the simpler form in Eq. (27) without having to impose the constraints introduced in Section 4. While removing these constraints would clearly affect the variational derivation presented earlier, we now show that the results found earlier remain nearly optimal for the case when both components are imaged.

Consider again the representative case of nearly unpolarized light ($\mathbf{s} \approx \mathbf{0}$), for which the uncertainty in the estimation of the Stokes parameters is largest, and for which the Fisher information matrix in Eq. (27) reduces to \mathcal{N} times $\overline{\boldsymbol{\mu}_m \boldsymbol{\mu}_n}$. Given the forms found for $\mathcal{I}_n^{(j)}(\mathbf{x})$ (antisymmetric in x and symmetric in y for $n = 1$, symmetric in x and antisymmetric in y for $n = 2$, and rotationally symmetric for $n = 3$), the 3×3 matrix with components $\overline{\boldsymbol{\mu}_m \boldsymbol{\mu}_n}$ is automatically diagonal. Further, because $\boldsymbol{\mu}(\mathbf{x})$ is a unit vector, the trace of this matrix is unity. The uncertainty in the measurement is then minimized if the three diagonal elements have equal magnitude, that is, if $\overline{\mu_n^2} = 1/3$ for $n = 1, 2, 3$. It turns out that the optimal BM solution derived in Section 7 nearly achieves this, giving $\overline{\mu_1^2} = \overline{\mu_2^2} = 0.32$ and $\overline{\mu_3^2} = 0.35$, while the SEO gives $\overline{\mu_1^2} = \overline{\mu_2^2} = 0.35$ and $\overline{\mu_3^2} = 0.30$. The cross-sections of the uncertainty ellipsoids for both the

optimal BM solution and the SEO are shown (in blue) in Fig. 6 for the case in which 1500 photons are detected in each PSF, giving a total number of photons of $\mathcal{N} = 3000$. The first thing one notices is that the uncertainties are indeed smaller by a factor of about $1/\sqrt{2}$ with respect to those where only the PSF for the \mathbf{e}_1 component was used (red). Also, the uncertainties are now exactly symmetric in s_3 , as is the coverage of s_3 shown in Fig. 7(b).

Finally, note that the relations $\overline{\mu_m \mu_n} = \overline{\mu_n^2} \delta_{mn}$ and $\overline{\mu_n} = \mathbf{0}$, valid for the optimal and SEO solutions (and whether only \mathbf{e}_1 or both \mathbf{e}_1 and \mathbf{e}_2 components are imaged), imply that the four basic PSFs, $\mathcal{I}_n^{(j)}(\mathbf{x})$, are orthogonal (and almost orthonormal since $\overline{\mu_n^2} \approx 1/3$) under the weight $1/\mathcal{I}_0^{(j)}(\mathbf{x})$. This orthogonality is useful in the retrieval of the parameters from the measured PSFs.

10. Concluding remarks

A variational calculation was performed to determine the spatial birefringence distribution to be placed at the pupil plane of an imaging system that optimizes the encoding of polarization in the PSF's shape while keeping the PSF as small as possible. This optimal birefringence pattern was found to be similar (both in distribution and performance) to that found naturally near the equilibrium points of a transparent window under stress: the mask is not birefringent at its center, but its retardance grows with the radial pupil coordinate u , while the orientation of the slow and fast axes rotates with the azimuthal pupil coordinate ϕ as $\pm\phi/2$. (As mentioned earlier, other solutions with equal performance correspond to birefringence distributions that result from the cascading of these solutions with uniform wave plates.) The only differences were in how the retardance grows with u away from the center, but these cause only small variations in the resulting PSFs. If the variational derivation had been based on a different measure of PSF size and/or requirements for polarimetric optimality (e.g. a measure of the uniformity of μ_3 in Fig. 7), the details of the optimal $\delta(u)$ would change slightly, these variations being comparable to those between any of these optimal solutions and the SEO. That is, the mechanics of stress-induced birefringence naturally provide a nearly optimal birefringence distribution for imaging polarimetry, without the need for nano-fabrication or combinations of spatial light modulators: by applying a pressure distribution with trigonal symmetry to the edges of a glass window, a nearly-optimal birefringence mask (the SEO) results. Note that the SEO was also found to be nearly optimal for generating focused beams with a highly localized zero of intensity [15].

Since the aim of this work was to show that the optimal BM distribution has very similar characteristics and performance to the SEO, we considered the continuous functional form of the PSFs. The performance of these devices in real applications, however, will also depend on the pixelation and general characteristics of the detector, as well as on the algorithms used for the retrieval of the Stokes parameters from the measured PSFs. Preliminary work on this topic can be found in [13]. These aspects are particularly important in applications such as fluorescence microscopy [9], where photons are scarce and the PSFs are used to extract information not only about the two-dimensional polarization studied here, but of the three-dimensional polarization (as well as the three-dimensional position) of the light emitted by a fluorophore. For such applications, the near-optimality of the SEO is of central importance.

Appendix

A. Point-spread function

The polarization-dependent PSF in Section 3 is derived as follows. Consider the pupil functions

$$g_{\pm}(\mathbf{u}) = A(\mathbf{u})[q_0(\mathbf{u}) \pm iq_3(\mathbf{u})], \quad h_{\pm}(\mathbf{u}) = A(\mathbf{u})[-q_2(\mathbf{u}) \pm iq_1(\mathbf{u})] \quad (28)$$

and their Fourier transforms

$$G_{\pm}(\mathbf{x}) = \iint g_{\pm}(\mathbf{u}) \exp[ik(\mathbf{u} \cdot \mathbf{x})] d^2u, \quad H_{\pm}(\mathbf{x}) = \iint h_{\pm}(\mathbf{u}) \exp[ik(\mathbf{u} \cdot \mathbf{x})] d^2u. \quad (29)$$

(Note that the “+” subscripts correspond to the quantities g , h , G , and H defined in Eqs. (4) and (5).) Then Eqs. (1) and (3) lead to the following expression for the PSF:

$$I^{(1,2)}(\mathbf{x}) = \left\langle |G_{\pm}(\mathbf{x})E_{1,2} \pm H_{\pm}(\mathbf{x})E_{2,1}|^2 \right\rangle_{\text{T}}, \quad (30)$$

where E_1 and E_2 are the right- and left-circular polarization components of the input field. This can be expanded to obtain

$$\begin{aligned} I^{(1,2)}(\mathbf{x}) &= \langle |G_{\pm}|^2 |E_{1,2}|^2 + |H_{\pm}|^2 |E_{2,1}|^2 \pm 2 \operatorname{Re}\{G_{\pm}^* H_{\pm} E_{1,2}^* E_{2,1}\} \rangle_{\text{T}} \\ &= |G_{\pm}|^2 \langle |E_{1,2}|^2 \rangle_{\text{T}} + |H_{\pm}|^2 \langle |E_{2,1}|^2 \rangle_{\text{T}} \\ &\quad \pm 2 \operatorname{Re}\{G_{\pm}^* H_{\pm}\} \langle \operatorname{Re}\{E_{1,2}^* E_{2,1}\} \rangle_{\text{T}} \mp 2 \operatorname{Im}\{G_{\pm}^* H_{\pm}\} \langle \operatorname{Im}\{E_{1,2}^* E_{2,1}\} \rangle_{\text{T}} \\ &= \frac{1}{2} \left[|G_{\pm}|^2 (S_0 \pm S_3) + |H_{\pm}|^2 (S_0 \mp S_3) \pm 2 \operatorname{Re}\{G_{\pm}^* H_{\pm}\} S_1 + 2 \operatorname{Im}\{G_{\pm}^* H_{\pm}\} S_2 \right], \end{aligned} \quad (31)$$

where in the last step, the electric field was rewritten in terms of the Stokes parameters of the incident field, given by

$$\begin{aligned} S_0 &= \langle |E_1|^2 \rangle_{\text{T}} + \langle |E_2|^2 \rangle_{\text{T}}, & S_1 &= 2 \langle \operatorname{Re}\{E_2^* E_1\} \rangle_{\text{T}}, \\ S_2 &= 2 \langle \operatorname{Im}\{E_2^* E_1\} \rangle_{\text{T}}, & S_3 &= \langle |E_1|^2 \rangle_{\text{T}} - \langle |E_2|^2 \rangle_{\text{T}}. \end{aligned} \quad (32)$$

Thus, the PSF can be written as

$$I^{(1,2)}(\mathbf{x}) = \frac{1}{2} \sum_{n=0}^3 S_n \mathcal{I}_n^{(1,2)}(\mathbf{x}), \quad (33)$$

where the normalized intensity contributions associated with each Stokes parameter are given by

$$\begin{aligned} \mathcal{I}_0^{(1,2)}(\mathbf{x}) &= |G_{\pm}(\mathbf{x})|^2 + |H_{\pm}(\mathbf{x})|^2, & \mathcal{I}_1^{(1,2)}(\mathbf{x}) &= \pm 2 \operatorname{Re}\{G_{\pm}^*(\mathbf{x})H_{\pm}(\mathbf{x})\}, \\ \mathcal{I}_2^{(1,2)}(\mathbf{x}) &= 2 \operatorname{Im}\{G_{\pm}^*(\mathbf{x})H_{\pm}(\mathbf{x})\}, & \mathcal{I}_3^{(1,2)}(\mathbf{x}) &= \pm (|G_{\pm}(\mathbf{x})|^2 - |H_{\pm}(\mathbf{x})|^2). \end{aligned} \quad (34)$$

These results are simplified by noting that $g_{-}(\mathbf{u}) = g_{+}^*(\mathbf{u})$ and $h_{-}(\mathbf{u}) = h_{+}^*(\mathbf{u})$. From Eq. (29), it follows that $G_{-}(\mathbf{x}) = G_{+}^*(-\mathbf{x})$ and $H_{-}(\mathbf{x}) = H_{+}^*(-\mathbf{x})$, leading to the form in Eq. (7).

B. Fisher information

The Fisher information matrix for a measurement of the normalized Stokes vector \mathbf{s} may be derived using the formalism shown in [16]. The probability density function for a measurement of the continuous PSF is defined as

$$P(\mathbf{x}|\mathbf{s}) = \frac{I^{(1)}(\mathbf{x})}{\iint I^{(1)}(\mathbf{x}) d^2x}. \quad (35)$$

Using Eq. (6), this expression can be written in terms of the normalized Stokes parameters as

$$\frac{\mathcal{I}_0^{(1)} \left(1 + \sum_{n=1}^3 s_n \mathcal{I}_n^{(1)} / \mathcal{I}_0^{(1)} \right)}{\iint \mathcal{I}_0^{(1)} \left(1 + \sum_{n=1}^3 s_n \mathcal{I}_n^{(1)} / \mathcal{I}_0^{(1)} \right) d^2x} = w(\mathbf{x}) \frac{1 + \boldsymbol{\mu}(\mathbf{x}) \cdot \mathbf{s}}{1 + \bar{\boldsymbol{\mu}} \cdot \mathbf{s}}, \quad (36)$$

where the abbreviations in Eq. (9) and the paragraph that follows it were used. The elements of the Fisher information matrix may be derived from the log-likelihood function

$$\ell(\mathbf{s}|\mathbf{x}) = \ln \left[w(\mathbf{x}) \frac{1 + \boldsymbol{\mu}(\mathbf{x}) \cdot \mathbf{s}}{1 + \bar{\boldsymbol{\mu}} \cdot \mathbf{s}} \right]. \quad (37)$$

The first two derivatives of this function with respect to the normalized Stokes parameters are

$$\frac{\partial \ell}{\partial s_n} = \frac{\mu_n}{1 + \boldsymbol{\mu} \cdot \mathbf{s}} - \frac{\overline{\mu_n}}{1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s}}, \quad \frac{\partial^2 \ell}{\partial s_m \partial s_n} = -\frac{\mu_m \mu_n}{(1 + \boldsymbol{\mu} \cdot \mathbf{s})^2} + \frac{\overline{\mu_m} \overline{\mu_n}}{(1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s})^2}. \quad (38)$$

Then the elements of the unit Fisher information matrix for a single-photon measurement are

$$\begin{aligned} [\mathbb{F}(\mathbf{s})]_{mn} &= - \iint \left(\frac{\partial^2}{\partial s_m \partial s_n} \ell(\mathbf{s}|\mathbf{x}) \right) P(\mathbf{x}|\mathbf{s}) d^2 x \\ &= - \iint w(\mathbf{x}) \frac{1 + \boldsymbol{\mu} \cdot \mathbf{s}}{1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s}} \left[-\frac{\mu_m \mu_n}{(1 + \boldsymbol{\mu} \cdot \mathbf{s})^2} + \frac{\overline{\mu_m} \overline{\mu_n}}{(1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s})^2} \right] d^2 x \\ &= \left(\frac{1}{1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s}} \iint \frac{\mu_m \mu_n}{1 + \boldsymbol{\mu} \cdot \mathbf{s}} w(\mathbf{x}) d^2 x \right) - \left(\frac{\overline{\mu_m} \overline{\mu_n}}{(1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s})^3} \iint (1 + \boldsymbol{\mu} \cdot \mathbf{s}) w(\mathbf{x}) d^2 x \right). \end{aligned} \quad (39)$$

The integral in the second term evaluates to $1 + \overline{\boldsymbol{\mu}} \cdot \mathbf{s}$, which cancels with one factor in the denominator. This simplification leads to Eq. (8).

Funding

National Science Foundation (NSF) (PHY-1507278); Excellence Initiative of Aix Marseille University - A*MIDEX, a French “Investissements d’Avenir” programme.

Acknowledgments

The authors would like to thank Thomas G. Brown, Philippe Réfrégier, and Valentine Wasik.

References

1. C. F. LaCasse, R. A. Chipman, and J. S. Tyo, “Band limited data reconstruction in modulated polarimeters,” *Opt. Express* **19**, 14976–14989 (2011).
2. R. Azzam, “Stokes-vector and Mueller-matrix polarimetry,” *J. Opt. Soc. Am. A* **33**, 1396–1408 (2016).
3. E. Collett, *Polarized Light: Fundamentals and Applications* (Marcel Dekker, New York, 1993).
4. J. Chang, N. Zeng, H. He, Y. He, and H. Ma, “Single-shot spatially modulated Stokes polarimeter based on a GRIN lens,” *Opt. Lett.* **39**, 2656–2659 (2014).
5. R. Chipman, “Chapter 22: Polarimetry,” in *Handbook of Optics, Vol. II*, M. Bass, ed. (McGraw-Hill, New York, 1995).
6. R. Ramkhalawon, A. M. Beckley, and T. G. Brown, “Star test polarimetry using stress-engineered optical elements,” in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIX*, vol. 8227 (International Society for Optics and Photonics, 2012), p. 82270Q.
7. R. D. Ramkhalawon, T. G. Brown, and M. A. Alonso, “Imaging the polarization of a light field,” *Opt. Express* **21**, 4106–4115 (2013).
8. S. Sivankutty, E. R. Andresen, G. Bouwmans, T. G. Brown, M. A. Alonso, and H. Rigneault, “Single-shot polarimetry imaging of multicore fiber,” *Opt. Lett.* **41**, 2105–2108 (2016).
9. V. Curcio, 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 (CHIDO),” arXiv preprint arXiv:1907.05828 (2019).
10. B. G. Zimmerman, R. Ramkhalawon, M. Alonso, and T. G. Brown, “Pinhole array implementation of star test polarimetry,” in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXI*, vol. 8949 (International Society for Optics and Photonics, 2014), p. 894912.
11. B. G. Zimmerman and T. G. Brown, “Star test image-sampling polarimeter,” *Opt. Express* **24**, 23154–23161 (2016).
12. A. Vella and M. A. Alonso, “Poincaré sphere representation for spatially varying birefringence,” *Opt. Lett.* **43**, 379–382 (2018).
13. A. Vella, “Description and applications of space-variant polarization states and elements,” Ph.D. thesis, University of Rochester (2018).
14. J. S. Tyo, D. L. Goldstein, D. B. Chenault, and J. A. Shaw, “Review of passive imaging polarimetry for remote sensing applications,” *Appl. Opt.* **45**, 5453–5469 (2006).
15. A. Vella, H. Dourdent, L. Novotny, and M. A. Alonso, “Birefringent masks that are optimal for generating bottle fields,” *Opt. Express* **25**, 9318–9332 (2017).
16. A. Vella, “Tutorial: Maximum likelihood estimation in the context of an optical measurement,” arXiv preprint arXiv:1806.04503 (2018).