Non-iterative aberration retrieval based on the spot shape around focus

Pieter Smid [a], Chung W. See [a], Michael G. Somekh [a,b], Amanda J. Wright [a,*]

[a] *Optics and Photonics Research Group, Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD, UK*
[b] *Nanophotonics Research Centre, Shenzhen Key Laboratory of Micro-Scale Optical Information Technology & Institute of Microscale Optoelectronics, Shenzhen University, Shenzhen, China*

ARTICLE INFO

ABSTRACT

A non-iterative, robust, aberration retrieval method to determine primary aberrations by utilizing the intensity distribution at and around focus is presented. The primary Zernike aberrations (coma, spherical aberration and astigmatism) are retrieved by fitting a set of orthogonal circle functions within the central region of the intensity distribution recorded at 3 different axial planes, typically taken at best focus and either side of focus. Aberration indicators are derived from these fits for each primary aberration and it is shown that these indicators can be used for aberration retrieval. The selected indicators vary almost linearly with the magnitude of aberration up to 0.13λ rms, corresponding to a Strehl ratio of 0.44. In the presence of multiple primary aberrations, the method is found to be reliable for a total rms wavefront deviation below 0.10λ (Strehl ratio of 0.68). This approach is linear and non-iterative and will therefore be beneficial for applications where speed and limiting photon exposure is important such as wavefront correction in biomedical imaging.

1. Introduction

The performance of optical imaging systems can be severely degraded by wavefront aberrations caused by system and sample induced aberrations. System induced aberrations typically result from manufacturing errors and misalignment during assembly that impact performance and require post-design aberration correction. Sample induced aberrations are sample specific (e.g. biological samples) and are often spatially variant [1]. In biological microscopy the penetration depth is often limited [2] because of sample aberrations, which makes it difficult if not impossible to obtain high resolution images when the sample induced aberrations are not corrected. Various aberration correction techniques and optimisation algorithms can be used to improve an optical system [3]. Passive optical elements (e.g. diffractive optical elements, lenses) or active elements (e.g. adaptive optics, deformable membrane mirrors, spatial light modulators) can be used to correct aberrations (both system and sample induced) and restore the performance of an optical imaging system [4]. Passive optical elements, however, are not suited to correcting dynamic and spatially varying aberrations. To fully utilize any imaging technique, it is important to first accurately characterise the system and sample aberrations present so that an appropriate aberration correction method can be implemented.

Information about the phase of a wavefront can be obtained by interferometry [5] or a Shack-Hartmann wavefront sensor [6] for exam-

ple. Also, indirect wavefront sensing methods can be used to measure and correct for aberrations with a wavefront shaping device (e.g. deformable mirror) by optimizing a metric related to image-quality, e.g. intensity [7,8,9]. To measure wavefront aberrations in optical imaging systems one can also make use of phase retrieval methods which recover aberrations from image intensity data recorded around focus. Two commonly used phase retrieval methods use Gerchberg-Saxton (GS) type algorithms and the Extended Nijboer-Zernike theorem (ENZ). The GS was developed to retrieve the phase from intensity measurements at the aperture (or pupil plane) and in the image plane iteratively [10]. The relationship between the electromagnetic field in the aperture and image plane is described by a Fourier transform [11]. The iterative GS algorithm computes forward and inverse Fourier transforms of estimated amplitude and phase distributions in the aperture and image planes until the calculated intensity patterns closely match the measured intensity data. The downside of using a GS algorithm is that it can be slow, it is iterative and is computationally demanding. Moreover, GS is beset with convergence issues and also there can be ambiguities in the results [12]. The ENZ theory is an iterative method that uses complex semi-analytical functions and employs intensity data from several axial planes (more than 3) [13, 14]. The method presented here will accommodate aberrations of 0.13λ rms which means Strehl ratios down to 0.44. Even smaller Strehl ratios can be addressed if the requirement for linearity is relaxed.
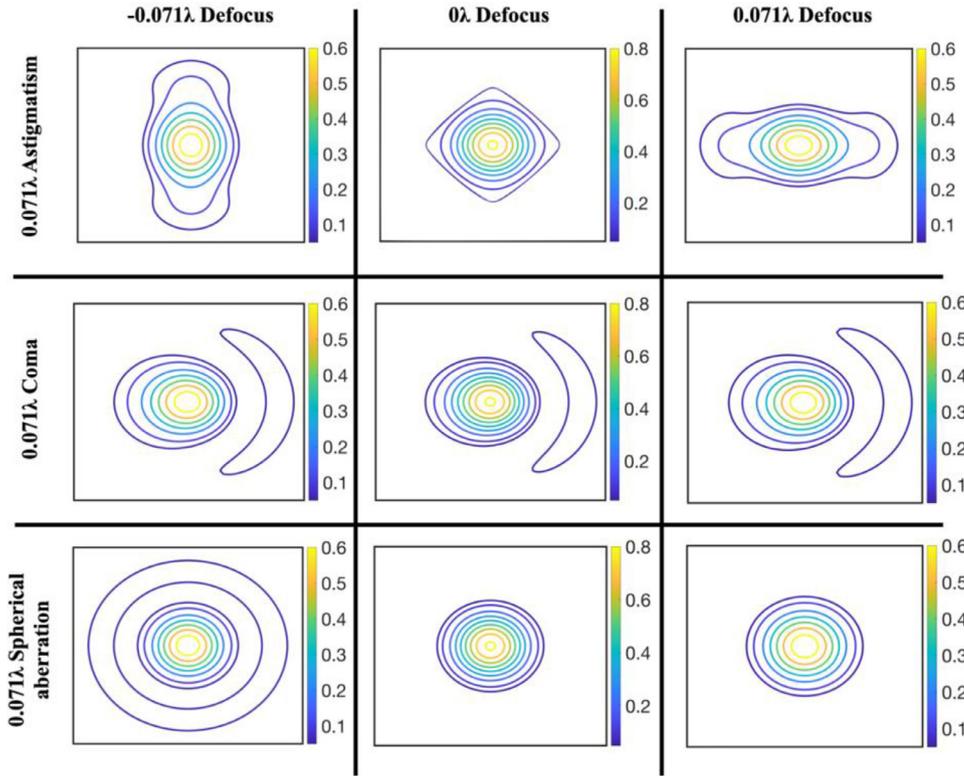
Figure 1. Simulated intensity distributions for three different axial planes, at focus and +/- 0.071λ either side of focus. First row: Astigmatic wavefront through focus. Second row: Coma aberrated wavefront through focus. Third row: wavefront with spherical aberration through focus. Aberration amplitudes are expressed in terms of λ rms.

The new method described here is a non-iterative, fast, and robust approach to aberration retrieval, requiring intensity information from only three axial planes. The principle behind this approach is that primary aberrations produce distinct and characteristic changes to the shape of the intensity distribution of a laser beam around focus in an imaging system. The method presented in this paper shows that by measuring the point spread function (PSF) in three planes it is possible to determine the most commonly encountered primary aberrations (astigmatism, coma and spherical aberration) based on the characteristic distortions of the PSF. Zernike polynomials are fitted to these three intensity distributions to determine the amount of shape change through focus. Specific combinations of the Zernike based fitting coefficients are then used to produce a set of aberration indicators. The aberration indicators are selected to be robust, almost linear and with minimal aberration cross-talk over as large an aberration range as possible. It should be noted that Zernike polynomials are used to evaluate the PSFs because they form a convenient set of orthogonal functions on a unit circle, but they are not used as direct measures of wavefront aberrations. Indeed, the Zernike polynomials are applied to intensity distributions rather than phase distributions as discussed below.

This manuscript describes a fitting procedure for intensity distributions that allows one to calculate aberration indicators for primary coma, astigmatism, and spherical aberration. Furthermore, the influence of defocus effects and the size of the fitting region is studied. The approach is found to work well when fitting to just the central circular region of the images taken near focus, over an area within 1 Airy Unit (AU). Therefore, the technique is less susceptible to poor signal to noise ratios than techniques that rely on information from the side-lobes of a PSF where the intensity is significantly lower. The approach is experimentally validated by looking at the linearity of the aberration indicators when applying known amounts of aberration to the system using a deformable membrane mirror. These experiments are performed for the case of a single aberration. The paper concludes with a set of numerical simulations assessing the capability of the approach when multiple primary aberrations are present in the system.

2. Aberration Retrieval

2.1. Aberrated Intensity distributions through focus

The characteristic intensity distributions at three positions near focus for different primary aberrations are shown in Fig. 1. For Fig. 1, the intensity distributions were calculated, under the Fraunhofer approximation, using Fourier transforms [11]. The analysis of the aberrations occurs in the image plane and therefore this approximation is applicable even for high numerical aperture (NA) objectives provided that the magnification is sufficiently large that NA of the imaging optics is low. In this case different polarization states can be treated independently [15]. To produce Fig 1 normalised optical coordinates were used, where the image plane coordinates (X, Y) were normalised using the diffraction unit (NA/λ) where λ is the wavelength of the spatially coherent light source and NA is the numerical aperture of the objective lens [16], giving (x, y). The pupil plane coordinates (ε, η) were normalised with respect to the pupil radius R, to give (ν, μ). The pupil function, $P(\rho,\theta)$, was described as follows,

$$P(\rho,\theta) = P_0(\rho,\theta)e^{-i2\pi\Phi(\rho,\theta)} \tag{1}$$

where P_0 is the amplitude of the electromagnetic field in the pupil plane, Φ is the phase function in the pupil and (ρ, θ) are cylindrical coordinates in the pupil plane. It was assumed that $P_0=1$ and that the amplitude was constant over the entire pupil. The phase function Φ is expressed in terms of Zernike polynomials,

$$\Phi(\rho,\theta) = \sum_{n,m} \alpha_n^m Z_n^m(\rho,\theta) \tag{2}$$

where α_n^m represents a Zernike amplitude coefficient and Z_n^m a Zernike mode, n and m are the radial and azimuthal orders, respectively. The Malacara normalization [17] was adopted for the Zernike polynomials and an (n, m) indexing scheme was used. The intensity distributions at either side of focus were determined by adding a Zernike defocus term Z_2^0 to the pupil function. Finally, the amplitude in the image plane of
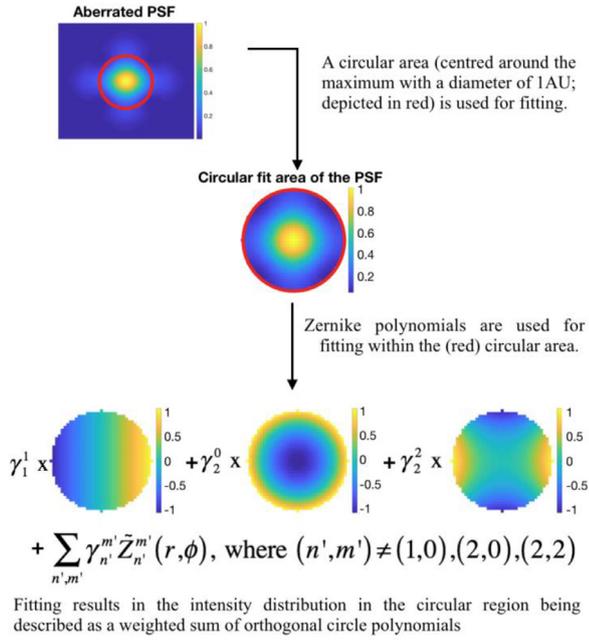
Figure 2. Flow chart of the fitting procedure. Zernike polynomials were used as the fitting polynomials to fit within the (red) circular area of the intensity PSF.

the pupil, U, can be obtained via a 2D Fourier transform of the pupil function,

$$U = \mathfrak{F}^{-1}\{P\} \tag{3}$$

where $\mathfrak{F}^{-1}$ represents the inverse 2D Fourier transform. The intensity distribution, I, is obtained by multiplying U with its complex conjugate.

$$I = U \cdot U^* \tag{4}$$

As can be seen from Fig. 1 primary aberrations generate characteristic focal spot distortion patterns through focus, as discussed by Nijboer in his thesis [18]. For example, in the presence of astigmatism the out of focus intensity distributions are elongated along one direction, with the direction of elongation rotating by 90 degrees when going through focus. For coma, the intensity patterns are similar at either side of focus and have a tail/comet shape oriented in a same direction in all axial planes. For spherical aberration, the intensity patterns before and after focus differ, with one side of focus having a central part that is narrow with a brighter outer intensity ring, whereas on the other side of focus the distribution broadens.

2.2. Fitting procedure of distorted PSFs for aberration retrieval

An aberration free PSF in the focal plane has more than 83% of the intensity concentrated within the first dark ring of the Airy pattern. The diameter of this circular region is often referred to as an Airy Unit. Aberrations broaden the intensity distribution of the PSF at focus. Zernike expressed the impact of aberrations at focus in terms of so-called Jinc functions [19] and it follows from Fig. 2 of his article that primary aberrations manifest themselves mainly within the first dark region of the Airy pattern. Therefore, the distortion of the intensity distribution by primary aberrations should be particularly noticeable within that region. Due to the circular shape of the Airy pattern and the fact that orthogonal functions are well suited for mathematical fitting problems, Zernike polynomials were chosen for fitting to the intensity distributions. The Zernike polynomials used for fitting will be referred to as the fitting polynomials to avoid confusion with the Zernike polynomials used to represent wavefront aberrations present in the pupil plane. For

the fitting procedure, best focus was identified as the point (x, y, z) in the vicinity of geometrical focus with the highest intensity. The three intensity distributions were then centered around best focus, normalized with respect to the highest intensity value in that particular axial plane, and a central circular region with a maximum diameter of 1AU drawn. The normalized, circular, distributions were fitted with the following function,

$$I_{central}(r, \phi) \approx \sum_{n', m'} \gamma_{n'}^{m'} Z_{n'}^{m'}(r, \phi) \tag{5}$$

where $\gamma_{n'}^{m'}$ is the coefficient of a fitting polynomial $Z_{n'}^{m'}$, n' and m' are the radial and azimuthal orders respectively, and r, ϕ the cylindrical coordinates in the image plane, the origin of the coordinate system lies at the point of maximum intensity. Fitting was restricted to polynomials up to radial order 4 (as well as the second order spherical polynomial Z_6^0) and Singular-Value-Decomposition (SVD) was used to minimize the sensitivity to noise. The fitting method is depicted in Fig. 2.

To investigate the change in the fitting coefficients, $\gamma_{n'}^{m'}$, as a beam propagates through focus, a varying defocus term was included in equation 2 and used to produce Fig. 3. The conversion from an axial displacement Z to the Zernike defocus amplitude α_2^0 and to the focal parameter f is given for low NA's (NA < 0.5) by [16],

$$Z = \frac{4\alpha_2^0}{NA^2} = -\frac{\lambda f}{\pi NA^2} \tag{6}$$

Primary astigmatism, coma and spherical aberration were set to a constant value of 0.07λ and each aberration considered separately. Fig. 3 shows the characteristic coefficients for spherical aberration (γ_2^0, γ_4^0), vertical astigmatism (γ_2^2, γ_4^2) and vertical coma (γ_1^1, γ_3^1) as a function of defocus. For oblique astigmatism and horizontal coma, plotting $(\gamma_2^{-2}, \gamma_4^{-2})$ and $(\gamma_1^{-1}, \gamma_3^{-1})$ respectively produced equivalent results to Fig 3a and Fig 3b. These coefficients have been selected because they are sensitive to the characteristic changes in intensity distribution through focus for a particular primary aberration (see Fig 1).

2.3. Derivation of aberration indicators for aberration retrieval

This section shows the rationale for producing a set of aberration indicators based on combinations of the characteristic coefficients that have been shown to vary through focus for each of the primary aberrations (see Fig. 3). The aberration indicators are selected to have minimal crosstalk and to be almost linear over a defocus range of $\pm 0.1\lambda$, with a deviation from linearity of less than 10% for coma and less than 5% for astigmatism and spherical aberration.

2.3.1. Aberration indicator for astigmatism

When imposing vertical (see Fig. 3a) or horizontal astigmatism, the coefficients are odd functions about focus and vary almost linearly over a defocus range of $\pm 0.1\lambda$. The change of sign of these coefficients through focus indicates a rotation of 90° of the intensity patterns. These characteristic coefficients, also vary when coma is present in the system, however these variations are symmetric about focus. To remove this dependency on coma, an astigmatism indicator can be produced by subtracting values of $\gamma_2^{\pm 2}$ measured in two axial planes $(-f, f)$. The vertical astigmatism indicator proposed is then given by,

$$VA_{ind} = (\gamma_2^2)_{f_0+f} - (\gamma_2^2)_{f_0-f} \tag{7}$$

and the oblique astigmatism indicator is given by,

$$OA_{ind} = (\gamma_2^{-2})_{f_0+f} - (\gamma_2^{-2})_{f_0-f} \tag{8}$$

2.3.2. Aberration indicator for coma

In the presence of coma, the coefficients $(\gamma_1^{\pm 1}, \gamma_3^{\pm 1})$ are symmetrical through focus, vary non-linearly (see Fig. 3b) and are insensitive to astigmatism and spherical aberration. The non-linearity, combined with the resulting sign ambiguity (due to the symmetry), makes developing
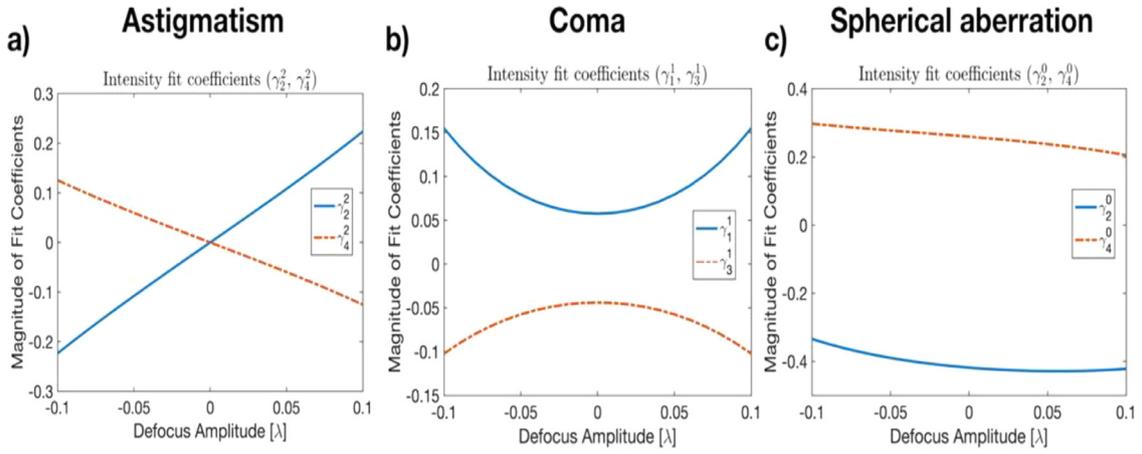
Figure 3. The characteristic coefficients plotted for a fixed amount (0.07λ) of vertical astigmatism, vertical coma and spherical aberration, with varying amounts of defocus α_2^0. The defocus range is $[-0.1\lambda, 0.1\lambda]$. Aberration amplitudes are expressed in terms of λ rms, corresponding to the Z_2^0 Zernike mode and are thus independent of NA.

a coma aberration indicator more challenging. However, if instead of changing focus, the fitting coefficients $(\gamma_1^{\pm 1}, \gamma_3^{\pm 1})$ are plotted at best focus for varying amounts of applied coma, an approximately linear relationship exists between the coma amplitude and the coefficients, as shown in Fig. 4.

A coma aberration indicator, which only requires analysing an intensity distribution at best focus, is proposed as:

$$C_{ind} = \left(\gamma_1^{\pm 1}\right)_{f=0} or \left(\gamma_3^{\pm 1}\right)_{f=0} \tag{9}$$

The change in values of the $\gamma_1^{\pm 1}$ coefficients are larger than the change in values of the $\gamma_3^{\pm 1}$ coefficients (see Fig. 3b) for a fixed change in coma and hence the $\gamma_1^{\pm 1}$ coefficients were chosen for the evaluation of C_{ind}.

2.3.3. Aberration indicator for spherical aberration

For spherical aberration, the γ_4^0 coefficient (Fig. 3c) varies almost linearly over a defocus range of $\pm 0.1\lambda$. The γ_2^0 coefficient is less linear, but it can still be used to retrieve spherical aberration. To remove the dependency on astigmatism and coma, an almost linear spherical aberration indicator can be produced by subtracting the values of γ_2^0 or γ_4^0 measured at two axial planes. Suitable spherical aberration retrieval indicators would be:

$$S_{ind} = \left(\gamma_2^0\right)_{f_0+f} - \left(\gamma_2^0\right)_{f_0-f} or \left(\gamma_4^0\right)_{f_0+f} - \left(\gamma_4^0\right)_{f_0-f} \tag{10}$$

Since the variation in γ_4^0 is more linear than the variation in γ_2^0, γ_4^0 was chosen for the evaluation of S_{ind}. The spherical aberration indicator is sensitive to radial changes in the intensity distributions on opposite sides of focus. In the presence of coma or astigmatism, the changes in (γ_2^0, γ_4^0) are symmetric about best focus and thus cause S_{ind} to be zero (Fig. 4).

Fig. 5 illustrates how the vertical astigmatism and vertical coma indicators would be evaluated based on intensity distributions at $[f_0 - f, f_0 + f]$. The reference plane f_0 should lie in the vicinity of best focus to avoid fitting intensity distributions lying far from best focus where variations in the fitting coefficients tend to become non-linear. The spherical aberration indicator is computed in a similar fashion to the astigmatism indicator (compare equation 10 with 7-8). As will be shown in section 4, the indicators are robust and allow aberration retrieval when more than one primary aberration is present.

Fig. 6 shows how the indicators for vertical astigmatism, spherical aberration and vertical coma vary as a function of the amplitude of the respective aberration. It can be seen that over the plotted aberration amplitude ranges the indicators are approximately linear. It is important
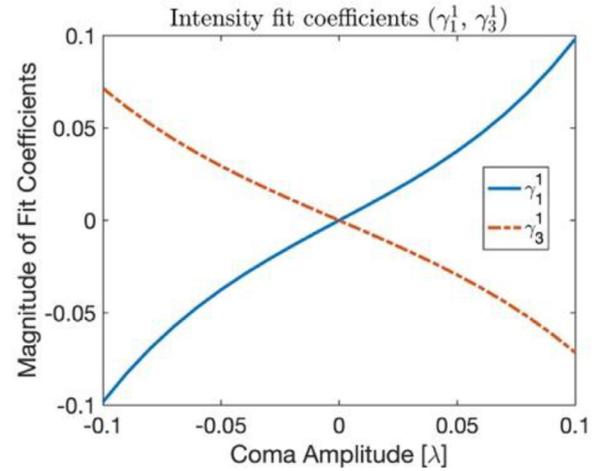
Figure 4. Fitting coefficients (γ_1^1, γ_3^1) for varying amounts of coma at best focus. Aberration amplitudes are expressed in terms of λ rms.

to note that all three aberrations may be extracted with just three measurements. Processing of these intensity distributions allows the three separate aberrations to be determined.

2.4. Influence of defocus on the aberration indicators

In the following section, the defocus or axial range over which the aberration indicators remain linear is discussed in order to determine the optimum distance from focus that the $\pm f$ intensity distribution should be recorded. There is a trade-off between sensitivity and linearity of aberration indicators. The ENZ theory [13,16,20] was used to obtain expressions which describe the intensity distribution for aberrated beams through focus (see Appendix A).

In the case of spherical aberration, if one allows a 10% deviation from the linear response, a linear range is achieved from $+f$ to $-f$ for $f \approx 1.53$ (or $\alpha_2^0 \approx 0.07\lambda$ rms). For an astigmatic wavefront, one gets a close to linear range for VA_{ind} over a defocus range of about $f = \pm 1.16$ (or $\alpha_2^0 = 0.054\lambda$). Since the C_{ind} is applied in the best focal plane, one can estimate the deviation in f from best focus that is allowed by permitting a 10% deviation in the value of C_{ind}. The tolerance for f is then ± 0.63 (or $\alpha_2^0 = 0.029\lambda$).
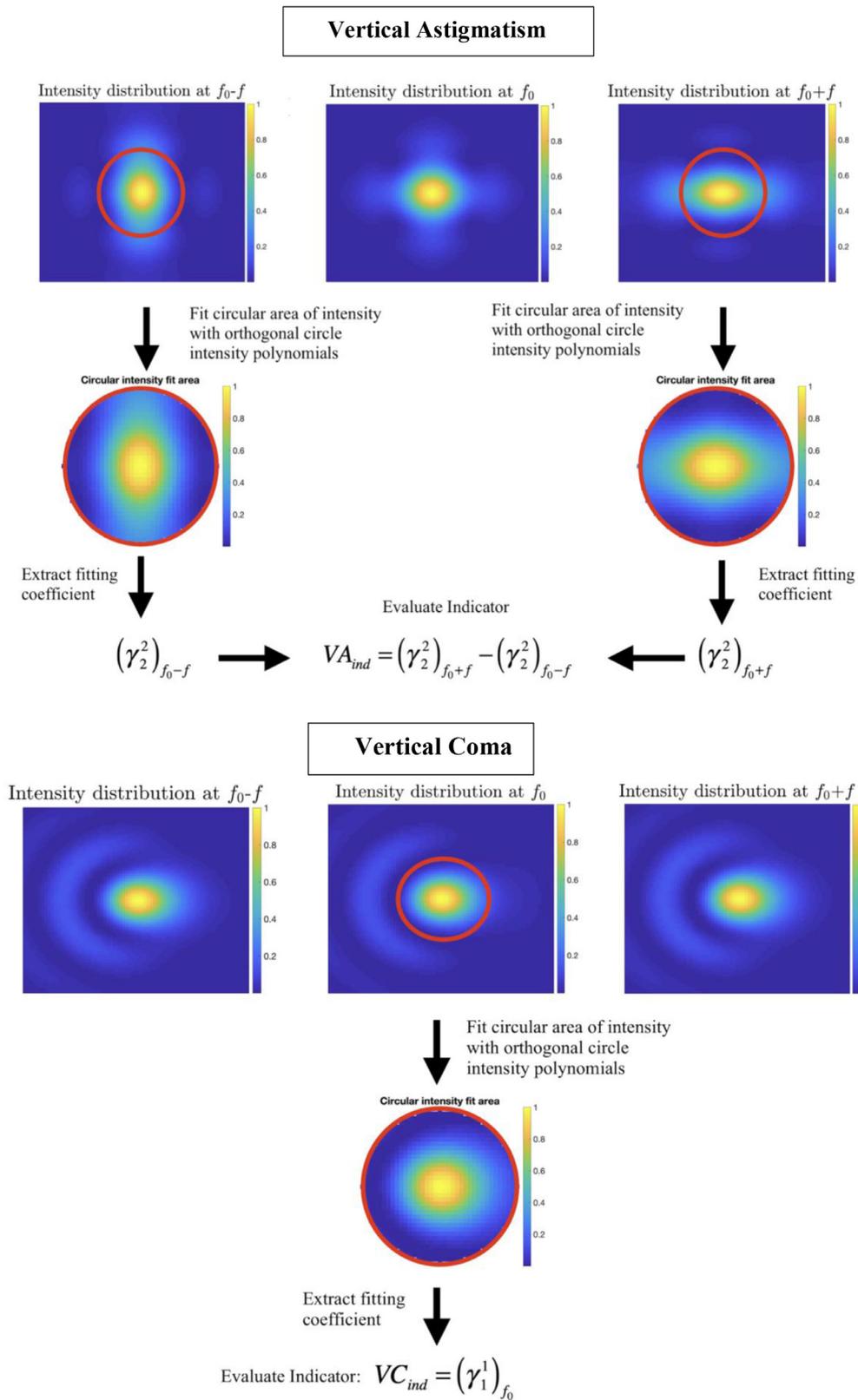
Figure 5. Illustration on how to evaluate the astigmatism and coma aberration indicators based on intensity distributions at different axial planes $[f_0 - f, f_0 + f]$.

Based on this analysis, for astigmatism and spherical aberration the recommended axial plane or defocus position to evaluate the aberration indicators for good sensitivity and linearity is f ∼ ±1.3 (which corresponds to the average found for astigmatism 1.16 and spherical aberration 1.53). For conversion to an axial distance reference is made to equation 6. For green light (wavelength = 532 nm) and an objective with an NA of 0.5 the axial displacement at which the two planes are measured should be no more than ∼ ±881nm or ∼ ±1.66λ. The coma aberration indicator should be evaluated at best focus.
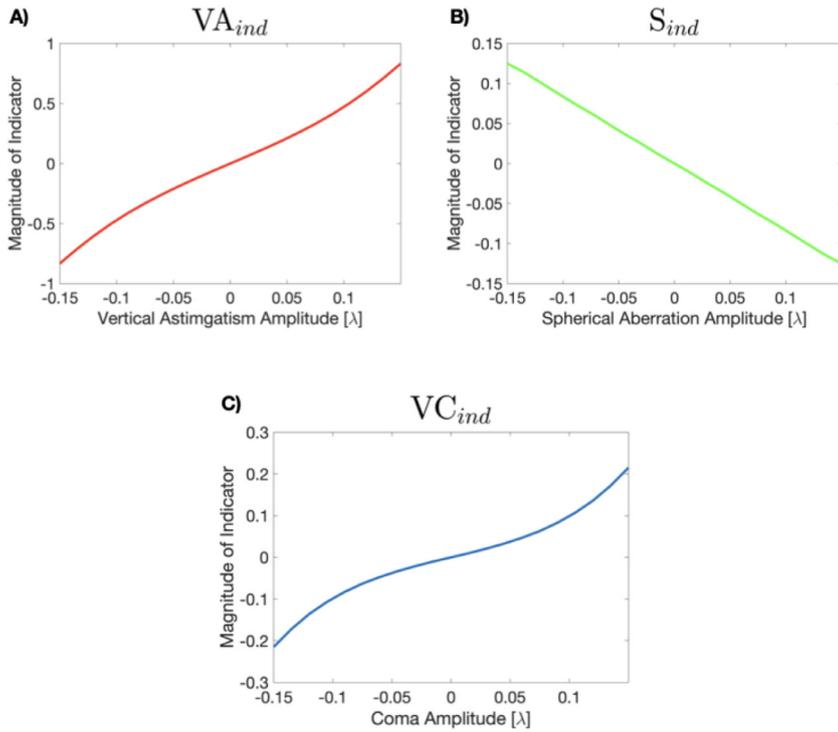
Figure 6. Indicators for A) vertical astigmatism, B) spherical aberration and C) vertical coma as a function of the amplitude of the respective aberration. The astigmatism and spherical aberration were evaluated at axial planes f = ±1.53 (or in terms of $\alpha_2^0 = 0.07\lambda$ rms). Aberration amplitudes are expressed in terms of λ rms.

2.5. Influence of the fit region size

In experiments one might be restricted to fit the intensity distribution using circular masks smaller than 1AU because of the area covered by sensor used for measurement. Reducing the size of the fit area affects the coefficients and hence the aberration indicators. These changes in the fitting coefficients are in fact the same as scaling down Zernike coefficients to smaller pupil sizes, as described by the Dai formula [23], given by,

$$\tilde{\gamma}_n^m = \varepsilon^n \left[\gamma_n^m + (n+1) \sum_{i=1}^{(N-n)/2} \gamma_{n+2i}^m \sum_{j=0}^{i} \frac{(-1)^{i+j}(n+i+j)!}{(n+j+1)!(i-j)!j!} \varepsilon^{2j} \right] \quad (11)$$

where N is the total number of radial orders used for the expansion, and ε is the pupil scaling factor where an ε of 1 is equivalent to 1 A.U. As can be seen from Equation 11, only higher order coefficients (with the azimuthal order m) affect $\tilde{\gamma}_n^m$. In Fig 8, Dai's formula was used to calculate the variations in the characteristic aberration indicators when reducing the radius of the fit region. For astigmatism, coma, and spherical aberration an aberration value of 0.07λ was used (the astigmatism and spherical aberration indicators, Equation 7 and 10, were evaluated at two planes a distance of $\alpha_2^0 = 0.07\lambda$ either side of best focus).

One can conclude that a circular fit region between 0.7-1 AU (represented by 0.7-1ε on Fig. 7) is best suited to retrieve spherical aberration, coma, and astigmatism because the respective indicators have the largest absolute values. Although not discussed here similar arguments can be used to show that the precise illumination profile is not critical in changing the effectiveness of the retrieval process.

3. Experimental Validation

To validate the aberration retrieval method, an optical setup was built. The setup is illustrated in Fig. 8.

Laser light (wavelength of 532 nm, continuous wave), injected in a single mode fiber, provides illumination for the experiments. Known aberrations were applied to the light after it had been collimated by lens L1 using a Mirao 52e (Imagine Optics, France) deformable membrane mirror (DMM). The DMM had previously been calibrated in closed-loop
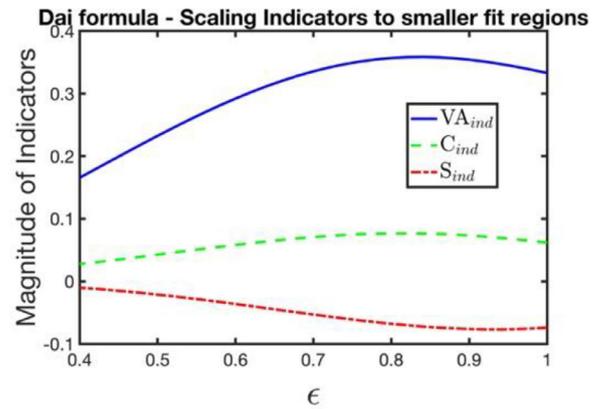
Figure 7. Scaling the astigmatism, coma, and spherical aberration indicators to a smaller pupil fit region. ε is the pupil scaling factor. For astigmatism, coma, and spherical aberration a value of $\alpha_n^m = 0.07\lambda$ was used (the astigmatism and spherical aberration indicator, Equation 7 and 10, were evaluated at two planes a distance of $\alpha_2^0 = 0.07\lambda$ either side of best focus).

mode [24] and the voltage combinations necessary to produce the first 15 Zernike modes were saved and used later for open-loop operation of the DMM. The DMM is conjugated with the pupil of lens L4 (500mm) via a Kepler telescope, L2 (100mm) and L3 (30mm), which reduce the beam diameter from ~12mm at the DMM to ~4mm at L4. The amplitude distribution at L4 was close to uniform (65% of the maximum at the edge) controlled via the adjustable aperture after L1 and we found that the indicators evaluated for a uniform pupil gave good agreement. L4 focuses the laser beam onto the electron multiplying charged coupled camera (EMCCD; iXon 885 Andor). System aberrations caused by lenses L1 to L4 were corrected before running the experiment. Aberrations were generated by the DMM thereafter.

The three primary aberrations of interest (astigmatism, coma, and spherical aberration) were applied independently using the DMM. Their amplitude was varied over a range of ~ [-0.13λ to +0.13λ] in steps
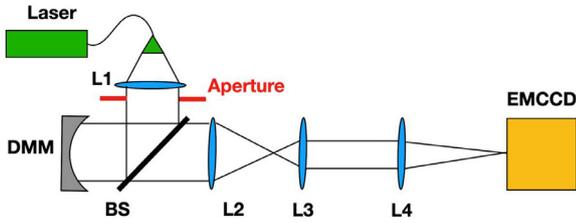
Figure 8. Schematic of the experimental setup used for aberration retrieval. L: Lens; BS: Beamsplitter plate; DMM: Deformable Membrane Mirror; EMCCD: Electron-Multiplying CCD. Showing the aperture used to alter the beam size.

of ~0.027λ. Defocus was also applied using the DMM by changing the value of the Zernike Z_2^0 coefficient. Defocus was varied over a range of ~ [-0.1λ to +0.1λ] in steps of ~0.025λ. The CCD sensor auto-exposure setting was used at each frame so that the maximum intensity for each frame was ~90% of the saturation value and the camera was cooled down to -50° to minimize thermal noise. The diameter of the first dark ring of the focal spot (1 AU) was measured to be ~32 pixels on the camera. The intensity distribution in each axial plane was fitted using the procedure outlined in Fig. 2.

Under small aberration assumptions the position of maximum intensity can be determined by interpolation with a second order 2D polynomial fit around the pixel with the highest intensity in any axial plane. This will be further referred to as the "maximum intensity fit" and the position of maximum intensity will be defined as the new centre of each image. Second order polynomials were used for the maximum intensity

fit, because contributions of higher order terms are small around a maximum (the fit was a 2D area having a width smaller than 0.25AU). For an aberration such as coma, the intensity distribution around a maximum may be better described by including higher order polynomials. However, finding the position of maxima in such circumstances is more complicated and would require more complex algorithms which are beyond the scope of this paper. The parabola should be fitted in a region < ~0.3AU (range where the intensity can be well described with a parabola). To satisfy this relationship and ensure at least three data points, the diameter of the first dark ring of the Airy Disk should be at least 9 pixels. One may interpolate and resample the data if there are less than 9 pixels.

For each individual aberration, the different indicators $VA_{ind}, OA_{ind}, C_{ind}, S_{ind}$ (see Figs. 2 and 5 and equations 7-10) were plotted to show to what extent the characteristic indicator is sensitive to its particular aberration (see Fig 9). The astigmatism and spherical aberration indicators were plotted for different out of focus image pairs to show the sensitivity and linearity behaviour. For coma only one orientation (vertical coma) is plotted (similar results were obtained for the horizontal orientation). For C_{ind} the average value of the coma indicators, measured at the three defocus values $\alpha_2^0 = (-0.025\lambda, 0, 0.025\lambda)$, was taken. Over such a small defocus range, the coma indicator did not change significantly. A circular mask with a diameter of 26 pixels was used for fitting (which corresponds to about 0.81AU). Cross-talk between indicators will be discussed in a later section.

The astigmatism (VA_{ind}, OA_{ind}) and spherical aberration (S_{ind}) indicators show an approximately linear behaviour close to zero and become less linear for larger defocus values (> 0.075λ) and for larger aberration
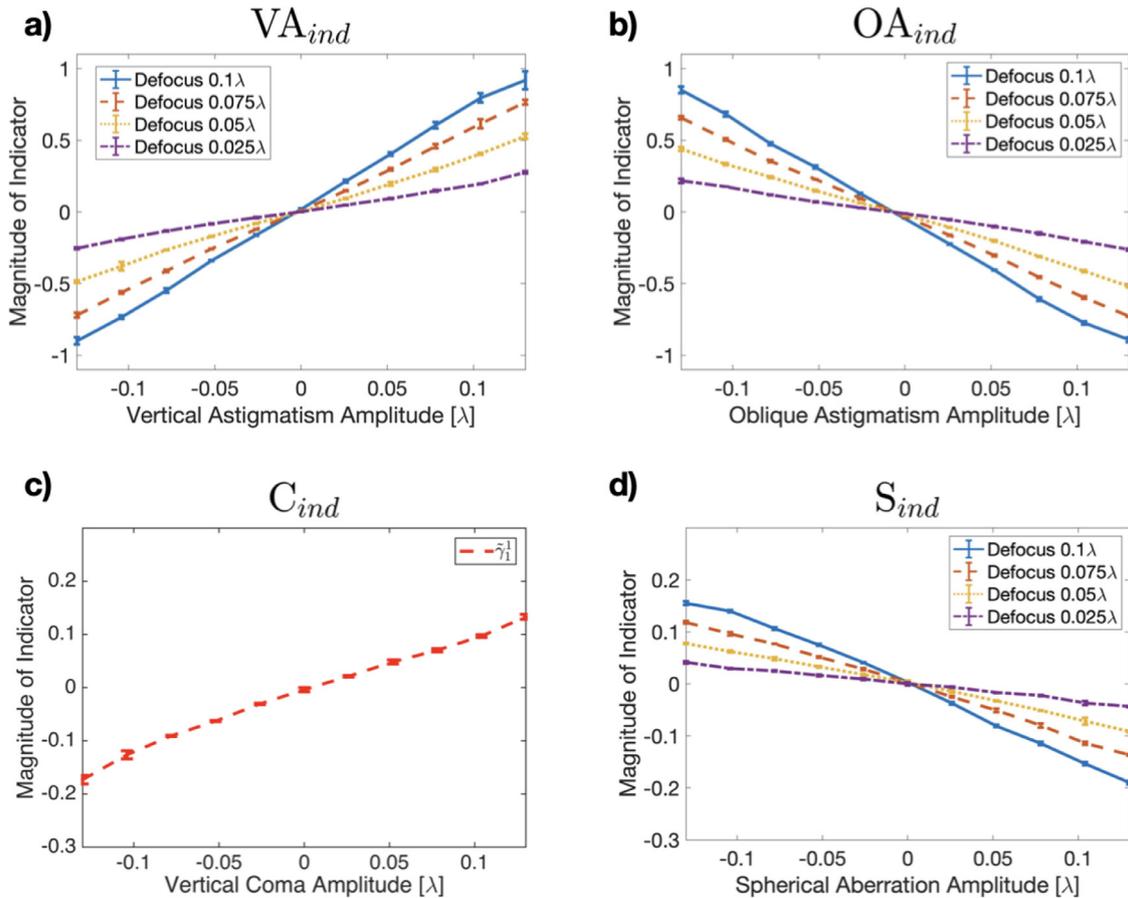
Figure 9. Experimentally determined aberration indicators. The three primary aberrations of interest: astigmatism, coma and spherical aberration, were applied independently from each other with the DMM. Their amplitude was varied over a range of ~ [-0.13λ 0.13λ] in steps of ~0.027λ. Defocus was also applied using the DMM. Defocus was varied over a range of ~ [-0.1λ to 0.1λ] in steps of about ~0.025λ for astigmatism and spherical aberration only. The coma indicator was determined taking from best focus plane only. Aberration amplitudes are expressed in terms of λ rms.
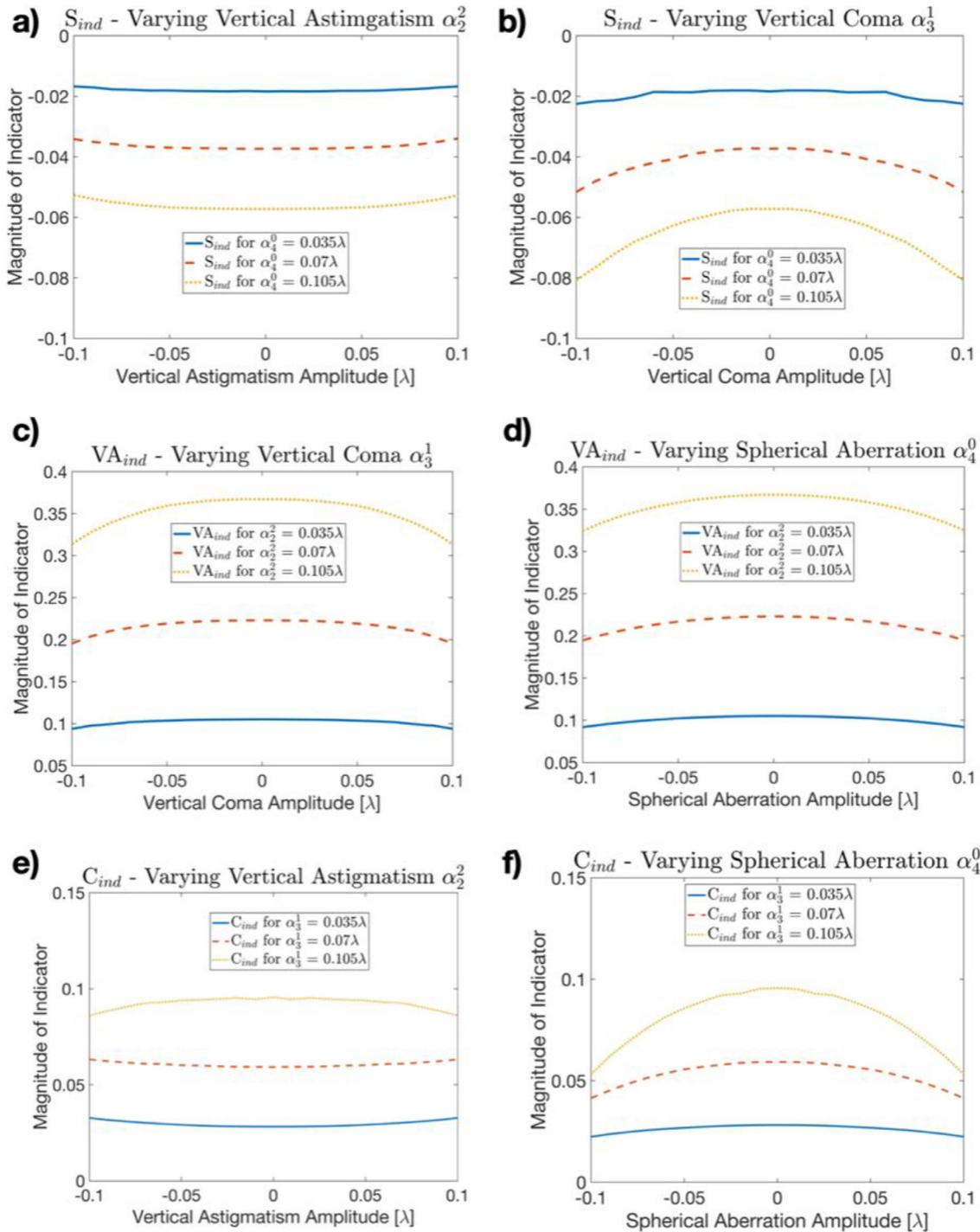
Figure 10. Investigating cross-talk, the influence of additional primary astigmatism, coma and spherical aberration on the 3 aberrations indicators, a-b) spherical aberration indicator, c-d) vertical astigmatism indicator, and e-f) vertical coma indicator. For each indicator the corresponding aberration was set to three different constant values: [0.035λ, 0.07λ, 0.105λ] while varying the other primary aberrations over a range of [-0.1λ, 0.1λ]. Aberration amplitudes are expressed in terms of λ rms.

amplitude ($> 0.13\lambda$) values (see Fig. 9a), b), and d)). To estimate the error in the indicators at larger defocus, a linear curve was fitted through the data points in the aberration amplitude range of [-0.075λ, 0.075λ] at 0.1λ defocus. The equation of the fitted curve ($R^2 > 0.98$) was used to calculate by means of extrapolation the indicator values at $\pm0.125\lambda$ which were then compared with the measured values. The difference (or error) in the indicator values at $\pm0.125\lambda$ is ~15%. The same estimation at a defocus value of 0.05λ results in an error $< 9\%$. Comparable errors are obtained for the coma indicator C_{ind}. In the single aberration

scenario, aberration amplitudes can be retrieved up to 0.13λ with a precision of roughly $\pm0.01\lambda$. It is noted that aberration smaller than 0.03λ barely affect the image quality in an imaging system.

These experimental results are in good agreement with the predications made in section 2.4 where the spherical aberration indicator was seen to have a linear response for $f \approx 1.53$, equivalent to $\alpha_2^0 \approx 0.07\lambda$, and the astigmatism indicator showed a linear response for $f \approx 1.16$ (or $\alpha_2^0 = 0.054\lambda$). To ensure a linear response when implementing this method, it will be important to work within this limit and to have a

fixed value of f for any experiment to achieve consistent and comparable results.

4. Multiple aberrations: Simulations

In optical systems, aberrations can sometimes appear in combination rather than individually. Simulations were carried out to assess the reliability of the derived primary aberration indicators in the presence of multiple primary aberrations.

To compare the FFT simulations with experimental results, the method for finding the position of the maximum intensity of the two-dimensional PSF described in Section 3 was used. For the spherical aberration indicator, for example, the cross-talk with respect to astigmatism (i.e. $\Phi = \alpha_4^0 Z_4^0(\rho, \theta) + \alpha_2^2 Z_2^2(\rho, \theta)$) and coma (i.e. $\Phi = \alpha_4^0 Z_4^0(\rho, \theta) + \alpha_3^1 Z_3^1(\rho, \theta)$) was studied. The value of α_4^0, the coefficient for spherical aberration, is kept constant (at values of $[0.035\lambda, 0.07\lambda, 0.105\lambda]$) while either α_2^2 or α_3^1, the coefficients for astigmatism and coma respectively, were varied within a range of $[-0.1\lambda, 0.1\lambda]$. S_{ind} was evaluated from two planes $\alpha_2^0 = 0.05\lambda$ either side of defocus. Fig. 10 shows the cross-talk simulation results for spherical aberration, coma and astigmatism.

Fig. 10 shows that cross-talk becomes more significant when the amount of additional aberration present increases. There is also more cross-talk between spherical aberration and coma. For primary aberration amplitudes below 0.07λ, the indicators are hardly affected by cross-talk. Beyond 0.07λ, the indicators are considered sufficiently constant when any additional primary aberration is within $\pm0.05\lambda$. The astigmatism indicator is the most robust (Fig. 10 c and d). There is mutual cross-talk between spherical aberration and coma (see Fig. 10 b and f). Based on these simulation results, aberration retrieval in the presence of multiple primary aberrations is reliable (error smaller than 10%) if, as rule of thumb, the total rms wavefront distortion of all primary aberrations is not larger than 0.1λ rms.

5. Conclusion and discussion

This paper presents three new aberration indicators suitable for aberration retrieval of primary aberrations (astigmatism, coma, and spherical aberration). The astigmatism and spherical aberration indicators are determined from intensity distributions recorded at paired defocused planes ($\pm0.05\lambda$ rms defocus), whereas the coma indicator is measured from a single intensity distribution recorded at best focus. The smallest aberration amplitude change that can be reliably detected is $\sim0.01\lambda$ rms, this was measured over a range of 0.03λ rms to 0.13λ rms. In the presence of two primary aberrations the indicators can, as a rule of thumb, be used for aberration retrieval for rms wavefront aberrations smaller than 0.1λ rms. The method can detect multiple primary aberrations if the rms wavefront aberrations is not considerably larger than 0.1λ rms. This approach is suitable for spatially coherent light, however if a broadband source was used, such as a supercontinuum source as opposed to a laser, filters would be required in the detection path to isolate the different wavelengths for analysis.

The aberration retrieval method described, based on fitting Zernike polynomials to intensity distributions, has the advantage of being robust in terms of least-square fitting (making use of orthogonal functions to fit experimental data) and mitigates the adverse effects of noise by limiting the fit to a central relatively high intensity region. The proposed method is relatively simply to implement. It is a non-iterative method, where the aberration indicators vary linearly with their respective aberration and allow primary aberrations in optical imaging systems to be measured from just three images. There is no need to compute large numbers of Fourier transforms as with the Gerchberg-Saxton algorithms, nor is it necessary to use the complex analytical expressions of the ENZ theory for computing aberrated intensity distributions through focus. It is recommended that there are at least 9 pixels within the Airy disk diameter for retrieving primary aberrations with this method. One may interpolate and resample the data were there are less than 9 pixels.

Other linear phase retrieval methods [25] are often limited to very small aberration ranges ($< 0.07\lambda$ rms) that is for systems which are already close to diffraction limited performance. The new linear approach presented here can retrieve aberration amplitudes up to 0.13λ rms. A value of 0.13λ rms corresponds to a Strehl ratio of ~0.44 which would translate to a decrease of the PSF peak intensity by about 56%. A Strehl ratio of 0.44 represents a significant departure from the diffraction limit and would correspond to a deterioration in image quality in any imaging system, non-linear and super-resolution microscopes being particularly susceptible to aberrations. Going forward, this method could be expanded to a larger aberration regime by, for example, providing look-up tables (based on simulations), making use of machine learning algorithms and/or deriving new aberration indicators specifically for the large aberration regime.

The presented aberration retrieval method is well suited for retrieving the most common aberrations encountered in optical imaging systems (such as microscopes and telescopes) or individual lenses. The necessity to perform only three measurements to recover the three primary aberrations renders the method highly efficient in terms of light utilization and speed. It is possible to envisage how this new method could form part of an Adaptive Optics wavefront correction system, where just three images taken close to focus would allow the user to determine the primary aberrations present in their imaging system. In super-resolution microscopes, such as PALM and STORM [26,27], the images of activated fluorescent point sources could be used to retrieve field dependent aberrations whilst imaging. Once retrieved these aberrations could be corrected for using a deformable membrane mirror or spatial light modulator to restore image quality and resolution. Current Adaptive Optics approaches typically rely on a wavefront sensor or an iterative routine for determining the wavefront correction required. Using the approach presented here would remove the need for a wavefront sensor and, most importantly, being iteration-free, would make the correction fast which is advantageous for biomedical imaging where limiting photon dose is paramount.

Declaration of Competing Interest

None.

CRediT authorship contribution statement

Pieter Smid: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Chung W. See:** Conceptualization, Resources, Supervision. **Michael G. Somekh:** Conceptualization, Writing – review & editing, Funding acquisition. **Amanda J. Wright:** Conceptualization, Resources, Supervision, Writing – review & editing, Funding acquisition.

Appendix A. The intensity distribution for aberrated beams through focus

Here we simply describe the most important steps for computing intensity through focus in the presence of a single primary Zernike aberration using spherical aberration as an example. For spherical aberration $\Phi = \alpha_4^0 R_4^0(r) = \alpha_4^0(6r^4 - 6r^2 + 1)$, the Taylor expansion of the complex exponential up to second order is,

$$e^{i\alpha_4^0 R_4^0(r)\cos\varphi} \approx 1 + i\alpha_4^0 R_4^0(r) - \frac{(\alpha_4^0)^2}{2!}A\left\{R_4^0(r)\right\}^2. \tag{A1}$$

The quadratic term can be expanded in terms of the radial Zernike polynomials as follows,

$$\frac{(\alpha_4^0)^2}{2!}\left\{R_4^0(r)\right\}^2 = \frac{(\alpha_4^0)^2}{2!}\left(\frac{1}{5}R_0^0 + \frac{2}{7}R_4^0 + \frac{18}{35}R_8^0\right) \tag{A2}$$

with α_4^0 being the spherical aberration amplitude. Substituting the latter in the diffraction integral while assuming homogeneous illumination of the pupil:

$$U = \frac{1}{\pi}\int_0^1 e^{if\rho^2}\rho\int_0^{2\pi}\left[1 + i\alpha_4^0 R_4^0(r) - \frac{(\alpha_4^0)^2}{2\cdot 2!}\left\{\frac{1}{5}R_0^0 + \frac{2}{7}R_4^0 + \frac{18}{35}R_8^0\right\}\right]$$
$$\times e^{i2\pi\rho r\cos(\phi)}d\theta d\rho \tag{A3}$$

The $R_n^m(r)$ terms in the inner integral can be expressed in terms of a Bessel series [21]. Bessel functions can in turn be expanded into polynomial series [22],

$$J_\upsilon(z) = \left(\frac{z}{2}\right)^\upsilon \sum_{k=0}^\infty \frac{1}{k!\Gamma(k+\upsilon+1)}\left(\frac{-z^2}{2}\right)^k \tag{A4}$$

where Γ is the gamma function. Since we are interested in the intensity distribution within the Airy disk region, the Bessel functions are replaced by their serial expansion up to the order r^8. The difference between the Airy spot intensity and the truncated series representation up to radial order 8 terms is less than 0.005. The last step is re-writing the obtained expression in terms of radial Zernike polynomials. For example, for r^4 we have

$$r^4 = \frac{1}{6}\left(R_4^0 + 3R_2^0 + 2R_0^0\right). \tag{A5}$$

By further substituting $r' = 3.832r$ the radius is expressed in Airy units (AU, i.e. for the first dark ring r' equals 1) and by multiplying the amplitude with its complex conjugate, we obtain the intensity distribution (omitting defocus f terms of power 5 and higher). We now consider the influence of defocus on the aberration indicators for the three primary aberrations.

Spherical aberration:

In the presence of small amounts of spherical aberration and close to best focus, the intensity distribution through focus may be approximated by

$$I_{spherical}\left(r,\phi;f,\alpha_4^0\right) \approx \left\{\begin{array}{c}(0.258 - 0.023\,f^2 + \ldots)+ \\ (-0.017 + (7.28\cdot 10^{-4})\,f^2 - \ldots)f\left(\alpha_4^0\right) \\ + \ldots\end{array}\right\}R_4^0(r) + \ldots \tag{A6}$$

Taking the coefficient γ_4^0 at two axial planes at a distance 2f apart, for example the two planes $(-f, f)$:

$$S_{ind} = (\gamma_4^0)_f - (\gamma_4^0)_{-f} \approx -\left\{0.017(1 - 0.043\,f^2)2f(\alpha_4^0) + \ldots\right\}$$
$$\left\{0.034(1 - 0.043\,f^2)f(\alpha_4^0) + \ldots\right\} \tag{A7}$$

where the $-0.043\,f^2$ term describes the deviation from the linear response.

Astigmatism:

Similar to spherical aberration, an expression for the intensity distribution of an astigmatic beam through focus can be derived using a phase aberration of the form $\Phi = \alpha_2^2 R_2^2(r) = \alpha_2^2 r^2 \cos(2\theta)$ and following the steps explained in Equations A1-A5.

Coma:

For a coma phase aberration $\Phi = \alpha_3^1 R_3^1(r)\cos(\theta) = \alpha_3^1(3r^3 - 2r)\cos(\theta)$ the intensity distribution at best focus is more relevant.

All intensity equations, derived using the ENZ theory, were not normalized with respect to the maximum intensity at each focal plane

(as was done for the FFT simulations, see Fig. 2-6). Normalizing with respect to the maximum intensity is not trivial. The value of maximum intensity in a focal plane depends on the aberration coefficients and on the value of the focal parameter f. For small aberrations however one can determine the maximum intensity value by computing the on-axis intensity $I(0, 0, f, \alpha_n^m)$ in each plane. It was found that normalizing the intensity distributions extends the range over which the indicators can be considered linear. This is because the indicators and the value of maximum intensity both display similar non-linear behaviour when the amount of aberration increases, therefore, to some extent, the non-linearity in the maximum intensity curve compensates for the non-linearity of the indicator for large amounts of aberration. Another advantage of normalising is that one compensates for intensity fluctuation of the source.

References

[1] Schwertner M, Booth MJ, Neil MA, Wilson T. Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry. J Microsc 2004;213(1):11–19.

[2] Girkin J M, Poland S, Wright AJ. Adaptive optics for deeper imaging of biological samples. Curr Opin Biotechnol 2009;20:106–10.

[3] Gross H. Handbook of optical systems: aberration theory and correction of optical systems (Wiley-VCH, 2007). Imaging optics. Braat J, Török P, editors. Cambridge: Cambridge University Press; 2019. doi:10.1017/9781108552264.

[4] Booth MJ. Adaptive optics in microscopy. Phil Trans R Soc A 2007;3652829–43.

[5] Tiziani HJ. Optical methods for precision measurements. Opt Quant Electr 1989;21(4):253–82.

[6] Beverage JL, Shack RV, Descour MR. Measurement of the three-dimensional microscope point spread function using a Shack-Hartmann wavefront sensor. J Microsc 2002;205(Pt 1):61–75.

[7] Neil MA, Booth MJ, Wilson T. Closed-loop aberration correction by use of a modal Zernike wave-front sensor. Opt Lett 2000;25(15):1083–5.

[8] Wright AJ, Burns D, Patterson BA, Poland SP, Valentine GJ, Girkin JM. Exploration of the optimisation algorithms used in the implementation of adaptive optics in confocal and multiphoton microscopy. Microsc Res Tech 2005;67(1):36–44.

[9] Wright AJ, Poland SP, Girkin JM, Freudiger CW, Evans CL, Xie XS. Adaptive optics for enhanced signal in CARS microscopy. Opt Express 2007;15(26):18209–19.

[10] Gerchberg RW, Saxton WO. A practical algorithm for the determination of phase from image and diffraction plane pictures. Optik 1972;35(2):237–50.

[11] Goodman JW. Introduction to fourier optics. Englewood, Co.: Roberts & Company. xviii; 2005. p. 491.

[12] Fienup J. Phase retrieval algorithms: a comparison. Appl Opt 1982;21:2758–69.

[13] Braat JJM, van Haver S, Janssen AJEM, Dirksen P. Chapter 6 Assessment of optical systems by means of point-spread functions. In: Wolf E, editor. Progress in optics. Elsevier; 2008. p. 349–468.

[14] Dirksen P, Braat JJM, Janssen A, Juffermans CAH. Aberration retrieval using the extended Nijboer-Zernike approach. J Micro/Nanolithography, MEMS, and MOEMS 2003;2(1).

[15] Shen M, Chow T, Shen H, Somekh M. Virtual optics and sensing of the retrieved complex field in the back focal plane using a constrained defocus algorithm. Opt Express 2020;28:32777–92.

[16] Braat, J., Dirksen, P., van Haver, S., Janssen, A., Detailed description of the ENZ approach, http://www.nijboerzernike.nl.

[17] Malacara D. Optical shop testing. In: Wiley series in pure and applied optics. Hoboken, N.J.: Wiley-Interscience; 2007. p. 862.

[18] Nijboer BRA. The diffraction theory of optical aberrations: Part II: Diffraction pattern in the presence of small aberrations. Physica 1947;13(10):605–20.

[19] Zernike F, Stratton FJM. Diffraction theory of the Knife-Edge test and its improved form, the Phase-Contrast method. Mon Noti Royal Astro Soc March 1934;94(5):377–84.

[20] Janssen AJ. Extended Nijboer-Zernike approach for the computation of optical point-spread functions. J Opt Soc Am A Opt Image Sci Vis 2002;19(5):849–57.

[21] Nijboer BRA. The diffraction theory of optical aberrations: Part II: Diffraction pattern in the presence of small aberrations. Physica 1947;13(10):605–20.

[22] Abramowitz M. Handbook of mathematical functions, with formulas, graphs, and mathematical tables. Dover Publications, Inc; 1974.

[23] Dai GM. Scaling Zernike expansion coefficients to smaller pupil sizes: a simpler formula. J Opt Soc Am A Opt Image Sci Vis 2006;23(3):539–43.

[24] Polo A, Haber A, Pereira SF, Verhaegen M, Urbach HP. An innovative and efficient method to control the shape of push-pull membrane deformable mirror. Opti expr 2012;20(25):27922–32.

[25] Polo A, Haber A, Pereira SF, Verhaegen M, Urbach HP. Linear phase retrieval for real-time adaptive optics. J European Opti Soc 2013;8 Rapid Publications2013.

[26] Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. Imaging intracellular fluorescent proteins at nanometer resolution. Science 2006.

[27] Dickson R, Cubitt A, Tsien R, et al. On/off blinking and switching behaviour of single molecules of green fluorescent protein. Nature 1997;388:355–8.