

International Conference on Space Optics—ICSO 2018

Chania, Greece

9–12 October 2018

Edited by Zoran Sodnik, Nikos Karafolas, and Bruno Cugny



Image-based wavefront correction for space telescopes

Orestis Kazasidis

Sven Verpoort

Ulrich Wittrock



icso proceedings



International Conference on Space Optics — ICSO 2018, edited by Zoran Sodnik,
Nikos Karafolas, Bruno Cugny, Proc. of SPIE Vol. 11180, 111807Z · © 2018 ESA
and CNES · CCC code: 0277-786X/18/\$18 · doi: 10.1117/12.2536206

Proc. of SPIE Vol. 11180 111807Z-1

Image-based wavefront correction for space telescopes

Orestis Kazasidis*^a, Sven Verpoort^a, Ulrich Wittrock^a

^aMünster University of Applied Sciences, Stegerwaldstrasse 39, 48565 Steinfurt, Germany

ABSTRACT

With a view to future large space telescopes, we investigate image-based wavefront correction with active optics. We use an image-sharpness metric as merit function to evaluate the image quality, and the Zernike modes as control variables. In severely aberrated systems, the Zernike modes are not orthogonal to each other with respect to this merit function. Using wavefront maps, the PSF, and the MTF, we discuss the physical causes for the non-orthogonality of the Zernike modes with respect to the merit function. We show that for combinations of Zernike modes with the same azimuthal order, a flatter wavefront in the central region of the aperture is more important than the RMS wavefront error across the full aperture for achieving a better merit function. The non-orthogonality of the Zernike modes with respect to the merit function should be taken into account when designing the algorithm for image-based wavefront correction, because it may slow down the process or lead to premature convergence.

Keywords: active optics, adaptive optics, sharpness metrics, aberration compensation, algorithm design

1. INTRODUCTION

Space telescopes with 10-m-class primary mirrors are currently being studied for astronomy, in particular for characterization of exoplanets, and for Earth observation from the geostationary orbit. Such telescopes will need to have segmented, lightweight primaries in order to reduce mass and stowed volume. Active optics at the primary mirror and/or in a plane conjugate to the primary mirror will be required to co-phase the segments, align the optical telescope, and correct for manufacturing errors and slow drifts caused by thermo-elastic effects and gravitational release. The method we discuss here requires a continuous surface of the active element, because we use the Zernike modes to describe the active element. Therefore, in the case of a segmented mirror, the segments should be already co-phased by using another technique.

Conventional adaptive optics measures the wavefront and applies its inverse to the corrective active element. Direct wavefront sensing using a dedicated wavefront sensor requires a bright guide star and results in non-common path errors. In addition, the angular separation between the science target and the guide star leads to anisoplanatism. In ground-based telescopes, anisoplanatism is reduced with sophisticated concepts such as multi-conjugate adaptive optics, by using several guide stars, wavefront sensors, and corrective elements. Indirect wavefront sensing uses the science camera and an iterative method, e.g., phase diversity, to retrieve the wavefront. Although phase diversity is technically image-based, here we use the term "image-based" to refer to correction methods that do not require the wavefront information. In this sense, phase diversity is not an image-based method. Our image-based method evaluates the image of the science camera and adapts the surface of the active element to increase a merit function. It can handle large aberrations, in contrast to wavefront sensing that has limited dynamic range. Using an image-based method, active optics would allow image optimization for different objectives, e.g., for maximization of contrast in high or in low spatial frequencies, depending on the science target.

In this paper we discuss the landscape of an image-sharpness metric used as merit function when we control the surface of the active element with Zernike modes. Zernike modes are orthogonal to each other over a unit circle with respect to the wavefront. They are also orthogonal to each other with respect to our merit function for small aberrations, but we demonstrate that this is not valid for aberrations of more than $\lambda/8$ RMS. We aim at the optimization of severely aberrated systems with several λ of aberration and low Strehl ratio. This is represented by the second transition in the graphic representation in Fig. 1. Once the optical system is near the Maréchal limit, the Zernike modes become orthogonal to each other with respect to the merit function.

*kazasidis@fh-muenster.de; Telephone: +49 2551-962-116; photonics-lab.de

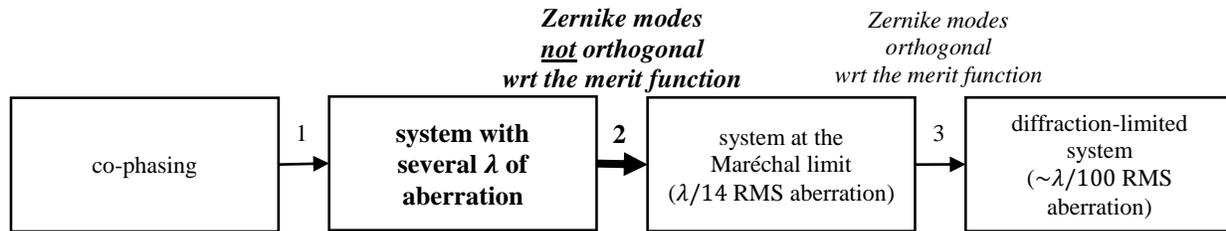


Figure 1. Graphical representation of the stages for the wavefront correction in a space telescope.

2. IMAGE-BASED WAVEFRONT CORRECTION

Our image-based wavefront correction without wavefront sensing is a blind optimization of a merit function that evaluates the quality of the image of the science camera. The configuration parameters for the correction are: 1. the merit function, 2. the control domain, and 3. the algorithm. In this section we discuss our selection of the merit function (2.1) and of the control domain (2.2). The design of the algorithm is determined by these parameters but is outside the scope of this paper.

2.1 Merit function

It is often desirable to state the performance of an optical system by a single number. This immediately allows ranking of different optical systems, optimization of an optical system during its design, or finding the optimum state of an active or adaptive optics system. Examples of performance metrics that deliver a single numerical value are the Strehl ratio (S), the RMS wavefront error (σ), and image-sharpness metrics. Examples of performance metrics that deliver more than a single number, and thus contain more information, are the point spread function (PSF), the modulation transfer function (MTF), wavefront maps, and spot diagrams.

Since all single-number performance metrics lack detailed information about the performance of an optical system, the question arises, which single-number metric is most suitable for a certain imaging scene (e.g., for Earth observation typical imaging applications are urban areas, forests, and maritime surveillance) and for a certain task (e.g. tracking fast moving objects), and what its limitations are. An image-based metric can be applied in different image regions and thus achieve optimal performance for different field angles. For example, when trying to resolve a double star, the region of interest will be a small region of the image. Thus the active optics should correct a narrow field of view. On the other hand, when observing star clusters and nebulae, a larger image region should be corrected, and the correction of the active optics should be balanced over a wide field of view.

We use the image-sharpness metric S_1 introduced by Muller and Buffington¹. This is a single-number metric: $S_1 := \int I(x, y)^2 dx dy$, where $I(x, y)$ is the irradiance at the point (x, y) of the image plane. Our merit function (MF) is a discretized adaptation of S_1 :

$$MF = - \frac{\sum_{n_x=1}^{N_x} \sum_{n_y=1}^{N_y} I(n_x, n_y)^2}{\left[\sum_{n_x=1}^{N_x} \sum_{n_y=1}^{N_y} I(n_x, n_y) \right]^2} N_x N_y . \quad (1)$$

In the above formula x and y are the axes of the image plane, N_x and N_y are the numbers of pixels in each axis, and I is the pixel value. The minus sign converts the sharpness maximization to a minimization problem. MF balances the influences from the contrast of all spatial frequencies.

2.2 Control domain

In optical systems design, the optimization variables are parameters of the optical elements, e.g., material, diameter, thickness, surfaces' curvatures, and position. Examining the performance of sets of values for all the parameters, the designer tries to optimize a merit function. In an adaptive optics system, the adaptive element offers additional degrees of freedom to compensate for aberrations. The optimization variables are the inputs of the adaptive element, i.e., the voltages of the electrodes for a deformable mirror or a spatial light modulator. But the relation of the voltages of a deformable mirror with the optical performance is complex. For this reason, we wish to transform the space of the voltages of the

deformable mirror to wavefront shapes, which can be expressed, e.g., as Zernike modes. Assuming a linear deformable mirror[†], this is done with the influence functions. Using the Zernike modes as variables, we gain insight into the optimization procedure and can possibly reduce the number of the variables in order to speed up the optimization. The Zernike modes used in this paper are listed in the Appendix. Having selected the merit function and the control domain, an algorithm to search the space of Zernike modes should be designed.

3. IMAGE SHARPNESS WHEN VARYING ZERNIKE MODES

3.1 Simulation method

We simulate the pupil wavefront of an imaging system with a uniformly circular aperture in MATLAB over a 300 pixels × 300 pixels grid. We obtain the PSF by using the 2-D fast Fourier transform of the wavefront. We choose the width of the diffraction-limit PSF to be 40 pixels, significantly larger than the required width of 5 pixels, according to the Nyquist sampling theorem. To reduce the computational cost, we limit the total grid of the PSF to 1201 pixels × 1201 pixels, 30 times larger than the diffraction-limited PSF. The error caused by this truncation is negligible. Finally, we generate the MTF of the system by the 2-D fast Fourier transform of the PSF.

We assume that the corrective active element is placed in a plane conjugate to the pupil and that it is controlled with Zernike modes. We normalize the aberrated PSFs to the maximum of the diffraction-limited PSF, to allow comparison among PSFs with different aberrations. Throughout the paper we use the Zernike notation of Wyant and Creath³. We call Z_i the i -th Zernike mode and z_i its coefficient.

3.2 RMS wavefront error and Strehl ratio

The RMS wavefront error (σ) is a common merit function in optical systems design. For small aberrations, the RMS wavefront error is directly related to the Strehl ratio and to the modulation transfer function (MTF). Maréchal formulated the following relation with the Strehl ratio: $S \approx [1 - 2\pi^2\sigma^2/\lambda^2]^2$, where λ is the wavelength. Shannon formulated an empirical formula with the MTF⁴: $MTF(\nu) = DTF(\nu)\{1 - (\sigma/0.18)^2[1 - 4(\nu - 0.5)^2]\}$, where ν is the normalized spatial frequency and $DTF(\nu)$ the diffraction-limited MTF. The Zernike modes are balanced with respect to the RMS wavefront error for every aberration. Therefore, for small aberrations, the Zernike modes are also balanced with respect to the Strehl ratio and to the MTF. This means that, as long as the total aberration is sufficiently small, adding any aberration to the wavefront leads to deterioration of all image quality metrics: increase of the RMS wavefront error, decrease of the Strehl ratio, and decrease of the MTF for all spatial frequencies.

For large aberrations, the Strehl ratio and the MTF can be multiple-valued for the same RMS wavefront error^{5,6}. Higher RMS wavefront error may thus lead to lower or higher Strehl ratio depending on the aberration modes contributing to the aberration. The same is also true for the MTF⁴. The term "large aberrations" commonly refers to $\sigma > \lambda/4$ or $S < 0.4$. To illustrate the complex relation between the RMS wavefront error and the Strehl ratio for large aberrations, we calculate them when defocusing (Zernike mode Z_3) in the presence of different values of astigmatism 0° (Zernike mode Z_4) and show the results in Fig. 2. The RMS wavefront error (Fig. 2a) increases monotonically when the aberration increases, because the Zernike modes are balanced with respect to the RMS wavefront error. On the other hand, the Strehl ratio (Fig. 2b) decreases monotonically for increasing $|z_3|$ only as long as $z_4 \leq 0.4\lambda$ (dark blue, orange and yellow curves). For $z_4 \geq 0.6\lambda$ (violet, green and light blue curves), the Strehl ratio has two maxima away from $z_3 = 0$. The multiple-valued Strehl ratio with respect to the RMS wavefront error is shown in Fig. 2c for $\sigma > 0.2\lambda$.

3.3 Merit function for combinations of Zernike modes

We recently showed⁷ that for large aberrations the Zernike modes are not orthogonal to each other with respect to the merit function defined by (1). Here, we further explore this dependence by running simulations for combinations of two Zernike mode aberrations. Incoherent images of extended objects can be generated by the 2-D convolution of an extended object and the PSF. Using an extended object would restrict the validity of the results in the spatial frequencies that are present in the object. However, the PSF contains all spatial frequencies and can be used to draw conclusions for every spatial

[†] Deformable mirrors often employ actuators which suffer from hysteresis, e.g., piezoelectric actuators. Hysteresis can be eliminated in closed-loop operation with a wavefront sensor. If no wavefront sensor is used, hysteresis can be reduced by using a nonlinear model. The control algorithm should cope with the residual hysteresis, caused by modelling errors².

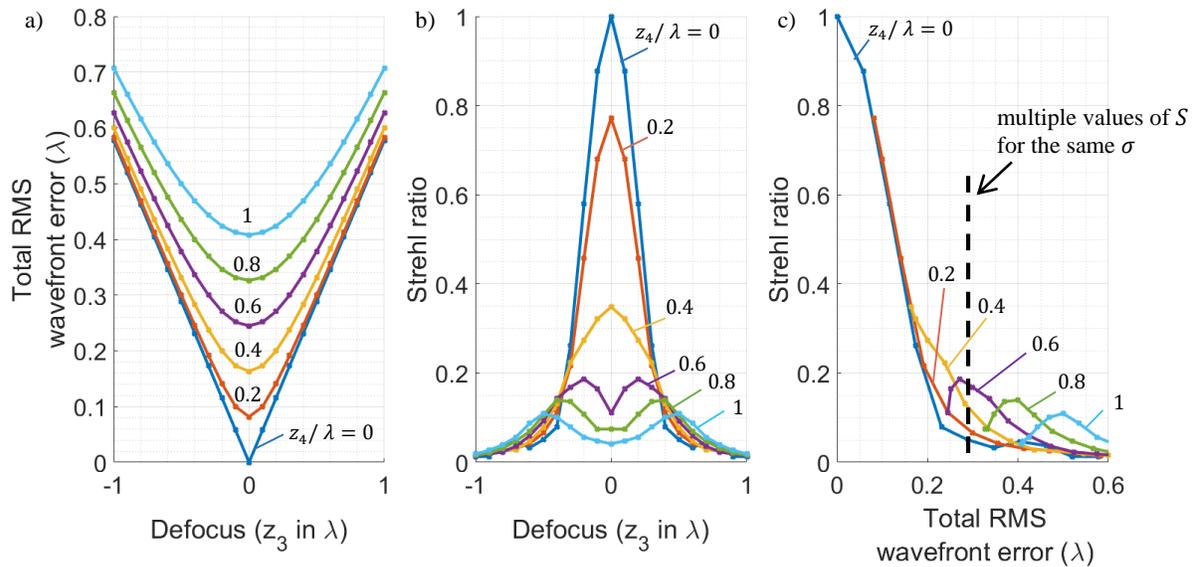


Figure 2. a) The total RMS wavefront error (σ) when varying defocus (Z_3) for different values of astigmatism 0° (Z_4) is monotonically increasing. b) The Strehl ratio (S) may decrease or increase adding defocus (Z_3) in the presence of astigmatism 0° (Z_4). c) Combined representation of plots a) and b). The Strehl ratio becomes multiple-valued with respect to the total RMS wavefront error for $\sigma > 0.2\lambda$. Each curve has a different constant value z_4 , and $|z_3|$ increases for increasing total RMS wavefront error.

frequency that may be present in the object. Therefore, we calculate the merit function for the $1201 \text{ pixels} \times 1201 \text{ pixels}$ image of the PSF. The combinations of Zernike modes shown in this paper are characteristic examples to investigate the physical causes for the non-orthogonality of the Zernike modes with respect to the merit function. To this end, we use wavefront maps, the PSF, and the MTF.

In section 3.3.1 we discuss the combination of defocus (Z_3) and astigmatism 0° (Z_4). In section 3.3.2 we discuss the combination of astigmatism 0° (Z_4) and secondary astigmatism 0° (Z_{11}), as an example for the combination of Zernike modes with the same azimuthal order. The next two sections discuss combinations of trefoil 0° (Z_9): the section 3.3.3 with coma x (Z_6), and the section 3.3.4 with astigmatism 0° (Z_4). The global minimum (optimum) of the merit function is for zero aberration, as expected, in all cases. We show that for large aberrations, the merit function can be improved by adding a Zernike mode, despite the fact that this increases the RMS wavefront error.

3.3.1 Defocus and astigmatism

Figure 3 shows our merit function calculated for the PSF under the same conditions as in Fig. 2, that is when defocusing in the presence of different values of astigmatism 0° . The merit function has a single minimum at $z_3 = 0$ as long as $z_4 \leq 0.2\lambda$, but has two minima for opposite values of z_3 when $z_4 \geq 0.4\lambda$. Although the progression of the curve resembles that of the Strehl ratio (Fig. 2b), it is different, because the Strehl ratio is just the maximum of the PSF, i.e., the value at a single point, whereas the merit function takes into account the whole PSF.

We examine two aberrations, marked as "Aberration 1" and "Aberration 2" in Fig. 3. They both have $z_4 = 0.6\lambda$. Aberration 2 has additional defocus $z_3 = 0.3\lambda$ and lower (better) merit function than aberration 1. Figure 4 shows the wavefront maps, the PSF and the MTF for these aberrations.

The Zernike modes of defocus (Z_3) and astigmatism 0° (Z_4) contribute to the first-order field-independent aberration of focus³: $W_{20} = 2z_3 - 6z_8 \pm \sqrt{z_4^2 + z_5^2}$. If only Z_3 and Z_4 exist in the system, the first-order field-independent focus becomes zero when $|z_3| = z_4/2$, the ratio of the coefficients for the aberration 2. Then the system suffers only from first-order field-independent astigmatism (W_{22}).

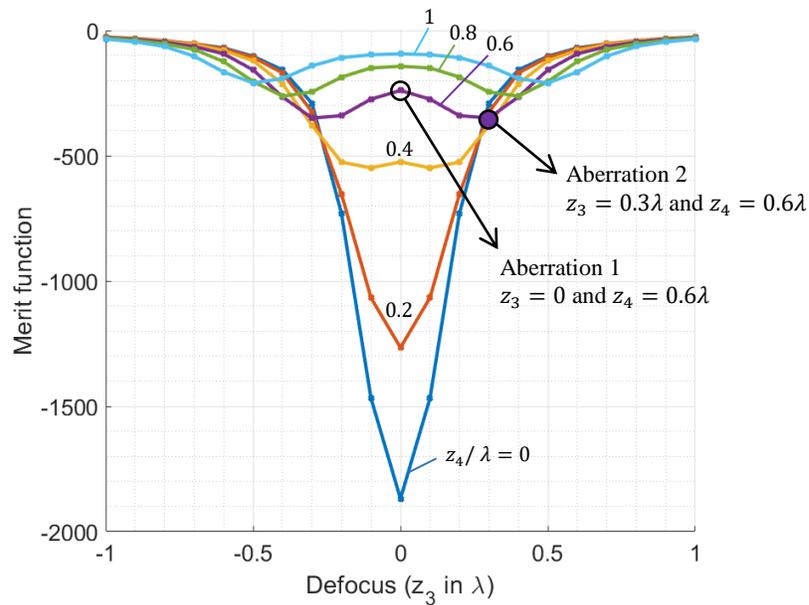


Figure 3. The merit function for the same conditions as in Fig. 1, that is when varying defocus (Z_3) for different values of astigmatism 0° (Z_4).

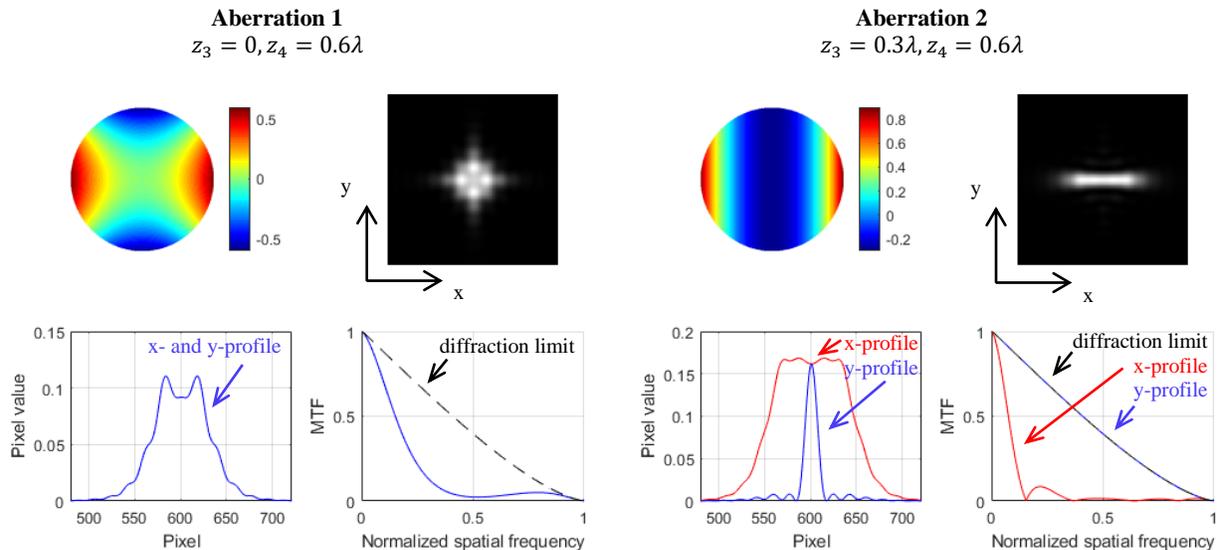


Figure 4. Plots for the aberrations 1 and 2 marked in Fig. 3. The aberrations have the same value of astigmatism 0° (Z_4), but different values of defocus (Z_3). From upper left to bottom right the plots show: the wavefront, the PSF, the PSF profiles in x and y axes, and the MTF in x and y axes.

Adding defocus of $|z_3| = z_4/2$ in the presence of astigmatism 0° slightly increases the width of the PSF in one axis of the image plane (the axis x for the aberration 2). This leads to deterioration of the contrast and of the resolution for spatial frequencies oriented in the direction of this axis. At the same time, this significantly shrinks the PSF in the other axis of the image plane, leading to diffraction-limited contrast and resolution for spatial frequencies oriented in the direction of that axis (the axis y for the aberration 2). This principle is applied to astigmatic systems which are focused differently depending on the object.

3.3.2 Zernike modes with the same azimuthal order

In our recent publication⁷, we showed that for large aberrations the Zernike modes for defocus (Z_3) and spherical aberration (Z_8) are not orthogonal to each other with respect to the merit function. Here, we show another example for combination of Zernike modes with the same azimuthal order, Z_4 and Z_{11} , i.e., the Zernike modes for astigmatism 0° and secondary astigmatism 0° , both of azimuthal order of +2. Figure 5 shows the merit function calculated for the PSF, when varying z_4 in the presence of different values of z_{11} .

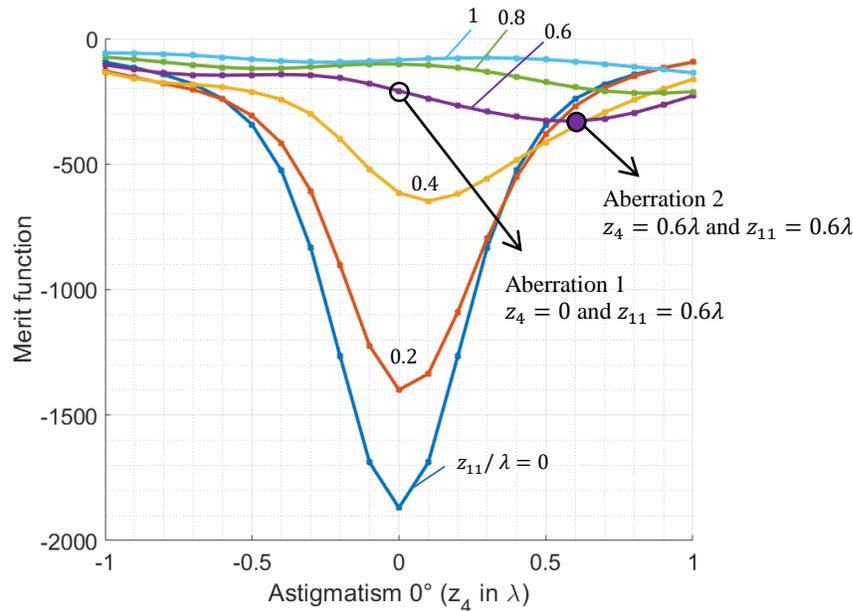


Figure 5. The merit function when varying astigmatism 0° (Z_4) for different values of secondary astigmatism 0° (Z_{11}).

For $z_{11} = 0$ there is a single minimum for the merit function at $z_4 = 0$. For $z_{11} > 0$ the global minimum shifts towards positive values of z_4 . Due to our step size for the Zernike coefficients, we first resolve this shift of the global minimum when $z_{11} = 0.4\lambda$ (yellow curve). In Fig. 6 we examine the aberrations marked as "Aberration 1" and "Aberration 2" in Fig. 5. Both have $z_{11} = 0.6\lambda$ (violet curve), but aberration 2 has additional astigmatism 0° $z_4 = 0.6\lambda$ and lower (better) merit function than aberration 1.

The addition of astigmatism 0° increases the wavefront deviation at the edges of the aperture, but smoothens the wavefront in the central part. This becomes obvious in Fig. 7 that shows the wavefront profiles for the aberrations 1 and 2 along the red dotted lines in Fig. 6. The wavefront with only secondary astigmatism 0° (aberration 1) has smaller variance, but the addition of astigmatism 0° (aberration 2) leads to a flatter wavefront in the central region of the aperture. We can calculate the Zernike modes over a smaller radius. Using the formulas for scaling the Zernike modes from the aperture where they are defined (radius r) to a smaller radius⁸, we find that for the aberration 2 there exists a radius $r' = 0.86r$ on which $z'_4 = 0$. On this radius the scaled Zernike modes comprise only $z'_{11} = 0.44\lambda$. For comparison, for the aberration 1 the scaled Zernike modes on the radius r' comprise $z'_4 = -0.45\lambda$ and $z'_{11} = 0.44\lambda$. The calculations are shown in the Appendix.

In the image plane, the relative heights of the side lobes of the PSF decrease from 41% to 13%. Consequently, energy is squeezed into the central lobe, slightly increasing its width but also increasing its peak (the Strehl ratio). The narrower central lobe of the PSF for the aberration 1 can be interpreted as a higher resolution limit. This is valid, provided that the detection routine doesn't misinterpret the side lobes (with 41% relative height) as distinct objects. Finally, the MTF for the aberration 1 falls fast for low spatial frequencies until 30% of the diffraction-limited cutoff frequency (ν_{cut}) and rises again with a peak near the diffraction-limited contrast at about 50% of the ν_{cut} . The MTF for the aberration 2 is in general smoother and achieves better contrast at low and mid spatial frequencies.

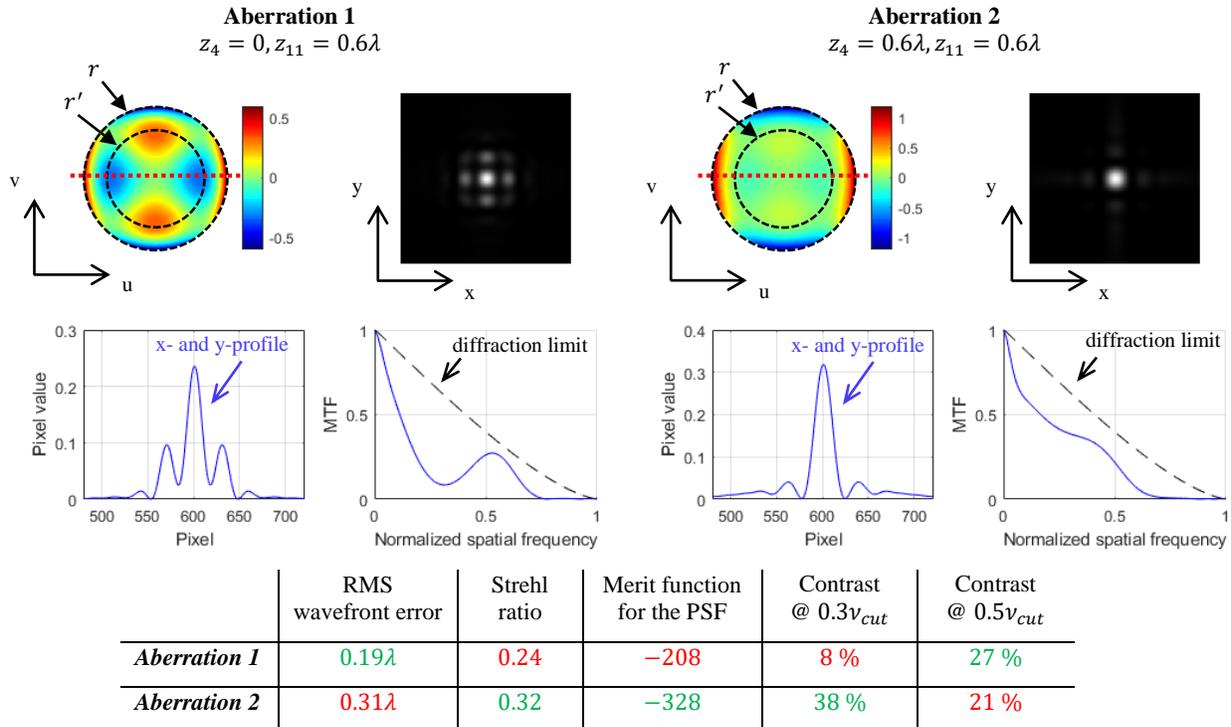


Figure 6. Plots for the aberrations 1 and 2 marked in Fig. 5. The aberrations have the same value of secondary astigmatism 0° (Z_{11}), but different values of astigmatism 0° (Z_4). From upper left to bottom right the plots show: the wavefront, the PSF, the PSF profile, and the MTF. The table below the plots records several metrics to compare the optical performance for the two aberrations. Improved (deteriorated) performance for each metric is marked with green (red).

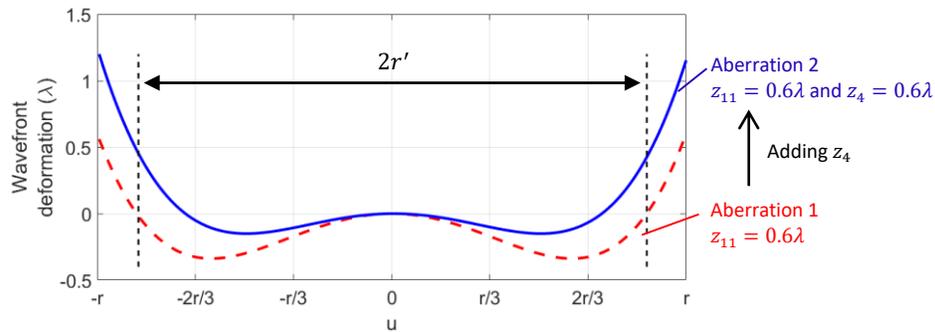


Figure 7. The wavefront profiles for the aberrations 1 and 2 along the red dotted lines of Fig. 6. The addition of astigmatism 0° (blue curve – aberration 2), in the presence of secondary astigmatism 0° , leads to a flatter wavefront in the central region of the aperture (marked as $2r'$). For $r' = 0.86r$ the wavefront comprises a smaller value of secondary astigmatism 0° , and the astigmatism 0° is zero.

3.3.3 Trefoil and coma

The Zernike mode of trefoil 0° is $Z_9 = \rho^3 \cos(3\vartheta) = 4\rho^3 \cos^3 \vartheta - 3\rho^3 \cos \vartheta$. The first term ($4\rho^3 \cos^3 \vartheta$) is the fifth-order aberration of trefoil. The second term ($3\rho^3 \cos \vartheta$) is the third-order coma (neglecting the field dependence) and is added to make the Zernike mode orthogonal to the lower order modes on the unit circle and to minimize the RMS wavefront error. The fifth-order aberration $\rho^3 \cos^3 \vartheta$ (neglecting the field dependence) is called "trefoil" when studying the wavefront. It is also called "elliptic coma", based on the image plane intensity: the circles that appear in the image spot for third-order coma turn into ellipses when fifth-order aberration of trefoil is added^{9,10}. This relation between trefoil and coma

is revealed when we vary the Zernike coma x (Z_6) in the presence of different values of Zernike trefoil 0° (Z_9). Figure 8 shows the merit function calculated for the PSF.

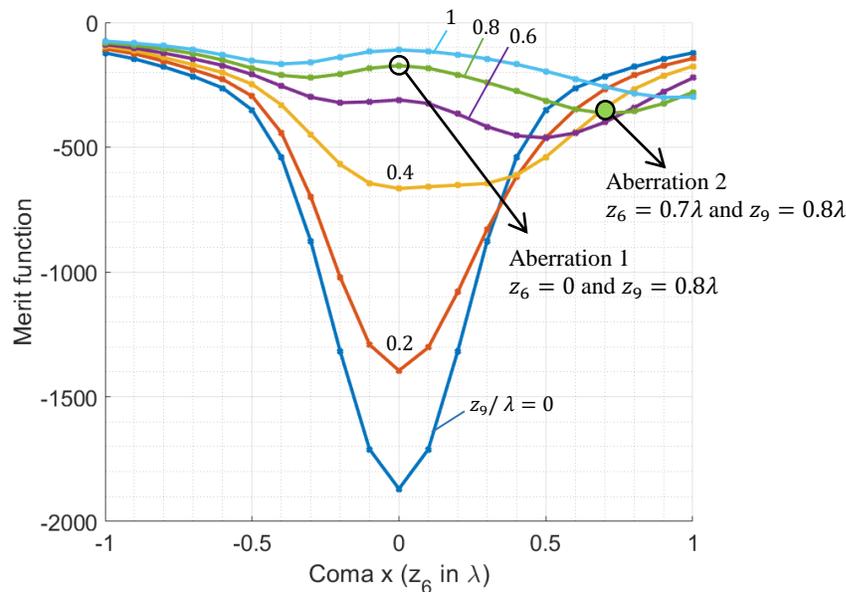


Figure 8. The merit function when varying coma x (Z_6) for different values of trefoil 0° (Z_9).

For $z_9 \leq 0.2\lambda$ there is a single minimum for the merit function at $z_6 = 0$. For $z_9 > 0.4\lambda$ the global minimum shifts towards positive values of z_6 . In Fig. 9 we examine the aberrations marked as "Aberration 1" and "Aberration 2" in Fig. 8. They both have $z_9 = 0.8\lambda$, but aberration 2 has additional coma x $z_6 = 0.7\lambda$ and lower (better) merit function than aberration 1.

Adding positive coma x in the presence of trefoil 0° (aberration 2) makes the wavefront resemble to a single ripple in the u axis. The wavefront is practically uniform in one axis, the v axis of the coordinate system of the aperture. The wavefront actually becomes completely independent of v when $z_6 = z_9$, in which case the wavefront is $W = z_9(4u^3 - 2u)$ and two of the three ripples of the wavefront vanish.

In the image plane, adding coma x increases the Strehl ratio and shrinks the PSF, for both x and y axes. The PSF width in the y axis decreases until the diffraction-limited width. Finally, at the cost of the contrast reduction for low spatial frequencies oriented in the x axis, the MTF increases at mid and high frequencies for spatial frequencies oriented in both x and y axes. It even reaches the diffraction limit MTF for spatial frequencies oriented in the y axis in the case of positive coma x of the same magnitude as the trefoil 0° ($z_6 = z_9$).

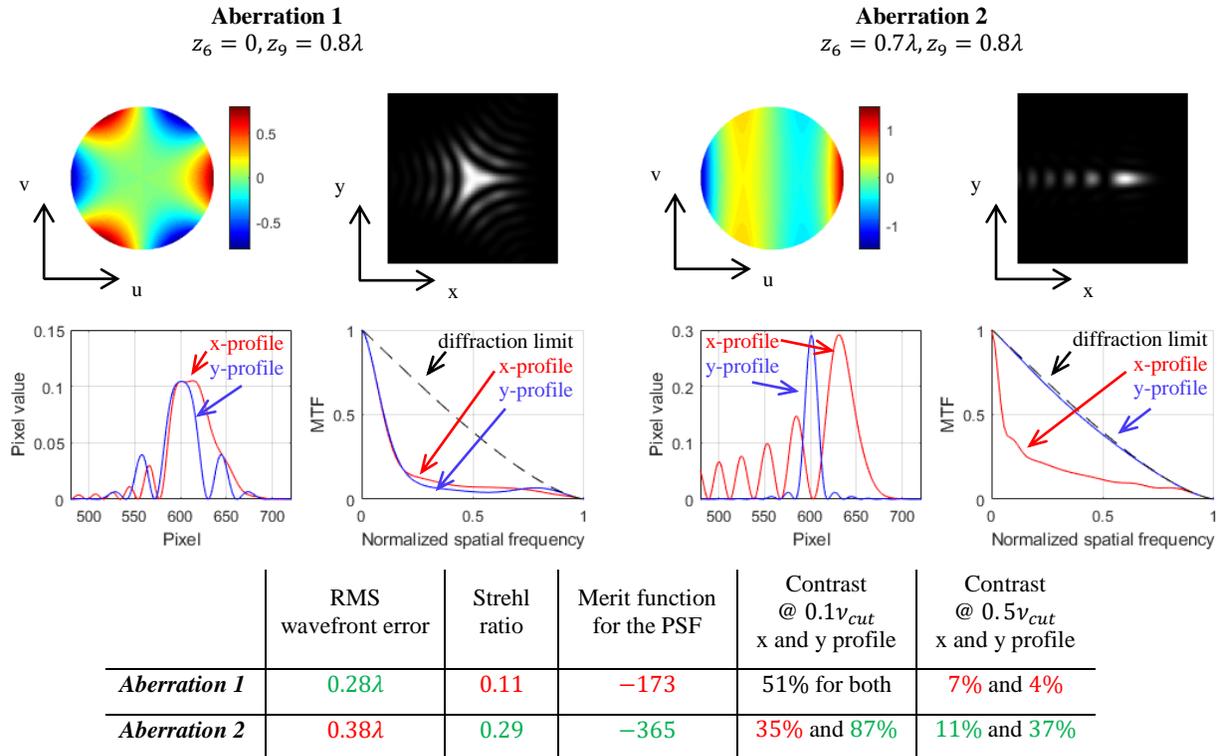


Figure 9. Plots for the aberrations 1 and 2 marked in Fig. 8. The aberrations have the same value of trefoil 0° (Z_9), but different values of coma x (Z_6). From upper left to bottom right the plots show: the wavefront, the PSF, the PSF profiles in x and y axis, and the MTF in x and y axis. The table below the plots records several metrics to compare the optical performance for the two aberrations. Improved (deteriorated) performance for each metric is marked with green (red).

3.3.4 Trefoil and astigmatism

Apart from "trefoil" and "elliptic coma", the fifth-order aberration $\rho^3 \cos^3 \theta$ (neglecting the field dependence) is also called "triangular astigmatism" in part of the literature¹⁰. This is connected to the image plane intensity: in the presence of third-order astigmatism adding fifth-order aberration of trefoil turns the image spot into a triangle. This motivated us to research the combination of the Zernike modes of astigmatism and trefoil. We varied the trefoil 0° (Z_9) in the presence of different values of astigmatism 0° (Z_4) and show the merit function calculated for the PSF in Fig. 10.

For $z_4 \leq 0.2\lambda$ there is a single minimum for the merit function at $z_9 = 0$. For $z_4 \geq 0.6\lambda$ two equal minima appear, for opposite values of z_9 . In Fig. 11 we examine the aberrations marked as "Aberration 1" and "Aberration 2" in Fig. 10. They both have $z_4 = 0.6\lambda$, but aberration 2 has additional trefoil 0° $z_9 = 0.4\lambda$ and lower (better) merit function than aberration 1.

The addition of trefoil 0° partially compensates the ripple caused by the astigmatism 0° at one half of the aperture. For the aberration 2 with positive trefoil 0° , it's the left half of the aperture ($\pi/2 \leq \theta \leq 3\pi/2$, negative u). The wavefront aberration is $W = z_4 Z_4 + z_9 Z_9 = (z_4 + z_9 u)(u^2 - v^2) - 2z_9 uv^2$ and its v -derivative is $\partial W / \partial v = -2v(z_4 + 3z_9 u)$. We notice that for $u = \pm 0.5$, the v -dependence of the wavefront vanishes when $z_4 / z_9 = \mp 3/2$. This is equal to the ratio of the Zernike coefficients of Z_4 and Z_9 for the aberration 2 ($0.6\lambda / 0.4\lambda$).

In the image plane, the PSF shrinks and its peak intensity increases. Ripples with relatively small peaks appear, the highest being 17% of the peak intensity. Finally, the MTF with additional trefoil 0° is slightly lower for spatial frequencies up to about 25% of the ν_{cut} , but is significantly higher for mid spatial frequencies between 25% and 70% of the ν_{cut} . This

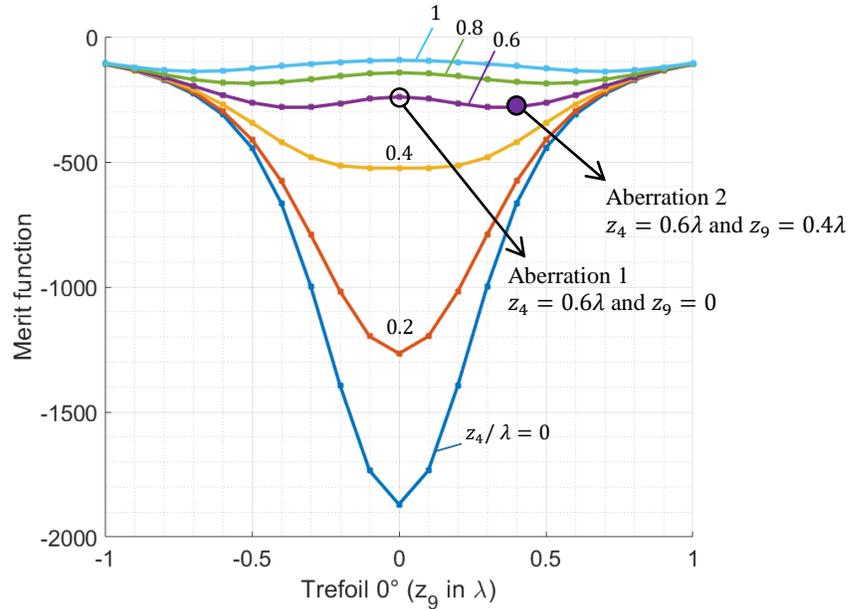
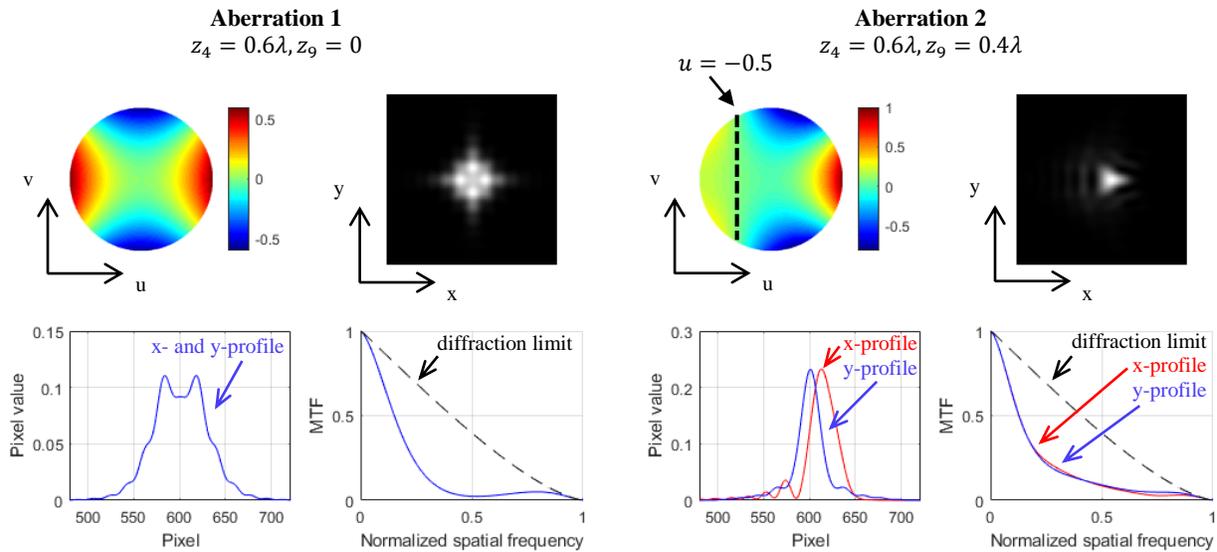


Figure 10. The merit function when varying trefoil 0° (Z_9) for different values of astigmatism 0° (Z_4).



	RMS wavefront error	Strehl ratio	Merit function for the PSF	Contrast @ $0.1\nu_{cut}$ x and y profile	Contrast @ $0.4\nu_{cut}$ x and y profile
Aberration 1	0.25λ	0.11	-240	70% for both	5% for both
Aberration 2	0.28λ	0.23	-282	63% for both	13% for both

Figure 11. Plots for the aberrations 1 and 2 marked in Fig. 9. The aberrations have the same value of astigmatism 0° (Z_4), but different values of trefoil 0° (Z_9). From upper left to bottom right the plots show: the wavefront, the PSF, the PSF profiles in x and y axis, and the MTF in x and y axis. The table below the plots records several metrics to compare the optical performance for the two aberrations. Improved (deteriorated) performance for each metric is marked with green (red).

leads to higher resolution. Setting the limiting resolution at about 10% contrast, the cutoff frequency is about $0.3\nu_{cut}$ for $z_4 = 0.6\lambda$ (aberration 1) and increases to $0.5\nu_{cut}$ by adding trefoil 0° $z_9 = \frac{2}{3}z_4 = 0.4\lambda$ (aberration 2).

4. CONCLUSIONS

We have shown that for aberrations of more than $\lambda/8$ RMS the Zernike modes are not orthogonal to each other with respect to the common image-sharpness metric of Muller and Buffington¹. The non-orthogonality of the Zernike modes should be taken into account when designing the algorithm for image-based wavefront correction, because it may slow down the process or lead to premature convergence. If the algorithm optimizes the Zernike modes separately, several iterations over all Zernike modes are required to ensure that the global minimum is found.

We discussed several combinations of two Zernike modes and investigated the physical causes for their non-orthogonality using wavefront maps, the PSF, and the MTF. We found that in certain cases when adding a Zernike mode, the merit function is improved, although the RMS wavefront error increases. In all the examples we discussed, the improvement of the merit function comes with an increase of the Strehl ratio. However, we cannot directly connect the merit function to the improvement of contrast at a certain range of spatial frequencies. In section 3.3.2 we have shown that for combinations of Zernike modes with the same azimuthal order, a flatter wavefront in the central region of the aperture is more important than the RMS wavefront error across the full aperture for achieving a low (good) merit function.

The results indicate that although the RMS wavefront error is an important metric for image quality, it can be misleading, especially for optical systems with several λ of aberration and low Strehl ratio. In this case, image-based active optics can improve the image quality by adding a low-order Zernike mode to partially compensate an uncorrectable higher-order Zernike mode. An example was discussed in section 3.3.2, where secondary astigmatism 0° (Z_{11}) was partially compensated by adding astigmatism 0° (Z_4). This improved the merit function by 58% and the Strehl ratio by 33%, although the RMS wavefront error increased by 63%.

ACKNOWLEDGMENTS

This work is supported by the funding programme "Qualifizierungsstelle" of Münster University of Applied Sciences.

REFERENCES

- [1] Muller, R. A. and Buffington, A., "Real-time correction of atmospherically degraded telescope images through image sharpening," *J. Opt. Soc. Am.* 64(9), 1200-1210 (1974).
- [2] Kazasidis, O., Verpoort, S., Soloviev, O., Vdovin, G., Verhaegen, M., and Wittrock, U., "Extended-image-based correction of aberrations using a deformable mirror with hysteresis," *Opt. Express*, accepted for publication (2018).
- [3] Wyant, J. C. and Creath, K., "Basic wavefront aberration theory for optical metrology," in [Applied Optics and Optical Engineering], editors Shannon, R. R. and Wyant, J. C., Academic Press, Inc., New York, vol. 11, ch. 1, 28-39 (1992).
- [4] Shannon, R., "Optical Specifications," in [Handbook of Optics Volume I], editor Bass M., In McGraw-Hill, Inc., ed. 2, ch. 35, 35.1-35.8 (1995).
- [5] King, W. B., "Dependence of the Strehl Ratio on the Magnitude of the Variance of the Wave Aberration," *J. Opt. Soc. Am.* 58, 655-661 (1968).
- [6] Mahajan, V. N., "Strehl ratio for primary aberrations: some analytical results for circular and annular pupils," *J. Opt. Soc. Am.* 72, 1258-1266 (1982).
- [7] Kazasidis, O., Verpoort, S., and Wittrock, U., "Algorithm design for image-based wavefront control without wavefront sensing," *Proc. SPIE* 10695, 1069502 (2018).
- [8] Schwiegerling, J., "Scaling Zernike expansion coefficients to different pupil sizes," *J. Opt. Soc. Am. A* 19, 1937-1945 (2002).
- [9] Thompson, K. P., "Multinodal fifth-order optical aberrations of optical systems without rotational symmetry: the comatic aberrations," *J. Opt. Soc. Am. A* 27, 1490-1504 (2010).
- [10] Malacara-Hernandez, D. and Malacara-Hernandez, Z., [Handbook of Optical Design], CRC Press, ed. 3, 163-165 (2013).

APPENDIX

Some low-order Zernike modes

		Polar coordinates	Cartesian coordinates*
Z_3	Defocus	$2\rho^2 - 1$	$2(x^2 + y^2) - 1$
Z_4	Astigmatism 0°	$\rho^2 \cos 2\theta$	$x^2 - y^2$
Z_6	Coma x	$(3\rho^2 - 2)\rho \cos \theta$	$3x^3 + 3xy^2 - 2x$
Z_8	Spherical aberration	$6\rho^4 - 6\rho^2 + 1$	$6(x^2 + y^2)^2 - 6(x^2 + y^2) + 1$
Z_9	Trefoil 0°	$\rho^3 \cos 3\theta$	$x^3 - 3xy^2$
Z_{11}	Secondary astigmatism 0°	$(4\rho^2 - 3)\rho^2 \cos 2\theta$	$4(x^4 - y^4) - 3(x^2 - y^2)$

* In the text the Cartesian coordinates of the aperture are (u, v) . (x, y) are the Cartesian coordinates in the image plane.

Scaling Zernike modes in smaller apertures (according to [8])

with reference to Fig. 6

z_i is the coefficient of the i -th Zernike mode in the aperture r

z'_i is the coefficient of the i -th Zernike mode in the aperture $r' < r$

$$z'_4 = \left(\frac{r'}{r}\right)^2 \left[z_4 - z_{11} \sqrt{15} \left(1 - \left(\frac{r'}{r}\right)^2 \right) \right]$$

$$z'_{11} = \left(\frac{r'}{r}\right)^2 z_{11}$$

Aberration 2: $z_4 = 0.6\lambda$ and $z_{11} = 0.6\lambda$. To find the radius r' on which the scaled z'_4 is zero:

$$z'_4 = 0 \Rightarrow \left[z_4 - z_{11} \sqrt{15} \left(1 - \left(\frac{r'}{r}\right)^2 \right) \right] = 0 \Rightarrow \frac{r'}{r} = 0.86$$

The scaled z'_{11} on the radius r' is:

$$z'_{11} = \left(\frac{r'}{r}\right)^2 z_{11} = (0.86)^2 0.6\lambda = 0.44\lambda$$

Aberration 1: $z_4 = 0$ and $z_{11} = 0.6\lambda$. The scaled Zernike modes on the radius r' are:

$$z'_4 = (0.86)^2 [0 - 0.6\lambda \sqrt{15} (1 - (0.86)^2)] = -0.45\lambda$$

$$z'_{11} = \left(\frac{r'}{r}\right)^2 z_{11} = (0.86)^2 0.6\lambda = 0.44\lambda$$