Calibration of quasi-static aberrations in exoplanet direct-imaging instruments with a Zernike phase-mask sensor

M. N'Diaye\textsuperscript{1,3,4}, K. Dohlen\textsuperscript{1,4}, T. Fusco\textsuperscript{2,1,4}, and B. Paul\textsuperscript{2,1,4}

\textsuperscript{1} Aix Marseille Universit\'e, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388 Marseille, France
e-mail: [mamadou.ndiaye,kjetil.dohlen]@aomp.fr
\textsuperscript{2} Office National d'Études et de Recherches Aérospatiales, Département d'optique théorique et appliquée,29 avenue de la division Leclerc, 92322 Châtillon, France
\textsuperscript{3} Space Telescope Science Institute, 3700 San Martin Drive, Baltimore MD 21218, USA
\textsuperscript{4} Groupement d’Intérêt Scientifique PHASE (Partenariat Haute Résolution Angulaire Sol-Espace)

Received 12 June 2012 / Accepted 14 May 2013

\textbf{ABSTRACT}

\textbf{Context.} Several exoplanet direct-imaging instruments (VLT-SPHERE, Gemini Planet Imager, etc.) will soon be in operation, providing original data for comparative exoplanetary science to the community. To this end, exoplanet imagers use an extreme adaptive optics (XAO) system to correct the atmospheric turbulence and provide a highly corrected beam to a near-infrared (NIR) coronagraph for suppressing diffracted stellar light. The performance of the coronagraph is, however, limited by the non-common path aberrations (NCPA) due to the differential wavefront errors existing between the visible XAO sensing path and the NIR science path and leading to residual speckles that hide the faintest exoplanets in the coronagraphic image.

\textbf{Aims.} Accurate calibration of the NCPA in exoplanet imagers is mandatory to correct the residual, quasi-static speckles remaining in the coronagraphic images after XAO correction in order to allow the observation of exoplanets that are at least $10^6$ fainter than their host star. Several approaches have been developed during these past few years to reach this goal. We propose an approach based on the Zernike phase-contrast method operating in the same wavelength as the coronagraph for the measurements of the NCPA between the optical path seen by the visible XAO wavefront sensor and that seen by the NIR coronagraph.

\textbf{Methods.} This approach uses a focal plane phase mask of size $\sim \lambda/\text{D}$, where $\lambda$ and D denote the wavelength and the telescope aperture diameter, respectively, to measure the quasi-static aberrations in the upstream pupil plane by encoding them into intensity variations in the downstream pupil image. The principle of this approach as described in several classical optical textbooks is simplified by the omission of the spatial variability of the amplitude diffracted by the phase mask. We develop a more rigorous formalism, leading to highly accurate measurement of the NCPA, in a quasi-linear way during the observation.

\textbf{Results.} With prospects of achieving subnanometric measurement accuracy with this approach for a static phase map of standard deviation $44$ nm rms at $\lambda = 1.625 \mu m$ ($0.026 \lambda$), we estimate a possible reduction of the NCPA due to chromatic differential optics by a factor ranging from 3 to 10 in the presence of adaptive optics (AO) residuals compared with the expected performance of a typical current-generation system. This would allow a reduction of the level of quasi-static speckles in the detected images by a factor 10 to 100, thus correspondingly improving the capacity to observe exoplanets.

\textbf{Key words.} instrumentation: high angular resolution – techniques: high angular resolution – telescopes – methods: numerical

1. Introduction

Following the recent images of the exoplanets HR 8799 b, c, d (Marois et al. 2008; Currie et al. 2011a; Soummer et al. 2011; Galicher et al. 2011) and $\beta$ Pic b (Lagrange et al. 2009, 2010; Currie et al. 2011b; Bonnefoy et al. 2011), the astronomical community has high expectations for future discoveries and studies in comparative exoplanetology. In the next few years, high-contrast imaging and spectroscopy will provide many clues to the frequency, diversity, and habitability of exoplanets, the formation and evolution of planetary systems, the relationship between brown dwarfs and planets, etc. (Oppenheimer & Hinkley 2009). The forthcoming instruments, VLT-SPHERE (Beuzit et al. 2008), Gemini Planet Imager (GPI, Macintosh et al. 2008), Subaru-SCExAO (Guyon et al. 2010a), and Palomar P1640 (Hinkley et al. 2011), will start lifting the veil with the direct imaging of young or massive gaseous planets that are $10^6$ fainter than their host star at a few tenths of an arcsecond. Space missions will also be concerned since the future James Webb Space Telescope (Clampin 2008) will form the images of planets of about two Jupiter masses around M-type stars, offering about $10^3$ contrast level at half an arcsecond from a host star (Clampin 2010). The next decade will see the emergence of new exoplanet imagers, such as EPICS (Kasper et al. 2008), for the future European Extremely Large Telescope (E-ELT, Gilmozzi & Spyroulidis 2008) on the ground and coronagraphic telescopes in space (Guyon et al. 2010b; Trauger et al. 2010; Boccaletti et al. 2012), for the study and spectral analysis of extrasolar planets from $10^7$ to $10^{10}$ times fainter than their host stars at a few hundredths of an arcsecond.

However, the huge contrast ratio at a small angular separation (from $10^6$ to $10^{10}$ in the visible and near-infrared (NIR) at less than one arcsecond) between a host star and its planetary companion makes the observation of such an object very challenging (Traub & Oppenheimer 2010). Implementation of several techniques, including extreme adaptive optics (XAO, Fusco et al. 2006; Macintosh et al. 2008; Guyon et al. 2010a; Hinkley et al. 2011), stellar coronography (e.g., see the review
of concepts in Guyon et al. 2006), and post-processing methods (e.g. Marois et al. 2000; Sparks & Ford 2002; Marois et al. 2006; Lafrenière et al. 2007; Mugnier et al. 2008; Vigan et al. 2008), is required to disentangle the photons of the planetary companions from those of its host star. An XAO system will form images at the resolution limit for a ground-based telescope with high Strehl ratio (better than 90%), compensating for the effects of the atmospheric turbulence. A coronagraph will strongly suppress the stellar signal, removing the telescope diffraction effects. Post-processing methods will attenuate the speckles present in the coronagraphic image due to ripples in the residual wavefront after adaptive optics correction to make further improvements, necessary to reduce these residual speckles which are quasi-static wavefront deformations due to non-common path aberrations (NCPA) between the optical path seen by the visible wavefront sensor and that seen by the NIR coronagraph. For instance with SPHERE, a pointing error smaller than 0.5 mas rms was specified to ensure a satisfying centering of the star image on the coronagraph and to reach the scientific objectives of the instrument (Dohlen et al. 2006a). In addition, Stapelfeldt (2006) estimates that wavefront errors (wfe) lower than λ/280 and λ/2800 with λ denoting the wavelength of study, or equivalently 6.0 nm rms and 0.6 nm rms in the H-band (λ = 1.65 μm), will lead to 10^7 and 10^8 contrast thresholds, respectively.

The current-generation instruments, SPHERE and GPI, will respectively rely on the phase-diversity method (Sauvage et al. 2007; Mugnier et al. 2008) and an interferometric approach (Wallace et al. 2010) to measure the non-common path aberrations. They are part of the many methods proposed in the past few years, while other concepts rely on interferometry (Angel 1994; Nishikawa et al. 2008; Galicher et al. 2008, 2010), speckle-nulling technique (Trauger & Traub 2007), coronagraphic phase diversity (Sauvage et al. 2010b), analysis of the light blocked by a coronagraph (Guyon et al. 2009; Vogt et al. 2011), or the Zernike phase contrast principle (Bloemhof & Wallace 2003; Wallace et al. 2011; N’Diane et al. 2012). Several post-processing methods have also been developed to exploit the chromatic behavior of speckles (Marois et al. 2000; Sparks & Ford 2002), using their rotation with respect to the sky (Marois et al. 2006) or based on algorithms of optimal recombination of images (Lafrenière et al. 2007).

We here consider the Zernike sensor based on the phase-contrast method (Zernike 1934). The interest of pursuing the study of this concept is illustrated well by Guyon (2005), who claims that it is “ideal” since it has a photon noise sensitivity factor (βP in his notation) of unity where competing concepts, such as the fixed pyramid sensor, have a sensitivity factor of 1/√2, and the Shack-Hartmann sensor, with the assumptions used in the paper, reaches a minimum factor of 3. We have analyzed the case of using this ideal sensor as a slow wavefront sensor applied to the measurement and real-time calibration of the chromatic component of the NCPA between the optical path seen by the visible XAO wavefront sensor and the one seen by the NIR coronagraph. Our approach is to use the Zernike wavefront sensor in the NIR to minimize these chromatic differential aberrations.

By taxing a small part of the science beam (<10%) as close as possible upstream of the coronagraph (see Fig. 1) and by using long exposures (>1 s), we benefit from the excellent noise-propagation properties of this sensor to correct these quasi static aberrations in parallel with science observations. Clearly, the use of a beam splitter before the coronagraph implies that the Zernike sensor does not see exactly the same aberration function as the coronagraph. However, since the differential optics are limited to a single, fixed component whose optical quality can be optimized by the use of high-quality optics, the differential aberrations can be very small (typically 1–2 nm rms) and stable in time. This point is particularly important and represents the main advantage over the classical approach, used notably in SPHERE, where the chromatic component of the NCPA is calibrated at intervals of several hours using phase diversity. Since the chromatic differential optics (represented as double arrows in Fig. 1) between the visible XAO wavefront sensor and the NIR coronagraph consists of physically long optical paths (several meters) containing fixed, as well as variable, optical elements (e.g., atmospheric dispersion correctors using rotating prism elements), their contributions to NCPA are necessarily time variable, leaving non-negligible, chromatic differential aberrations in the coronagraph plane. In the proposed Zernike sensor, we will still rely upon classical calibration methods such as phase diversity (new developments in coronagraphic phase diversity (Paul et al. 2012) can be of particular interest in this context), but the lifetime of this calibration is expected to be much longer, probably extending well beyond one night.
The GPI instrument approaches this problem by implementing a real-time, post-coronagraphic wavefront sensing system (HOWFS) based on a Mach-Zehnder interferometer setup. Again, a beam splitter taxes a fraction of the science light, but since the separation occurs after the coronagraph, it does not introduce differential aberrations. Potentially, this system therefore does not require external calibration. The cost of such a system lies in its complexity, however, the Mach-Zehnder interferometric system is very alignment sensitive, and a piezo-electric phase modulator does not require external calibration. The cost of such a system again, a beam splitter taxes a fraction of the science light, but since the separation occurs after the coronagraph, it does not introduce differential aberrations. Potentially, this system therefore does not require external calibration. The cost of such a system lies in its complexity, however, the Mach-Zehnder interferometric system is very alignment sensitive, and a piezo-electric phase modulator does not require external calibration. The cost of such a system

Bloomhof & Wallace (2003) propose a practical implementation of the Zernike sensor for real-time AO operation in combination with a Shack-Hartmann wavefront sensor for bootstrapping, and Wallace et al. (2011) propose a phase-shifting version using an original piston system. N’Diaye et al. (2012) presented the first results of a development of the sensor based on fixed, ion-beam machined phase masks, arguing the obtention of high-quality masks with small $\lambda/D$ dimensions. A similar concept has been tried earlier for the measurement of phasing errors in segmented telescopes (Dohlen et al. 2006b) and demonstrated on the sky with the VLT (Surdej et al. 2010; Vigan et al. 2011), using a mask with a size roughly equal to the atmospheric seeing-disk diameter. It is interesting to note that the Zernike sensor can be seen as a compact equivalent of the Mach-Zehnder interferometer with an amplitude or phase mask inserted in the interfering arms (Angel 1994; Langlois et al. 2002; Dohlen 2004). It is also noteworthy that the Zernike sensor can be used with any aperture shape and pupil obscuration (shadow of the secondary mirror and spiders, gaps between segments, etc.) and that it is sensitive to phasing errors in segmented telescope systems. It therefore constitutes a promising wavefront-sensing option for future space and ground-based instruments like EPICS, receiving light through segmented primary collectors such as the European Extremely Large Telescope. On a shorter timescale, its insertion into current high-contrast imagers like SPHERE on the VLT will allow online coronagraph-plane wavefront sensing as a complement to the currently implemented phase diversity method at a minimal cost in terms of system modifications.

This paper addresses the performance of the Zernike sensor in the context of a SPHERE upgrade as described above in order to move forward practical implementation and on-sky demonstration of the concept. A recollection of the mathematical formalism is followed by an analysis of the impact of various error sources, such as chromatic effects and AO residuals, leading to constructing a complete error budget. While the principles of the Zernike phase-contrast method are explained in classical textbooks (Hecht & Zajac 1987; Malacara 1992; Goodman 1996; Born & Wolf 1999), the mathematical description used there is not fully complete. By introducing our notation we provide a comprehensive treatment, considering also the case of deviations from the classical $\pi/2$ phase shift, which can be particularly useful for measuring very small aberrations.

To estimate the precision obtained with this method, we have identified error sources, including detection noise, AO residuals, and chromatic effects, and built an error budget as a function of target flux. This detailed study of measurement errors constitutes a major contribution to the knowledge of the Zernike sensor.

We have excluded investigation of online measurements of variable amplitude aberrations since these are not considered limiting for current-generation systems (SPHERE, GPI, etc.) and since these instruments do not have means for correcting such aberrations (i.e., a second deformable mirror). However, given a slight increase in complexity, allowing observation of a pupil image not seen through the Zernike mask simultaneously with the Zernike pupil image, online amplitude measurements will be possible.

In Sect. 2, we describe the formalism of the Zernike phase-mask sensor, underlining its simplicity, sensitivity, and quasi-linearity for a direct reconstruction of the phase map related to the measurement of the intensity in the pupil. Numerical simulations illustrate the reconstruction of a static phase map with this approach in Sect. 3, while phase-map reconstruction in the presence of AO residuals is investigated in Sect. 4. Chromaticity effects, such as finite spectral bandwidth or the difference between design and measurement wavelengths, are analyzed in Sect. 5. We finally discuss the application of the Zernike sensor to an exoplanet direct imaging instrument in Sect. 6, proposing an error budget in order to estimate the ultimate performance limits of the system.

### 2. The Zernike phase-mask sensor

#### 2.1. Principle

The scheme of the Zernike phase mask concept is given in Fig. 2. Residual wavefront errors after the XAO system are contained in the complex amplitude of the electric field at the entrance pupil plane A of the Zernike sensor. A phase mask in the form of a small circular depression in a glass plate is placed at the following focal plane B where the star image is formed, introducing a phase change for the complex amplitude of the central part of the star image going through the mask. This leads to interference between the electric fields going through and outside the phase disk in the relayed pupil in plane C, producing an intensity pattern that is related to the wavefront aberrations. The exact intensity encoding of wavefront errors depends on the size and depth of the mask, and quasi-linearity is achieved with an adequate choice of these parameters.

We note that this interferometric phenomenon is closely related to the nulling observed in the phase mask stellar coronagraph (Roddier & Roddier 1997) where a circular $\pi$-phase disk is used to remove diffracted starlight. The objective of the Zernike sensor, however, is different since it aims at a coding of the residual aberrations as pupil plane intensity variations rather than optimal coronagraphic extinction.

![Fig. 2. Layout of the Zernike sensor for an accurate reconstitution of the residual phase in the framework of a high-contrast exoplanet imager.](image-url)
2.2. Formalism

The classical textbook description of the mathematical formalism of the Zernike sensor (Hecht & Zajac 1987; Malacara 1992; Goodman 1996; Born & Wolf 1999) as reported in Wallace et al. (2011) is often somewhat simplified, neglecting the variability of the reference wave’s complex amplitude across the pupil.

In the following, we describe the formalism of the Zernike sensor. For the sake of clarity, we omit the position vectors \( r \) and \( \rho \) in the pupil and focal planes, their modulus \( \rho \), and the wavelength \( \lambda \), and \( \mathcal{F} \) symbolizes the Fourier transform operator in which we include the Fourier optics scaling factor \( 1/\lambda f \), with \( f \) the telescope focal length (Goodman 1996). The Fourier transform is also shorthand written with a hat: \( \mathcal{F}[A] = \hat{A} \). The operator \( \otimes \) denotes the convolution product.

The complex amplitude of the electric field \( \Psi_A \) at the aperture (plane \( A \)) is given by

\[
\Psi_A = P e^{i\phi} = P_0 (1 - \epsilon) e^{i\phi},
\]

in which the real functions \( P \) and \( \phi \) describe the amplitude and phase, respectively, of the entrance pupil. The phase function is assumed to have a zero mean, and \( P \) is normalized such that \( P_0 \) defines the telescope aperture shape, equal to 1 inside the pupil and 0 elsewhere and \( \epsilon \) is a zero-mean amplitude error function.

The phase mask is located in the following focal plane, denoted plane B. Its amplitude transmission function \( t \) can be written as

\[
t = 1 - (1 - e^{i\phi}) M, \tag{2}
\]

in which \( M \) defines the top-hat function of the phase mask equal to 1 for \( |\rho| < d/2 \) and 0 otherwise, \( d \) denoting the mask diameter. The term \( \theta \) represents the phase shift introduced by the mask.

The complex amplitude of the field \( \Psi_B \) after the mask is then given by

\[
\Psi_B = t \Psi_A, \tag{3}
\]

and so the complex amplitude in the exit pupil following the mask is

\[
\Psi_C = \Psi_B = \Psi_A - (1 - e^{i\phi}) \hat{M} \otimes \Psi_A. \tag{4}
\]

For masks \( M \) smaller than the Airy disk, \( \hat{M} \) becomes a smooth function that is broader than the pupil. In highly corrected systems, and since \( \epsilon \) and \( \phi \) have zero mean, the convolution product can be approximated to

\[
b = \hat{M} \otimes \Psi_A \approx \sqrt{S} \hat{M} \otimes P_0 = \sqrt{S} b_0, \tag{5}
\]

where \( S = (1 - \sigma^2/2)^2 \) is the Strehl ratio (Maréchal 1947) and \( \sigma^2 \) the wavefront variance. This is a real function, smoothly varying across and beyond the pupil, see Fig. 3. Knowing both pupil geometry and mask geometry, \( b_0 \) can be computed once and for all. With typical mask diameters close to \( \lambda/D \), \( b \) has a profile similar to that of an Airy pattern twice the size of the pupil, see Fig. 3. The presence of \( S \) in this expression is problematic for the absolute measurement of finite wfe, but in the vicinity of zero aberration, which is the regime we seek through closed-loop correction, it will be close enough to unity to be ignored. We note that the spatial variability of \( b \) is ignored both by classical authors (Malacara 1992; Born & Wolf 1999) and by contemporary authors (Wallace et al. 2011), treating this as a constant. Maintaining the spatial distribution of \( b \) is important in our analysis since ignoring it will introduce spurious aberrations.

We can now rewrite the exit pupil electric field as

\[
\Psi_C = \Psi_A - (1 - e^{i\phi}) b = a + b e^{i\phi}, \tag{6}
\]

with \( a = \Psi_A - b \). The electric field in pupil plane C is seen to represent the interference of the wavefront \( a \) emanating from outside of the image plane mask with \( b \), a smooth reference wavefront, emanating from within the mask. As in classical interferometry we therefore expect, as noted and exploited by Zernike, gaining him a Nobel prize, that phase modulations in plane \( A \) show up as intensity variations in plane \( C \).

Corresponding to phase error \( \phi \), Eq. (6) can be represented in the Argand diagram as a shifted circle. The unit circle centered on the origin representing \( \Psi_A \) is shifted to a new center at \( -b (1 - e^{i\phi}) \), see Fig. 4. The intensity in the exit pupil \( I_C \) is equal to the square of the length of the vector \( \Psi_C \) joining the corresponding point on the shifted circle to the origin:

\[
I_C = |\Psi_C|^2. \tag{7}
\]

As expected, the length of this vector is indeed related in a quasi-linear way to the phase errors in the entrance pupil for a certain range of \( \phi \). Figure 5 plots the exit pupil intensity as a function of phase error for a typical mask geometry.

By accounting for amplitude errors in the pupil, the scheme described above becomes somewhat more cluttered, but noting that if we remove the mask from plane B (actually just shifting it a distance equal to a few times its diameter), we obtain an exit pupil intensity of \( I_{CD} = |\Psi_A|^2 = P^2 = P_0^2 (1 - \epsilon)^2 \), from which the pupil amplitude function \( P \), hence \( \epsilon \), can be extracted during the calibration stage.

Inverting Eq. (7) yields \( \phi \) as a function of \( I_C \). Expressing the complex exponentials as cosines and sines and using Eqs. (1) and (6), we find that

\[
I_C = P^2 + 2b^2 (1 - \cos \theta) + 2Pb [\sin \phi \sin \theta - \cos \varphi (1 - \cos \theta)]. \tag{8}
\]

For very small phase errors, we can consider a linear case where we only maintain first-order terms of \( \phi \) in the Taylor expansion. Then we find

\[
I_C = P^2 + 2b^2 (1 - \cos \theta) + 2Pb [\sin \phi \sin \theta - (1 - \cos \theta)]. \tag{9}
\]
A more accurate expression is obtained by maintaining the second order of \( \phi \) in the Taylor expansion. Then \( I_C \) can be written as

\[
I_C = P^2 + 2b^2(1 - \cos \theta) + 2bP \left[ \phi \sin \theta - (1 - \varphi^2/2)(1 - \cos \theta) \right].
\] (12)

Solving this second-order equation is tedious and not very rewarding since it does not simplify well. For the purpose of illustration we therefore provide here the solution corresponding to the classical case where \( \theta = \pi/2 \) and ignoring amplitude errors:

\[
I_C = 1 + 2b^2 + 2b(\varphi^2/2 + \varphi - 1),
\] (13)

for which we get the solution

\[
\varphi = -1 + \sqrt{3 - 2b - (1 - I_C)/b}.
\] (14)

For \( b = 0.5 \), the phase error is simply given by \( \varphi = -1 + \sqrt{2I_C} \). Figure 5 compares the exact expression (Eq. (8) in blue solid line) with this simplified expression (Eq. (14), second-order in green dashed line) and its linear version (Eq. (11), first-order in red dotted line) in the case where \( P = 1, b = 0.5 \), and \( \theta = \pi/2 \).

Clearly, the second-order approximation represents the exact expression in the \( \pm \pi/4 \) range well. The linear expression is significantly less accurate, but could be interesting in a closed-loop, null-error context.

A value of \( b \) close to 0.5 over the pupil is achieved using a phase mask with angular diameter of 1.06\( \lambda/D \) in the case of a circular aperture, see Fig. 3. This recalls the case of a phase-mask coronagraph, where a mask of this dimension but with a \( \pi \) phase shift achieves perfect nulling in the center of the coronagraphic exit pupil (Roddier & Roddier 1997).

Our study is limited to circular apertures, but the formalism of the Zernike sensor is valid for any aperture shape and obscuration (shadow of the secondary mirror and spiders, gaps between segments, etc.) since the intensity measurement is made inside the geometric pupil. With the Zernike phase mask, like the Roddier & Roddier phase mask, the light diffracted by an obscuration in the entrance pupil and modified by the mask interference effects will remain within the re-imaged obstructions in the relayed pupil, making the geometric pupil free of light contamination, unaltered for the exit pupil intensity measurements and wavefront reconstitution inside it. Phase discontinuities caused by segment phasing errors are also perfectly coded by this approach, making the Zernike sensor a very promising system for the NCPSA and piston measurements in EPICS, the future exoplanet imager for the segmented E-ELT (Kasper et al. 2008).

### 2.3. Noise Propagation

To investigate the ultimate performance of this concept, we consider error propagation due to measurement noise and other effects in the case of operation near zero phase error. In this case we are safely within the linear regime described by Eq. (11). Since the intensity used in this expression is normalized by the average entrance pupil flux, \( I_C = S/S_0 \), where \( S \) is the measured signal in the Zernike exit pupil and \( S_0 \) is the average signal in the entrance pupil (both in number of photo-electrons per pixel):

\[
\varphi = S/S_0 - 0.5.
\] (15)

By differentiating this expression we get the following relationship for noise on the phase due to small signal fluctuations \( \delta S \):

\[
\delta \varphi = \delta S/S_0.
\] (16)
Different noise sources can be considered, in particular detector readout noise \( \delta S_D \) and photon noise \( \delta S_p = S \cdot 0.5 \delta \nu \) when \( \nu = 0 \). Denoting \( \sigma^2 \) the variance of the phase across the pupil and denoting the different contributors by the subscripts used above, we can write the following error budget for detection noise:

\[
\sigma^2 = \sigma_D^2 + \sigma_p^2 = \delta S_D^2 / S_0^2 + 0.5 / S_0.
\]

In terms of noise propagation, our concept proves to be the most optimal wavefront sensor among the existing systems as shown by Guyon (2005).

### 3. Static phase map reconstruction

To illustrate wavefront reconstruction with the Zernike sensor by numerical simulations, we first consider its response to low-order aberrations. We apply increasing amounts of each of the first eight Zernike coefficients to an otherwise flat wavefront, see Fig. 6. For all the considered modes, a linear response is observed for values lower than 48 nm (0.03 \( \lambda \)) rms. Within this range and except for the tip/tilt aberrations, the curve of the Zernike sensor response to low-order aberrations is linear with a slope of unity. For tip/tilt aberrations (lateral image movement), a slope of 0.81 is observed. This curious exception is a standard deviation of 9.8 nm (6 \( \lambda \)) rms wavefront error. The error map shows a standard deviation of 9.8 nm (6 \( \times 10^{-3} \lambda \)) rms. In addition to the expected random error map, we clearly see a deterministic component correlated with the original phase map. This reminiscence is due to the nonlinearity of the response curve, leading to an imperfect averaging of the AO residual phase.

Figure 10 shows the standard deviation of the error map as a function of the AO-residual wavefront error for 100 and 1000 phase screens (solid lines). Three ranges of AO residuals can be identified in this plot: a constant level for very small AO residuals corresponding to the static reconstruction error (1.9 nm rms) reported above, and a nearly linear increase up to a saturation level corresponding to the aberration of the original static screen (44 nm rms). In the intermediate regime, corresponding to a range of AO residuals between about 12 nm and 240 nm rms, the error is approximately equal to 1/10 of the amount of the AO residual independently of the number of phase screens. This level is close to what would be expected when averaging 100 phase screens (broken lines) but too high in the case of 1000 screens, probably due to the nonlinearity effect observed in Fig. 9.

In practical cases, the measurement made by the aid of this sensor will be fed back into the AO system in the form of offset voltages for the deformable mirror, leading to a near zero NCPA. Then the nonlinearity effect will be strongly reduced. To illustrate this case, Fig. 10 also plots the standard deviation of the error map as a function of the AO residual wavefront error for zero NCPA, showing a perfect fit with the expected 1/\( \sqrt{N} \) relationship (broken lines) up to about 240 nm (0.15 \( \lambda \)) rms. Assuming 90 nm (0.055 \( \lambda_0 \)) AO residual aberrations for zero NCPA, an error of 1.7 \( \times 10^{-3} \lambda_0 \) is estimated for 1000 AO phase screens. Extrapolating this result for \( N \) phase screens gives us a reconstruction error \( \sigma = 5.4 \times 10^{-2} \lambda_0 / \sqrt{N} \). The error for very large
residuals, for which the wavefront sensor breaks down, is independent of the number of screens.

The number of phase screens \( N \) that are averaged during an exposure time \( T \) is \( N = T/\tau \), where \( \tau \) is the lifetime of the phase screen. In current-generation XAO systems such as SPHERE, operating at around 1 kHz, the total adaptive optics wavefront error budget amounts to some 55 nm rms (Fusco et al. 2006), of which about \( 50 \) nm rms is attributed to sources such as Shack-Hartmann wavefront sensor noise propagation effects. The wavefront sensor noise is an uncorrelated noise related to the operating frequency \( v_{\text{XAO}} \) of the XAO system. Considering \( v_{\text{XAO}} = 1 \) kHz, the aberrations due to the wavefront sensor noise has a lifetime \( \tau_{\text{SH}} = 1/v_{\text{XAO}} = 1 \) ms. A 1 s exposure therefore corresponds to 1000 noise error screens.

A part of the remaining AO errors, estimated to 20 nm rms in the SPHERE budget (Fusco et al. 2006), is attributed to lag in the servo loop. As described in Macintosh et al. (2005), these errors create long-lived (order of 1 s) speckles in coronagraphic images due to aberrations caused by time lag in classical AO systems as atmospheric phase screens blow across the telescope pupil. While advances in AO control technology, in particular the predictive Fourier control method (Poyneer et al. 2007), hold good promise of reducing the importance of such aberrations, this technique is not implemented in the SPHERE baseline.

It has been pointed out to us that since this effect appears at a frequency similar to the one we propose for measuring NCPA using the Zernike sensor, this residual aberration will neither be corrected nor averaged out, hence constituting a limiting noise source. This is not the case, however, since although the speckles in the image plane have a long lifetime due to the translation of a quasi-static phase screen across the pupil, wavefront error observed by an individual detector pixel located in the pupil plane is not static. Indeed, while the corresponding speckle lifetime is on the order of \( D/v \), with \( D \) the telescope diameter and \( v \) the wind speed, the life-time of an aberration observed in a pupil-plane pixel is \( D/(\eta N_{\text{pix}}) \), where \( N_{\text{pix}} \) is the number of pixels across the pupil. In the case of SPHERE, assuming \( N_{\text{pix}} = 40 \) for the Zernike sensor in order to match the spatial sampling of the XAO system, the pupil-plane lifetime of this effect is on the order of \( \tau_{\text{lag}} = 25 \) ms.

Denoting aberration lifetime as \( \tau \) and exposure time as \( T \), the Zernike sensor measurement error variance due to an aberration \( w \) is \( \sigma^2 = \tau^2/\tau_{\text{lag}} \). We therefore identify two main contributors of AO residual aberrations to the error budget:

\[
\sigma_{\text{AO}}^2 = \sigma_{\text{lag}}^2 + \sigma_{\text{SH}}^2 = (\tau_{\text{lag}}^2 + \tau_{\text{SH}}^2)/T. \tag{18}
\]

5. Sensitivity to chromatic effects

The above results of the Zernike sensor have been given for a monochromatic light source emitting at wavelength \( \lambda_0 \), which has been used for the definition of the phase mask and for the data analysis. We now study the impact of chromatic effects. We first study the impact of a source wavelength \( \lambda \neq \lambda_0 \) on the Zernike sensor measurement for an analysis done at \( \lambda_0 \), see Fig. 11. In the case of zero aberration, the measurement error is minimal at \( \lambda_0 \) and increases quickly when the analysis wavelength moves away from the design wavelength. However, if we know the source wavelength precisely and recalculate the \( \theta \) and \( b \) terms accordingly, the residual error is reduced to the numerical noise level for the entire wavelength range considered here.

When the nonzero aberration case is considered, as shown by the red solid-line curve in Fig. 11, a steeper rise from the optimal wavelength is observed, but, more remarkably, the optimal wavelength is blue-shifted by about 1%. This can be explained by the fact that the reduced Strehl ratio is partly compensated for by an increased concentration of flux within the Zernike mask. Assuming a realistic situation where the analysis wavelength is known and taken into account for the zero-aberration case, the measurement errors in the presence of 0.027\( \lambda_0 \) static aberrations are smaller than 1.5 \times 10^{-2}\( \lambda_0 \) within the range considered, see Fig. 11. Possible improvements in this value can be obtained, working at an analysis wavelength ~1% shorter than the source wavelength as shown in the plot where the red-line curve line is found below the blue-line one within the range between 0.98\( \lambda_0 \) and \( \lambda_0 \), offering possible advanced strategies of analysis, but this goes beyond the scope of this paper.

In broadband, assuming a flat spectral distribution over a spectral range \( \Delta \lambda \) centred at \( \lambda_0 \), the zero-aberration error has the

Fig. 9. Same as Fig. 7 but in the presence of 100 phase screens with 0.05 \( \lambda \) rms wavefront errors in the entrance pupil.

Fig. 10. Reconstitution error for a static phase map (solid line) and zero NCPA (dashed line) as a function of the amount of AO residuals. The dotted line corresponds to the static reconstitution error for the 44 nm rms NCPA map.

Fig. 11. Wavefront error measurement as a function of the source wavelength \( \lambda \) for an analysis at \( \lambda_0 \).

shape of the monochromatic curves observed in Fig. 11: the bandwidth-induced shift is caused by the asymmetry of these curves, more pronounced in the unaberrated case, and the measurement error is related to their steepness, more pronounced in the aberrated case.

Assuming a realistic case of 20% bandwidth, corresponding roughly to the atmospheric H-band centered on 1625 nm,
and using the central wavelength as the analysis wavelength, the chromatic measurement error for zero aberrations is \( \sigma_\lambda = 4.1 \times 10^{-3} \lambda_0 \), corresponding to 0.65 nm, when using the central wavelength as the analysis wavelength, but can be reduced to 3.0 \times 10^{-5} \lambda_0 if the filter function and object spectral type are known, see Fig. 13. As discussed in the introduction, we must account for the calibration of a small differential aberration between the Zernike sensor plane and the coronagraph plane of 1 to 2 nm rms. This is small enough not to have a significant impact on this budget, as can be estimated from the curves in Fig. 13. We therefore assume the value of \( \sigma_\lambda = 3.0 \times 10^{-3} \lambda_0 \) for the error budget explained below.

6. Application to a real system, error budget

We here consider the application of this sensor in a typical XAO system similar to the ones currently being constructed (Macintosh et al. 2008; Sauvage et al. 2010a). These are optimized to work in the NIR around \( \lambda = 1.6 \mu m \) with atmospheric correction operating at frequencies exceeding 1 kHz. NCPA are calibrated off-line at intervals limited by practical and observing efficiency considerations to more than one hour, perhaps as much as one day. Even though the instrument is expected to be quite stable during this time frame, evolutions in NCPA due to thermo-elastic effects, rotating elements such as the atmospheric dispersion correctors, and air-mass related chromatic distortion. To this end, an error budget is built in the context of a paper: separating the measurement beam from the science beam just upstream of the coronagraph mask. When aberrations are measured during observations and fed back to the XAO system in the form of updated reference slopes, the aberrations measured by the sensor will essentially be zero so that we can ignore error terms related to the absolute level of wavefront distortion. To this end, an error budget is built in the context of a SPHERE upgrade based on the terms identified throughout this paper:

\[
\sigma_\text{err}^2 = \sigma_\text{D}^2 + \sigma_\text{p}^2 + \sigma_\text{AO}^2 + \sigma_\lambda^2 \\
= \delta S_D^2/S_0^2 + 0.5/S_0 \\
+ \left( \tau_{\text{Lag}} w_{\text{Lag}} \right)^2/T + \left( 3.0 \times 10^{-5} \lambda_0 \right)^2 .
\]

(19)

The different parameters for our calculation are listed in Table 1 while the different contributors to the measurement errors for our concept are reported in Table 2. The AO cut-off frequency of SPHERE is \( v_C = 1/2p \) with \( p = 20 \) cm, and so a pupil of 40 \times 40 pixels (or subpupils) is required for the Zernike sensor to perform measurement up to this frequency. We consider two different exposure times in this study, 1 s and 10 s.

Figure 14 displays the resulting measurement error of the NCPA map as a function of number of photo-electrons per pixel. At low flux, the measurement error is dominated by readout noise. At intermediate flux range, all the noise sources contribute equivalently, leading to a total noise of 0.01 \( \lambda \) (16 nm) rms in the presence of 600 e\(^{-}\) per pixel. At high flux, the error measurement reaches a plateau of 2 \times 10^{-3} \( \lambda \) (3 nm) rms for 1 s exposure and 5 \times 10^{-3} \( \lambda \) (1 nm) for 10 s exposure, representing the time-lag part of the AO residuals that dominates the noise in this range. It should be noted that this limit is set by the XAO system rather than the Zernike sensor and that, as mentioned above, techniques such as predictive Fourier control method for AO control are expected to efficiently reduce such AO residuals (Poyneer et al. 2007). Implementation of this concept in SPHERE as an upgrade path appears as an efficient and

| Table 1. Parameters used for estimating the error budget. |
|----------------|----------------|----------------|
| Parameters      | Relative value | Physical value |
| Zero mag. flux density in H-band | 1080 Jy | |
| Telescope diameter \( D \) | 8 m | |
| Telescope transmission \( T_{\text{int}} \) | 40% | |
| Beamsplitter transmission \( T_{\text{BS}} \) | 5% | |
| Mask design wavelength \( \lambda_0 \) | 1.625 \mu m | |
| Focal ratio F/# | 40 | |
| Mask diameter \( d \) | 1.062 \( F_0 \) | 69.0 \mu m |
| Mask phase shift \( \theta \) | \( \pi/2 \) | |
| Refractive index \( n \) (Fused silica) | 1.4431 @ \( \lambda_0 \) | |
| Mask depth \( z \) | 0.25 \( \lambda_0/(n - 1) \) | 0.917 \mu m |
| Filter wavelength \( \lambda_0 \) | 1.625 \mu m | |
| Filter bandwidth \( \Delta \lambda \) | 20% | 0.325 \mu m |
| DM inter-actuator pitch \( p \) | 20 cm | |
| AO residual amount | 0.055 \( \lambda_0 \) | 90 \mu m |
| Wind speed \( v \) | 10 m s\(^{-1}\) | |
| HAWAII-2RG readout noise | 18 e\(^{-}\) | |
| HAWAII-2RG pixel full well | 10\(^3\) e\(^{-}\) | |
| Pupil size | 40 \times 40 pixels | |

Notes. The term \( S_0 \) denotes the average entrance pupil flux. \((a)\) From Sect. 2.3, \((b)\) From Table 1, \((c)\) From Sect. 4, \((d)\) From Sect. 5.

| Table 2. Error budget. |
|----------------|----------------|----------------|
| Error | Value in e\(^{-}\) |
| \( a \) Photon noise \( \sigma_\text{P} \) | \( \sqrt{0.5 S_0} \) | |
| \( a \) Readout noise \( \sigma_\text{R} \) | 18 e\(^{-}\) | |
| \( c \) AO residuals error \( \sigma_{\text{AO}} \) | \( 2\pi \times 5.4 \times 10^{-2} S_0 / \sqrt{N} \) | |
| \( c \) Chromaticity \( \sigma_\lambda \) | \( 2\pi \times 3.0 \times 10^{-5} S_0 \) | |

Fig. 14. Wavefront error measurement in the presence of different noise sources as a function of the number of photo-electrons for a 40 \times 40 pixel pupil size.
realistic means of obtaining subnanometric precision for NCPA compensation in the H-band, representing an improvement in terms of wavefront error by a factor of 3 to 10. According to Perrin et al. (2003), the residual speckles in a highly corrected coronagraphic image are proportional to the power spectral density of the wavefront, hence proportional to the square of the rms wavefront error by a factor of 3 to 10. According to A94, page 10 of 11 of these quasi-static aberrations due to chromatic dispersion in SPHERE images by a factor ranging from 10 to 100.

7. Conclusion

In an exoplanet imager, non-common path aberrations between the visible XAO sensing path and the NIR scientific path induce static aberrations and demonstrated its efficiency for online measurements of static or quasi-static aberrations in a typical XAO system. In the presence of AO residuals of 81 nm, representing the phase-contrast method of Zernike (1934) in the context of exoplanet direct imaging. This system uses a focal plane wavefront error by a factor of 3 to 10. According to A94, page 10 of 11 of these quasi-static aberrations due to chromatic dispersion in SPHERE images by a factor ranging from 10 to 100.

We established a formalism for this approach, considering the spatial variability of the wave diffracted by the mask. The importance of this parameter, often ignored or simplified in classical textbooks (Hecht & Zajac 1987; Malacara 1992; Goodman 1996; Born & Wolf 1999) and by contemporary authors (Wallace et al. 2011), is underlined here, allowing us to reach an accurate calibration of residual wavefront errors with the Zernike sensor. A quasi-linear relationship between the NCPA map and the intensity variations in the exit pupil is obtained with this approach, providing a simple algorithm for the reconstitution of a static phase map.

We analyzed the accuracy of the Zernike sensor for calibrating static aberrations and demonstrated its efficiency for online measurements of static or quasi-static aberrations in a typical XAO system. In the presence of AO residuals of 81 nm, representing the phase-contrast method of Zernike (1934) in the context of exoplanet direct imaging. This system uses a focal plane wavefront error by a factor of 3 to 10. According to A94, page 10 of 11 of these quasi-static aberrations due to chromatic dispersion in SPHERE images by a factor ranging from 10 to 100. The simplicity of the Zernike sensor design makes its implementation possible as an upgrade path for the forthcoming E-ELT (Kasper et al. 2008). Its ability to work in the presence of a telescope aperture with segmented primary mirror, central obstruction, and spider arms makes the technique particularly promising for complex aperture geometries.

Acknowledgements. This research was supported by funding from CNRS-INSU and the Institute Carnot STAR. The authors would like to thank the anonymous referee for the careful reading, suggestions, and comments on the manuscript, Jean-François Sauvage and Laurent Mugnier for fruitful discussions as well as Marc Ferrari for his support, the Région Provence-Alpes-Côte d’Azur and ONERA for financial support with B. Paul’s scholarship. M.N. acknowledges Laurent Pueyo for his insightful contributions to the manuscript and Rémi Soummer for his support. This work is partially based upon work supported by the National Aeronautics and Space Administration under Grant NNX12AG05G issued through the Astrophysics Research and Analysis (APRA) program.

References

Clampin, M. 2008, in SPIE Conf. Ser., 7010
Dohlen, K., Langlois, M., Lanzoni, P., et al., 2006b, in SPIE Conf. Ser., 6267
Fusco, T., Rousset, G., Sauvage, J.-F., et al., 2006, Optics Express, 14, 7515
Guyon, O., Martinache, F., Garrel, V., et al. 2010a, in SPIE Conf. Ser., 7736

A94, page 10 of 11
Surdej, J., Yaitskova, N., & Gonte, F. 2010, Appl. Opt., 49, 4052
Traub, W. A., & Oppenheimer, B. R. 2010, Direct Imaging of Exoplanets (S. Seager), 111