A Novel Wavefront Sensor for Interferometry

Nazim A. Bharmal[a], David F. Buscher[a], Christopher A. Haniff[a], and Justin I. Read[b]

[a]Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK
[b]Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

ABSTRACT

A new design for a wavefront sensor suitable for low-order low-light correction is shown. The hybrid modal sensor, the Nine Element (NE) sensor, is compared with a curvature sensor and quadcell under single aperture applications. The design of the NE sensor allows the use of readily-available array detectors. We discuss the optimisation of the design to maximise its performance with respect to the number of Zernike polynomials to detect and optical parameters, using a simulated annealing technique. Numerical simulations show the good SNR response low-light levels, and indicate a reduction in wavefront variance from $6.41\,\mathrm{rad}^2$ to $2.01\,\mathrm{rad}^2$. The sensitivity to tip/tilt errors is demonstrated to be comparable to a quadcell. Successful closed feedback loop operation results in corrected Strehl ratios of over 0.5. Improvements and future work is discussed.

Keywords: adaptive optics, optical interferometry, wavefront sensing

1. INTRODUCTION

Adaptive Optics has matured technologically in the last decade to become a integral part of new telescopes. From initial applications on 4m sized apertures, the science and engineering drivers for new developments are now focused on optimising image quality for apertures sizes of 8m and greater, a necessary step toward application on the next generation of large telescopes. Optical interferometers achieve their high resolution by an array of small telescopes. The first prototypes used apertures that were effectively immune to atmospheric phase errors (beyond piston and tip/tilt). However, new arrays under construction will extend aperture sizes to approximately 1m, where phase errors become an important issue in the design.

Since an interferometer does not suffer the loss of resolution from phase errors that a single aperture telescope would, the reduction in visibility, and subsequent drop in SNR, is the major problem to overcome. Simply retrofitting an existing AO system to an array telescope would restrict targets to the brightest due to minimum flux requirements. This requirements would effectively reduce the SNR of observations.

While spatial filters[1] can be a solution to the atmospheric phase errors, one problem is the coupling losses from phase aberrations. This can reduce throughput significantly, lowering the SNR. Most of the power in turbulence induced phase fluctuations is in the lowest spatial frequencies. This is where AO for an interferometer will provide the biggest gains, by correcting those errors which have the largest effect on beam propagation,[2] in conjunction with spatial filtering.

Compared to AO systems currently installed on single aperture telescopes, an interferometric environment will have a reduced number of photons to work with. This is partially due to the consequence of having small apertures, and we also wish to maximise the throughput into the science instruments. The regime in which we wish to apply AO can therefore be described as the low-light low-correction regime.

Wavefront sensors (WFSs) are the key to successful low-light low-correction AO, since these are the subsystems which are affected most by both constraints. Two categories of sensor exist, Modal and Zonal. The zonal approach is best typified by the Shack-Hartmann[3] sensor which measures the phase in separate spatial zones. Modal sensors operate by separating the phase into orthogonal spatial frequency modes and measuring the amplitude of each one, an example being the curvature wavefront sensor.[4]

In this paper we use numerical simulations to show how a new modal design, the NE sensor, obtains low-light low-correction, with the emphasis on stable feedback loop. Performance in comparison with existing wavefront sensor designs are examined and a discussion of possible further developments are made.

Further information contact: e-mail: Nazim Bharmal nab26@mrao.cam.ac.uk

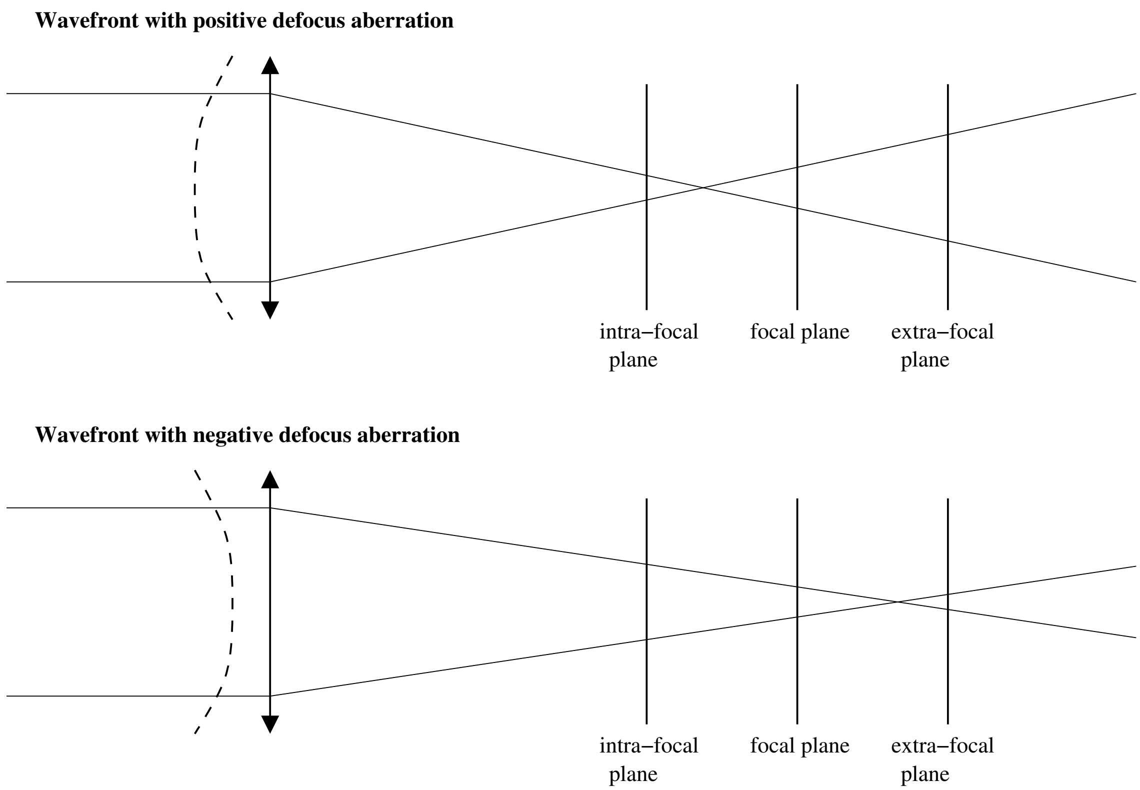

Figure 1. The degeneracy when attempting to measure defocus aberrations at the focal plane. At the focal plane, the same image is formed with either defocus aberration. The degeneracy is broken when taking measurements before (intra-focal) and after (extra-focal) the focus on the optical axis.

2. DESIGN AND IMPLEMENTATION

2.1. Theory

Over a circular aperture with no obstructions, we can decompose a wavefront using Zernike circle polynomials,[5] with a useful expressions given by Noll[6] for the typical variance given by each polynomial, see table 1.

Table 1. Residual variance after correction of successive Zernike polynomials, Δ_i in units of $(D/r_0)^{5/3}\mathrm{rad}^2$

$\Delta_1 = 1.03$	$\Delta_2 = 0.58$	$\Delta_3 = 0.13$	$\Delta_4 = 0.11$	$\Delta_5 = 0.09$	$\Delta_6 = 0.07$

Current interferometers use a tip/tilt system, which measures and corrects the first two Zernike polynomials (Z_2,Z_3) by using a Quadcell. This type of sensor is a 2x2 array detector which cannot detect further modes.

The next mode, defocus (Z_4), is detectable as an aberration at the focal plane but a degeneracy exists in the sign of the amplitude (figure 1) since an equal but opposite defocus produces the same intensity distribution. This degeneracy can be broken by making measurements out of focus, similar in concept to using a phase diversity technique. Changing the defocus is relatively easy to perform with optical components making it a practical implementation to simulate. The two equal but opposite defocused planes are called the intra-focal plane and extra-focal plane.

Using the defocus aberration as an example, it can be seen that a zero amplitude aberration gives two equally defocused images. A positive (in the sense of positive amplitude of the Z_4 term) defocus will produce a smaller,

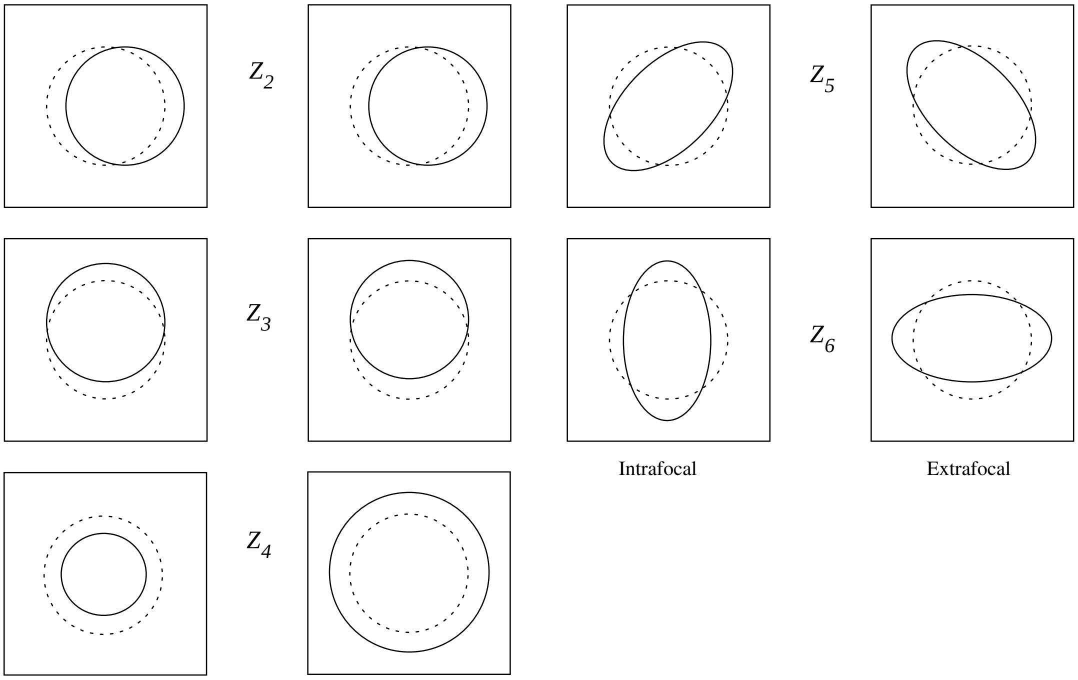

Figure 2. Distortion expected with the first 5 Zernikes after piston for the intra-focal and extra-focal planes. The solid line shows the outline of an Airy disc for a given aberration and the dotted line shows the same with no aberrations. Note that for Z_2 and Z_3, the aberrated image gives the same signal in both planes.

and thus more intense, image in the intra-focal plane and a larger, and less intense, image in the extra-focal plane. The converse is true for a negative defocus, a larger, less intense image in the intra-focal plane, and a smaller, more intense image in the extra-focal plane. By taking a differential measurement of the intensity in the two planes the amplitude of the aberration can therefore be determined.

A similar result applies to the astigmatic aberrations (Z_5,Z_6). By defining complementary areas on each plane, we can distinguish between particular aberrations and a pictorial summary of the signals measurable is shown in figure 2. We define the areas to be square pixels arranged as a 3 $\times$ 3 array, with the width of a pixel being of the order of half an Airy diameter, which corresponds to $\simeq 0.5\lambda f$, where f is the focal ratio of the optics concerned. This can allow easy use of an arbitrary sized array detector by simply altering the f ratio. The pixels can then correspond directly to elements of an array detector such as a CCD. Changing the size of a pixel with fixed f ratio changes both the sensitivity and linearity of the sensor.

This new configuration, dubbed the Nine Element (NE) sensor (figure 3) now has the ability to detect the changes in size of the image in horizontal, vertical, and both diagonal axes. The first two Zernike polynomials, Z_2 and Z_3, give the same displacement of the image in the intra-focal and extra-focal planes. The differential measurement between corresponding detector pixels is therefore zero for these aberrations. However, the sum of intensity in both planes will null the other aberrations and maintain the displacement. By measuring the difference between pixels either side of the central pixel, for vertical and horizontal directions, the displacement can be obtained (which we call $\mathbf{\Delta x}$ and $\mathbf{\Delta y}$). The method chosen here actually returns three displacements for each direction.

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix} = \begin{pmatrix} P_3 - P_1 \\ P_6 - P_4 \\ P_9 - P_7 \end{pmatrix}, \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \Delta y_3 \end{pmatrix} = \begin{pmatrix} P_1 - P_7 \\ P_2 - P_8 \\ P_3 - P_9 \end{pmatrix} \tag{1}$$

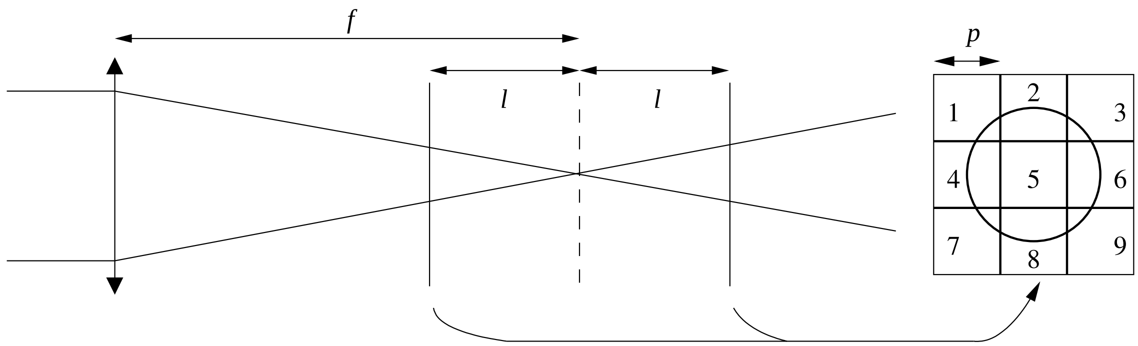

Figure 3. Optical layout of NE sensor, the parameters l and p are as described in the text, the circle represents the outline of an Airy disc at the focus. The diagram is not shown to scale. The number refer to individual detector elements, referred as P_n in the text.

where P_i refers to the summed intensity of array element i in both planes. From the sensor, a measurement vector **a** of length 15 is returned, which contains 9 differential and 6 displacement values. Compared to the Quadcell, we have increased the pixels over which the light is distributed from 4 to 18 (9 in each plane), and have increased the number of detectable Zernike polynomials to 5.

2.2. Linearity

To recover the Zernikes amplitudes **z** from the measurements **a**, we first make the assumption of linearity. For small amplitudes ($< 1\,rad$), this is valid but we must be careful to assess the range for which this is true. These assumption imply there exists a matrix **M** such that $\mathbf{z} = \mathbf{Ma}$.

Constructing **M** is more difficult than its inverse $\mathbf{M}^{-1}$, which can be obtained from

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{z} \tag{2}$$

and letting $\mathbf{z_i} = 0.1\delta_{ik}$. The index k then refers to a list of Zernikes we wish to attempt to measure. We can construct $\mathbf{M}^{-1}$ row-wise using each Zernike polynomial individually with a suitably small amplitude of 0.1 rad.

The inverse of **M** may not exist, and this can be determined by the singularness of $\mathbf{M}^{-1}$. The inversion of this matrix and the measure of its singularity can be done simultaneously by a Singular Value Decomposition[7] (SVD) algorithm. This procedure decomposes any matrix, **A**, into three components, **U**, **W**, and **V**. Two matrices, **U** and **V**, are unitary and **W** is pseudo-diagonal (some elements on the diagonal may be zero). Inversion is done by standard techniques except setting elements which are zero on the diagonal of **W** to zero in $\mathbf{W'}^{-1}$, equivalent to reducing the range of $\mathbf{M}'$. The resultant matrix give the following relationship,

$$\mathbf{z}' = \mathbf{M}'\mathbf{a} \tag{3}$$

where $\mathbf{z}' \to \mathbf{z}$ as $a \to 0$.

The reduced range of $\mathbf{M}'$ corresponds to aberrations which are unreachable through the measurement space and hence were not useful anyway. This singularity information can be used to therefore select those Zernikes which can be usefully detected i.e. not marginally detected. and those which are best ignored.

The range of linearity for a given Zernike polynomial is obtained by inputting that aberration with varying input amplitudes and comparing this to detected amplitudes. This range is taken as adhoc; as long as the sensor does not return amplitude values larger or opposite in sign from those input then they will not contribute to instabilities in a feedback loop.

2.3. Noise

Using SVD, the algorithm assumes a variance of 1 on each row of the initial matrix and subsequently performs a least-squares fit equivalent with the decomposed matrices. The decomposition allows an estimate of the variance on each element of the returned vector $\mathbf{z}'$ since the noise adds in quadrature and depends only on the scaling of vectors by $\mathbf{M^{-1}}$. By scaling rows of $\mathbf{M'}^{-1}$ appropriately, effects of detector noise and photon noise can be obtained quickly without the need for Monte Carlo-type simulations. Early optimisation steps can be performed quickly to remove those modes which may suffer from excess noise, implying low sensitivity, as mentioned in the previous section. The quoted noise values are for a signal of 0.5 photons per defocused image plane i.e. an input of one photon.

2.4. Analysis & optimisation

Once the matrix $\mathbf{M}'$ has been obtained, the NE sensor's performance can be quantified. Noise on measurement has been described above, as has linearity in the measurement of individual Zernikes. Two further parameters are critical to successful operation; the cross-talk between detectable modes and the aliasing of higher frequency modes onto those detected.

The wavefront variance can be better quantified by the Strehl ratio which is a standard measure of wavefront aberration in an imaging system. As is usual in AO analysis, imaging to obtain the Strehl ratio and the wavefront sensing are done at two different wavelengths to take advantage of the $\lambda^{6/5}$ scaling.

Evolving the phase screen temporally, by using Taylor's hypothesis of a frozen layer of turbulence moving past an aperture, can be used in conjunction with a closed loop (feedback) system to obtain a realistic operational model. Iterating allows an averaged corrected and uncorrected Strehl ratio with other parameters to be determined. The simulation code was written in a modular fashion, allowing the change of wavefront sensor type without changing other components. This was used to perform the comparison of the results against established designs.

By fixing the focal length of the system at f/30 with a 1 m diameter aperture, the only two parameters to optimise upon were sensor pixel size, p, and length of defocus, l. Using a fixed size computational grid with a fixed aperture on the grid, the sensor pixel size given is in fractions of an Airy disk diameter and the defocal length in metres.

The optimisation process was carried out using a simulated annealing algorithm[7] which used an adhoc cost function. To calculate the cost, a generic algorithm was employed which took a set of consecutive Zernikes, from Z_2 to Z_{16} with the same amplitude, and input each as wavefront phase individually. From this a $M \times N$ matrix $\mathbf{Q}$ of values was created, each row of which consists of the difference between measured and input amplitudes for the measured Zernikes. The dimension M equals the 14 Zernikes input and dimension N equals the number of measured Zernikes.

$$\mathbf{Q}_{ij} = |\mathbf{z}'_{j,Z_i} - \mathbf{z}_{j,Z_i}| \tag{4}$$

with $\mathbf{z}_{j,Z_i} = \delta_{ji} \times a$, given an amplitude a for an input of Z_i.

Considering just the the measurable Zernikes, we can take the upper N rows to form a $N \times N$ matrix $\mathbf{Q}'$. The trace of $\mathbf{Q}'$ will reveal the departure from linearity and the off-diagonal components will be non-zero for cross-talk between measurable polynomials.

The remaining $M - N \times N$ matrix $\mathbf{Q}''$ is representative of aliasing and any components which are non-zero are a measure of this effect. This places the requirement that $M > N$ which is clearly achieved for $M = 14$ with this sensor. A sum of components this matrix is a measure of aliasing.

$$\mathbf{Q} = \left(\begin{array}{c} \left(\begin{array}{c} \uparrow \\ Q' \\ \downarrow \end{array} \right) \\ \left(\begin{array}{c} \uparrow \\ Q'' \\ \downarrow \end{array} \right) \end{array} \right)$$

To calculate the linearity cost we use an amplitude of 1 rad per Zernike and return the trace of $\mathbf{Q}'$, and for aliasing we use an amplitude of 2 rad and return the sum of the elements of $\mathbf{Q}''$. Cross-talk was ignored since it was found to be negligible even at high amplitudes ($\gg 1$ rad). The total cost is the sum of both costs in quadrature.

3. RESULTS

3.1. Optimisation

3.1.1. Noise

From noise results the first 5 Zernikes, Z_2 to Z_6 are most favourably detected with a noise of 2.47 rad. Including coma (Z_7 and Z_8) increases the noise to 14.93 rad and modes higher than Z_9 would be undetectable due to the limited number of pixels. The explanation for the poor sensitivity with coma is that the Z_2 and Z_3 polynomial terms are present in the Z_8 and Z_7 polynomial terms, respectively, creating significant confusion between them.

Table 2. Noise in rad^2 for the first 5 Zernikes. The input signal is 1 photon. Total noise is 2.47 rad rms

Z_2	Z_3	Z_4	Z_5	Z_6
1.74	1.74	0.20	1.63	0.79

3.1.2. Simulated annealing

The optimisation procedure used was not guaranteed to give the optimal solution. However using the parameters returned, table 3, to examine linearity and in closed-loop simulations, gave results that were no worse than those found by manually searching parameter space. They can be taken as acceptably optimal but clearly further work remains in this area.

Table 3. Parameters used in numerical simulations of the NE sensor. Pixel size p is in Airy disc diameters

p	0.5	computational array size	128×128
l	1.5 m	detection wavelength	700 nm
f/#	30	imaging wavelength	1 μm
aperture diameter, D	1 m	r_0	$D/3.0$
wind velocity, v	3 ms^{-1}	feedback iteration time, T	$r_0/10v$
detector noise	10e$^-$		

3.1.3. Linearity

The linearity is shown in figure 4 and can be seen to be better for Z_2 and Z_3 than the other polynomials. However, the turnover occurs for the astigmatic modes, Z_5 and Z_6 at 1.4 rad and 1.1 rad respectively and for Z_4 at 0.9 rad. These may seem initially disappointing, but in a closed feedback loop the amplitudes being detected are planned to be < 1 rad and therefore these limits are acceptable. The difficulty may come from initial measurements to close the loop and dealing with large departures from the predicted amplitudes. Although not investigated, an adaptive algorithm may have to be employed to prevent instabilities with strong turbulence.

3.2. Closed-loop

A wavefront sensor is typically engaged in a positive closed feedback loop, and to simulate this the NE sensor was input with a temporally evolving phase screen. The screen simulates turbulence with parameters in table 3 corresponding to those for a site and instrument such as COAST.[8] The aperture phase is extracted using a circular mask, offset from the origin, on a single square phase screen (length $\gg$ aperture mask diameter),

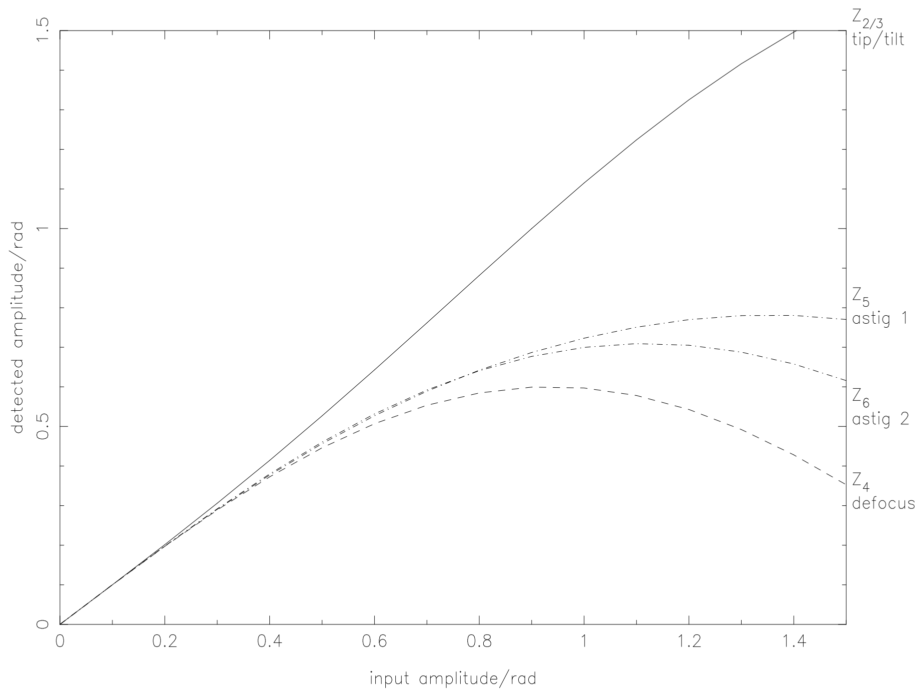

Figure 4. Measured amplitudes versus input amplitudes for the first 5 Zernike polynomials, Z_2 to Z_6.

constructed via a Fourier transform technique.[9] The mask is defined as the circular aperture (no obstructions) and the offset is updated with each iteration to simulate the passing of the phase screen at constant velocity across the aperture. The correction "element" is a simply a sum of Zernikes. For comparison, using identical turbulence and temporal parameters, a quadcell (QC) and a curvature wavefront sensor (CWFS), with 19 radial sensor elements, were also simulated.

Table 4. Comparison of the NE sensor with quadcell and curvature wavefront sensor for varying photon numbers. The noise incorporated both detector and photon noise terms. The photon rates are those input into the sensor, not at the imaged planes. The uncorrected variance is 6.41 rad^2. The corrected Strehl ratios are those for residual tip/tilt after correction and with the residual removed. The CWFS and NE failed to converge with a 10^2 photon flux.

Sensor	**Number photons/iteration**	**Corrected Strehl** with $Z_{2/3}$	without $Z_{2/3}$	**Variance remaining** rad^2
CWFS	10^6	0.532	0.792	1.57
	10^3	0.435	0.700	2.36
	10^2	-		
QC	10^6	0.474	0.589	1.89
	10^3	0.473	0.589	1.90
	10^2	0.458	0.584	2.12
NE	10^6	0.545	0.643	2.02
	10^3	0.518	0.624	2.01
	10^2	-		

The results, table 4, show the expected intermediate performance of the NE sensor. Compared to quadcell, the Strehl ratios are higher but the variance is lower. The NE sensor failed to close the loop for 10^2 photons/iteration, as did the CWFS. That intensity of photons is the same as or below the simulated detector readout noise of $10\,e^-$ per sensor pixel for the two sensors so the failure is not surprising.

The CWFS shows better performance at high light levels than the NE sensor since it can detect 19 Zernikes, but the tip/tilt performance is relatively poor. Considering that these are the aberrations with the highest amplitudes, we can see the QC, which is just a tip/tilt detector, has a similar improvement to the NE sensor with the removal of remaining tip/tilt errors. This suggests that the NE sensor is at least as good as the quadcell in correcting tip/tilt. The drop in performance when operating with 10^3 photons/iteration is greater with the CWFS than the NE sensor, suggesting the NE sensor is more robust to noise and therefore better suited to low-light conditions.

3.3. Aliasing and higher-order Zernikes

Although not explicitly considered here, the effect of high-order Zernikes present in the atmospheric phase screen do not alias to an extent that causes errors preventing the closure of the feedback loop. This can be understood by considering the aberrations caused in the focal plane, close to where we are taking images. The high-order aberrations are converted to high frequency aberrations which are averaged out over the size of the detector pixels we use. Therefore by choosing a relatively low-resolution array, we are biased toward those low frequency aberrations we wish to detect and can ignore otherwise noisy signals.

4. CONCLUSION

The numerical simulations performed with the NE sensor show that this design should be suitable for use in the low-light level environment of an interferometer to perform low-order correction. The detection of the first 5 Zernike polynomials can be made reliably with a relatively straightforward design, which is implementable with commonly available array detectors, such as CCD cameras. The design does not have any major impact on the optical design of a telescope, requiring only an input of the beam received from the sky.

The performance of the sensor is expected to work best in conjunction with a spatial filtering system where the combination will maximise the useful light throughput and SNR, compared to application of either technique alone.

The optimisation technique used did not necessarily converge on the optimal solution and prior to work on a prototype, this problem must be solved or at least the range of error in optimisation more quantitatively established.

ACKNOWLEDGMENTS

Thanks are due to Justin Read for initial code used in the development of this paper and François Rigaut for making publicly available his CWFS simulations on the Internet.

REFERENCES

1. J.W.Keen, D.F. Buscher, and P.J. Warner *Mon. Not. R. Astron. Soc.* **326**, pp. 1381–1386, 2001.
2. D. Buscher and S. Shaklan, "Low-order adaptive optics and single-mode fibers in stellar interferometry," in *Adaptive Optics in Astronomy*, *Proc. SPIE* **2201**, pp. 260–271, 1994.
3. J. Hardy, *Adaptive Optics for Astronomial Imaging*, OUP, Oxford, 1997.
4. F. Roddier, C. Roddier, and N. Roddier, "Curvature sensing: a new wavefront sensing method," *Proc. SPIE* **976**, pp. 203–209, 1988.
5. M. Born and E. Wolf, *Principles Of Optics*, Pergamon Press, New York, 1980.
6. R. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**(3), pp. 208–211, 1976.
7. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes*, CUP, Cambridge, 1988.
8. J.E. Baldwin, C.Roger, G.C. Cox, C.A. Haniff, J. Rogers, P.J. Warner, J. Peter, D.M. Wilson, and C.D. Mackay in *Amplitude and Intensity Spatial Interferometry II*, *Proc. SPIE* **2200**, pp. 118–128, 1994.
9. B. McGlamery, "Computer simulation studies of compensation of turbulence degraded images," in *Image processing*, *Proc. SPIE* **74**, pp. 225–233, 1976.