COMPARATIVE THEORETICAL AND EXPERIMENTAL STUDY OF A SHACK-HARTMANN AND A PHASE DIVERSITY SENSOR, FOR HIGH-PRECISION WAVEFRONT SENSING DEDICATED TO SPACE ACTIVE OPTICS

A. Montmerle Bonnefois[1], T. Fusco[1,2], S. Meimon[1], V. Michau[1], L. Mugnier[1], JF Sauvage[1]
C. Engel[2], C. Escolle[2], M. Ferrari[2], E. Hugot[2]
A Liotard[3], M Bernot[3], M Carlavan[3], F Falzon[3]
T. Bret-Dibat[4], D. Laubier[4]

[1] *ONERA, France,* [2] *Laboratoire d'Astrophysique de Marseille (LAM), France,*
[3]*Thalès Alenia Space (TAS), France,* [4] *CNES, France*

I. INTRODUCTION

A. *General points*

Earth-imaging or Universe Science satellites are always in need of higher spatial resolutions, in order to discern finer and finer details in images. This means that every new generation of satellites must have a larger main mirror than the previous one, because of the diffraction. Since it allows the use of larger mirrors, active optics is presently studied for the next generation of satellites. To measure the aberrations of such an active telescope, the Shack-Hartmann (SH), and the phase-diversity (PD) are the two wavefront sensors (WFS) considered preferentially because they are able to work with an extended source like the Earth's surface, as well as point sources like stars.

The RASCASSE project was commissioned by the French spatial agency (CNES) to study the SH and PD sensors for high-performance wavefront sensing. It involved ONERA and Thales Alenia Space (TAS), and LAM. Papers by TAS and LAM on the same project are available in this conference, too **[1,2]**.

The purpose of our work at ONERA was to explore what the best performance both wavefront sensors can achieve in a space optics context. So we first performed a theoretical study in order to identify the main sources of errors and quantify them – then we validated those results experimentally.

The outline of this paper follows this approach: we first discuss phase diversity theoretical results, then Shack-Hartmann's, then experimental results – to finally conclude on each sensor's performance, and compare their weak and strong points.

B. *Approach*

The modelisation of the direct problem had to be as realistic as possible to be relevant, so a diffractive model was used.

The wavefronts (and telescope pupils) came directly from TAS studies about lightweight mirrors for spatial active optics, and described realistically the kind of wavefronts the WFS will have to measure. They took into account the deformations of the mirror caused by gravity on the mirror structure, thermomechanical effects, and polishing residues.

The extended scenes we used were high-resolution air images provided by CNES. We considered realistic photometry conditions, with typical and unfavourable scene's mean luminances (resp. 14 W/m²/sr/µm and 50 W/m²/sr/µm), and realistic detector read-out noise, (sub)sampling, and transfer fonction. We considered the typical scanning speed (hence exposure time and motion blur), and resolution of a LEO satellite.

We didn't take into account the polychromaticity for the SH, since this sensor is achromatical – but we did consider the spectral bandwidth for the PD sensor. However, in this case, the scenes were supposed to be grey, which seemed a sensible hypothesis as long as our bandwidth was shorter than 100 nm.

We then used several variations of the inverse problem, ie. of the wavefront measurement based on the modeled images, in order to establish an error budget of the sensors.

Performance criteria were defined to establish these error budgets and WFS comparisons, for two types of measurements:

- Full wavefront measurement on point source (like stars or calibration sources): in this case, we considered that the active mirror could compensate 21 Zernike modes, and that our performance criterion had to reflect that fact. So the metric we used was the root-mean-square error on the 21 first Zernike modes (RMSE$_{21}$), defined as follow:

$$RMSE_{tot}^2 = RMSE_{21}^2 + RMSE_{HO}^2 \qquad (1)$$

Where $RMSE_{tot}$ was the total error (rms error between the true wavefront and the estimated wavefront), and $RMSE_{HO}$ the error on orders higher than 21.

- Low-order aberration measurement on extended scenes (earth satellite images): the metric we used was the error on defocus and comas modes ($RMSE_{4,7,8}$).

On extended scene, we supposed for both SH and PD that non-cooperative scenes, i.e. scenes with an insufficient high spatial frequency content, could be discarded before any phase reconstruction were performed, through simple algorithms based on their spatial frequency properties, so only cooperative scenes were taken into account.

Two wavefront maps were considered: Map 1and Map 2, Map 1 bis and Map 2 bis are deduced from Maps 1 and 2 (cf. Table 1).

| **Map 1**	RMS_{tot} = 74 nm rms, RMS_{21} = 57 nm rms	**Map 2**	RMS_{tot} = 127 nm rms and RMS_{21}= 104 nm rms
Map 1 bis	Map 1 HO + up to ±12 nm defocus or coma	**Map 2 bis**	Map 2 HO + up to ±104 nm defocus or coma

Table 1: RMS of the wavefront maps used in this study. RMS_{tot} is the total rms, RMS_{21} is the rms on the 21 first Zernike modes. HO stands for high-order residues of the map, i.e in our case Zernike modes >21.

II. PHASE DIVERSITY SENSOR: THEORETICAL STUDY

A. Data modelling

We made simulations of realistic phase diversity images in a spatial active telescope context. We study the impact of the following parameters: the type of the object (point source or extended scene), the spectral bandwidth (polychromatic simulation), the sampling (Nyquist and half Nyquist rates were considered). Moreover, realistic pupil and aberrant wavefront (see §1.), motion blur, and noise were considered.

Some examples of these images can be seen in Fig. 1.
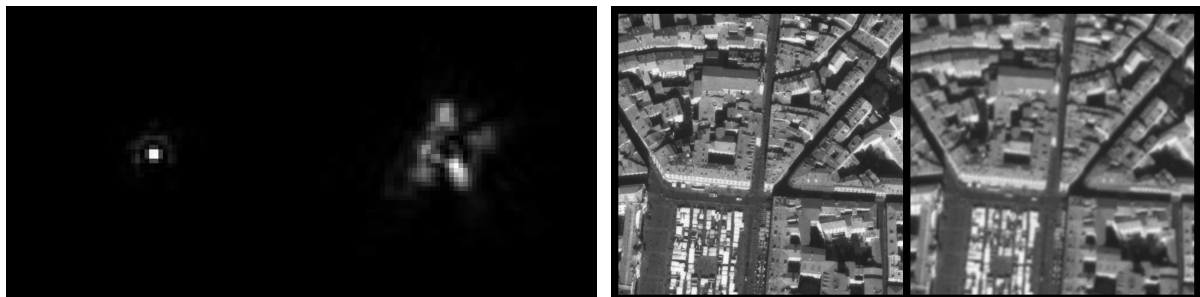
Fig. 1: Example of simulated phase diversity images – left: point source – right: extended scene

We then used phase diversity on these simulated images, and compared the estimated wavefronts with the true wavefront.

B. Data reduction

For spatial applications, the speed of the algorithms is an important factor. So a compromise must be found between performance and complexity of the algorithm. That's why point to point phase diversity algorithms **[3]** were not considered, as they require far too much execution time.

For point source phase diversity, we used algorithms with known objects – as the images can be subsampled by a factor 2. Aliasing is so important in that case that it has to be explicitly taken into account in the algorithms, along the lines of **[4]**. On the contrary, on extended scenes, the object is by definition unknown, but aliasing is generally not a problem due to the lack of high spatial frequency information in typical satellite earth images. So in the latter case, we up-sampled the images to the Shannon-Nyquist rate, then used phase diversity algorithms with both unknown object and wavefront **[5,6,7].**

Our phase diversity algorithm has two options to deal either with monochromatic or polychromatic PSF (with a known spectre and a grey object). Of course, the option that takes into account the spectral bandwidth is slower than the one supposing a monochromatic sensitivity.

C. Main conclusions and error budget

We tested our phase diversity algorithms on many images, simulated with varying parameters, and different versions of the algorithms (see §II.A.), in order to establish an error budget: errors result from the gap between the realistic direct problem and the simplified inverse problem. Our main conclusions are as follows:

Aliasing

In point source mode, the major source of error is by far the high-order modes of the wavefront, especially when they are localized, and not randomly distributed: mirror fixation devices (MFD) footprints on the wavefront for example can be particularly problematic. The bias due to the aliasing of these high-order modes on the lower orders depends on the number of Zernike modes used in the phase diversity algorithm (hence the complexity, and speed, of the algorithm), and on the nature of the wavefront. Instabilities (local minima) can also appear in some cases.

Hence the main conclusion of this theoretical study is that the frequency distribution of the residual wavefront of spatial active mirrors really matters and should be taken into account in their design

In extended source mode, since only very low orders are measured (from defocus to coma), this bias is negligible.

Polychromatic PSF

Taking into account the fact that the PSF is polychromatic in the phase diversity algorithm is possible but adds complexity – and computation time. We found that, with a 80 nm bandwidth, we could hope to gain at most 4 nm in point source mode by using a polychromatic algorithm – that is, if we know perfectly the spectral response of the star we observe. In extended source mode, the gain is negligible, and the grey object hypothesis is debatable.

Therefore, a simpler algorithm, using a monochromatic PSF, may be considered efficient enough depending on the aimed performances.

Noise

The standard deviation of $RMSE_{21}$ due to the noise draw in the images depends on the signal-to-noise ratio of the images.

In point source mode, there is no design constraint about the exposure time – as the whole range of the sensor can always be used.

In extended scene mode, the luminance of the scene is imposed, as well as the exposure time. The only parameters that can be optimized to improve the SNR of the images are the transmission of the phase diversity device (hence design optimization), and the spectral bandwidth.

We found that a 80-100 nm bandwidth is a good compromise between the validity of the "grey object" assumption (see §2.c) and the SNR of the images.

Error budget

We can resume the sources of error we identified as follows:
- b_{HO}: bias due to the aliasing of higher order modes on lower orders
- σ_N: standard deviation of the results due to the image noise **[9]**. It depends also on the scene spatial power spectral density, but in this work we always chose a favorable case.
- b_λ: bias due to the fact that the estimation of the wavefront was supposed monochromatic, while the images were polychromatic. This bias can be reduced, or even suppressed, by taking the polychromaticity into account in the phase diversity algorithm.

Other sources of error were proven to be negligible – in particular, results were the same with well-sampled and subsampled (by a factor 2) images.

Since these errors can reasonably be assumed to be independent, the error budget is:

$$RMSE_{21}^2 = b_{HO}^2 + b_\lambda^2 + \sigma_N{}^2 \tag{2}$$

The value of each of these errors depends on design hypothesis and wavefronts. By way of example, Table 1 gives the results we obtained with our parameters on point source, and Table 2 gives typical results with extended scenes

Type of object	Point source		Point source	
Subsampling factor	2		2	
Aberration (WFE)	Map 1		Map 2	
$RMSE_{tot}$	20 nm rms		30 nm rms	
$RMSE_{21}$	5,8 nm rms	b_{HO} = 2,8 nm rms	7,3 nm rms	b_{HO} = 5,8 nm rms
		σ_N = 3,2 nm rms		σ_N = 3,7 nm rms
		b_λ = 3 nm rms		b_λ = 3 nm rms

Table 2: typical results of our theoretical study on point source phase diversity (for the notations, see §IB and §IIC).

Type of object	Extended scenes	Extended scenes
Subsampling factor	2	2
Aberration (WFE)	Map 1 bis	Map 2 bis
$RMSE_4$	<1 nm rms, depending on the scene	<5 nm rms , depending on the scene
$RMSE_{4,7,8}$	<2,5 nm rms, depending on the scene (σ_N = 2 nm)	2 to 10 nm rms, depending on the scene (σ_N = 2 nm)
linearity	Result independent of the strength of the aberration on all the considered range.	Result independent of the strength of the aberration on all the considered range.

Table 3: Typical results of our theoretical study on extended scene phase diversity (for the notations, see §IB and §IIC).

III. SHACK-HARTMANN SENSOR: THEORETICAL STUDY

A. Data modelling

As for the PD sensor, we modeled Shack-Hartmann (SH) images to be as close as possible to experimental images. We took into account diffraction and central obstruction by a realistic spatial telescope pupil, individual spreading of subaperture's PSFs due to local variations of the wavefront to measure, field iris diaphragm, sensor noise, and sampling – Nyquist and half of Nyquist on the microlenses focal plane were considered. The same high-resolution earth images as the PD theoretical study were used (see §II), the final object resolution being adapted accordingly to the SH sensor design. An example of simulated Shack-Hartmann image can be seen in Fig. 2.

B. Data reduction

We tested several algorithms in order to establish an error budget: errors contributions results from the gap between the realistic direct problem and the simplified inverse problem.

Slopes measurement algorithms

We encountered no difficulty to measure slopes in the point source case – since the SNR is excellent, a simple center of gravity with threshold is efficient enough.

With extended scenes we have to deal with model errors : the usual way to deal with slopes measurement in this case is to measure the correlation peak between the image of a reference subaperture and the image of the aperture where the slope has to be measured **[10,11]**. This assumes that both images are identical, which is a valid hypothesis in most applications where high sensitivity is not required. However in this context, the images are too different because of the partial occultation of some sub-apertures by the pupil, and of the local differences of the wavefront. In the end, with this method the sensitivity of the SH sensor is not compatible with the high-performance requirement (this single post of error is > 10 nm rms).

There are two ways to deal with this problem:
- We can keep only the sub-apertures without occultation by the telescope pupil: the advantage is that calculations remain simple. The drawback is that a lot of information has to be discarded this way, and the number of sub-apertures may have to be increased in order to keep the required sensitivity. Localized wavefront errors can also be lost with this method,

- Or we can measure the intercorrelation, not between two images, but between crossed-convolutions of images and PSFs: indeed, if we consider the ith sub-aperture, where the slope δ is measured, and a reference sub-aperture, then we have:

$$image_i * PSF_{ref} = image_{ref} * PSF_i * \delta \tag{3}$$

i.e the crossed-convolution are identical, apart for the shift we want to measure, δ.

The PSFs can be calibrated before the satellite is launched, or on board thanks to a reference point source. With this method, the calculations for one sub-aperture are a bit longer, since convolutions have to be made before the correlation maximum calculations. But on the other hand, this method allows us to keep a small number of sub-apertures – hence a more compact sensor and a more effective use of the available flux (see 3.d).

Wavefront reconstruction

We mentioned in §IIB that local high-order aberrations added complexity to phase-diversity wavefront sensing. We made the same observation when we reconstructed the wavefront from the measured slopes. Calculating the Zernike-to-slope matrix, M, is immediate. However, questions arise when we have to invert this matrix: we found that the usual maximum-likelihood inversion can lead to very important errors, because of the aliasing of higher order modes upon the low-order modes.

The best way to reduce, or even avoid aliasing is to use a maximum a posteriori (MAP) inversion for M matrix **[12]**.For the *a priori*, we only supposed a quadratic decrease of the PSD of high-order modes.
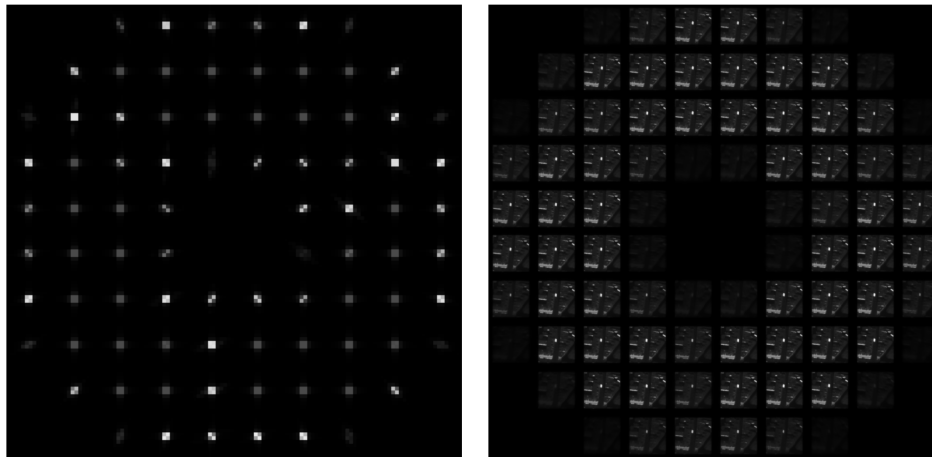
Fig. 2: example of simulated Shack-Hartmann images – left: point source – right: extended scene

C. *Theoretical study conclusions: Error budget and design considerations*

The sources of error we established on point source measurement were:
- b_{model}: bias due to the fact that slope measurement algorithms use simplifying hypothesis of the SH imaging. We lowered significantly this bias value by using crossed convolutions (see Eq 3 and Table 5)
- b_{HO}: bias due to the aliasing of higher order aberration modes on lower orders due to the subsampling of the wavefront by the microlenses array. With the wavefront maps we used, the usual maximum likelihood wavefront reconstruction method leaded to a very poor sensivity – while MAP provided the best performance.
- σ_N: standard deviation of the results due to the image noise It depends also on the scene spatial power spectral density, but in this work we used favorable pictures (see **[13]** for a study of σ_N).

Since these errors can reasonably be assumed to be independent, the error budget is:

$$RMSE^2 = b_{HO}^2 + b_{model}^2 + \sigma_N^{\ 2} \tag{4}$$

The value of each of these errors depends on design hypothesis and wavefronts. By way of example, Table 4 and Table 5 give the results we obtained with our parameters.

Once the algorithms for the sensor have been optimized, the overall performance of the SH sensor is only a matter of design choices: number of sub-apertures, sampling in the focal-plane of the microlenses, and, in the case of extended sources, field of view of the sub-apertures and details of the scene.

Table 4 and Table 5 show how the sampling in the focal plane and the number of microlenses influence both the sensitivity of the sensor and its size. It shows among other things that a subsampled, 10x10 Shack-Hartman has a good enough sensitivity, consistent with high-performance requirement, and by far the most compact design of the configurations we tested.

Type of object		Point source				Point source			
Aberration (WFE)		Map 1				Map 2			
Number of microlenses		10x10		16x16		10x10		16x16	
Subsampling factor		1	2	1	2	1	2	1	2
RMSE$_{21}$ (in nm rms)	b_{HO}	7	5,7	3	4,7	13,4	8,7	0,6	7,4
	b_{model}	<<	<<	<<	<<	<<	<<	<<	<<
	σ_N	1	1	1	1	1	1	1	1

Table 4: Typical results of our theoretical study on point source Shack Hartmann (for the notations, see §IB).

Type of object		Extended scene				Extended scene			
Aberration (WFE)		Map 1 bis				Map 2 bis			
Number of microlenses		10x10		16x16		10x10		16x16	
Subsampling factor		1	2	1	2	1	2	1	2
RMSE$_{4,7,8}$ (in nm rms)	b_{HO}	<<	<<	<<	<<	<<	<<	<<	<<
	b_{model}	<1	<1	<1	<1	<5	<3	<1	<1
	σ_N	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1

Table 5: Typical results of our theoretical study on extended scene for 3 modes reconstruction with a Shack-Hartmann (for the notations, see §IB). Our field of view was supposed to be 32 diffraction spots wide in the microlenses focal plane.

IV. EXPERIMENTAL RESULTS

A. Experiment

Once this theoretical study was finished, an experiment was performed **[2]** – one Shack-Hartmann sensor and one phase-diversity sensor were implemented. Two kinds of pupils were tested: a round pupil and a realistic telescope pupil (with central obscuration). The same aberrant wavefronts as in the simulations were carved on seven phase plates – one of them being a reference phase plane. These phase plates were previously measured with a Zygo interferometer and were therefore known with a nanometric precision. The Zygo measurements were used as our true wavefronts.

For a detailed description of the bench, see **[2].**

Since the bench had its own residual aberrations, which were unknown and changing during the day due to thermo-mechanical effects, all the measurements we present here are differential measurements, the seventh phase plate, the plane plate being used as a reference wavefront.

Note that the point source was polychromatic (white fibered source), while the extended scene had a 60 nm bandwidth (OLED screen).
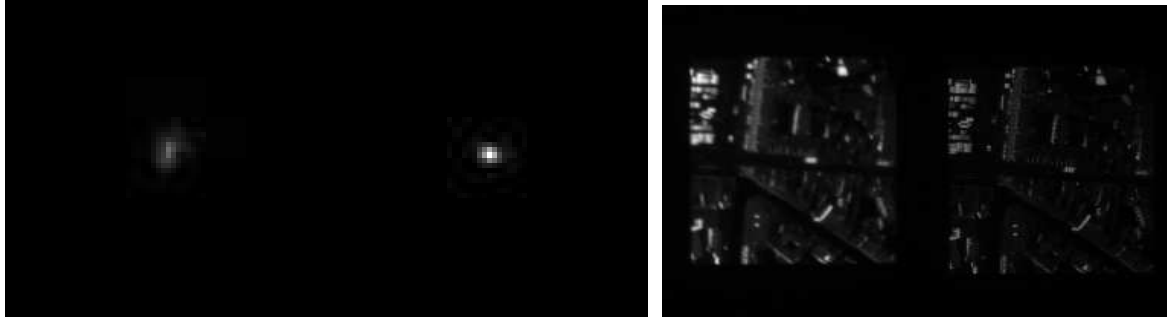
A. Phase diversity experimental measurements

Examples of phase diversity experimental images are shown Fig. 3. And experimental root-mean-square error between the true wavefront and the measurements are shown in Table 6 and Table 7.

The main difficulty we encountered was the apparition of instabilities when the phase diversity was not perfectly calibrated, due to local minima in the minimization process of the phase diversité. Therefore, the defocus between the two arms of the phase diversity sensor was measured using temporal phase diversity with a better than 2% precision – and similarly for the numerical aperture in the image space. We also had to use experimental images of the pupil (with central obscuration) and use a parametric model for the detector transfer function, in order to obtain the best sensitivity.

However, despite all these efforts, the experimental point source sensitivity was never better than 10-15 nm – which is good per se, but is higher than we expected according to the theoretical study.

Extended scene phase diversity wavefront sensing gave good results – the algorithms were robust, precision was better than 3 nm (on the 6 first Zernike modes) and linearity was better than 1%. The only remaining problem was the residual field aberrations of the bench – depending on the part of the field we use in the calculations, a disparity of the measured RMSE of up to 3 nm was observed because of them.
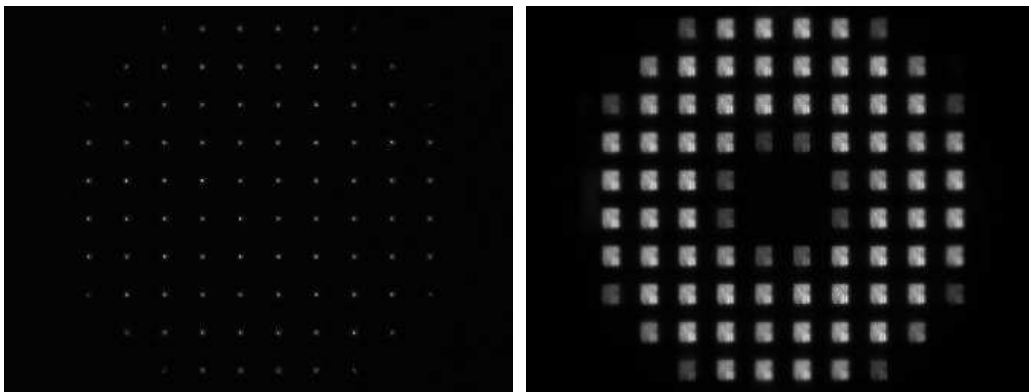

Fig. 3: example of experimental Shack-Hartmann images – left: point source – right: extended scene.

B. Shack-Hartmann experimental measurements

Examples of Shack-Hartmann experimental images are shown Fig. 4. And experimental root-mean-square error between the true wavefront and the measurements are shown in Table 6 and Table 7.

Shack-Hartman experimental data reduction was much easier than phase diversity's. Once the sensor was calibrated with standard procedures, results were immediate. It's interesting to note that the Shack-Hartmann sensor achieved the same sensitivity as the phase diversity sensor, both for point source and extended scenes.

Fig. 4: example of experimental Shack-Hartmann images – left: point source – right: extended scene

C. Comparison between both wavefront sensors

Both sensors provided the same performance experimentally.

Phase diversity requires a very simple experimental device, but the experimental parameters (numerical aperture, phase diversity, pupil and detector transfer function) must be calibrated carefully in order to provide good sensitivity. With some wavefronts, in our configuration, local minima can appear – but this could be avoided through further design and algorithmic optimization. Moreover, classical phase diversity algorithms remain too slow for space applications – real-time phase-diversity wavefront sensing [14] could be a solution, but needs further studies.

On the contrary, the Shack-Hartmann sensor's complexity is experimental. The design must be optimized depending on the required sensitivity, and needs an internal reference – while phase diversity's measurements are absolute. But then, the SH sensor is very robust experimentally, and the algorithms can be simple and efficient at once. And much quicker than phase-diversity's – which is an important factor for a spatial sensor.

Type of object	Point source		Point source	
Aberration (WFE)	Map 1		Map 2	
Sensor	SH	PD	SH	PD
$RMSE_{21}$	13 nm rms	14 nm rms	18 nm rms	12 nm rms

Table 6: experimental results with point source wavefront sensing, with Shack-Hartmann and Phase diversity.

Type of object	Extended scene		Extended scene	
Aberration (WFE)	Map 1 bis		Map 2 bis	
Sensor	SH	PD	SH	PD
$RMSE_{21}$	< 2 nm rms	< 1 nm rms	< 2 nm rms	< 2 nm rms

Table 7: experimental results with extended scene wavefront sensing, with Shack-Hartmann and Phase diversity.

V. CONCLUSION AND PERSPECTIVES

We showed theoretically and experimentally that Phase diversity and Shack-Hartmann sensors, whose sensitivities are comparable, are both precise enough for space active optics applications.

In the short term, we haven't managed (yet) to achieve experimentally the same sensitivity we had predicted theoretically. It means that new sources of error must be added to our error budget. New experiments and in-depth studies are planned in order to better understand the causes of these experimental errors, and how to minimize them. Possible sources of error include, among other things: calibration errors, phase plate positioning, field aberrations, and environment instability.

In the longer terms, both WFS would need further optimizations in order to be sent to space, *ie.* design simplification (for SH), acceleration of the algorithms (for PD) and robustness enhancement (for both). More, spatial active mirrors design and WFS should be co-designed in order to optimize the telescope overall performance: for instance, we showed that wavefronts with low-order aberrations are easier to measure and compensate than those which have localized, high-order ones.

REFERENCES

[1] A. Liotard et. al, *"wave-front sensing for space active optics : RASCASSE project",* this conference.

[2] C. Engel et al., *"The LAM space active optics facility",* this conference.

[3] L. M. Mugnier, A. Blanc and J. Idier., *"Phase Diversity: a technique for Wave-Front Sensing and for Diffraction-Limited Imaging,"* chap. 1 of Advances in Imaging & Electron Physics, Vol. 141, pp. 1-76, edited by Peter Hawkes, Elsevier, 2006.

[4] J. R. Fienup, *"Phase retrieval for undersampled broadband images"*, JOSA A 16(7): 1831–1837, 1999.

[5] A. Blanc*, "Identification de réponse impulsionnelle et restauration d'images : apports de la diversité de phase,"* PhD thesis, Paris 11, 2002.

[6] Idier, L. Mugnier, and A. Blanc, *"Statistical behavior of joint least square estimation in the phase diversity context,"* IEEE Trans. Image Processing, 14(12):2107–2116, December 2005.

[7] A. Blanc, L. M. Mugnier, and J. Idier, *"Marginal estimation of aberrations and image restoration by use of phase diversity"*, J. Opt. Soc. Am. A, 20(6):1035–1045, 2003.

[8] Escolle et al., *"Adapting large lightweight primary mirror to space active optics capabilities"*, Proc. SPIE (2014)

[9] Meynadier, L et al., *"Noise propagation in wave-front sensing with phase diversity",* Applied optics, 38(23), 4967-4979, 1999

[10] Michau, V., Rousset, G., & Fontanella, J. , *"Wavefront sensing from extended sources"* in Real Time and Post Facto Solar Image Correction (Vol. 1, p. 124), 1993

[11] L. Poyneer, *"Scene-based Shack-Hartmann wave-front sensing: analysis and simulation,"* Applied Optics 42(29), pp. 5807–5815, 2003.[12] Mugnier, L. M., Robert, C., Conan, J. M., Michau, V., & Salem, S. , *"Myopic deconvolution from wave-front sensing". JOSA A, 18*(4), 862-872, 2001.

[13] Michau, V. et al, *"Shack-Hartmann wavefront sensing with extended sources",* in SPIE Optics+ Photonics (pp. 63030B-63030B), International Society for Optics and Photonics, 2006.

[14] I. Mocœur, L. M. Mugnier, and F. Cassaing, *"Analytical solution to the phase-diversity problem for real-time wavefront sensing"*, Opt. Lett., 34(22):3487–3489, November 2009.