# End-to-end design of Wavefront Sensor and Reconstruction Using Neural Networks

Francisco OYARZUN[1] Pauline TROUVE[2] Frederic CHAMPAGNAT[2] Thierry FUSCO[1,2] Benoit NEICHEL[1]

[1]Laboratoire d'Astrophysique de Marseille, 38 Rue Frédéric Joliot Curie, 13013 Marseille, France

[2]DTIS, ONERA, Université Paris Saclay, 91123 Palaiseau, France

**Résumé –** Nous proposons une optimisation jointe des paramètres optiques d'un analyseur de front d'onde et d'un réseau de neurones utilisé pour analyser le signal et reconstruire le front d'onde. Nous montrons que cette approche permet d'obtenir de meilleures performances que l'entraînement du seul réseau de reconstruction.

**Abstract –** We propose a joint optimization of the optical parameters of a wavefront sensor and a neural network used to decode the signal and perform wavefront reconstruction. We show that it is possible to achieve better performance with this technique than training only the reconstruction network.

## 1 Introduction

Ground-based telescopes suffer from resolution loss due to atmospheric turbulence. Instead of achieving the diffraction-limited resolution $\theta_{\text{diffraction-limited}} = \lambda/D$, the effective resolution is degraded to $\theta_{\text{seeing}} = \lambda/r_0$, where $r_0$ is the Fried parameter [6], typically ranging from 3 cm to 15 cm in good observatories. This results in a resolution loss of 50 to 250 times for an 8 m telescope. To mitigate this, adaptive optics (AO) is used, consisting of a wavefront sensor (WFS) to measure aberrations, a deformable mirror (DM) to correct them, and a control system to compute corrections. This work focuses on optimizing the WFS and the phase reconstruction algorithm.

A WFS converts phase distortions $\phi(x, y)$ into an intensity signal $I(\phi)$ measurable by a detector. The transformation depends on the WFS type : Shack-Hartmann WFS [15] measures the wavefront gradient, Zernike WFS [22] encodes phase as a sinusoidal signal, and Pyramid WFS [16] relates intensity and phase via the Hilbert transform. While traditional reconstruction assumes a linear response, Fourier Filtering WFSs (FFWFS), such as the Zernike or Pyramid WFSs, [20, 1] relate the input phase and the intensity in the detector with the following non-linear relation [4] :

$$I(\phi) = \left|\mathcal{F}\{\mathcal{F}\{\mathbb{I}_p e^{i\phi}\} \cdot m\}\right|^2, \tag{1}$$

where $\mathcal{F}\{\}$ is the Fourier transform, $\mathbb{I}_p$ the pupil indicative function and $m$ is a filtering mask. While these sensors provide high robustness to noise and linear response for small phases, their nonlinear response to high amplitudes of phase requires advanced reconstruction techniques such as iterative solvers [3] or machine learning methods.

Neural networks (NNs) have shown promise as wavefront reconstructors. Previous studies have optimized either the reconstructor [10, 21] or the optical parameters of the WFS [7] separately. In this work, we propose a co-design method where the reconstructor and the optical parameters are optimized. This approach allows the system to adapt its optical design for better signal codification in the WFS, which is tailored to be decoded by the network. Figure 1 shows the general schematic used in out system. A similar co-design framework has been explored in the context of 3D passive imaging, where both optical parameters and image processing algorithms were jointly optimized to enhance depth estimation from defocus [19].

The masks selected for optimization are the Pyramid and Zernike WFS masks. The Pyramid mask was chosen due to its central role as the primary WFS in the single-conjugate AO mode of the ELT instrument HARMONI [14]. The Zernike mask was selected for its growing popularity as a second-stage WFS in extreme AO systems, particularly in the context of direct exoplanet imaging [13]. In this study, we utilize WFSs beyond the small-phase regime, where linear reconstruction is typically valid. To ensure robustness, we train our system using full atmospheric turbulence scenarios, covering both high and low $r_0$, as well as open- and closed-loop data.

## 2 Methods

### 2.1 Wavefront sensor modeling

To simulate the wavefront sensor (WFS), we model the electromagnetic field as a discretized wavefront with a telescope aperture of 1.5 m and a sensing wavelength of 500 nm [11]. Following Eq. 1, the signal is determined by the choice of the filtering mask $m = e^{i\Delta}$. For a Zernike wavefront sensor, the mask consists of a central dot with depth $\alpha$ and diameter $\beta$, approximated as :

$$\Delta_{zernike}(x, y) \approx \frac{\alpha}{2}\left[tanh\left(\gamma\left(\frac{\beta}{2} - \sqrt{x^2 + y^2}\right)\right) + 1\right], \tag{2}$$

with $\gamma \gg 1$ to emulate a discontinuous function. The normal values for these coefficients are $\alpha = \pi/2$, and $\beta = 1\,\lambda/D$. For the Pyramid WFS, the mask is a four-sided pyramid with slopes $\alpha$ and $\beta$ :

$$\Delta_{pyramid}(x, y) = |\alpha x| + |\beta y|. \tag{3}$$

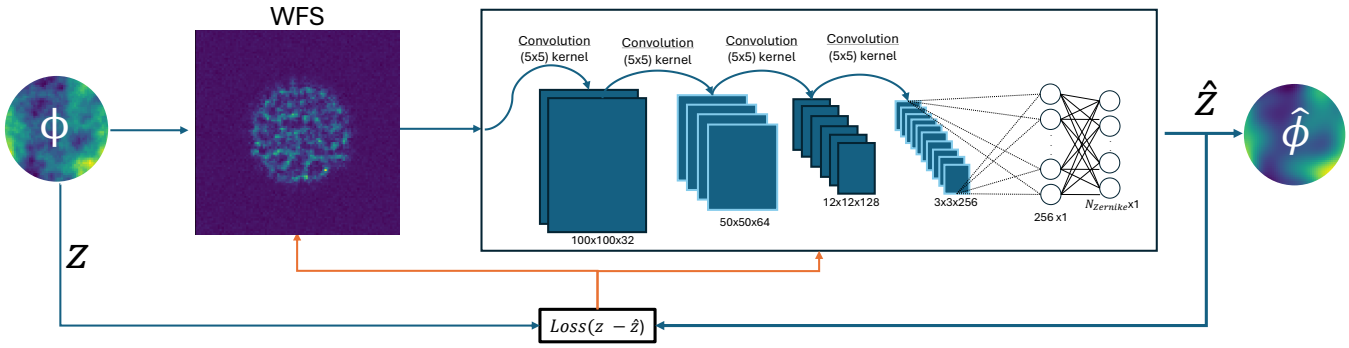


FIGURE 1 : Diagram of the system. The input phase $\phi$ is propagated through the WFS, and the image obtained is the input of the NN. The design of the NN includes also activation functions and Max-Pooling between every layer. The sizes of the features are given in height x width x channels.

The normal values for these coefficients are $\alpha = \beta = \pi/2$. In this work, instead of using the values of $\alpha$, $\beta$ and $\gamma$ as fixed, we optimize them jointly with the reconstruction network, which will be described in the next subsection.

### 2.2 Neural network reconstructor

We use a Convolutional Neural Network (CNN) for wavefront reconstruction. The network, as shown in Fig. 1, processes a $100 \times 100$ grayscale image using four convolutional layers with 32, 64, 128, and 256 filters, applying GELU activation and max-pooling for downsampling after each layer. An adaptive average pooling layer extracts global features, followed by a fully connected layer that maps them to the first $N_{\text{Zernike}}$ Zernike modes.

### 2.3 Dataset

To generate the dataset to train the system, we take advantage of the fact that the wavefront has a known power spectral density (PSD), which we can use to generate an arbitrarily large number of samples. The PSD of the wavefront can be constructed by combining the PSD of the atmosphere, which follows a Kolmogorov distribution [9], the PSD of the deformable mirror, which corresponds to a high-pass filter [17], that allows us to generate open and closed loop data, and the PSD of the temporal errors, since there is wind moving the atmosphere and the AO loop takes time between making a measurement and sending the corrections to the DM [17]. To generate each wavefront sample, we start by generating a grid of complex Gaussian random variables in the spatial frequency domain. This random field is then modulated by the square root of the combined PSDs, which include contributions from the atmospheric turbulence, deformable mirror dynamics and temporal lag. The wavefront sample can be obtained as the real (or imaginary) part of the inverse Fourier transform of the resulting spectrum. Once a wavefront is generated, it is propagated through the WFS to produce the input image for the neural network, while the ground truth is obtained by computing the Zernike decomposition of the wavefront. The dataset contains wavefronts with $r_0$ between 5 - 20 cm, wind speeds ranging from -15 to 15 m/s in any direction, open loop, closed loop and intermediate residuals, photon noise considering between $10^3$ and $10^6$ photon per frame and detector noise between 0 and $3\,e^-/pix/frame$. We generated a test dataset that contains 256 samples that will be used to compare the different scenarios after the training.

### 2.4 Joint Optimization and Training

Instead of training the WFS and the reconstructor separately, we perform a co-optimization, allowing the WFS mask parameters and the neural network weights to evolve simultaneously. The loss function used was a modified version of the mean squared error (MSE) between the ground truth $z$ and the estimated values $\hat{z}$ for the $N$ Zernike coefficients, in which we normalized by the $r_0$ value of the sample (the strength of the turbulence increases as $r_0$ decreases), to improve the generalization and robustness of the NN. For each sample in the batch, the loss function is computed as

$$\mathcal{L} = r_0 \frac{1}{N} \sum_N (z - \hat{z})^2. \tag{4}$$

Given that we have a fully differentiable model for the optical system, it is possible to use the same back propagation algorithm to perform the optimization for both the NN and the filtering mask. In this work, the model is trained using Adam optimizer [8]. The training strategy consisted of 30 000 epochs with batches of 64 images each and a learning rate of $1e-4$ for both the reconstruction network and optical parameters.

## 3 Training the Zernike wavefront sensor

We trained three models. In the first, the reconstruction network was trained with a fixed mask of $\alpha = \pi/2$ and $\beta = 1\,\lambda/D$, while in the second, it was trained with a fixed mask with $\beta = 2\,\lambda/D$. In the third case, both the network and the phase mask parameters were optimized together. Figure 2 shows the results from these runs, in which blue and orange curves show the fixed optical parameters, and the green curve is the run where the optical parameters were optimized along with the network.

From Fig. 2, we observe that both fixed-parameter configurations show similar loss evolutions, with the $2\lambda/D$ diameter offering only a slight improvement. In contrast, co-optimizing

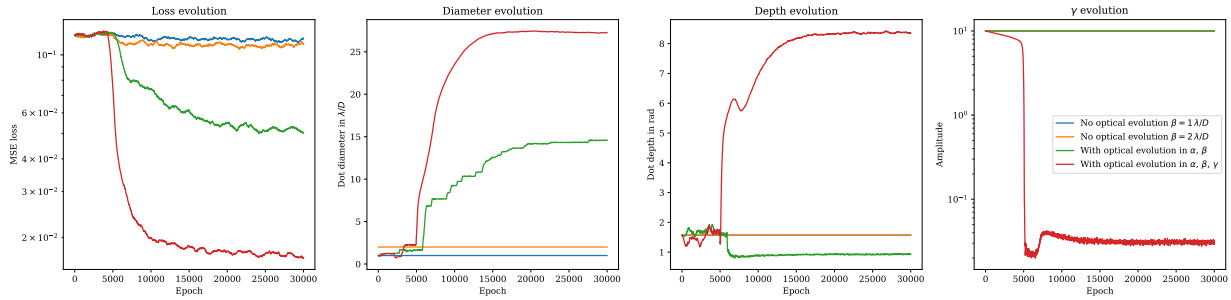


FIGURE 2 : Evolution during training of the Zernike wavefront sensor. In blue and orange the optical parameters are fixed and the network trained. In green both the network and the optical parameters of the regular Zernike WFS are trained. In red, an extra free parameter was added that does not correspond to regular Zernike masks. Left panel : loss evolution through the epochs ; middle left panel : diameter evolution of the Zernike dot ; middle right panel : depth evolution of the Zernike dot. The blue and orange curves overlap ; right panel : evolution of the parameter $\gamma$. Blue, orange and green overlap

the optical parameters with the network led to a significant loss reduction, particularly around epoch 5 000, where the diameter and depth of the mask evolved rapidly.

The optimized diameter converged to over $14\,\lambda/D$, far exceeding the conventional values (1 - 2 $\lambda/D$) typically used for small-phase regimes [2]. These small values are suitable for nearly diffraction-limited star images, but the signal saturates quickly with larger amplitudes of phase. The optimizer enlarged the dot diameter to match the average size of the star during the training, which included both open- and closed-loop phases, and therefore increasing the dynamic range of the sensor.

While conventional Zernike WFSs set the depth to $\alpha = \pi/2$ for a signal proportional to the sine of the phase, a different depth introduces a cosine component. Traditional linear reconstruction can't use this, but a non-linear reconstruction can leverage both.

Evaluating the test dataset, it gives the results presented in table 1, in which it is possible to confirm the improvement of the loss function when the optical parameters were optimized along with the network. The table also includes reference values for the loss for the two masks using state-of-the-art (SOTA) linear reconstruction techniques [18], in which it is possible to observe that the co-optimization performed better than the SOTA methods.

Manufacturing a $14\lambda/D$ diameter mask is feasible using standard photolitographic reactive ion etching techniques [12], similar to those for regular Zernike masks. The $\lambda/D$ value represents an angular size, allowing the physical spot size to be adjusted by the F-number of the beam.

# 4 An extra free parameter for the Zernike wavefront sensor

To have a differentiable model of the mask of the Zernike wavefront sensor, we had to approximate it using a $\tanh$ function, with a $\gamma$ factor high enough to emulate a discontinuity. Since we had a parametric model and optimization algorithm that could optimize the optical parameters of a WFS, we let the system evolve not only the diameter and the depth but also this extra factor. Figure 2 shows in red the results of the training using the extra parameter. It is possible to observe in the loss evolution that adding this free parameter improved the performance of the system. The value of this parameter converged to around $\gamma = 0.03$, therefore the approximation for the Zernike mask of $\gamma \gg 1$ does not hold anymore and the resulting mask no longer resembles a standard Zernike WFS. Also, the diameter and the depth converged to values much greater than before. Again, it is important to remark that the Zernike WFS is designed to work with small amplitudes of phase, and the training done here was done with a wide variety of wavefronts. When evaluating the test dataset in table 1, this training resulted in a lower loss when compared to the previous trainings of the Zernike WFS. This suggests that adding degrees of freedom in the WFS can have a significant impact on the overall wavefront reconstruction process.

# 5 Training the Pyramid wavefront sensor

The usual design of the Pyramid WFS consists of a glass pyramid that separates the light into four distinct pupils. It has been shown that it is possible to use a flattened version of the Pyramid [5], in which the pupils overlap. These overlapped pupils can bring benefits, such as signal coming from the interference of the pupils and a smaller detector footprint.

To test the Pyramid wavefront sensor, we ran two trainings : in the first only the reconstructor was trained, and in the second the angle of the pyramid was allowed to change, which could result in overlapping pupils. When the optical parameters were allowed to evolve, the system always converged to an overlap of around 90% of the pupils. This was independent of the starting conditions, as we ran several iterations of the second training, and it always converged to that value. Nevertheless, the overall performance of this optimized system is similar to the classic version of the Pyramid, in which the pupils are 0% overlapped. It is interesting to note that the PWFS, with and without optimization of the optical parameters, did not outperform all the other models on the test dataset, as can be observed in table 1, given that the Zernike with the extra free parameter had a lower total loss.

TABLE 1 : Evaluation of the test dataset. Top : different training scenarios ; bottom : State-of-the-art linear reconstruction

| Training configuration | Loss |
|---|---|
| Zernike WFS no optical evolution $\beta = 1\,\lambda/D$ | 0.122 |
| Zernike WFS no optical evolution $\beta = 2\,\lambda/D$ | 0.105 |
| Zernike WFS with optical evolution in $\alpha$, $\beta$ | 0.047 |
| Zernike WFS with optical evolution in $\alpha$, $\beta$, $\gamma$ | 0.016 |
| Pyramid WFS no optical evolution | 0.022 |
| Pyramid WFS with optical evolution in $\alpha$, $\beta$ | 0.021 |
| Linear reconstruction | |
| Zernike WFS $\beta = 1\,\lambda/D$ | 0.119 |
| Pyramid WFS | 0.058 |

# 6 Conclusion

In this work, we proposed a co-optimization framework where both the wavefront sensor and the neural network reconstructor are trained together. The system adapts the optical parameters of the WFS to improve the signal for the neural network, leading to a better reconstruction. This joint optimization results in improved performance compared to a fixed WFS with a separately trained neural network. For the Zernike wavefront sensor, we found that the optimal values for depth and width are vastly different from the usual design values, with the diameter of the dot 7 times larger than previously found values, and the depth a value different than $\pi/2$, which indicates that the reconstruction algorithm is taking advantage of the non-linear response of the Zernike WFS when the phase shift is a different value. For the Pyramid wavefront sensor, it showed that the current design of four distinct pupils is already close to the optimal solution, but it is possible to overlap the pupil up to 90% and achieve a similar performance, with a significant reduction in the footprint of the detector. Finally, we showed that adding an extra parameter to the Zernike mask allowed the system to better adapt to the wide variety of wavefronts used in the training, yielding a new mask that could even compete with the Pyramid WFS as a first-stage WFS. Future work will focus on testing these WFS in closed-loop AO configuration, and adding extra free parameters to the filtering masks of the WFS.

# Références


[1] V. CHAMBOULEYRON, O. FAUVARQUE, C. PLANTET et AL. : Modeling noise propagation in Fourier-filtering wavefront sensing, fundamental limits, and quantitative comparison. A&A, 670:A153, février 2023.

[2] V. CHAMBOULEYRON, O. FAUVARQUE, J. F. SAUVAGE et AL. : Variation on a Zernike wavefront sensor theme : Optimal use of photons. A&A, 650:L8, juin 2021.

[3] V. CHAMBOULEYRON, A. SENGUPTA, M. SALAMA et AL. : Using the Gerchberg-Saxton algorithm to reconstruct nonmodulated pyramid wavefront sensor measurements. A&A, 681:A48, janvier 2024.

[4] Olivier FAUVARQUE, Benoit NEICHEL, Thierry FUSCO et AL. : General formalism for Fourier-based wave front sensing. *Optica*, 3(12):1440, décembre 2016.

[5] Olivier FAUVARQUE, Benoit NEICHEL, Thierry FUSCO et AL. : General formalism for Fourier-based wave front sensing : application to the pyramid wave front sensors. *Journal of Astronomical Telescopes, Instruments, and Systems*, 3:019001, janvier 2017.

[6] D. L. FRIED : Optical Resolution Through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures. *Journal of the Optical Society of America (1917-1983)*, 56(10):1372, octobre 1966.

[7] Felipe GUZMÁN, Jorge TAPIA, Camilo WEINBERGER et AL. : Deep optics preconditioner for modulation-free pyramid wavefront sensing. *Photonics Research*, 12(2):301–312, 2024.

[8] Diederik P KINGMA et Jimmy BA : Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*, 2014.

[9] A. N. KOLMOGOROV : The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Numbers. *Proceedings of the Royal Society of London Series A*, 434(1890):9–13, juillet 1991.

[10] R. LANDMAN et S. Y. HAFFERT : Nonlinear wavefront reconstruction with convolutional neural networks for Fourier-based wavefront sensors. *Optics Express*, 28(11):16644, mai 2020.

[11] E. MUSLIMOV, N. LEVRAUD, V. CHAMBOULEYRON et AL. : Current status of PAPYRUS : the pyramid based adaptive optics system at LAM/OHP. *In Optical Instrument Science, Technology, and Applications II*, volume 11876 de *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 118760H, septembre 2021.

[12] M. N'DIAYE, K. DOHLEN, S. CUEVAS et AL. : Experimental results with a second-generation Roddier & Roddier phase mask coronagraph. A&A, 509:A8, janvier 2010.

[13] M. N'DIAYE, A. VIGAN, B. ENGLER et AL. : Cascade adaptive optics with a second stage based on a Zernike wavefront sensor for exoplanet observations : Experimental validation on the ESO/GHOST testbed. A&A, 692:A157, décembre 2024.

[14] B. NEICHEL, T. FUSCO, J. F. SAUVAGE, C. CORREIA et AL. : The adaptive optics modes for HARMONI : from Classical to Laser Assisted Tomographic AO. *In* Enrico MARCHETTI, Laird M. CLOSE et Jean-Pierre VÉRAN, éditeurs : *Adaptive Optics Systems V*, volume 9909 de *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 990909, juillet 2016.

[15] Ben C PLATT et Roland SHACK : History and principles of shack-hartmann wavefront sensing, 2001.

[16] Roberto RAGAZZONI : Pupil plane wavefront sensing with an oscillating prism. *Journal of Modern Optics*, 43(2):289–293, février 1996.

[17] Francois J. RIGAUT, Jean-Pierre VERAN et Olivier LAI : Analytical model for Shack-Hartmann-based adaptive optics systems. *In Adaptive Optical System Technologies*, volume 3353 de *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 1038–1048, septembre 1998.

[18] Noah SCHWARTZ, Jean-François SAUVAGE, Edgard RENAULT et AL. : Design of the HARMONI Pyramid WFS module. *arXiv e-prints*, page arXiv :2003.07228, mars 2020.

[19] Pauline TROUVÉ : *Conception conjointe optique/traitement pour un imageur compact à capacité 3D*. Theses, Ecole Centrale de Nantes (ECN), décembre 2012.

[20] Christophe VÉRINAUD : On the nature of the measurements provided by a pyramid wave-front sensor. *Optics Communications*, 233(1-3):27–38, mars 2004.

[21] Camilo WEINBERGER, Jorge TAPIA, Benoît NEICHEL et Esteban VERA : Transformer neural networks for closed-loop adaptive optics using nonmodulated pyramid wavefront sensors. A&A, 687:A202, juillet 2024.

[22] F. ZERNIKE : Diffraction theory of the knife-edge test and its improved form, the phase-contrast method. MNRAS, 94:377–384, mars 1934.