A Non-Linear Control Method with Reinforcement Learning for Adaptive Optics with Pyramid Sensors

ESA/ESO SCIOPS WORKSHOP 2022

B. Pou, J. Smith, E. Quiñones, M. Martín, D. Gratadour

Presentation

Bartomeu Pou

- PhD student in Artificial Intelligence in the **Barcelona Supercomputing Center** (BSC) and the **Universitat Politècnica de Catalunya** (UPC).
- Interested in Machine Learning and its application in Adaptive Optics.

Presentation

Jeffrey Smith
- PhD student.
- School of Computing, Australian National University (ANU).

Eduardo Quiñones
- Senior researcher.
- Barcelona Supercomputing Center (BSC).

Mario Martín
- Associate professor at Universitat Politècnica de Catalunya (UPC).

Damien Gratadour
- Associate professor at LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris.
- Australian National University (ANU).

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

l'Observatoire de Paris

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

.AMPERE
A Model-driven development framework for highly Parallel and EneRgy-Efficient computation supporting multi-criteria optimisation

Rising STARS

Introduction

Adaptive Optics (AO)

- **AO**: correct distortions on incoming wavefronts caused by the atmosphere.
- **Real time controller (RTC)**: based on the **WFS** image **reconstructs** the wavefront and adapts the **DM** to compensate the distortions.
  - Many sources of error (temporal, noise, aliasing, …).
- AO controller is a real-time cyber-physical system.

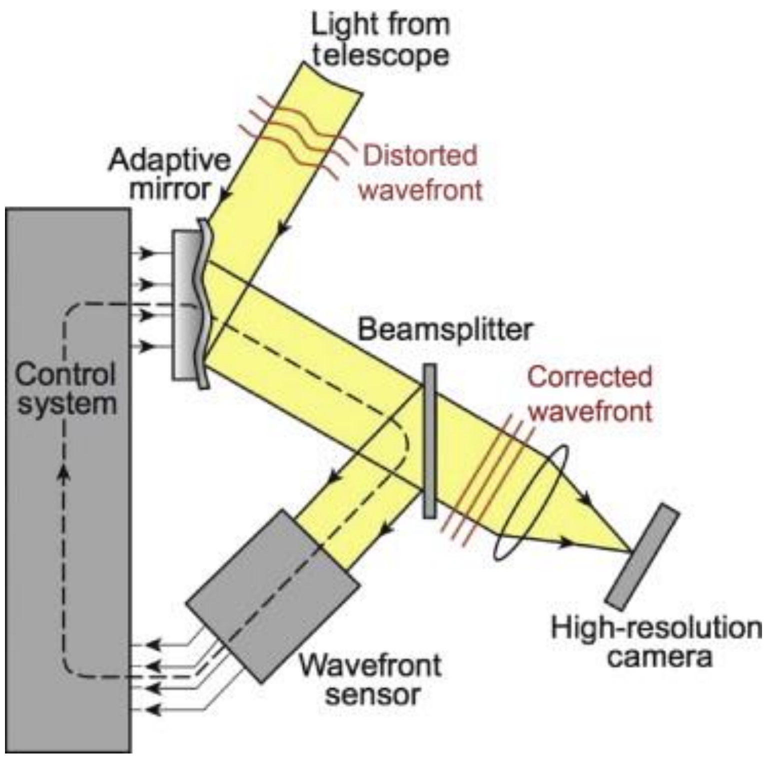


Fig 1. Closed-loop. Credit: Claire E. Max, UCSC.

Motivation

- State-of-the-art: model-based controllers.
  - Assumptions required (e.g. Kolmogorov Turbulence).
  - Need to calibrate to changing atmospheric conditions.
- Create an adaptive controller without any assumption based on **Reinforcement Learning (RL)**.
- Previous work of RL on AO with SH-WFS. [1, 2, 3].
- In this work we focus on AO with Pyramid WFS.

Background in RL

1. Framework for **Sequential Decision Making**.
2. **Learn** the optimal policy.
   - $\boldsymbol{a} \sim \pi^*(\boldsymbol{s})$
   - $\pi^*(\boldsymbol{s})$ maximises cumulative reward, r.
3. **Trial and error**.

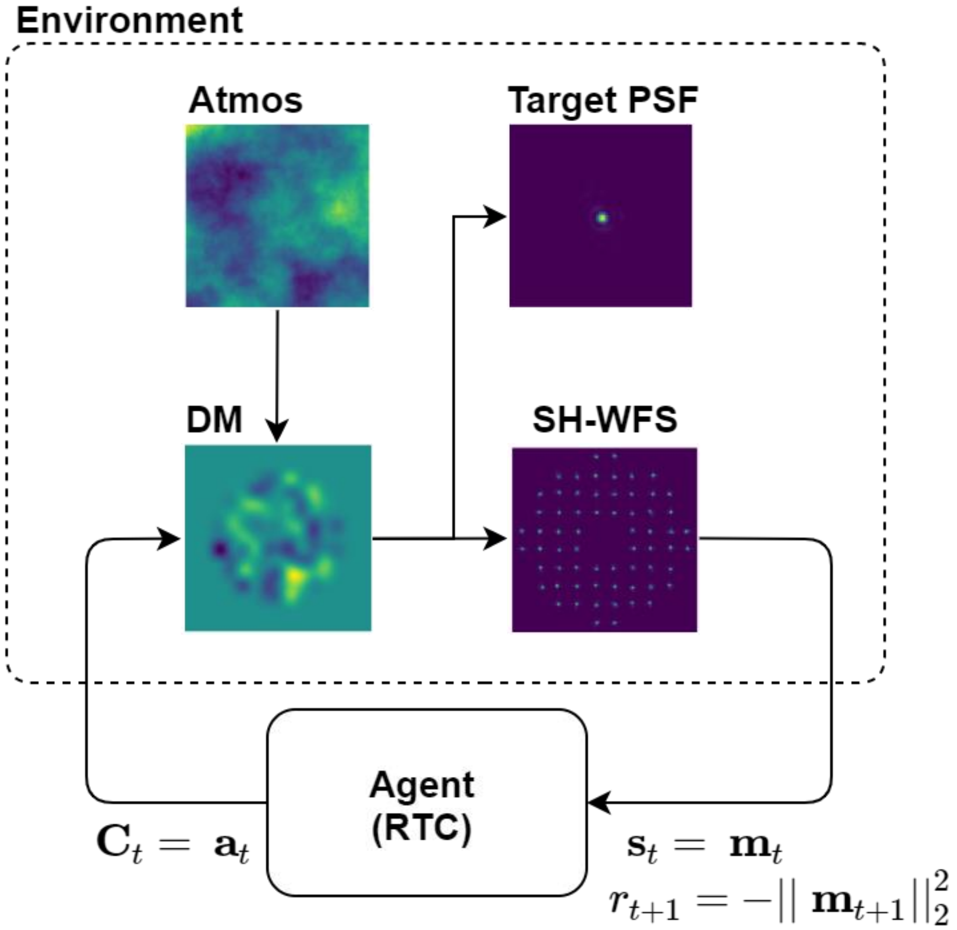
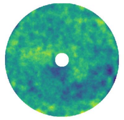
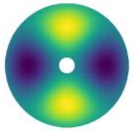
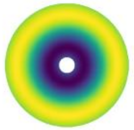
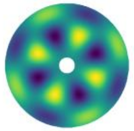


Fig 2. Agent – Environment interaction in AO. Extracted from [3].

I. Multi-Agent Reinforcement Learning (MARL) [3]

1. Avoid **curse of dimensionality** → from single-agent RL to multi-agent RL.
2. How?
   - Every agent control a set of global **orthogonal modes.**
   - $\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{r}$ are built considering only the controlled modes.

 $= c_0$  $+ \ldots + c_4$  $+ \ldots + c_{21}$  $+ \ldots$

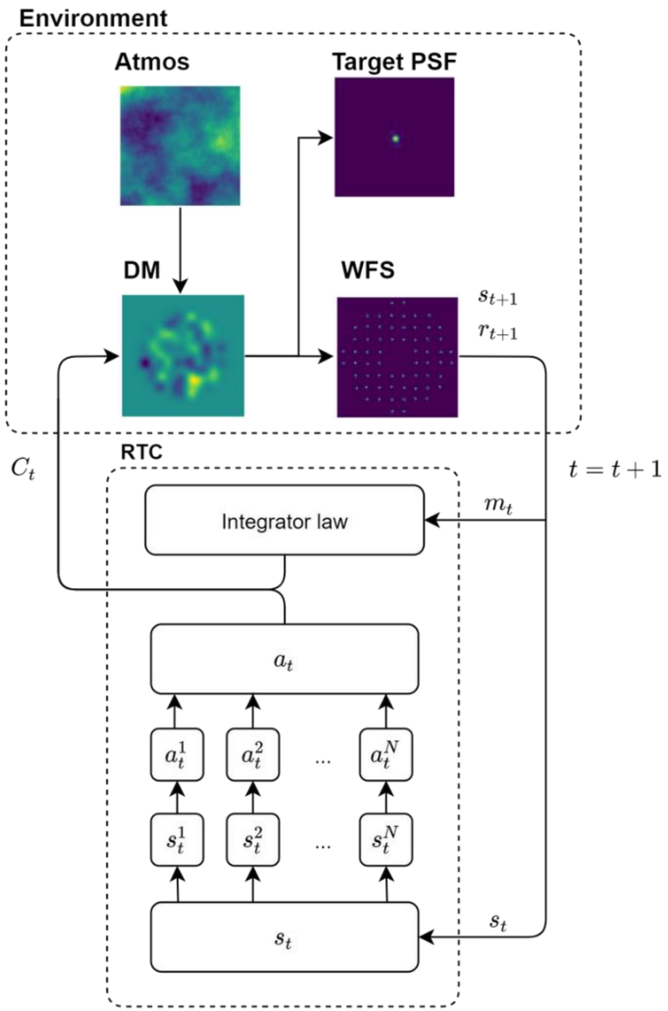


Fig 3. Multi-Agent – Environment interaction in AO.

I. Multi-Agent Reinforcement Learning (MARL) [3]

1. Avoid **curse of dimensionality** → from single-agent RL to multi-agent RL.
2. How?
   - Every agent control a set of global **orthogonal modes.**
   - $\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{r}$ are built considering only the controlled modes.

3. To build $\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{r}$, first we need to reconstruct the phase with a linear MVM approach and project it to the modal basis.

$$c_t = P_{a2m}(D\, m_t)$$

$$s_t^i = (c_t^i, c_{t-1}^i, c_{t-2}^i, c_{t-3}^i)$$

$$r_t^i = -\frac{1}{|M^i|} \sum_{m \in M^i} (c_t^m)^2$$

D: control matrix.
m_t: SH-WFS centroids at time t.
P_{a2m}: projection actuator commands to modal basis.

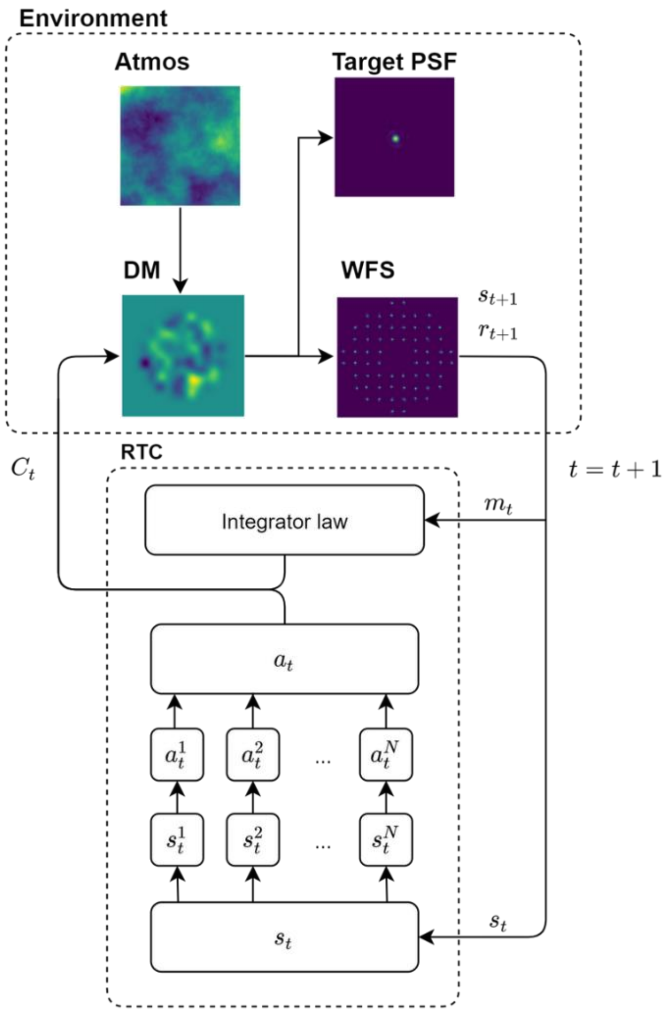


Fig 3. Multi-Agent – Environment interaction in AO.

Extension to the Pyramid WFS (PWFS)

The PWFS **linear** reconstruction (with MVM) for a mode depends on the status of the other modes.

- Modal basis: $(\phi_0, \dots, \phi_N)$
  - E.g. KL or Btt.
- $Rec(\phi_i) = f(\phi_0, \dots, \phi_N)$

We can not separate the problem into multiple independent problems!

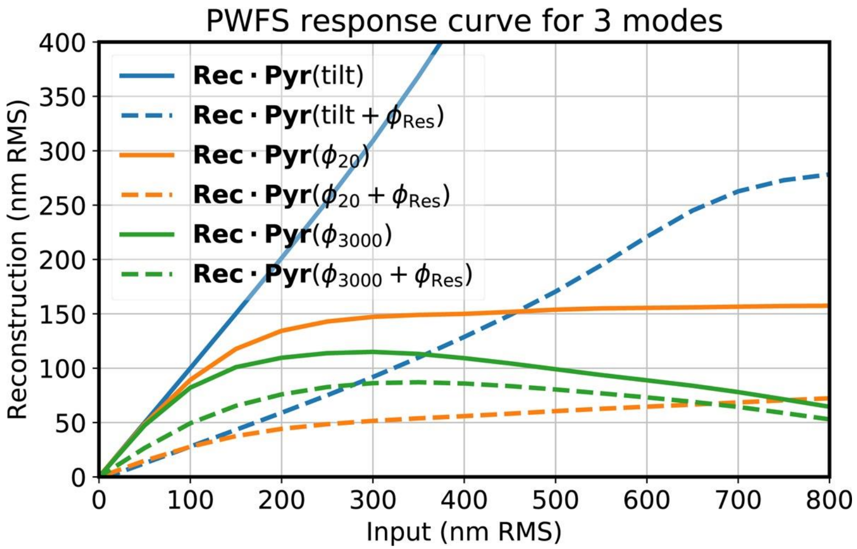


Fig 4. Reconstruction of 3 KL modes (tilt, mode 20 and mode 3000) with/without a residual phase. Extracted from [4].

Methods

Conditional Generative adversarial networks (C-GAN)

- Smith et al. [5] phase prediction from SH-WFS images with C-GAN [6] (**Image to image translation**).
  - **Supervised Learning**: requires dataset of pairs (WFS image, phase)

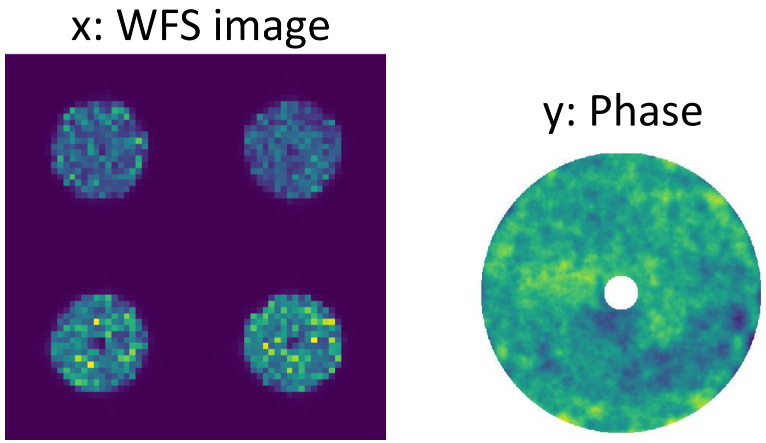


Fig 5. Datapoint

Conditional Generative adversarial networks (C-GAN)

- Game theoretical approach on learning neural network weights.
  - **Generator (G):** learn to predict output image (conditioned on an input image).
  - **Discriminator (D):** learn to predict if an output image is real or fake (conditioned on an input image).
- Process leads to improvement of G and D until equilibrium is reached.

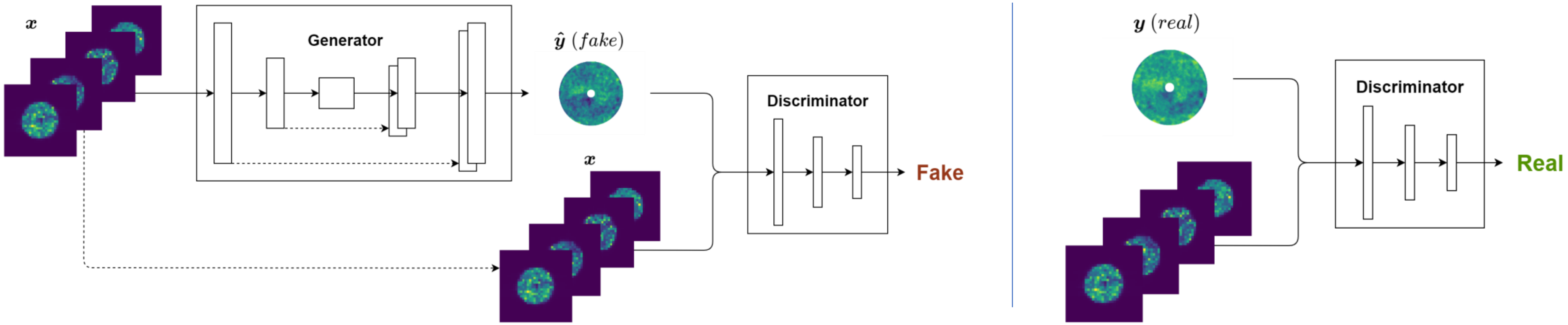


*Fig 6. C-GAN training. **Left**: the discriminator predicts fake because the phase is artificially generated. **Right**: vice versa.*

C-GAN + MARL

- Once the C-GAN is trained we inject the generator in a closed-loop to predict the phase.
- The phase is projected to the modes to derive s and r.

$$c_t^i = (P_{ph2m} G(x_t))^i$$

- The controller can be understood as composed by two components: **a non-linear reconstructor** (**C-GAN**) and **a predictive controller** (**MARL**).

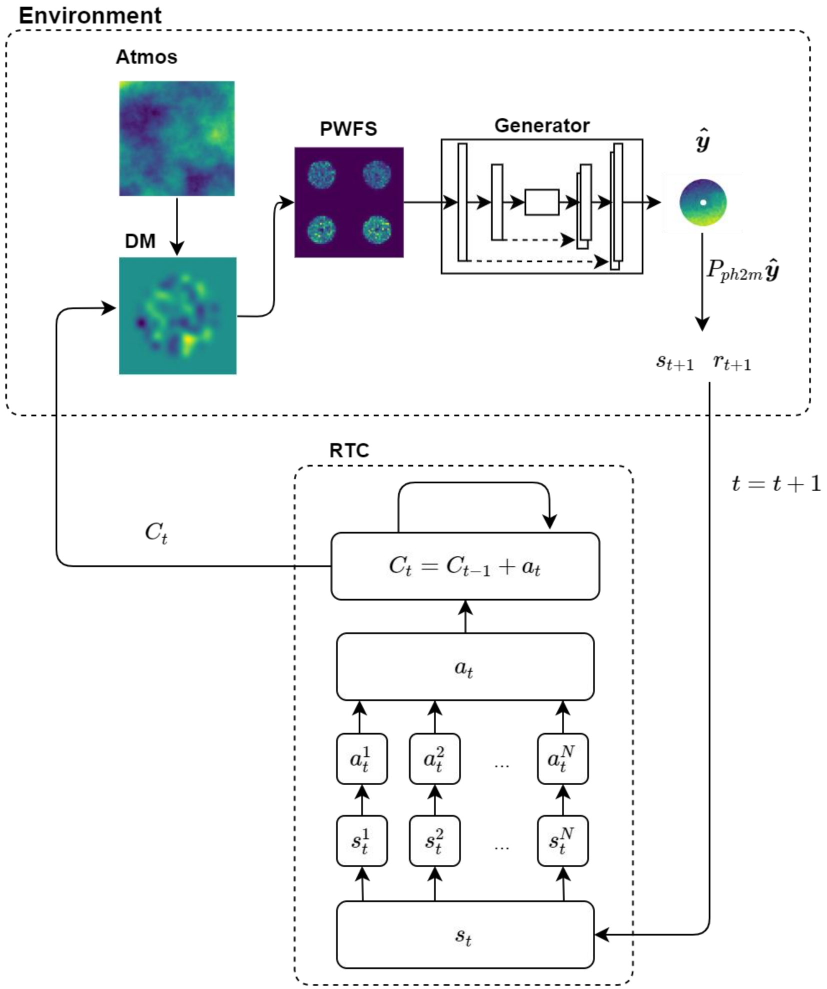


Fig 7. Multi-Agents – Environment interaction with C-GAN included in the environment.

Experiments

Simulation

High-performance GPU-enabled simulator of AO: COMPASS [7].

1. Simulation of a system with a 8m telescope equipped with 16x16 PWFS.
2. Comparison against an integrator optimised for the PWFS with CLOSE [8].
3. 6 agents controlling 34 modes each and a dedicated TT agent.

Telescope parameters		**Atmospheric parameters**	
Diamenter (m)	8	Num layers	1
AO loop parameters		Altitude (km)	0
Loop frequency (Hz)	500	r_0 (m)	0.16 @ 500 nm
Frames of delay	2	L_0 (m)	10^5
Target parameters		Wind speed (m/s)	20
λ_{target} (μm)	1.65	**PWFS parameters**	
DM parameters		Num. Subapertures	16x16
Mirrors	PZT and TT	Field stop size ($''$)	1.5
Coupling (PZT)	0.2	λ_{wfs} (μ m)	0.5
Num. of modes	206	Modulation Amplitude (λ/D)	3

Table 1. Parameters for the experiment.

Difficulties in training the C-GAN

Data-diversity problem. Final dataset consist of:

- Integrator + noise (1 M).
- Free turbulence + noise (1 M) .

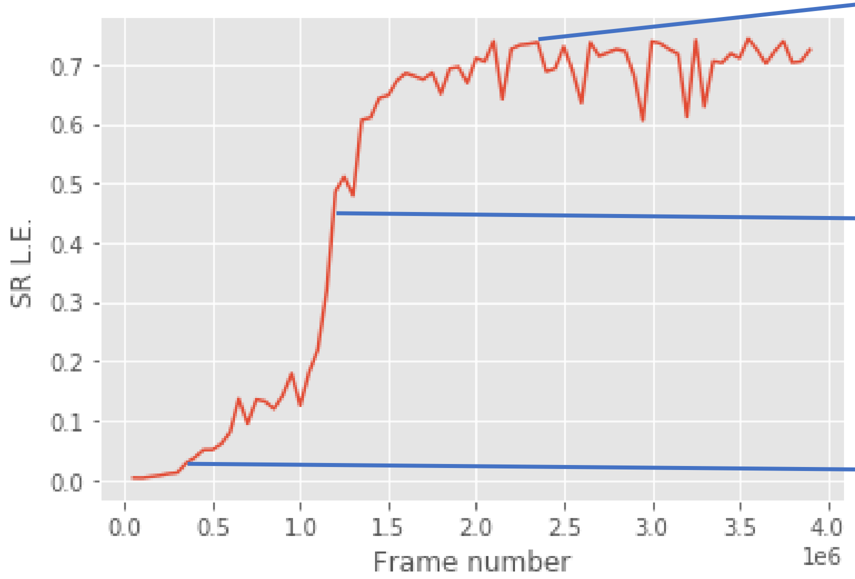
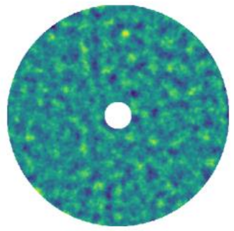
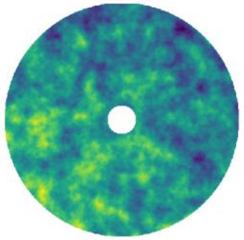


Fig 8. Difficulty in training the C-GAN.

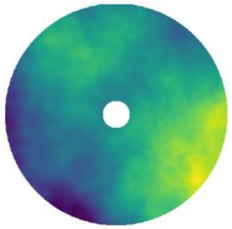
Variance of the residual phase



0.102 um RMS



0.216 um RMS



1.01 um RMS

Results I: MARL + C-GAN

- RL learns from scratch a predictive controller based on the C-GAN inferences.
- 8 points improvement over the integrator performance in SR L.E.

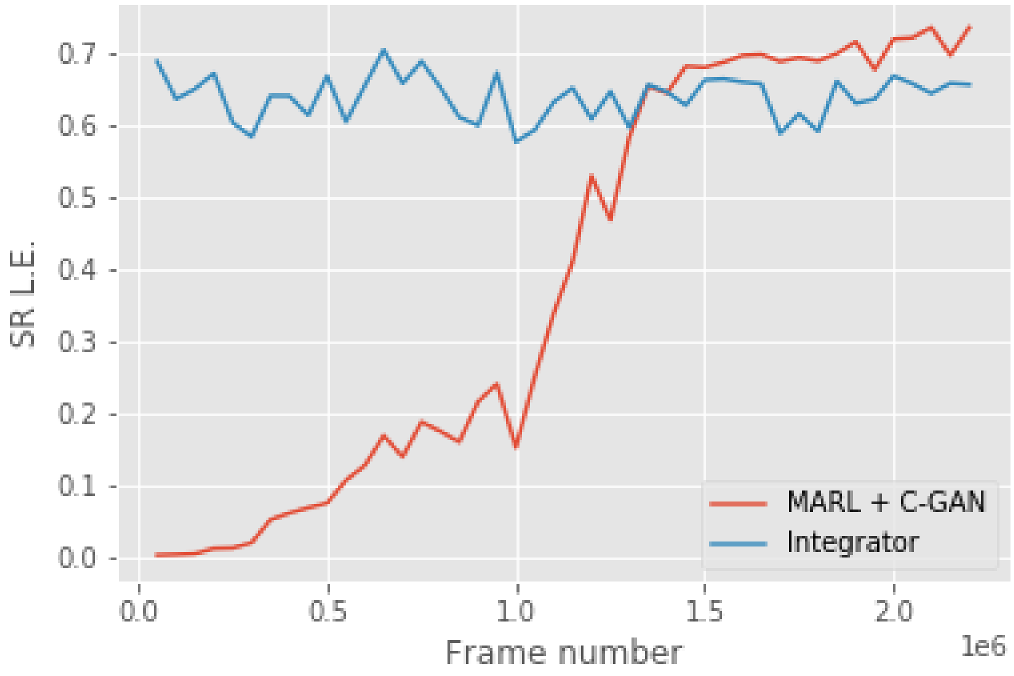


Fig 9. Training curve.

Results II. C-GAN

- Two examples of phase prediction during closed-loop.
- The percentage of error is lower when the amplitude of the phase is higher.
- At low amplitudes the GAN starts to fail.

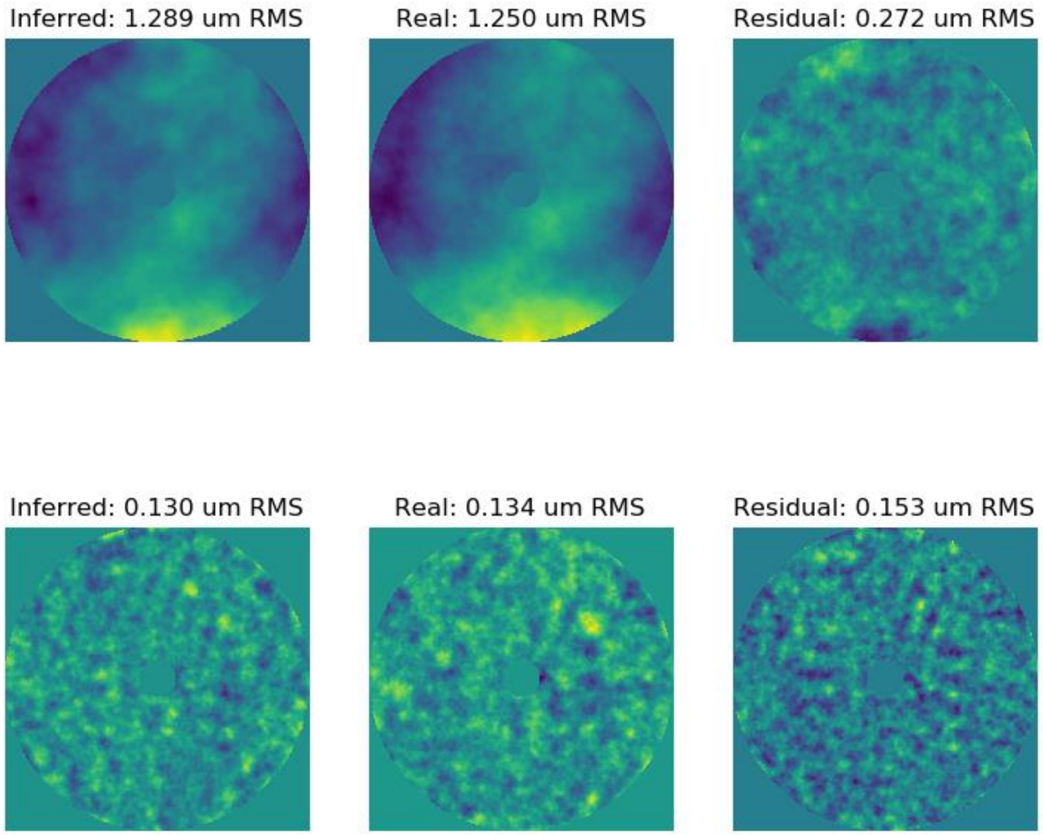


Fig 10. Examples of C-GAN prediction.

Conclusion and future work

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

l'Observatoire de Paris

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

.AMPERE
A Model-driven development framework for highly Parallel and EneRgy-Efficient computation supporting multi-criteria optimisation

Rising STARS

Conclusions

1. We have developed a new controller for AO with the PWFS based on Machine Learning.
2. The controller uses a **non-linear reconstructor** (C-GAN component) and a **predictive controller** (RL component).
3. We outperform an optimised integrator controller in the test experiment.

Future work

1. Repeat for a higher number of subapertures.
2. Investigate the effects of the dataset size for each regime. How much data should we gather from each amplitude to get the best final MARL performance?
3. Test on changing atmospheric conditions. (¿Can the C-GAN generalize to different values of r_0?).

Bibliography

[1] Nousiainen, Jalo, et al. "*Adaptive optics control using model-based reinforcement learning*." *Optics Express* 29.10 (2021): 15327-15344.

[2] Landman, Rico, et al. "*Self-optimizing adaptive optics control with reinforcement learning for high-contrast imaging*." *Journal of Astronomical Telescopes, Instruments, and Systems* 7.3 (2021): 039002.

[3] Pou, B., et al. "*Adaptive optics control with multi-agent model-free reinforcement learning*." *Optics Express* 30.2 (2022): 2991-3015.

[4] Deo, V., et al. "*A telescope-ready approach for modal compensation of pyramid wavefront sensor optical gain*." *Astronomy & Astrophysics* 629 (2019): A107.

[5] Smith, J., et al. "*Enhanced Adaptive Optics Control with Image to Image Translation*", Uncertainty in Artificial Intelligence, (2022), (in press).

[6] Isola, Phillip, et al. "*Image-to-image translation with conditional adversarial networks.*" *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[7] Gratadour, D., et al. "*COMPASS: an efficient, scalable and versatile numerical platform for the development of ELT AO systems.*" *Adaptive Optics Systems IV*. Vol. 9148. International Society for Optics and Photonics, 2014.

[8] Deo, Vincent, et al. "*A correlation-locking adaptive filtering technique for minimum variance integral control in adaptive optics.*" *arXiv preprint arXiv:2103.09921* (2021).

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

l'Observatoire de Paris

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

A Model-driven development framework for highly Parallel and EneRgy-Efficient computation supporting multi-criteria optimisation

Rising STARS

Acknowledgements

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No **871669** and **873120**.

.AMPERE

A Model-driven development framework for highly Parallel and EneRgy-Efficient computation supporting multi-criteria optimisation

Rising STARS