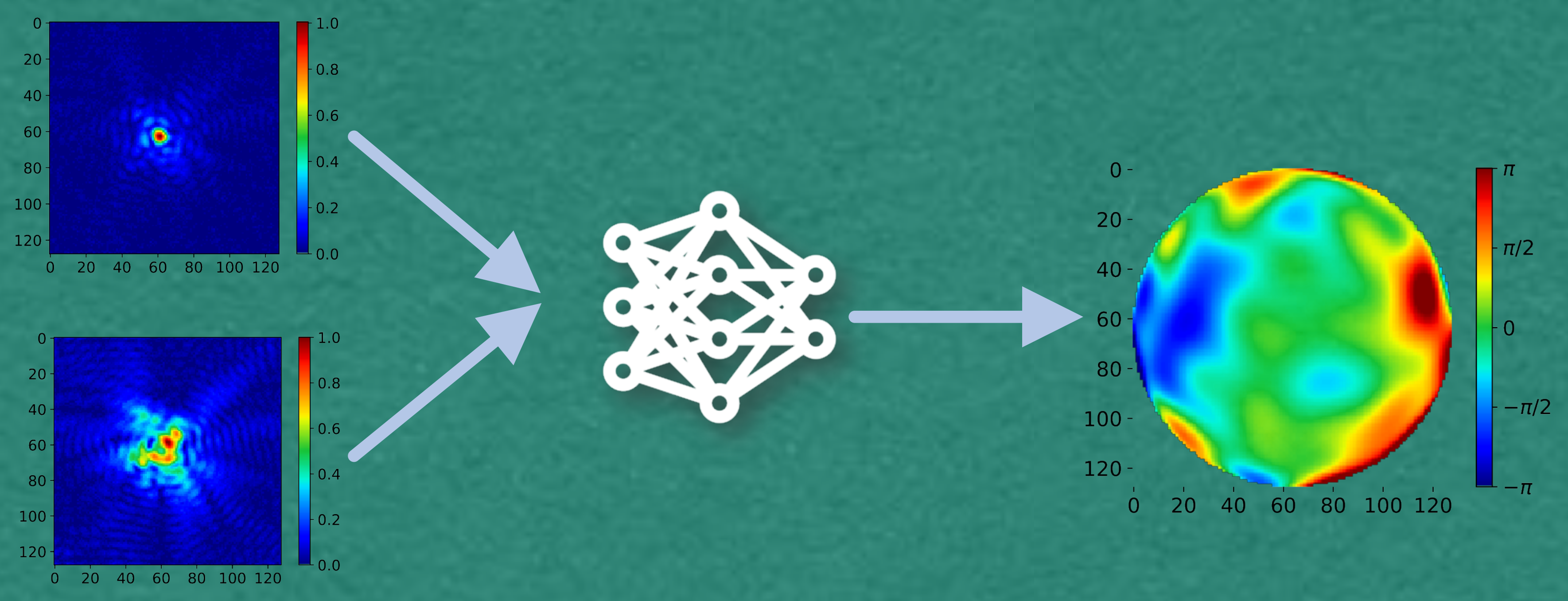
Focal-plane wavefront sensing using deep learning: review & perspectives



Gilles Orban de Xivry — gorban@uliege.be
December 2021

WFS WORKSHOP 2021

Outline

- Introduction: FPWFS, deep learning, review
- CNN-based framework FPWFS
- Practical considerations
- Future avenues

Focal plane wavefront sensing

Advantages

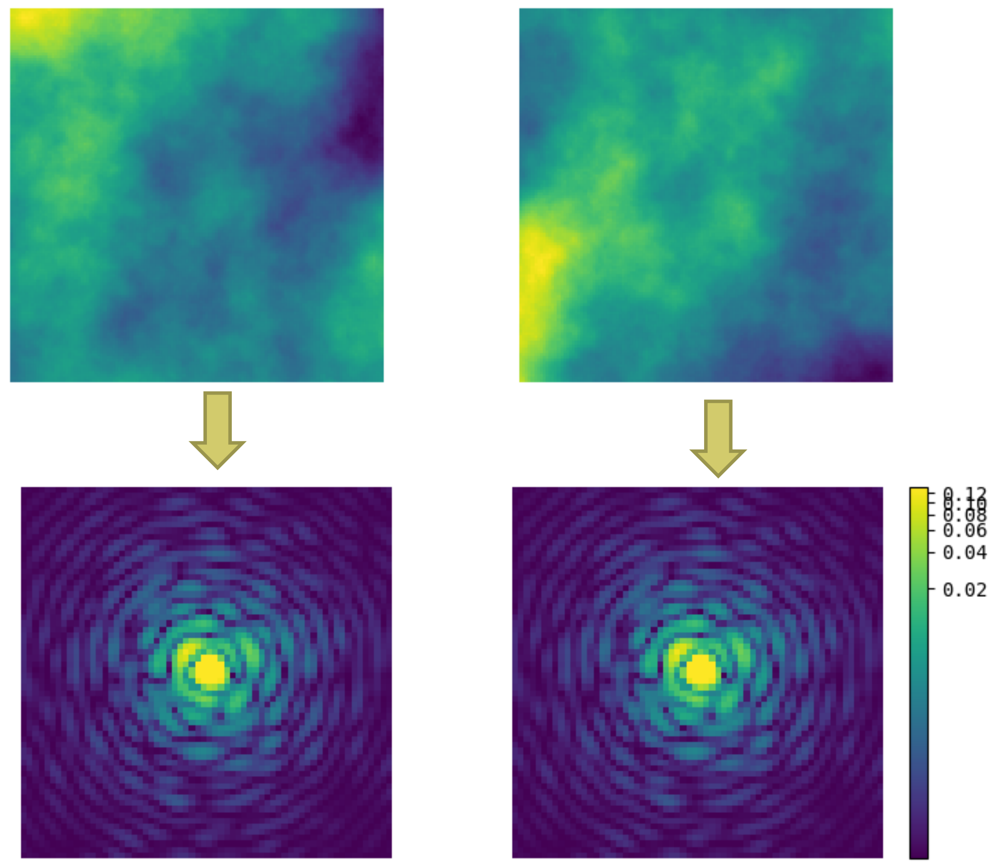
- No NCPA or chromatic errors
- High sensitivity, incl. to phase discontinuities
- Simple opto-mechanically

Disadvantages

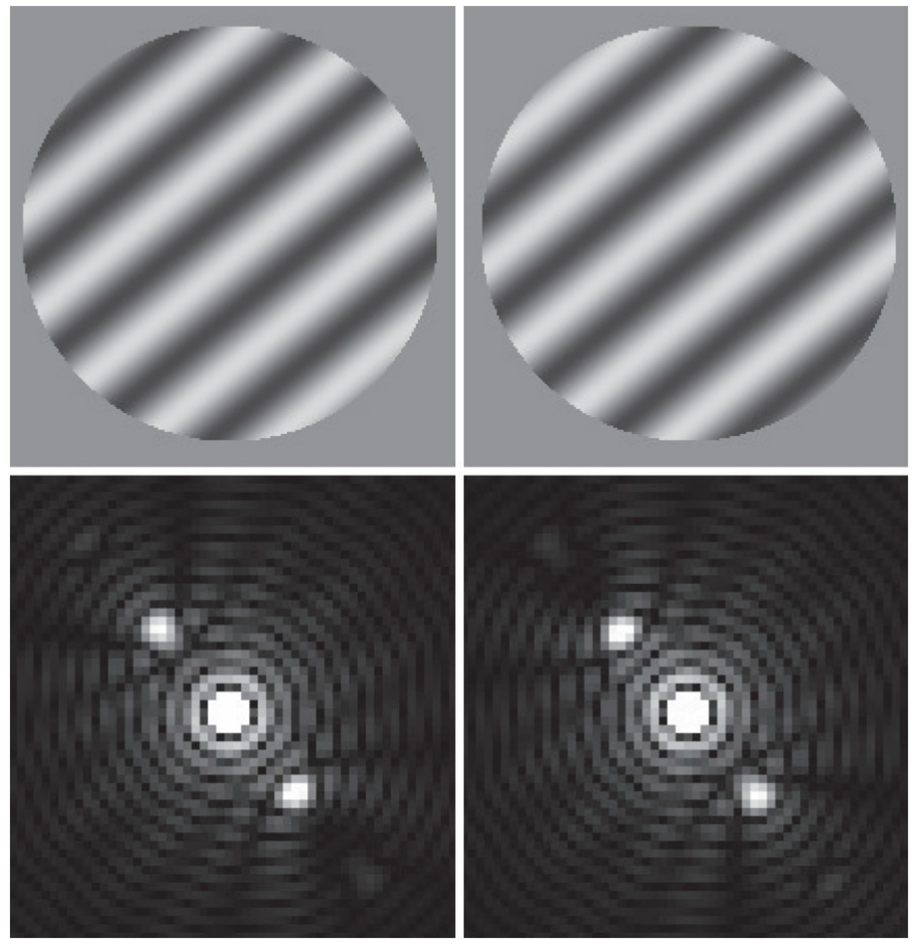
- High computational cost and/or limited to small aberration
- Intensity measurements result in phase ambiguity

➡ Moving the hardware complexity to the software

Sign ambiguity on even modes



Phase of sine wave difficult to infer



Guyon 2018

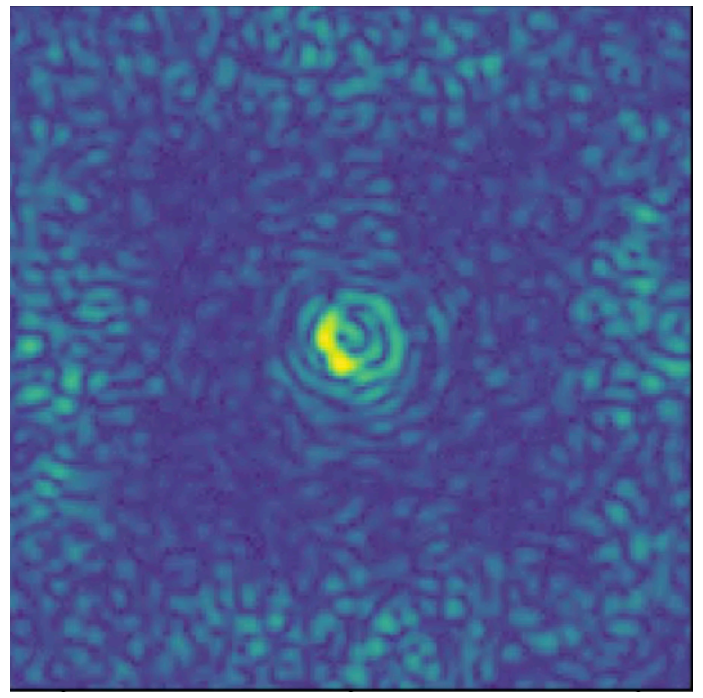
Focal plane wavefront sensing in astronomy

Two (selected) regimes

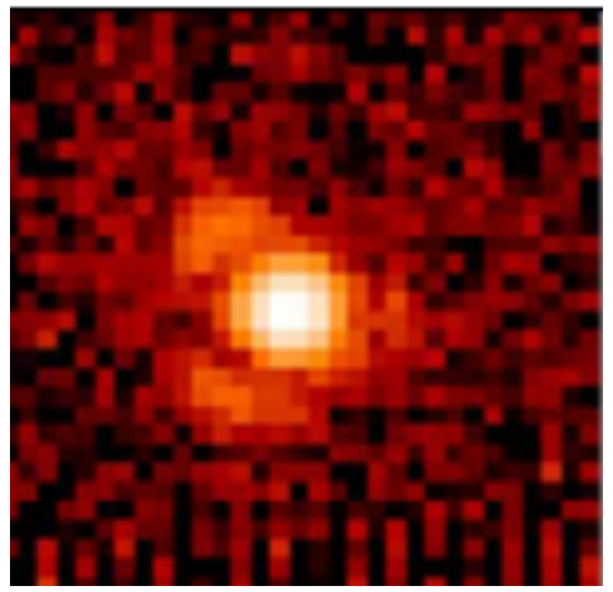
	NCPA	AO
Aberration level	100-500nm rms	1-5µm rms
Correction timescale	>1sec	1ms
Spatial frequency [number of modes]	~20 on a VLT ~100 on a ELT	~100 on 2-4m ~>400 on a 8m > 4000 on a 40m
Expected residuals	~20nm rms	~100nm rms

✳Also cophasing (JWST, ELT)

High-contrast image



Courtesy to M. Willson @ULiege



LWE, Vievard et al. 2019

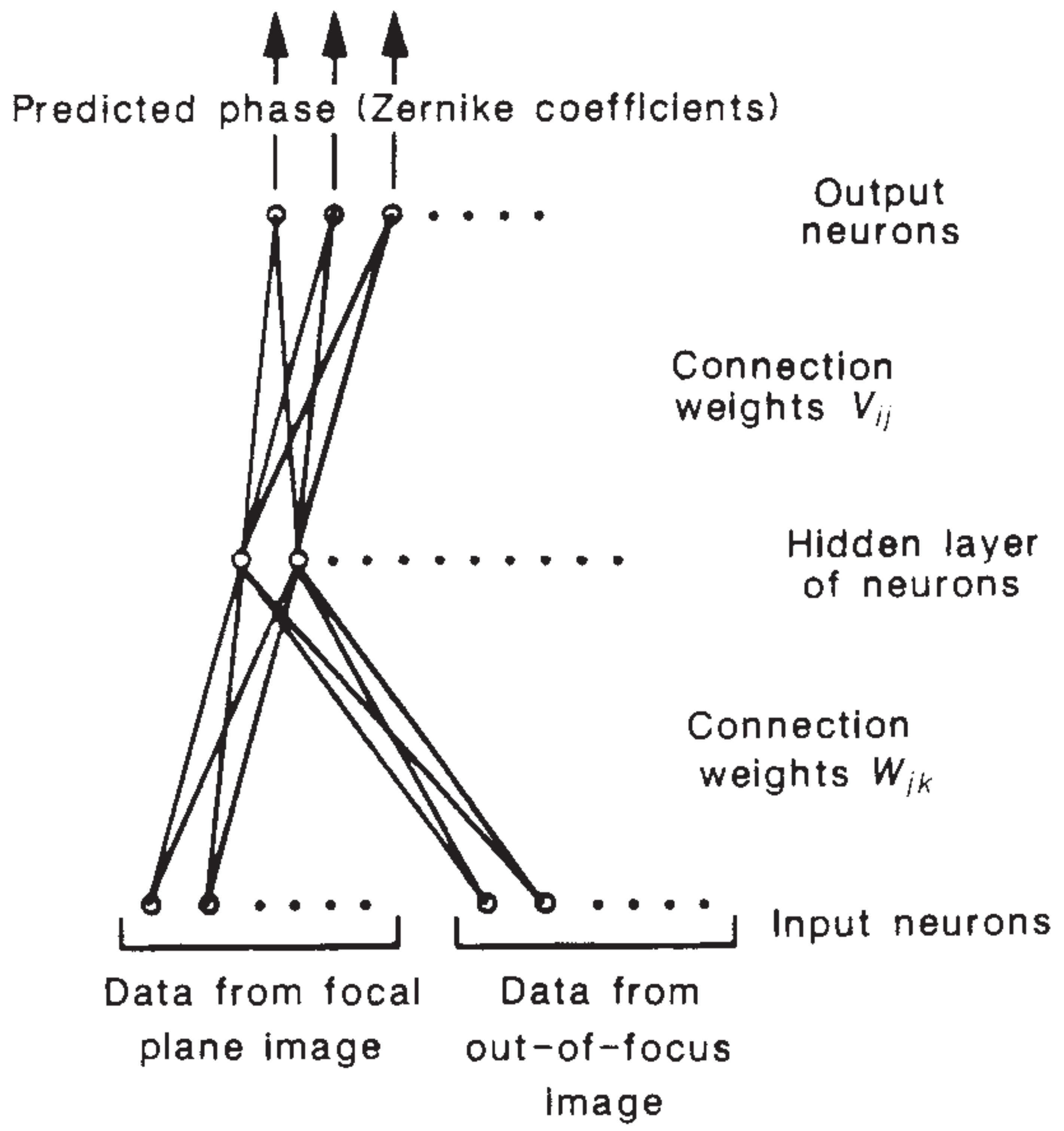
Deep learning ?

- **Why deep learning ?**
  - Handle the non-linear nature of the problem
  - Inference speed (no iterative process)
  - Robustness (no local optimum)

- **Deep learning can also mean in practice:**
  - Lessens the need for a deep mathematical grasp / precise formulation
  - Leverage modern hardware, GPUs

Early work in 1990's in astronomy

- **Co-phasing** (piston-tip-tilt) of 6 mirrors [1, 2]
- **Adaptive optics** [3]
- **HST aberration** [4]



- Using a shallow *multilayer perceptron* with sigmoid activation function.

  ➜ limited in their learning capacity (vanishing gradient problem)

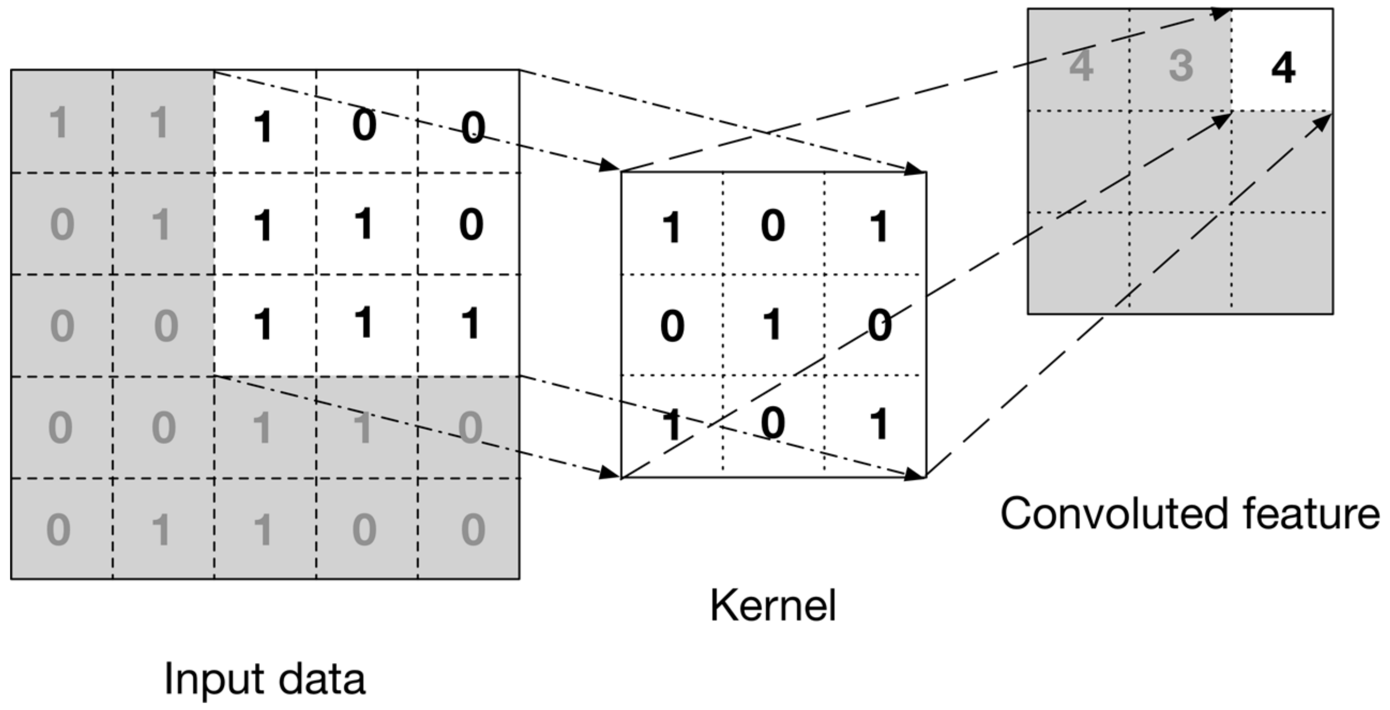
[1] Angel et al. 1990
[2] Lloyt-Hart et al. 1991
[3] Sandler et al. 1991
[4] Barrett & Sandler 1993

Convolutional Neural Networks

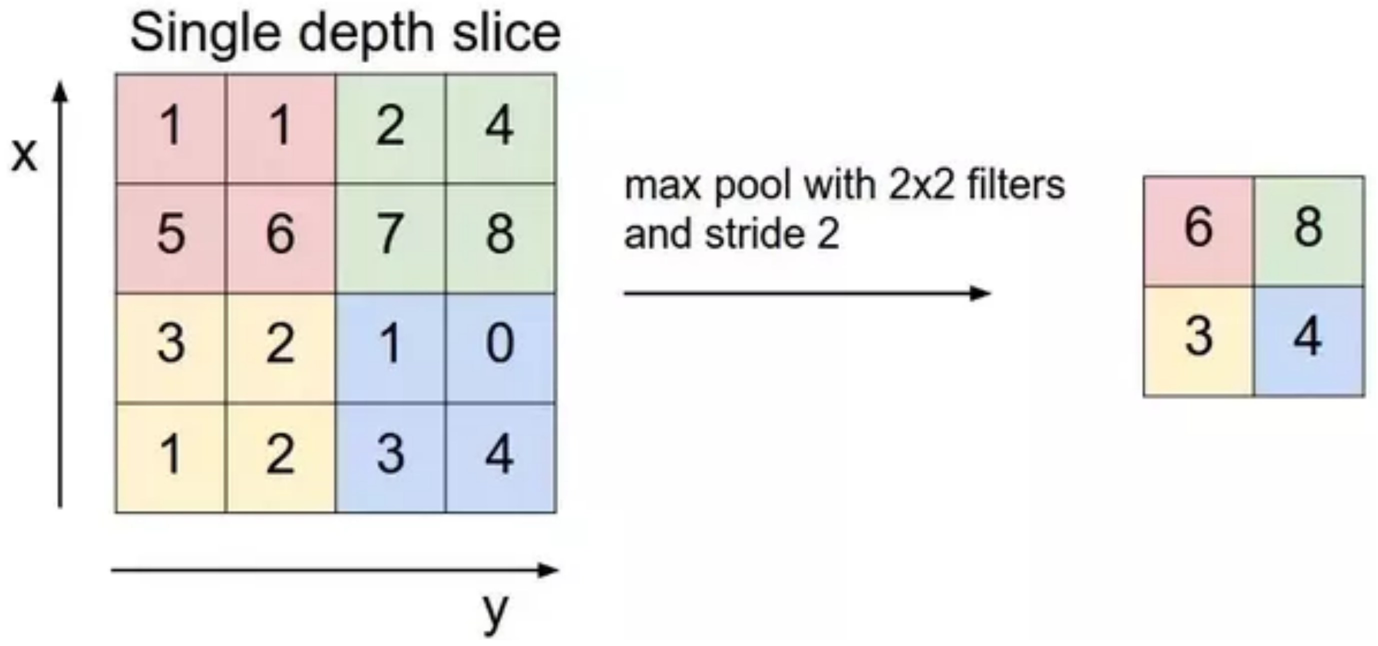
At the heart of all recent works

CNN are composition of

- convolution,
- pooling layer,
- activation function (ReLU)
- Normalisation layers
- fully connected layer



Convolution



Max pooling

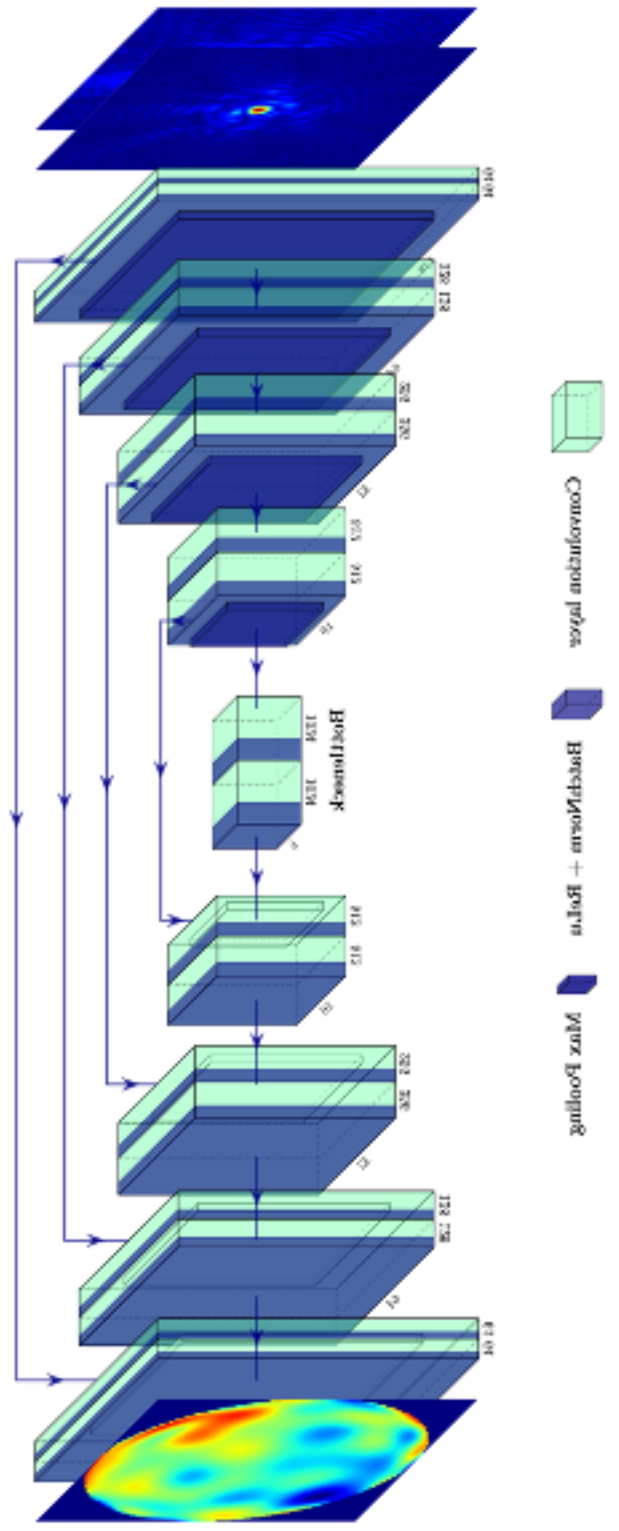
- Hierarchical composition provide a range of *receptive fields*
- Solutions to vanishing gradient allow training of deep networks (incl. automatic differentiation)
- Rapidly evolving field since 2010 (beginning of the ImageNet contest)

CNN-based framework FPWFS

Supervised learning.
Labelled dataset used for training.

PSF-based approach

- *Laser communication* (Guo et al. 2019)
- *Co-phasing of JWST* (Paine & Fienup 2018)
- *WFS for AO* in astronomy (Andersen et al. 2020)
- NCPA in astronomy and performance limit (Orban de Xivry et al. 2021)



But also:

- **imaged-based** (microscopy, e.g. Krishnan et al 2020, Wu et al. 2020)
- **metric-based** (Fourier-space metric that are *object agnostic*; Naik et al. 2020),
- **preconditioned intensity image** (Nishizaki et al. 2019)

▹ *Mild accuracy, few Zernike modes, compact CNN and fast inference*

CNN-based framework FPWFS

Enlarging dynamic range

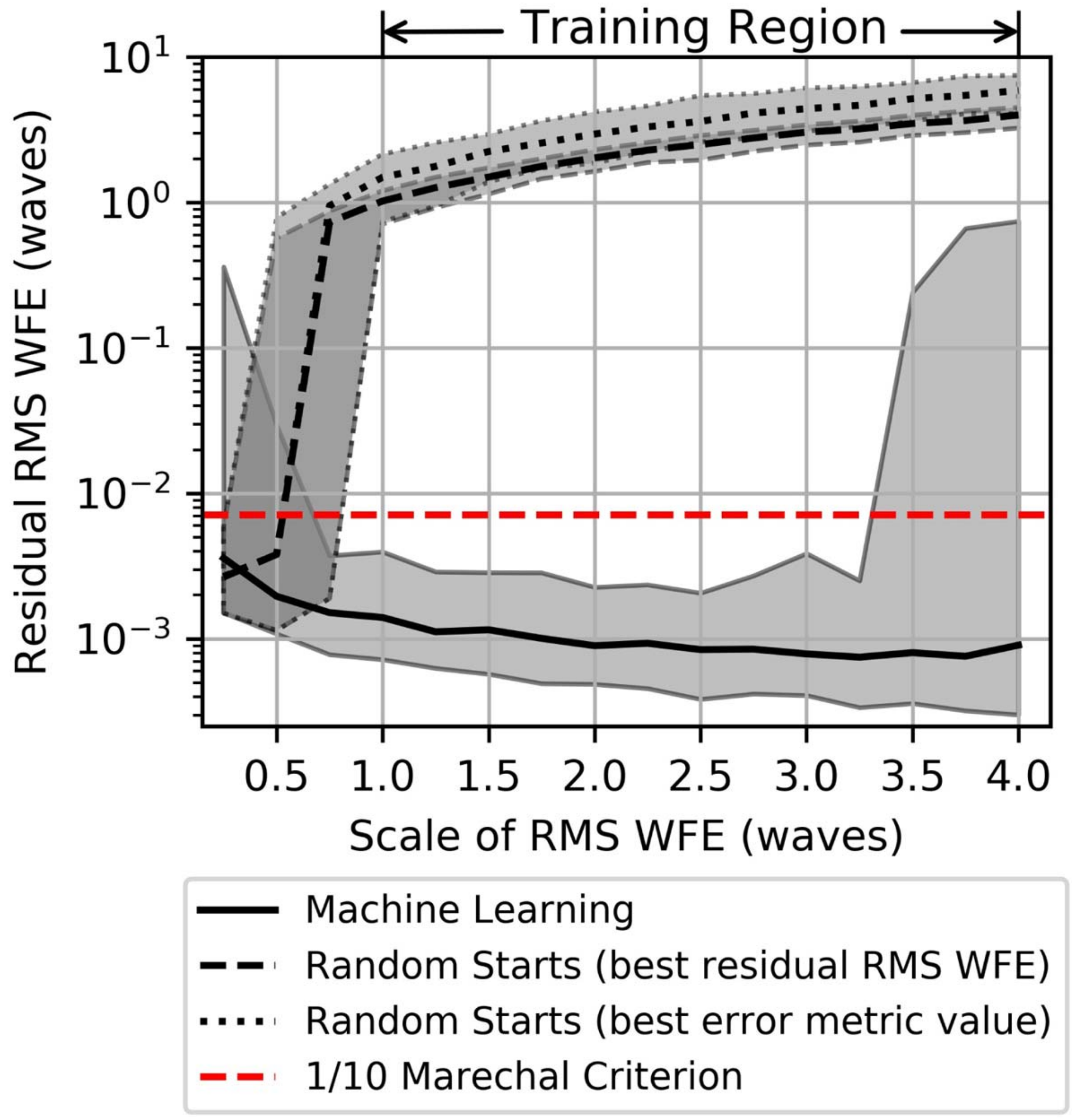
Co-phasing of JWST (**Paine & Fienup 2018**) in two steps:

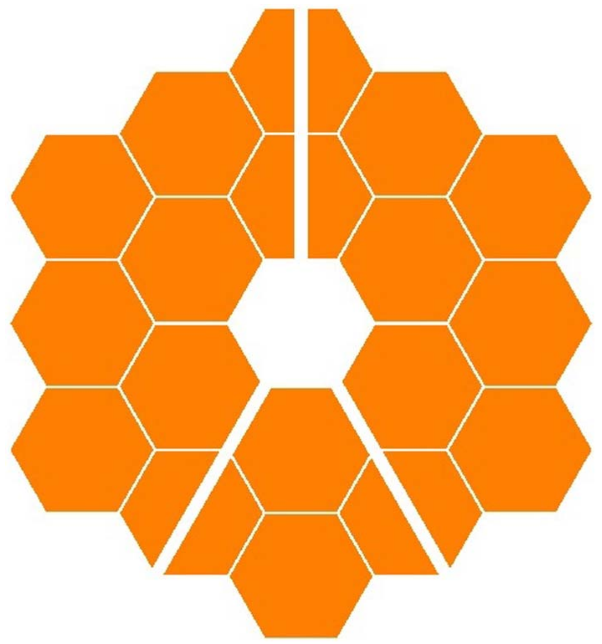
1. CNN to provide initial estimates

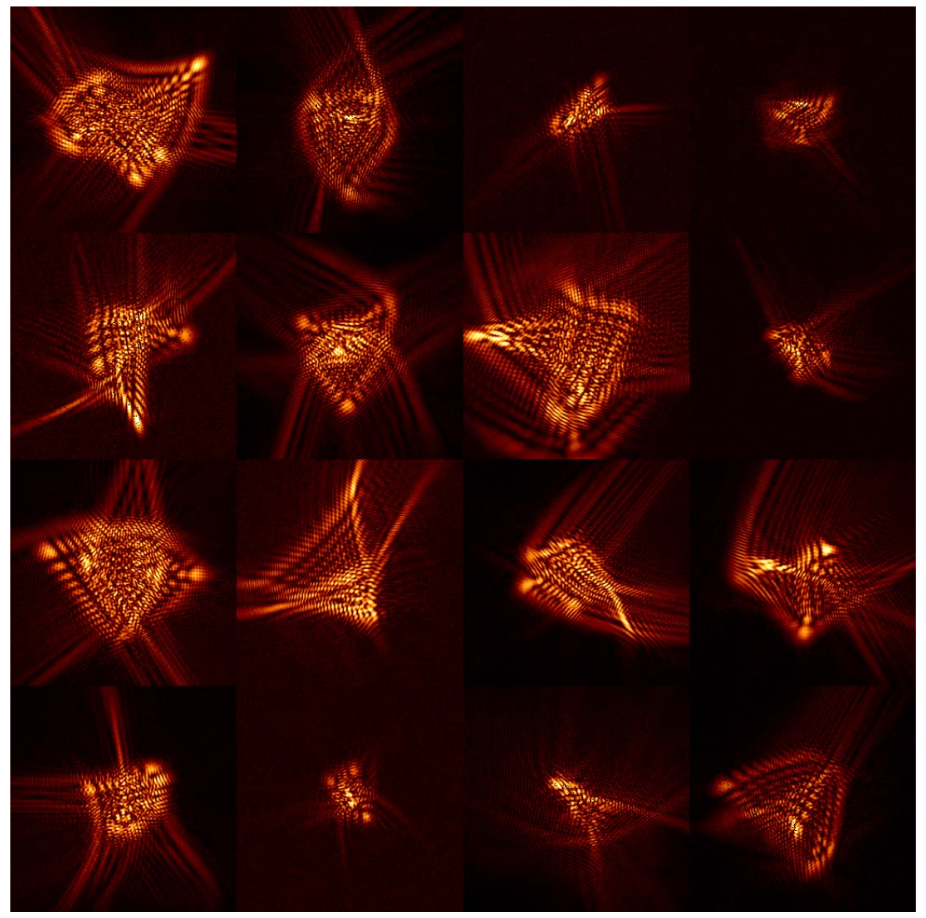
2. Gradient-based optimisation

▹ Greatly enlarge the capture range

- Gradient-based optimiser can use algorithmic differentiation (Jurling & Fienup 2014). *Automatic differentiation* provides the gradients of your model for ‘free'







Paine & Fienup 2018

★Other examples :

- *Zernike WFS* and the *Lyot-based low-order WFS* (Allan+2020a, Allan+2020b)
- *Reconstruction of PyWFS signal* (Landman & Haffert 2020)

Neural networks for adaptive optics

Setup explored by **Andersen et al. 2020**

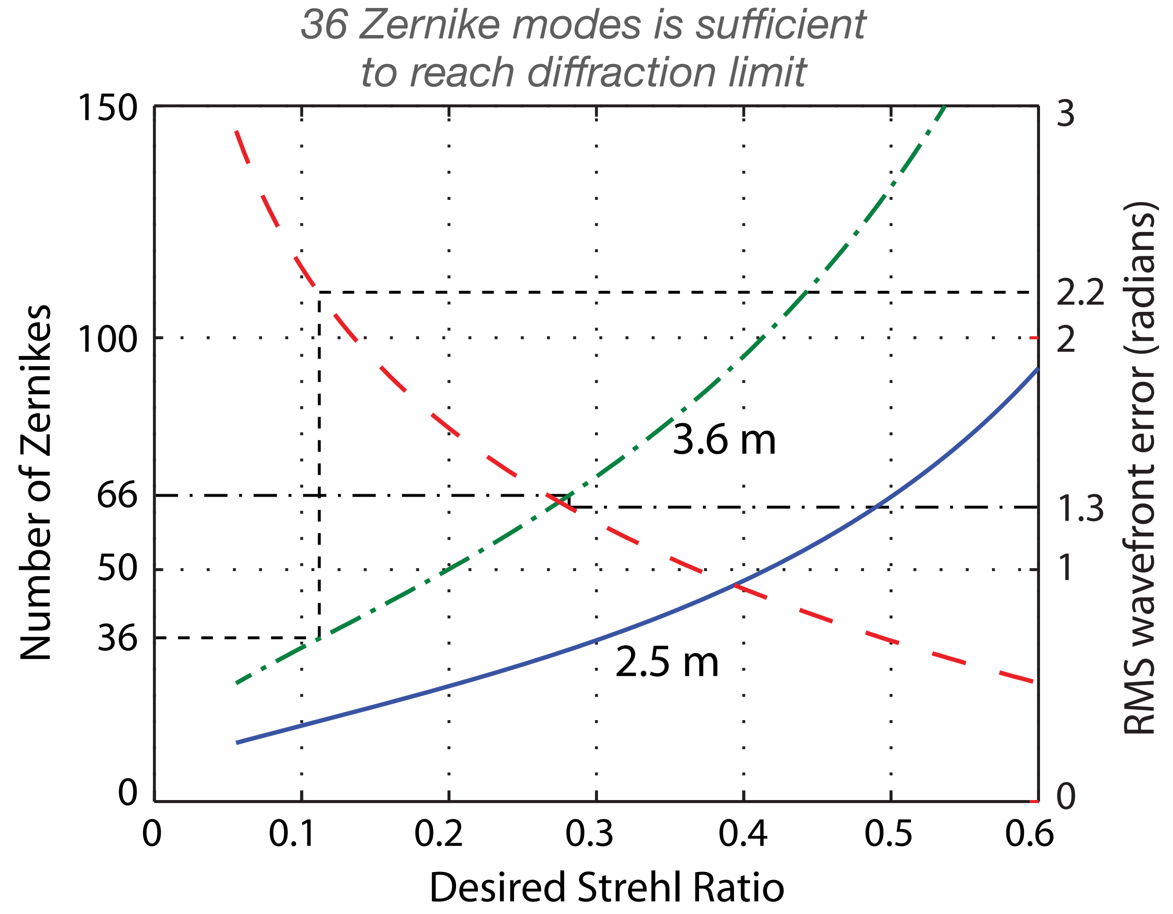
- Realistic simulations (noise, star magnitude, bit depth, polychromaticity)
- D=3.6m, r_0= 17cm, correction 36-66 Zernike
- Using in-focus and out-of-focus images

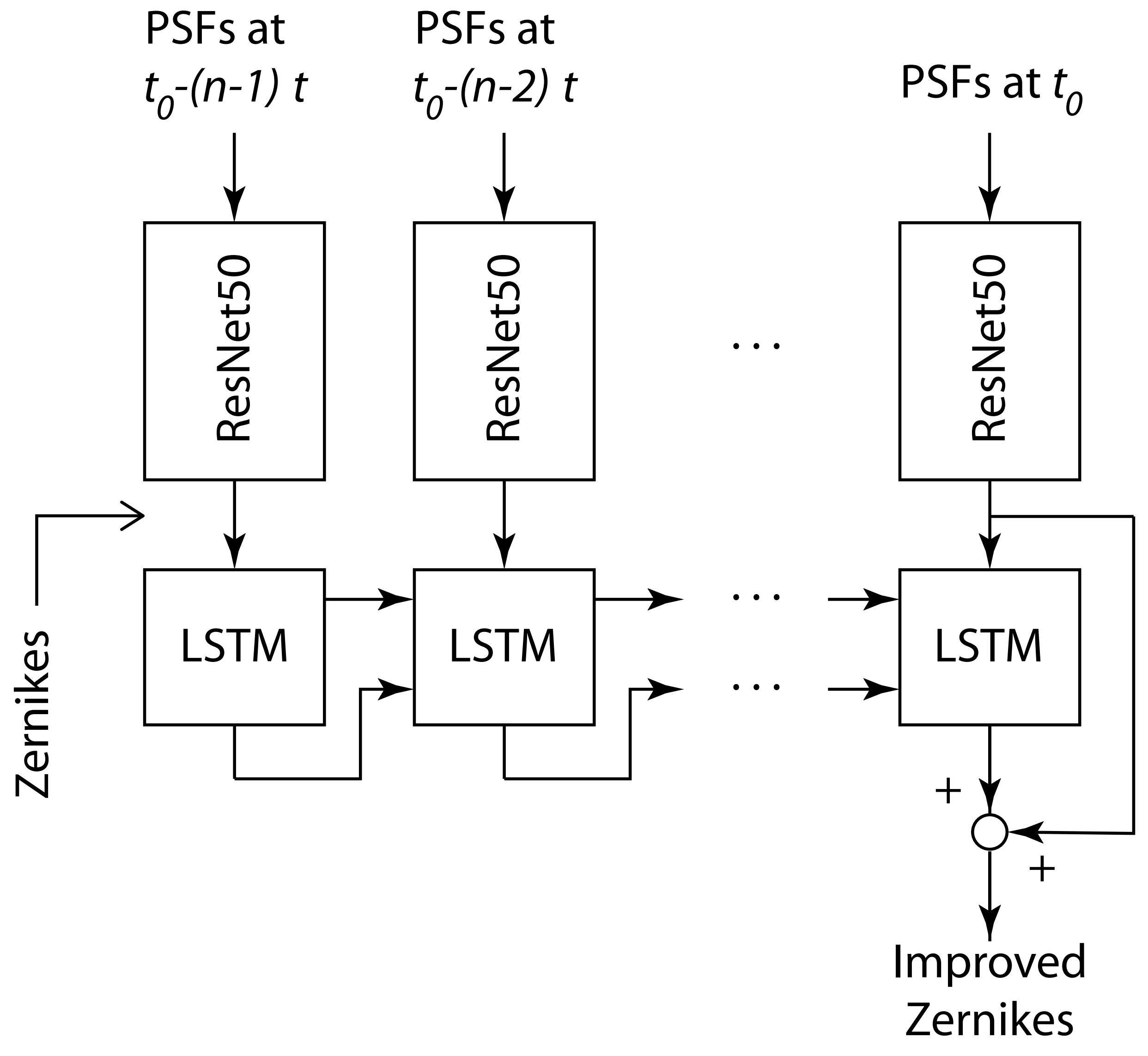
Single-shot wavefront sensing with ResNet-50:

- ~180nm rms WFE with 36 Zernike

Multishot WFS:

- *Closed-loop*, down to 150nm rms (~fitting error), *sampling freq 50Hz*
- Using RNN to improve the ResNet-50 Zernike estimate





★See also Allan+2020, SPIE, using GRU for the ZWFS

FPWFS with ML

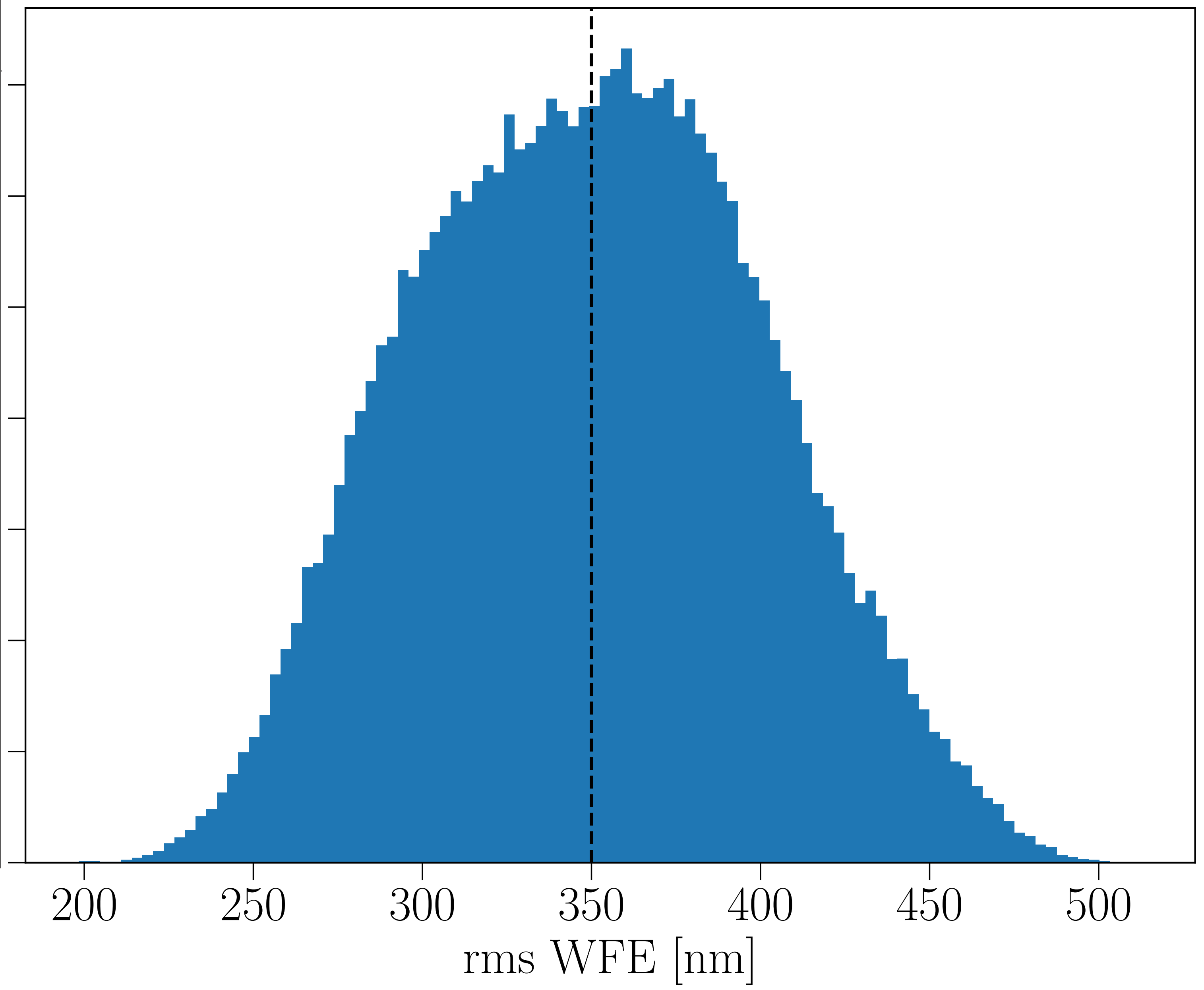
NCPA and performance of CNN

Simulation setup, **Orban de Xivry et al. 2021**

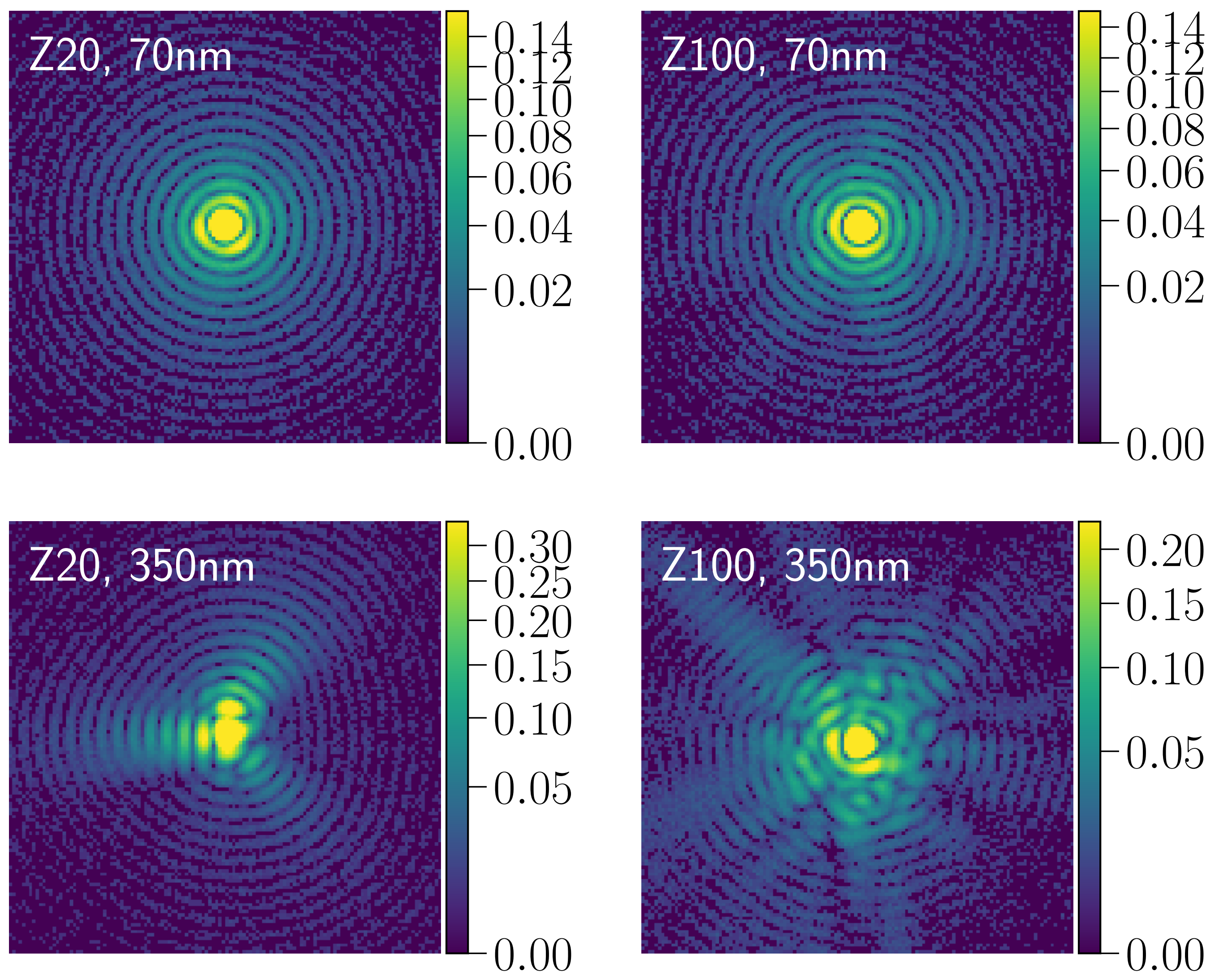
Wavelength, tel. Diameter	2200nm, 10m
Input WFE	70-350nm rms (0.2-1 rad)
# of modes	20 - 100
Pixel scale FoV	0.2 λ/D (0.01”/pix) 28.5 λ/D (1.4”)
Defocus diversity	λ/4

Each Zernike coefficients is draw from a uniform distribution

Distribution of WFE



Example of PSF



- Can CNN-based framework for FPWFS be applied to NCPA measurements?
- Performance limit, robustness of the approach ?

Metrics

Fundamental limit for wavefront sensing

- Particular nature of light: photon noise
- Fisher information matrix [1] $\sigma_j^2 \geq 1/(4N_{ph})$ per independent mode *j*
- Most sensitive: Zernike wavefront sensor [2, 3] $\sigma_j^2 \geq 1/(2N_{ph})$
- Focal plane sensitivity is further reduced

$$\sigma_{FP} = \sqrt{\frac{N_{zern}}{n_{img} N_{photons}}} \text{ [rad]}$$

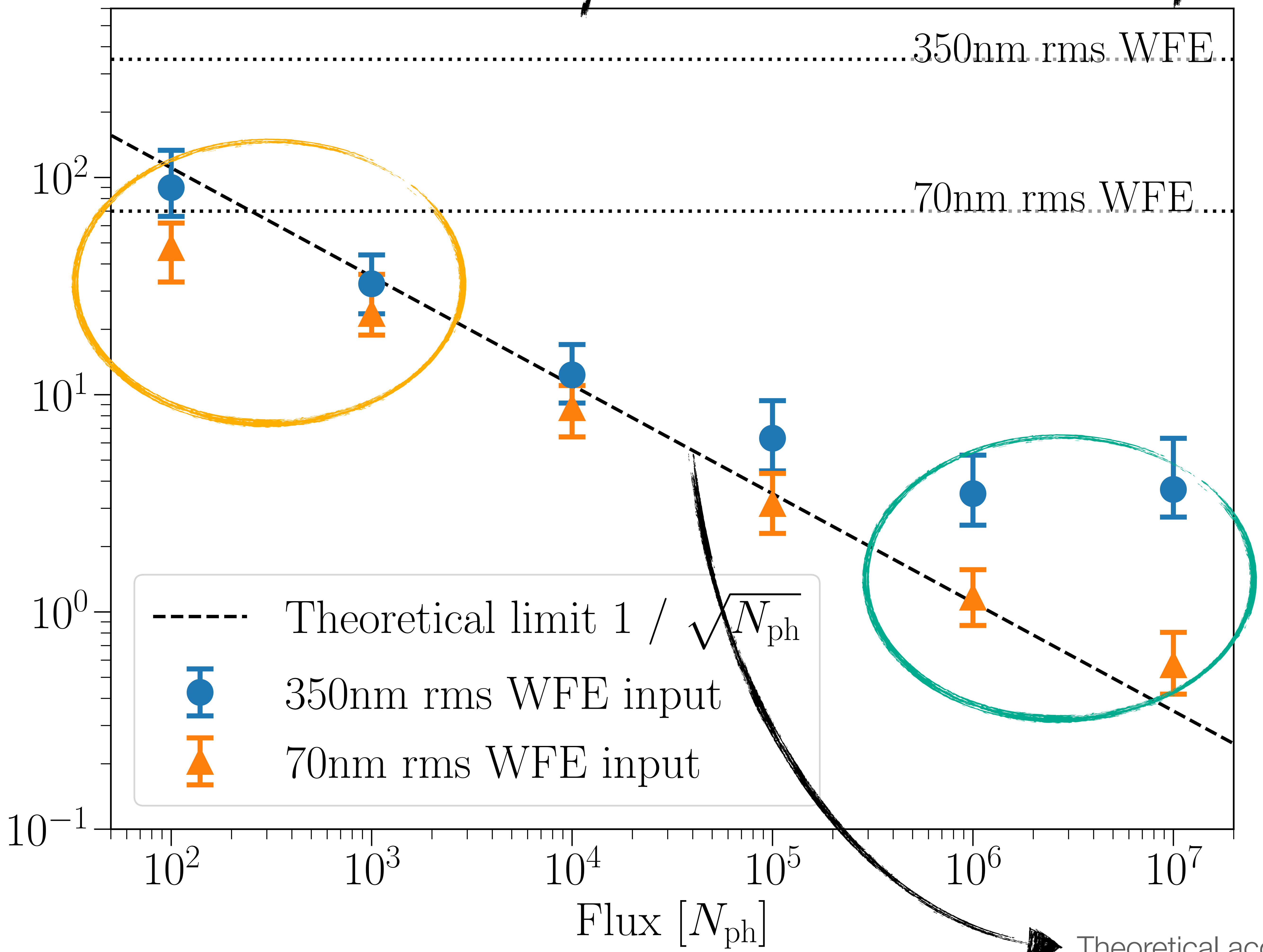
[1] Paterson 2008, 2013
[2] N'Diaye et al. 2013
[3] Chambouleyron et al. 2021
[4] Guyon 2005

Results

Performance limit

$m_K = 14.5$ $m_K = 7$ for T_i = 1sec

ResNet-50 for 20 Zernike's



Theoretical accuracy

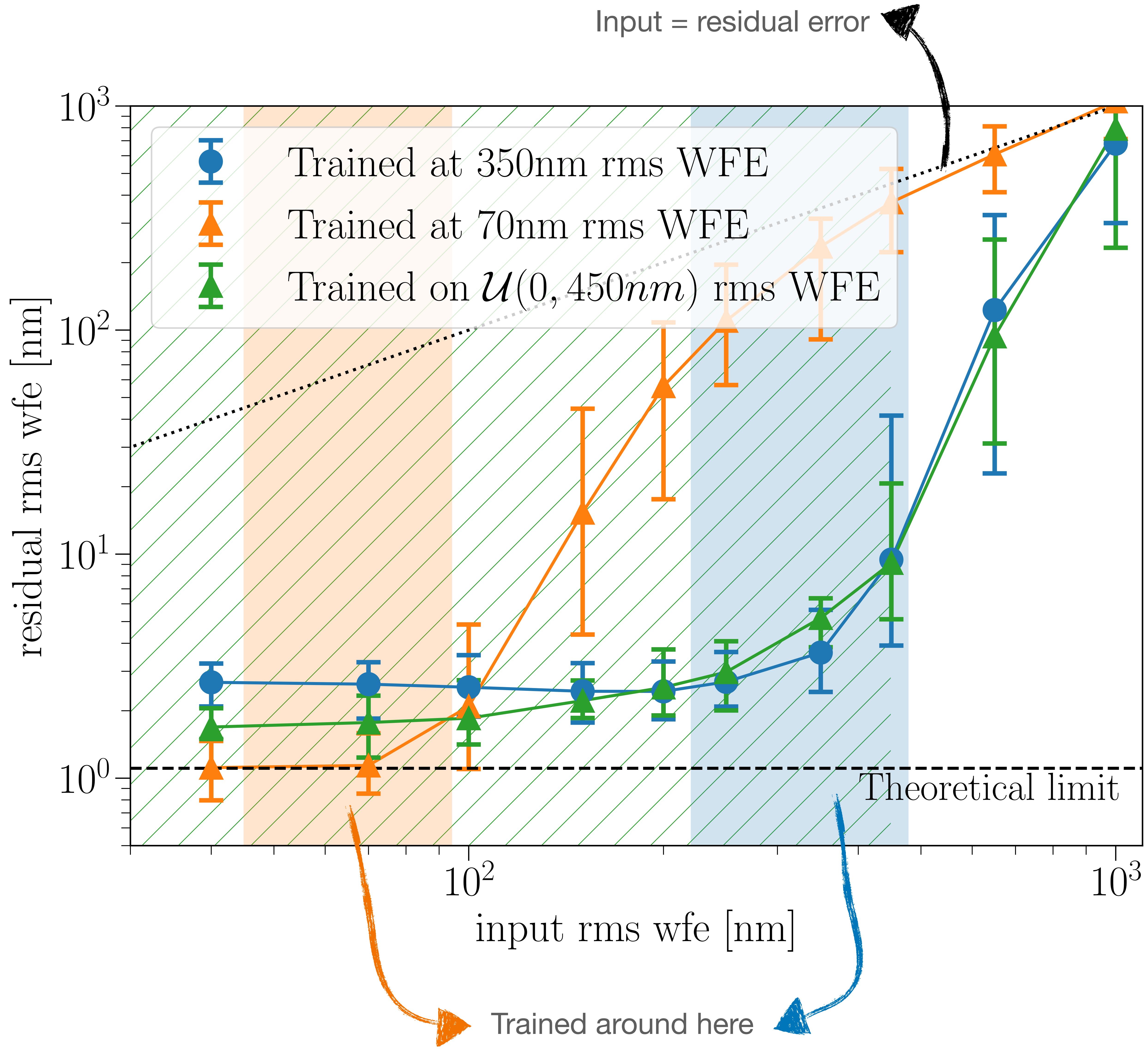
- ✳ Every point uses a different model
- ✳ Evaluation on 100 entries
- ✳ 'Excess' error for larger level of aberrations and large flux: numerical limitation
- ✳ Implicit regularisation at low flux level (due to data distribution)

Results

Dynamic range

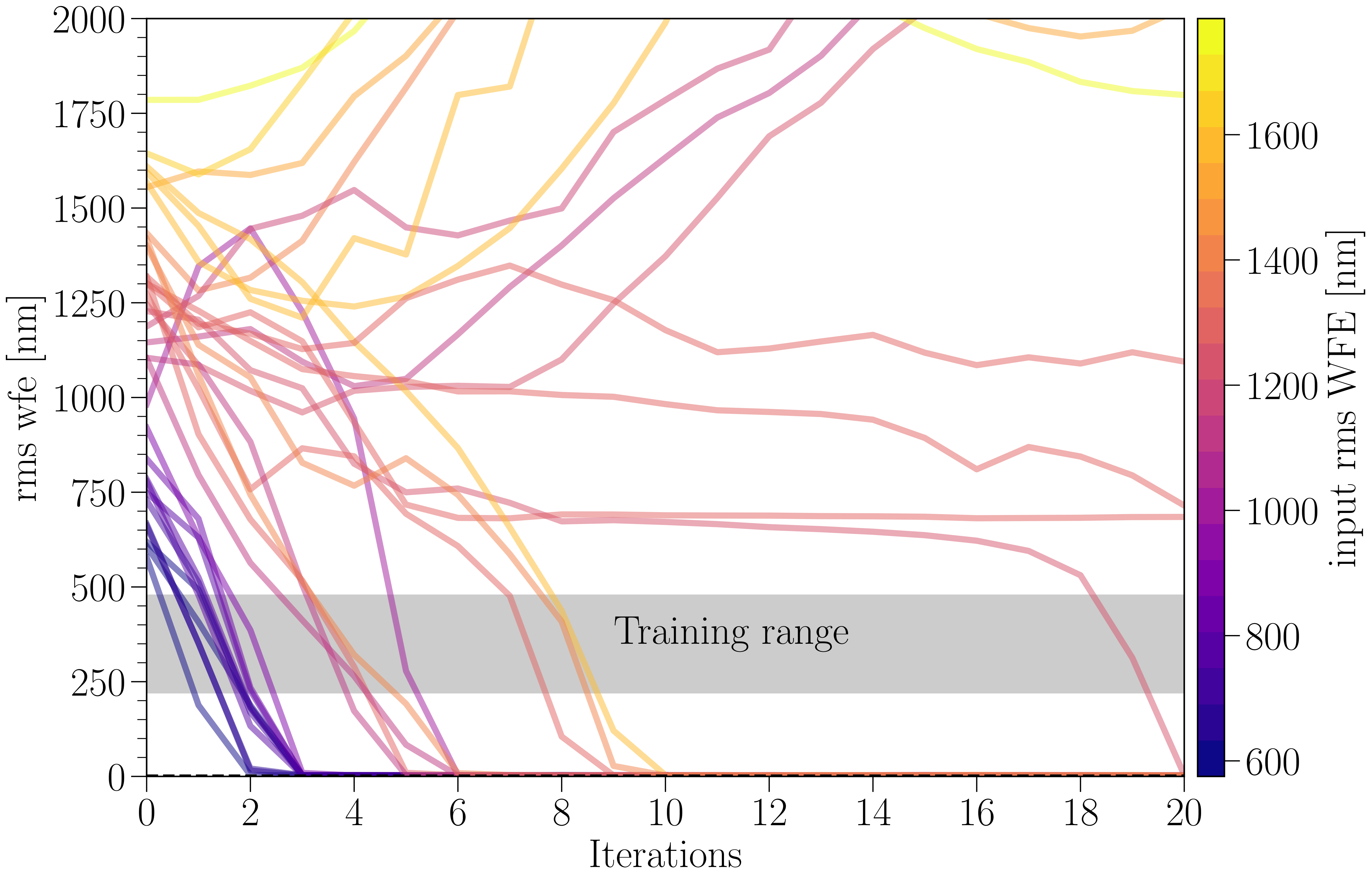
✓ Below training : constant accuracy

~ Above training : quickly increasing

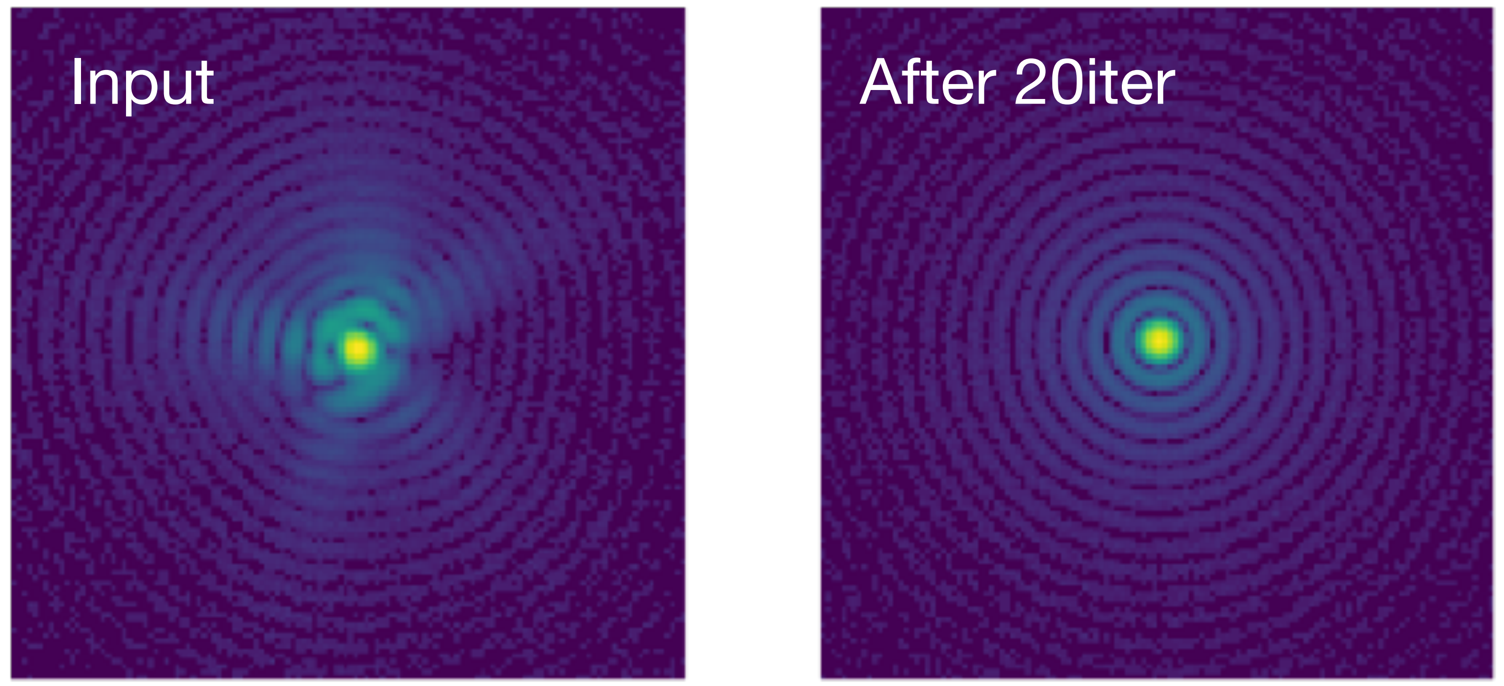


Results

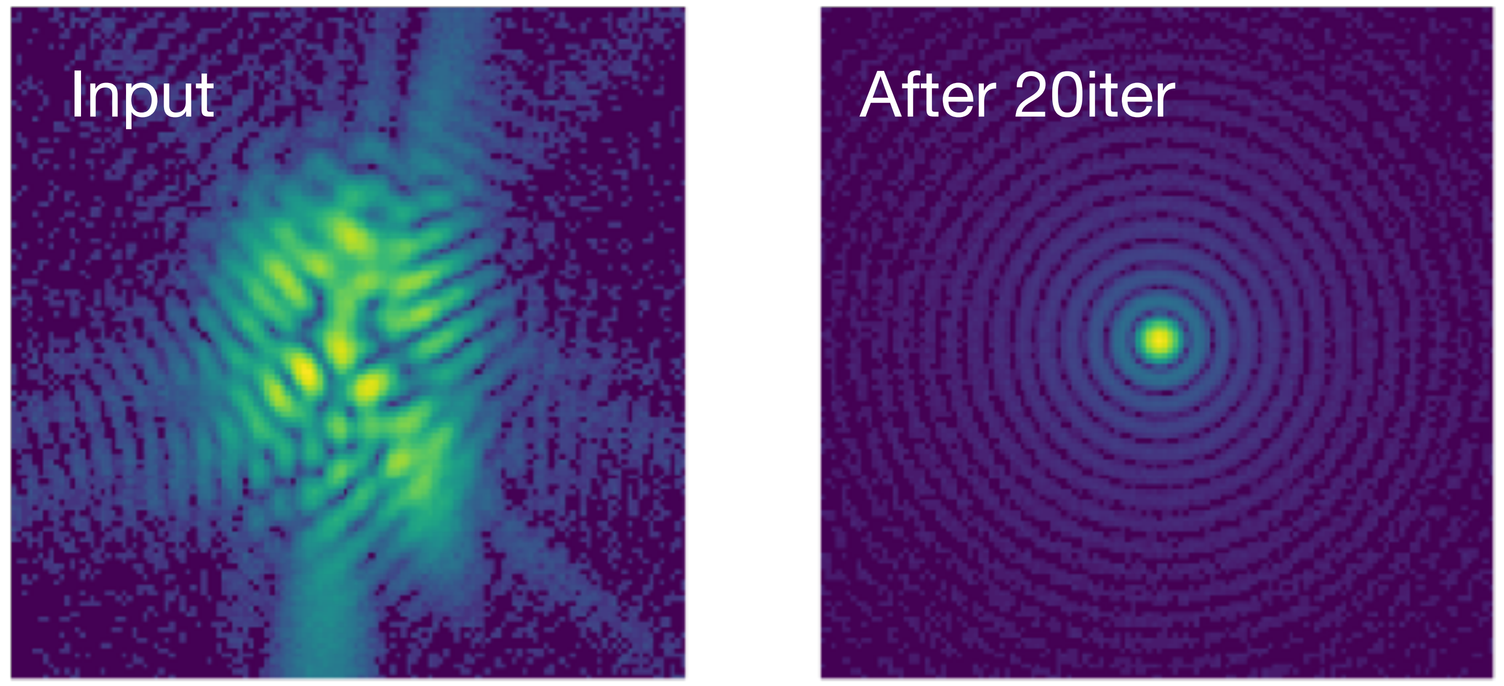
Dynamical range: application in closed-loop



~320nm rms WFE input



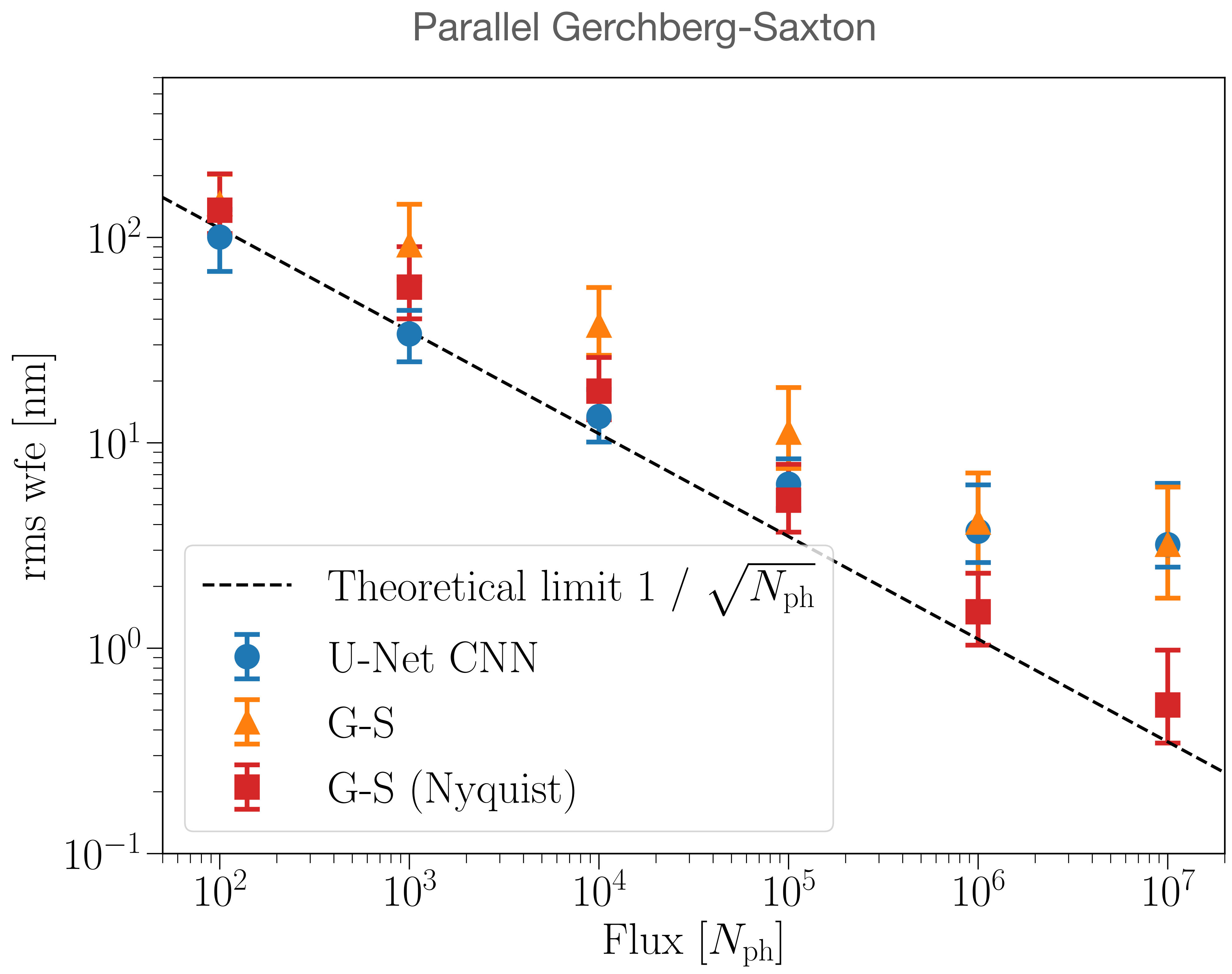
~1µm rms WFE input

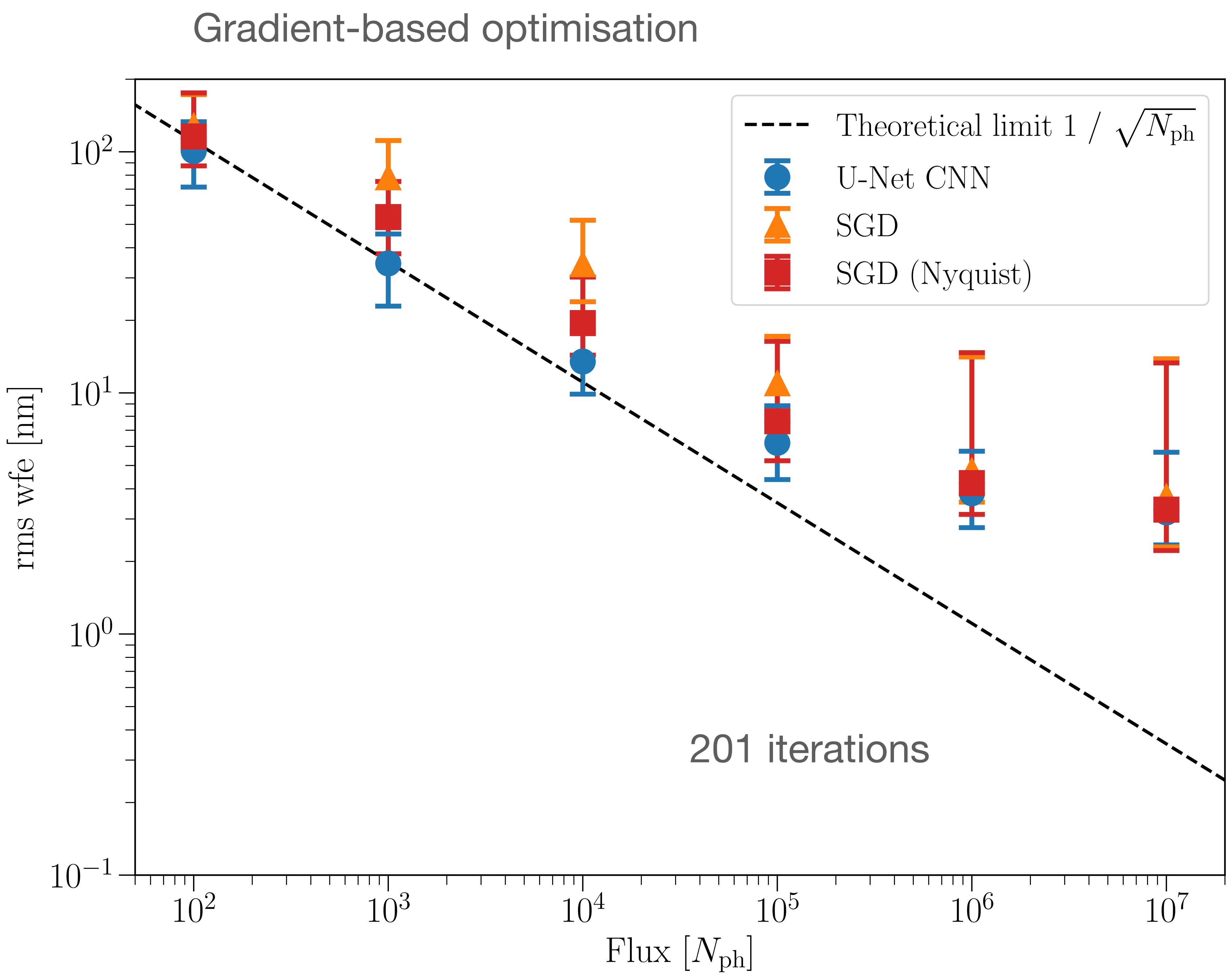


- Works well beyond training range
- Other strategy would imply to train on a larger range of aberrations or use different network architectures

Comparison

With Gerchberg-Saxton phase retrieval &
Gradient-based optimisation with automatic differentiation





★Using automatic differentiation for phase retrieval
- (Jurling & Fienup 2014),
- Peng et al. 2020
- Wong A. et al. 2021

★ See also MORPHINE on Github (Pope B et al.)

Practical consideration

Computational cost

- ✦ Typical training set size: 100,000
- ✦ Training time : several hours to days
- ✦ Typical inference time of typical architecture

  ~ several ms (with 1 x RTX2080Ti)

- ✦ But should only be seen as *upper bound*:
  - ✦ Faster training and inference (<1ms) with **lighter architecture**, and downsampling # of pixels.
  - ✦ Compression technique, pruning, etc.

Architectures	Number of parameters (M)	FLOP (G)	Model size (MB)
ResNet-50	23.71	8.22	91
U-Net	13.40	15.54	52

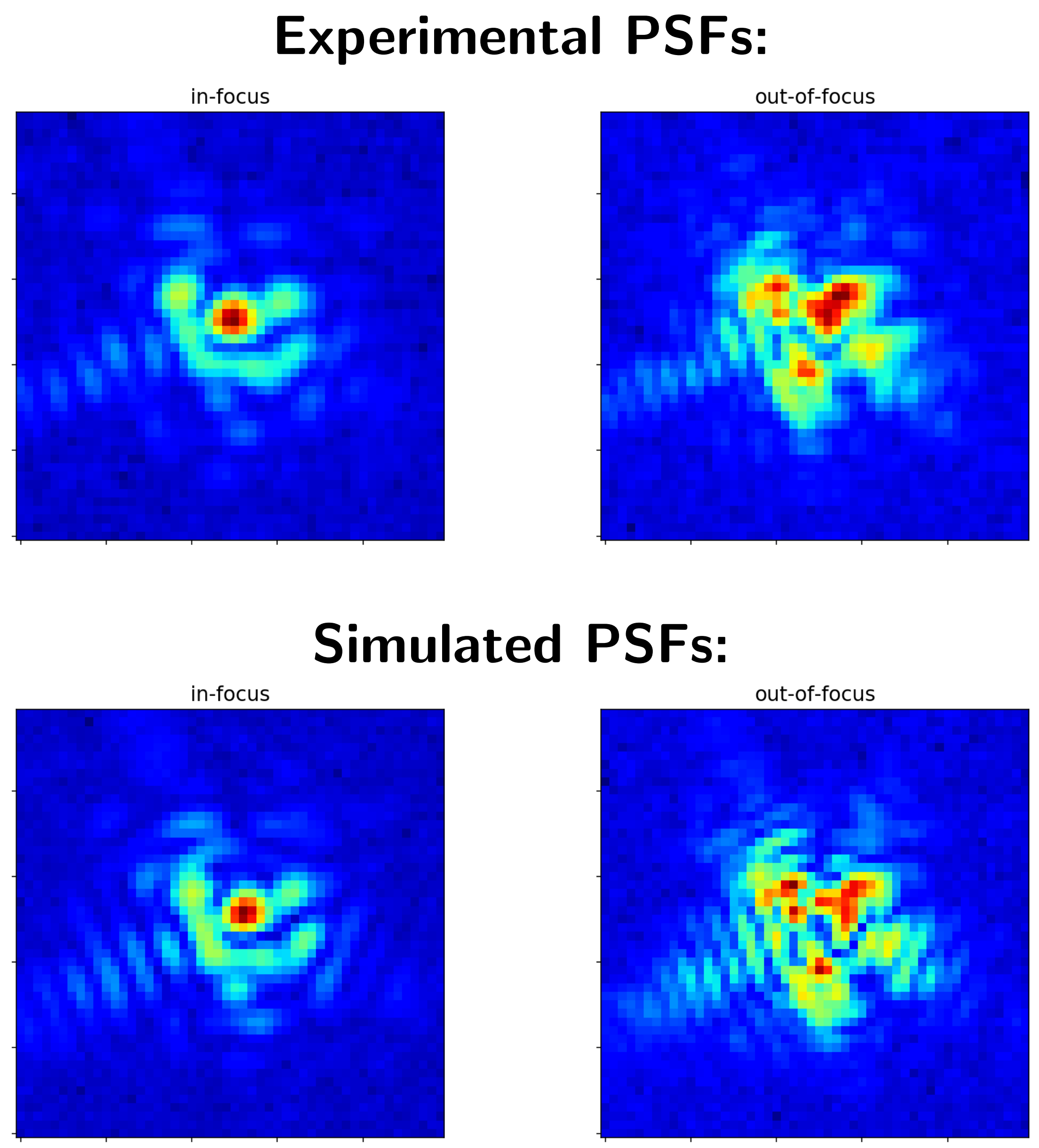
For 128x128 gridsizes and 100 Zernike's

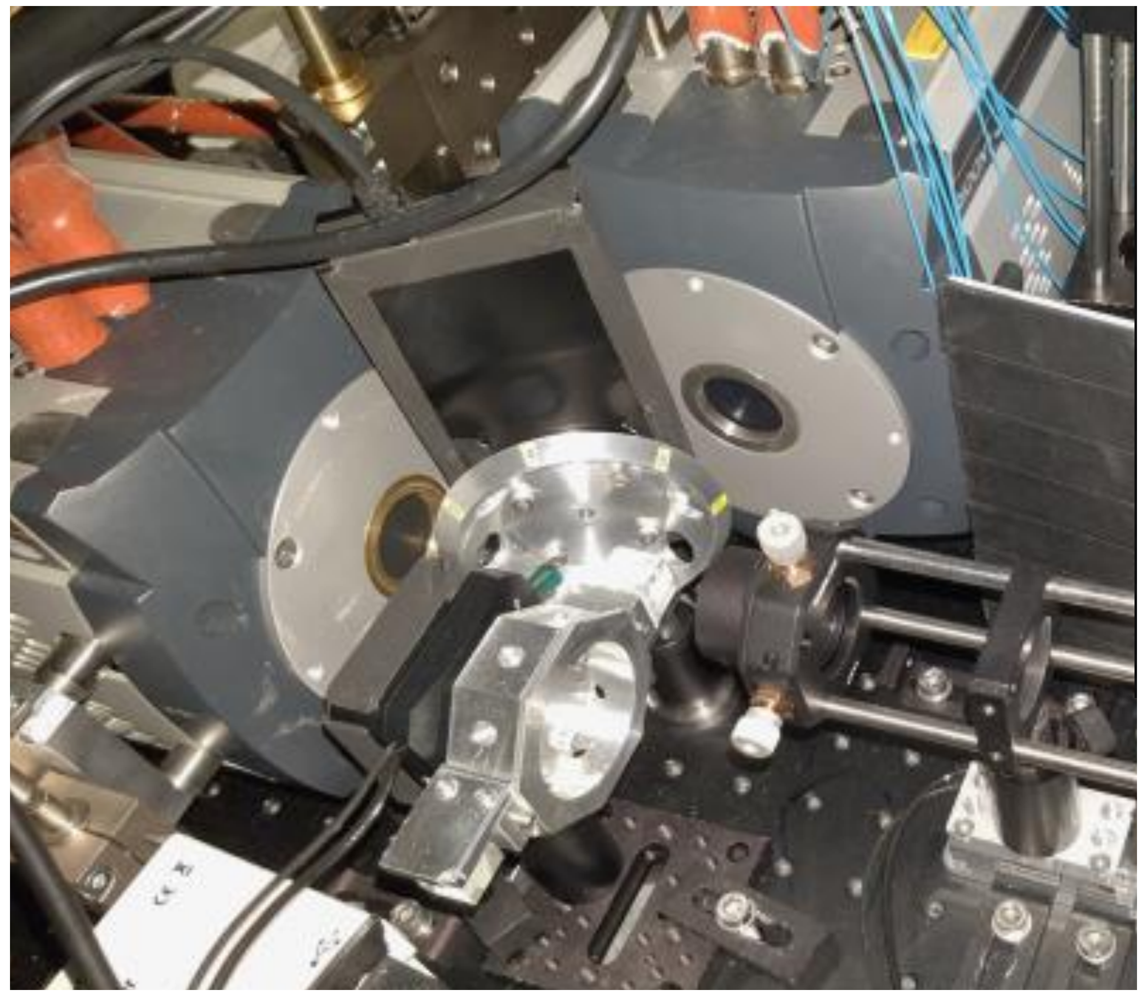
★See also Weinberger et al 2020

In practice

Application to lab data on SCExAO

- Training on lab data — see ***Kyohoon***'s talk
- Transfer learning / fine-tuning of model using real data — see ***Maxime's*** talk

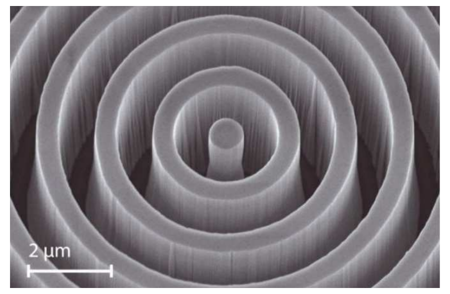


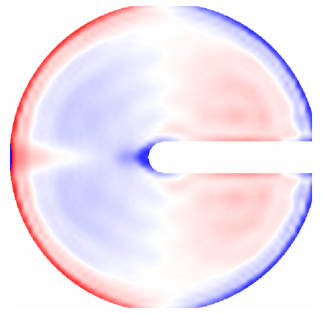


[1] Martinache 2013
[2] Vievard et al. 2019
[3] Bos et al. 2019
[4] Codona & Kentworthy 2013
[5] Gonsalves 2002
[6] Bos et al. 2020
[7] Cranney et al. 2021, this workshop
[8] Norris et al. 2020

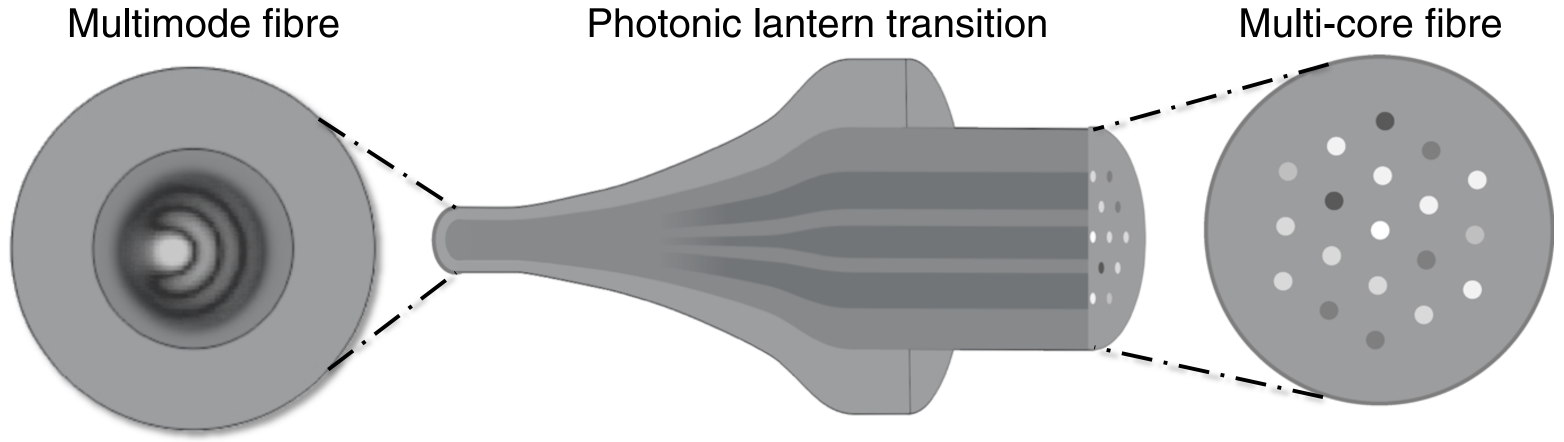
Lifting the sign ambiguity

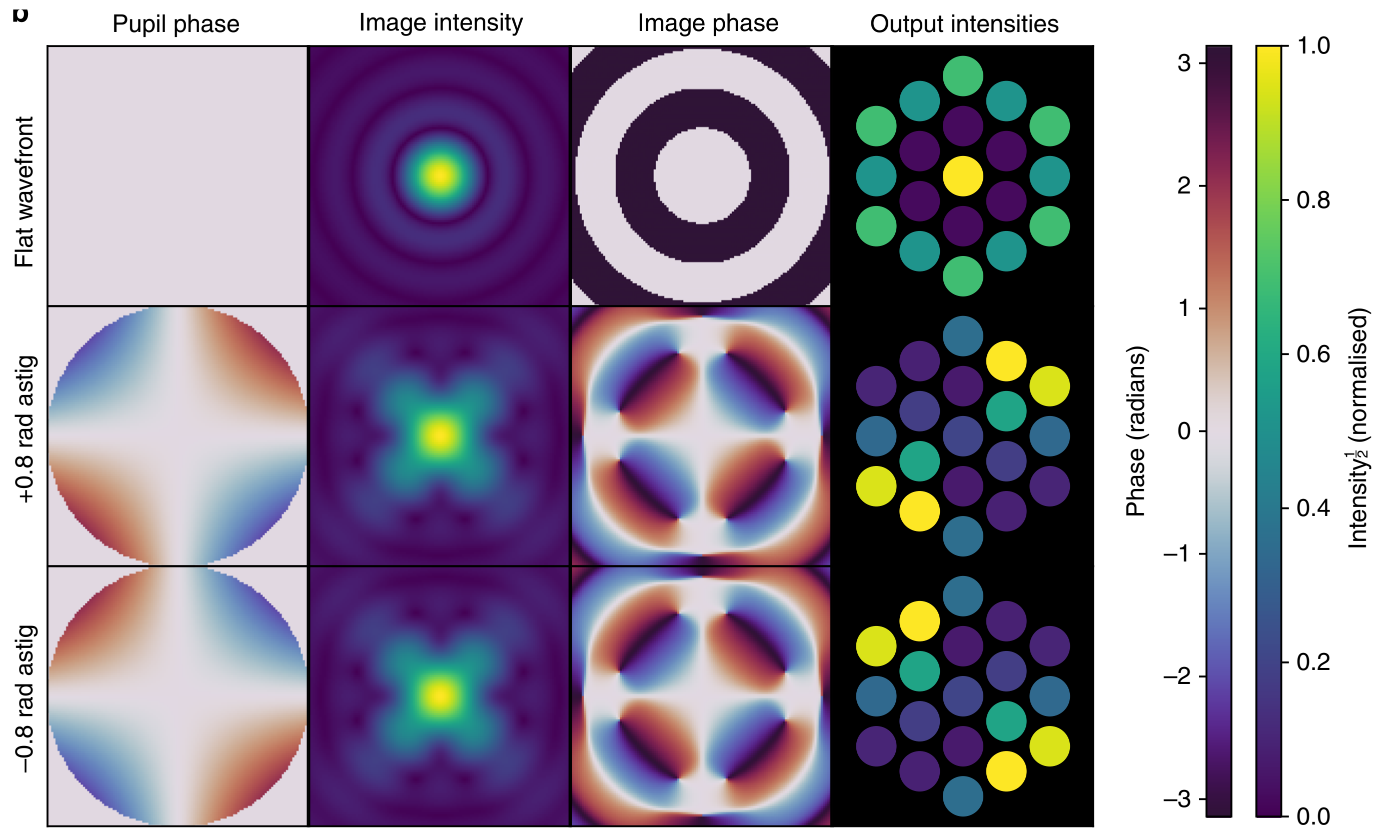
Selected approaches





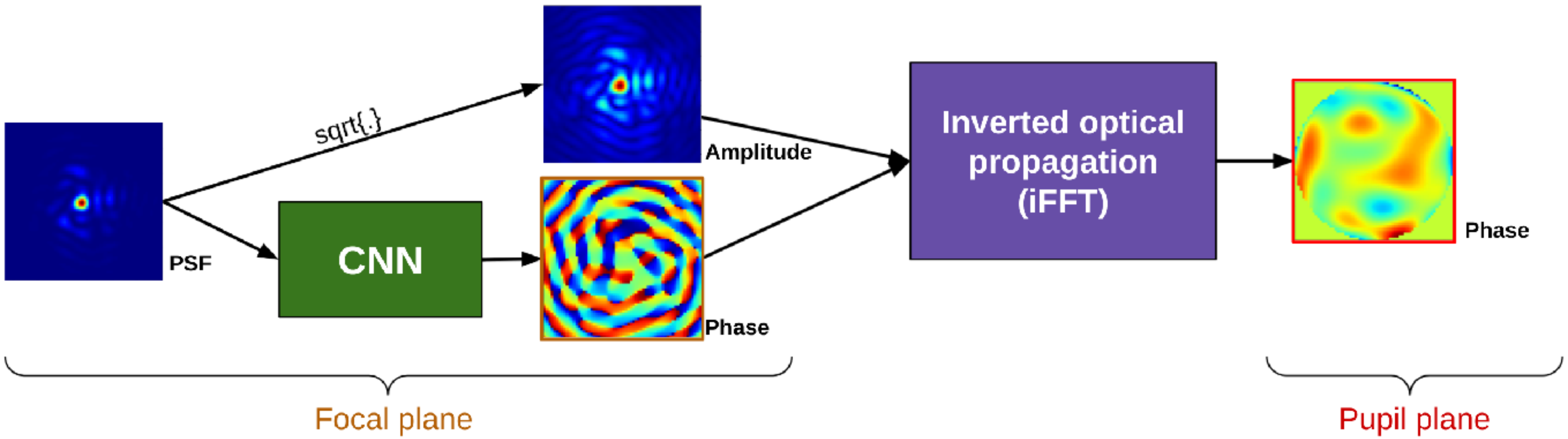
- *Vortex coronagraph* : see Maxime's talk
- *Asymmetric pupil* : LWE [1,2], FPWFS with the vAPP [3]
- Using the *WFS telemetry* as known diversity probes introduced by the turbulence : PSI [4]
- Sequential diversity using the *DM telemetry* [5, 6, 7]
- *Measure the phase*: All-photonic WFS [8]





NN coupled with physical model

Backward approach



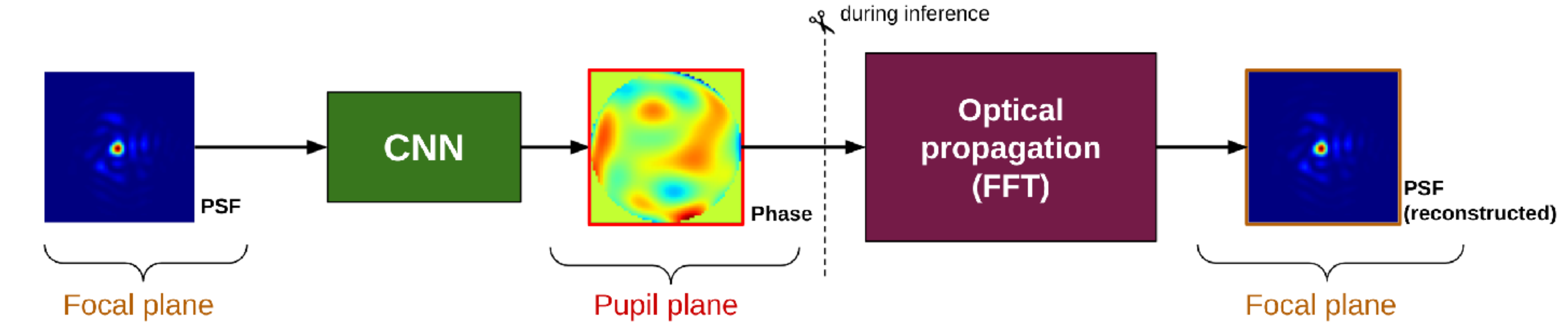
Courtesy to Maxime Quesnel

- Faster inference (generalisation), more insight on the optical system
- Still a supervised framework

★See also Peng at al. 2020,
for a somewhat similar implementation applied to holography

NN coupled with physical model

Forward “auto-encoder” approach



Courtesy to Maxime Quesnel

- Provides an *unsupervised framework*
- Possible application: trained or untrained
- Possibility to parametrised the physical model ?

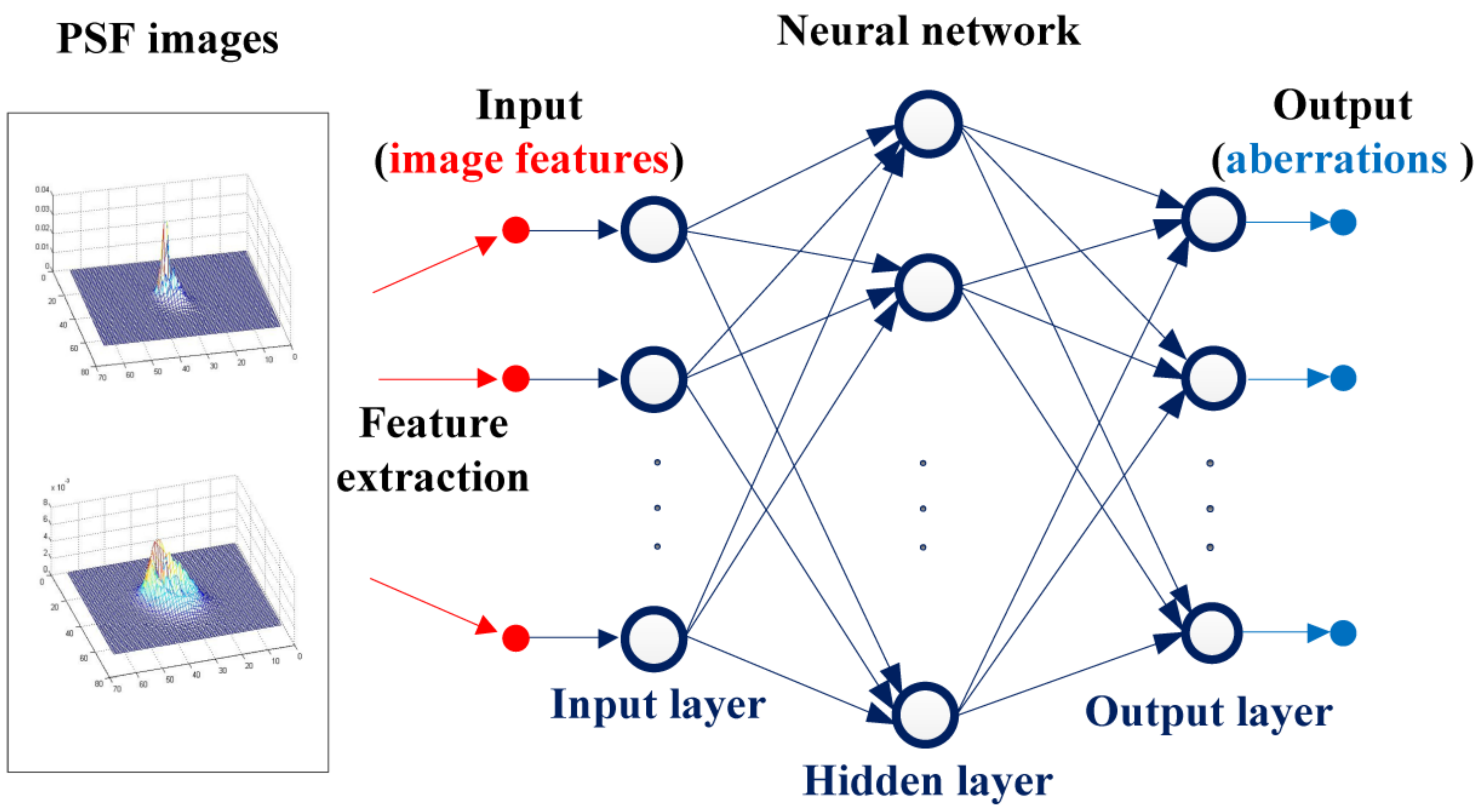
★ *See also*
  - Fei Wang el al. 2020, Phase imaging with an untrained neural network
  - Liaudat Tobias et al., 2021, Instrument response for Euclid
  - Emrah Bostan et al. 2020, Phase microscopy

★ Other application of autoencoder:
  - Pou et al. 2020, denoising of WFS images

Focal-plane decomposition

1. Decomposition of the PSF in a set of features -> compression of PSF image

2. Use a NN to map the image-plane coefficients to the pupil-plane coefficients



Ju G. Et al. 2018

Image decomposition techniques, e.g.:

- via PCA [1]
- Nijboer-Zernike basis [2, 3, 4]
- Tchebichef moment features [5]

Alternative : use an auto-encoder ?

[1] Terreri et al. 2019, WFS4ELT
[2] Magette 2010,
[3] Riaud et al. 2012a, b
[4] Antonello, Verhaegen, 2015
[5] Ju G. et al. 2018

Conclusions

- CNN-based FPWFS is a viable option, being confirmed experimentally
- Other - promising and rather uncharted - approaches:
  - Physics-based frameworks
  - Exploit sequential data via RNN (residual minimisation)
  - Coupling linear model with NN for modelling non-linearity
  - And beyond…