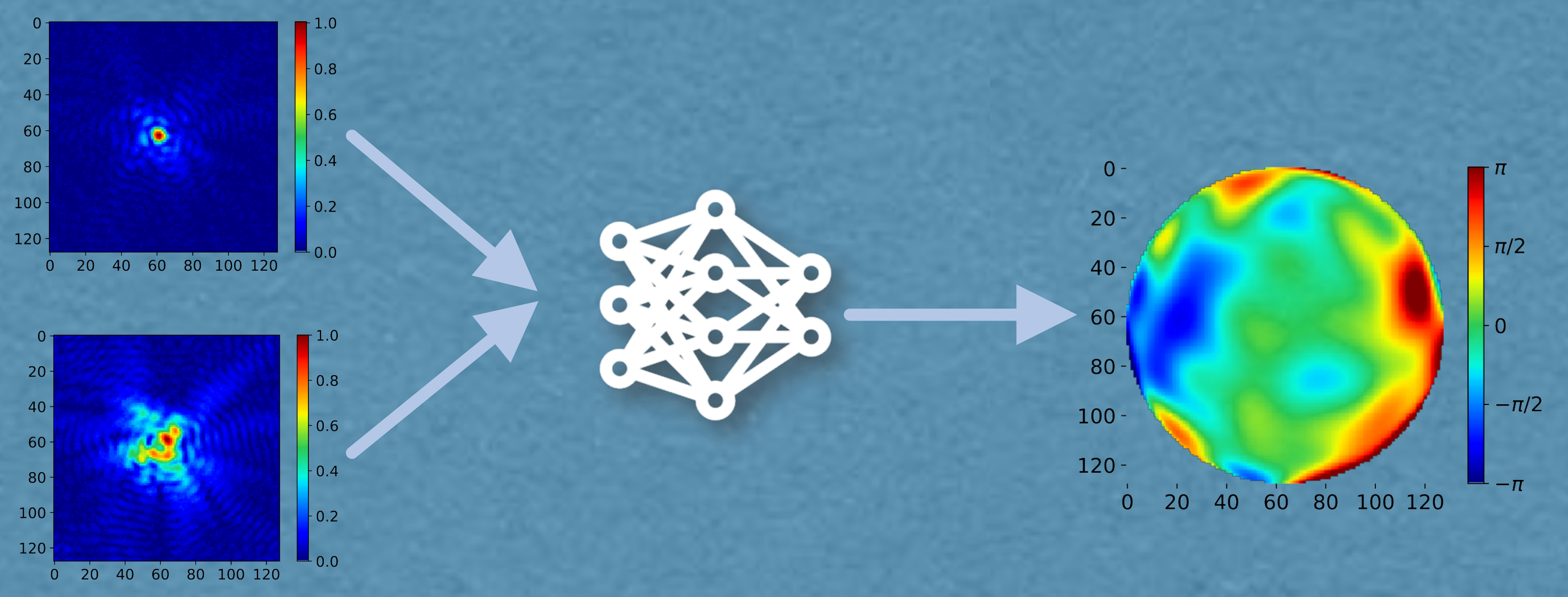
Focal-plane wavefront sensing using machine learning

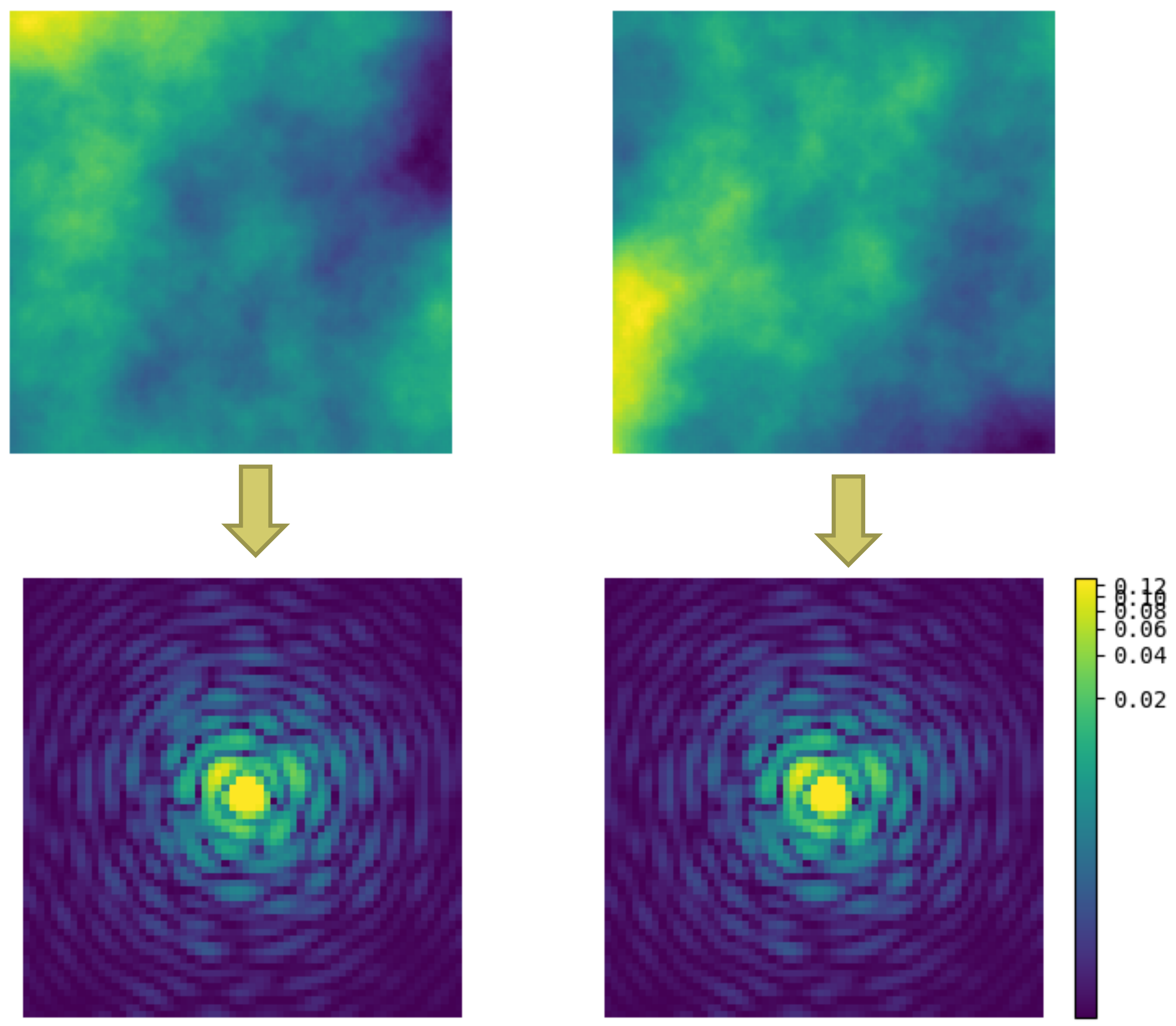


Gilles Orban de Xivry, M. Quesnel, G. Louppe, O. Absil

Wavefront Sensing in the VLT/ELT era V & AO Workshop Week II — October 2020

Focal plane wavefront sensing

Pro's	***Con's***
• High sensitivity	• High computational cost
• Simple opto-mechanically	• Phase ambiguity
• No NCPA or chromatic errors	

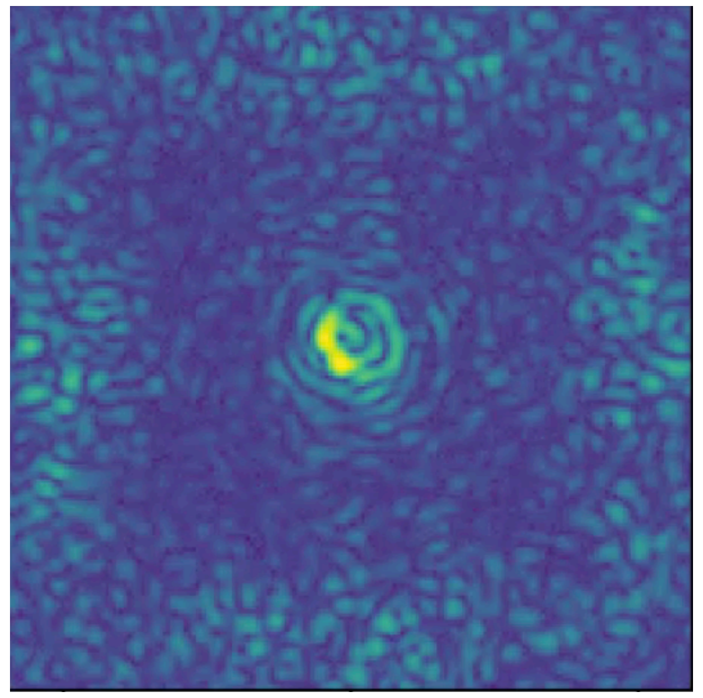


Focal plane wavefront sensing

Two 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

High-contrast image



M. Willson @ULiege

✴Also cophasing (JWST, ELT)

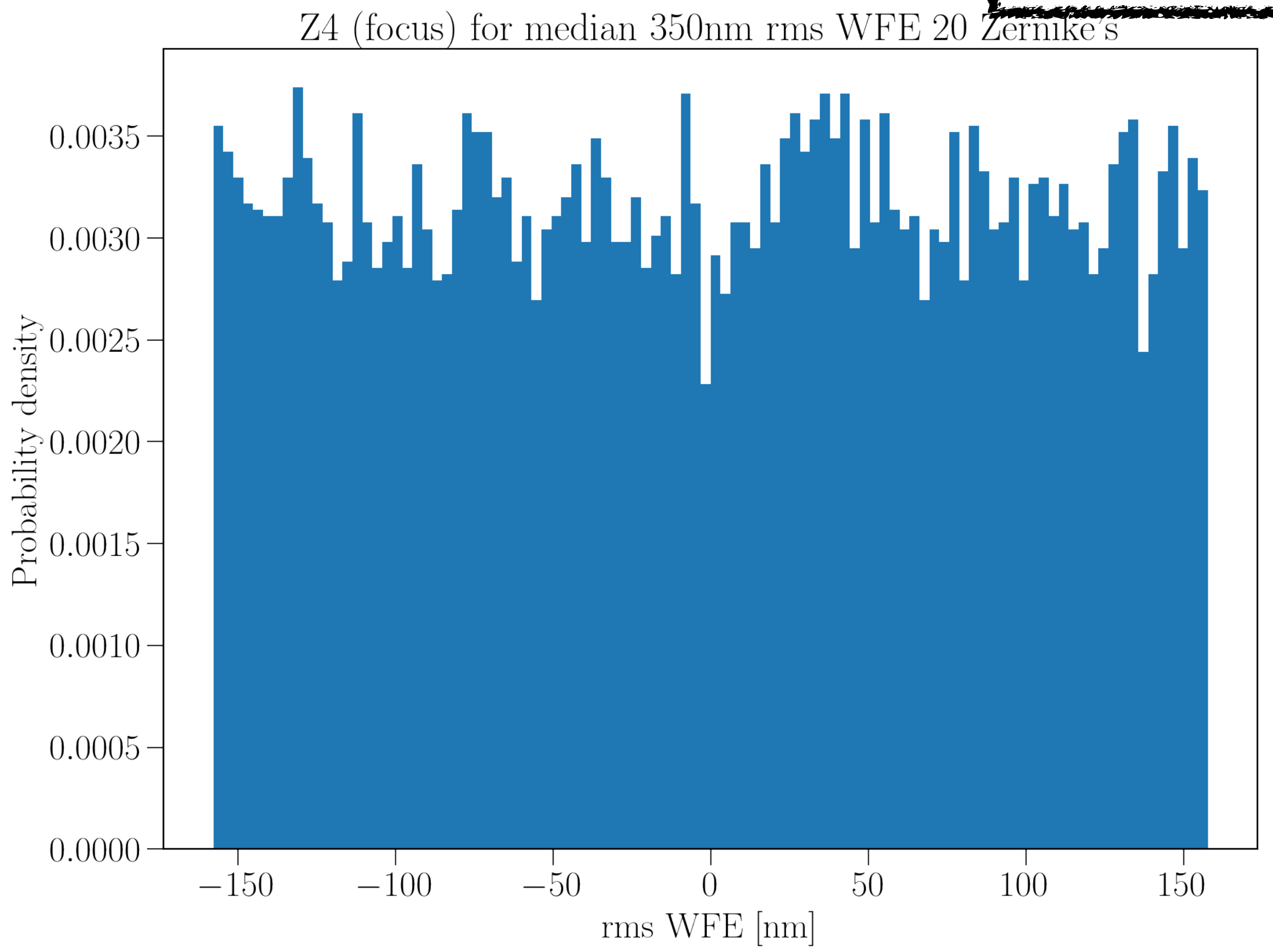
Simulation setup

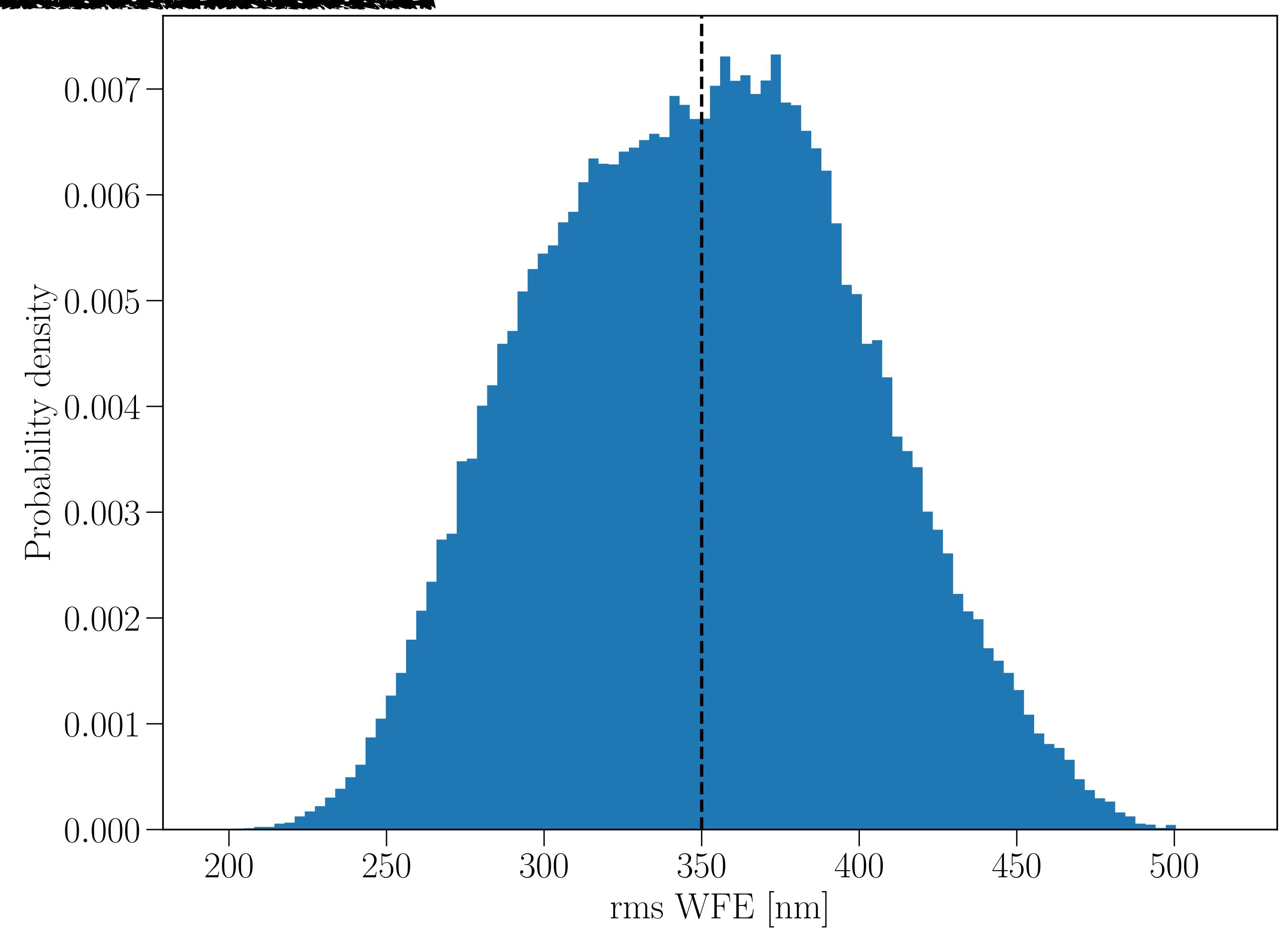
“Labelled" dataset generation

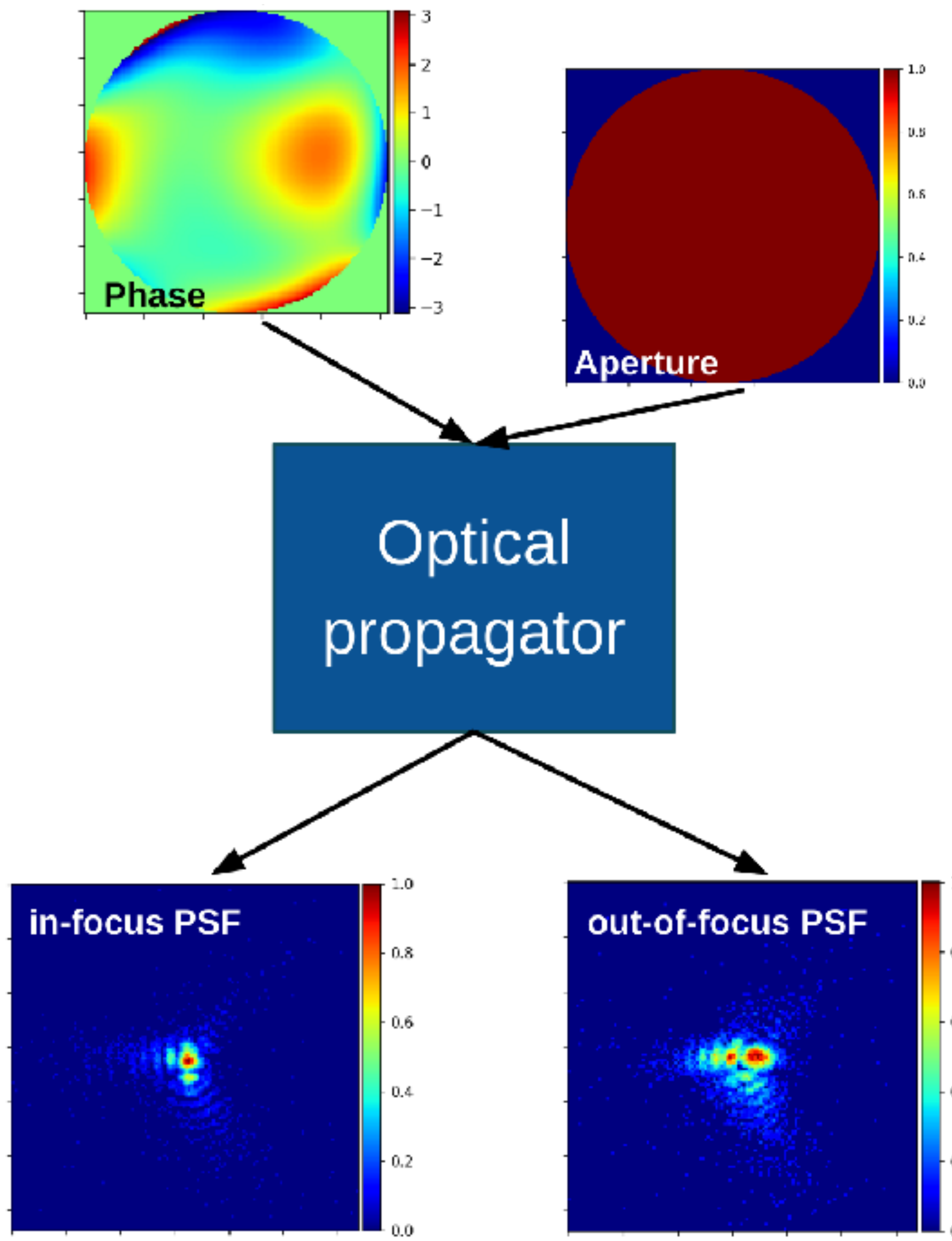
$$\phi(x, y) = \sum_{j}^{N_{zern}} c_j Z_j(x, y)$$

- ✳λ = 2200nm
- ✳Diameter = 10m
- ✳Input WFE = 70-350nm (0.2 - 1 rad)
- ✳Nb modes = 20 - 100 Zernike ; Power law over spatial frequency
- ✳Pixel scale = 0.2 λ/D. (0.01”/pix)
- ✳FoV = 28.5 λ/D (1.4”)
- ✳Size : 100,000 entries

Distributions in the data matter



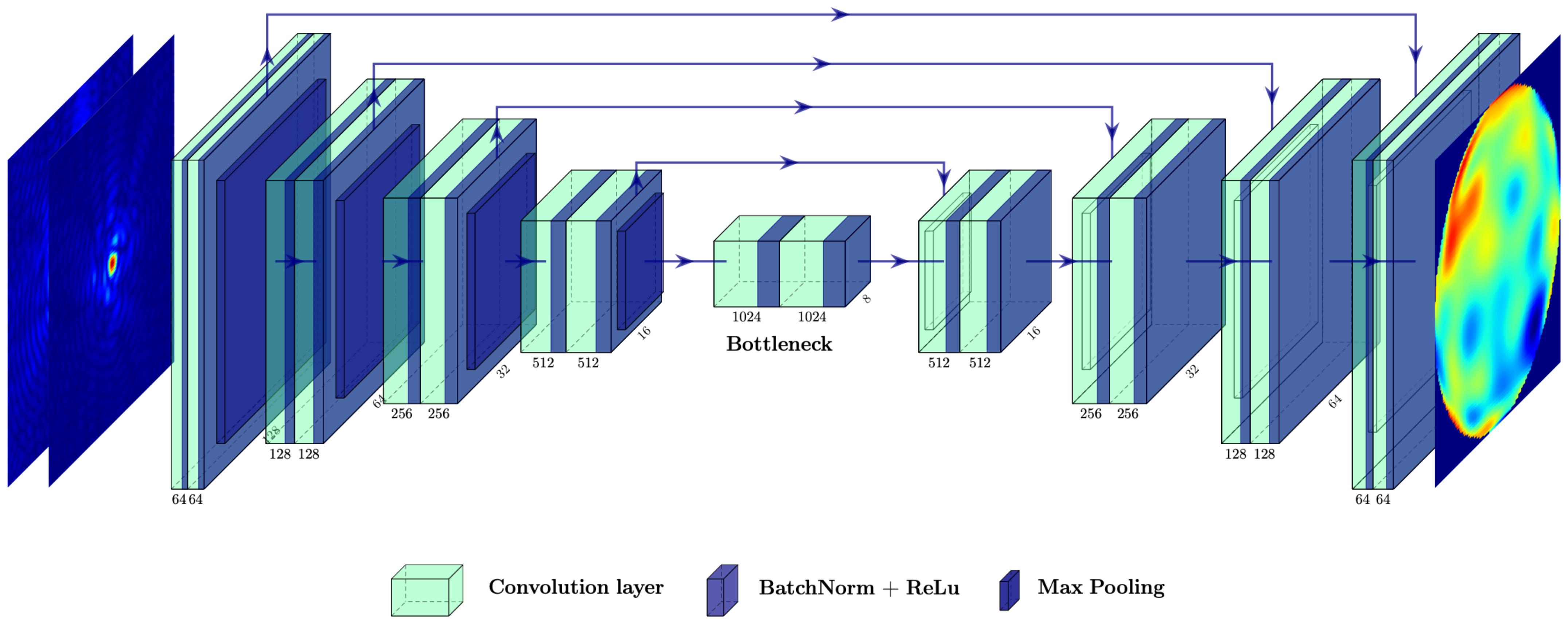




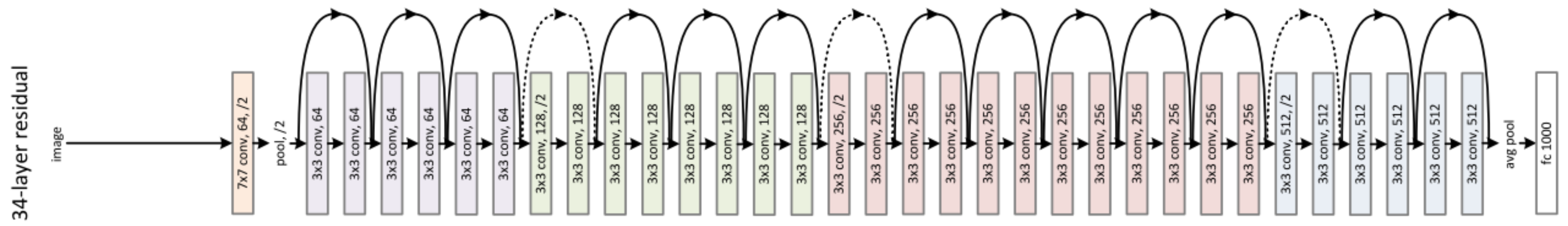
Simulation setup

Network architectures

U-Net



ResNet



- CNN readily available
- Last layer modify to perform regression instead of classification

ResNet-50 → vector c_i

U-Net → phase map. $\phi(x, y)$

Loss function

$$RMSE = \sqrt{\frac{1}{N_{pix}} \sum_{i,j}^{N_{pix}} \left[\phi(x_i, y_j) - \hat{\phi}(x_i, y_j)\right]^2}$$

Metrics

Fundamental limit and robustness

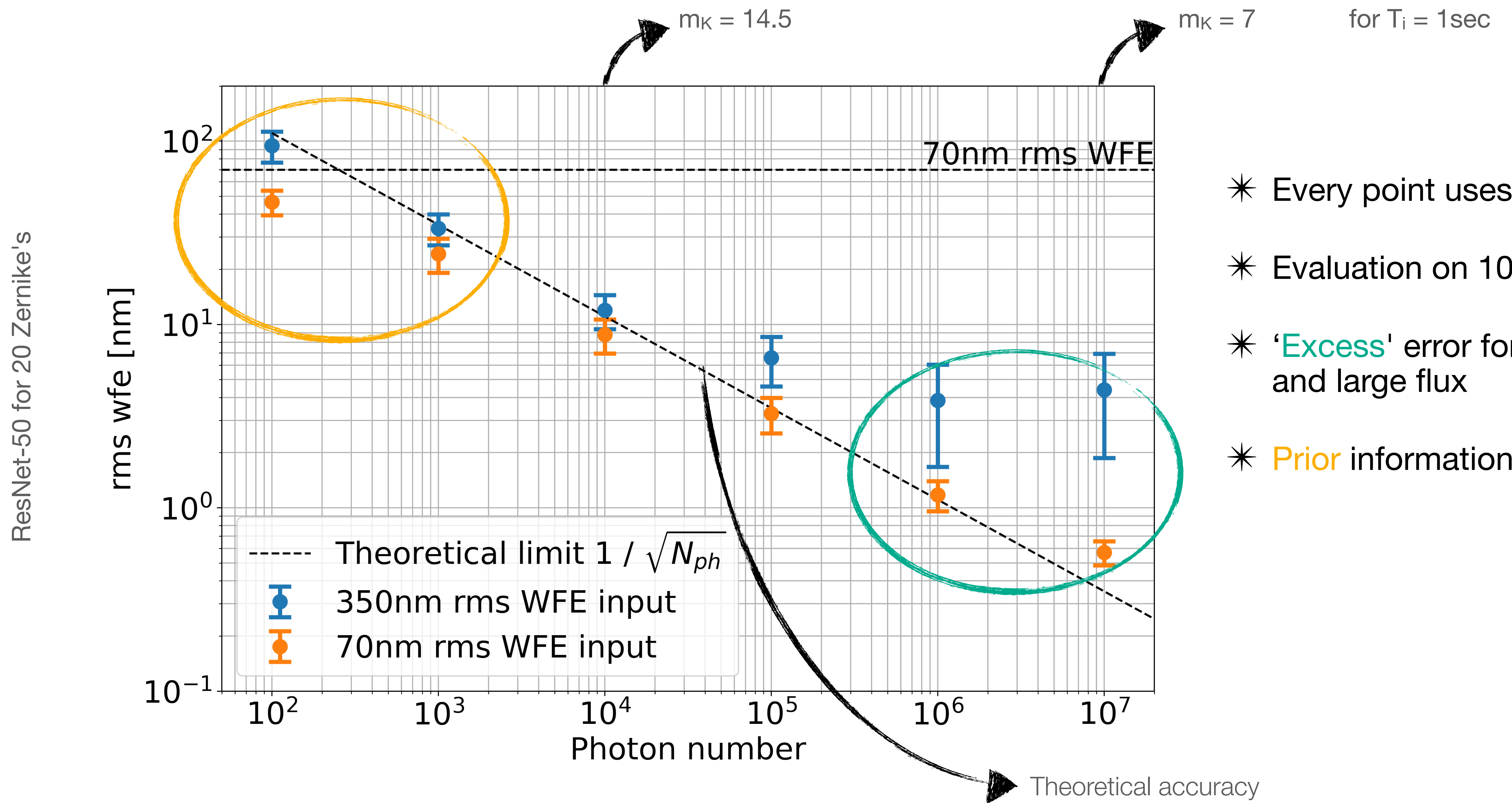
- Particular nature of light: photon noise
- Fisher information matrix [1] $\sigma_j^2 \geq 1/(4N_{ph})$ per independent mode *j*
- Most sensitive: Zernike wavefront sensor [2] $\sigma_j^2 \geq 1/(2N_{ph})$
- Focal plane sensitivity is further reduced

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

[1] Paterson 2008, 2013
[2] N'Diaye et al. 2013

Results

Fundamental limit



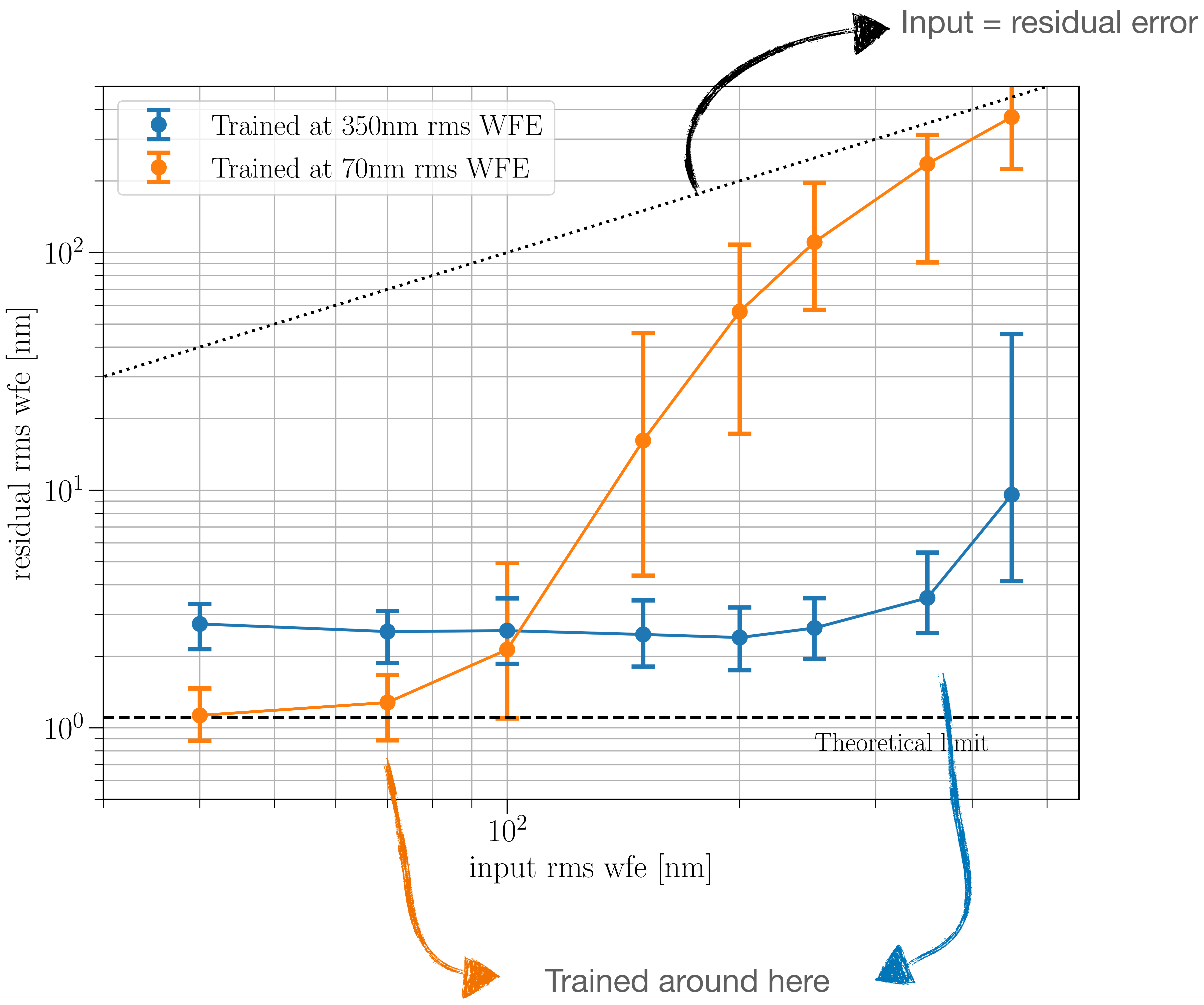
- Every point uses a different model
- Evaluation on 100 entries
- 'Excess' error for larger level of aberrations and large flux
- Prior information at low flux level

Results

Dynamical range

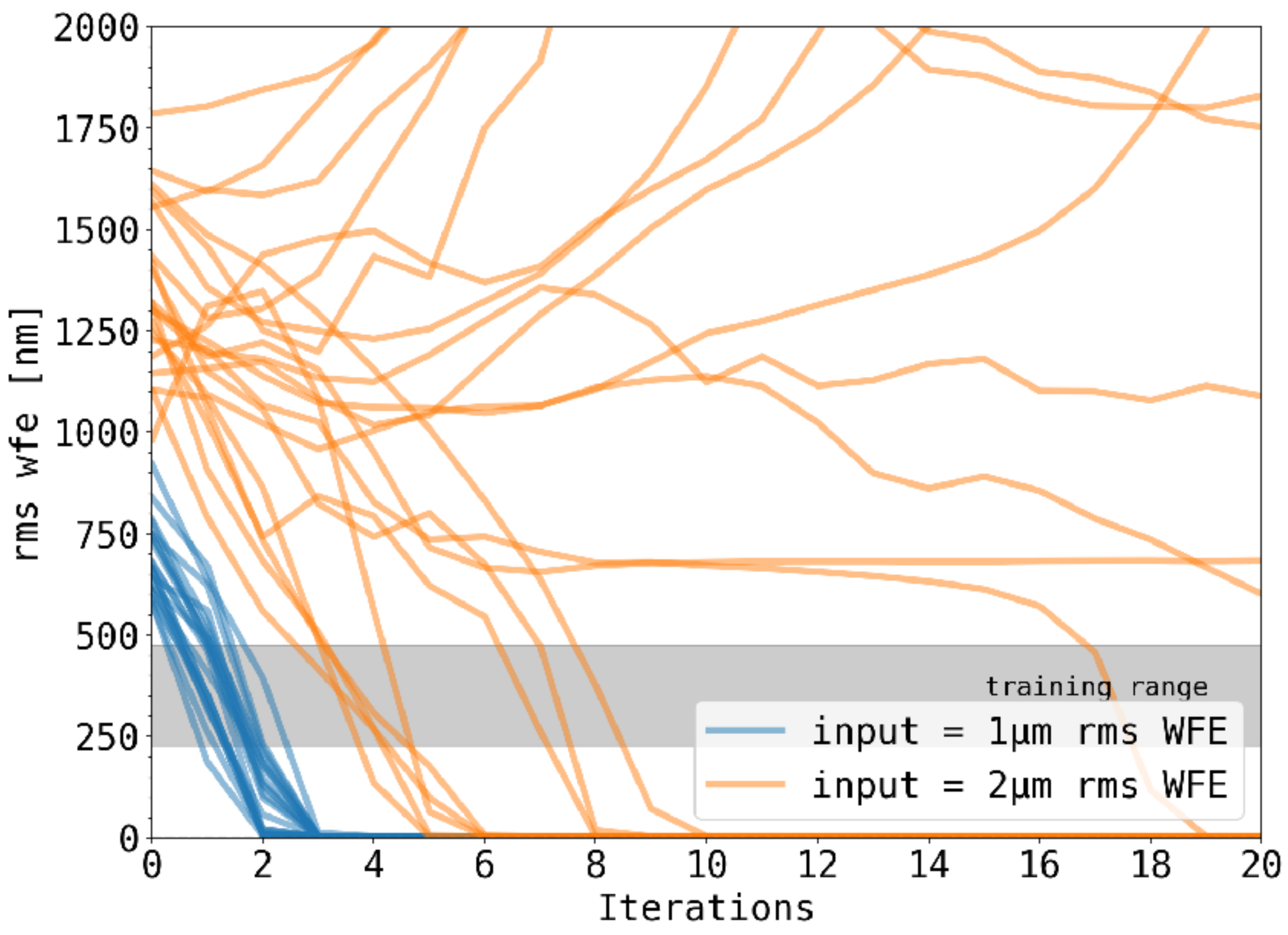
✓ Below training : constant accuracy

~ Above training : quickly increasing

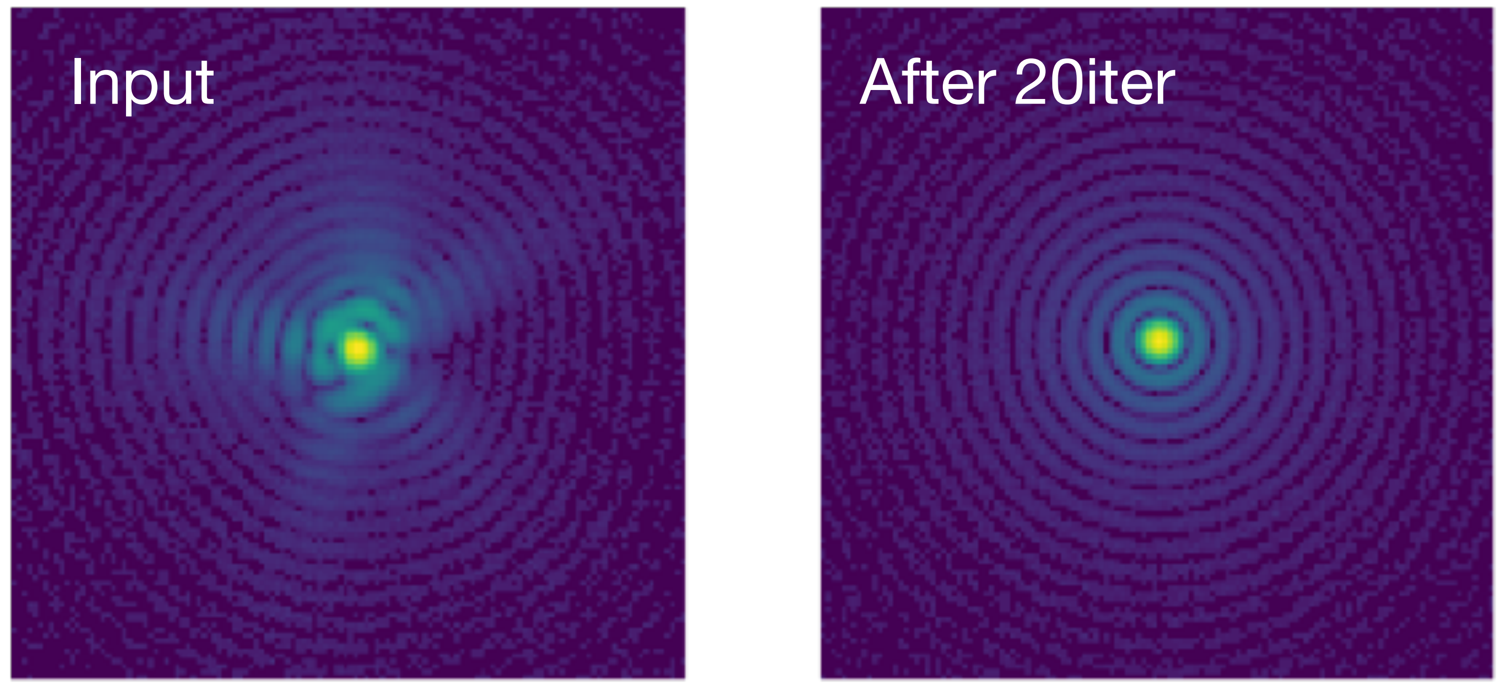


Results

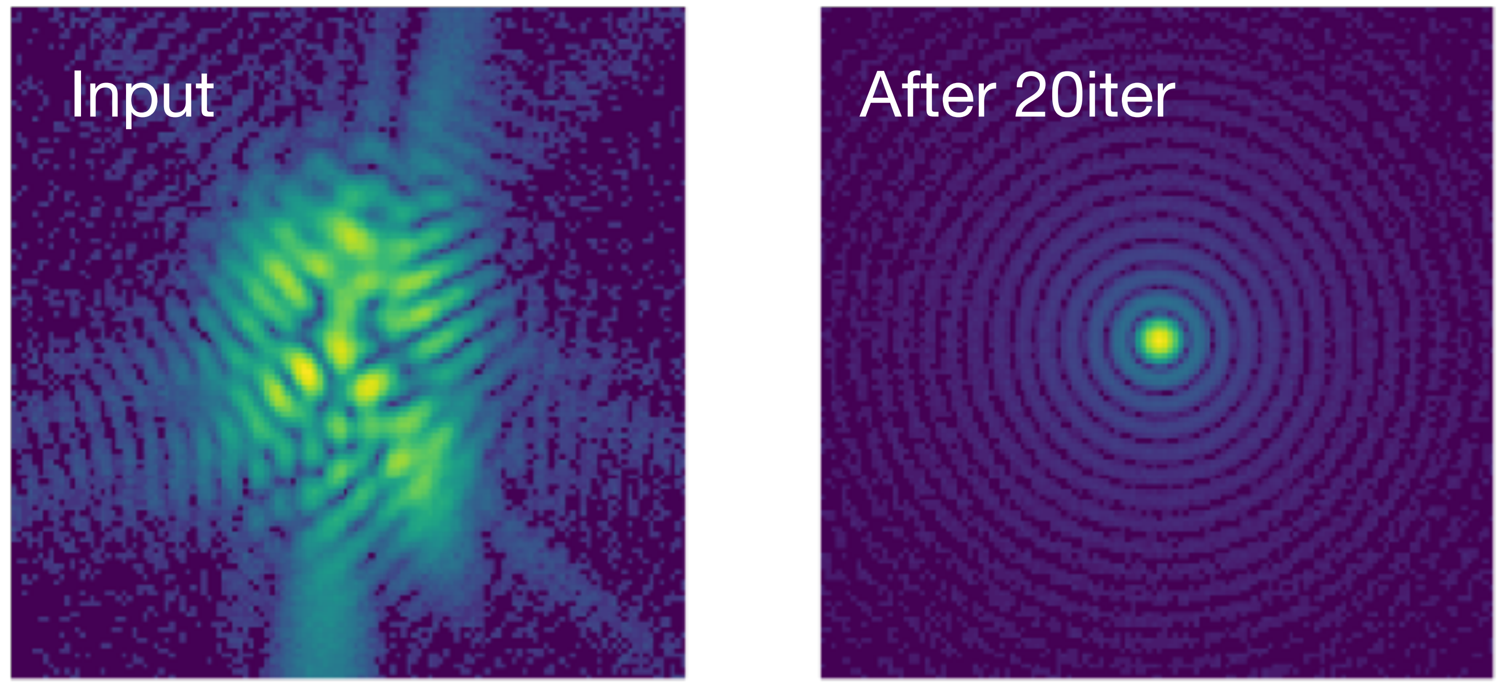
Dynamical range: application in closed-loop



~320nm rms WFE input



~1µm rms WFE input

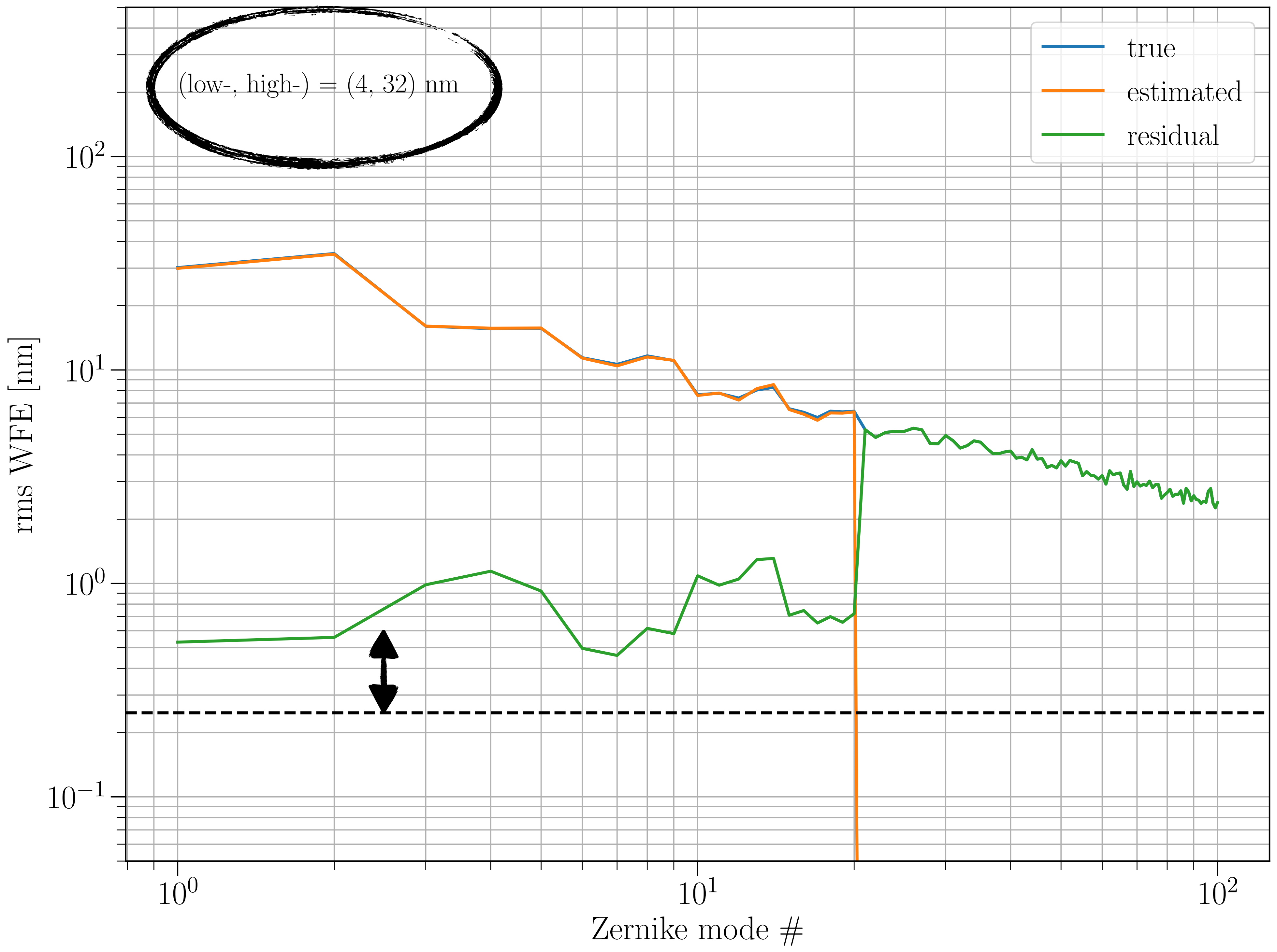


- Works well beyond training range

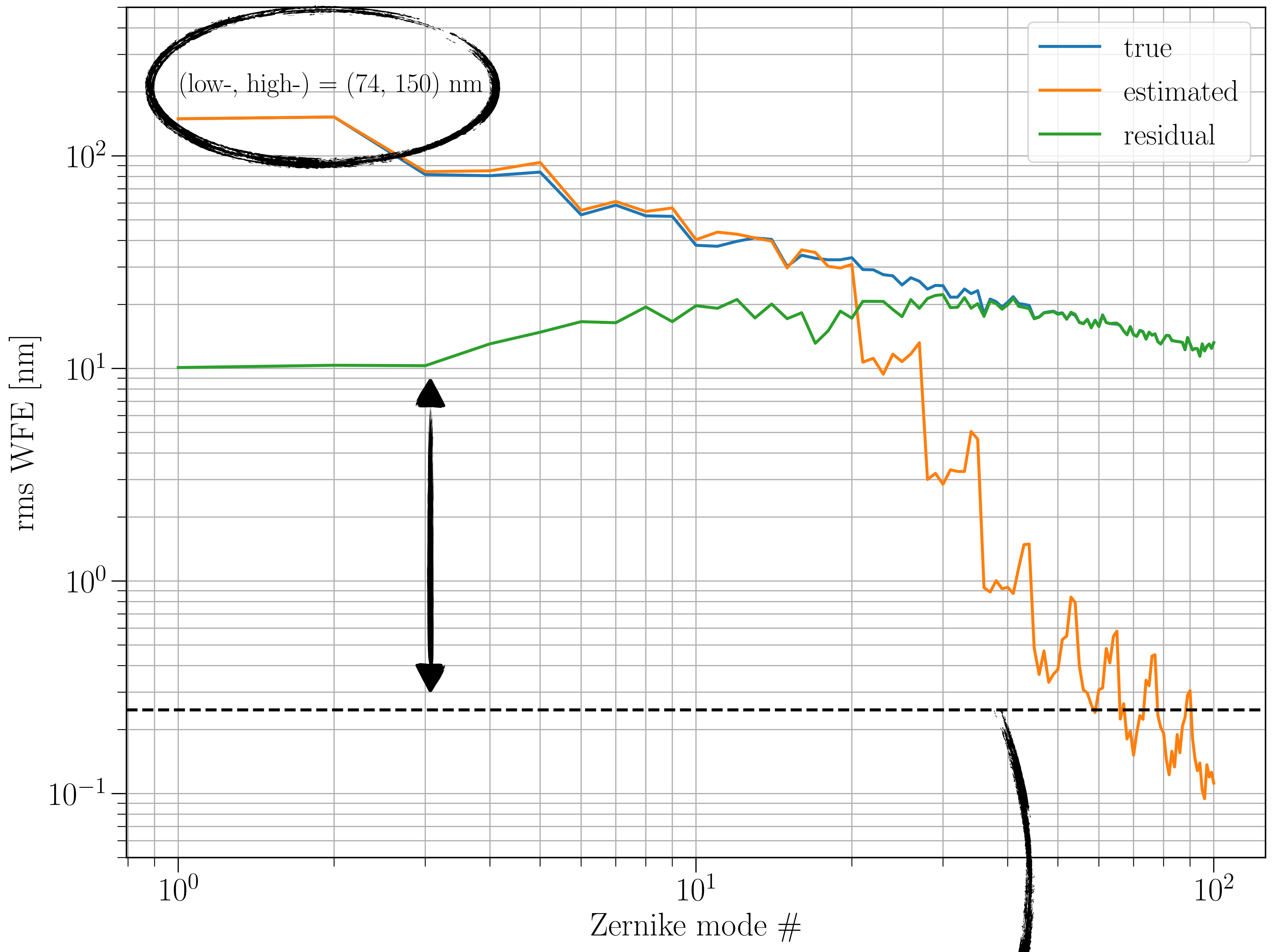
Results

Higher order disturbances

ResNet-50 ; 70nm rms WFE input



U-Net ; 350nm rms WFE input



✳Better drowning the fish with photon noise rather than revealing disturbances

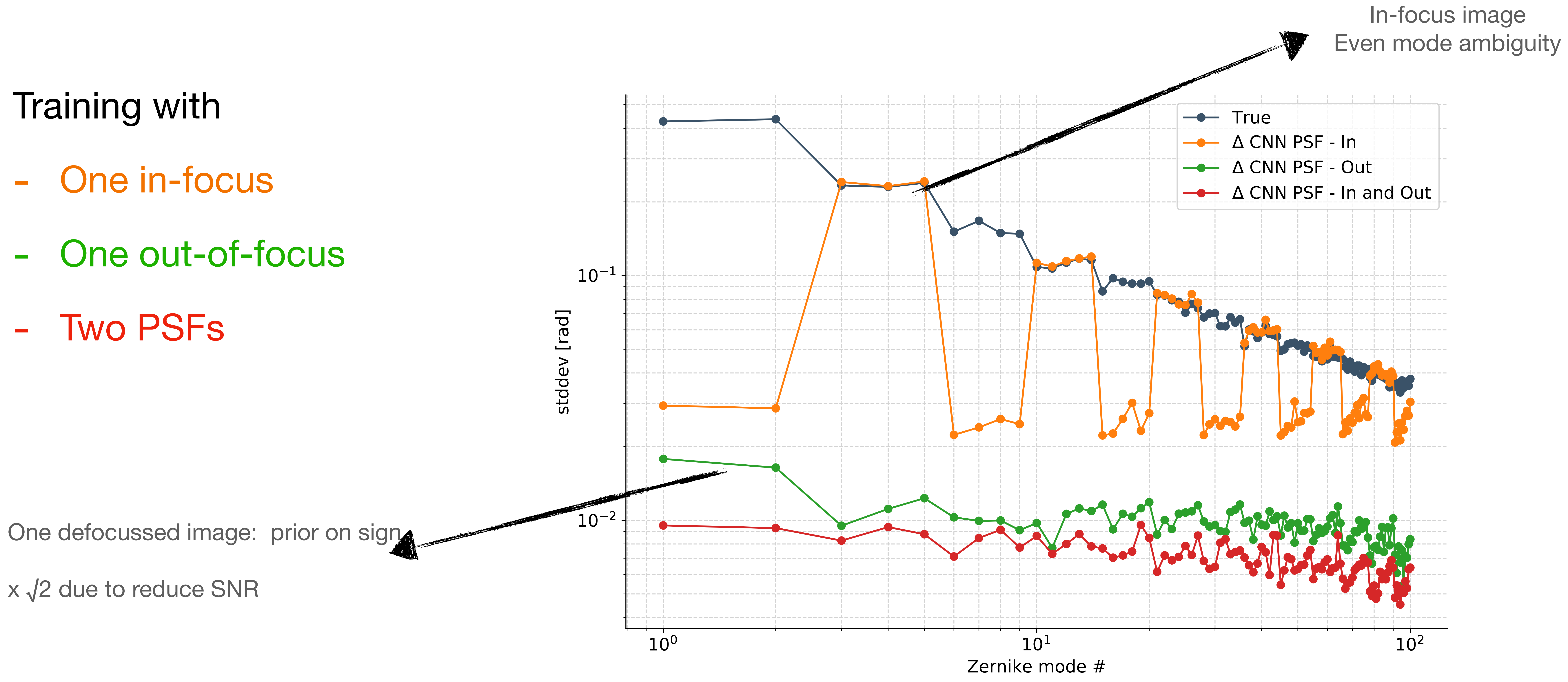
✳Adapt learning strategy

Theoretical accuracy

Results

Phase diversity and sign ambiguity

- Training with
  - One in-focus
  - One out-of-focus
  - Two PSFs



x √2 due to reduce SNR

Analysis and discussion

Computational cost

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

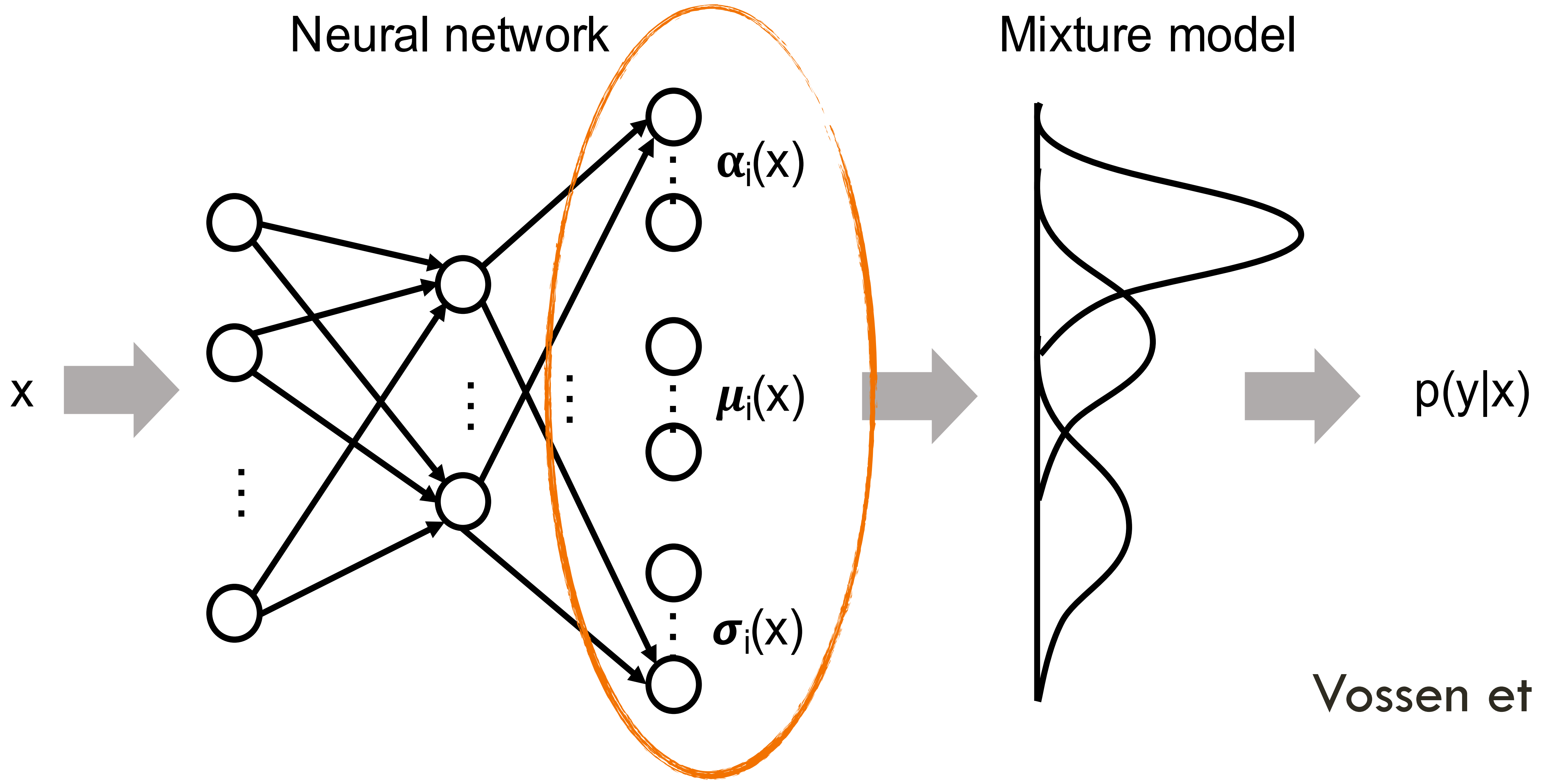
For 128x128 gridsizes and 100 Zernike’s

- ✦ AO application at 1kHz → 8-16 TFLOPs — RTX2080Ti provides >13TFLOPs
- ✦ Training time for a sample of 100,000 entries (2xGPU RTX2080Ti): <350sec / epochs, or <20hr for 200 epochs

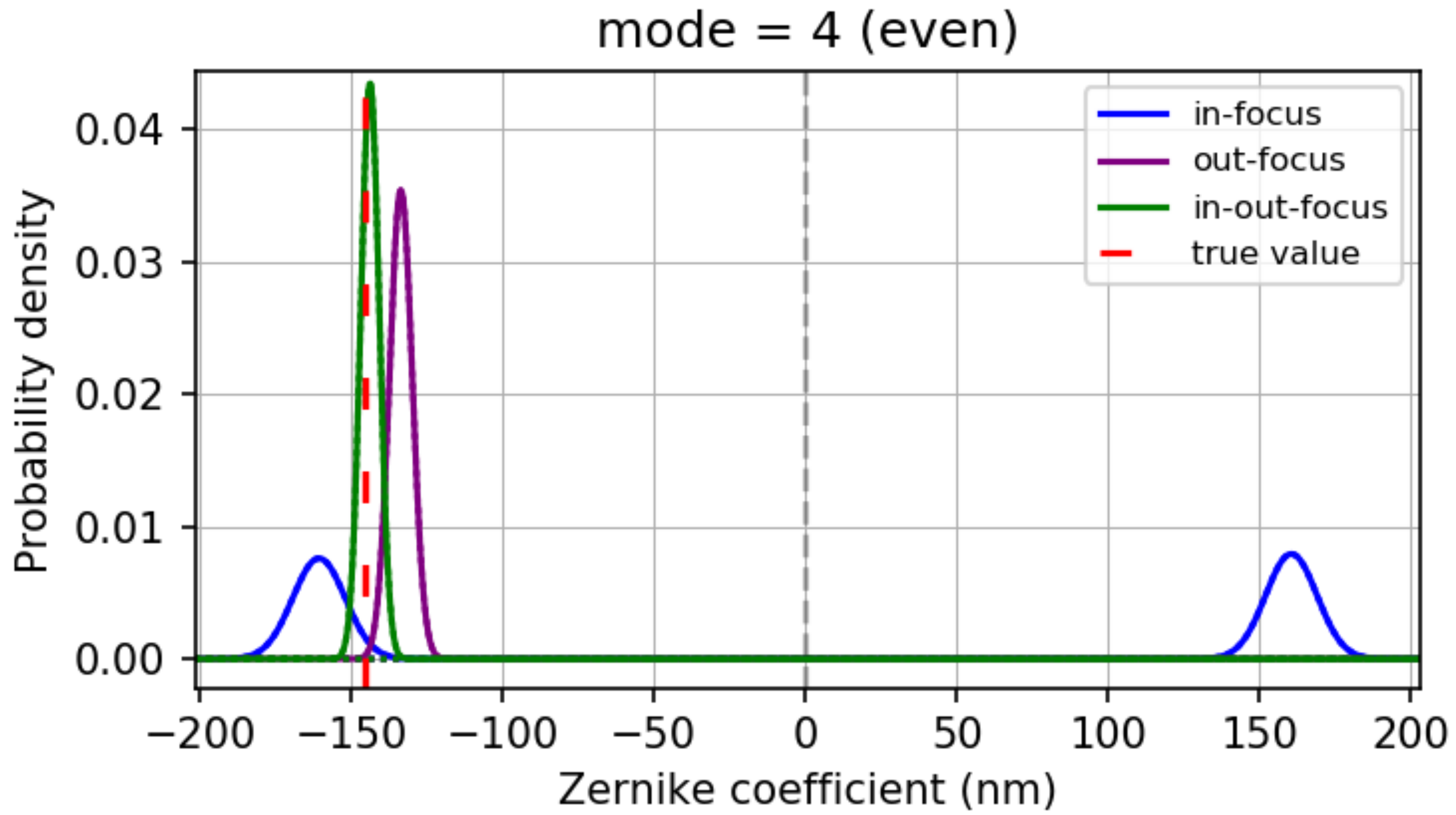
Analysis & discussion

Mixture density networks

- Adding a mixture density layer to ResNet
- Predict probability distributions
  - ➔ Degeneracy becomes explicit (i.e. sign ambiguity)
  - ➔ Information on error
- Con's : requires larger training dataset



Vossen et al. 2018



More analyses

- Pixel scale : mild sensitivity
- Varying SNR : relative robustness to be improved by adapting dataset
- Influence of training dataset size
- Application to vector vortex coronagraph
- Comparison to iterative algorithm (Gerchberg-Saxton type)

Conclusions

- CNN : optimum sensitivity, robustness, flexible
- Adapt your training strategy
- Lab & on-sky : simulated vs real data for training
- Ensure a source of diversity or prior