VolkswagenStiftung
Daimler und Benz Stiftung
STRUCTURES CLUSTER OF EXCELLENCE
Universität Heidelberg
Zukunft. Seit 1386.

Networks Learning the Universe:

From 3D (cosmological inference) to 1D (classification of spectra)

Caroline Heneka
Freigeist Fellow, Junior Research Group Leader

'Computer Vision Astrophysics' and observational cosmology group
@Institute for Theoretical Physics, Heidelberg University

EuCAIFcon2024, Amsterdam, April 30th 2024

Collaborators and students* (non-comprehensive):
Lara Alegre, Benedikt Schosser*, Tilman Plehn, Yannic Pietschke*, Tim Ullrich*,
Steffen Neutsch*, Marcus Brüggen, Fucheng Zhong*, Nicola Napolitano, Simon Barton*

VolkswagenStiftung

Daimler und Benz Stiftung

STRUCTURES CLUSTER OF EXCELLENCE

ITP Universität Heidelberg Zukunft. Seit 1386.

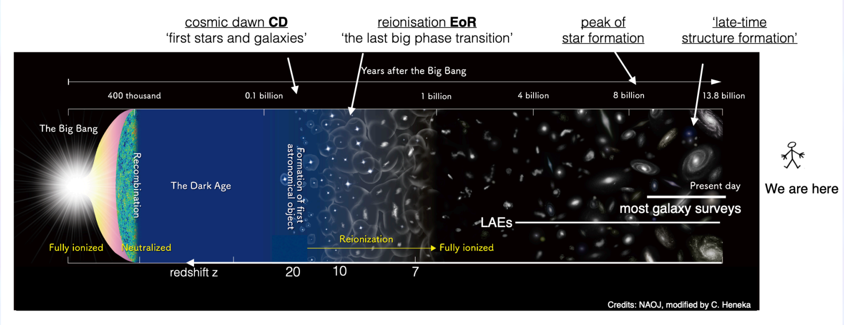
Our goal

Learn about cosmology, large-scale structure, and galaxy evolution using data from large astronomical surveys and developing the suitable modern ML & AI toolkit.

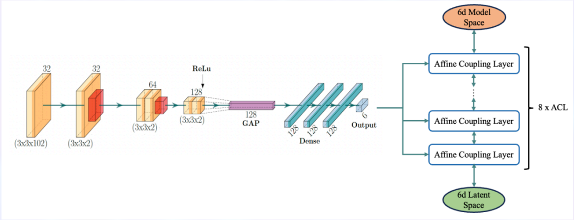
Our research:
- Computational astrophysics / cosmology
- Intensity mapping
- Large (radio) surveys, SKA & LOFAR

Specifically the modern ML toolkit for cosmology and large-scale surveys:
- Emulation, generation
- Inference

Also:
- Classification, anomaly detection
- **Computer Vision tasks in astronomy**



+

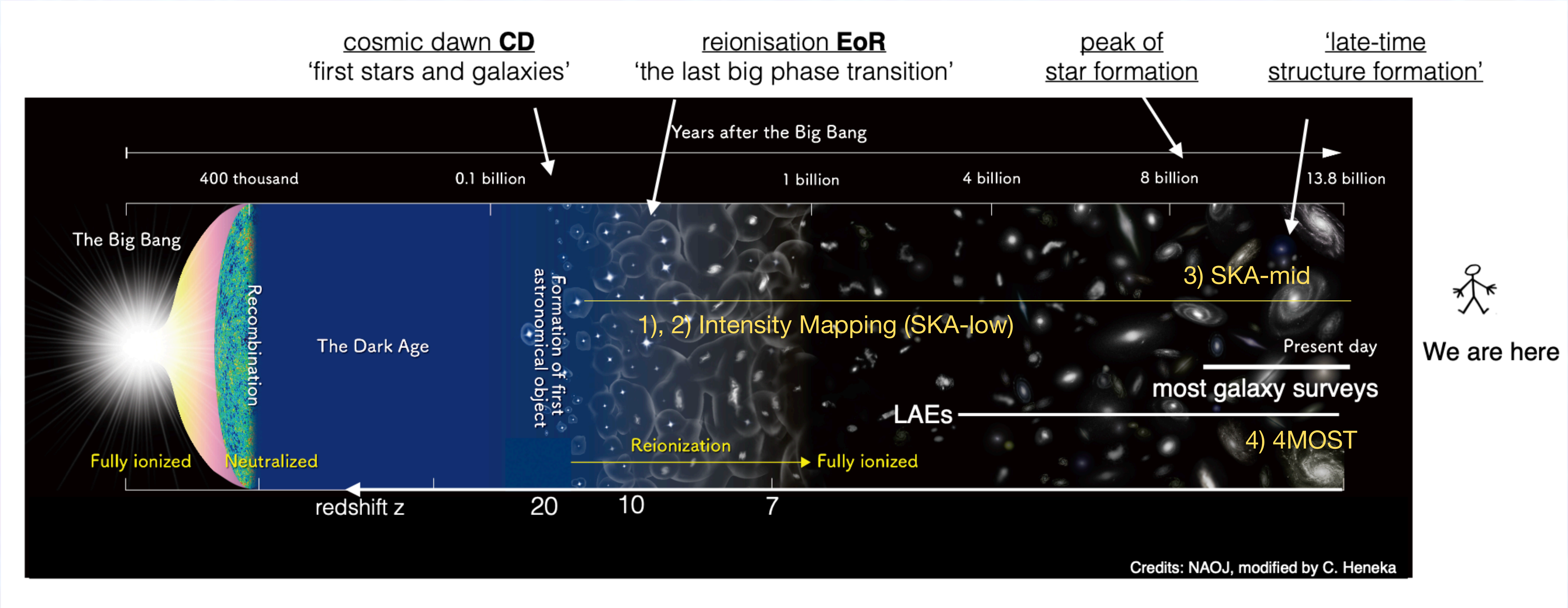


<u>Group 'Computer Vision Astrophysics' and observational cosmology:</u>

Lara Alegre (postdoc), Yannic Pietschke (PhD student), Vrund Patel (PhD student), Tom Schlenker (M.Sc. student), Abdulmalik Kara (M.Sc. student), Thomas Blankenburg (M.Sc. student), Marius Booz (B.Sc. student), Maitri Purohit (student assistant)

This Talk

1) Simulation-based inference (SBI) in 3D
2) Generative methods for simulation
3) Source detection & characterisation
4) Classification / Triggering



1) Simulation-based inference (SBI) for intensity mapping (3D)

21cm signal

a tracer of neutral hydrogen:

$$\delta T_{\mathrm{b}}(\nu) = \frac{T_{\mathrm{S}} - T_{\gamma}}{1+z}\left(1 - e^{-\tau_{\nu_0}}\right)$$

$$\propto \color{red}{x_{\mathrm{HI}}}\,(1 + \color{red}{\delta_{\mathrm{nl}}})\left(\frac{H}{\color{red}{\mathrm{d}v_{\mathrm{r}}/\mathrm{d}r} + H}\right)$$

Why care?

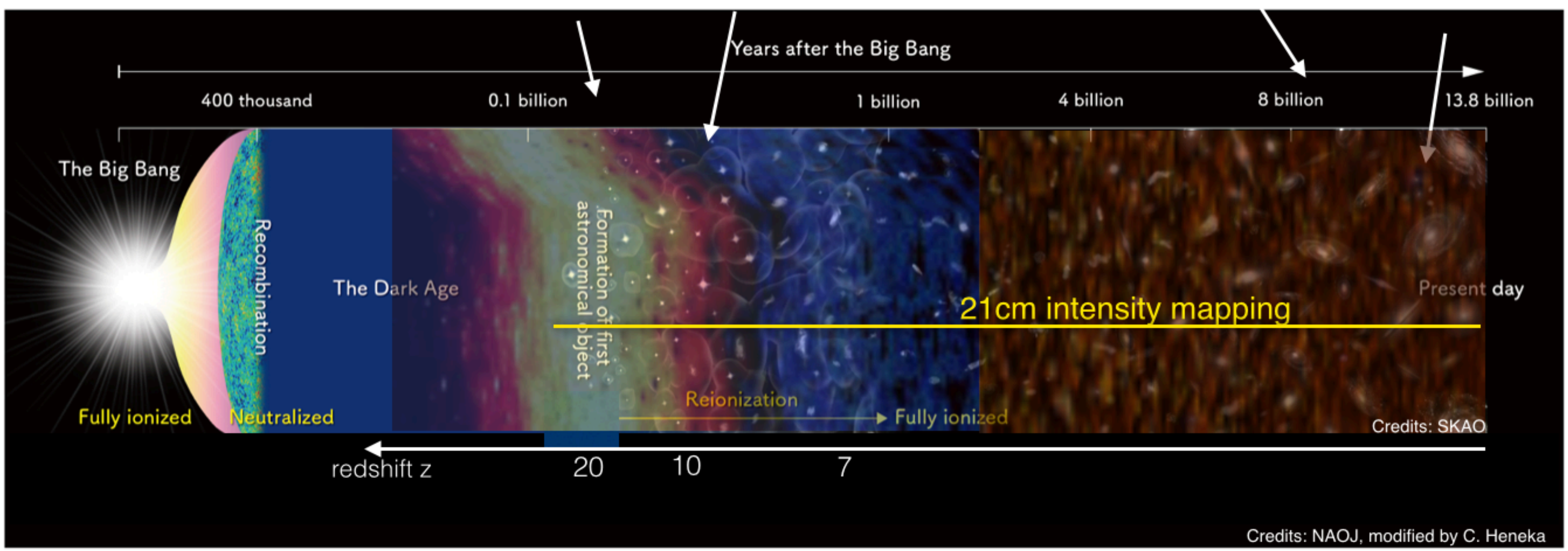
Tomography of >80% of the Universe

cosmic dawn **CD**
'first stars and galaxies'

reionisation **EoR**
'the last big phase transition'

peak of
star formation

'late-time
structure formation'



We are here

1) Simulation-based inference (SBI) for intensity mapping (3D)

Expected data SKA rate:
TB/s, few EB/day
Archive: ~700 PB/yr

Why care?

Tomography of >80% of the Universe

Square Kilometre Array - true ‘Big Data’

non-linear, non-Gaussian signal

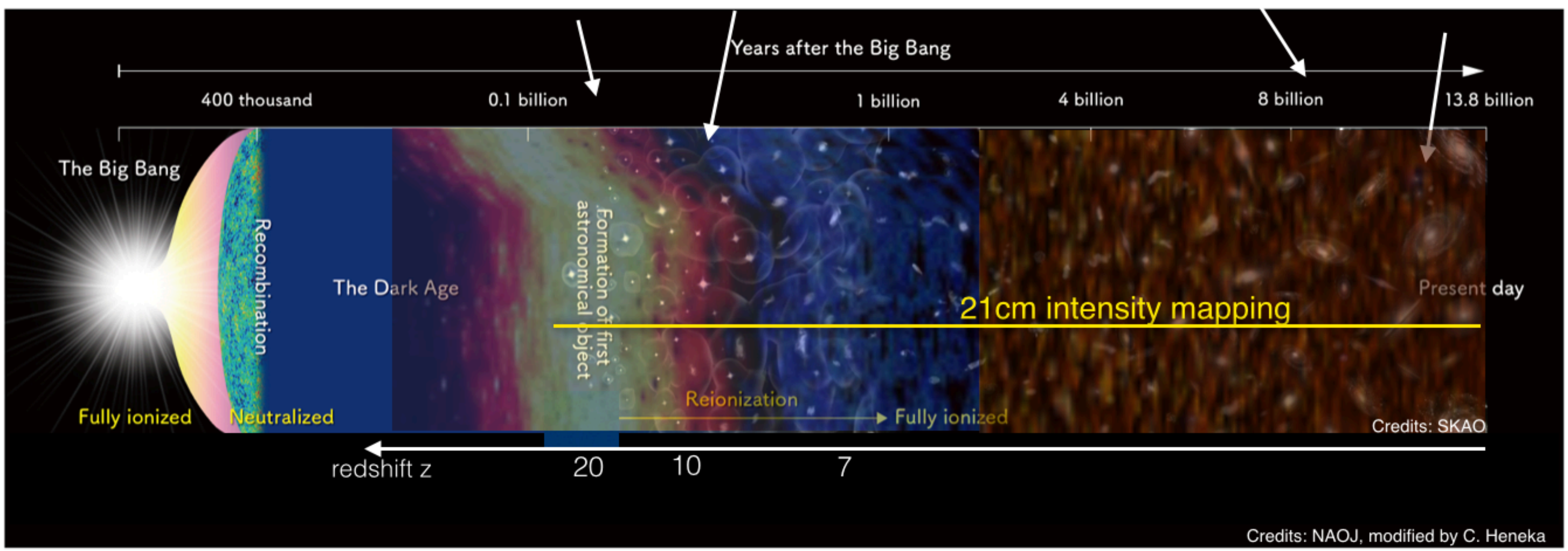
→ Move to full likelihood inference with networks

cosmic dawn **CD**
‘first stars and galaxies’

reionisation **EoR**
‘the last big phase transition’

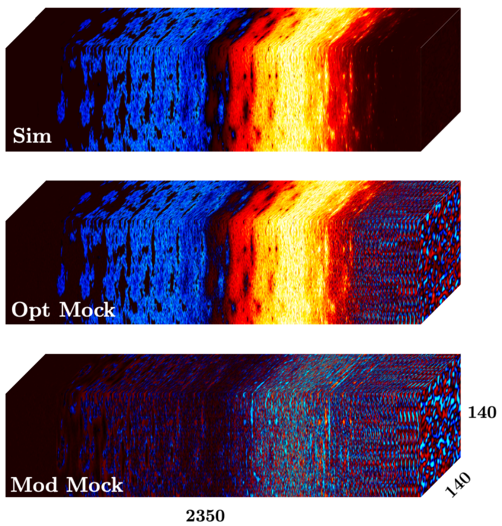
peak of
star formation

‘late-time
structure formation’

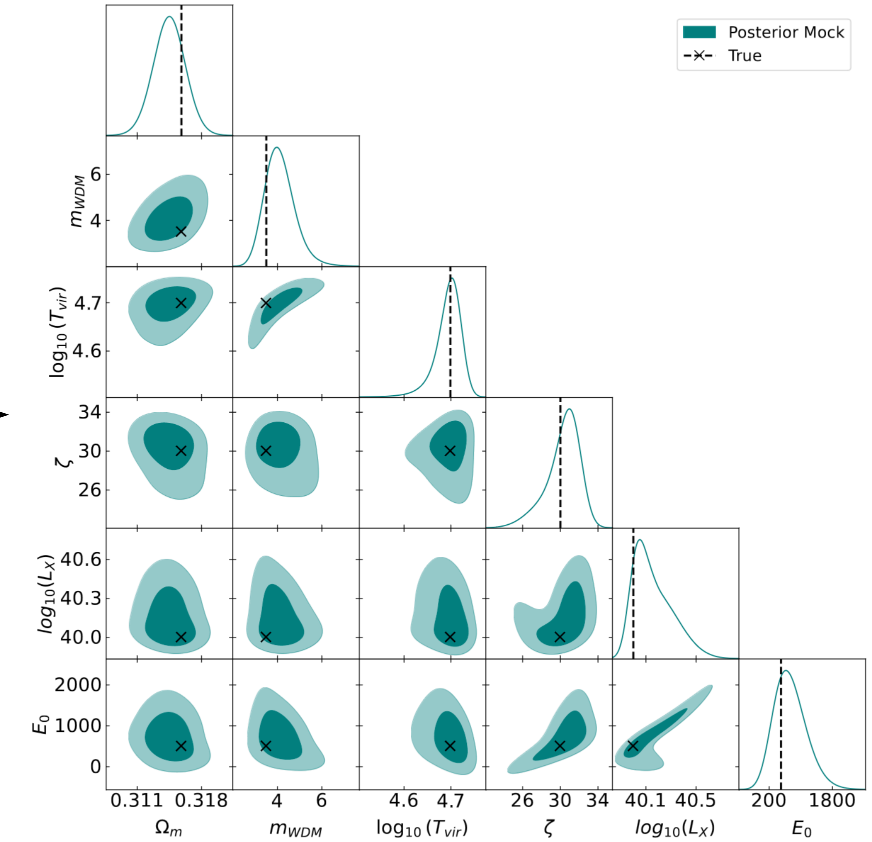


We are here

1) Simulation-based inference (SBI) for intensity mapping (3D)



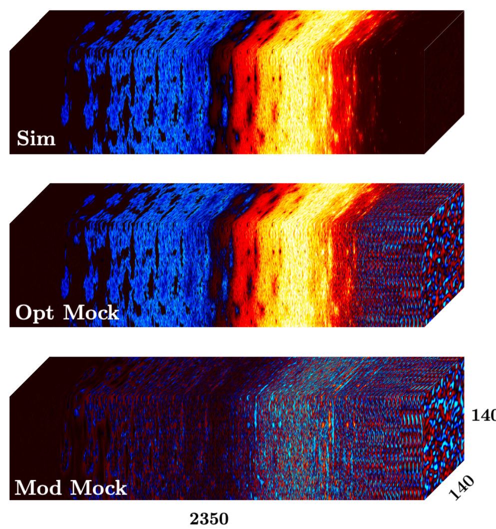
?



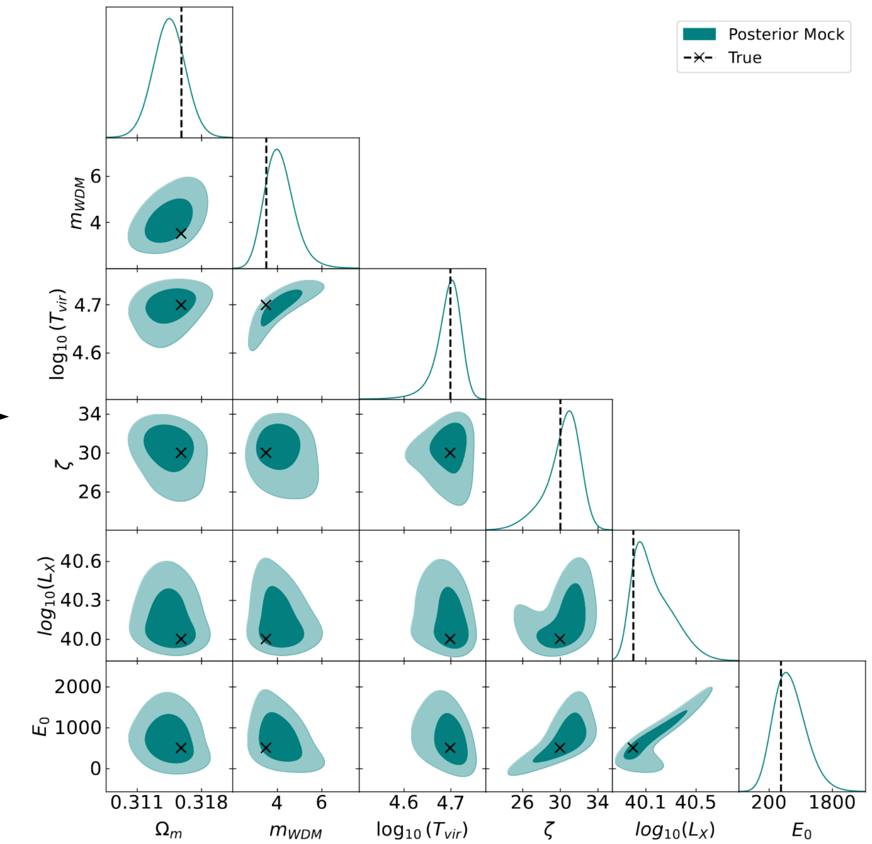
Neutsch, Heneka, Brüggen (2022), arXiv:2201.07587

Schosser, Heneka, Plehn, arXiv:2401.04174

1) Simulation-based inference (SBI) for intensity mapping (3D)



SBI with flows/cINN
based on BayesFlow
(arXiv:2003.06281)



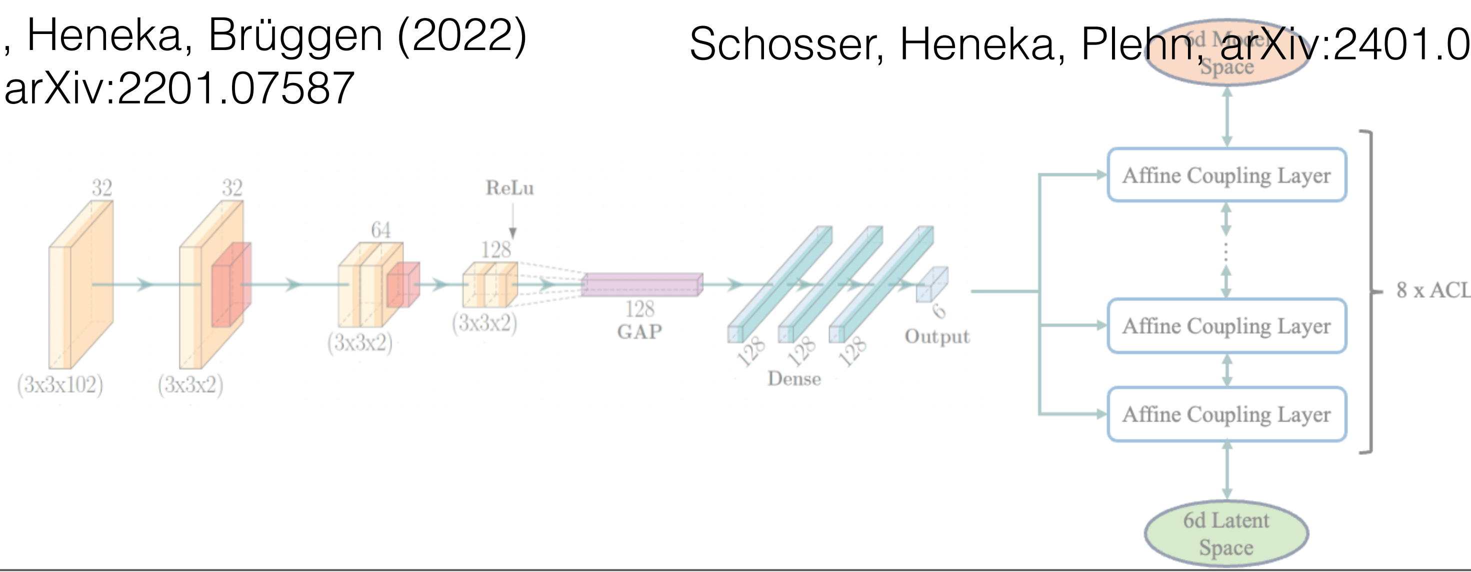
3D-21cmPIE-Net (public)

Neutsch, Heneka, Brüggen (2022)
arXiv:2201.07587

+

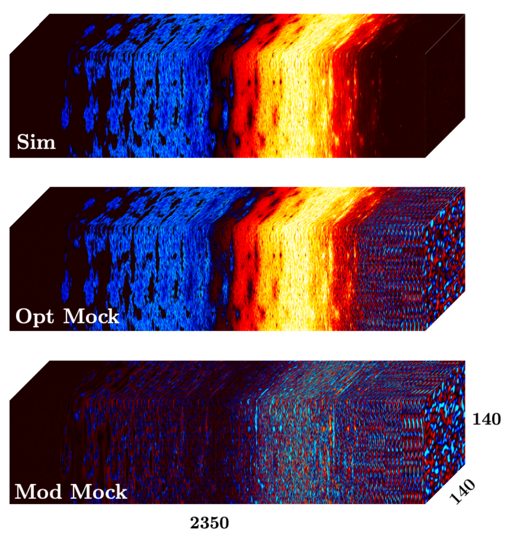
21cm-cINN (public when published)

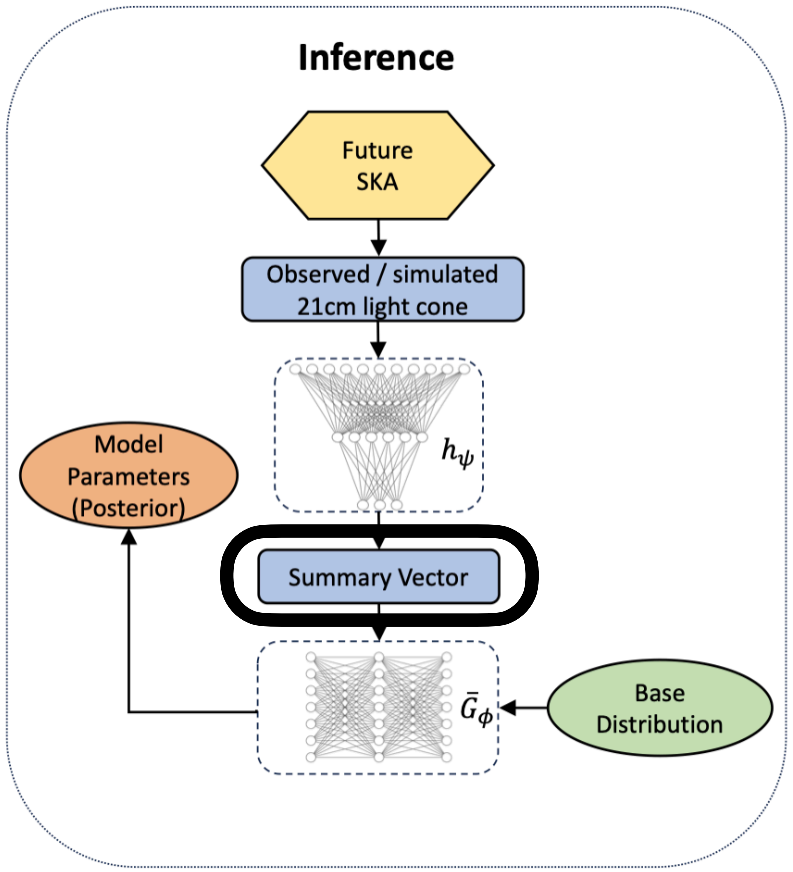
Schosser, Heneka, Plehn, arXiv:2401.04174

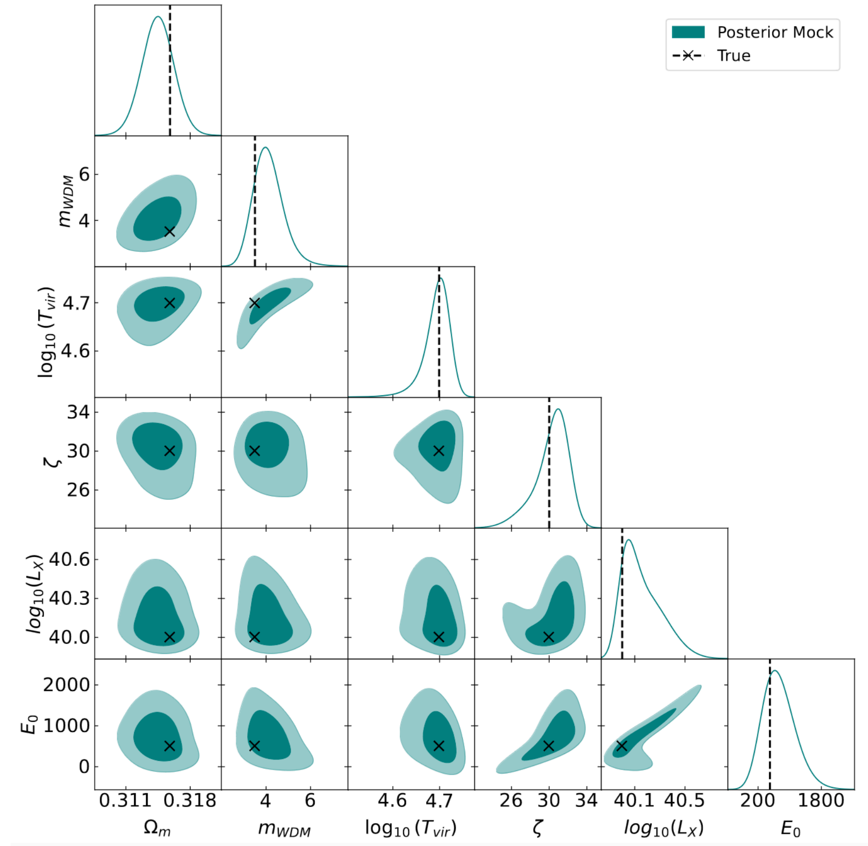


1) Simulation-based inference (SBI) for intensity mapping (3D)

more noise







BayesFlow, arXiv:2003.06281

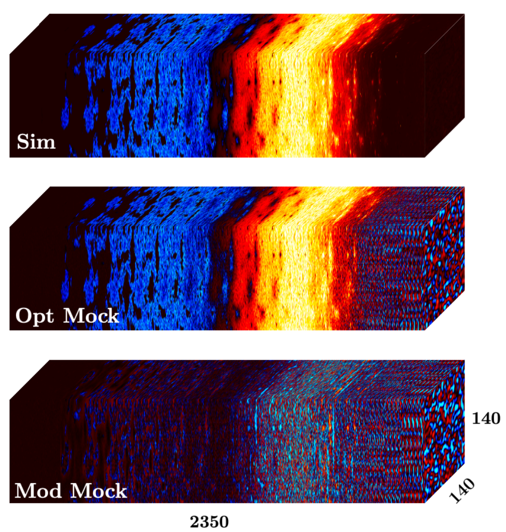
Neutsch, Heneka, Brüggen (2022), arXiv:2201.07587

Schosser, Heneka, Plehn, arXiv:2401.04174

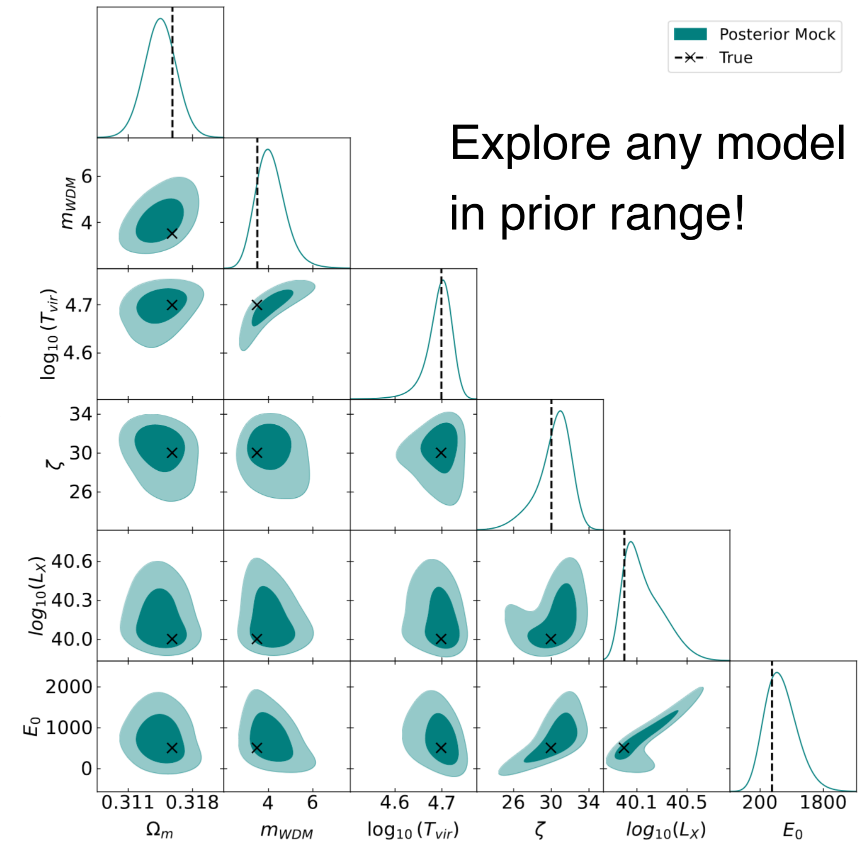
Sim: Summary stays close to original
Mock: Heavy adjustment of summary vector

We profit from learned summary in presence of noise (more).

1) Simulation-based inference (SBI) for intensity mapping (3D)



SBI with flows/cINN
based on BayesFlow
(arXiv:2003.06281)



Explore any model in prior range!

3D-21cmPIE-Net (public) + 21cm-cINN (public when published)

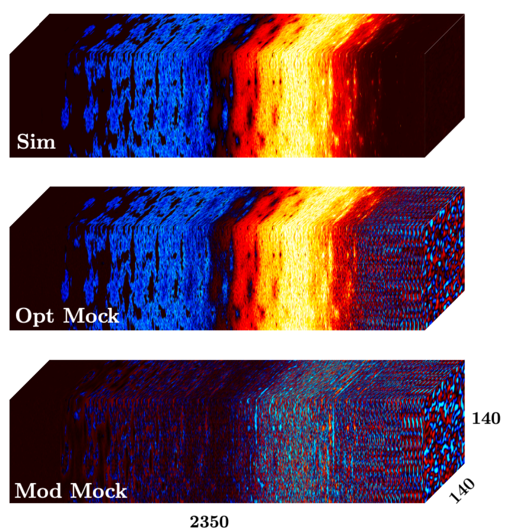
Neutsch, Heneka, Brüggen (2022)
arXiv:2201.07587

Schosser, Heneka, Plehn (2024), arXiv:2401.04174

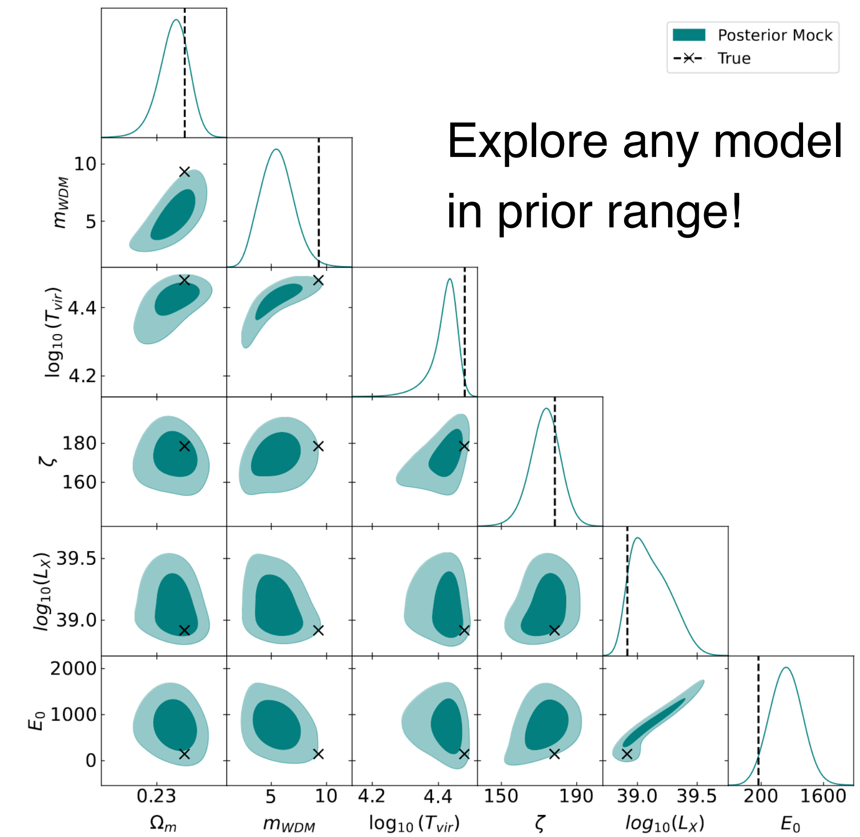
Learn more here: Poster location 108, Poster by Benedikt Schosser

'Optimal, fast, and robust inference of reionization-era cosmology with the 21cmPIE-INN'

1) Simulation-based inference (SBI) for intensity mapping (3D)



SBI with flows/cINN
based on BayesFlow
(arXiv:2003.06281)



Explore any model
in prior range!

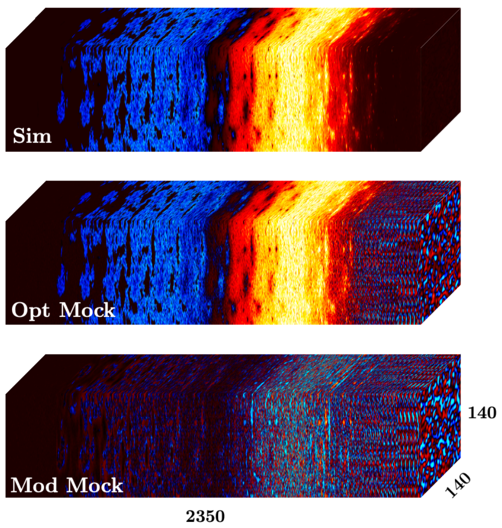
3D-21cmPIE-Net (public) + 21cm-cINN (public when published)

Neutsch, Heneka, Brüggen (2022)
arXiv:2201.07587

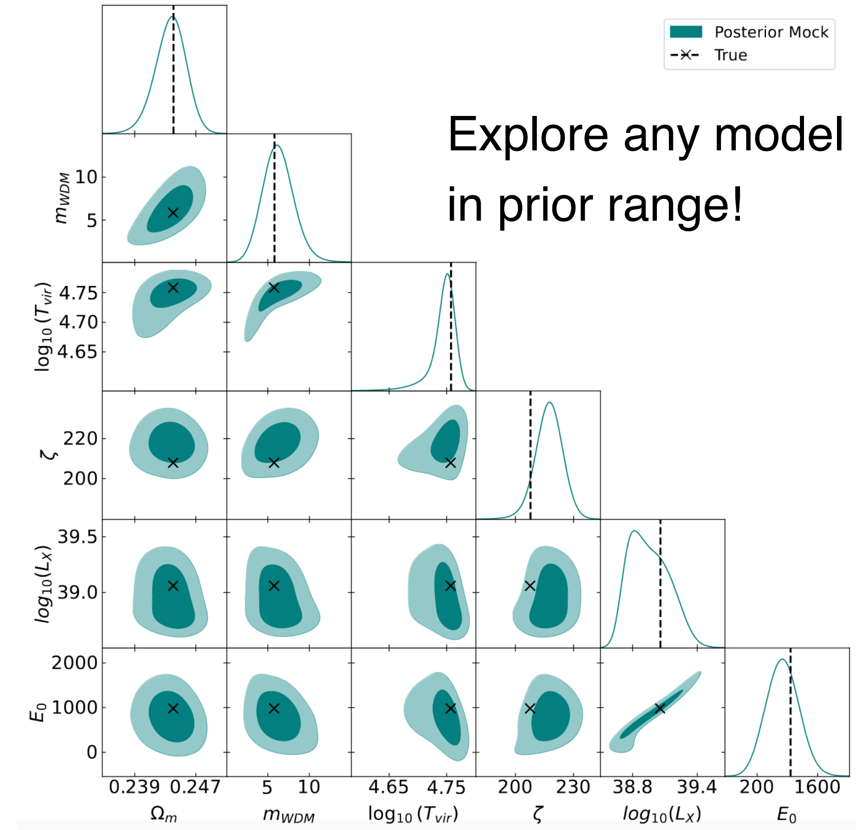
Schosser, Heneka, Plehn (2024), arXiv:2401.04174

Learn more here: Poster location 108, Poster by Benedikt Schosser
‘Optimal, fast, and robust inference of reionization-era cosmology with the 21cmPIE-INN'

1) Simulation-based inference (SBI) for intensity mapping (3D)



SBI with flows/cINN
based on BayesFlow
(arXiv:2003.06281)



Explore any model
in prior range!

3D-21cmPIE-Net (public) **+** 21cm-cINN (public when published)

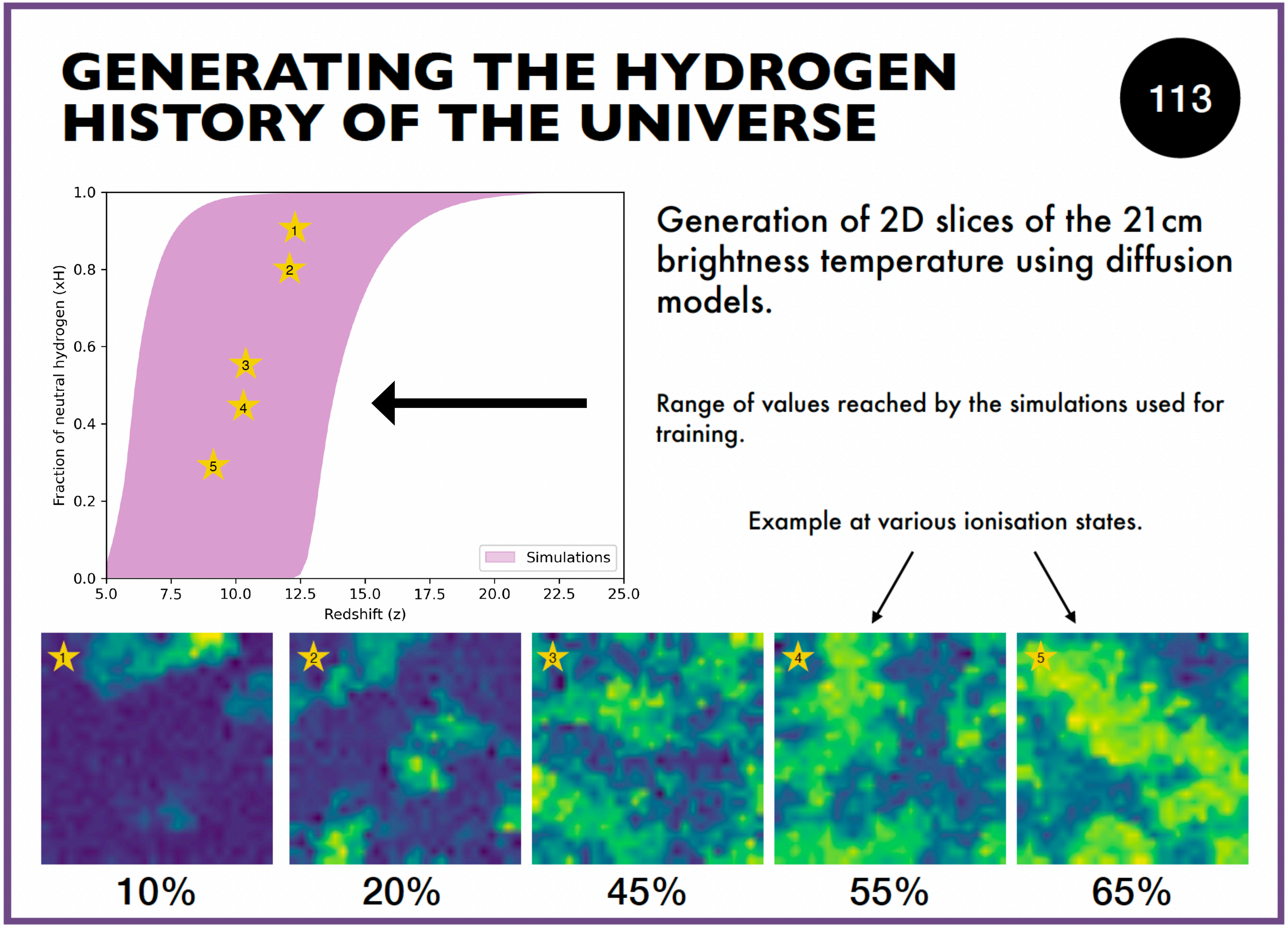
Neutsch, Heneka, Brüggen (2022)
arXiv:2201.07587

Schosser, Heneka, Plehn (2024), arXiv:2401.04174

Learn more here: Poster location 108, Poster by Benedikt Schosser

'Optimal, fast, and robust inference of reionization-era cosmology with the 21cmPIE-INN'

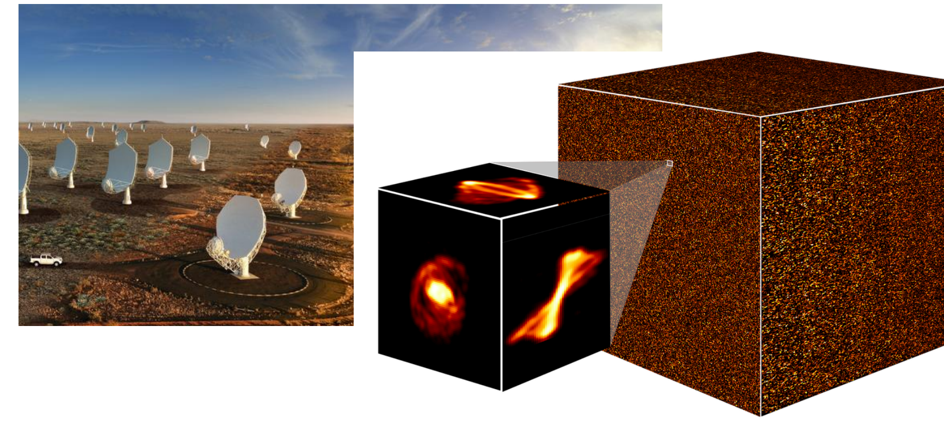
2) Generative methods for simulation (of the 21cm signal)

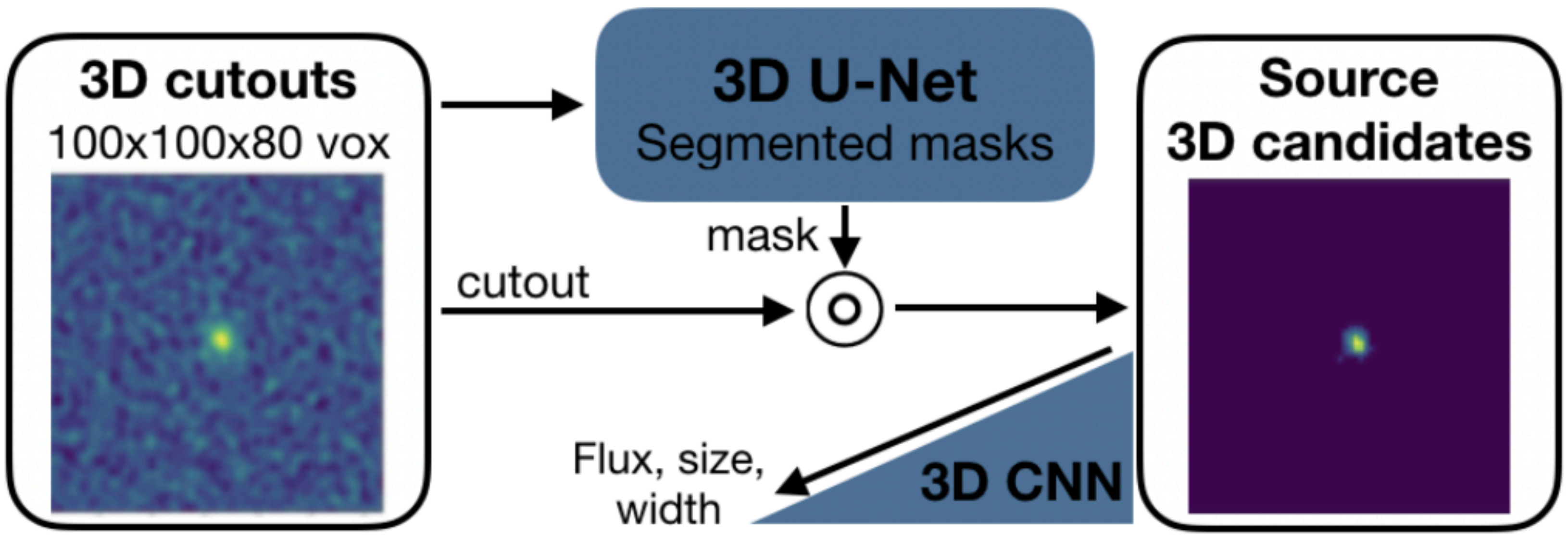


To learn more, see poster by: Lara Alegre, poster location 113

3) Radio source detection and characterisation

Example: Source detection in 3D tomographic data





<u>Main take-aways:</u>

- 3D better than stitching 2D (spatial) + 1D (frequency/time)
- High-fidelity 3D reconstructions, unbiased prior for characterisation
- Push to low signal-to-noise regime
- High S/N training data, ensemble decision

Hartley+ 23 (incl. Alegre, Heneka), arXiv:2303.07943
Heneka 22, arXiv:2311.17553

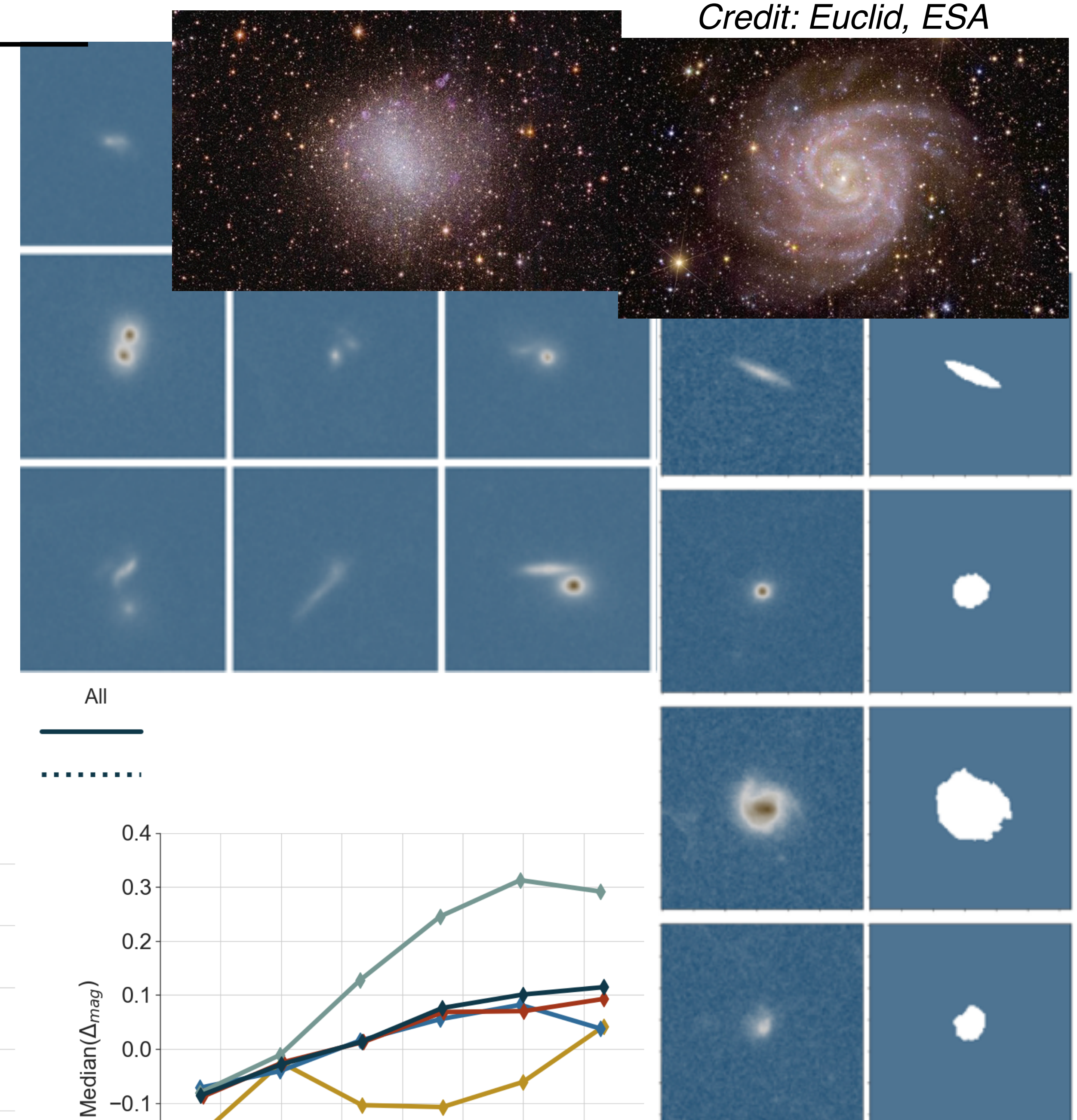
3) Optical source detection and characterisation

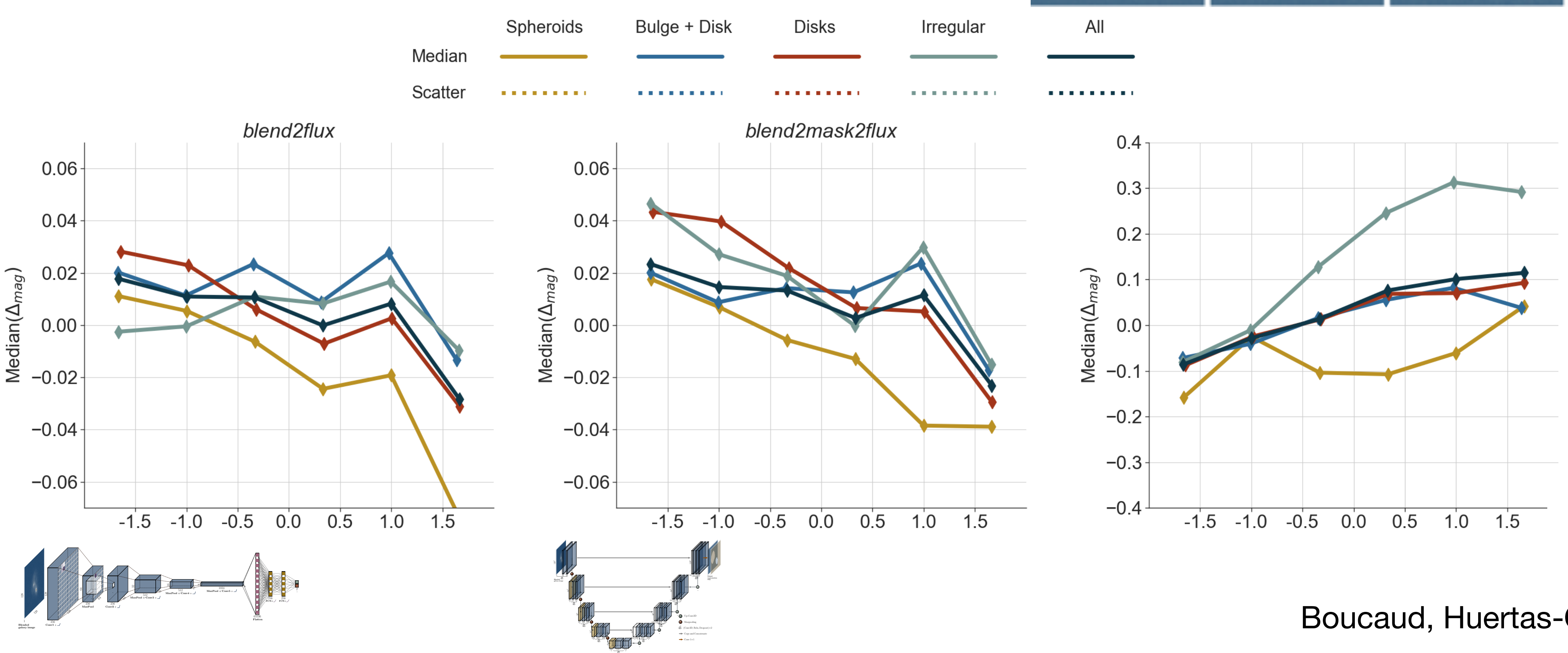
Example: Source detection in imaging data

Main take-aways:

- ~factor 5 more precise photometry
- Order of magnitude less bias!
- Good performance also for irregular type

Credit: Euclid, ESA





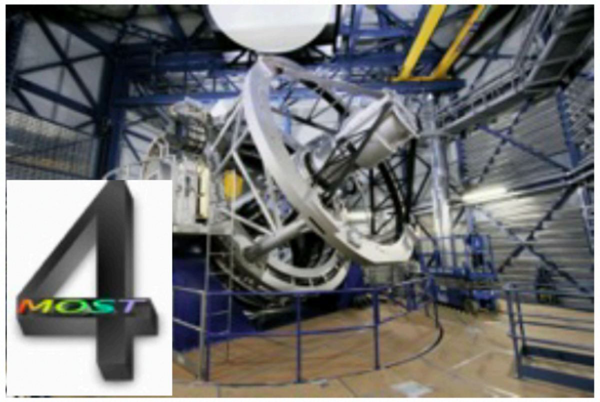
coindeblend

Boucaud, Huertas-Company, Heneka+ 20,
arXiv:1905.01324

4) Classification and triggering for large astronomical surveys

4MOST: On-the-fly classification of spectra (1D)

- 5-year survey
- wide-field, fibre-fed, optical spectroscopy
- on ESO's 4-m-class telescope VISTA
- 2.5-degree diameter field-of-view, 2436 fibres
- HRS R ≈ 18000 – 21000, LRS R ≈ 4000 – 7500
- 20mio. (LRS), 3mio. (HRS) sources



https://www.4most.eu Credit: ESO

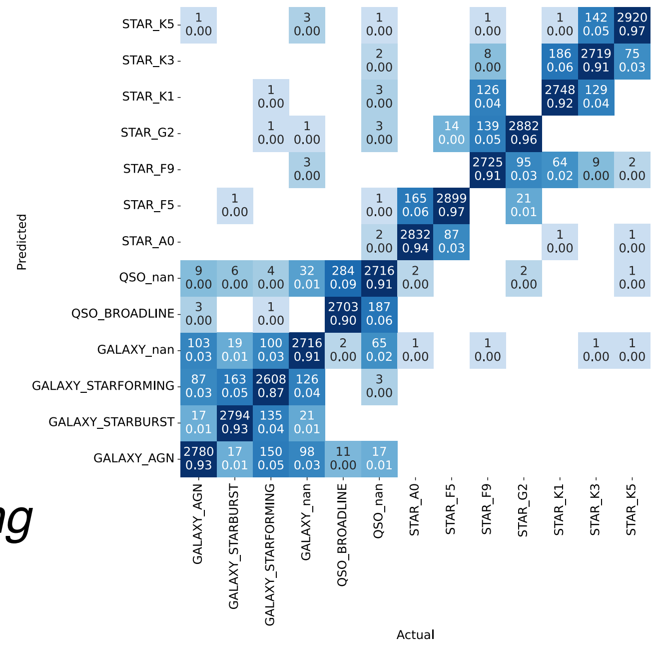
Goal: Data-driven classification pipeline layer (galactic & extragalactic sources)

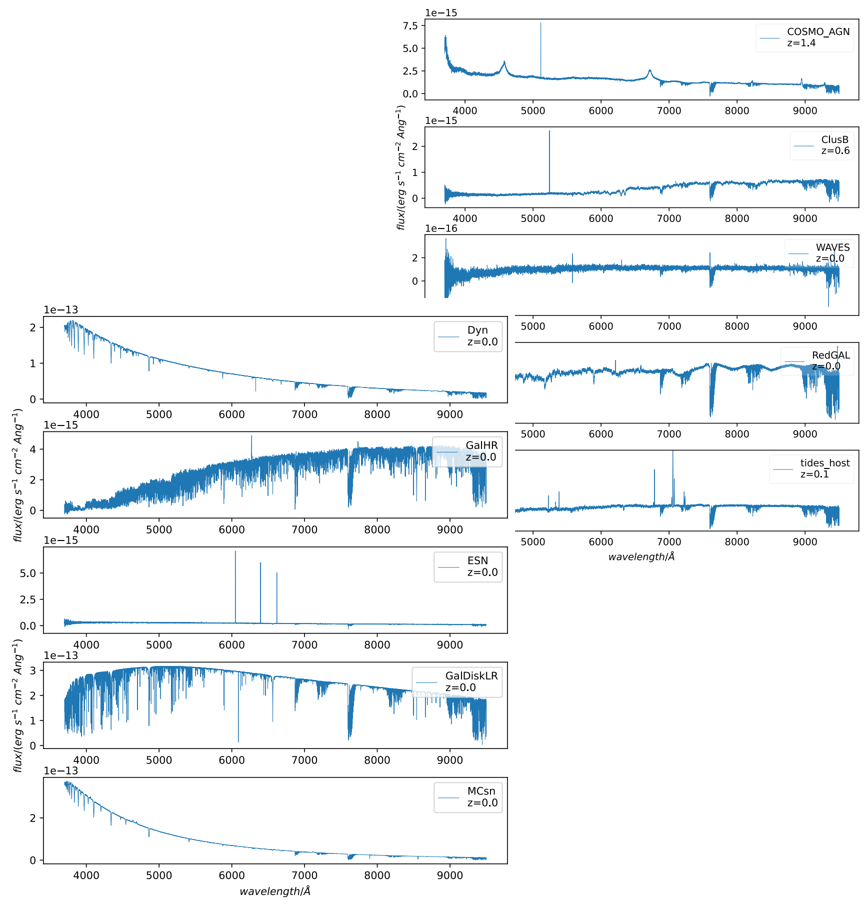
Classification infrastructure working group, led by: N. Napolitano & C. Heneka

→ Probabilistic multi-classifier

Benchmark with SDSS archival spectra: Convolutional network variants, BNN + contrastive learning for class uncertainties

- competitive with template fitting





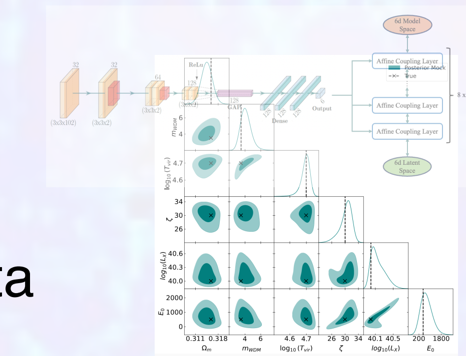
Zhong, Napolitano, Heneka+ arXiv:2311.04146

Summary and conclusions

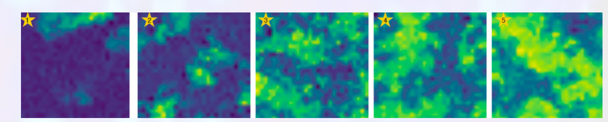
Use data from large astronomical surveys to learn about astrophysics + cosmology.

Important tasks are

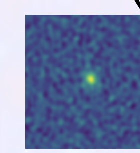
- Inference: SBI in 3D from non-Gaussian tomographic data

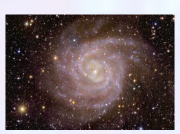


- Simulation: Produce large range of reionization topologies

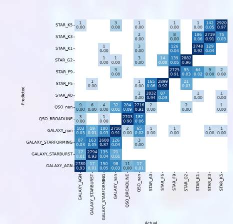


- Detection & Characterisation: Unbiased measurements from diverse sources (galaxies)





- Classification: Online classifier and triggering



ML/DL/AI has come to stay when dealing with astronomical survey data.

Neutsch, Heneka, Brüggen 22, arXiv:2201.07587
Schosser, Heneka, Plehn, arXiv:2401.04174
Hartley+ 23 (incl. Alegre, Heneka), arXiv:2303.07943
Heneka 22, arXiv:2311.17553
Boucaud, Huertas-Company, Heneka+ 20, arXiv:1905.01324
Zhong, Napolitano, Heneka+, arXiv:2311.04146

Thank you for your attention!