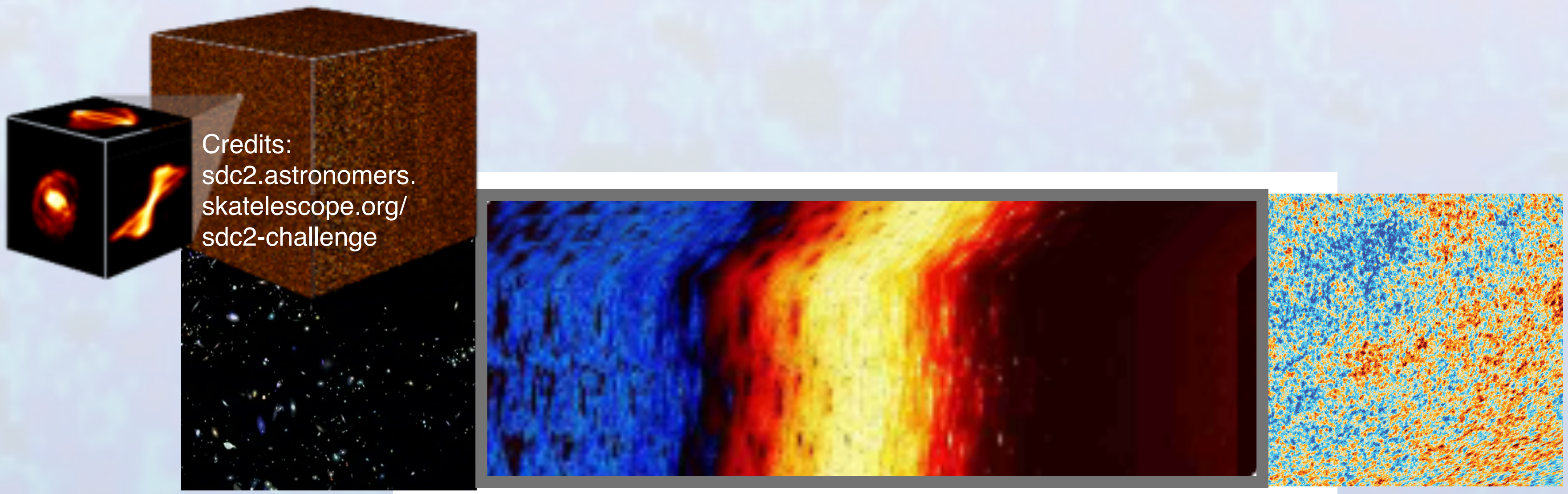
Learning from 3D tomographic 21cm intensity data



Credits: ESO

Credits: Planck, ESA

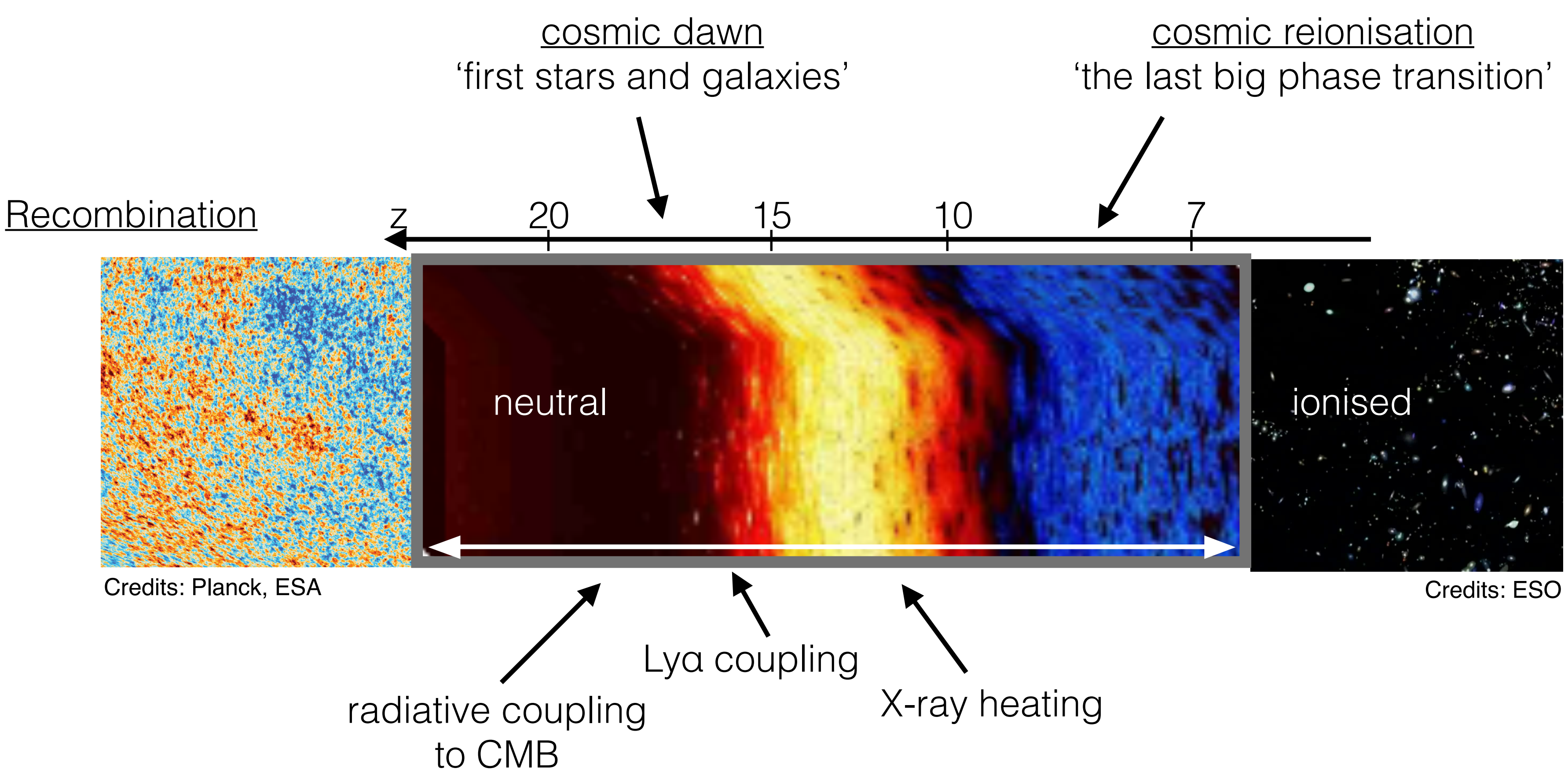
Caroline Heneka, Hamburg University, EXC Quantum Universe & Observatory

AG2021 - online, Splinter eScience, September 17th 2021

Collaborators: Steffen Neutsch (UHH), Marcus Brüggen (UHH), Michele delle Veneri (U. Naples), Bernardo Fraga (CBPF Brazil), Andrew Soroka (CMC MSU), Fedor Gubanov (CMC MSU), Clecio de Bom (BBPF Brazil), Alex Meshcheryakov (CMC MSU)

3D lightcones to track the history of the Universe

What is the cosmology at high redshifts? ‘gap’ between CMB and galaxy surveys
What properties do the very first stars and galaxies have?



21-cm lightcone*
generated with 21cmFAST

*lightcone = 2D spatial +
1D frequency information

blue = emission
red / yellow = absorption
against the CMB

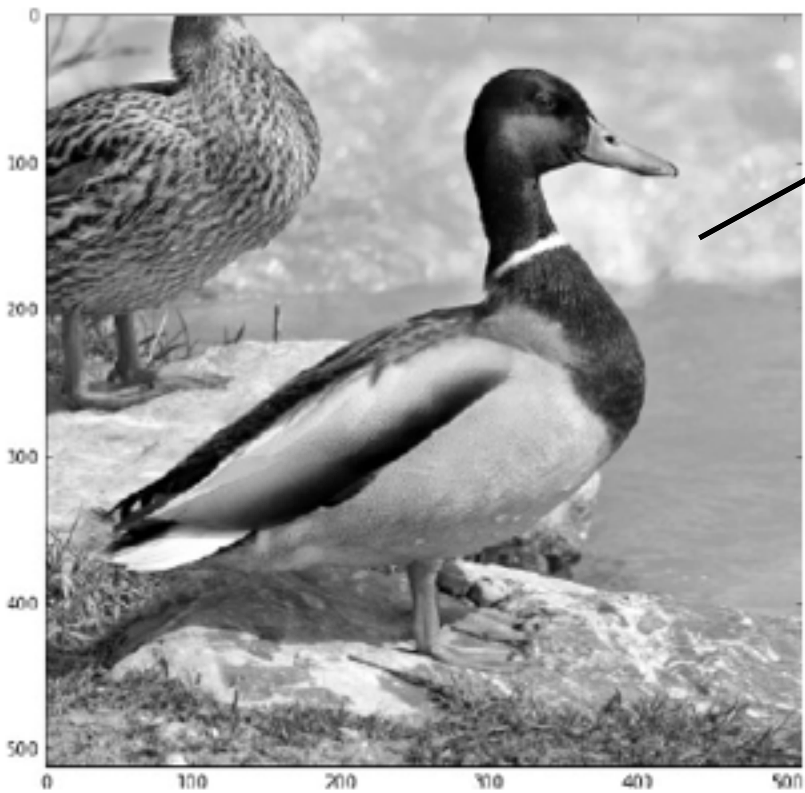
Galaxy surveys: Identify individual sources

IM: Measure emission fluctuations over ‘large’ areas

Example: Planck satellite for the CMB

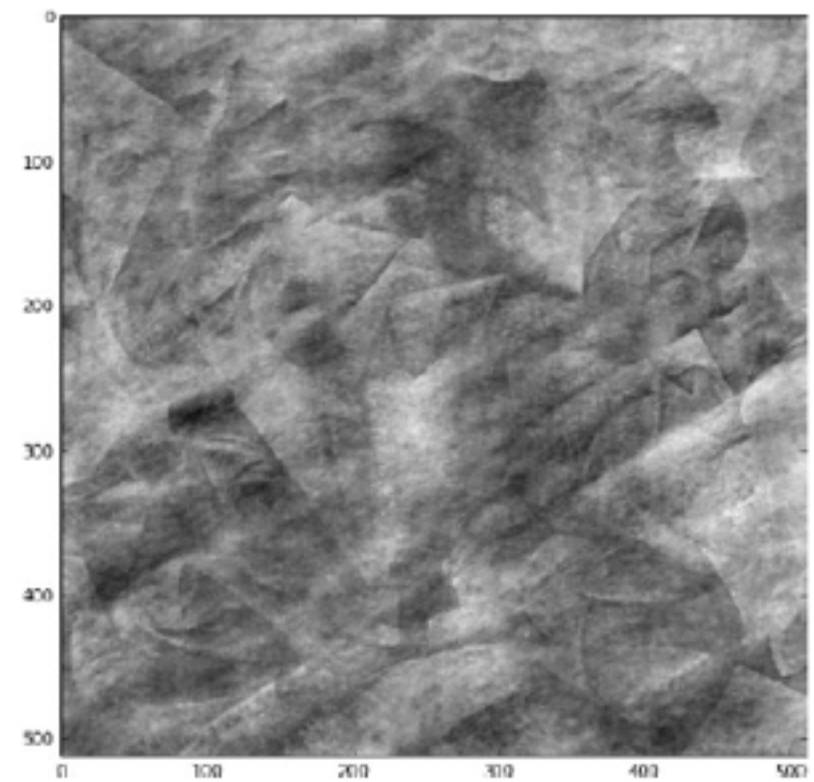
Why (deep) learning?

The duck example of (Non-)Gaussianity



parameters (size, color, background)

randomise phases



Credit: G. Bernardi

The duck: highly non-Gaussian

Same 2D power spectrum

The Gaussian duck

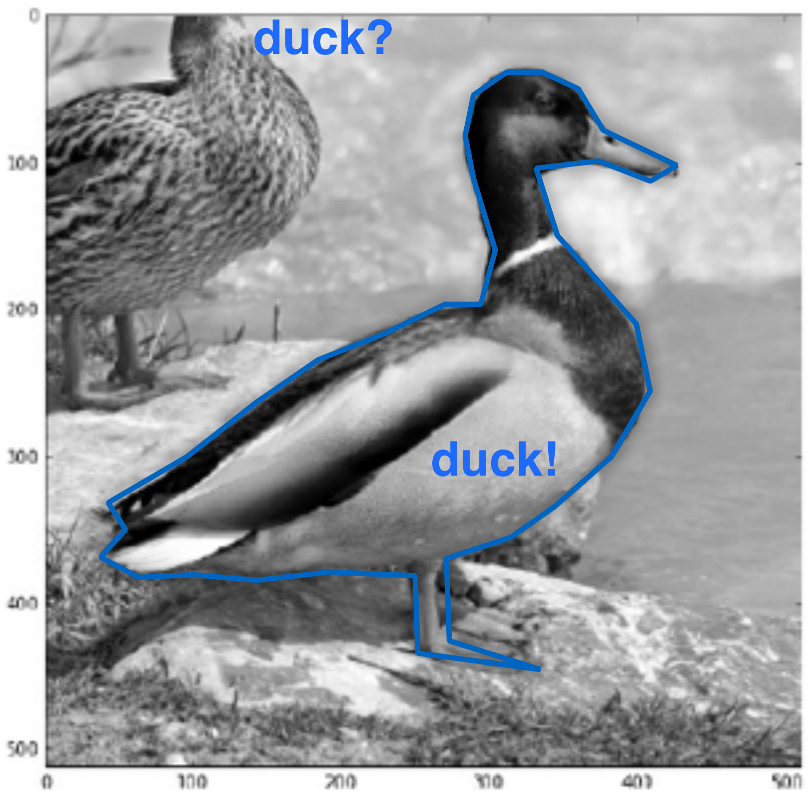
- Picks up non-Gaussian information
- Representation learning

Applications:

1. **Inference** (what duck? what properties? what shapes?)
2. Detect the duck (or galaxy, or signature)

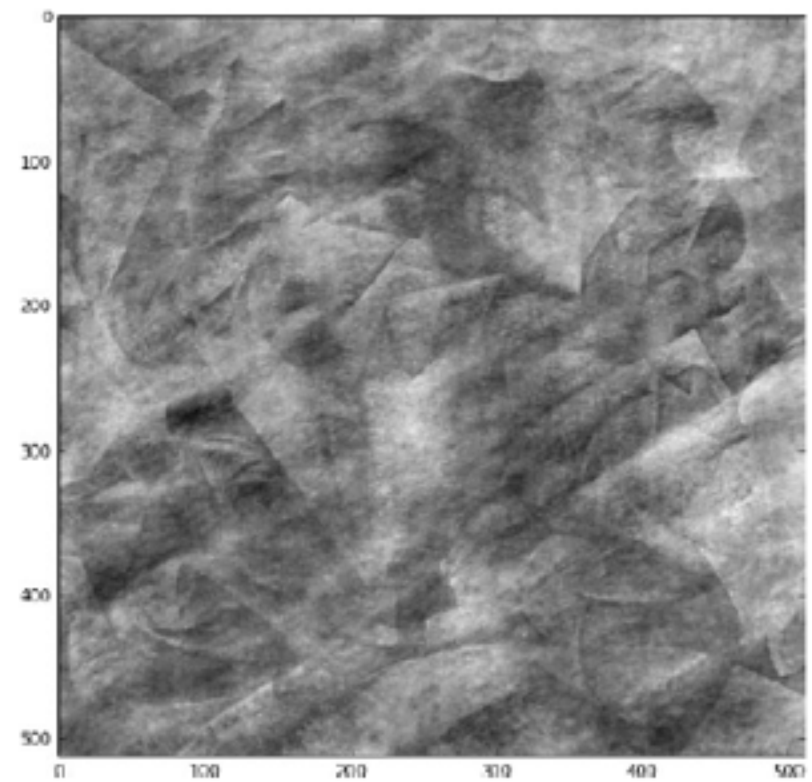
Why (deep) learning?

The duck example of (Non-)Gaussianity



The duck: highly non-Gaussian

randomise phases



Credit: G. Bernardi

The Gaussian duck

Same 2D
power spectrum

- Picks up non-Gaussian information
- Representation learning

Applications:

1. Inference (what duck? what properties? what shapes?)
2. **Detect** the duck (or galaxy, or signature)

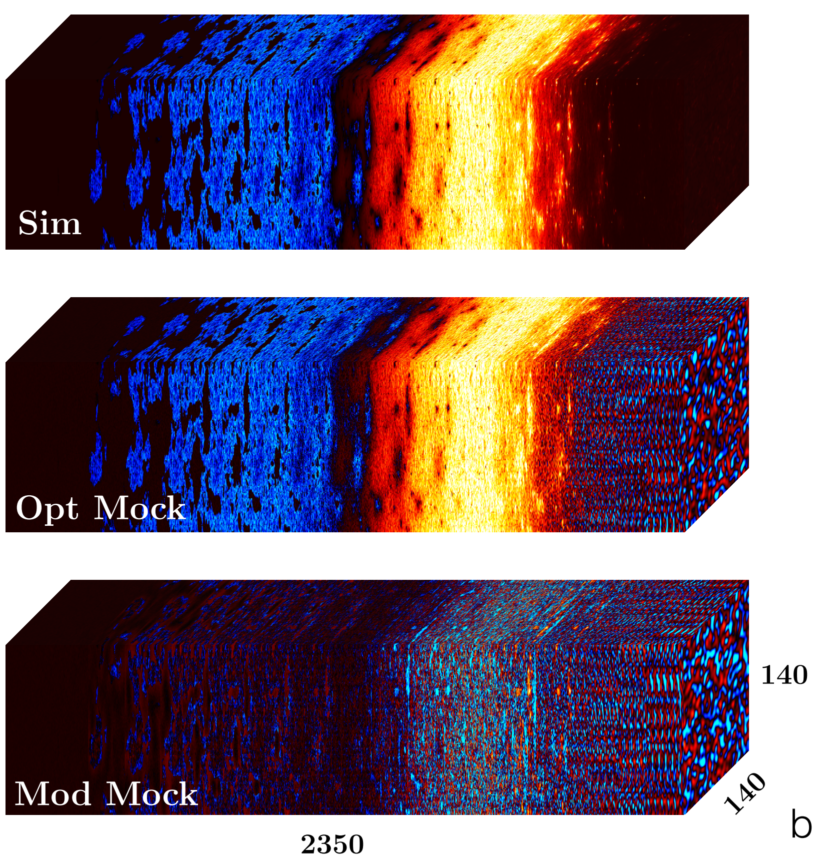
Inference from (mock) 3D tomographic cubes

Direct likelihood-free inference from 3D tomographic mock cubes (21cm IM)

Goal to infer astrophysical and cosmological key parameters
directly from intensity lightcones $(\Omega_{\mathrm{m}},\, \zeta,\, T_{\mathrm{vir}},\, L_{\mathrm{X}},\, E_0,\, m_{\mathrm{WDM}})$

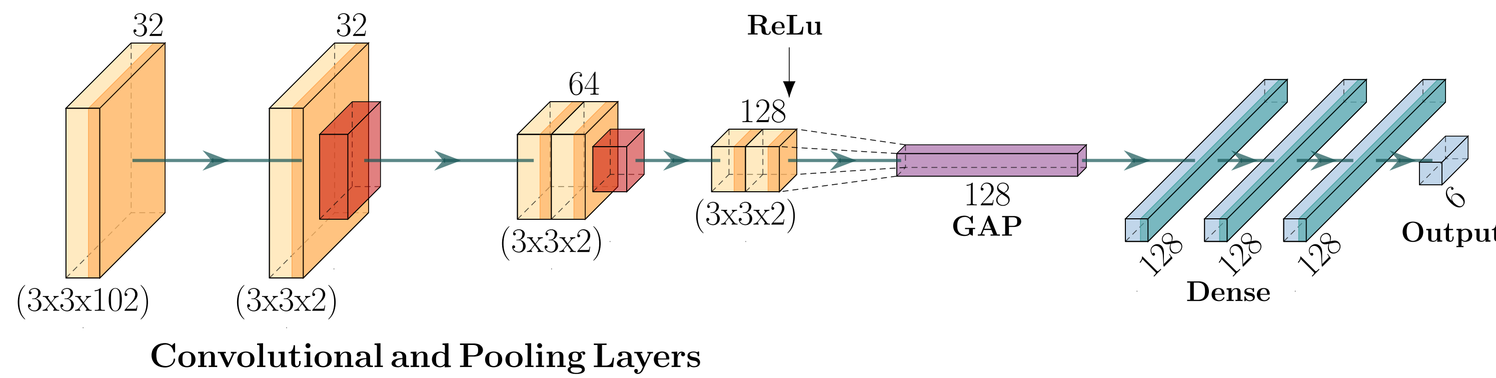
<u>cosmic reionisation</u>
'the last big phase transition'

<u>cosmic dawn</u>
'first stars and galaxies'



21-cm lightcone
generated with 21cmFAST

Mocks created with
21cmSense for SKA1-Low



blue = emission
red / yellow = absorption
against the Cosmic Microwave Background

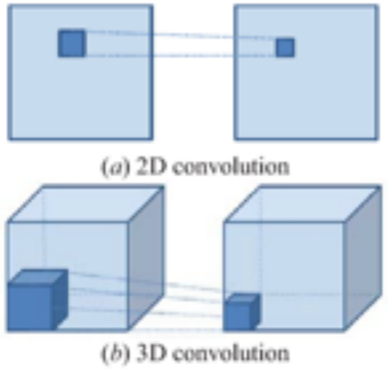
Neutsch, Heneka, Brüggen, in prep

Inference from 3D tomographic cubes

Direct likelihood-free inference from 3D tomographic cubes (21cm IM)

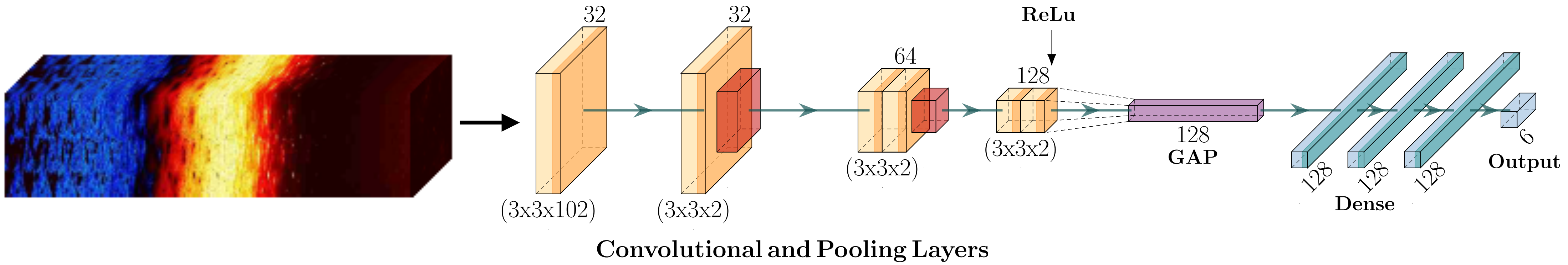
Various Options:

- Slicing and treatment with 2D CNN as image → 'standard' 2D CNN, residual (skip connections) 2D CNN
- Time series (frequency) of co-eval images → LSTM network
- Full 3D convolution → 3D CNN:



Moving from **2D to full 3D convolution**

Best-performing: simple Conv3D architecture



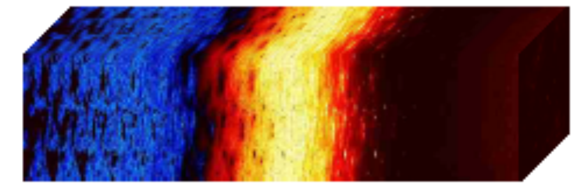
Database of ~5000 lightcones
140x140x2350 pix, 1.4 Mpc resolution

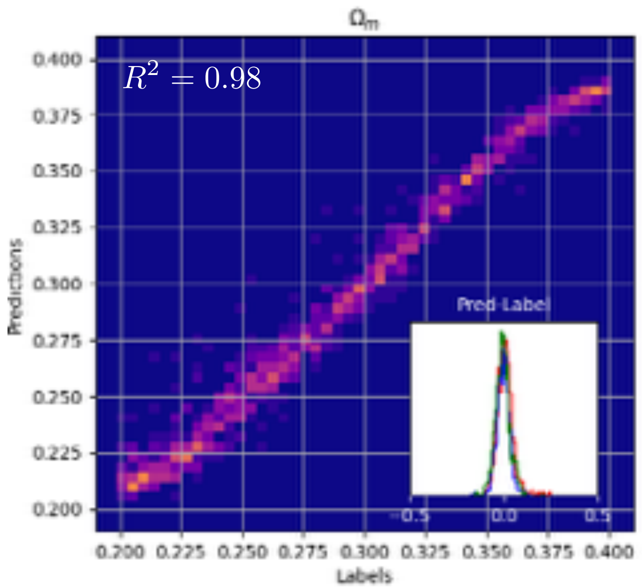
Neutsch, Heneka, Brüggen, in prep

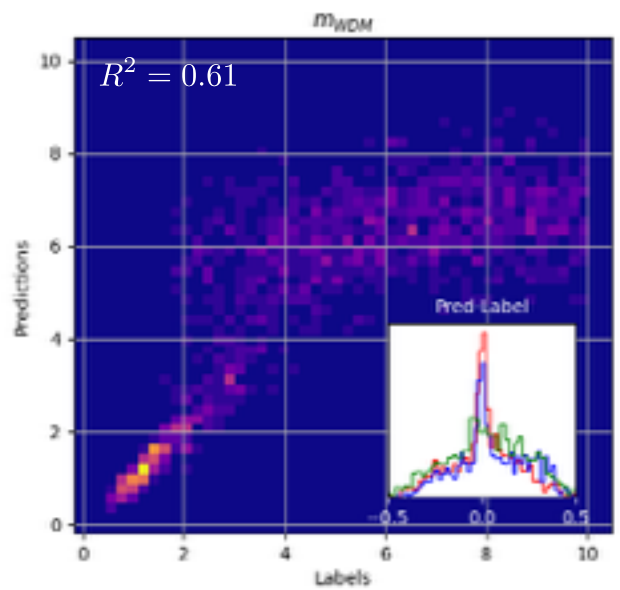
Inference from 3D tomographic cubes

Direct likelihood-free inference from 3D tomographic mock cubes (21cm IM)

$$(\Omega_{\rm m},\, \zeta,\, T_{\rm vir},\, L_{\rm X},\, E_0,\, m_{\rm WDM})$$

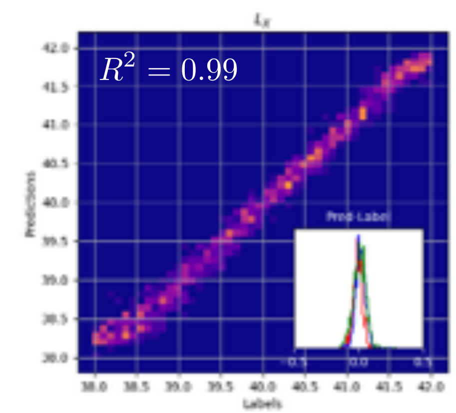


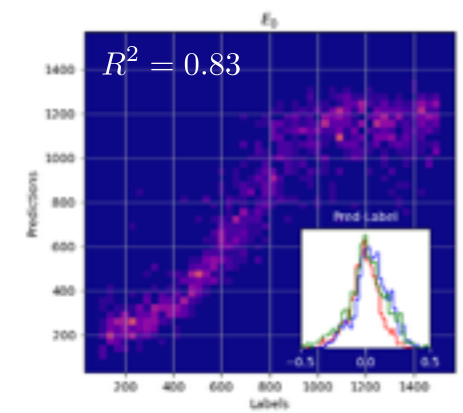




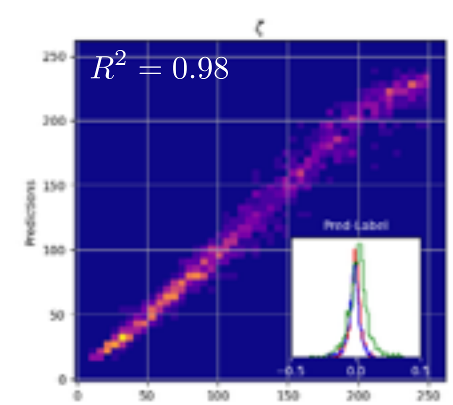
Cosmology

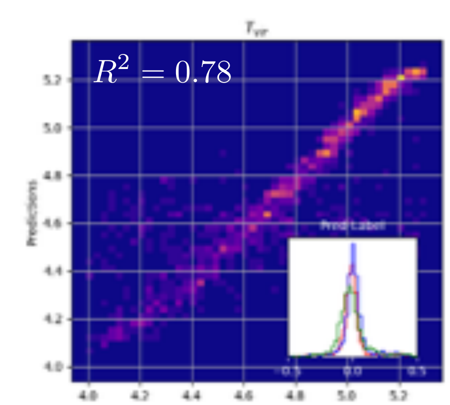
Cosmic Dawn





Reionisation





Directly constrain cosmology, cosmic dawn and reionisation astrophysics

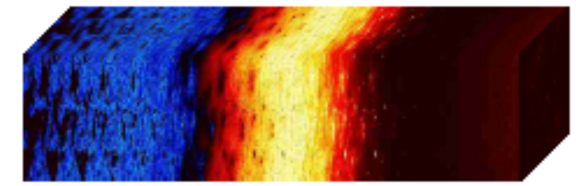
Also tested: ‘astro-only’ inference

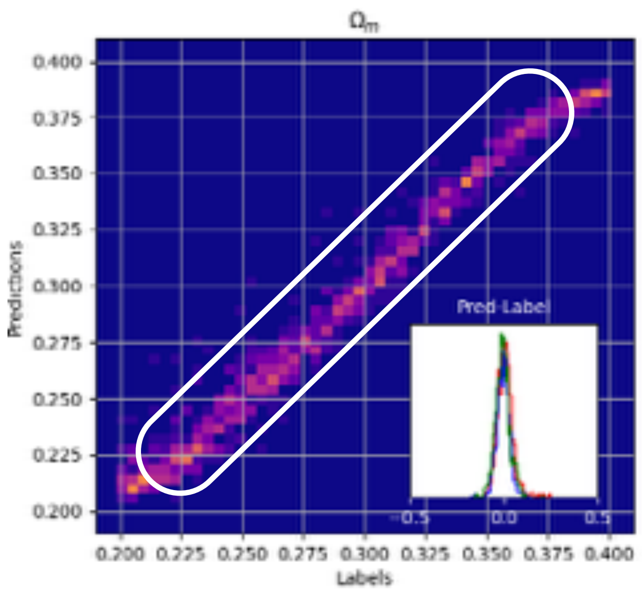
Neutsch, Heneka, Brüggen, in prep

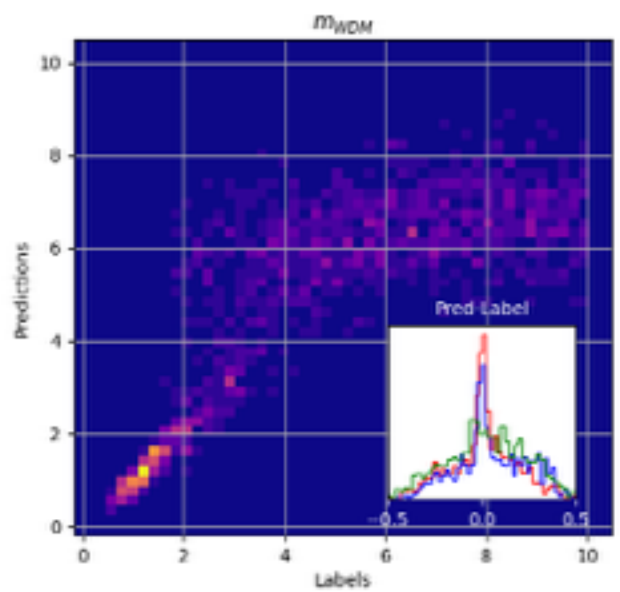
Inference from 3D tomographic cubes

Direct likelihood-free inference from 3D tomographic mock cubes (21cm IM)

$$(\Omega_{\rm m},\, \zeta,\, T_{\rm vir},\, L_{\rm X},\, E_0,\, m_{\rm WDM})$$



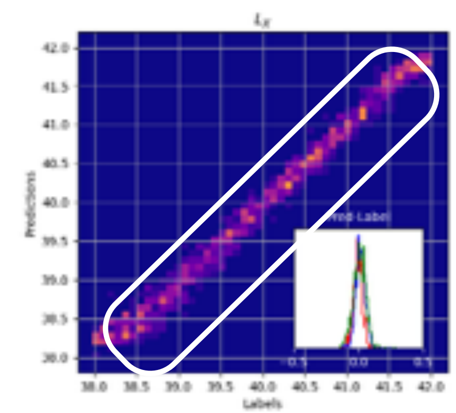


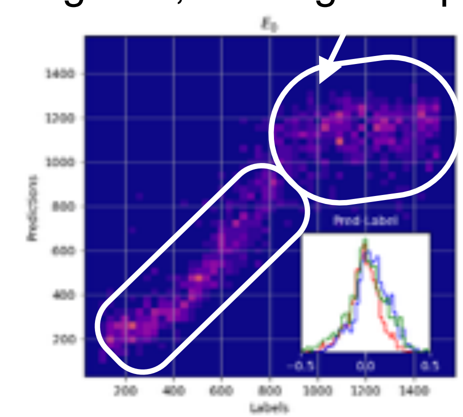


Cosmology

Cosmic Dawn

high E0, nothing escapes

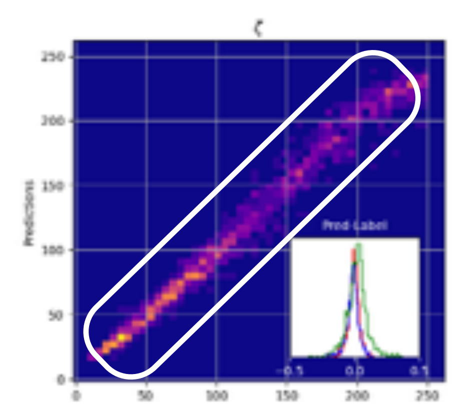


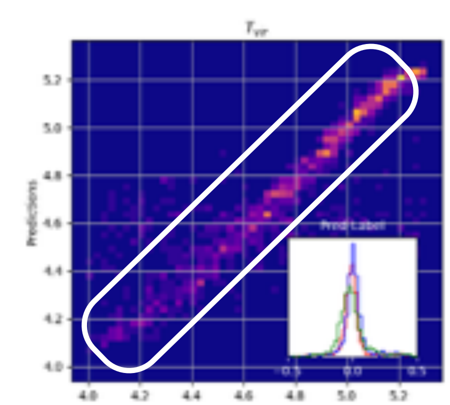


Directly constrain cosmology, cosmic dawn and reionisation astrophysics

Also tested: ‘astro-only’ inference

Reionisation

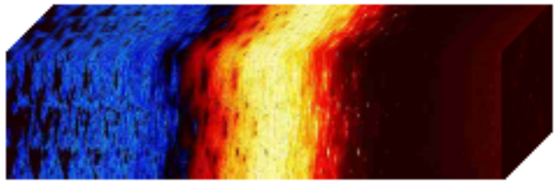




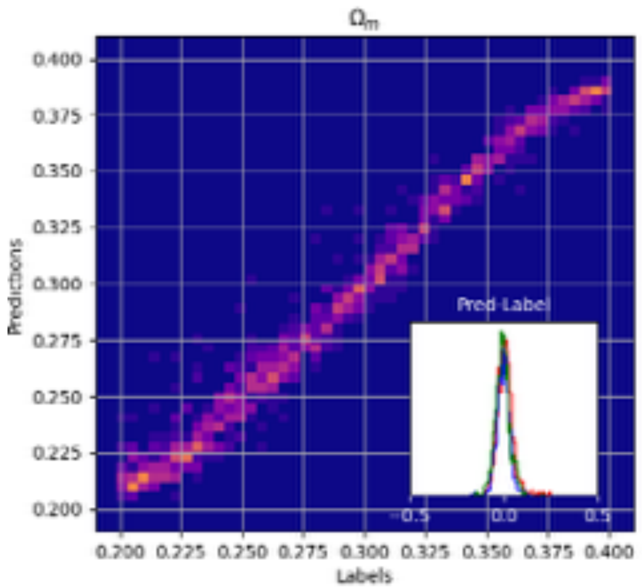
Neutsch, Heneka, Brüggen, in prep

Inference from 3D tomographic cubes

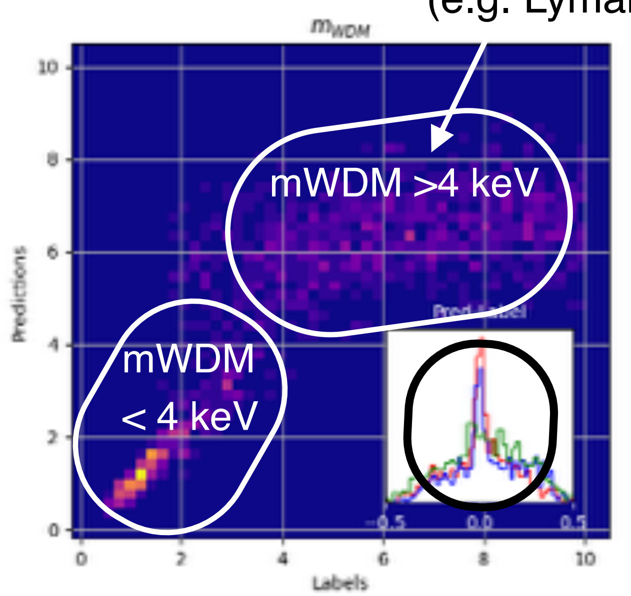
Direct likelihood-free inference from 3D tomographic mock cubes (21cm IM)



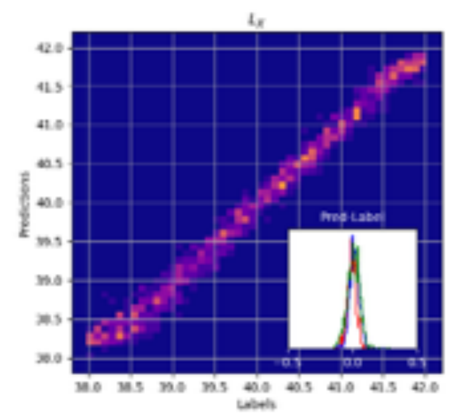
$$(\Omega_\mathrm{m},\, \zeta,\, T_\mathrm{vir},\, L_\mathrm{X},\, E_0,\, m_\mathrm{WDM})$$

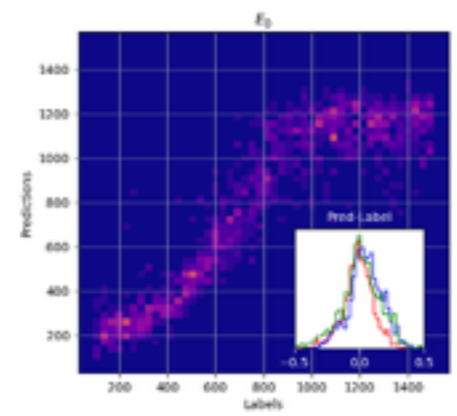


~results from other studies
(e.g. Lyman-forest)



Cosmic Dawn



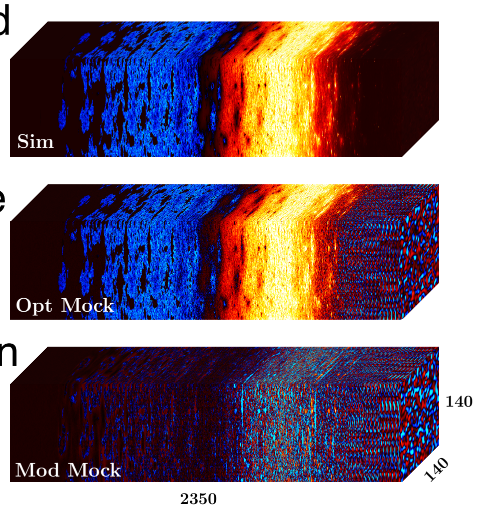


Cosmology

red

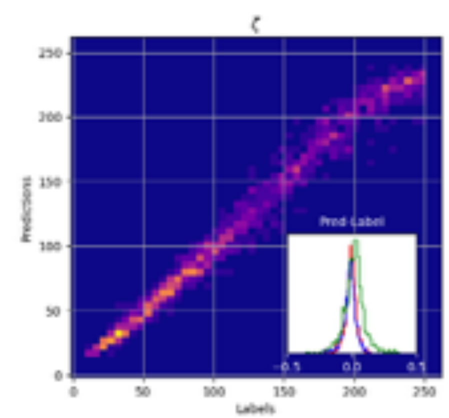
blue

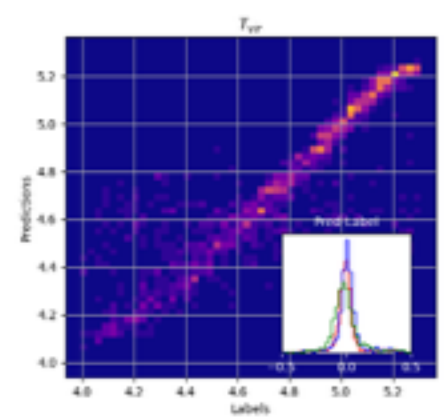
green



Directly constrain cosmology, cosmic dawn and reionisation astrophysics

Reionisation

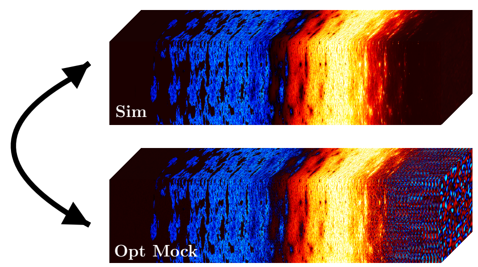




Neutsch, Heneka, Brüggen, in prep

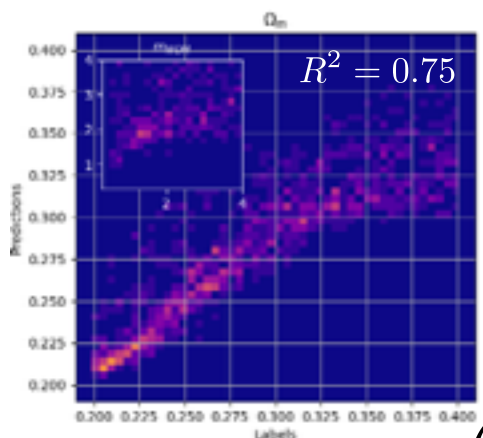
Testing robustness & interpretability

1) Transfer learning Sims & Mocks

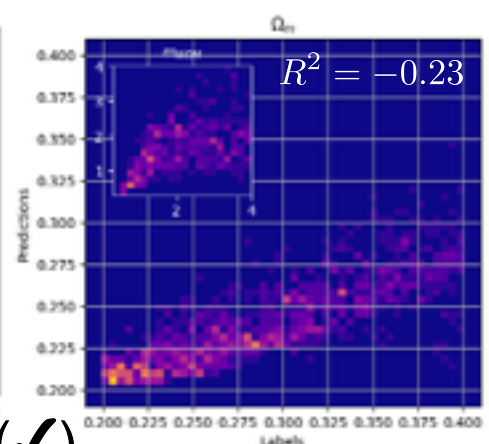


Sim -> Mock

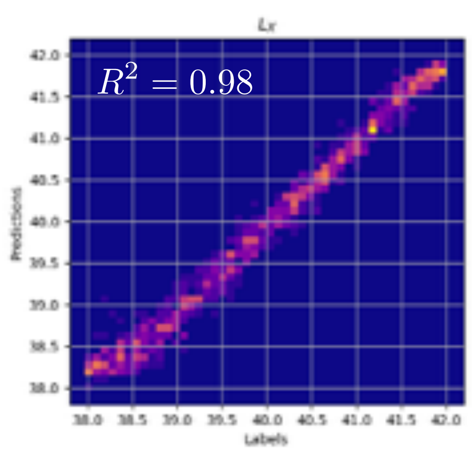
Mock -> Sim



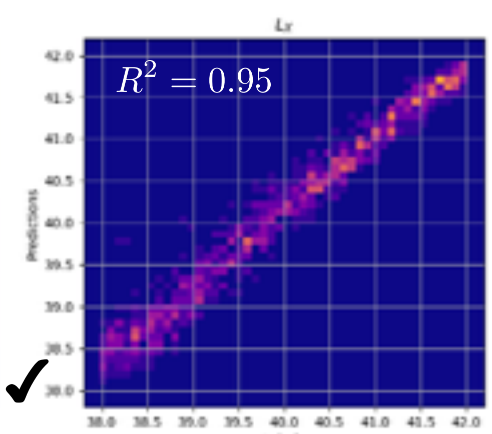
(✓)



✗



✓

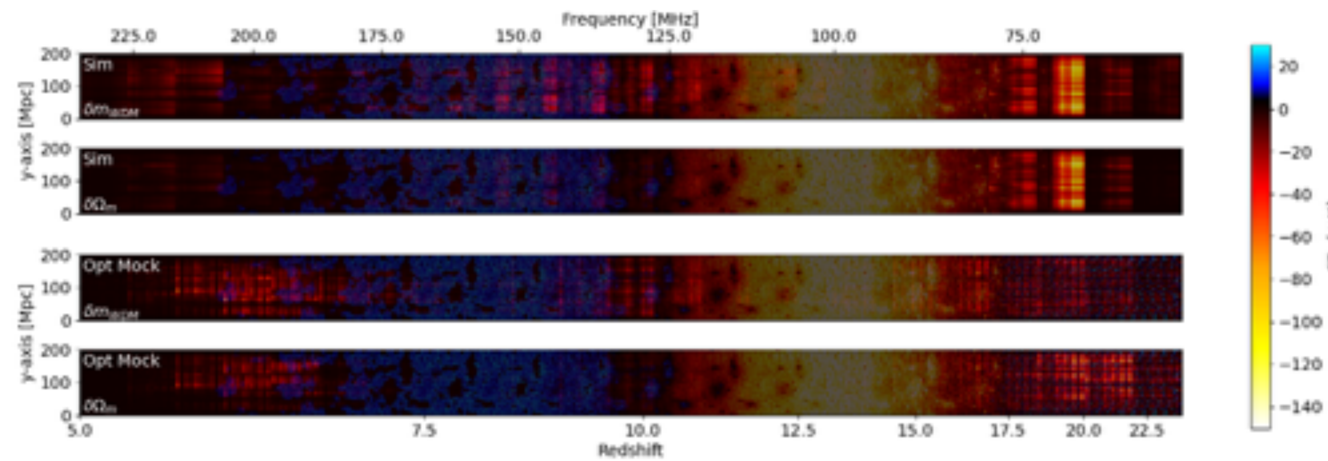


✓

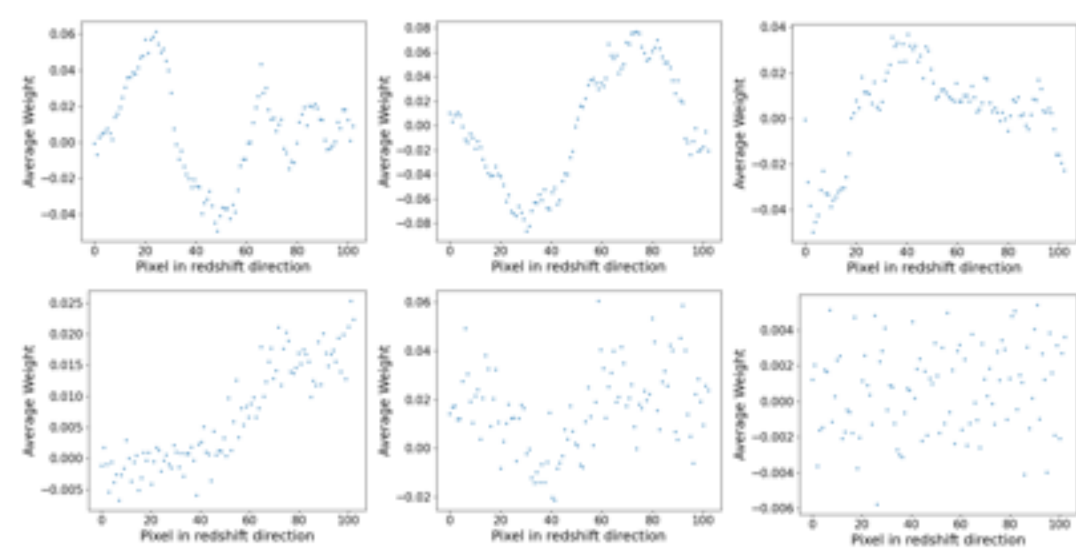
2) Gradient-based saliency maps

Sim

Mock



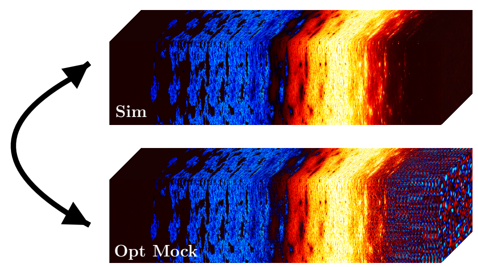
+ check your filters



Neutsch, Heneka, Brüggen, in prep

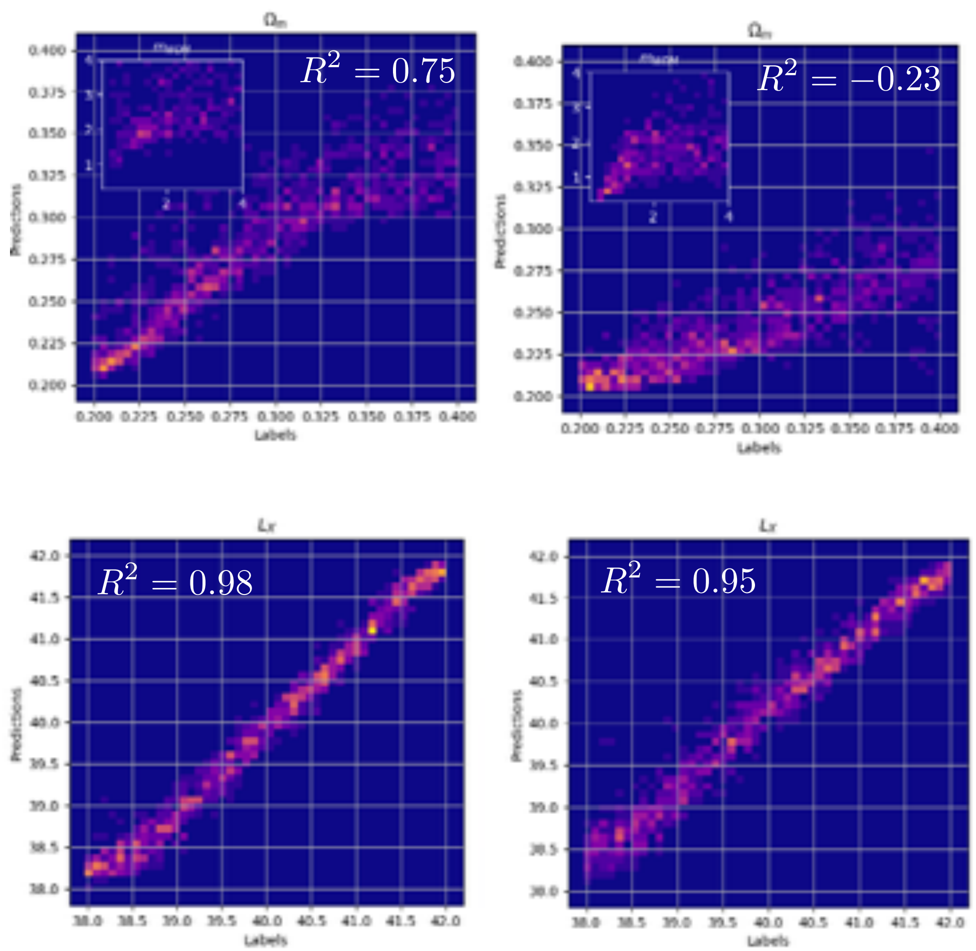
Testing robustness & interpretability

1) Transfer learning Sims & Mocks



Sim -> Mock

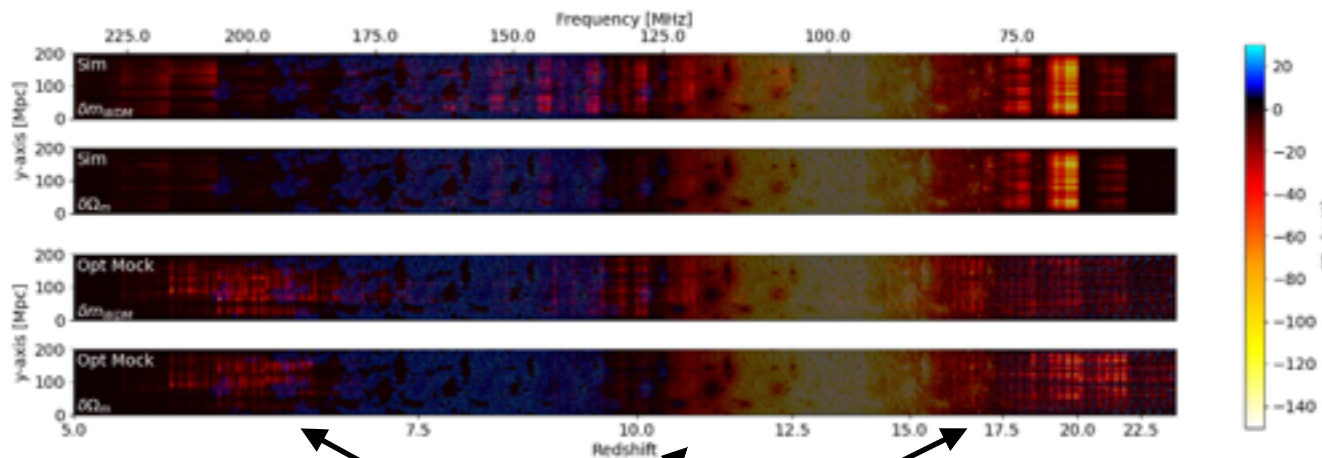
Mock -> Sim



2) Gradient-based saliency maps

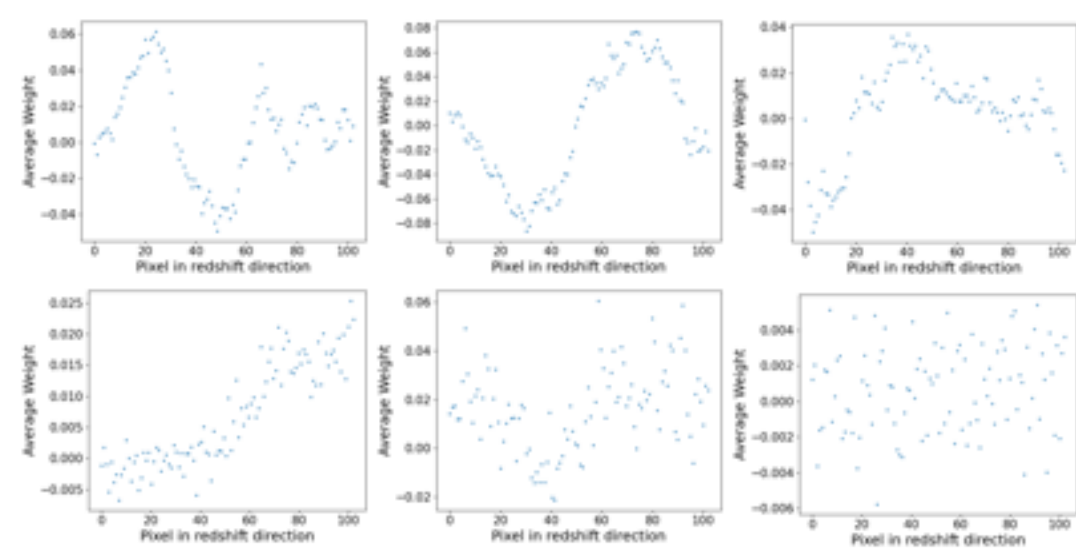
Sim

Mock



‘transitions’

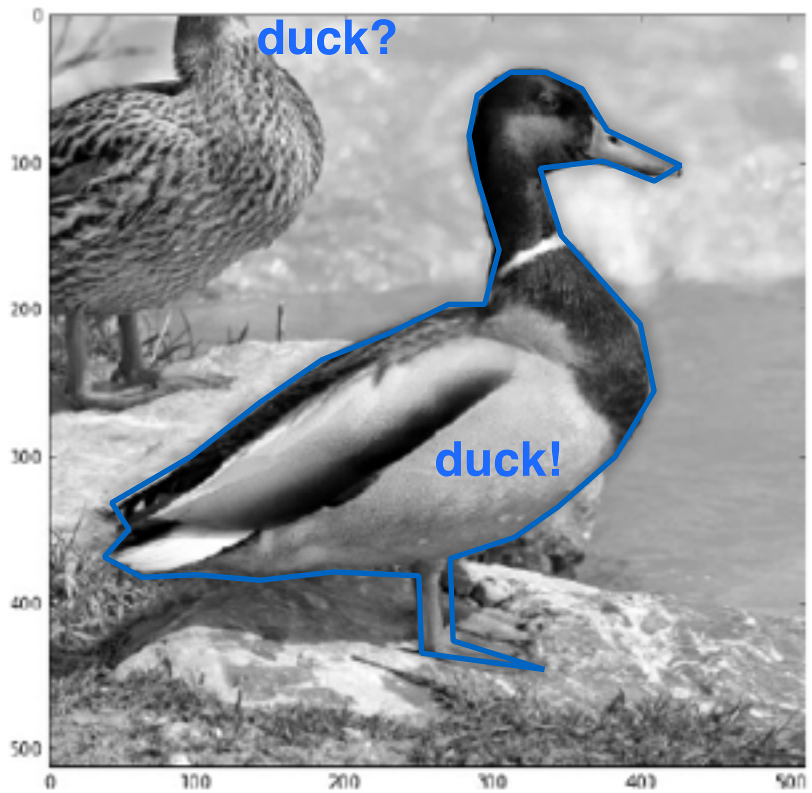
+ check your filters



Neutsch, Heneka, Brüggen, in prep

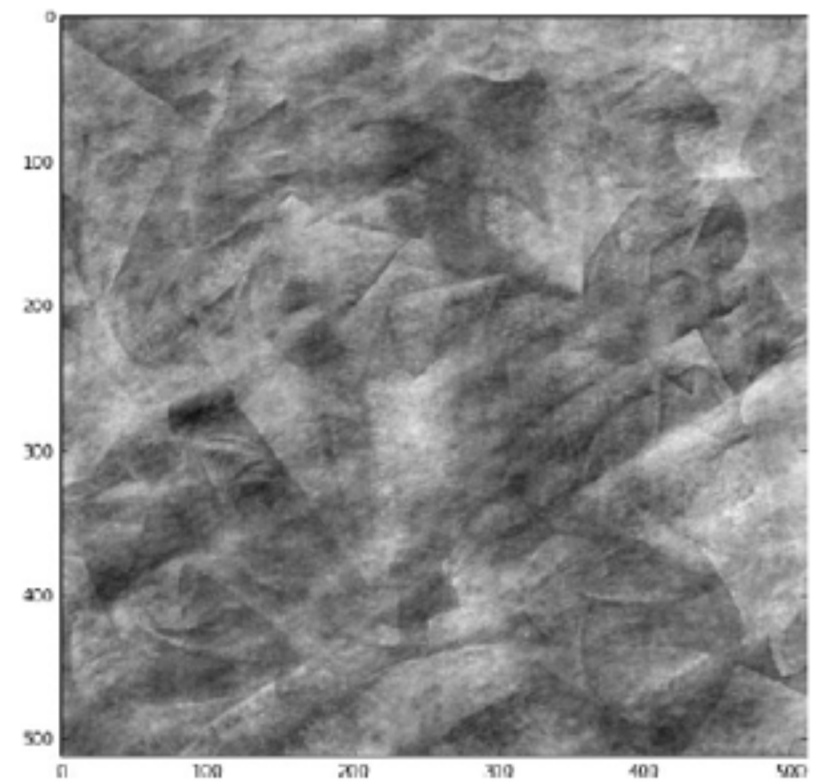
Why (deep) learning?

The duck example of (Non-)Gaussianity



The duck: highly non-Gaussian

randomise phases



Credit: G. Bernardi

The Gaussian duck

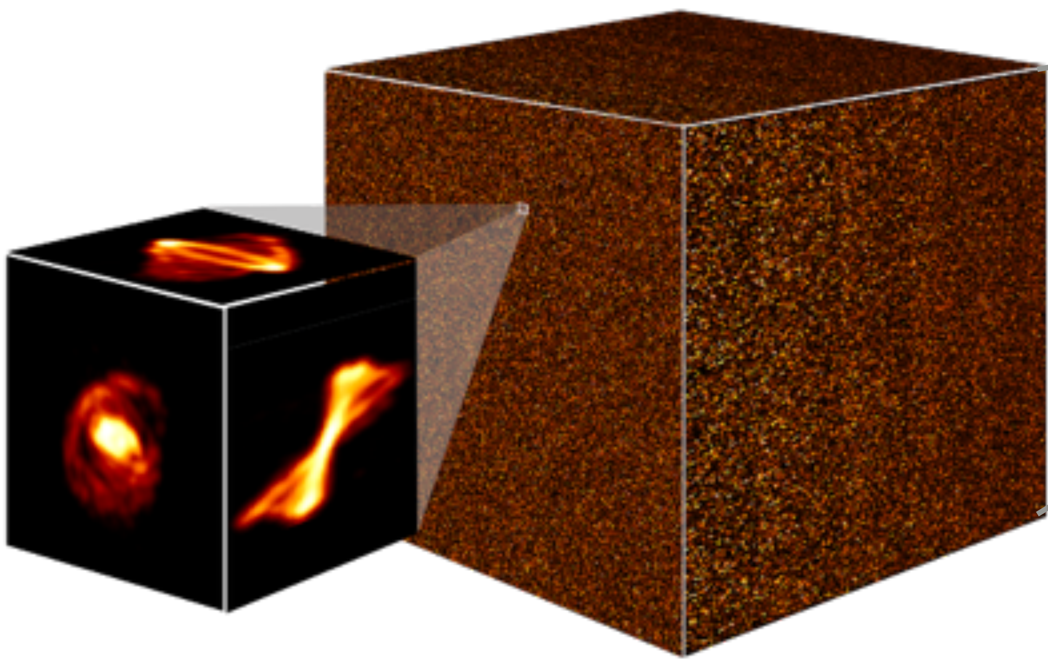
Same 2D
power spectrum

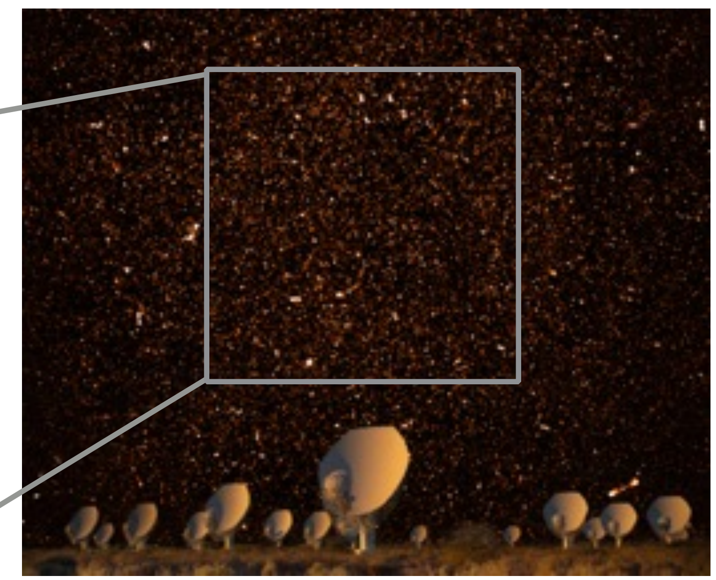
- Picks up non-Gaussian information
- Representation learning

Applications:

1. Inference (what duck? what properties? what shapes?)
2. **Detect** the duck (or galaxy, or signature)

Source Finding: SKA Science Data Challenge





Composite MeerKAT dishes and observations.
Credit: South African Radio Astronomy Observatory (SARAO)

SKA -
The Square Kilometre Array

An international effort to build the world's largest radio telescope

Expected data rate in full operation: 1 TB/s

Key science goals include:
Galaxy Evolution, Reionisation, Cosmology, Astroparticles

Credit: https://sdc2.astronomers.skatelescope.org/sdc2-challenge/data

Goal is source finding and characterisation

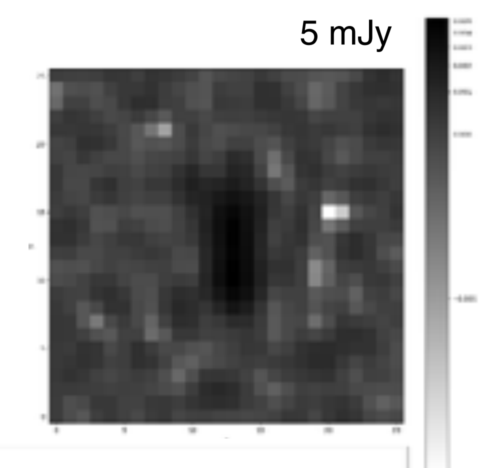
(+ test of computing nodes on the way to SKA)

https://sdc2.astronomers.skatelescope.org/computational-resources

The challenging HI sources:

- **low S/N**
- **small spatial size**
- **systematics**

The brightest
HI source



see also: 1905.01324 (optical, detection & de-blending)

SKA Science Data Challenge

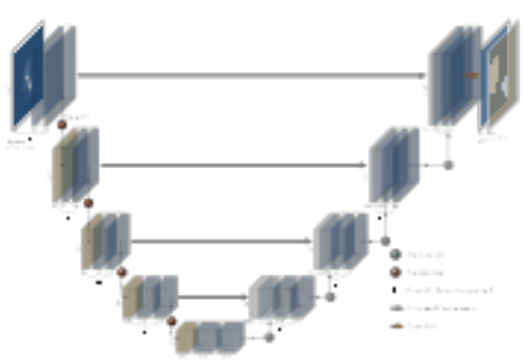
Machine learning and deep learning come together

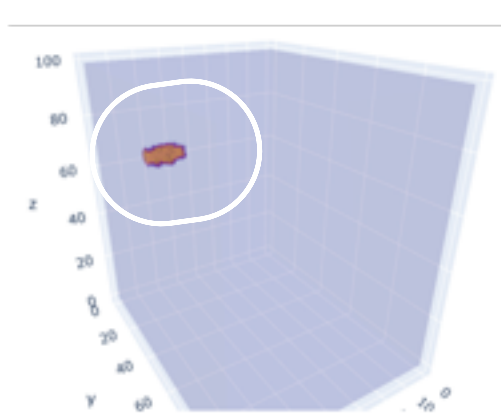
Team: Michelle delle Veneri, Andrew Soroka, Bernardo Fraga, Fedor Gobanov, Clecio de Bom, Alex Meshcheryakov

DL source detection & characterisation:

Best performing: full 3D approaches (U-Net type)

Trials: 2D / 3D variants of U-Net, R-CNN, inception network



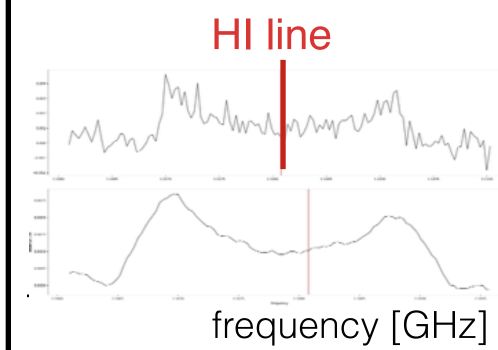


50:50 tp:fp

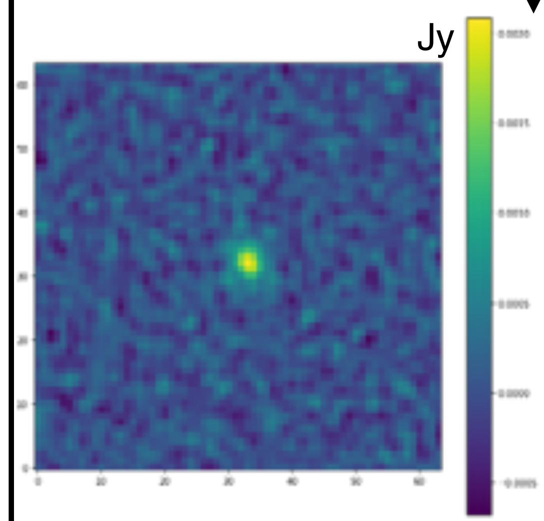
Pitfalls:
- **Pre-processing, noise model(s)**
- **High sparsity**
- **Needs multi-step & ensemble process**
- **Choice of training set**

Source detection baseline

Wavelet denoising & Multi-scale model:

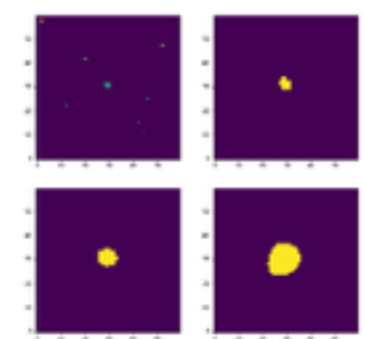


Gaussian filtering in frequency direction

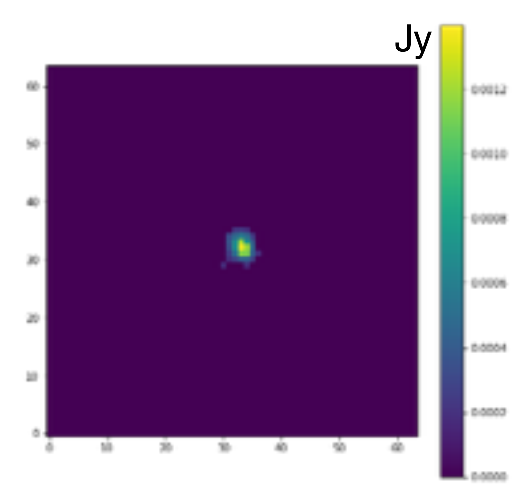


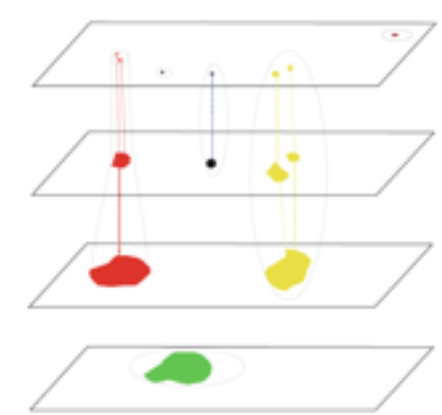
$$c_0[k,l] = c_J[k,l] + \sum_{j=1}^{J} w_j[k,l]$$

1) Transformation
2) Thresholding



3) Connectivity
4) Reconstruction





Inter-scale connectivity graph. Starck+2010

+ Iteration (‘CLEAN’ algorithm)

Conclusion & Outlook: 21cm tomography with nets

Main take-aways:

- Beyond Gaussianity: Direct inference from tomography with nets
- Avenue to jointly constrain astrophysics and cosmology at Cosmic Dawn and Reionization
- 3D net for 3D data
- SKA source detection: pitfalls in low S/N regime

Ongoing & future steps:

- Code will be made public in Github (@Steffen Neutsch)
- Test of Bayesian network for errors on parameters
- Test on data from SKA precursors & improved mocks

Thank you!

caroline.heneka@uni-hamburg.de, www.caroline-heneka.de