Painting with baryons

Augmenting N-body simulations with gas using deep generative models

Tilman Tröster

ttr@roe.ac.uk github.com/tilmantroester

Institute for Astronomy, University of Edinburgh

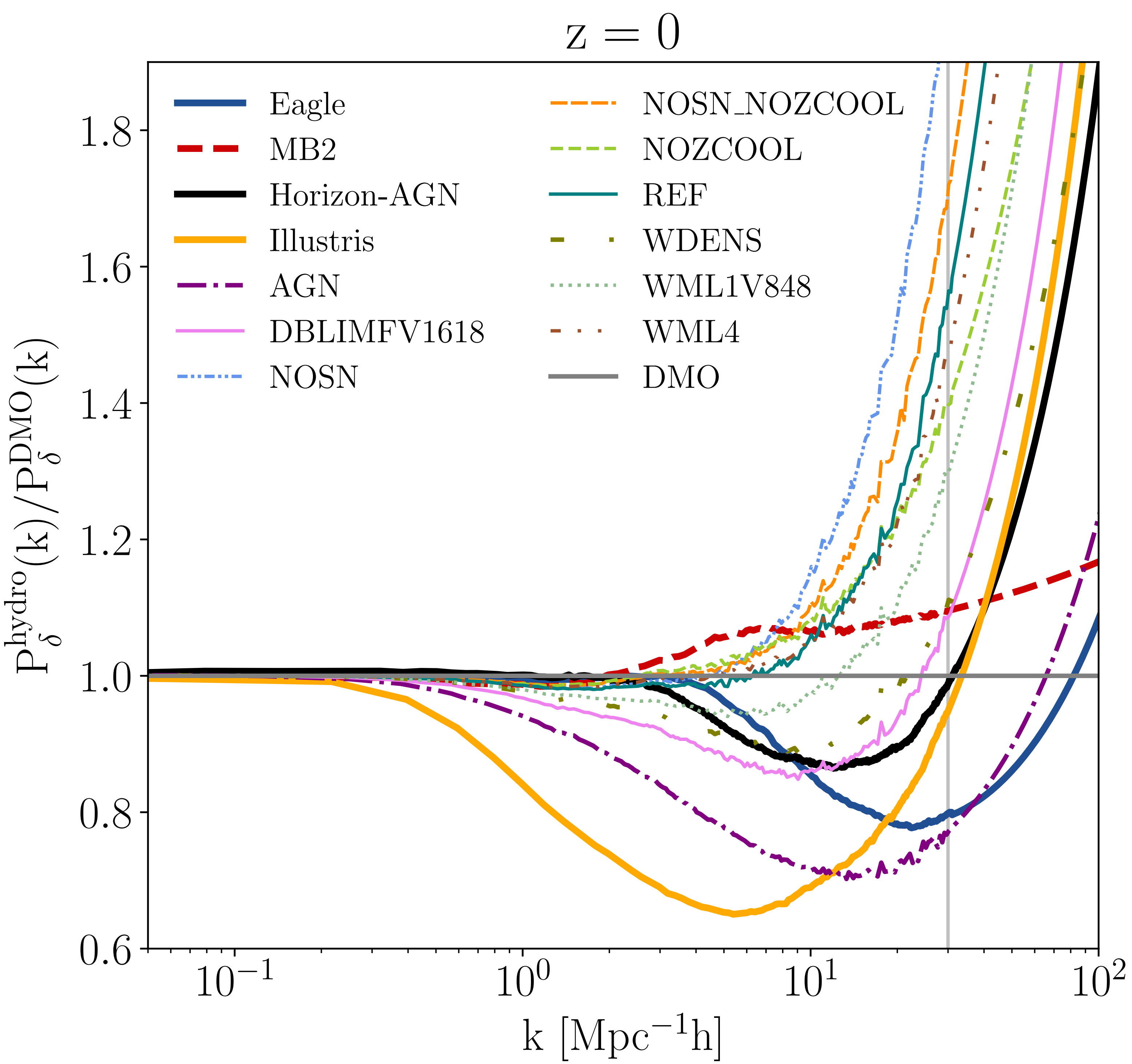
AIA 2019, Garching, 26 July 2019

Gravitational lensing

Gravitational lensing probes the clustering of matter

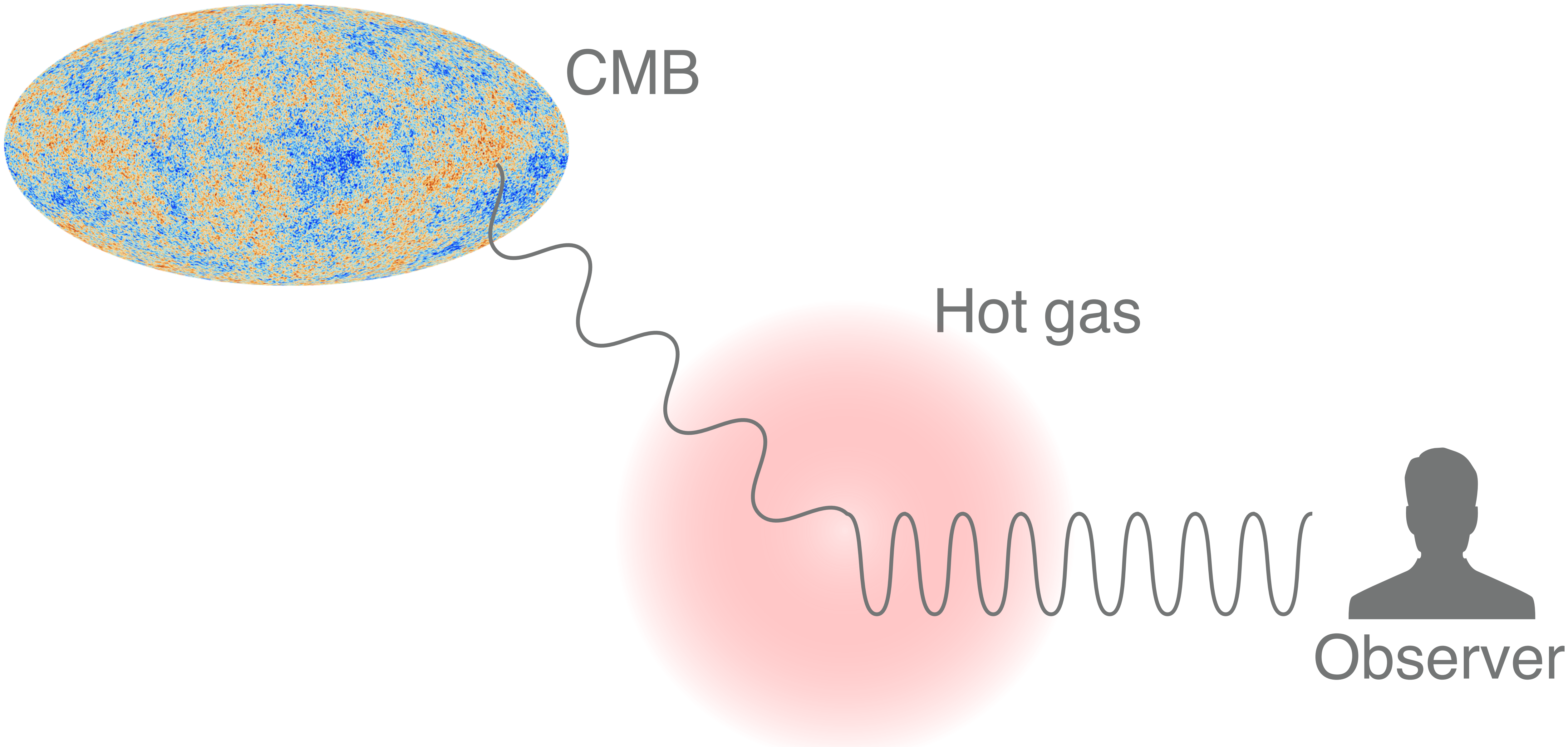
- ~80% of matter is dark matter
- ~20% is baryons
- Baryons are complicated!

Effect of baryons on the matter power spectrum

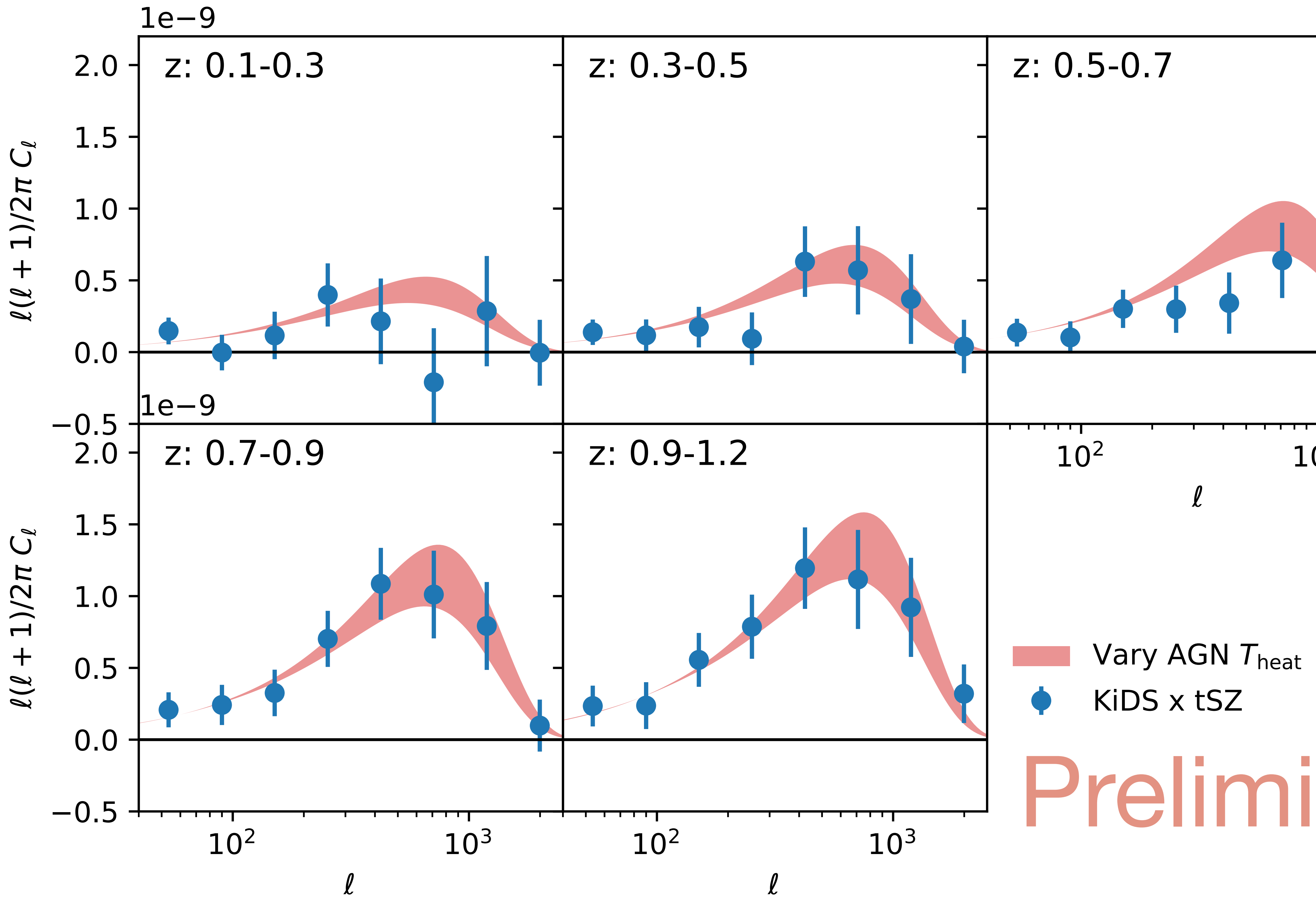


Huang+ 2018

Thermal Sunyaev-Zel'dovich (tSZ) Effect



Cross-correlate tSZ with lensing (Planck x KiDS-1000)



Preliminary!

Challenge: Covariance matrices

Use simulations

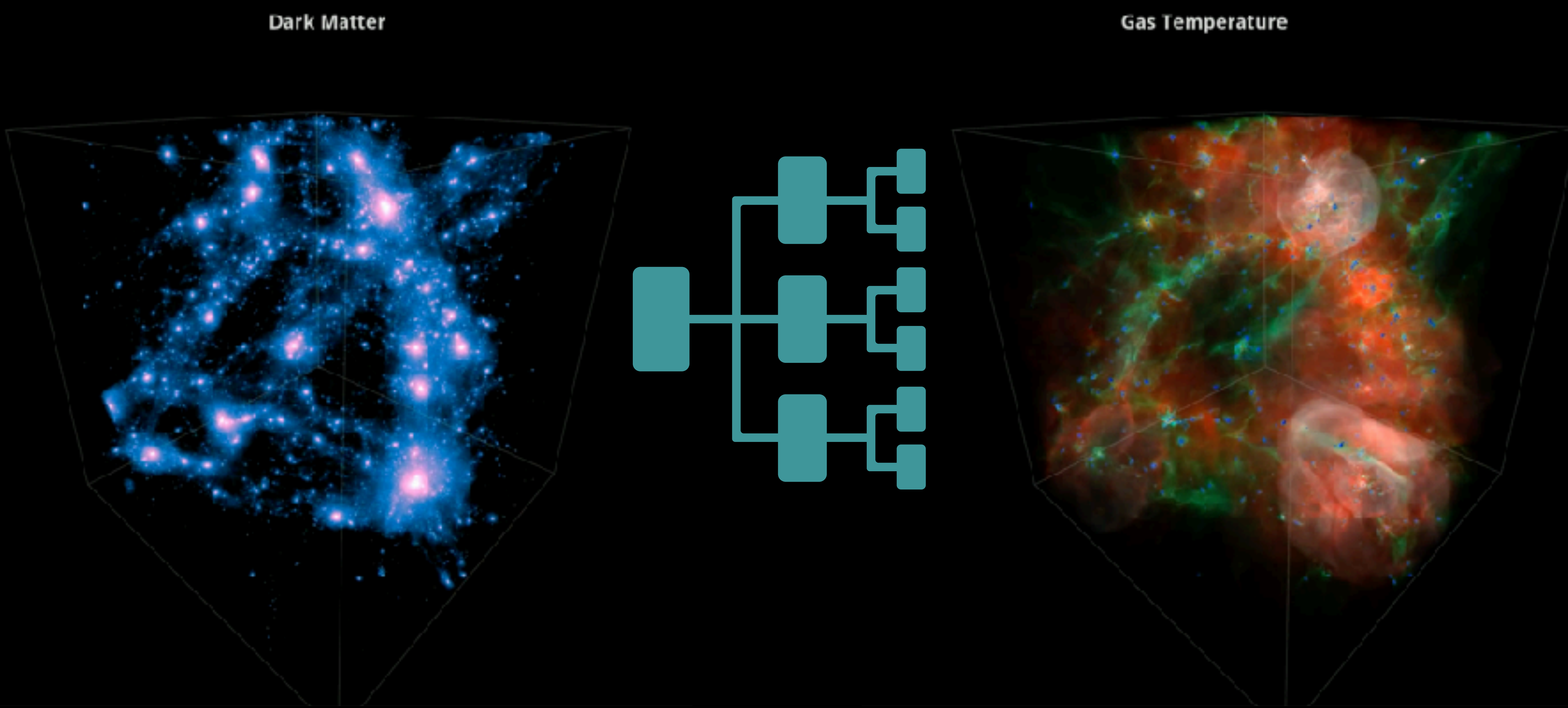
- Need $O(10^3)$ hydrodynamical simulations for tSZ+lensing
  - Expensive (~10^5 CPU hours)
  - Dark matter-only simulations are cheap (in comparison)

Why are hydro sims hard?

Feedback couples large and small scales

- Simulating large and small scales at the same time is hard
- But we don’t care about the small scales

Use machine learning?

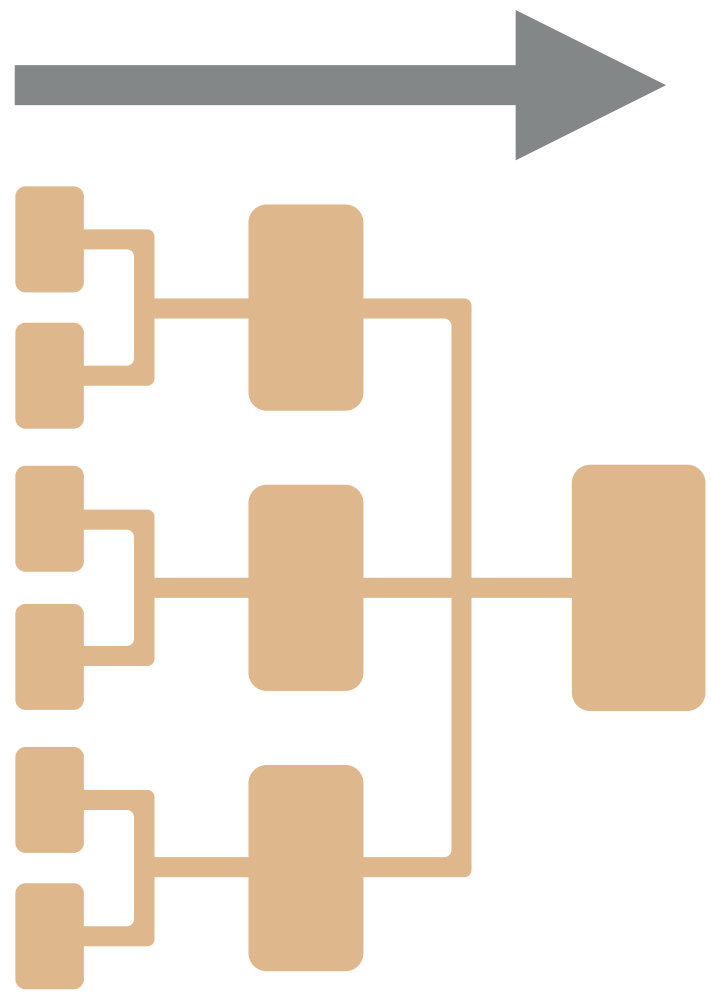


Vogelsberger+2004

Classification

Input

Output

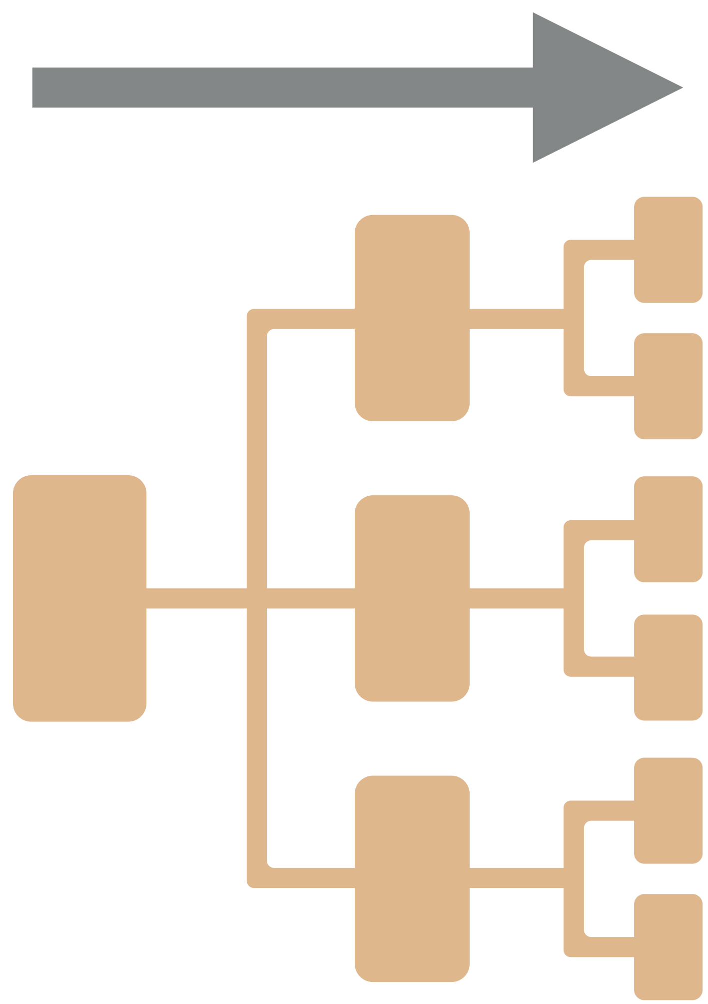


"cat"

Generative model: reverse classification

Input

“cat”



Output

Karras+2018

Karras+ 2019

Kingma+2018

Zhu+2017

Generative models

Variational auto-encoder (VAE)

- Easy to train
- Can predict variance of output

Generative adversarial network (GAN)

- Tends to give better results
- Training is more challenging; often unstable

Conditional Variational Auto-Encoder (CVAE)

Basic problem: given dark matter, sample pressure

- x is pressure, y is dark matter
    - $x \sim p(x|y)$

Introduce latent variable z

- $p(x|y) = \int \mathrm{d}z \; p(x, z|y) = \int \mathrm{d}z \; p(x|y, z)p(z|y)$
- Infinite mixture model

Conditional Variational Auto-Encoder (CVAE)

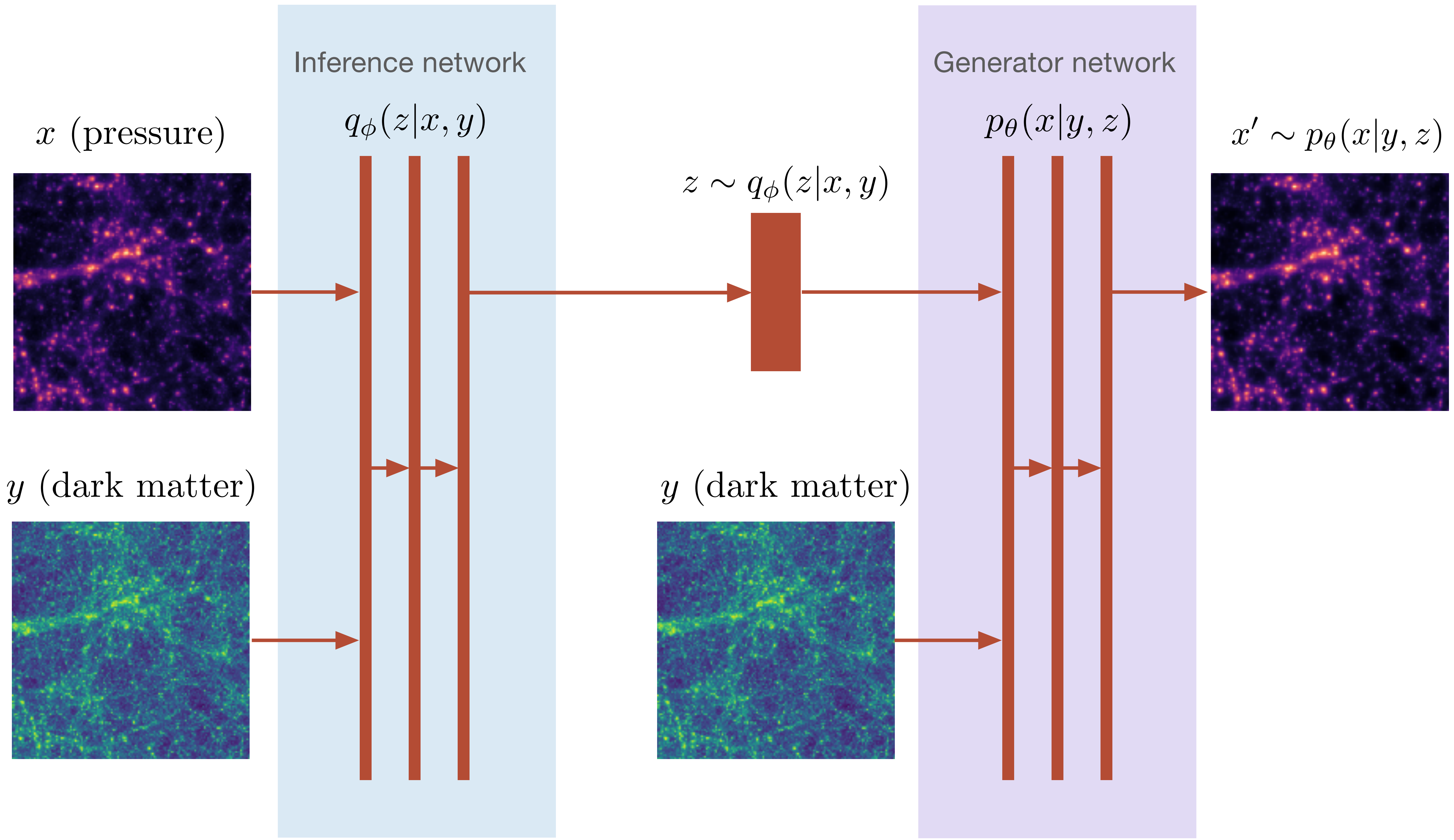
Parameterize as multivariate Gaussians

- Generator network $p_{\theta_2}(x|y,z)$
- Prior network $p_{\theta_1}(z|y)$
- Inference network $q_\phi(z|x,y)$

Variational lower bound

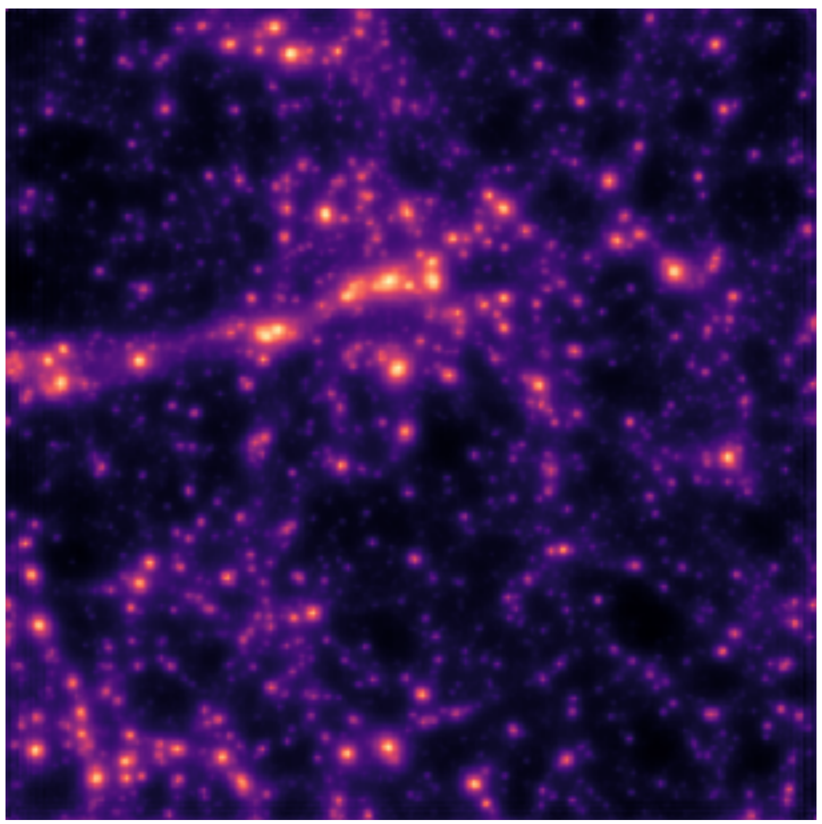
$$\log p(x|y) \geq \underbrace{-\mathbb{D}_{\mathrm{KL}}(q_\phi(z|x,y)||p_{\theta_1}(z|y))}_{\text{KL-term}} + \underbrace{\mathbb{E}_{z\sim q_\phi(z|x,y)}[\log p_{\theta_2}(x|y,z)]}_{\text{Reconstruction}}$$

Conditional Variational Auto-Encoder (CVAE)

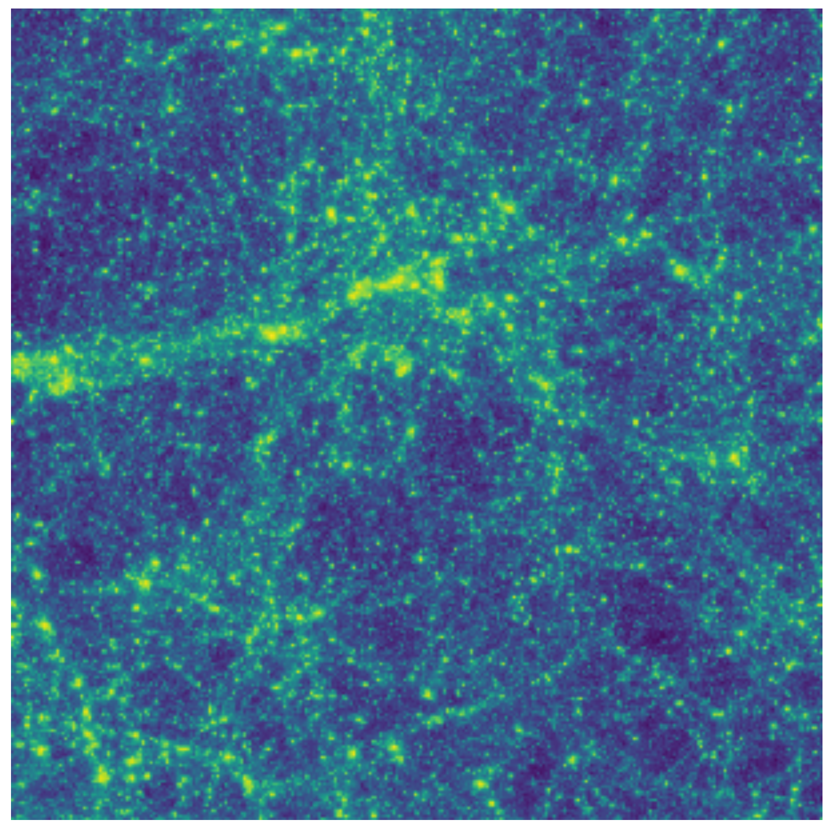


Results

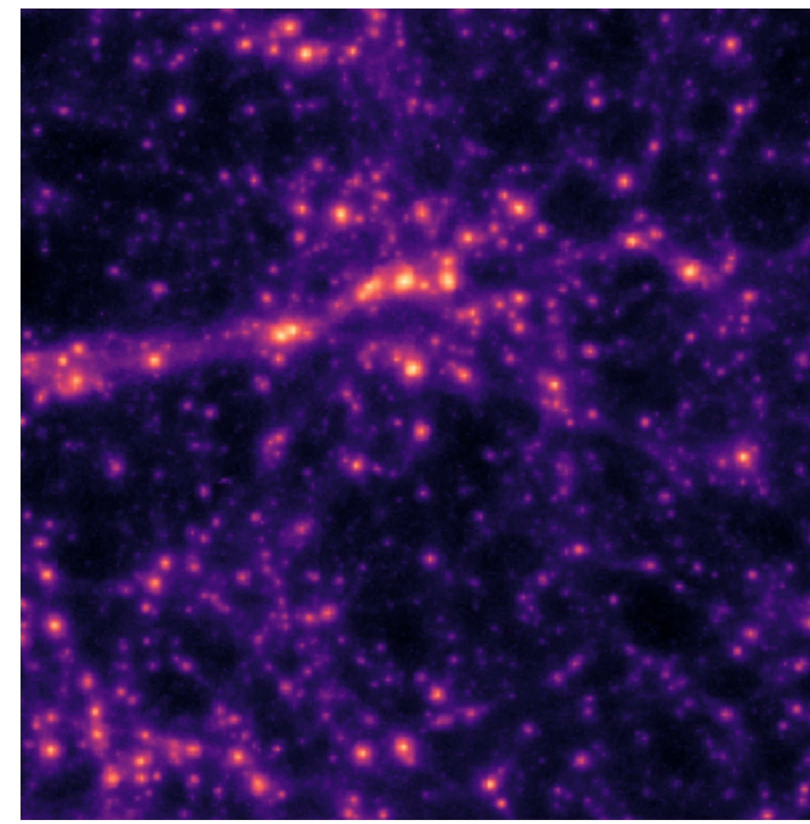
Pressure (fake)



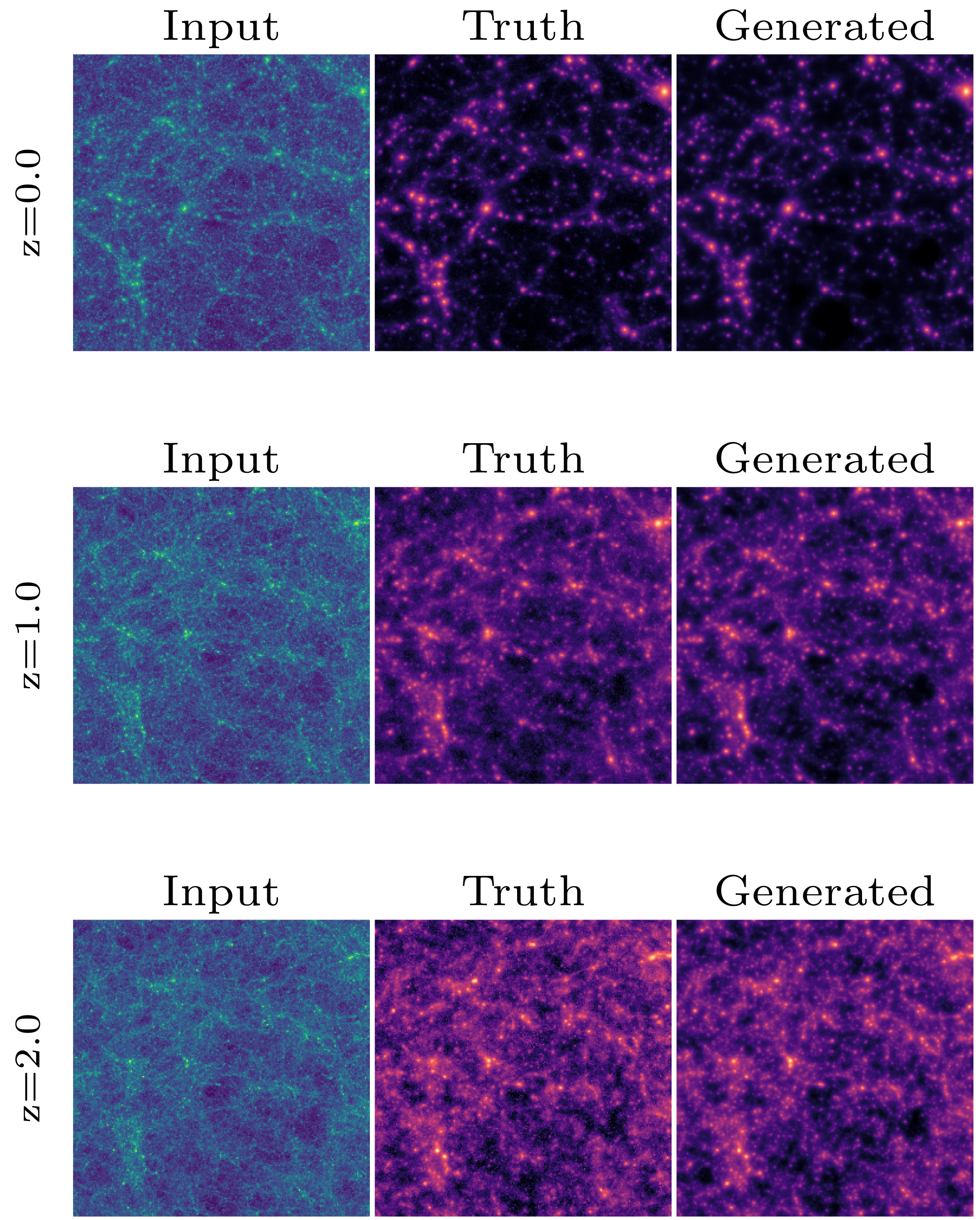
Dark matter (input)



Pressure (truth)

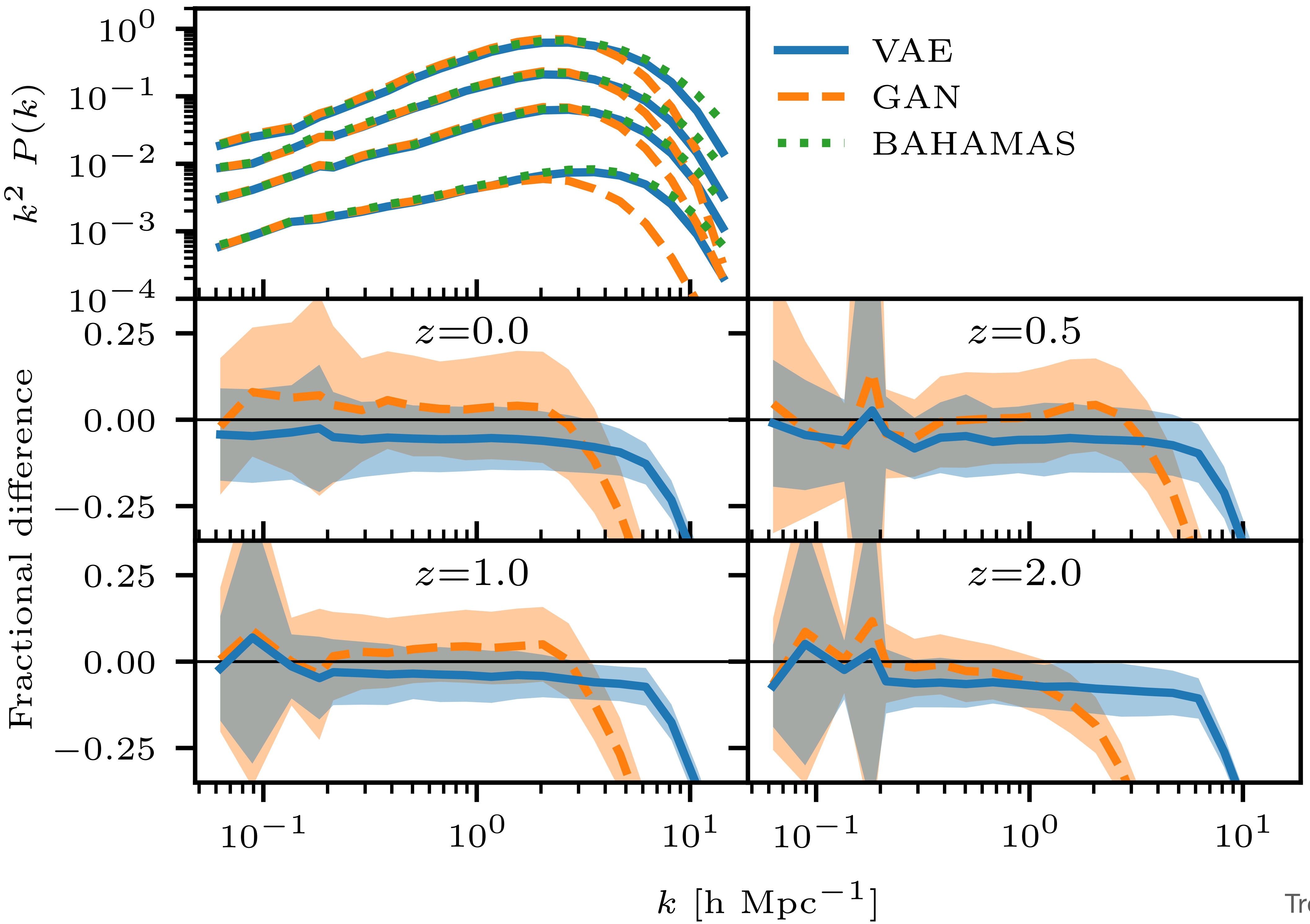


Results



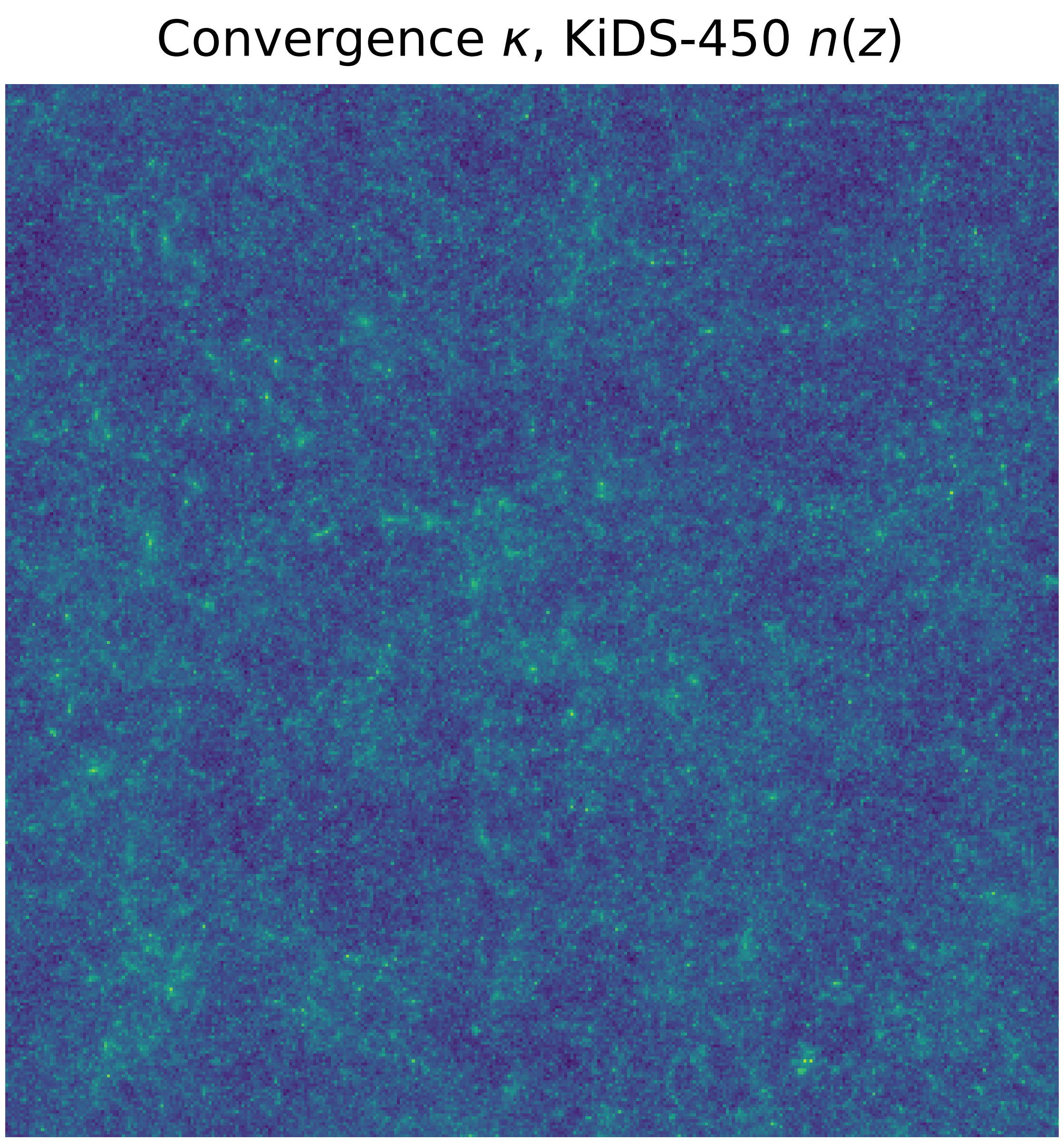
Tröster+2019

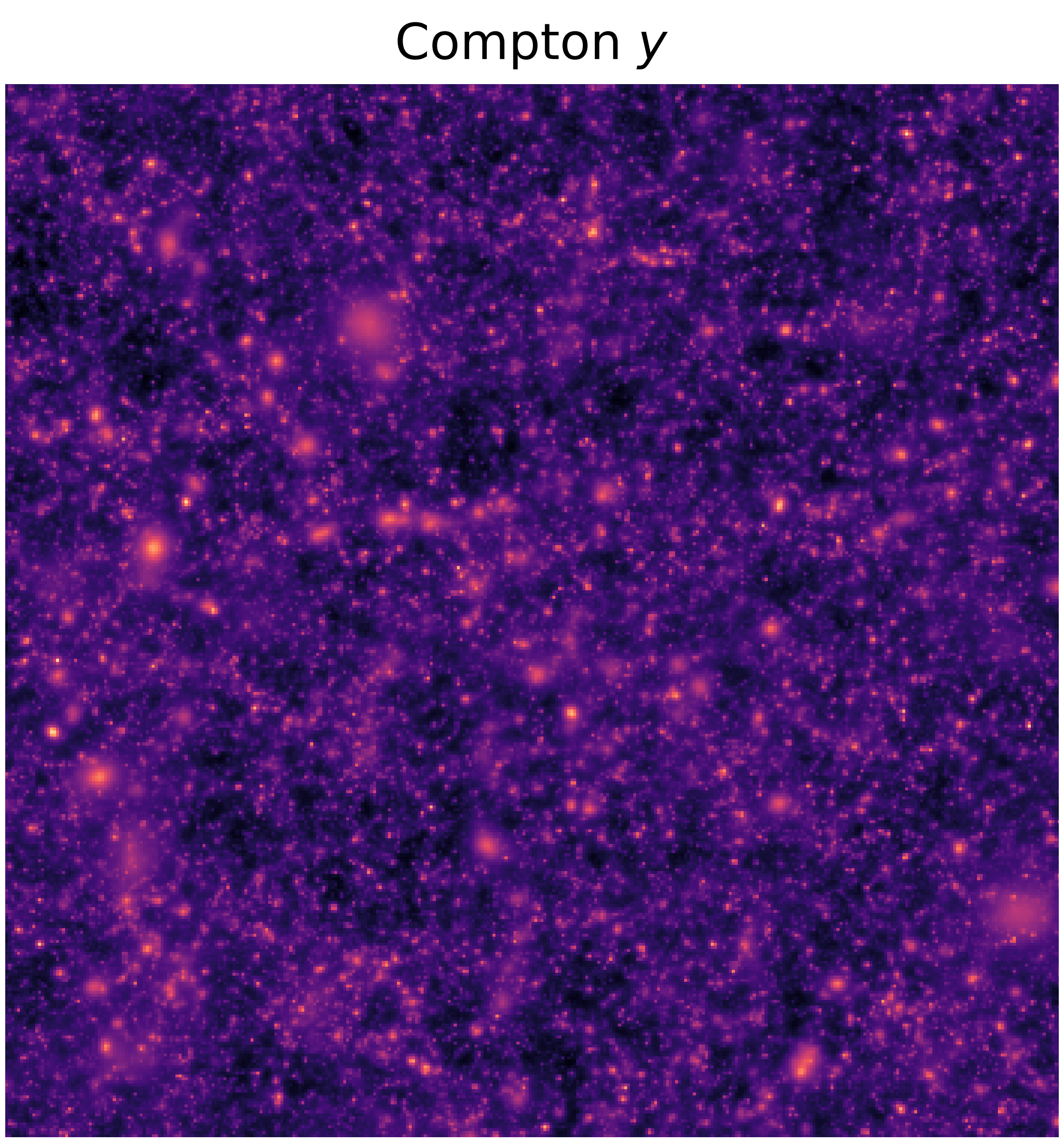
Cross-power spectra



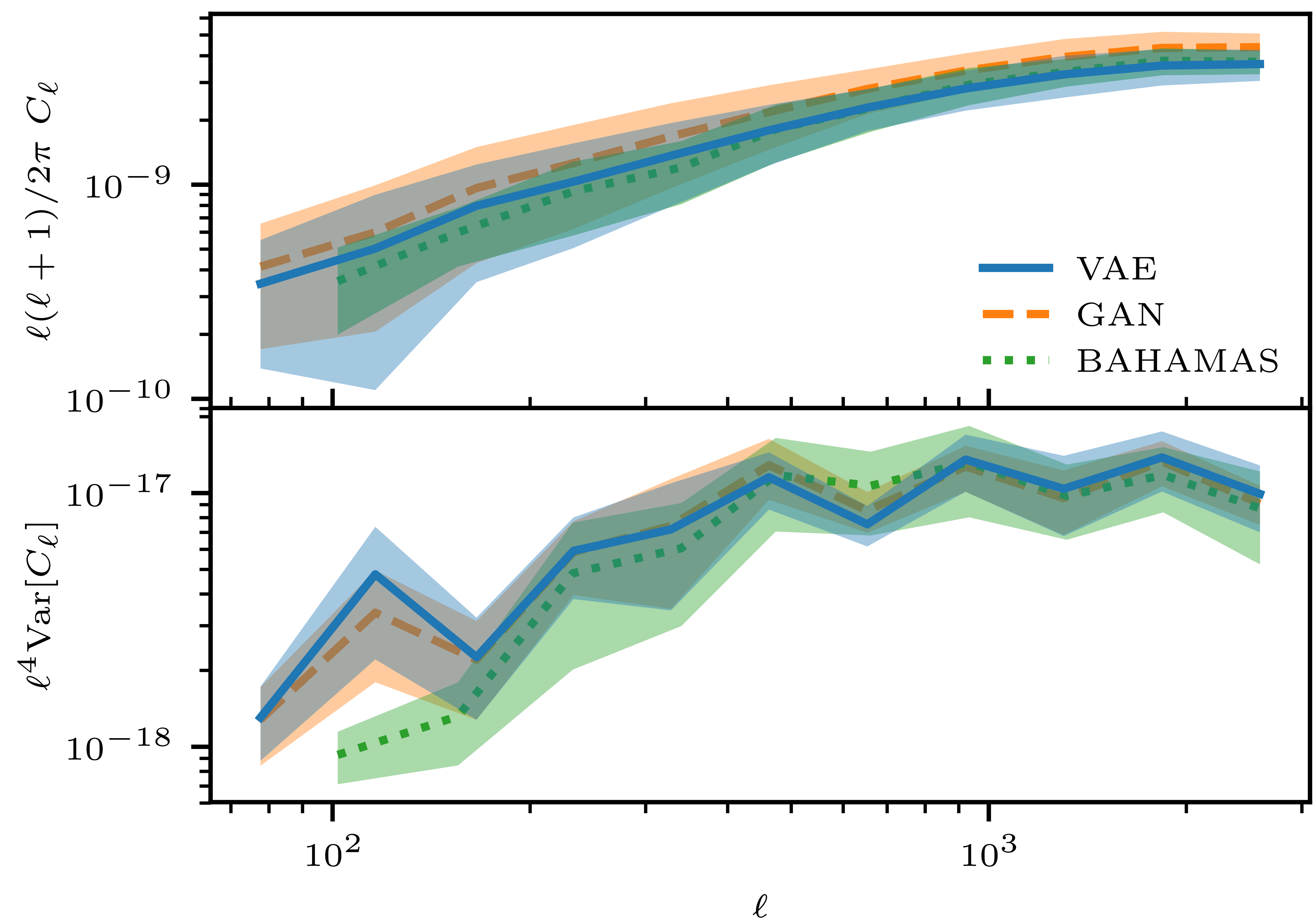
Tröster+2019

Convergence vs Compton-y





tSZ-shear cross spectra



Tröster+2019

Where to go from here

Physicality

- Use physical models where they exists; replace effective models and approximations

Exploit locality and symmetries

- Generating training data is expensive; increasing sample efficiency is key

Data representation

- Space is mostly empty. Grids are inefficient at representing cosmic fields; we need to move on from simple convolutional layers.

Thank you.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 797794