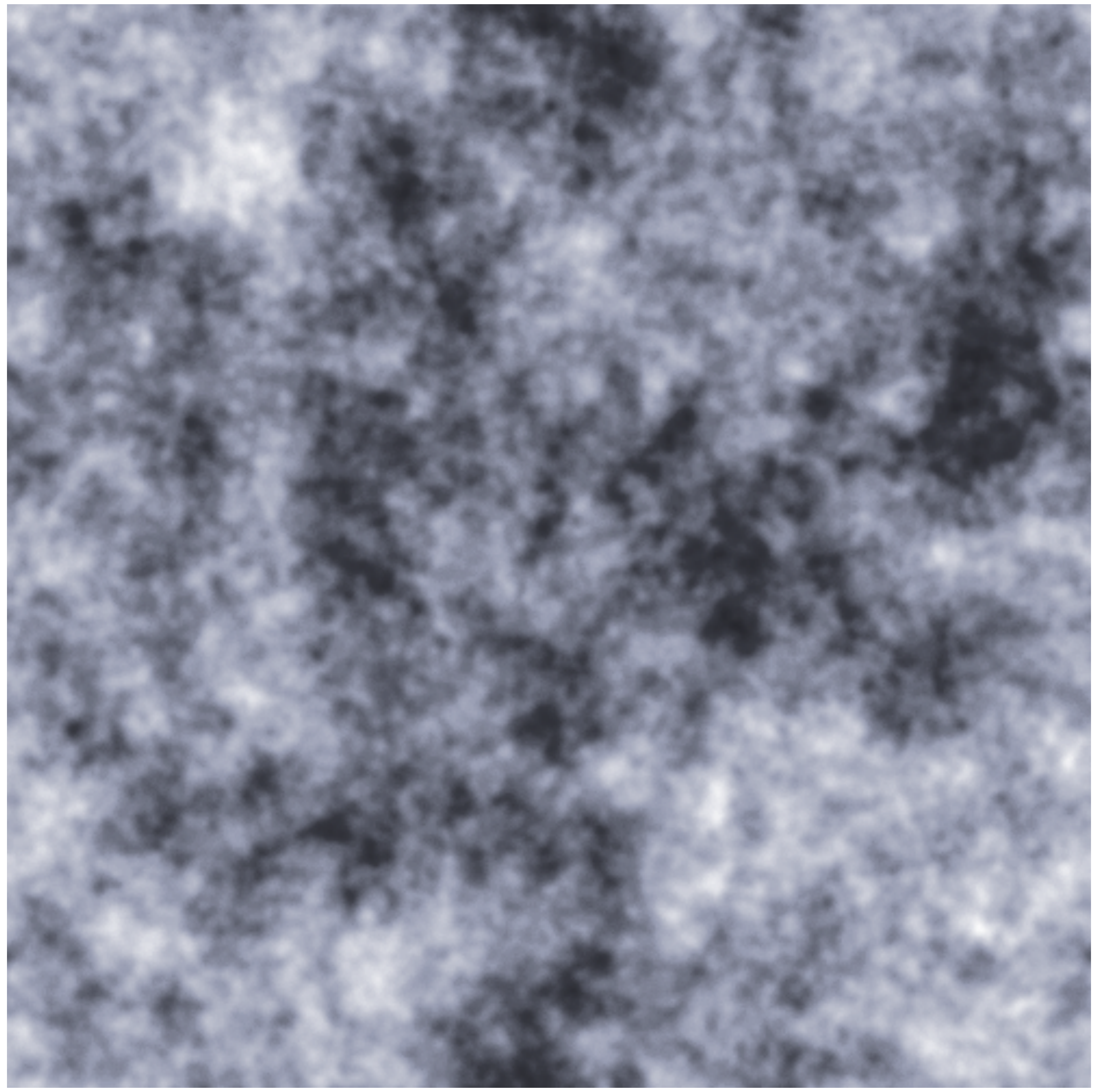
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Explainable deep learning models in Cosmology

Luisa Lucie-Smith
University of Hamburg

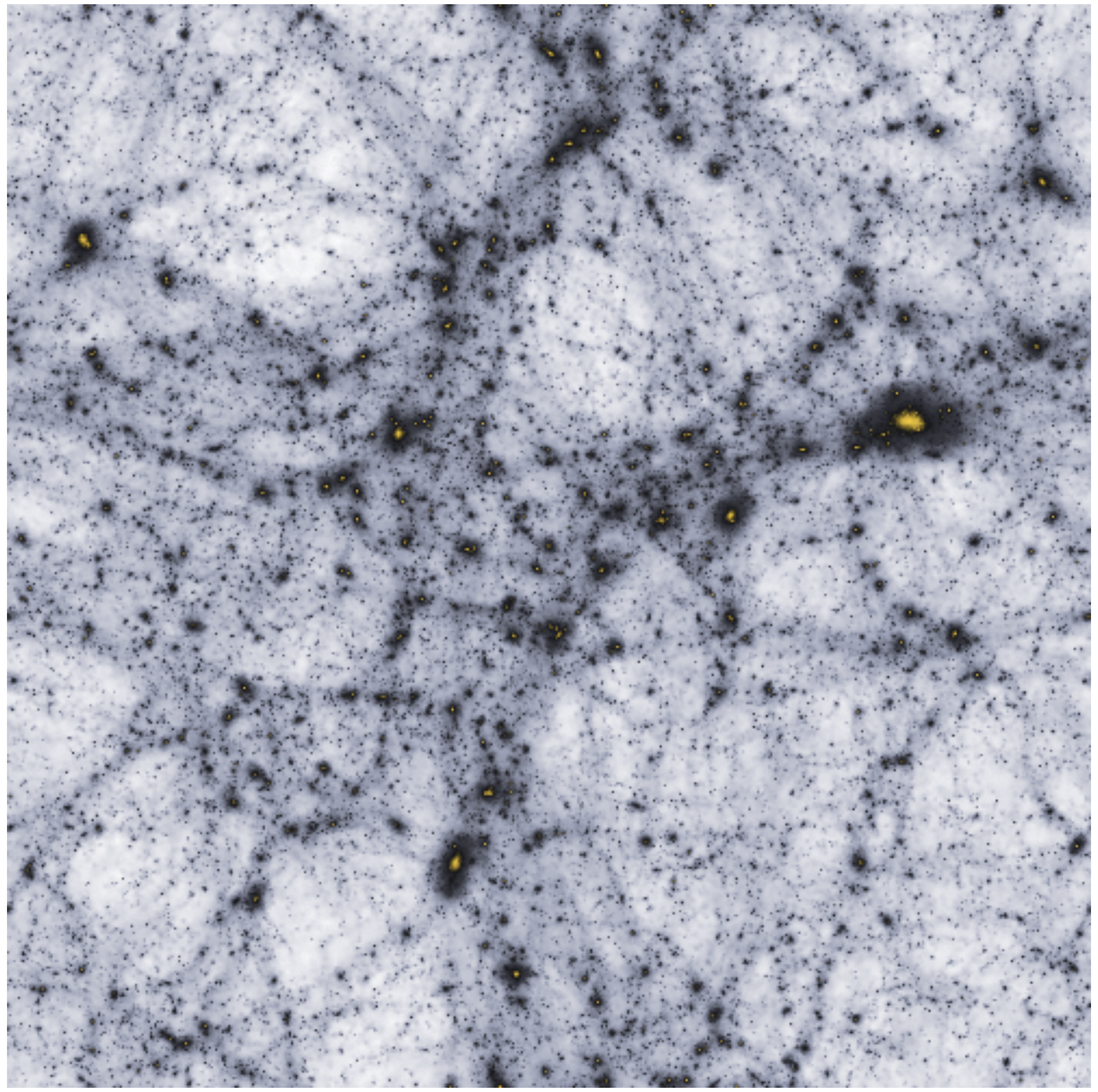
‘ML for Physics’ PUNCH4NFDI meeting,
Hamburg, 20th January 2025

Cosmological structure formation



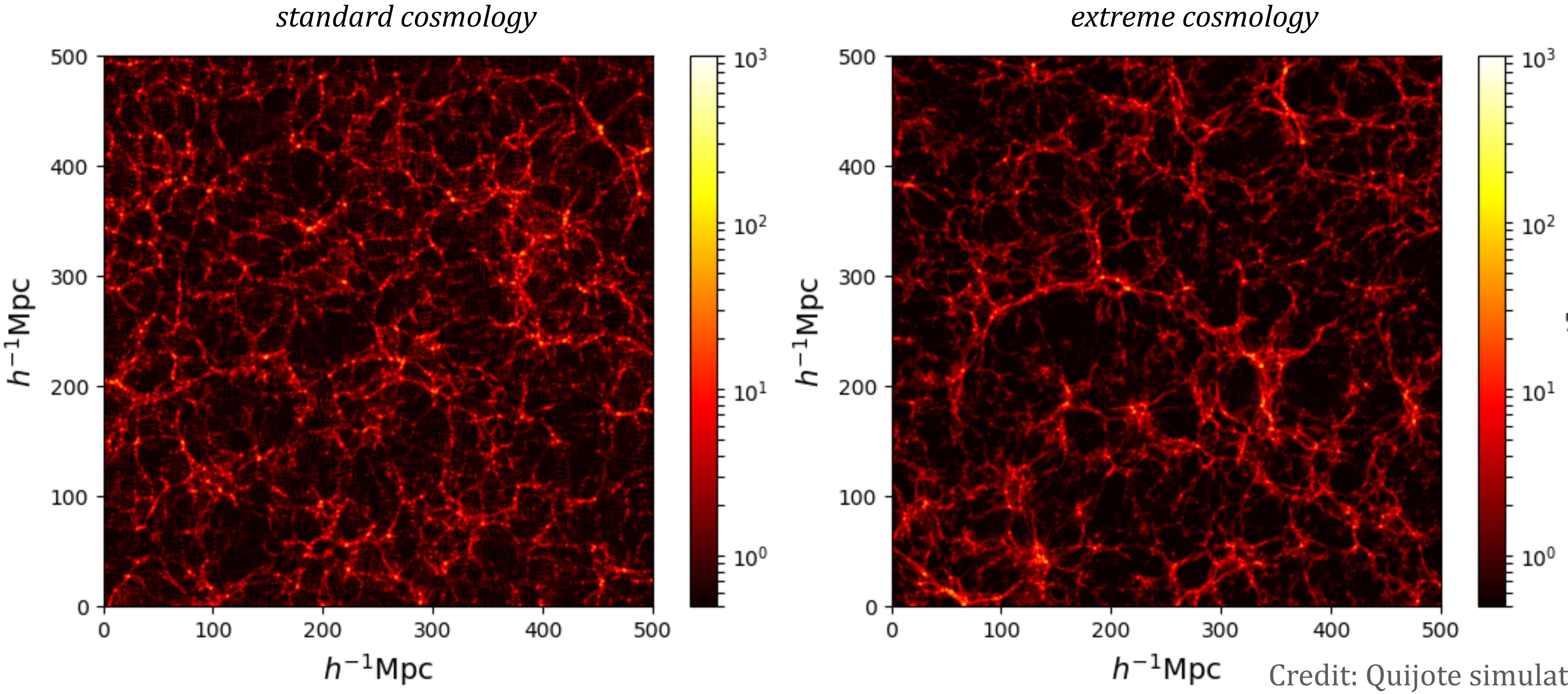
Initial density perturbations
(inferred from CMB constraints)

[time]



Late-time large-scale structure

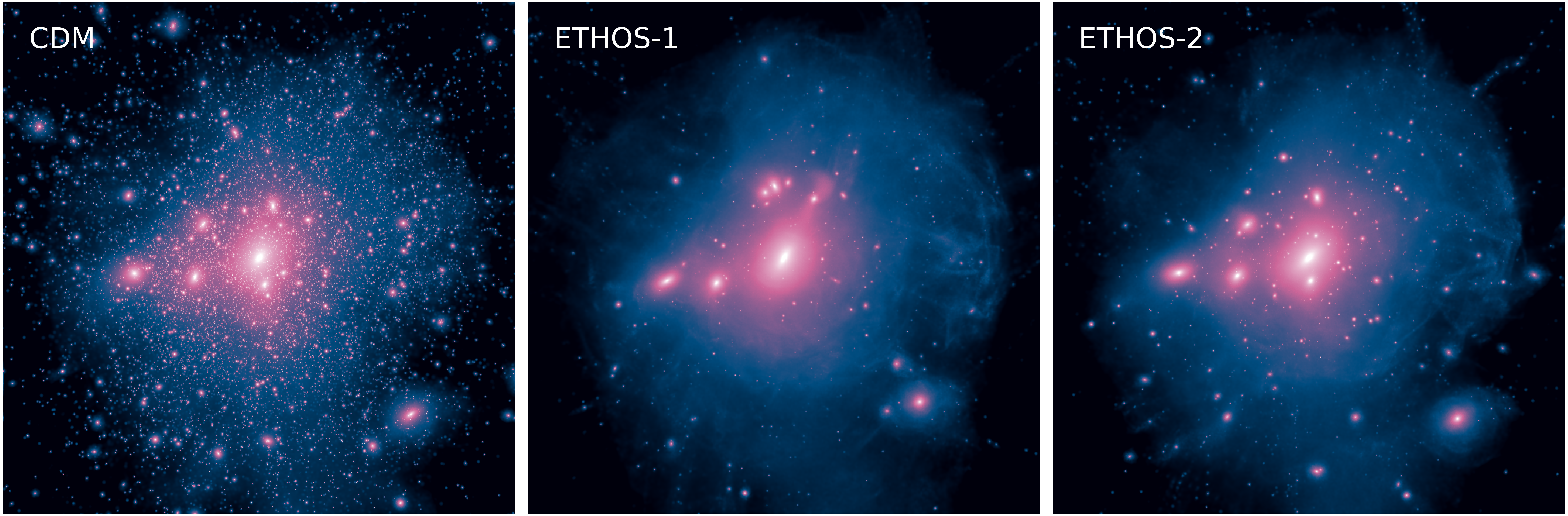
The large-scale structure contains signatures of fundamental physics



Credit: Quijote simulations

Dark matter halo structure contains signatures of nature of dark matter

Vogelsberger et al. (2016)



Characteristics of (good) models for large-scale structure:

- *Extract information from complex , non-linear data*
  —> cosmology from the large-scale structure on small, non-linear scales

- *Depend on minimal set of parameters*
  —> avoid large set of correlated parameters

- *Depend on physically interpretable & explainable parameters*
  —> set well-motivated priors & gain physical insights

- *Generalise beyond strict setting in which it was fitted*
  —> useful to explain a wide range of phenomena

Characteristics of (good) models for large-scale structure:

- *Extract information from complex , non-linear data*
  —> cosmology from the large-scale structure on small, non-linear scales

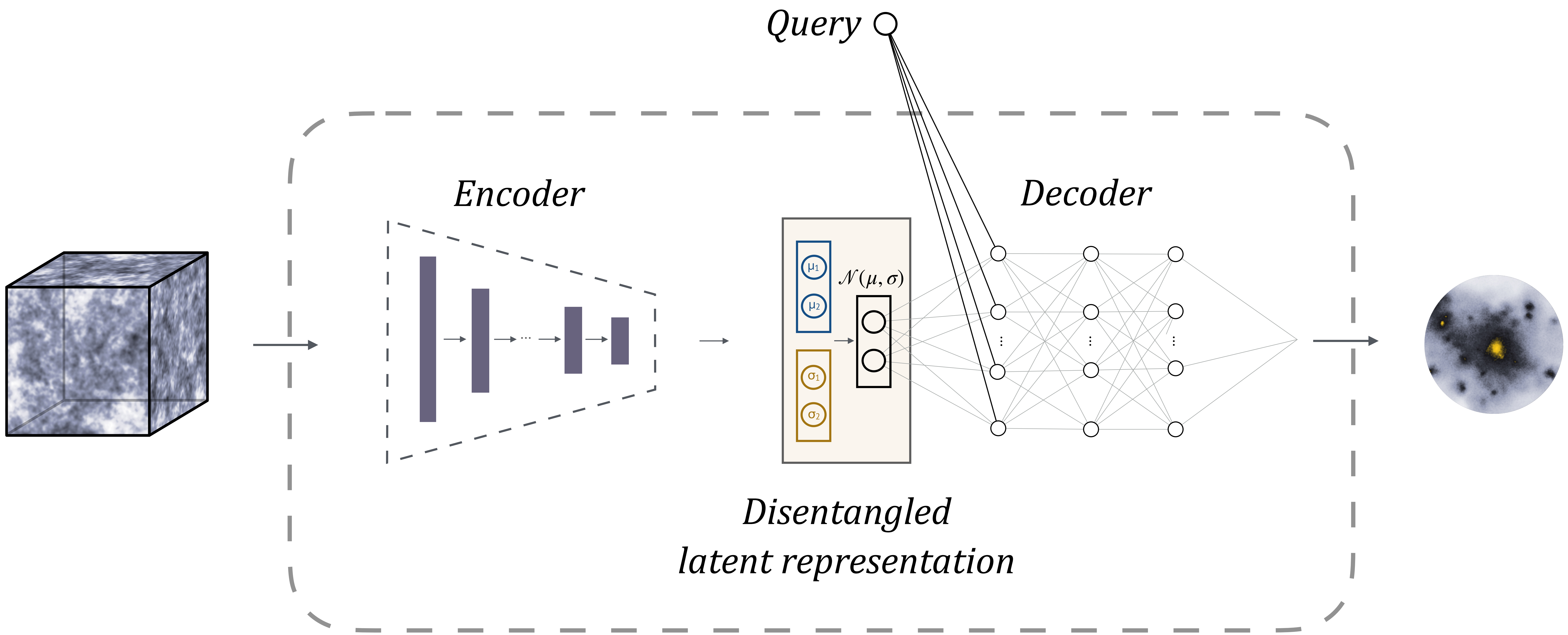
- *Depend on minimal set of parameters*
  —> avoid large set of correlated parameters

- *Depend on physically interpretable & explainable parameters*
  —> set well-motivated priors & gain physical insights

- *Generalise beyond strict setting in which it was fitted*
  —> useful to explain a wide range of phenomena

Can we use AI to build such models?

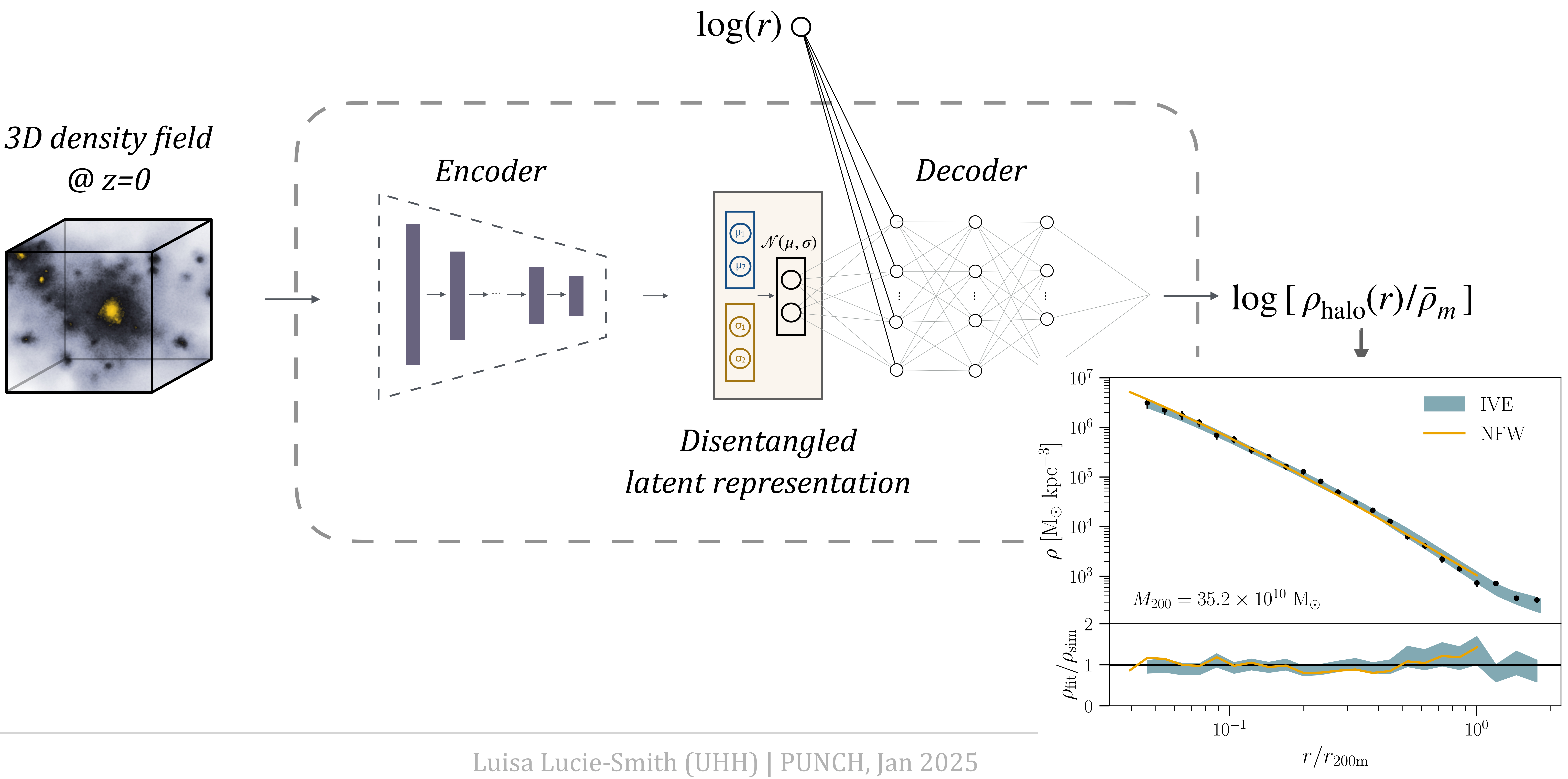
Interpretable Variational Encoder (IVE) for explainable AI



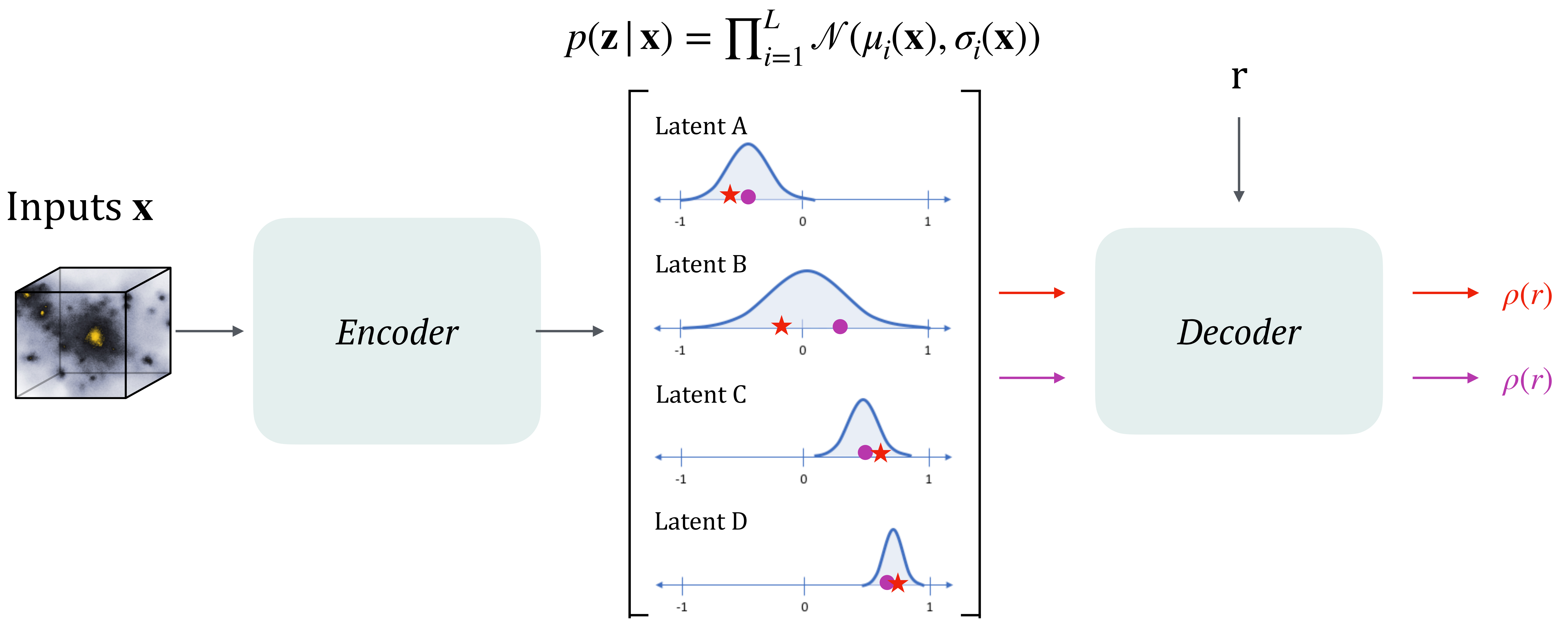
Model compression enables "explainability"

Iten et al. (PRL, 2020); Lucie-Smith et al. (PRD, 2022); Lucie-Smith et al. (PRL, 2024)

Modelling halo density profiles out to the halo outskirts



Desired latent representation properties for interpretability



- **Interpretability** can be achieved if latent space is **disentangled:** independent factors of variation in profiles captured by different, independent latents
- Disentanglement encouraged via **loss function** optimised during training

IVE loss function

β must be carefully fine-tuned
to balance accuracy with disentanglement

Predictive term *KL-divergence term*

$$\mathscr{L} = \mathscr{L}_{\text{pred}}(\rho_{\text{true}}, \rho_{\text{pred}}) + \beta\,\mathscr{D}_{\text{KL}}(p(\mathbf{z}\,|\,\mathbf{x}); q(\mathbf{z}))$$

(Higgins+, 2017)

MSE/Gaussian likelihood:

$$\mathscr{L}_{\text{pred}} = \frac{1}{N}\sum_{i=1}^{N}\left[\log_{10}\rho_{i,\text{true}} - \log_{10}\rho_{i,\text{pred}}\right]^2$$

How close are the predictions to the ground truths

Learnt latent distribution:

$$p(\mathbf{z}\,|\,\mathbf{x}) = \prod_{i=1}^{L}\mathscr{N}(\mu_i(\mathbf{x}), \sigma_i(\mathbf{x}))$$

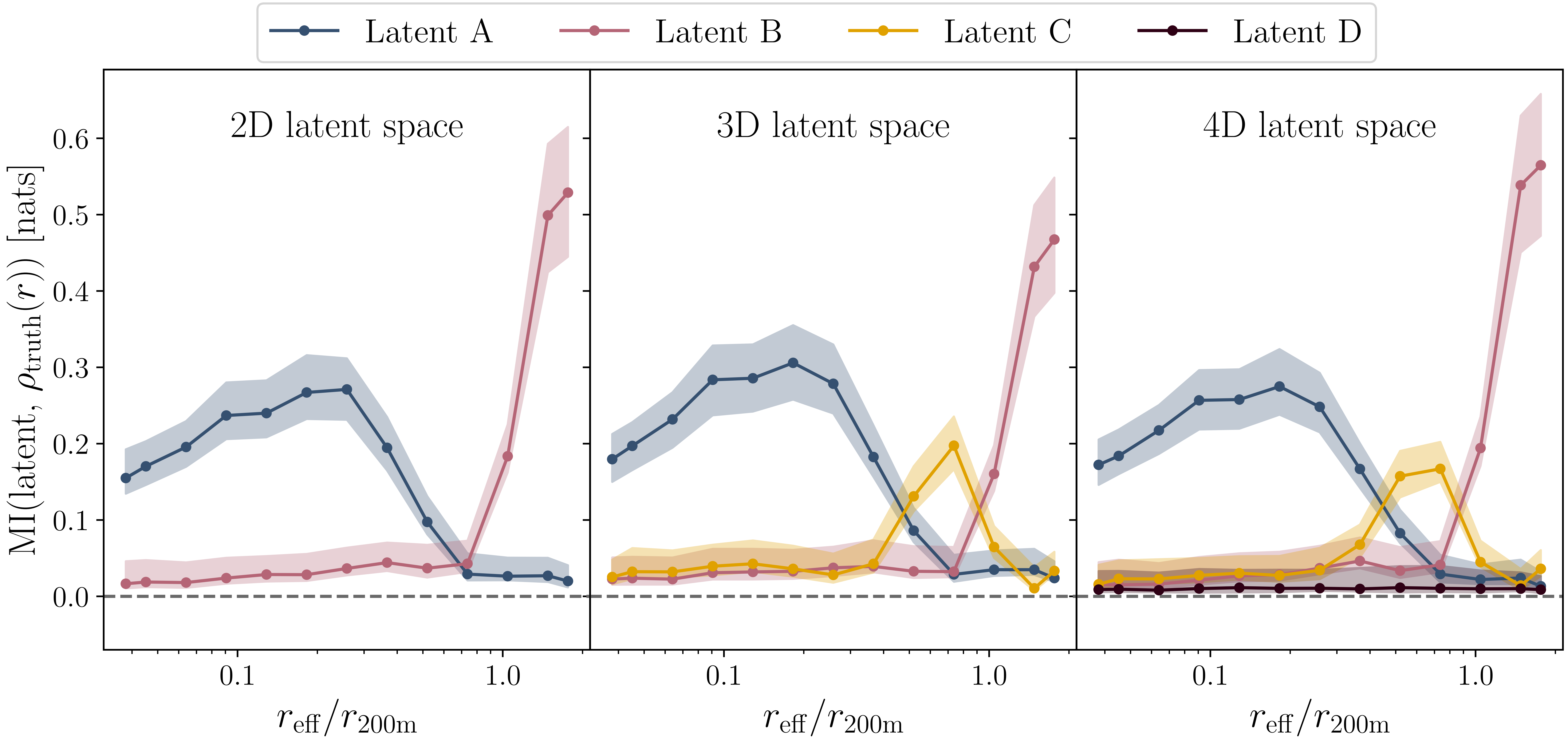
Prior:

$$q(\mathbf{z}) = \prod_{i=1}^{L}\mathscr{N}(0,1)$$

How close is the latent distribution to set of independent unit Gaussians

Interpreting the latent representation using mutual information

MI between latents & output



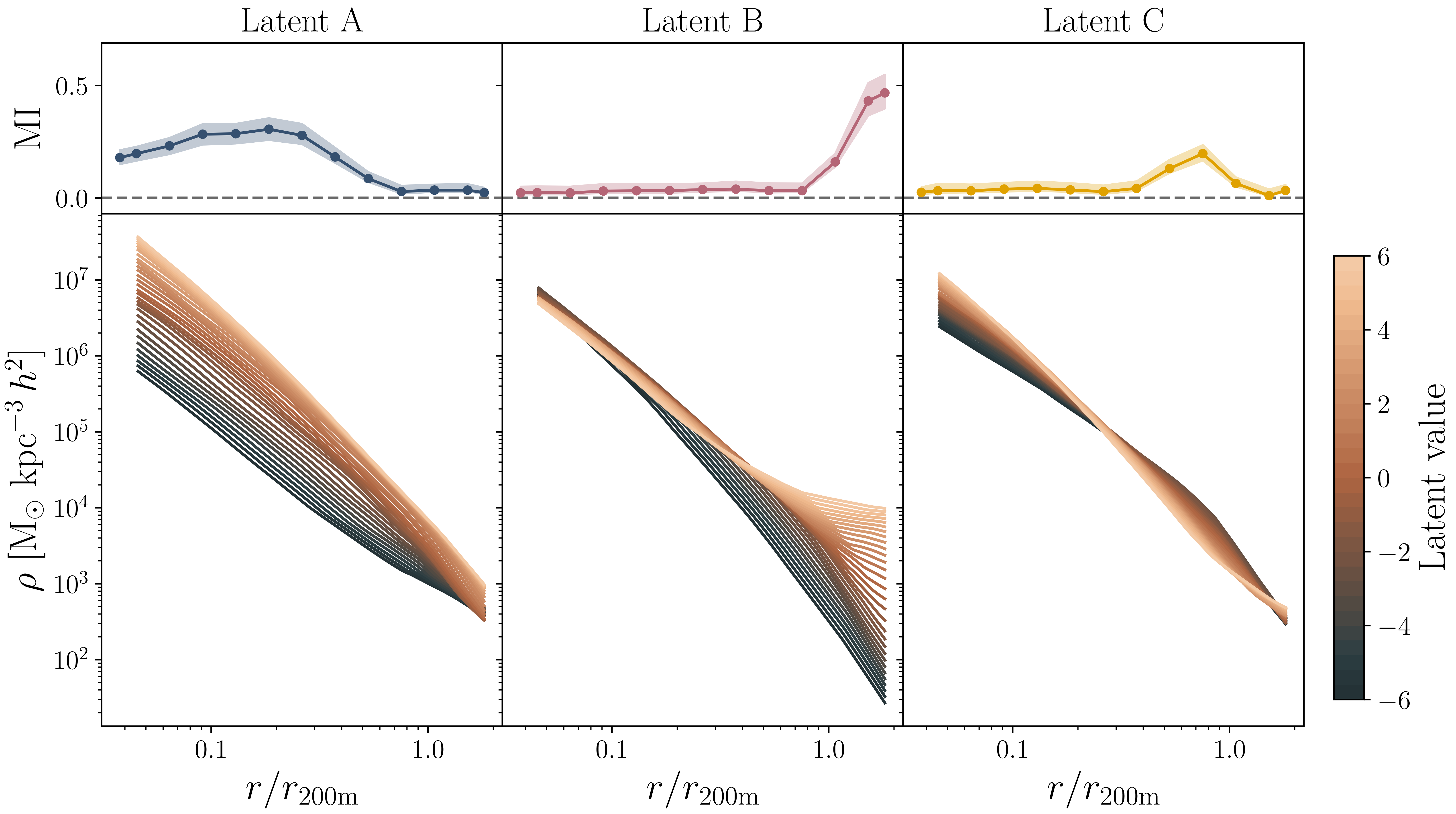
$$\mathrm{MI}(X,Y) = \int\int p(x,y)\,\log\left[\frac{p(x,y)}{p(x)p(y)}\right]\mathrm{d}x\mathrm{d}y$$

HTTPS://GITHUB.COM/DPIRAS/GMM-MI

Lucie-Smith et al. (PRD, 2022)

Piras, .., Lucie-Smith et al. (2023, MLST)

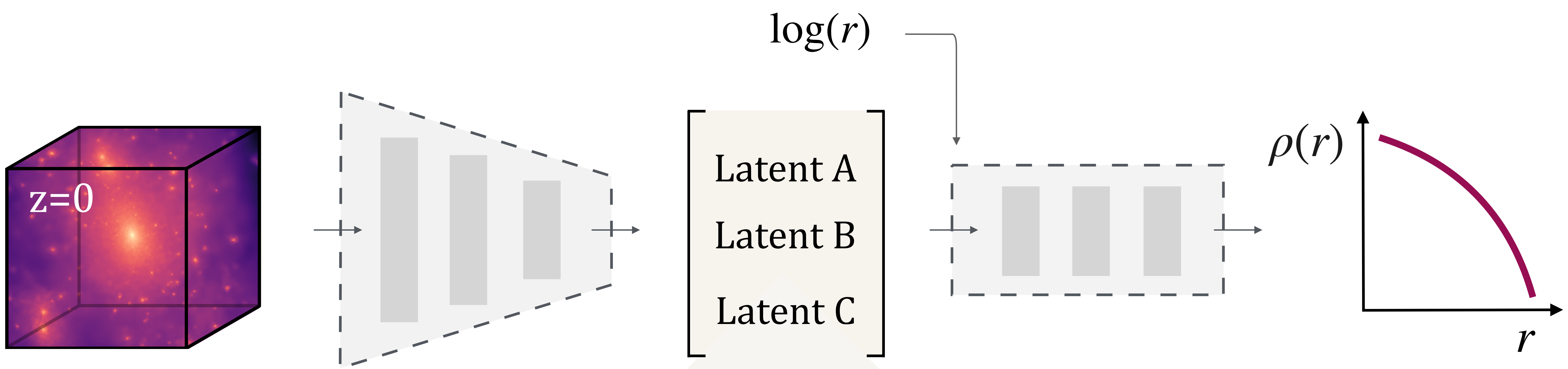
Systematically varying one latent at a time



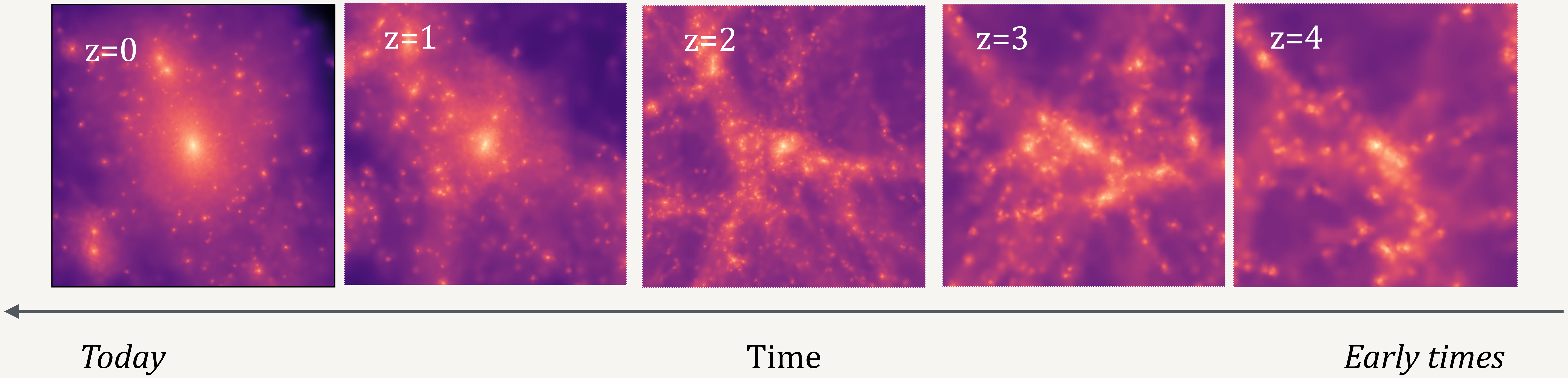
Latent A = ***normalisation***; Latent B = ***outer slope;*** Latent C = ***inner slope***

Lucie-Smith et al. (PRD, 2022)

Generalization: exploiting latent space beyond its original training task

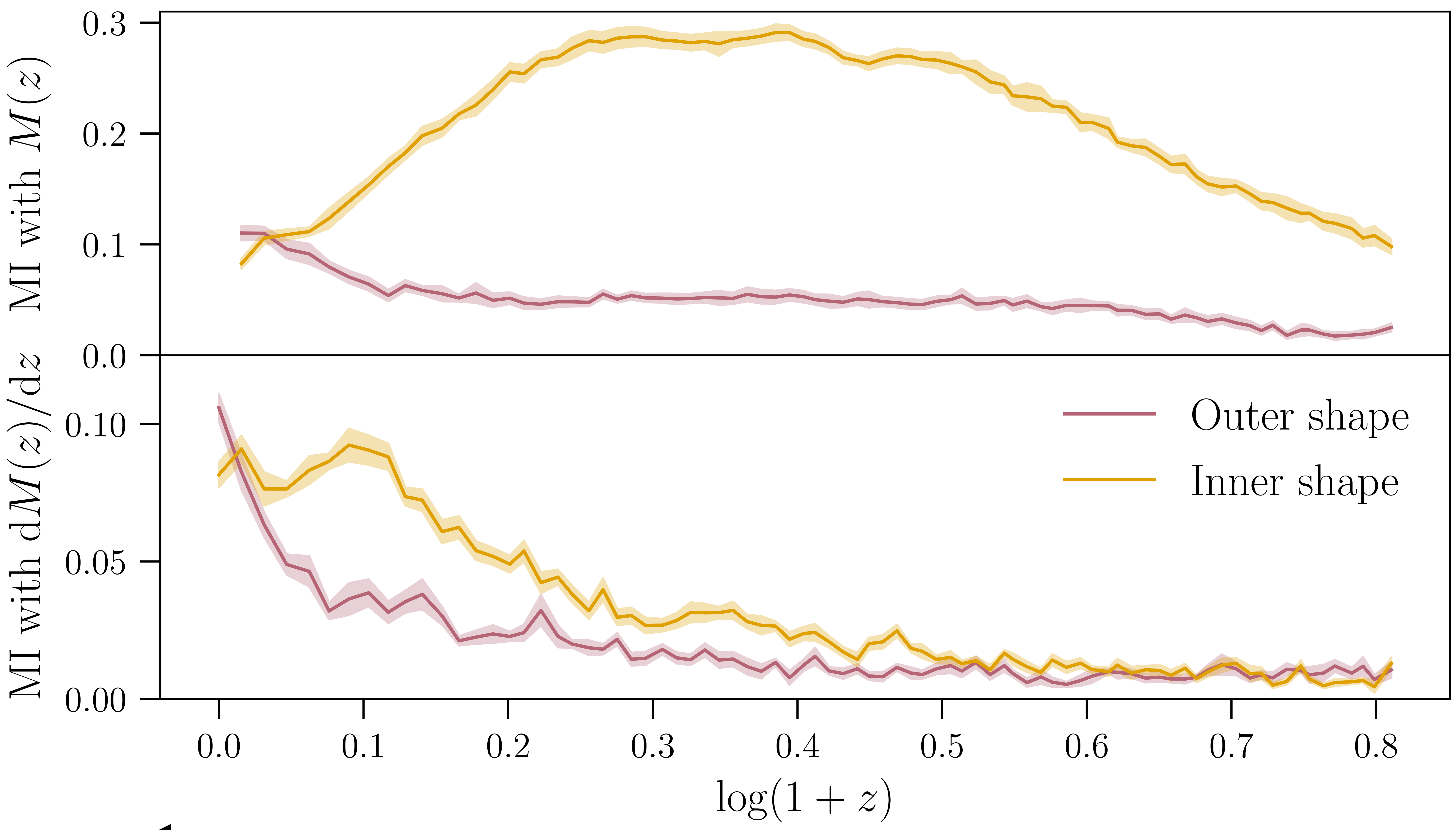


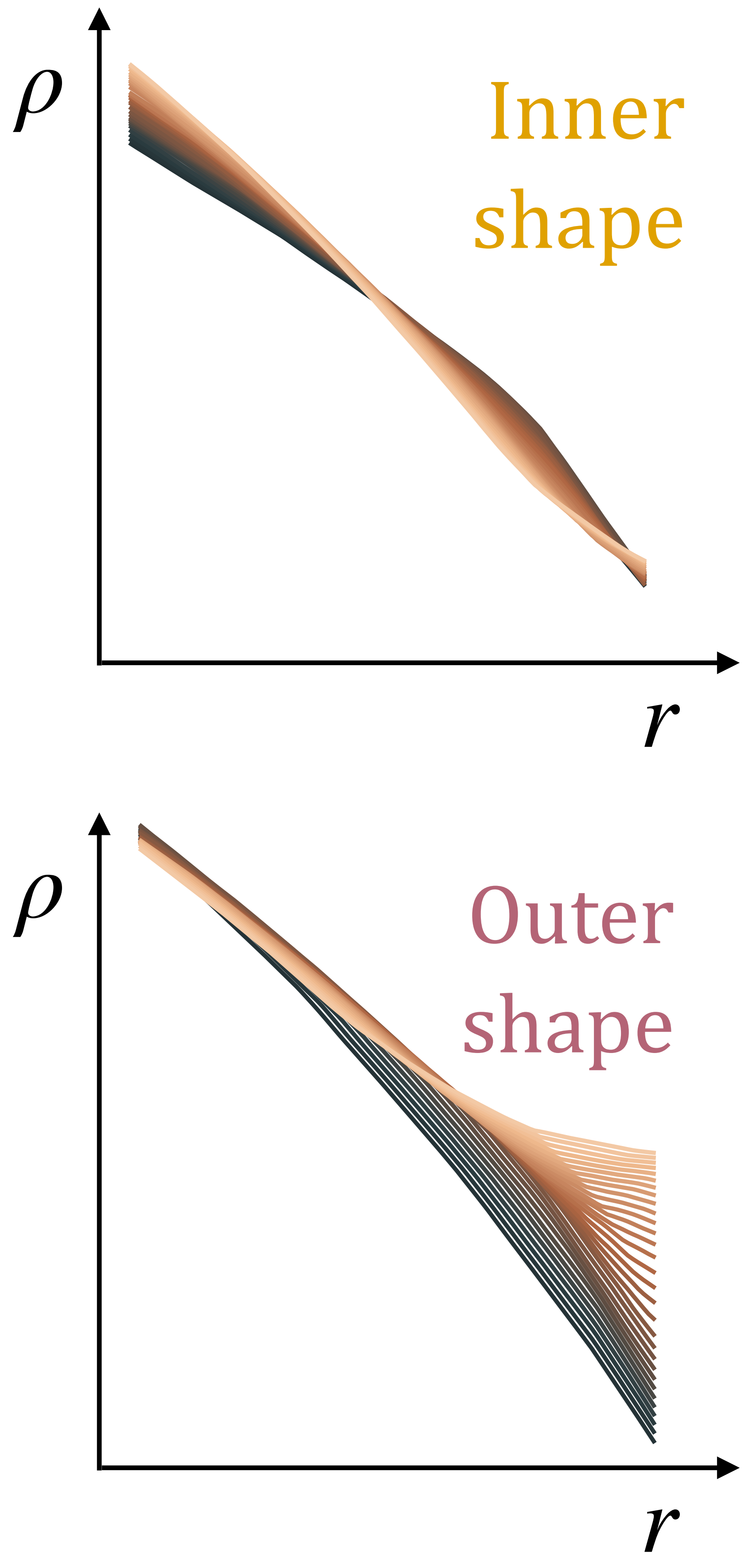
Does the latent space contain information about the origin of the halo density structure?



Lucie-Smith, Peiris, Pontzen (PRL, 2024)

Connection between the latents and the halo evolution history





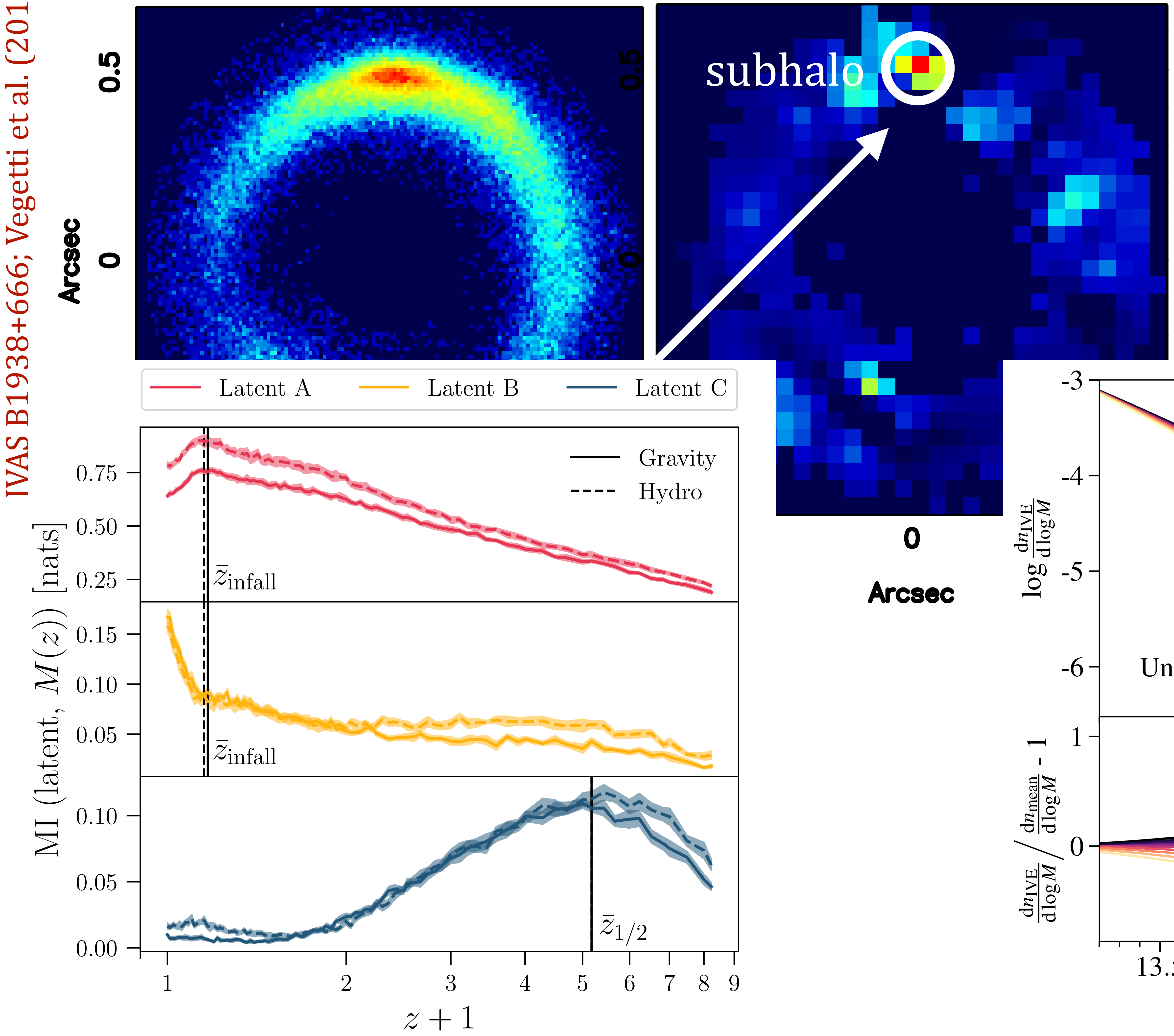
today Time *early times*

Lucie-Smith, Peiris, Pontzen (PRL, 2024)

Appliations to a variety of cosmological probes

Subhalo profile for strong lensing observations

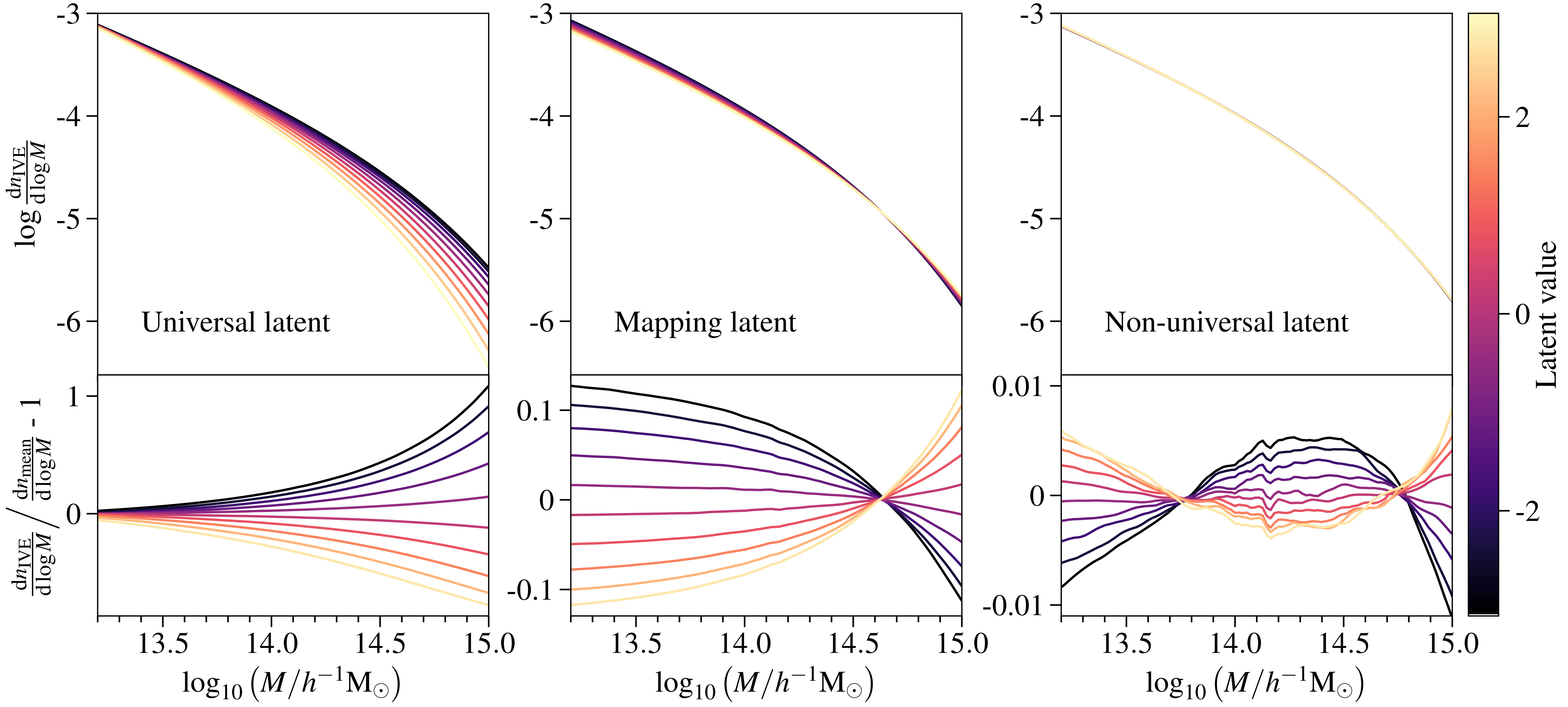
IVAS B1938+666; Vegetti et al. (2012)



Lucie-Smith, Despali, Springel (MNRAS, 2024)

Guo, Lucie-Smith+(MNRAS, 2024)

Non-universality of the halo mass function

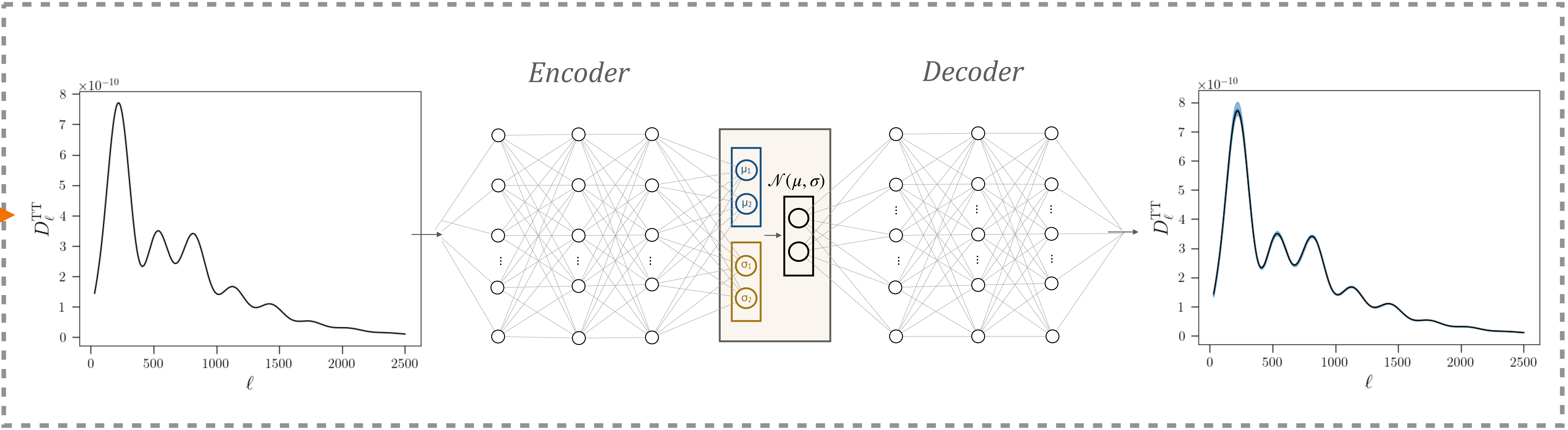


Can neural-based model compression be used to discover & constrain most important degrees of freedom of data?

What information is encoded in CMB for cosmologies beyond ΛCDM?

'H0 tension' may be resolved if there exists a component of "early dark energy" (EDE)

5 ΛCDM + 3EDE params



CMB temperature power spetrum

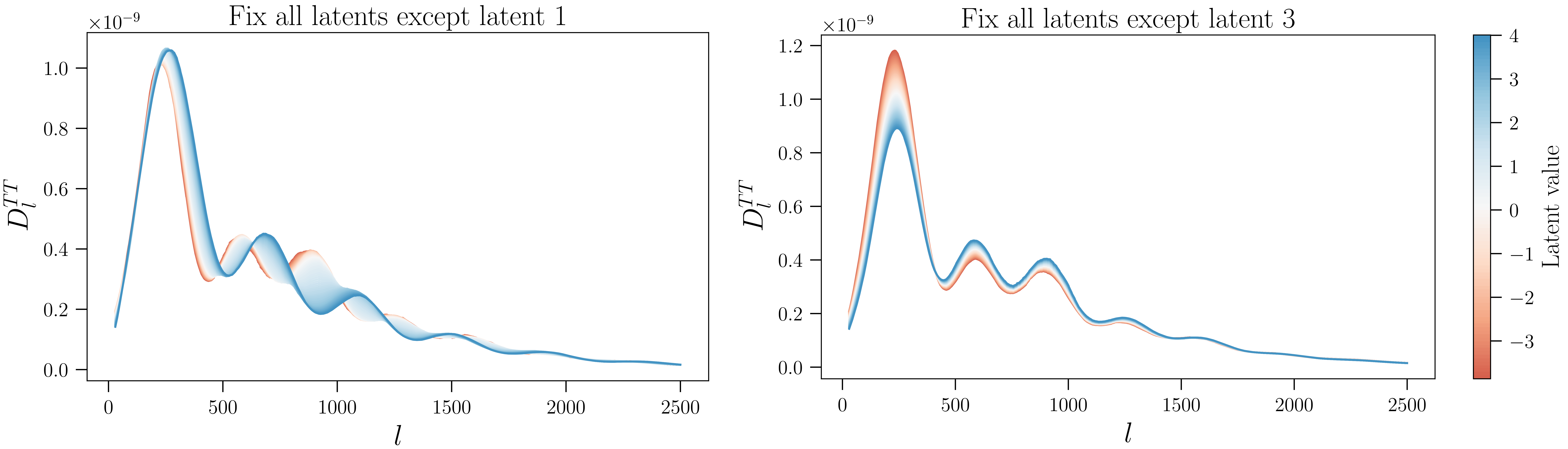
Latent representation

Reconstructed CMB temperature power spetrum

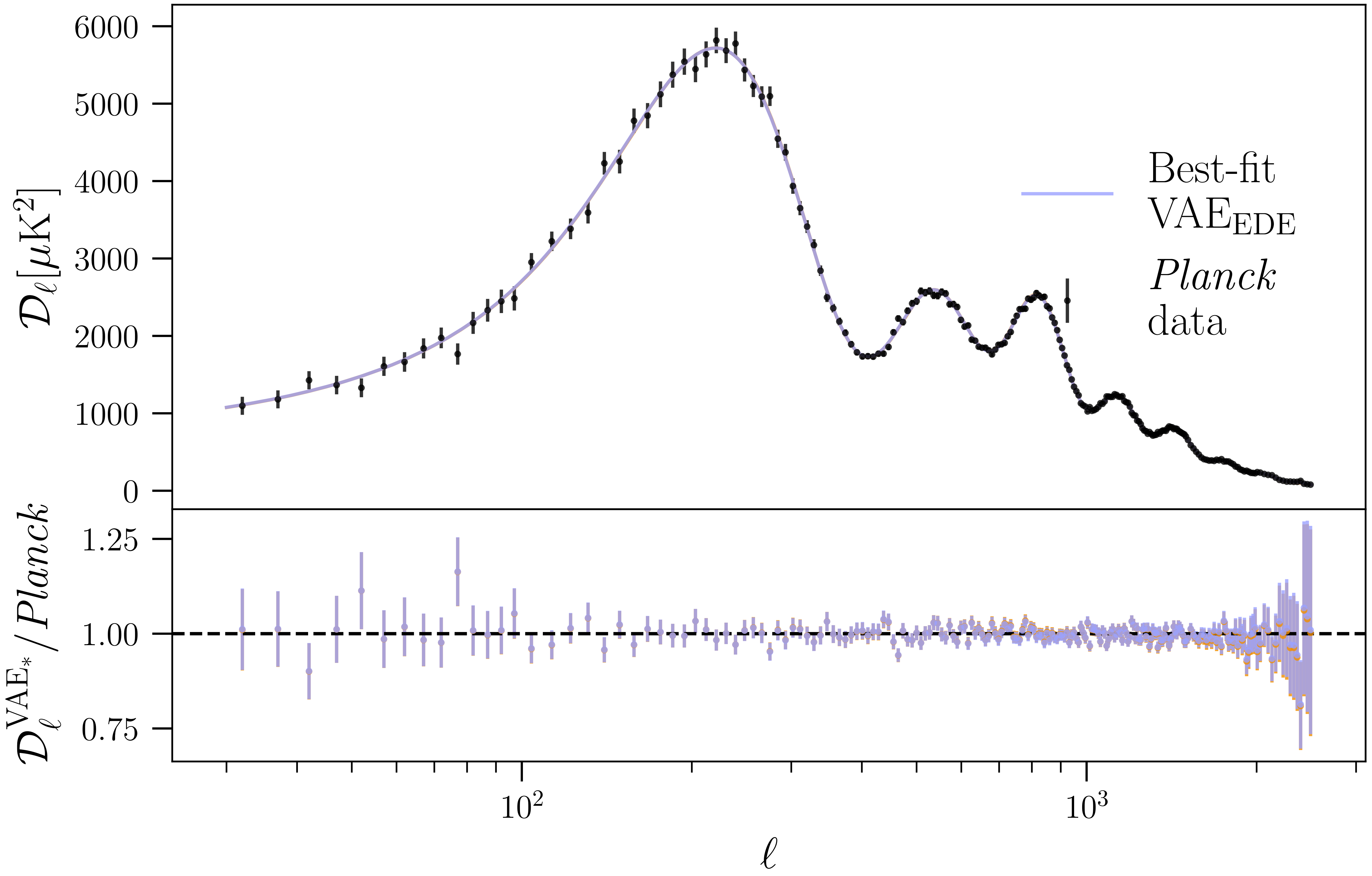
Piras, Lucie-Smith, Herold, Komatsu (in prep)

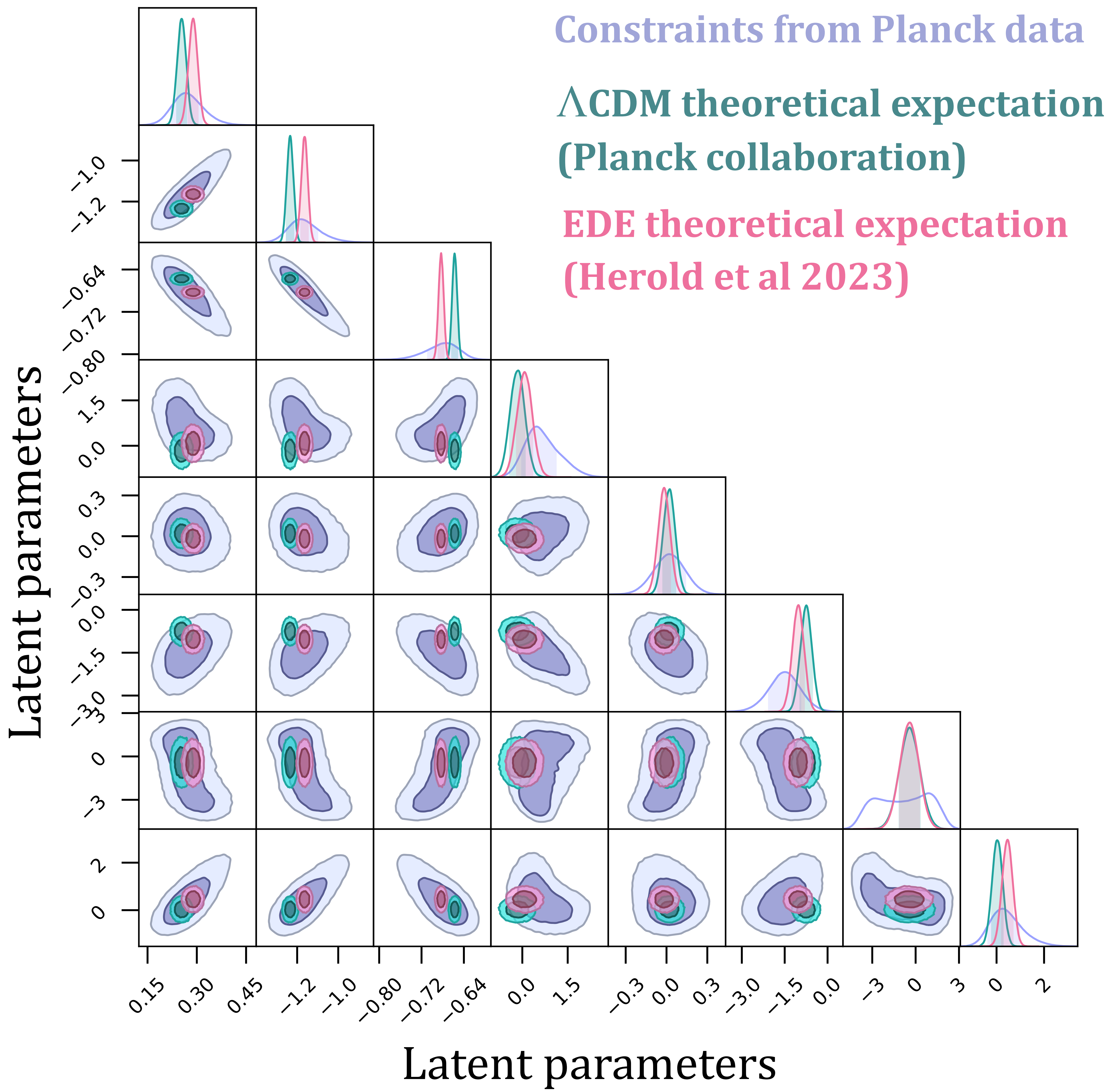
Independent dof in CMB assuming EDE Universe

8 independent degrees of freedom (of which 4 explain most of the variations in the CMB)



Cosmology in latent space: constraints from Planck data





Constraints from Planck data

ΛCDM theoretical expectation (Planck collaboration)

EDE theoretical expectation (Herold et al 2023)

Conclusions

- Interpretable variational encoders (IVE) provide new avenue to provide robust, physically interpretable models of cosmic structures
- IVE disentangles different physical effects in minimal set of ingredients & generalizes beyond its original training task
- Applications to density profiles, halo mass function and cosmological data vectors

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Luisa Lucie-Smith, luisa.lucie-smith@uni-hamburg.de