Reconstructing the cosmological distribution of dark matter from HI

Robert Feldmann & Mauro Bernardini
Department of Astrophysics
University of Zurich (UZH)

in collaboration with Daniel Anglés-Alcázar (CCA), Mike Boylan-Kolchin (UT Austin), James S. Bullock (USC), Claude-André Faucher-Giguère (Northwestern), Philipp Denzel (ZHAW), Jindra Gensior (Edinburgh), Joachim Stadel (UZH), Lucio Mayer (UZH), and the FIRE collaboration

Universität Zürich[UZH]

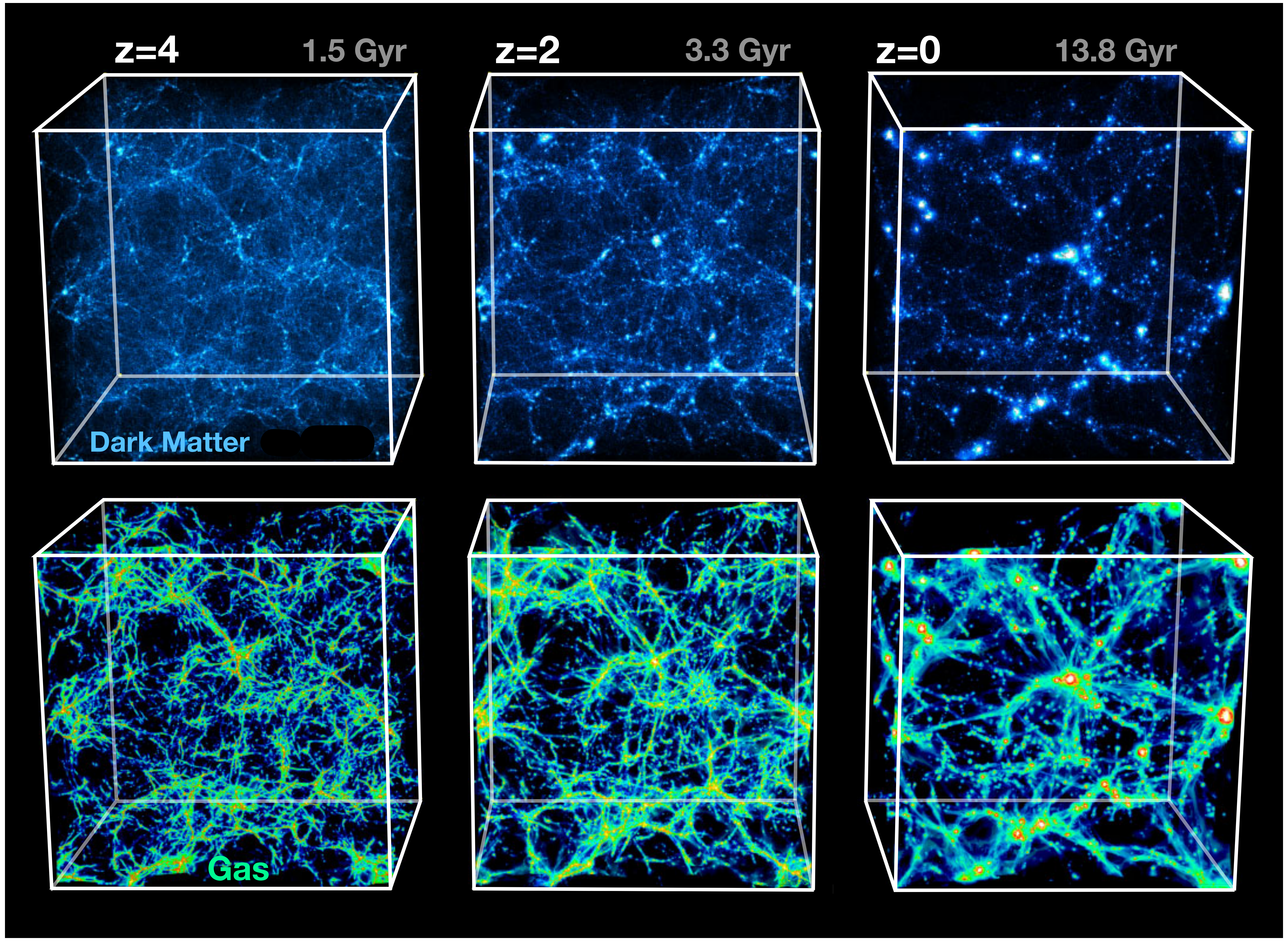
Swiss National Science Foundation

SKACH Winter Meeting 2026
Lugano, Switzerland – January 26–27, 2026

References:
Bernardini et al. 2022, arXiv:2110.11970
Feldmann et al. 2023, arXiv:2205.15325
Bernardini et al. 2025, arXiv:2502.15875
Bernardini et al. subm., arXiv:2507.05339

The Dark Matter Inference Problem

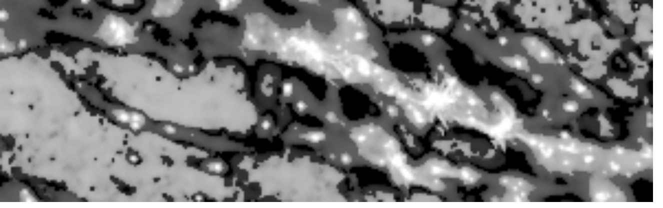
FIREbox (RF et al. 2023, arXiv:2205.15325)



- Growth of structure in the Universe mainly driven by dark matter (DM)
- DM is fundamental but invisible

- Can learn about DM from baryonic tracers: Galaxies, stars, cosmic gas
- Challenge: Tracers are biased and subject to complex physical processes

<u>How well can we infer the distribution of dark matter from baryonic tracers (especially from HI)?</u>



Why atomic hydrogen is special

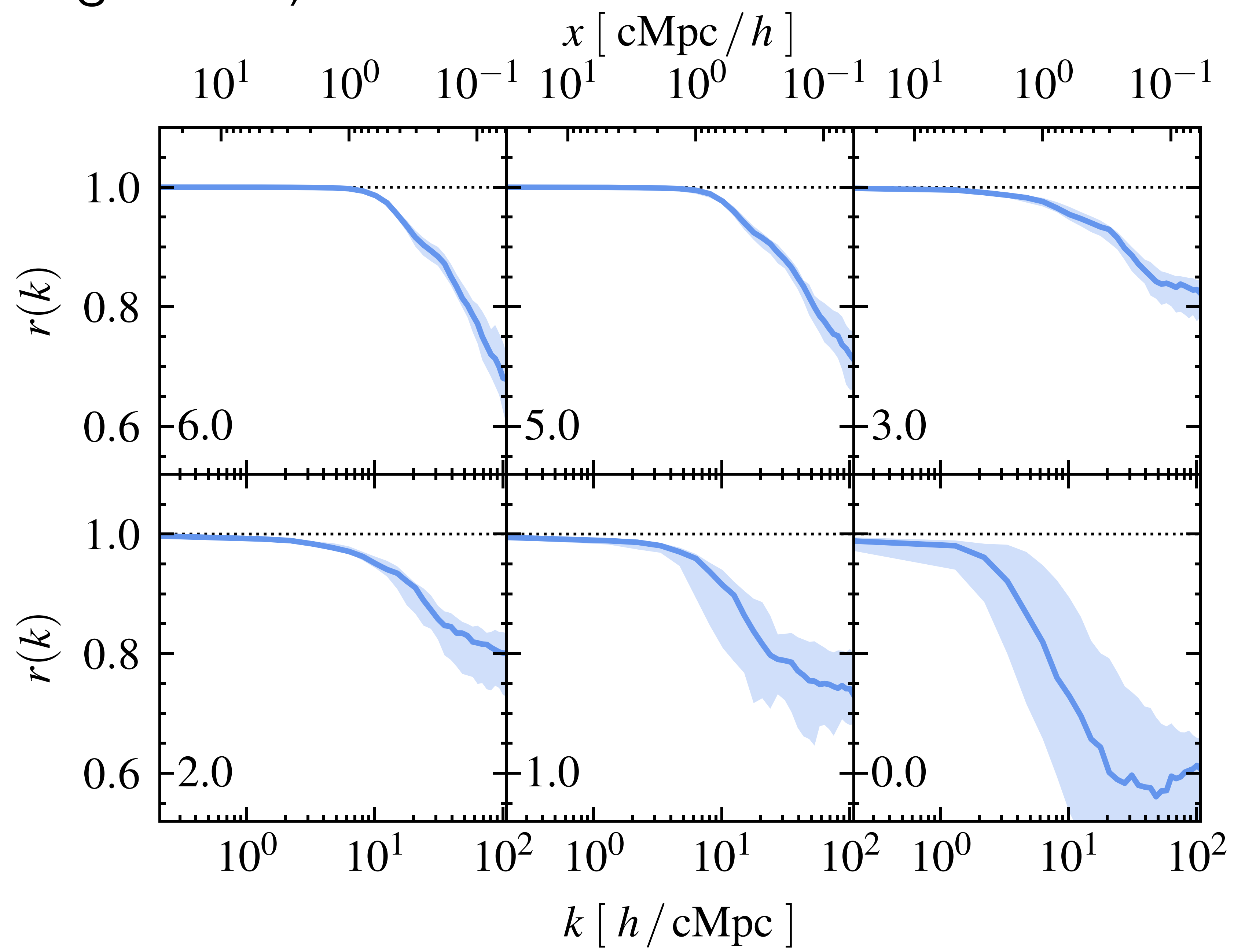
Atomic hydrogen (HI)

- traces gas in various environments: in/around galaxies but also in/between halos
- 21 cm line encodes redshift and dynamical information (radial velocity)
- observational target of SKA and other radio surveys (intensity mapping, resolved galaxies)

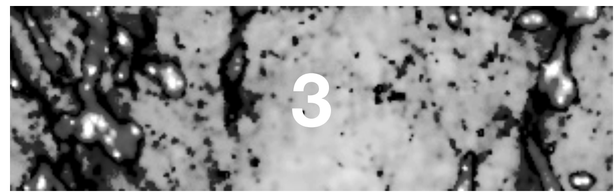
Link between HI – DM is non-linear and environment-dependent

- E.g., reduced HI – DM cross-correlation
- Baryonic Processes: Gas cooling, feedback, ram-pressure stripping, etc
- Challenges increase with decreasing spatial scale

➜ Accurate mapping requires spatially aware modeling



Cross-correlation between DM and total gas density at z=0-6 (Bernardini, RF, et al. 2025)



HI observations and their modeling

Standard approach is to forward-model from the underlying DM to HI

- Halo-based models: HI – halo mass relations (+ HI radial profiles)
- Hydrodynamical & semi-analytic models: principled but expensive, model-degeneracy
- Perturbative approaches: bias-expansion, valid on large & mildly non-linear scales
- Machine Learning methods

Limitations of traditional approaches

- Low-dimensional summary statistics of the HI – DM connection
- Information of spatial & kinematic structure reduced or lost

➜ Modeling on the field level

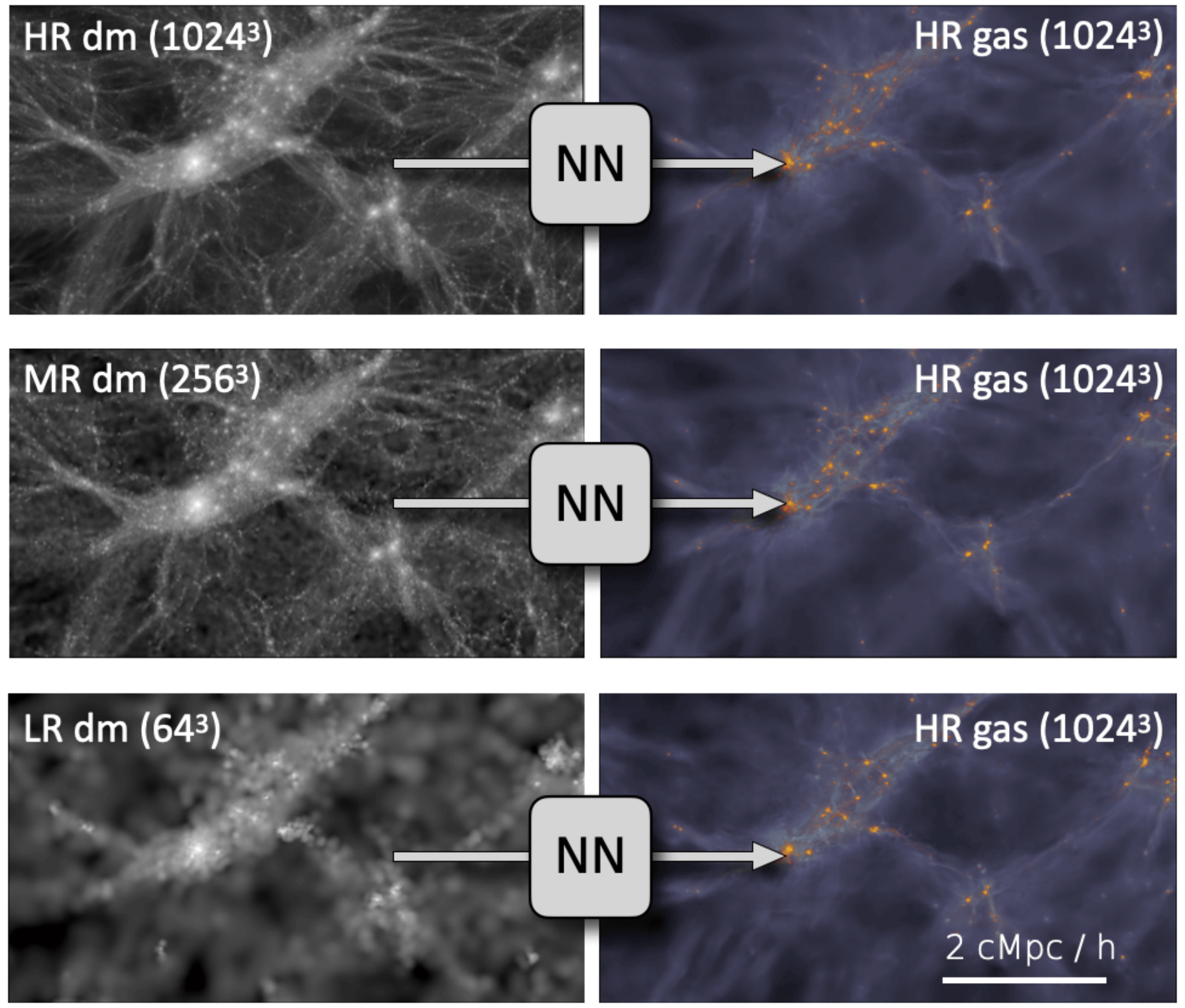
- Minimize information loss: use all spatial and kinematic information simultaneously
- Deep Learning well suited

EMBER

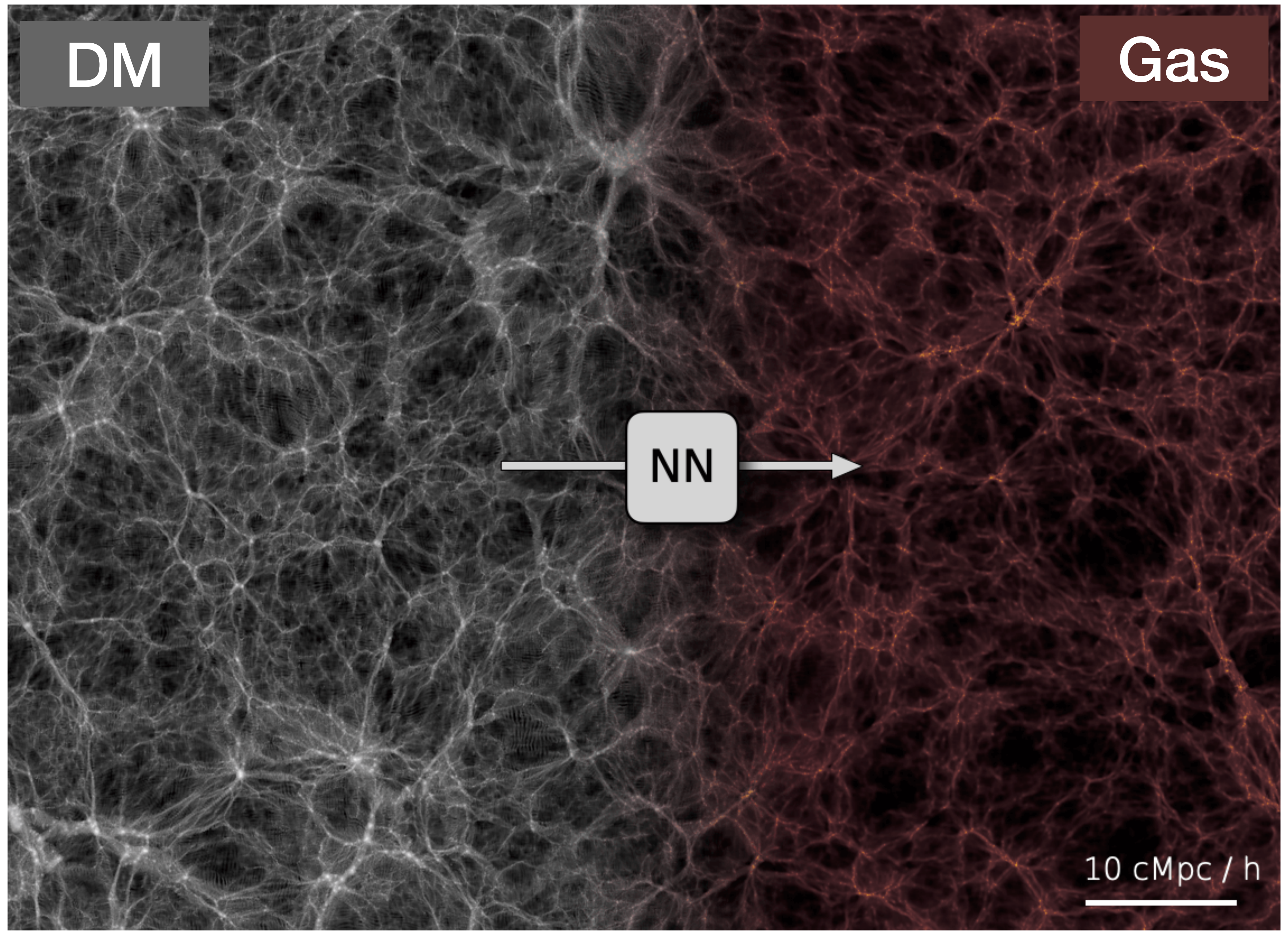
EMulating **B**aryonic **E**n**R**ichment Network (Bernardini, RF, et al. 2022)

- Deep learning framework to sample “realistic” baryon maps for given DM maps (2d) $p(\Sigma_{\rm gas}\,|\,\Sigma_{\rm DM})$
- Fully Convolutional Neural Network (CNN) + conditional Generative Adversarial Network (cGAN)
- Generator with U-Net (encoder/decoder) architecture, Wasserstein metric as adversarial loss
- Trained on cosmological hydrodynamical simulations from the FIRE project (FIREbox, MassiveFIRE)

Training
on small boxes + zoom-ins

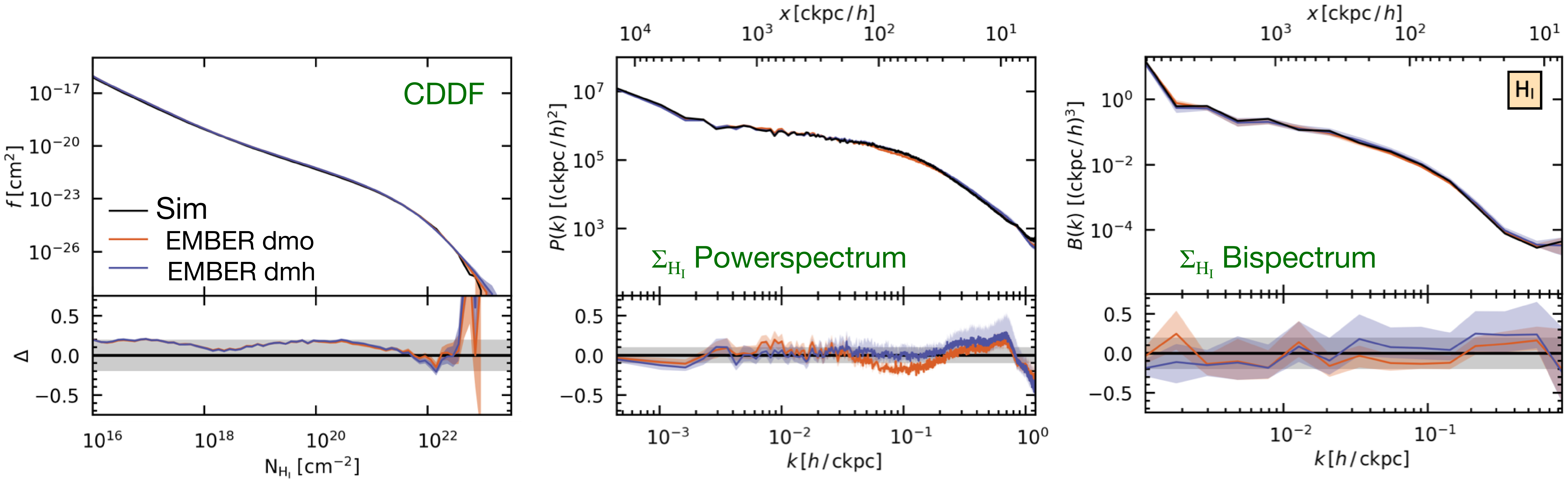


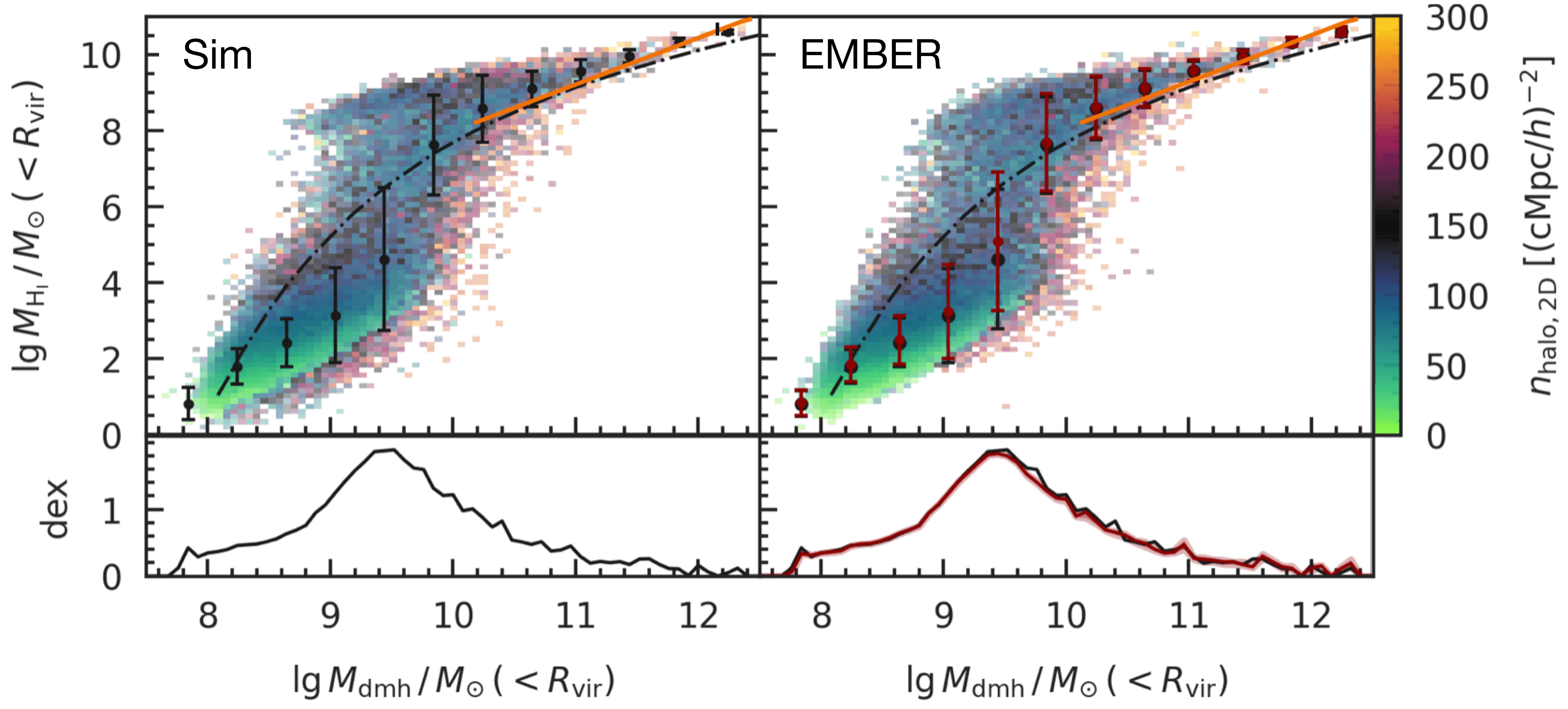
Application
for enriching large dark matter simulations with baryons



Bernardini, RF, et al. 2022, arXiv:2110.11970

How well does it work?





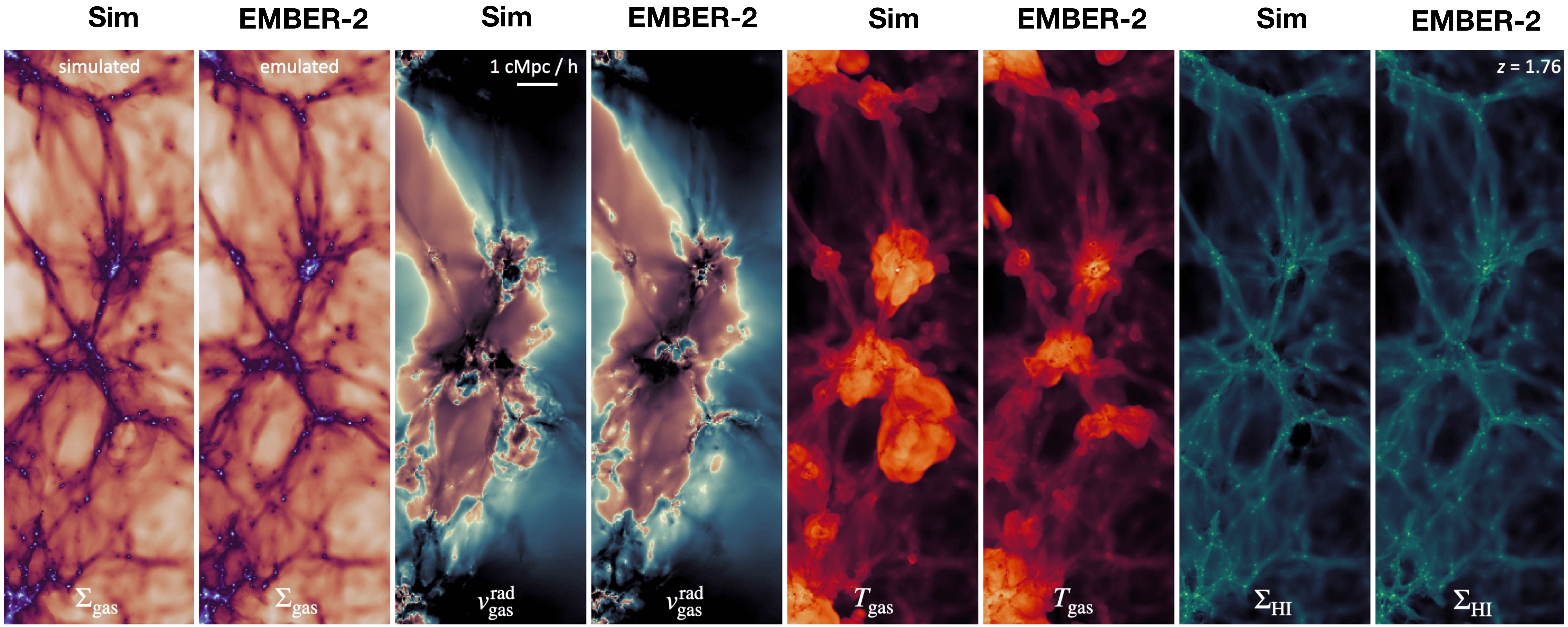
- Predicts CDDF, PS, and Bispectrum to within 10-20% down to galactic scales!
- Reproduces (projected) H_I masses in halos and its scatter!
- Trained at one specific redshift (z=2)!

EMBER-2

Designed to mitigate some of the short-comings of EMBER v1

- Multi-redshift predictions: Combines data from continuous range of redshifts (e.g., z=0-6)
- Leaner architecture ➜ easier to train, faster predictions
- Multi-field predictions: allows both multiple input and output fields ➜ consistent predicted T – density fields

$$(\Sigma_{\rm DM}, v_{\rm DM}^{\rm los}) \to (\Sigma_{\rm gas}, v_{\rm gas}^{\rm los}, T_{\rm gas}, \Sigma_{\rm HI})$$



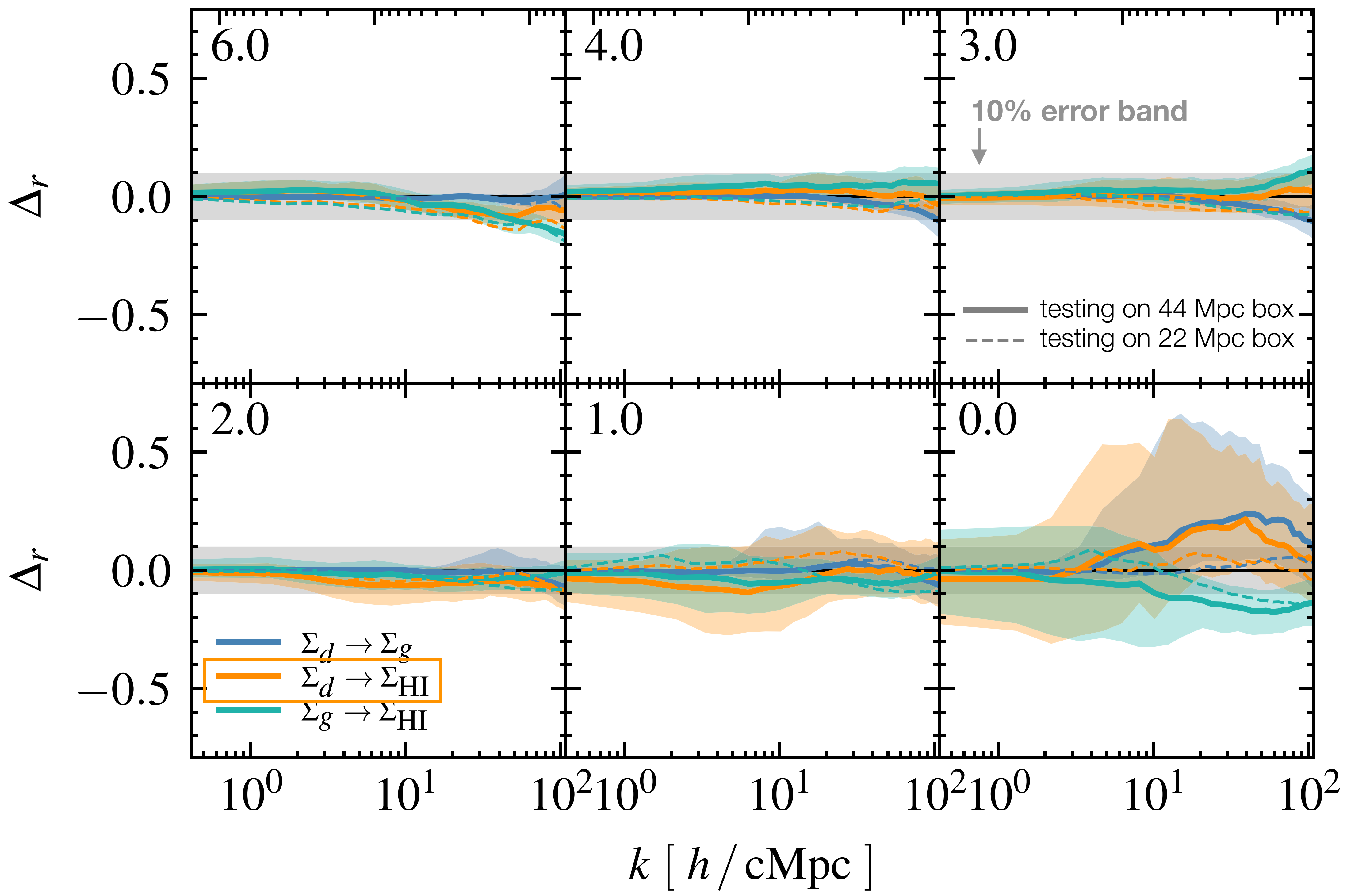
Bernardini, RF, et al. 2025, arXiv:2502.15875

EMBER-2 Results

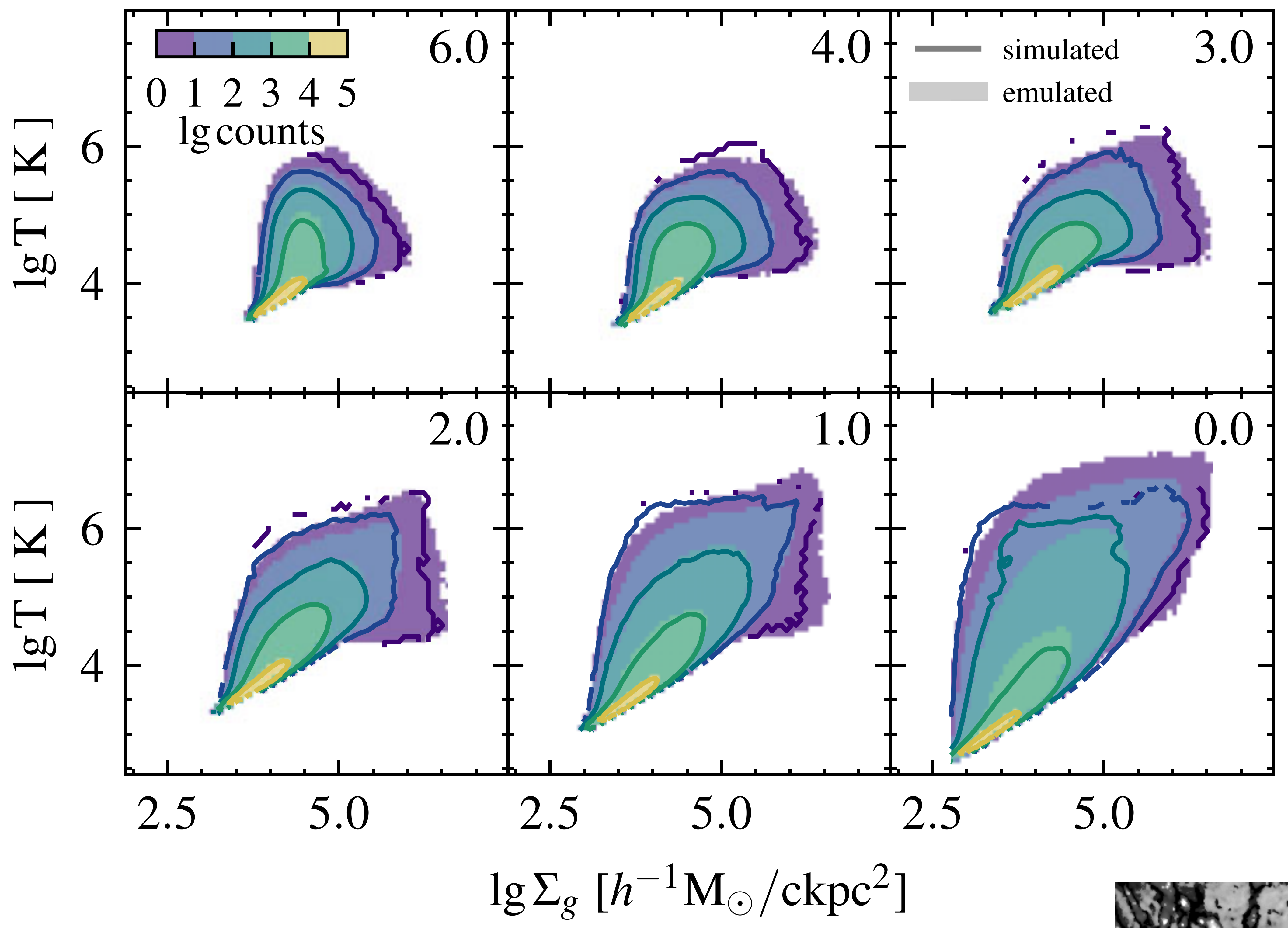
Bernardini, RF, et al. 2025, arXiv:2502.15875

- Prediction performance for CDDF and PS on par with or better than EMBER-1
- Useful metric: Cross-correlation coefficient between $\Sigma_{\rm HI}$ and $\Sigma_{\rm dm}$: $r(k) = \dfrac{P^{\times}_{\rm dm,HI}}{\sqrt{P_{\rm dm} P_{\rm HI}}}$
- Recovers well correlations between different baryonic fields, e.g. density and temperature

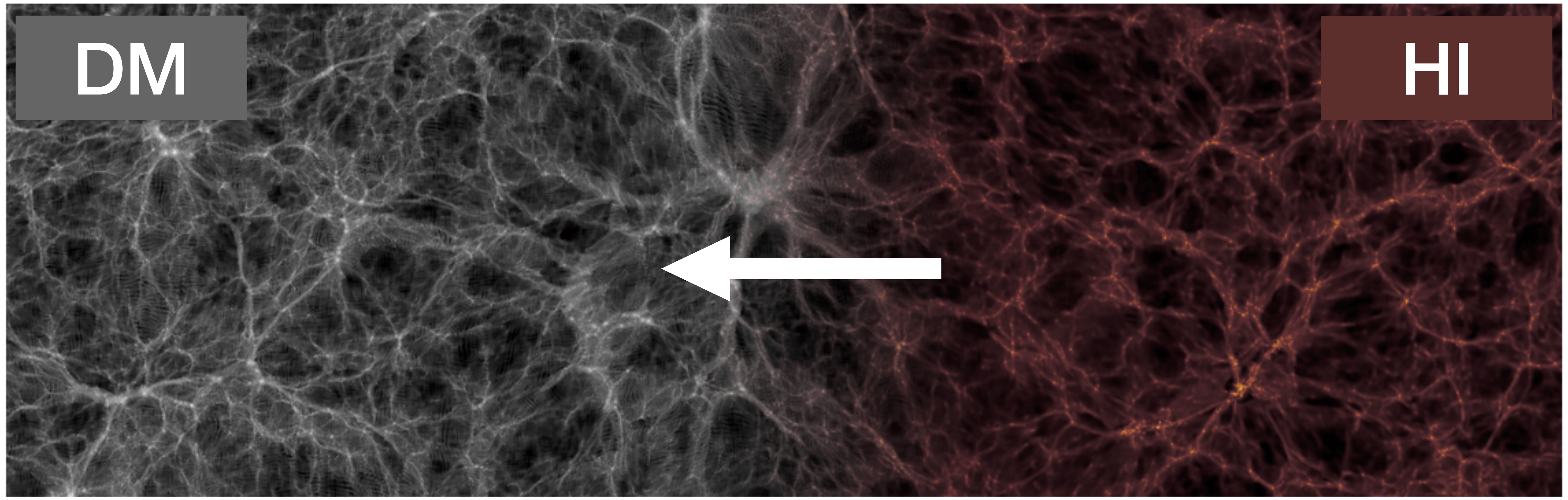
Difference of c.c. coefficient between simulation and reconstruction $\Delta_r(k) = \dfrac{r^{\rm rec}}{r^{\rm sim}} - 1$



Simulated and reconstructed phase diagram



The inverted problem: Predicting DM field from HI



Method:

- EMBER-2 w/ $(\Sigma_{\rm HI}, v_{\rm HI}^{\rm los}) \rightarrow (\Sigma_{\rm DM}, v_{\rm DM}^{\rm los})$, multi-z (z=0-6)

Simulations:

- 22 + 44 cMpc cosmological hydrodynamic simulations
- $m_b \sim 5 \times 10^5\ M_\odot$, $m_{DM} \sim 2.7 \times 10^6\ M_\odot$
- Planck-15 cosmological parameters
- FIRE-2 baryonic physics

Training

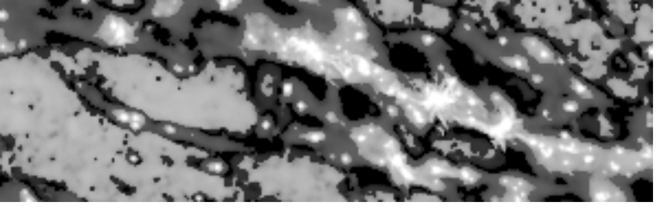
- 44 cMpc volume divided into 20 slabs (1024^2 pixels)
- each slab split randomly into 128^2 pixel tiles, augmented
- field data scaled with modified logarithmic functions
- same hyper-parameters and training strategy as default EMBER-2

Testing:

- tiles from 22 cMpc volume (all axes) and/or unused projections from 44 cMpc (for 3-fold CV)

Reconstructing DM from HI: Visual impression

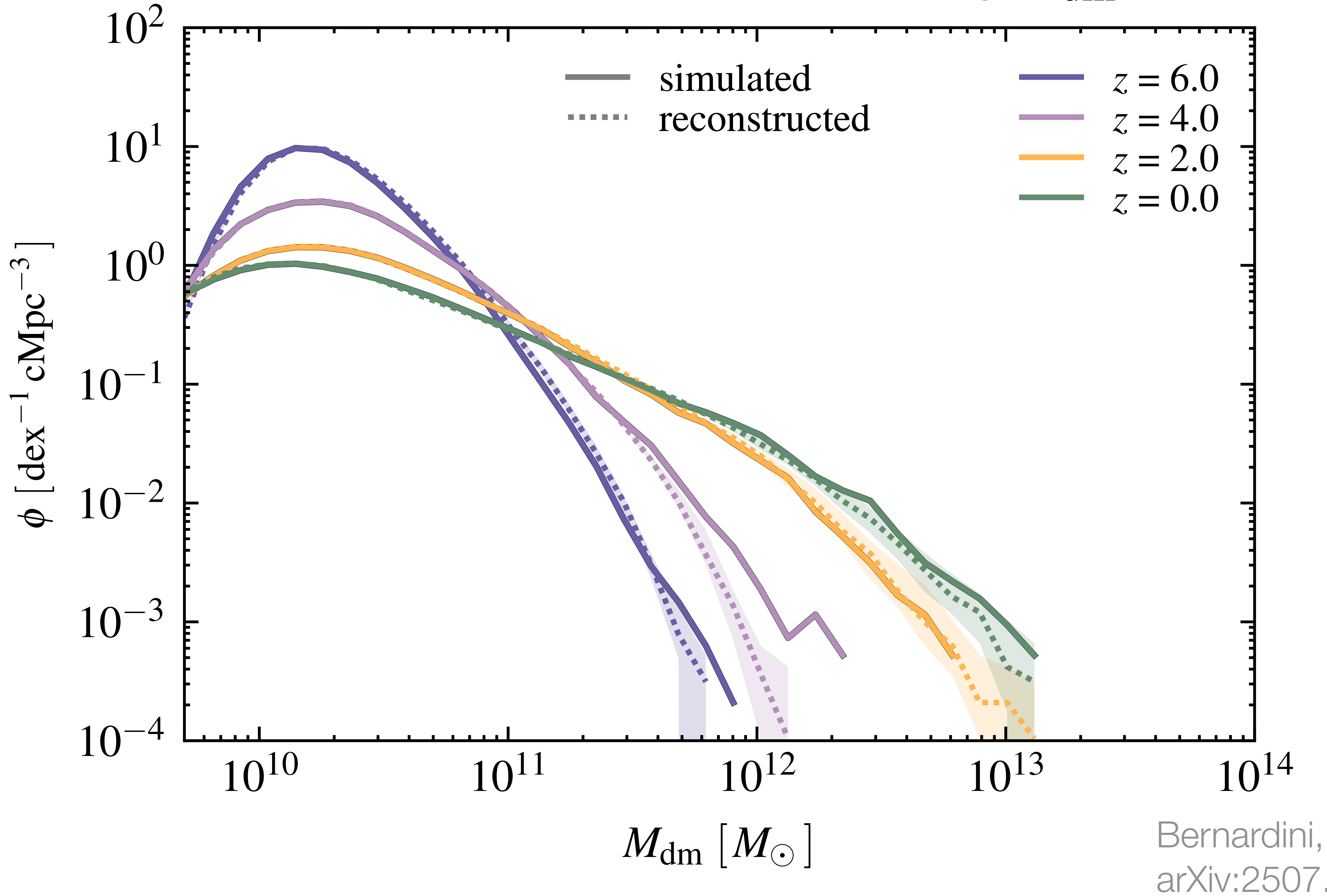
- Good reconstruction of DM field both for larger scales (~10 Mpc) and on scales of individual halos (< 1 Mpc)



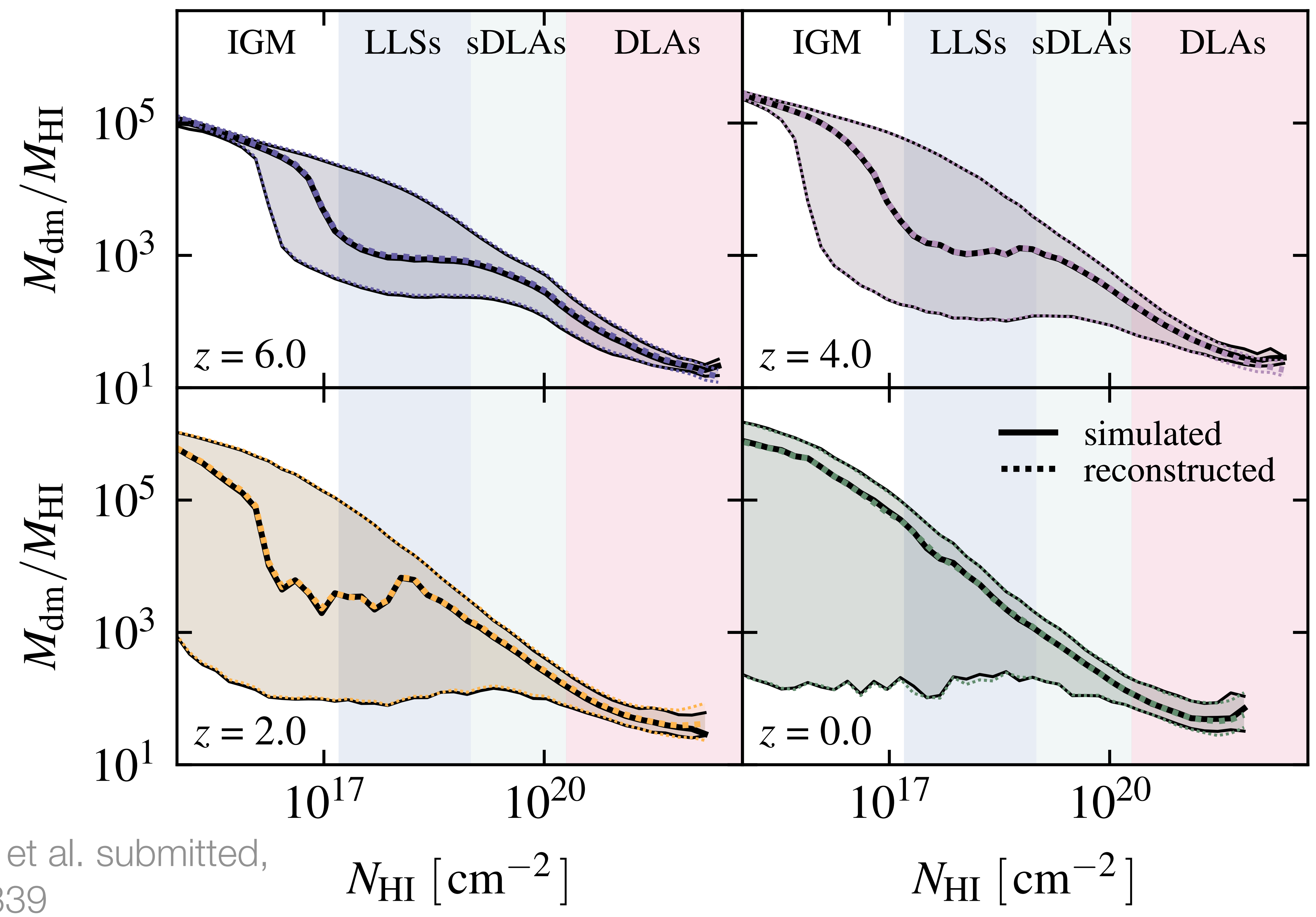
Reconstructing DM from HI: Properties of structures

- **Structures** are defined as **peaks in HI field** above some threshold HI column N_{HI}
- Each such structure has:
  - an associate HI and DM **mass**: projected mass within fixed aperture radius of 150 ckpc
  - a **peak HI column** N_{HI}

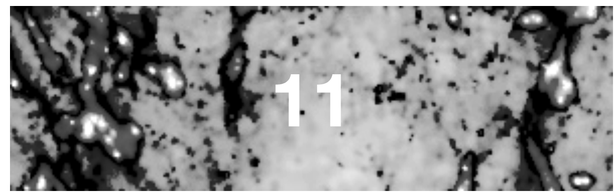
Structure mass function = $\dfrac{dn}{d \lg M_{\mathrm{dm}}}$



Dark matter to HI mass ratio



Bernardini, RF, et al. submitted,
arXiv:2507.05339



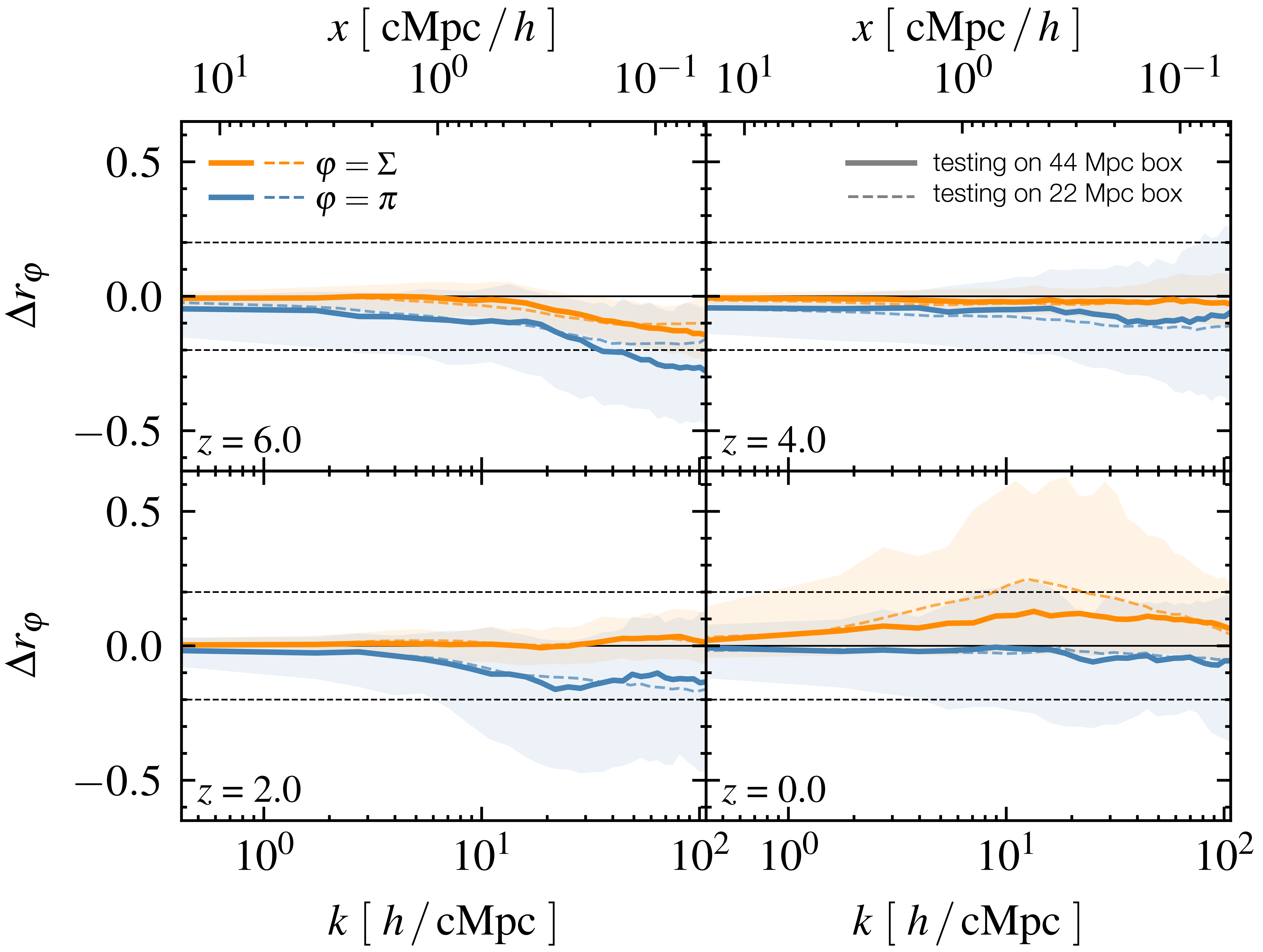
Reconstructing DM from HI: Cross-correlation

Cross-correlation coefficient

- Between surface density ($\Sigma_{\rm HI}$ & $\Sigma_{\rm dm}$)
- Between kinetic surface energy density ($\pi_{\rm HI}$ & $\pi_{\rm dm}$, where $\pi = v^2\Sigma/2$)
- As before, calculate the difference Δ_r of cross-correlation coefficient between simulation and reconstruction

DM — HI correlation reconstructed to better than 20% over z=0-6 and k up to ~100 h/Mpc.

Good reconstruction deep into the highly non-linear regime!



Strengths & Limitations

Strengths of the approach

- Field-level inference
- Fast inference

Limitations

- Trained on a single galaxy-formation model (baryonic physics not yet marginalized)
- Proof-of-concept study: assumes idealized HI maps (no noise, beam, or limited coverage)

Summary

- Large HI surveys will provide field-level information that encodes the underlying dark matter distribution

- **EMBER-2** is a deep-learning (CNN-GAN) based approach that enables **direct field-level inference of dark matter from HI** (and vice versa)
  - Structure based measures (e.g., structure mass functions) are reconstructed extremely well over z=0-6
  - HI — DM cross-correlations of surface density and surface energy density recovered to ~20% accuracy out to k ~ 100 h/cMpc

- Proof-of-concept: to be adapted and integrated into future HI analysis pipelines

Thank you!

References

Bernardini et al. 2022, arXiv:2110.11970

Feldmann et al. 2023, arXiv:2205.15325

Bernardini et al. 2025, arXiv:2502.15875

Bernardini et al. subm., arXiv:2507.05339

EMBER-2 source code:
https://maurbe.github.io/ember-2