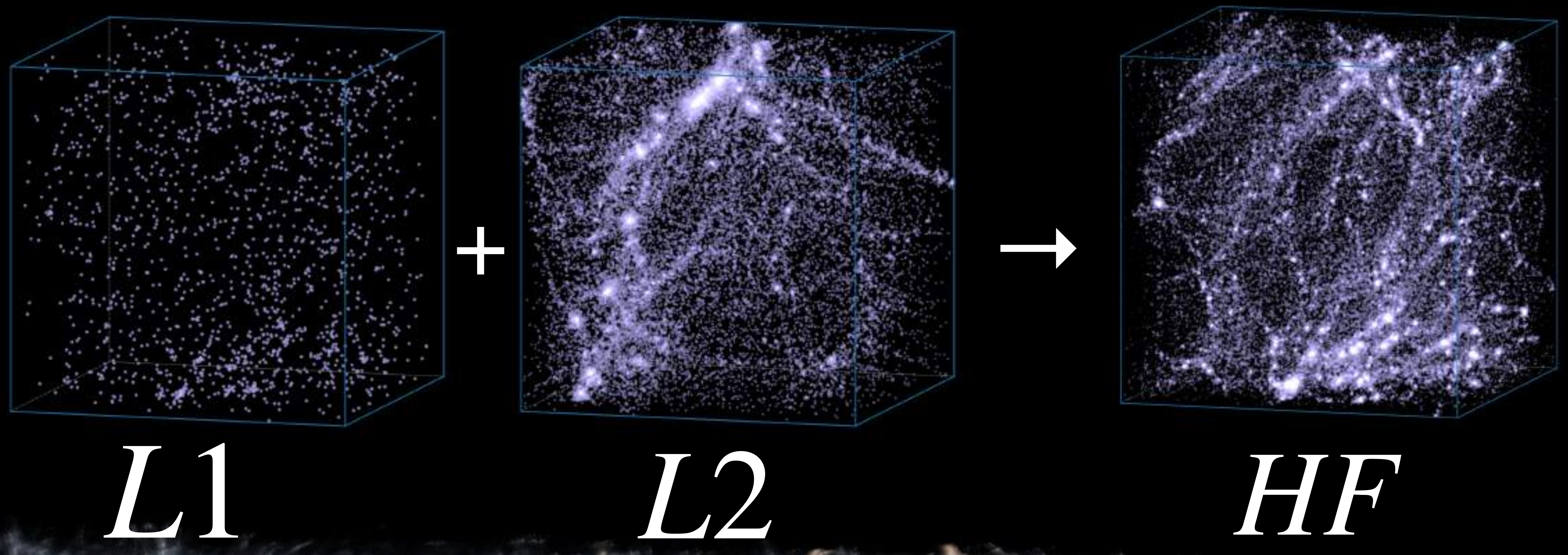
MF-Box: Multi-fidelity and multi-scale emulation, and PRIYA's large *Lya* emulator using Astrid simulations

Ming-Feng Ho
UC Riverside
NASA FINESST FI
jibancat.github.io

MF-Box (arXiv:2306.03144)
PRIYA (arXiv:2306.05471)



Simeon Bird (UCR Astro)

Martin A. Fernandez (UCR Astro → CSU climate science)

Christian Shelton (UCR CS)

Astrid simulation collaboration:

Yueying Ni (CfA), Nianyi Chen (CMU), Patrick Lachance (CMU), Xiaowen Zhang (CMU), James Davies (Scuola Normale Superiore), Mahdi Qezlou (UCR), Yu Feng, Tiziana Di Matteo (CMU), Rupert Croft (CMU)

Pippi Longstocking, created by Astrid Lindgren

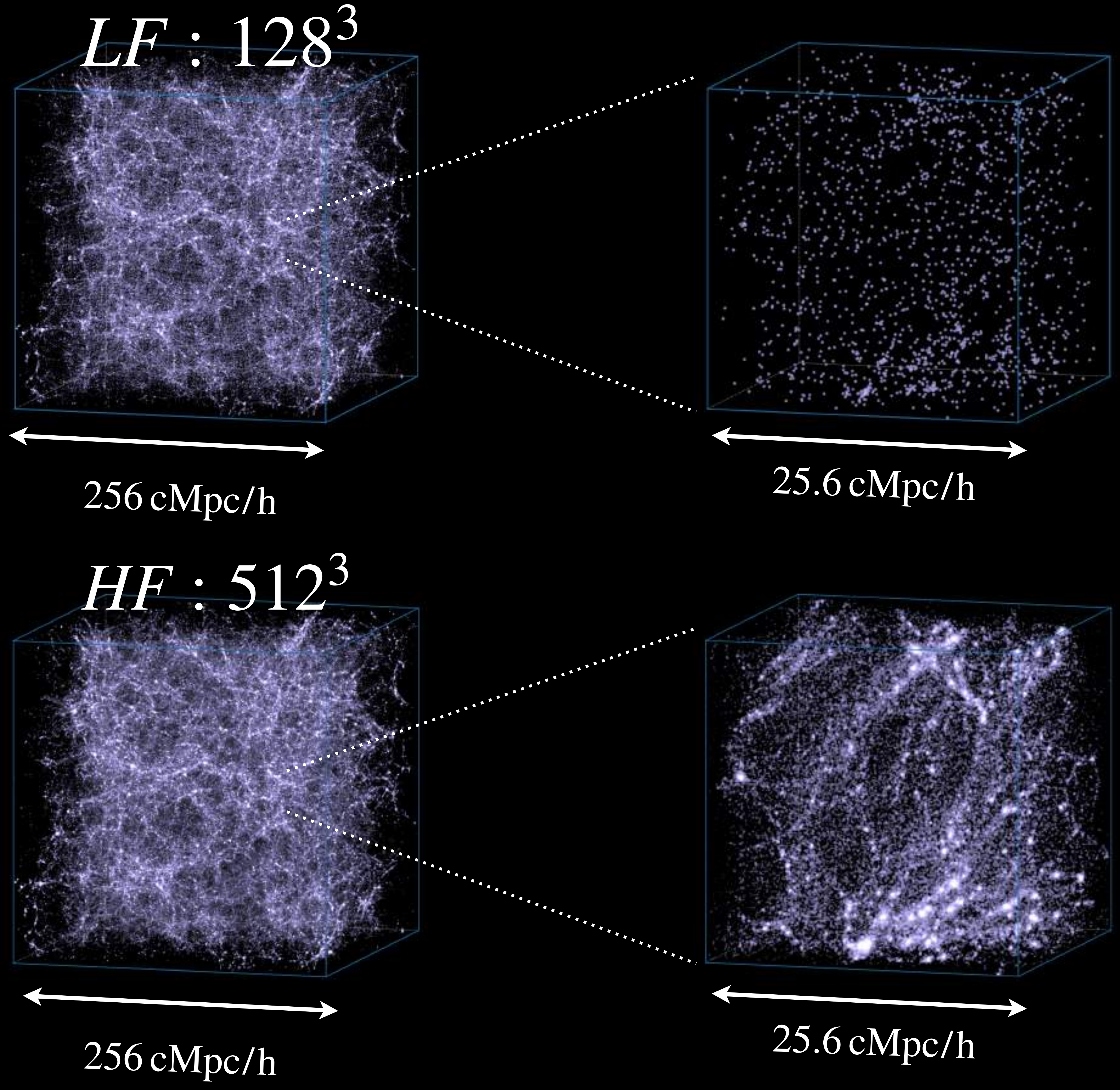
Jul 2023 | Cyberland

1

Need more small scales?

Common approach: Run higher resolution

- Me: I can see lots of structures



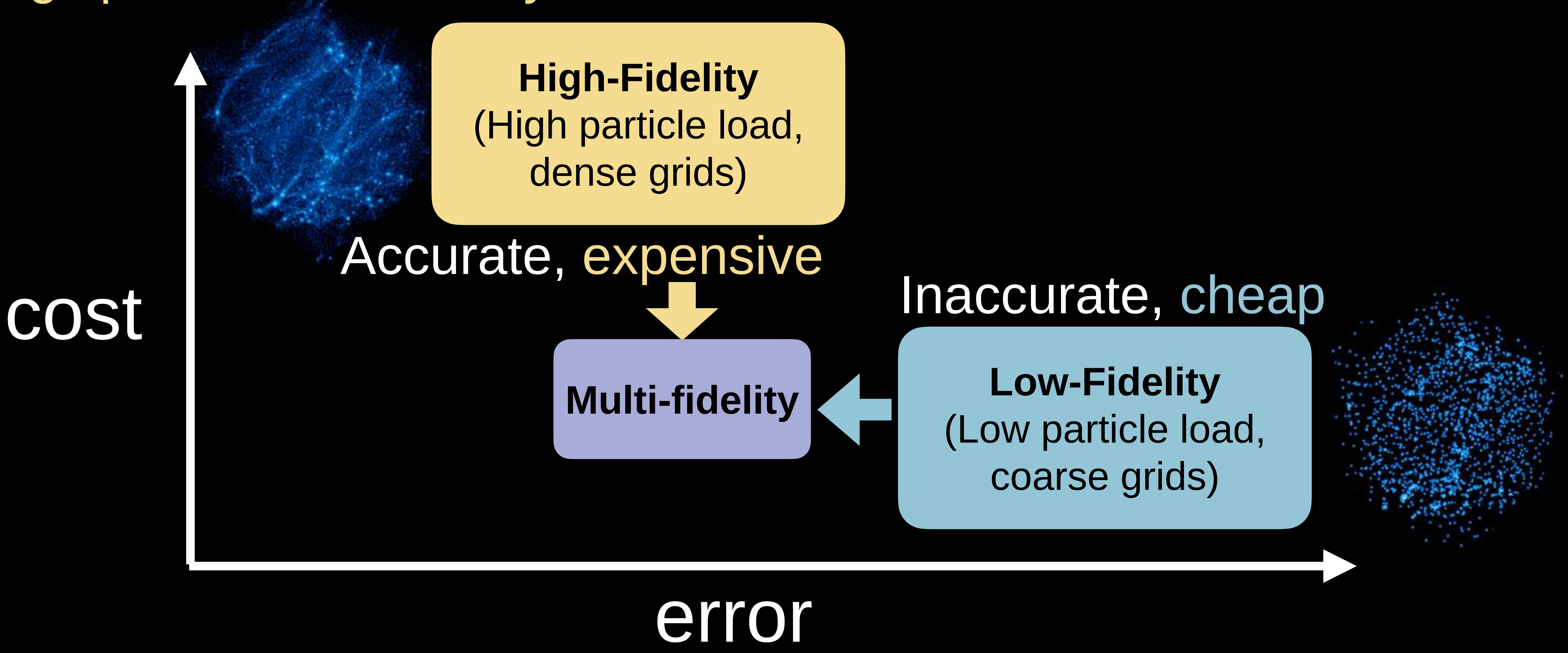
- a lack of structure at small scales

- Simulators say:
  Just increase resolution (more particles or denser grids)

- More small-scale structures
- You'll have almost the same information at large scales
- Simulators say: But be cautious, as you might hit the budget limitation for computation time. Increasing particle loads is proportional to O(N log(N))

Multi-fidelity: The solution?

Balancing speed and accuracy



Use tons of Low-fidelity to understand large scales,
Only a few of High-fidelity to refine information at small scales

See Ho, Bird, Shelton (2022) on multi-fidelity matter power spectrum emulation

But, simulators usually take a shortcut

Simulators: Separating scales for efficiency


Me: Are the physics at large scales correct?


Simulators: no money/time. Just run a large volume low-quality test simulation.

- We can examine large-scale statistics in the test simulations. For example, matching linear theory at large scales.


Is my subgrid model at small scales correct?

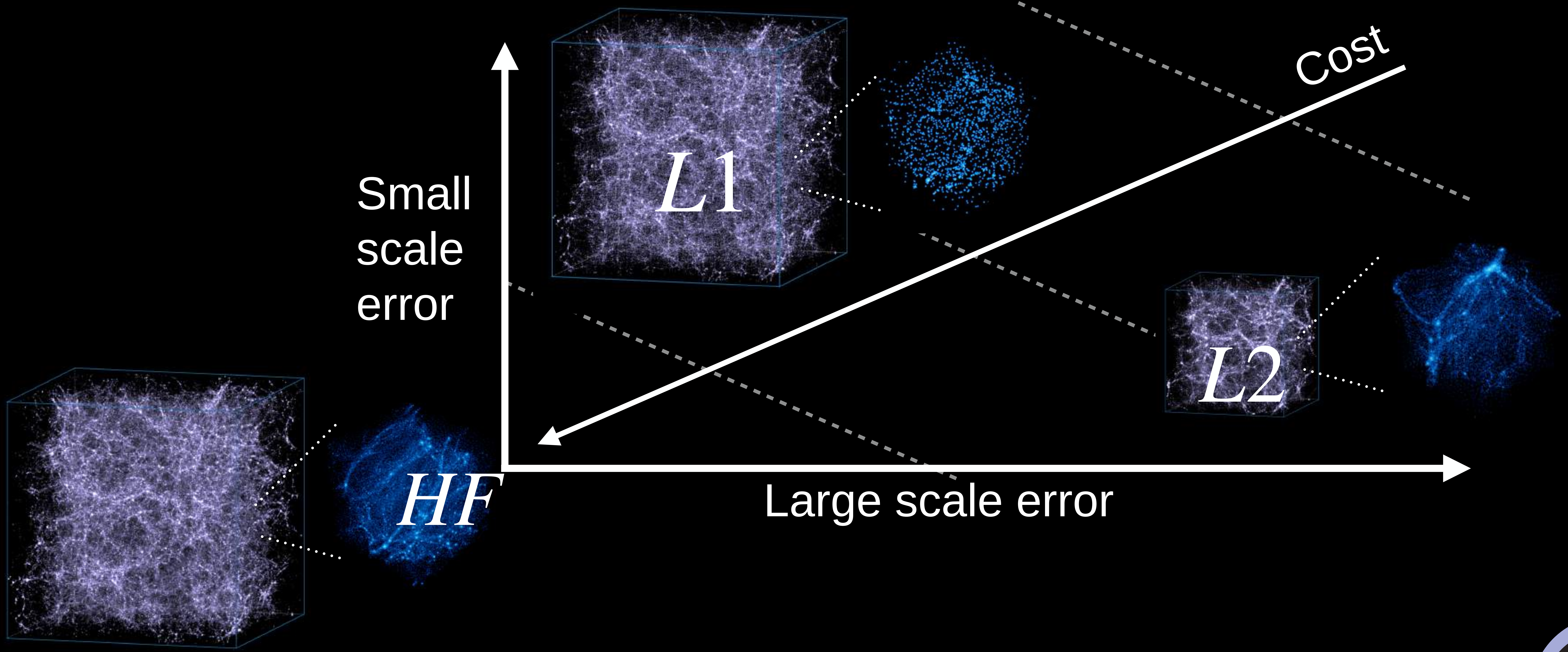

Simulators: no money/time. Just run a small volume low-quality test simulation.

- We can evaluate small-scale statistics in the test simulations. For example, tuning the stellar wind model by matching small-scale statistics.

- Question: How can we integrate this thought process into a coherent Bayesian model for future decision-making?

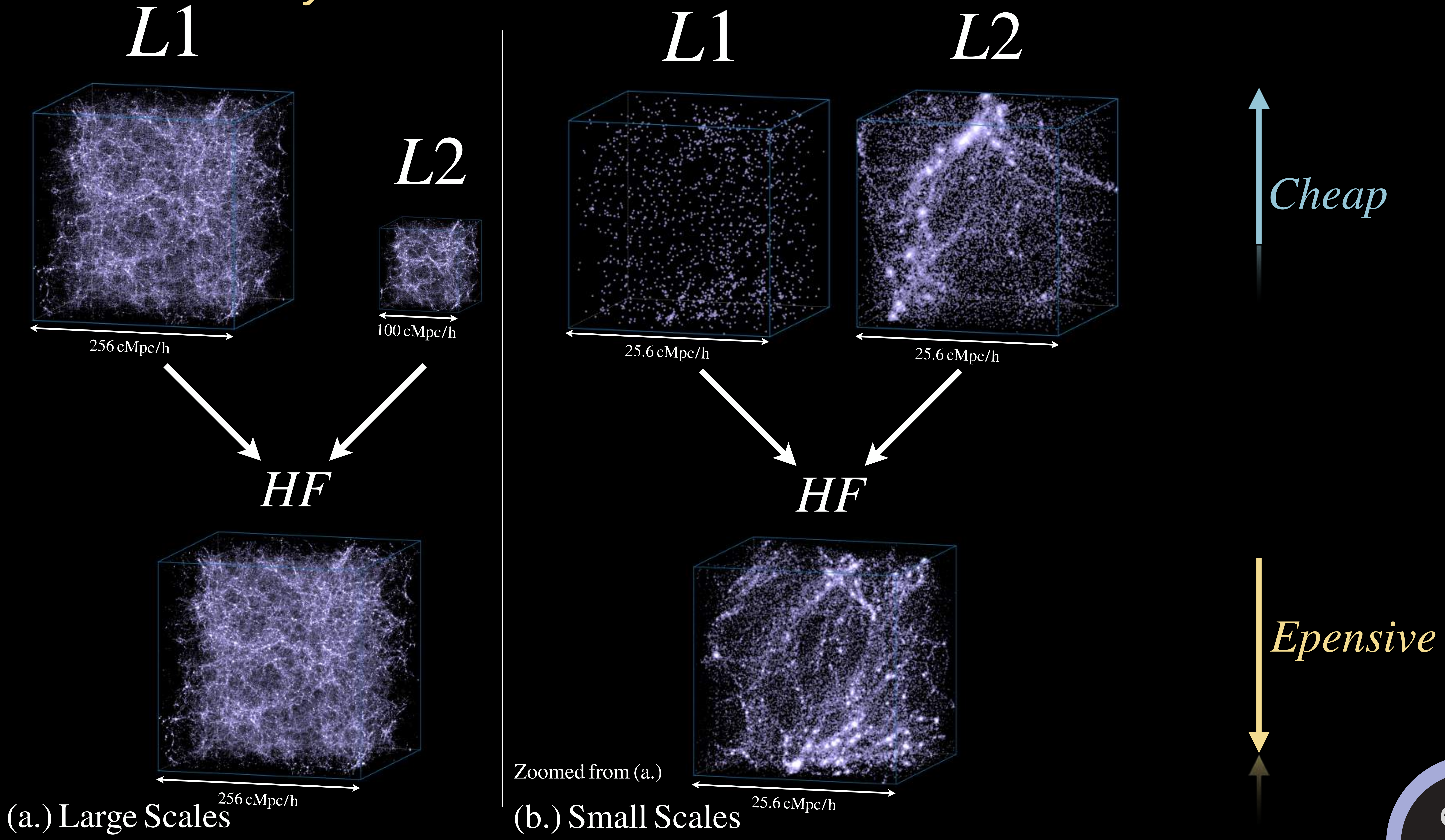
Multi-scale problem

Balancing different ***scales***



MF-Box: Multi-fidelity and multi-scale emulation

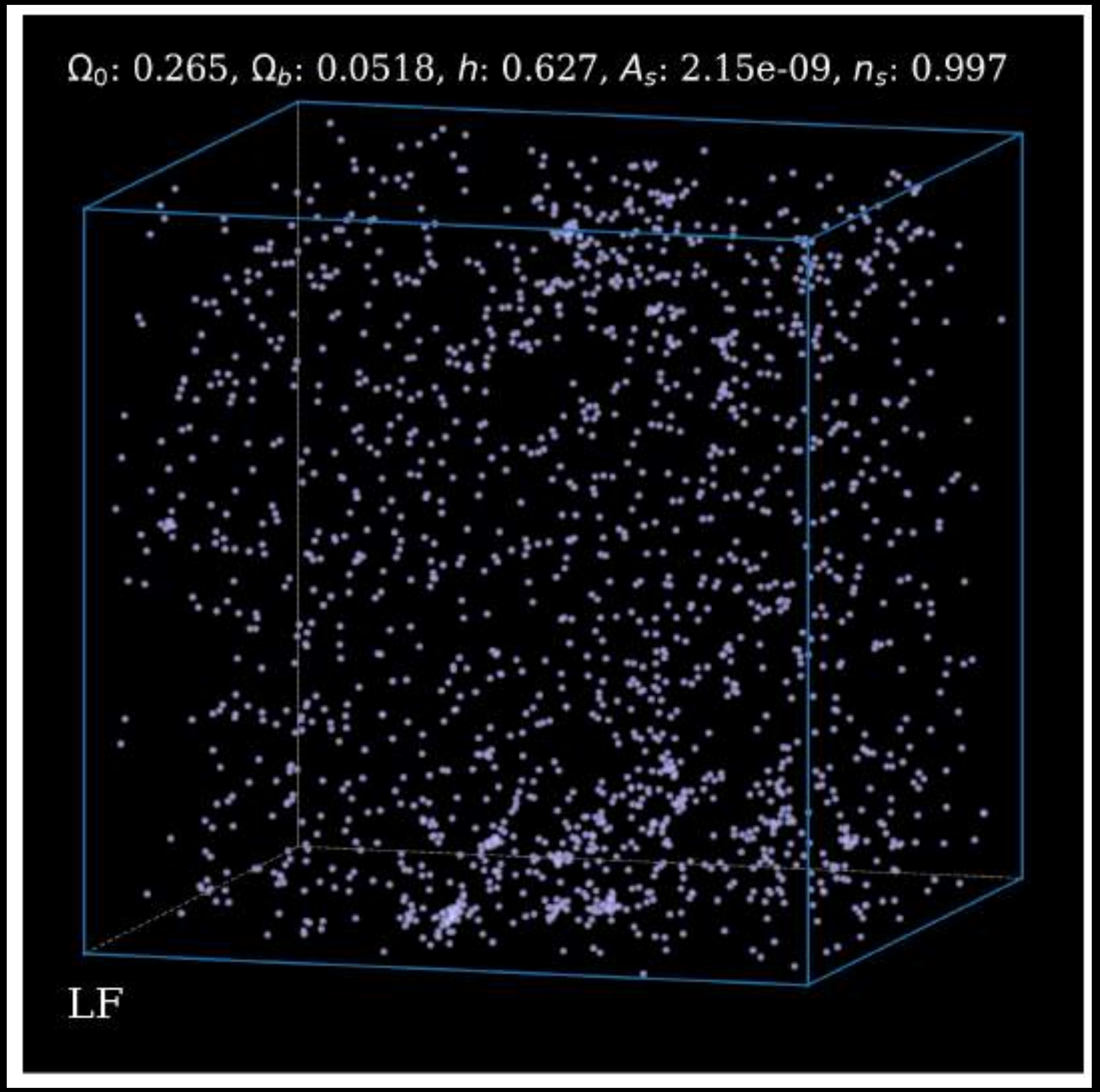
Combining scales in one Bayesian model

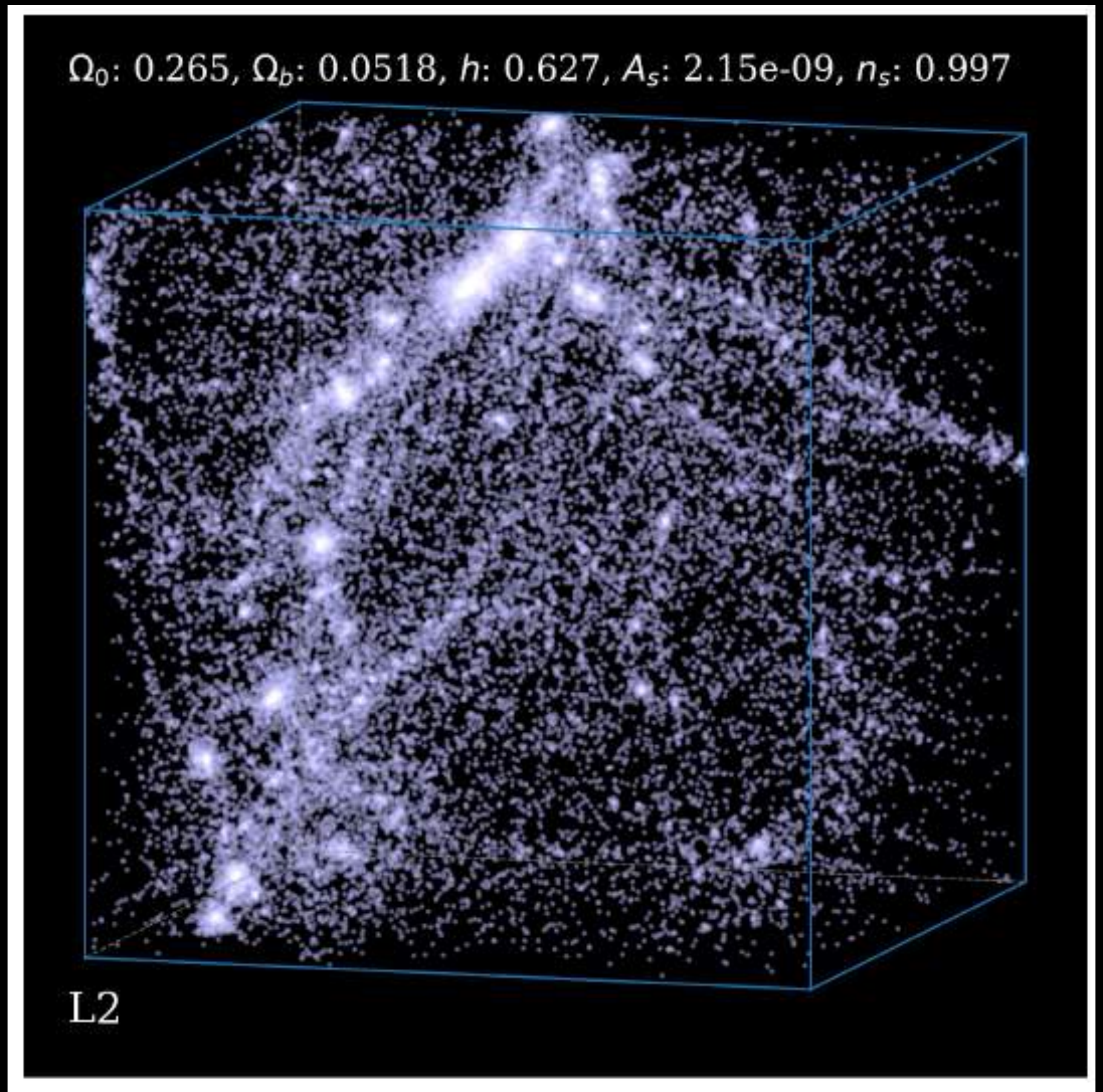


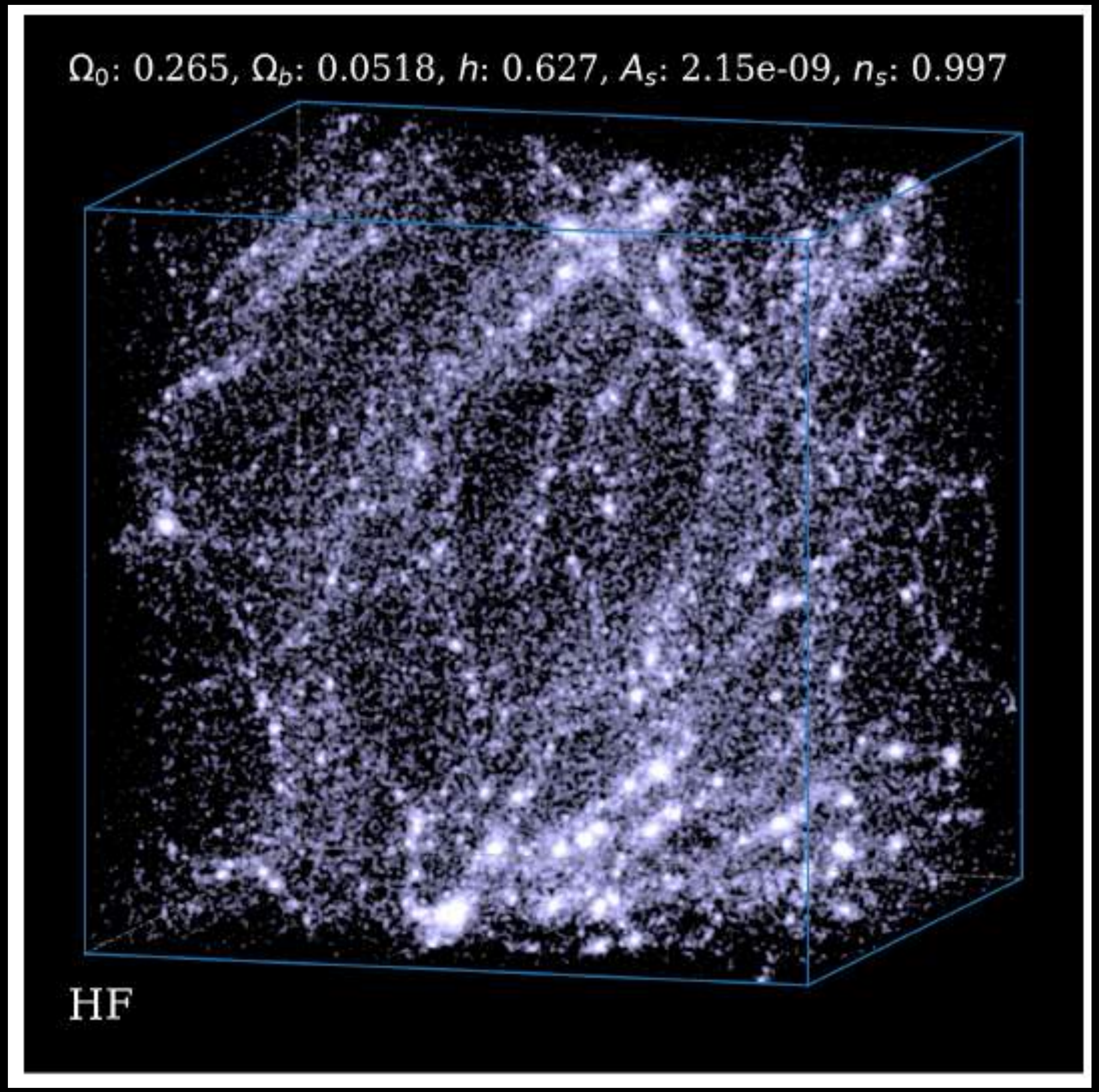
MF-Box (arXiv:2306.03144)

Interpolating cosmology with limited HF simulations

Simulations as a Bayesian model(θ, resolution, volume)



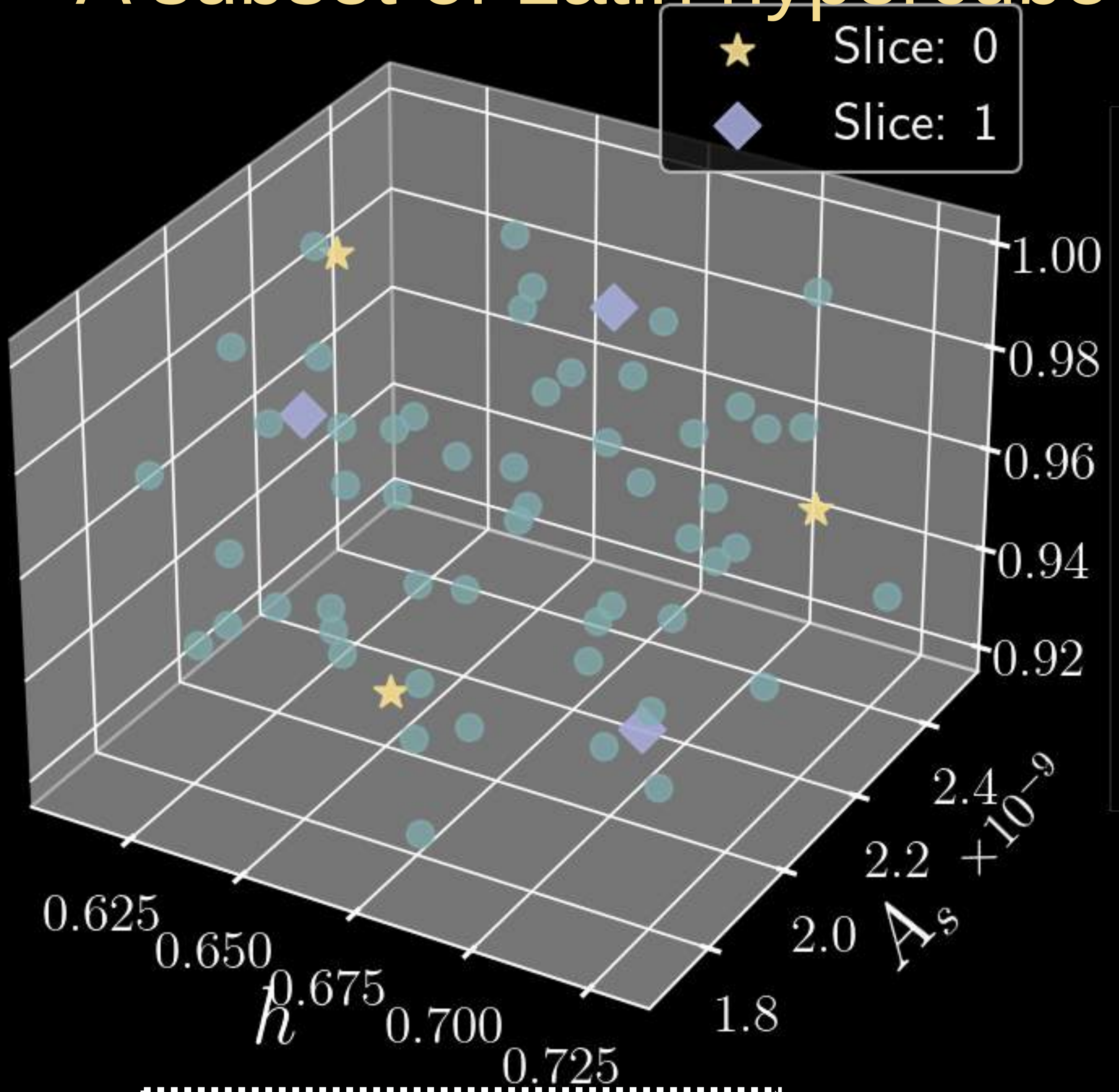


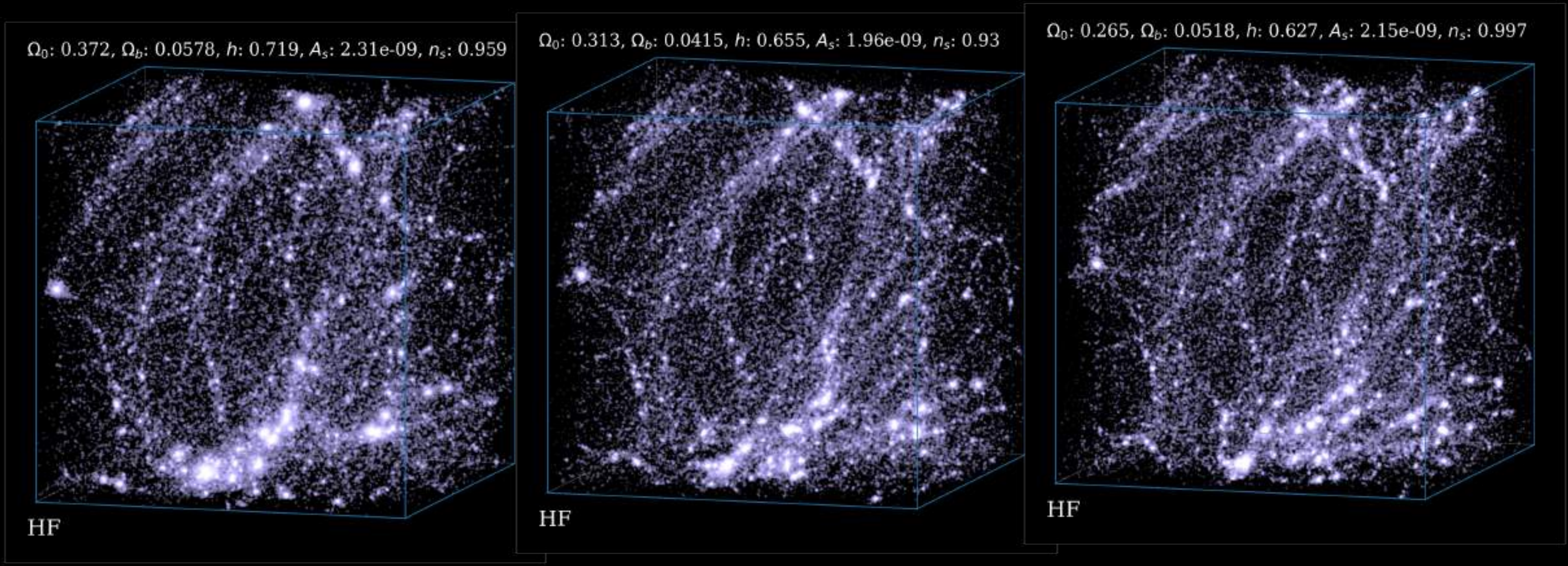


- Cheap information (L1, L2)
- Expensive information (HF)

Sliced Latin hyper cube (SLHD)

A subset of Latin hypercube that is space-filling





$$\text{Slice } 0 : \theta_{\mathrm{HF},0}, \theta_{\mathrm{HF},1}, \theta_{\mathrm{HF},2}$$

$$\boldsymbol{\theta}_{\mathrm{LF}} : SLHD$$
$$\boldsymbol{\theta}_{\mathrm{HF}} : Slice(s)$$

$$\Rightarrow \boldsymbol{\theta}_{\mathrm{HF}} \subset \boldsymbol{\theta}_{\mathrm{LF}}$$

Maximize the *diversity* within a subset

Probabilistic resolution correction with Gaussian processes

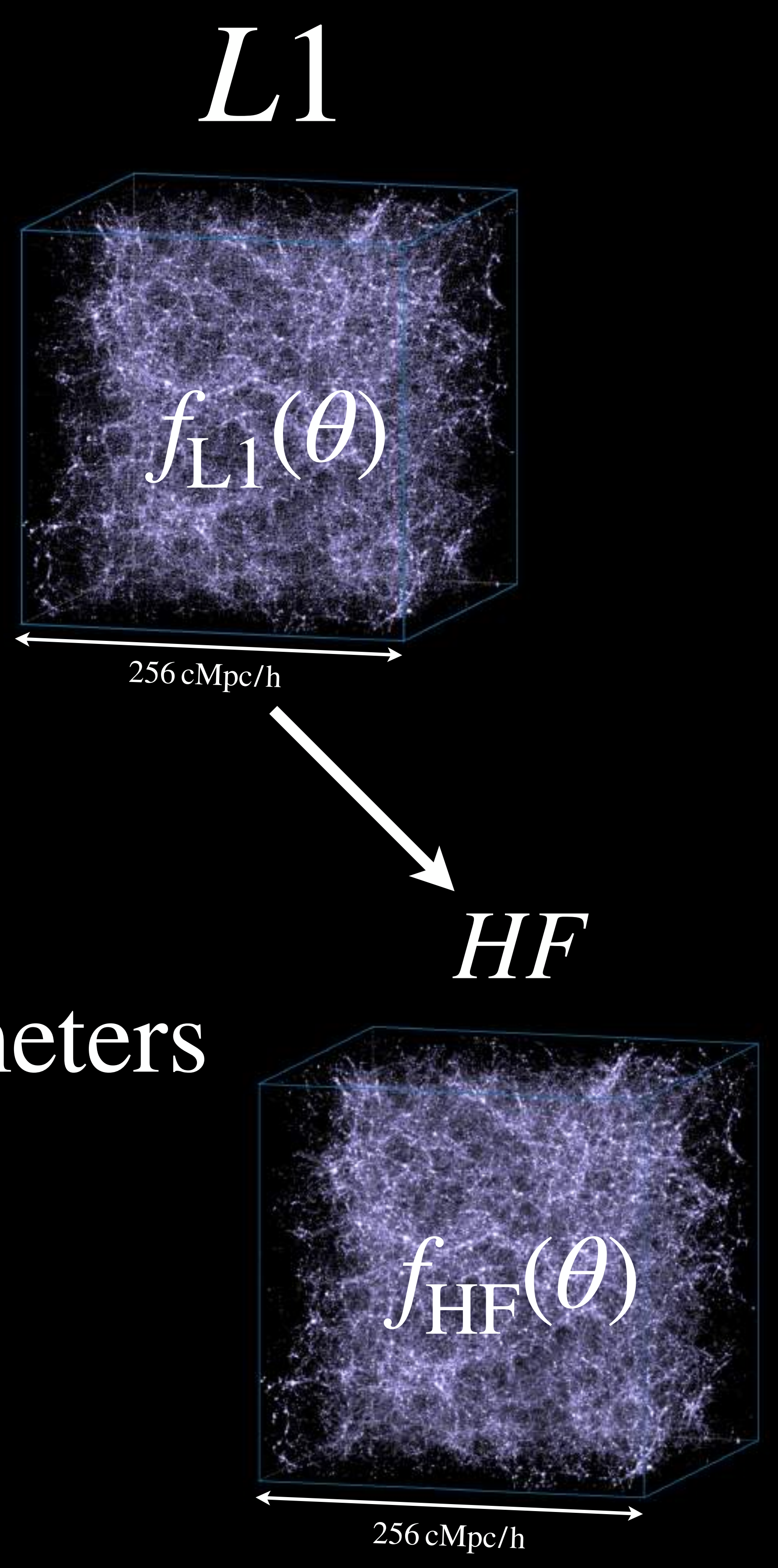
Comparing MFEmulator and MF-Box

- Works for summary statistics
- GP requires minimal training/tuning time (almost effortless, just a click)

θ : simulation input parameters

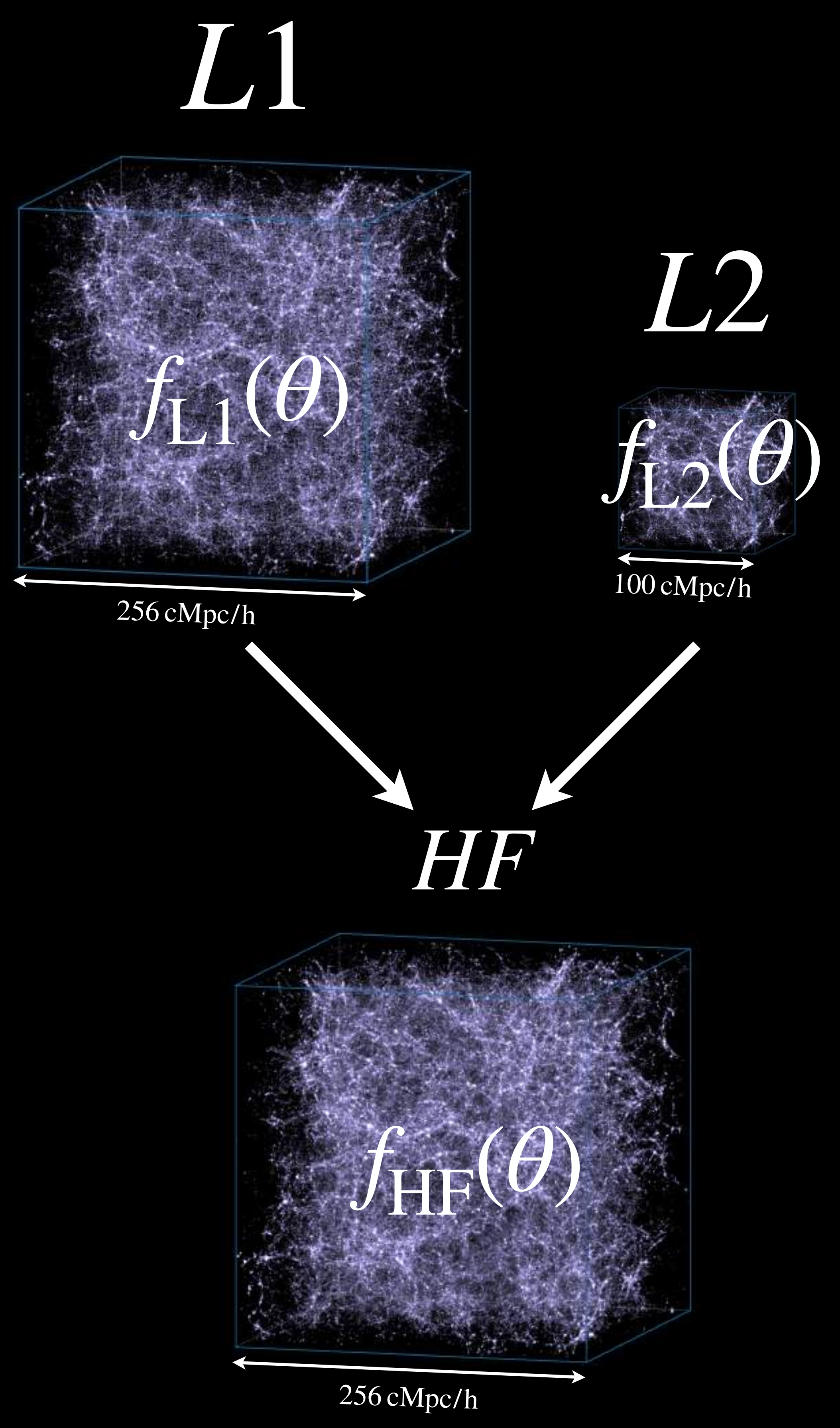
f : summary statistics

MFEmulator (Ho (2022))



Deep GP: $f_{HF}(\theta) = \rho(\theta, f_{L1}(\theta)) + \delta(\theta)$

MF-Box (Ho (2023))



$f_{HF}(\theta) = \rho(\theta, f_{L1}, f_{L2}(\theta)) + \delta(\theta)$

MF-Box as a decision-making tool

Optimizing computational budget allocation

- Objective: Maximize the accuracy of predicted summary statistics

$$\text{Emulator Error} \sim \mathcal{O}(\rho_{\mathrm{L1}} \cdot n_{\mathrm{L1}}^{-\frac{\nu_{\mathrm{L1}}}{d}} + \rho_{\mathrm{L2}} \cdot n_{\mathrm{L2}}^{-\frac{\nu_{\mathrm{L2}}}{d}} + n_{\mathrm{HF}}^{-\frac{\nu_{\mathrm{HF}}}{d}})$$

- Condition: Subject to a limited budget (C: Cost)

$$n_{\mathrm{L1}} \cdot C_{\mathrm{L1}} + n_{\mathrm{L2}} \cdot C_{\mathrm{L2}} + n_{\mathrm{HF}} \cdot C_{\mathrm{HF}} \leq C : \text{limited budget}$$

- Lagrangian, solving Karush–Kuhn–Tucker (KKT) conditions:

$$\mathcal{L}(n_{\mathrm{L1}}, n_{\mathrm{L2}}, n_{\mathrm{HF}}, \lambda) = \eta(\rho_{\mathrm{L1}} \cdot n_{\mathrm{L1}}^{-\frac{\nu_{\mathrm{L1}}}{d}} + \rho_{\mathrm{L2}} \cdot n_{\mathrm{L2}}^{-\frac{\nu_{\mathrm{L2}}}{d}} + n_{\mathrm{HF}}^{-\frac{\nu_{\mathrm{HF}}}{d}})$$
$$+\lambda(n_{\mathrm{L1}} \cdot C_{\mathrm{L1}} + n_{\mathrm{L2}} \cdot C_{\mathrm{L2}} + n_{\mathrm{HF}} \cdot C_{\mathrm{HF}} - C)$$

ν :smoothness

ρ :fidelity correlation

d :parameter dimension

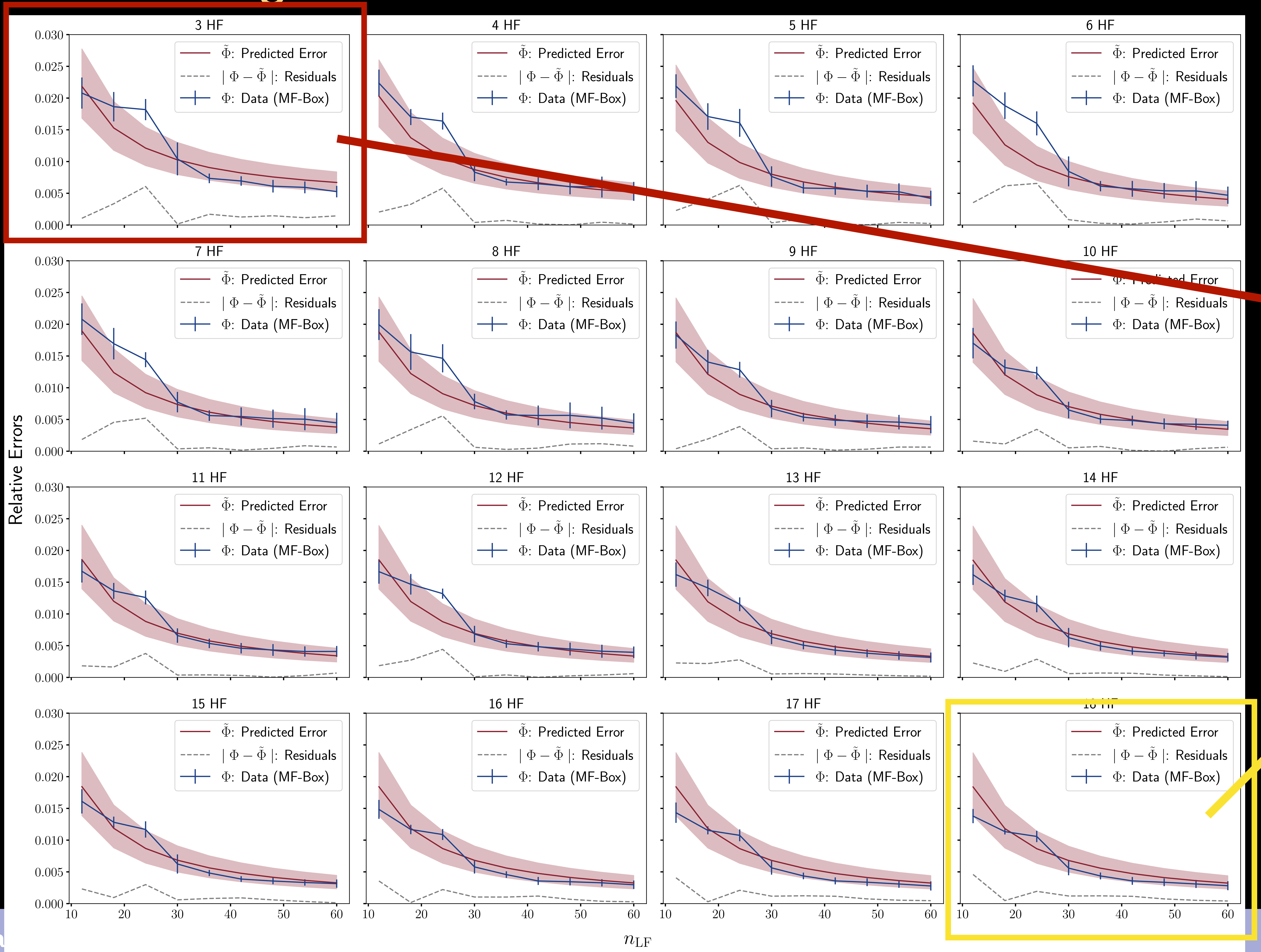
$$n_{\mathrm{L1}} \propto \left(\frac{\nu_{\mathrm{L1}}\rho_{\mathrm{L1}}}{C_{\mathrm{L1}}}\right)^{\frac{d}{\nu_{\mathrm{L1}}+d}}$$
$$n_{\mathrm{L2}} \propto \left(\frac{\nu_{\mathrm{L2}}\rho_{\mathrm{L2}}}{C_{\mathrm{L2}}}\right)^{\frac{d}{\nu_{\mathrm{L2}}+d}}$$
$$n_{\mathrm{HF}} \propto \left(\frac{\nu_{\mathrm{HF}}}{C_{\mathrm{HF}}}\right)^{\frac{d}{\nu_{\mathrm{HF}}+d}}.$$

Simply: Number of sims per node
$\propto$ *correctness / cost*

Ho (2023), original formulation: Ji (2021)

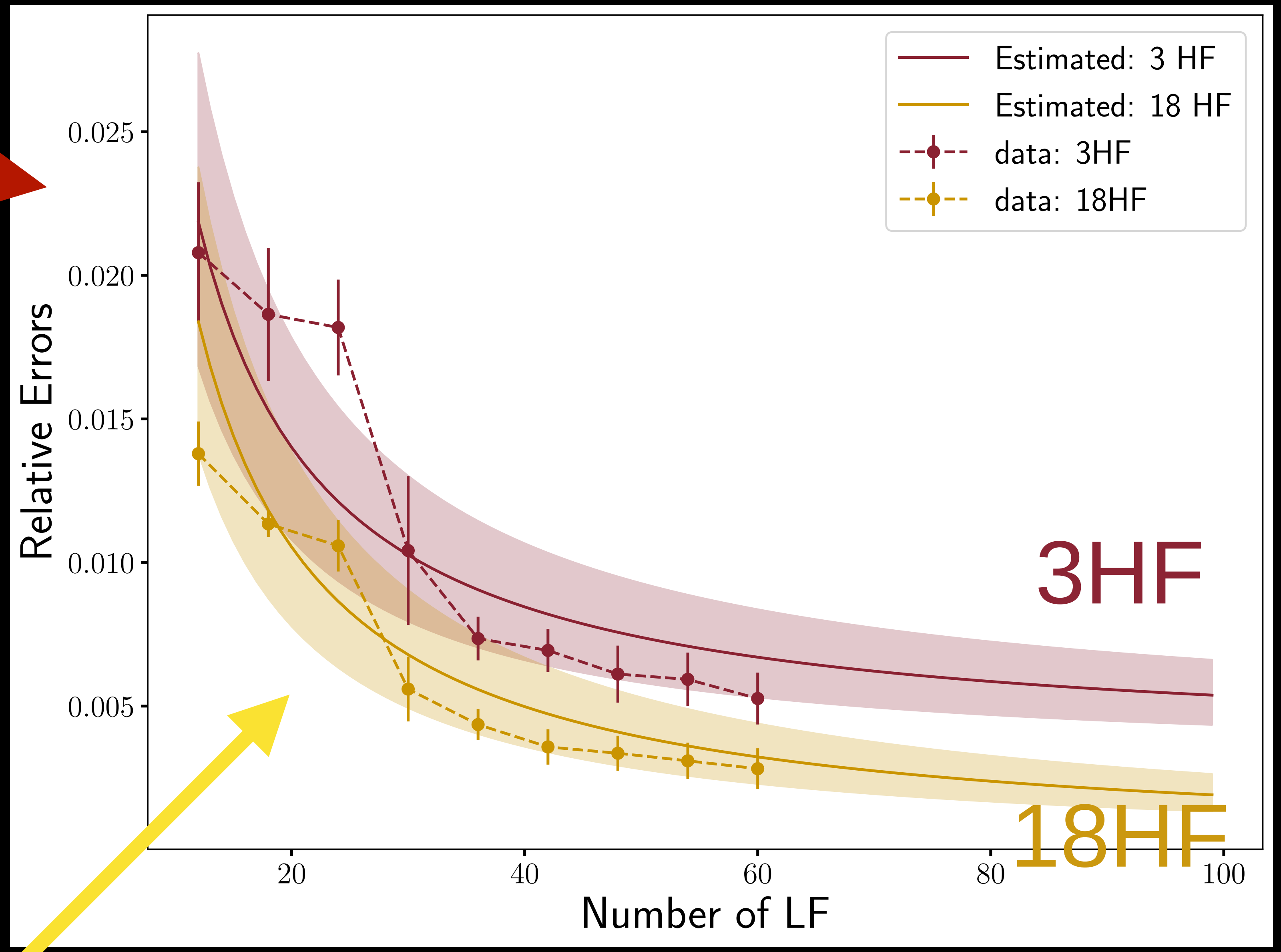
Empirical error analysis with MF-Box

Estimating errors from simulation runs



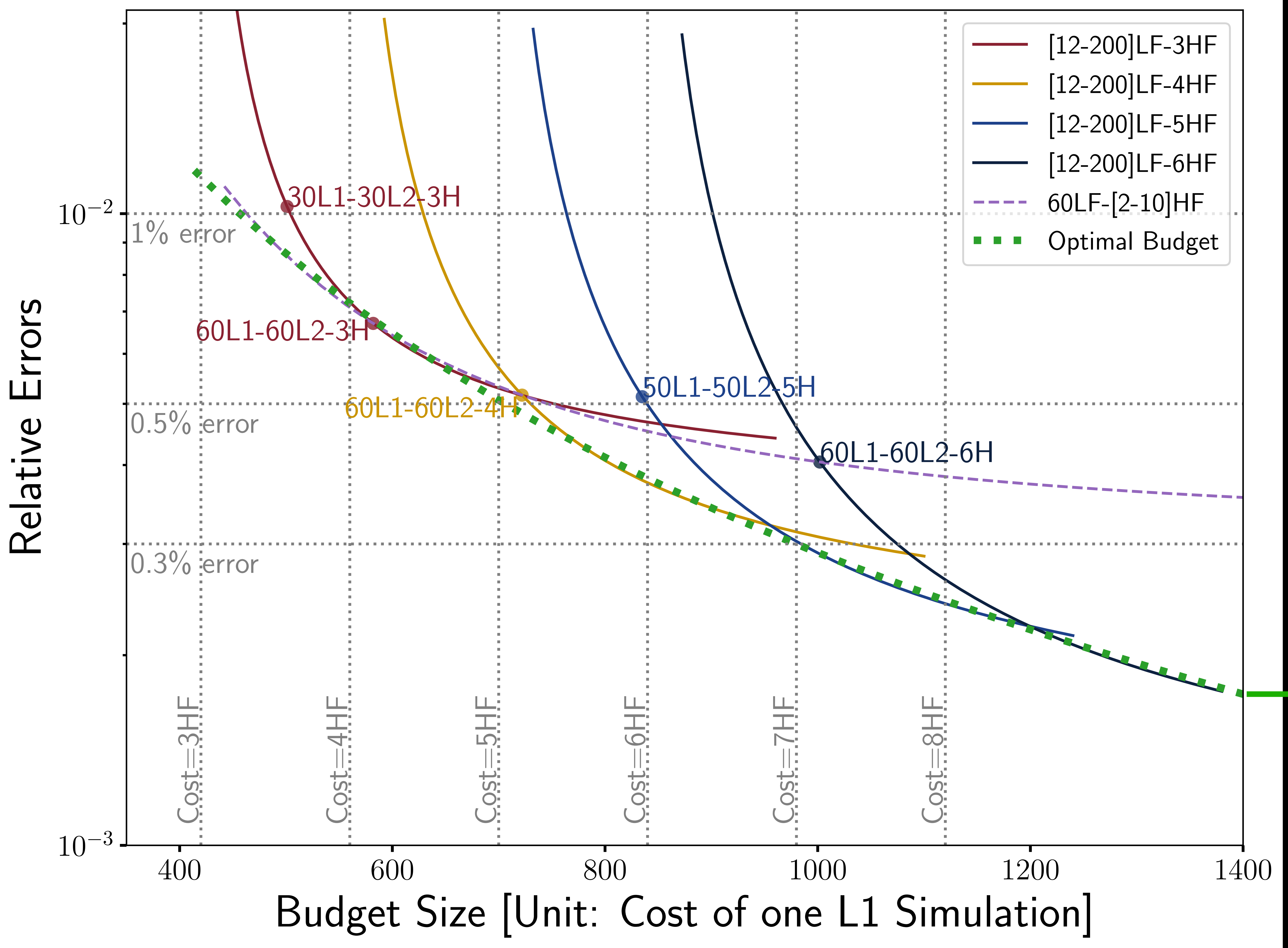
Emulator Error

$$\sim \mathcal{O}(\rho_{\mathrm{L1}} \cdot n_{\mathrm{L1}}^{-\frac{\nu_{\mathrm{L1}}}{d}} + \rho_{\mathrm{L2}} \cdot n_{\mathrm{L2}}^{-\frac{\nu_{\mathrm{L2}}}{d}} + n_{\mathrm{HF}}^{-\frac{\nu_{\mathrm{HF}}}{d}})$$



MF-Box for future budget allocation

Optimize resources for different resolution/volumes



$$n_{L1} \propto \left(\frac{\nu_{L1}\rho_{L1}}{C_{L1}}\right)^{\frac{d}{\nu_{L1}+d}}$$

$$n_{L2} \propto \left(\frac{\nu_{L2}\rho_{L2}}{C_{L2}}\right)^{\frac{d}{\nu_{L2}+d}}$$

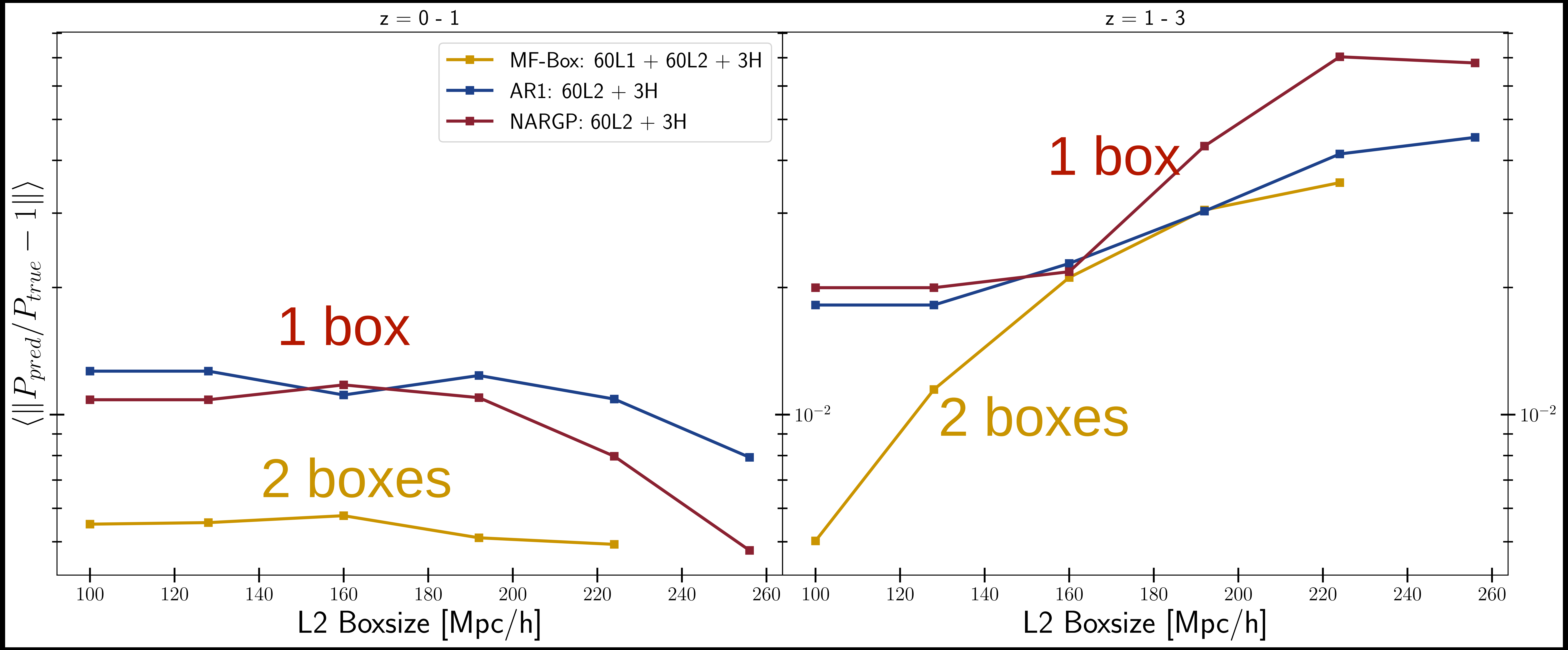
$$n_{HF} \propto \left(\frac{\nu_{HF}}{C_{HF}}\right)^{\frac{d}{\nu_{HF}+d}}.$$

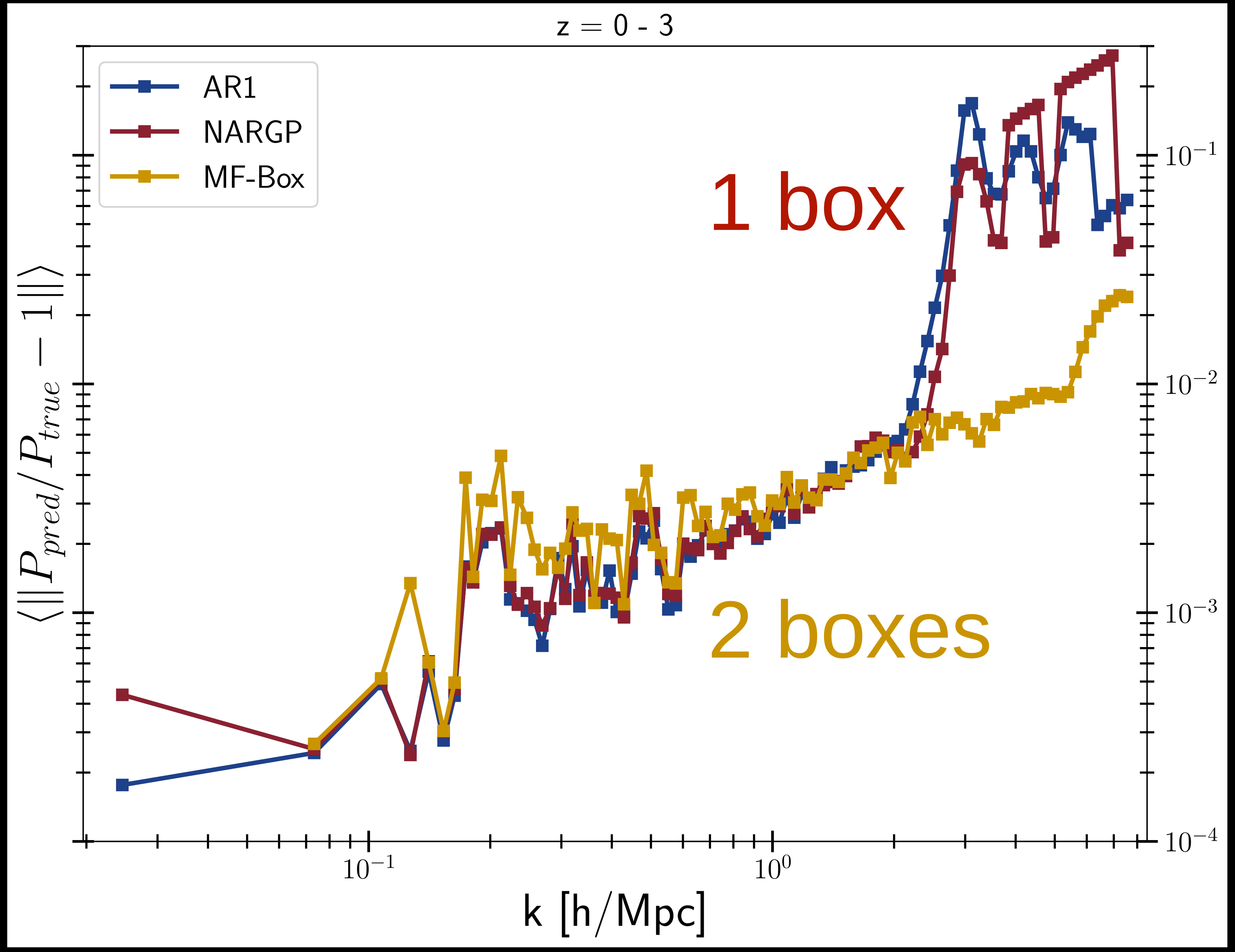
MF-Box (arXiv:2306.03144)

MF-Box's performance in combining scales

Enhancing small scales with smaller boxes

Low-z *high-z* *all-z*



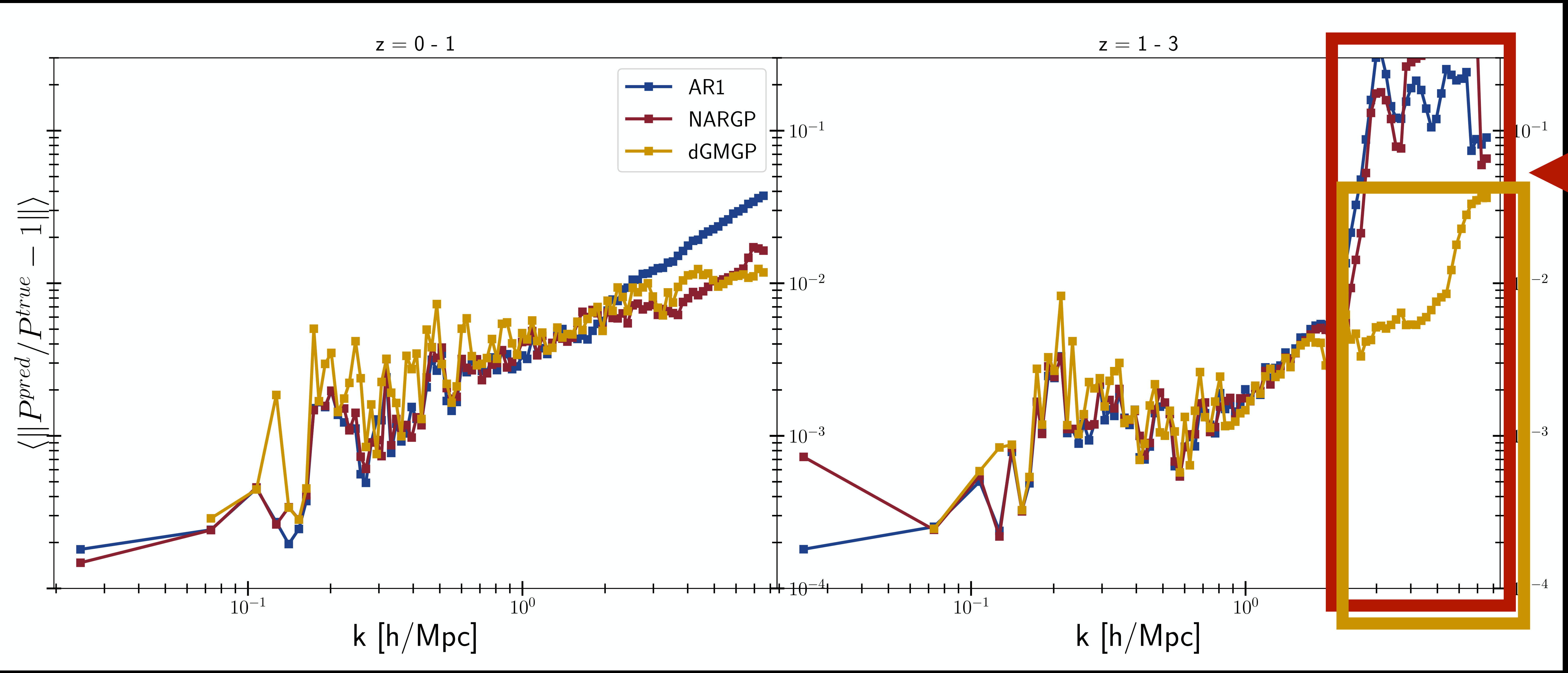


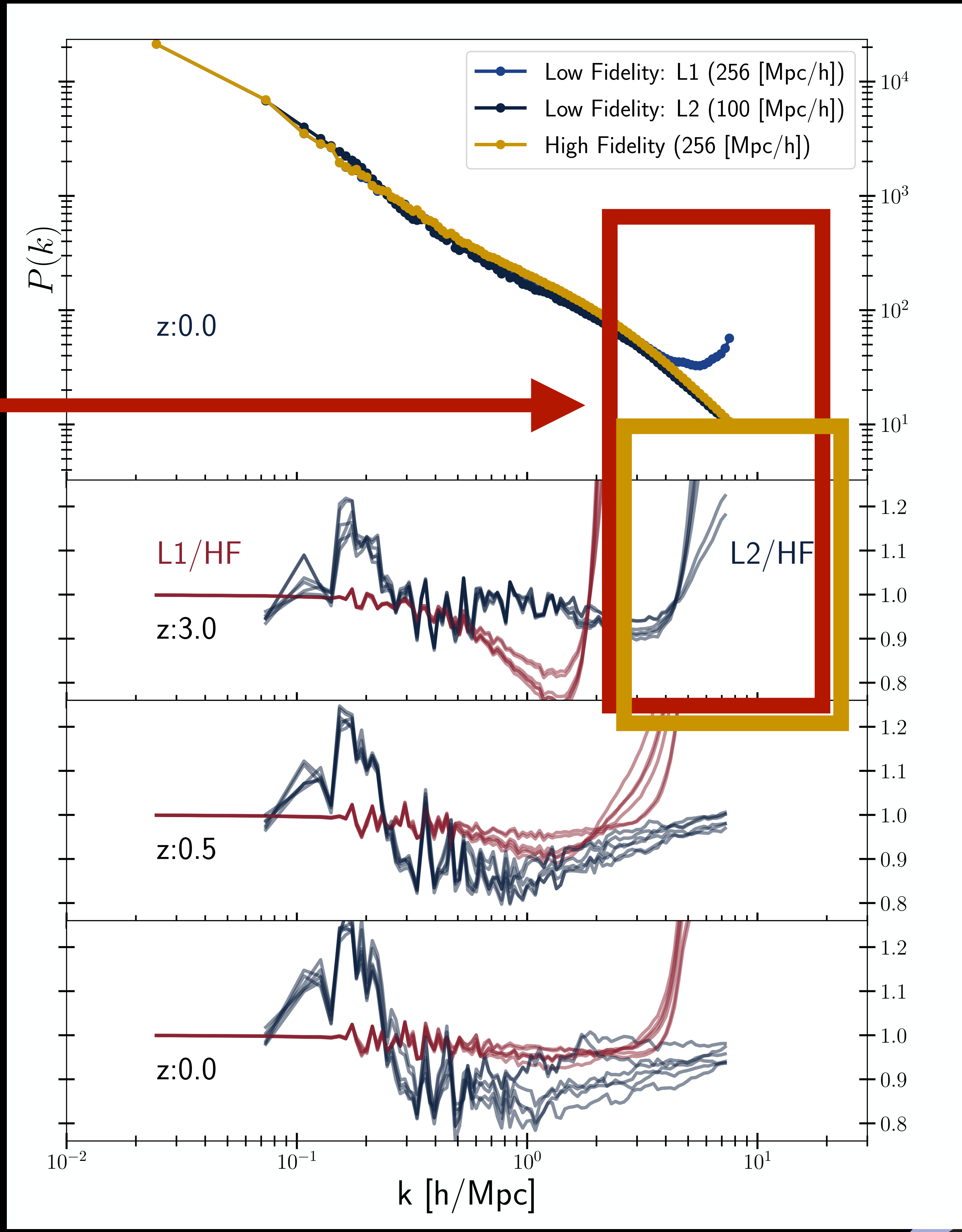
small ← large volume *small ← large volume*

MF-Box (arXiv:2306.03144)

Leverage accurate information from L2

Selecting optimal information and canceling systematics



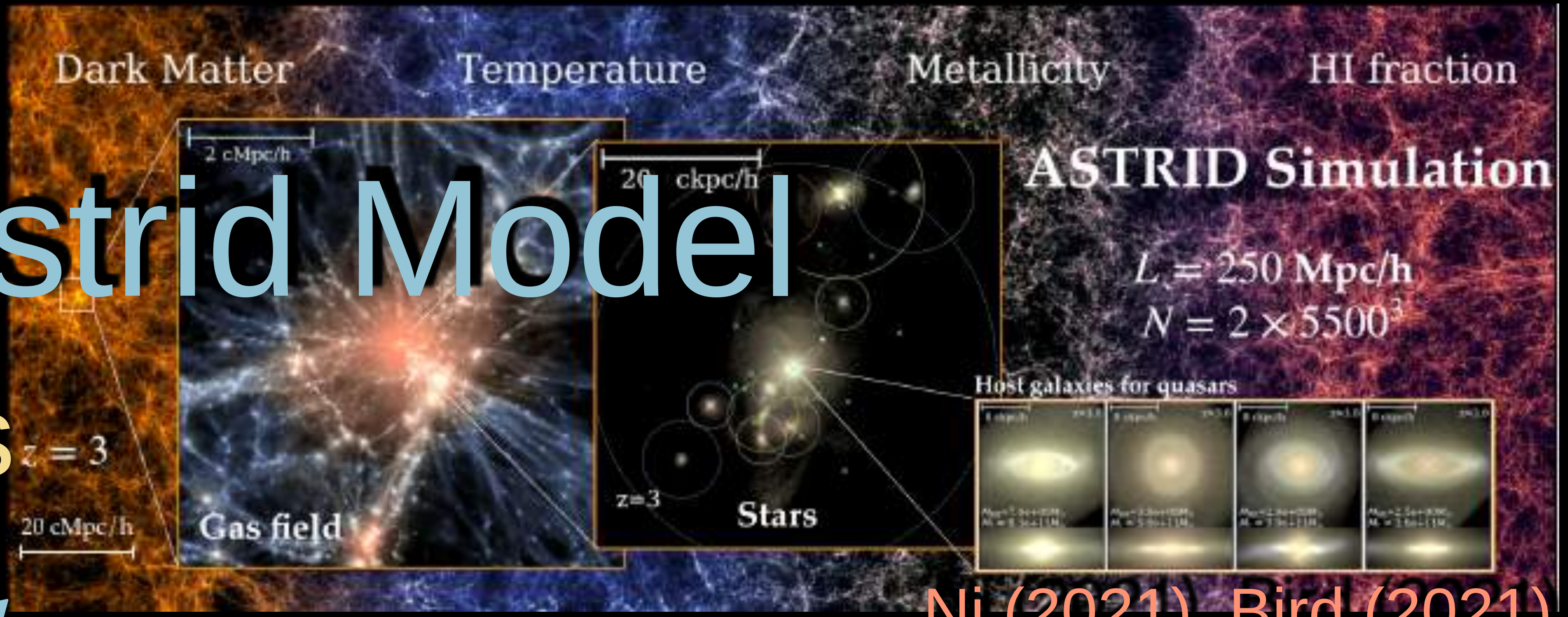


Shared systematics cannot be eliminated, but non-shared ones can be canceled

MF-Box (arXiv:2306.03144)

PRIYA: New *Lyα* Simulations from Astrid Model

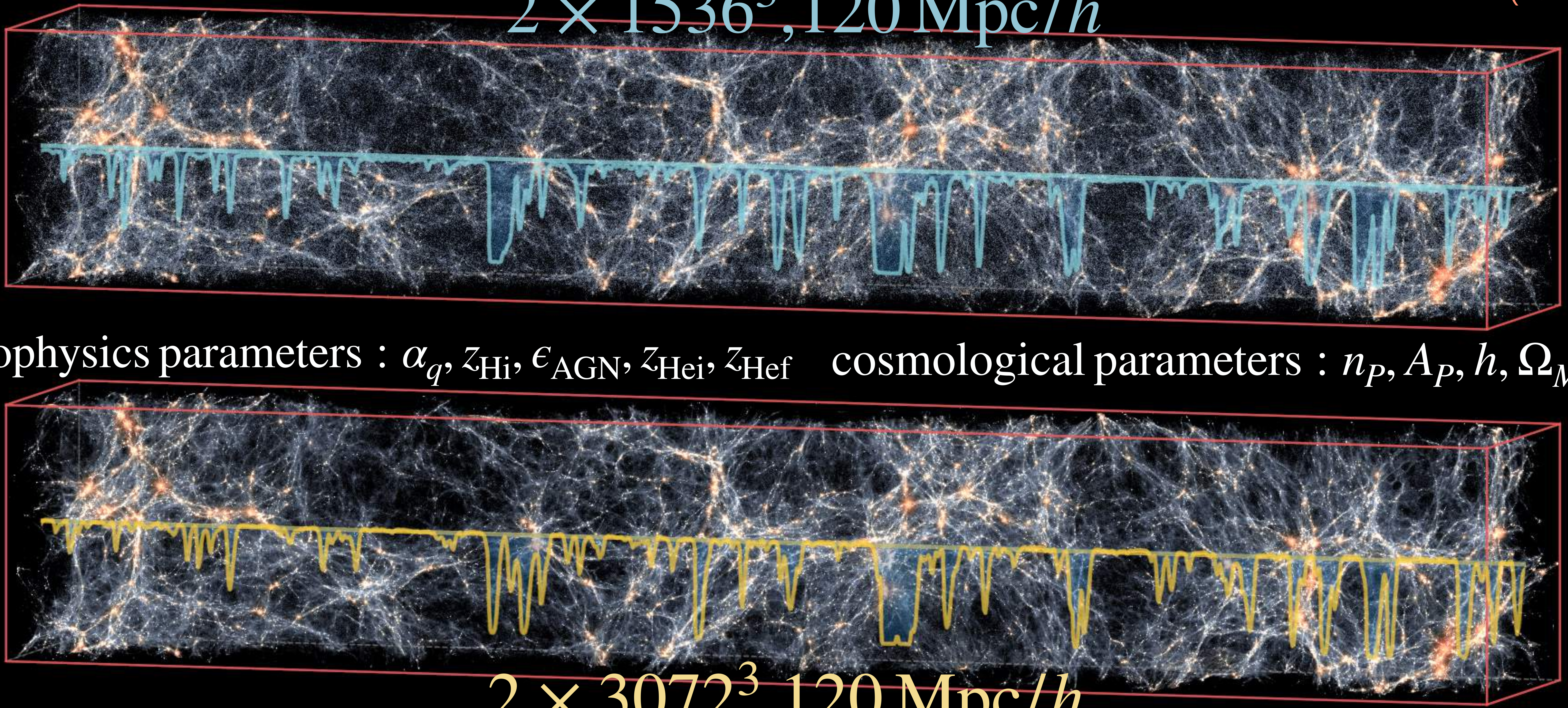
Multi-fidelity method for galaxy formation simulations



Ni (2021), Bird (2021)

$2 \times 1536^3, 120\ \mathrm{Mpc}/h$

LR



$\times 48$

astrophysics parameters : $\alpha_q, z_{\mathrm{Hi}}, \epsilon_{\mathrm{AGN}}, z_{\mathrm{Hei}}, z_{\mathrm{Hef}}$ cosmological parameters : $n_P, A_P, h, \Omega_M h^2$

HR

$\times 3$

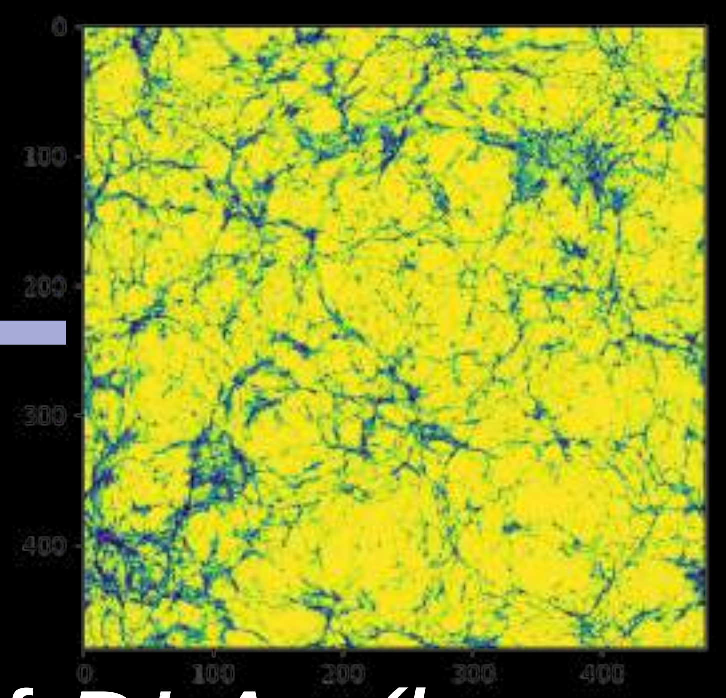
$2 \times 3072^3, 120\ \mathrm{Mpc}/h$

Bird 2023, arXiv:2306.05471

PRIYA: New *Lya* Simulations with realistic DLAs

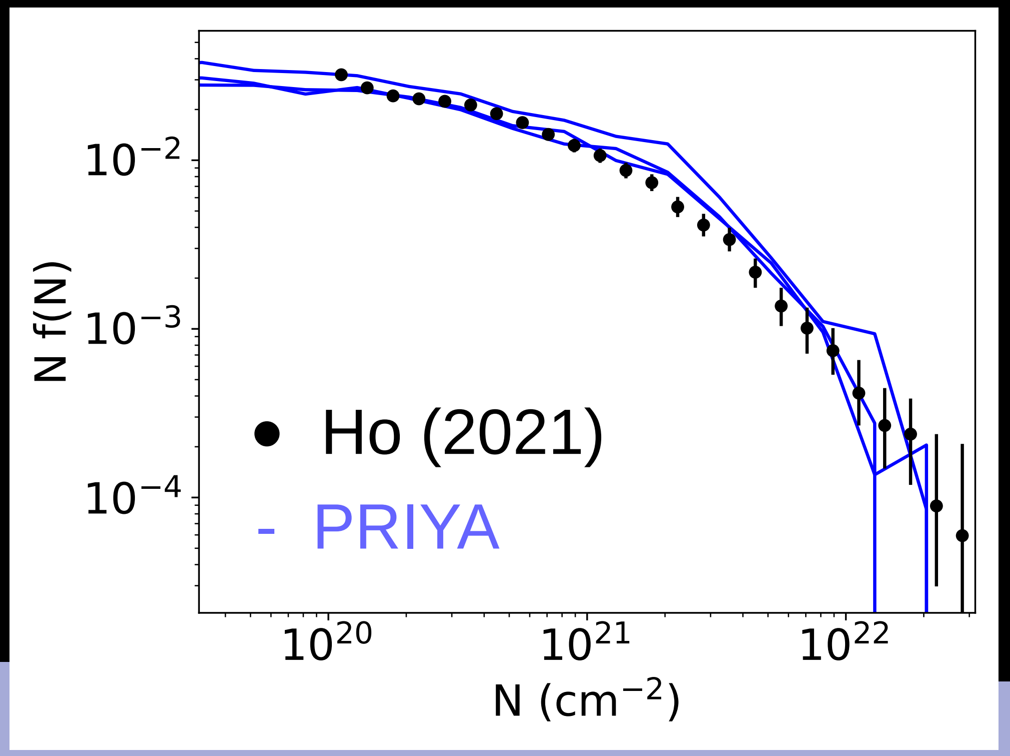
Mock *Lya* observations with realistic damped Lyman-α absorber population

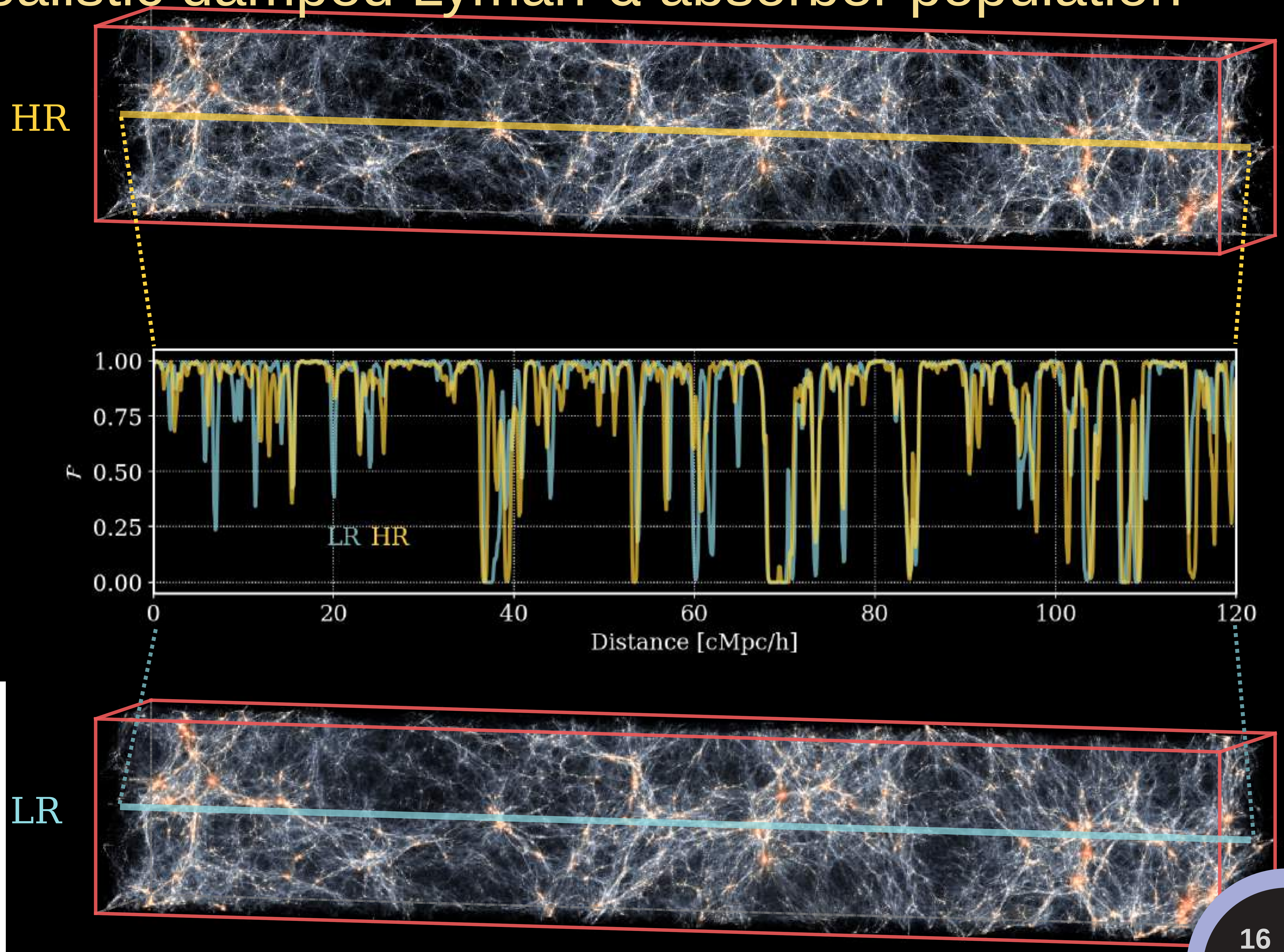
- 480 x 480 x 3 very dense slightlines in grids to beat down the noise:



Directly plot Lyα absorptions in 2D. Lots of skewers!

- Realistic population of *DLAs (known contamination of 1D flux)*, match SDSS DR16's CDDF (Ho 2021):



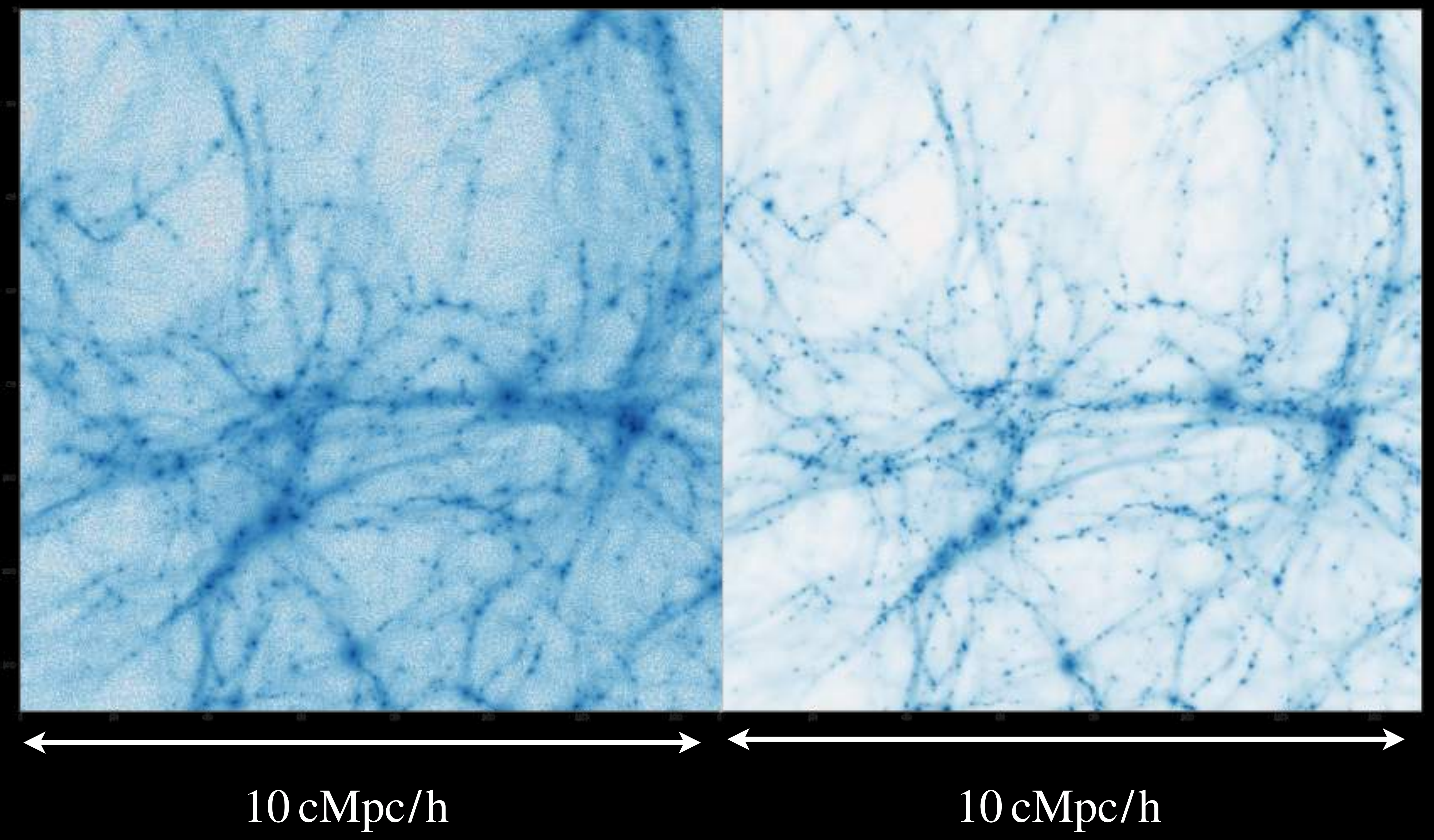


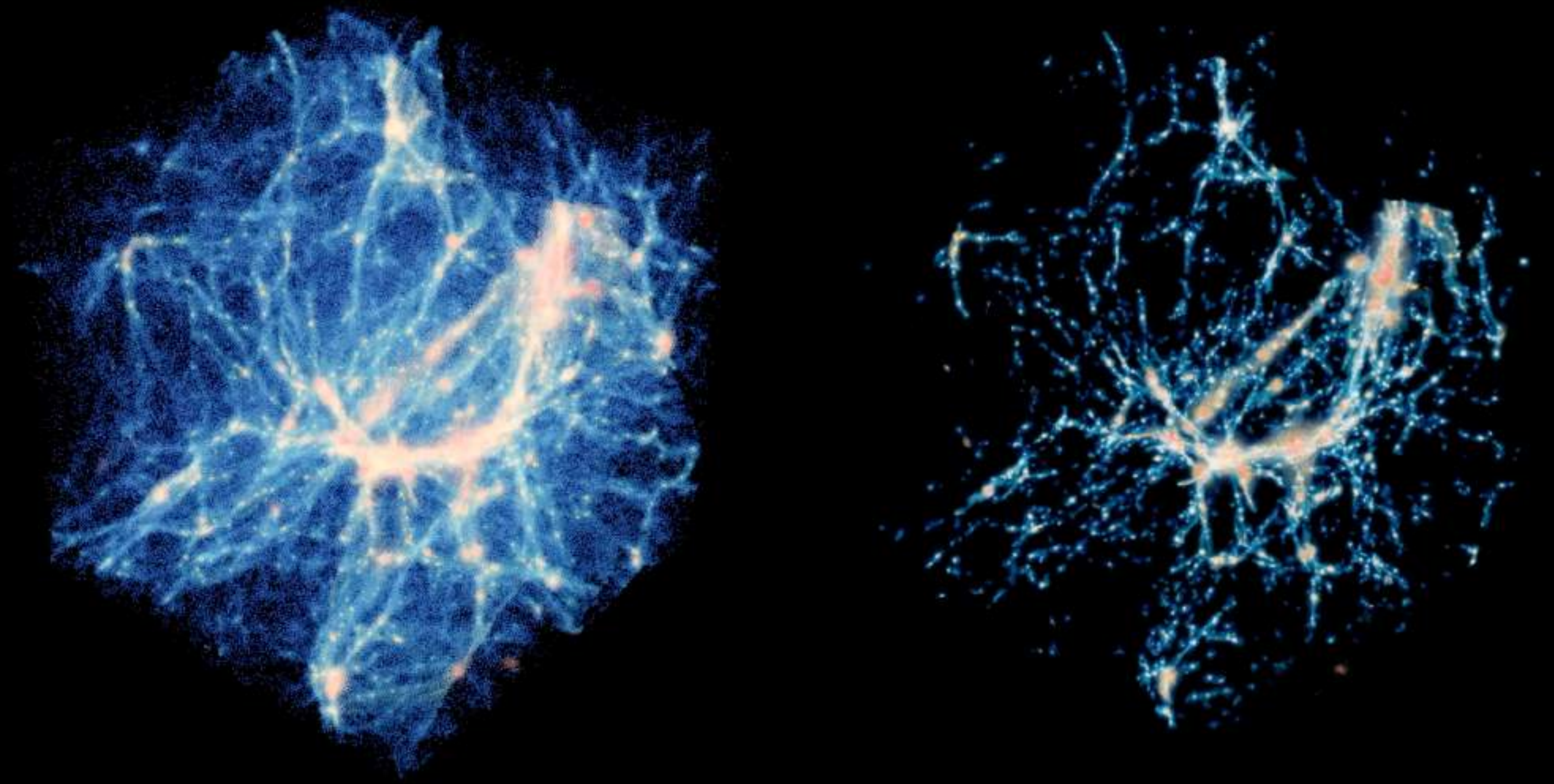
Bird 2023, arXiv:2306.05471

Subtle differences at small scales

A small volume comparison

LF *HF*

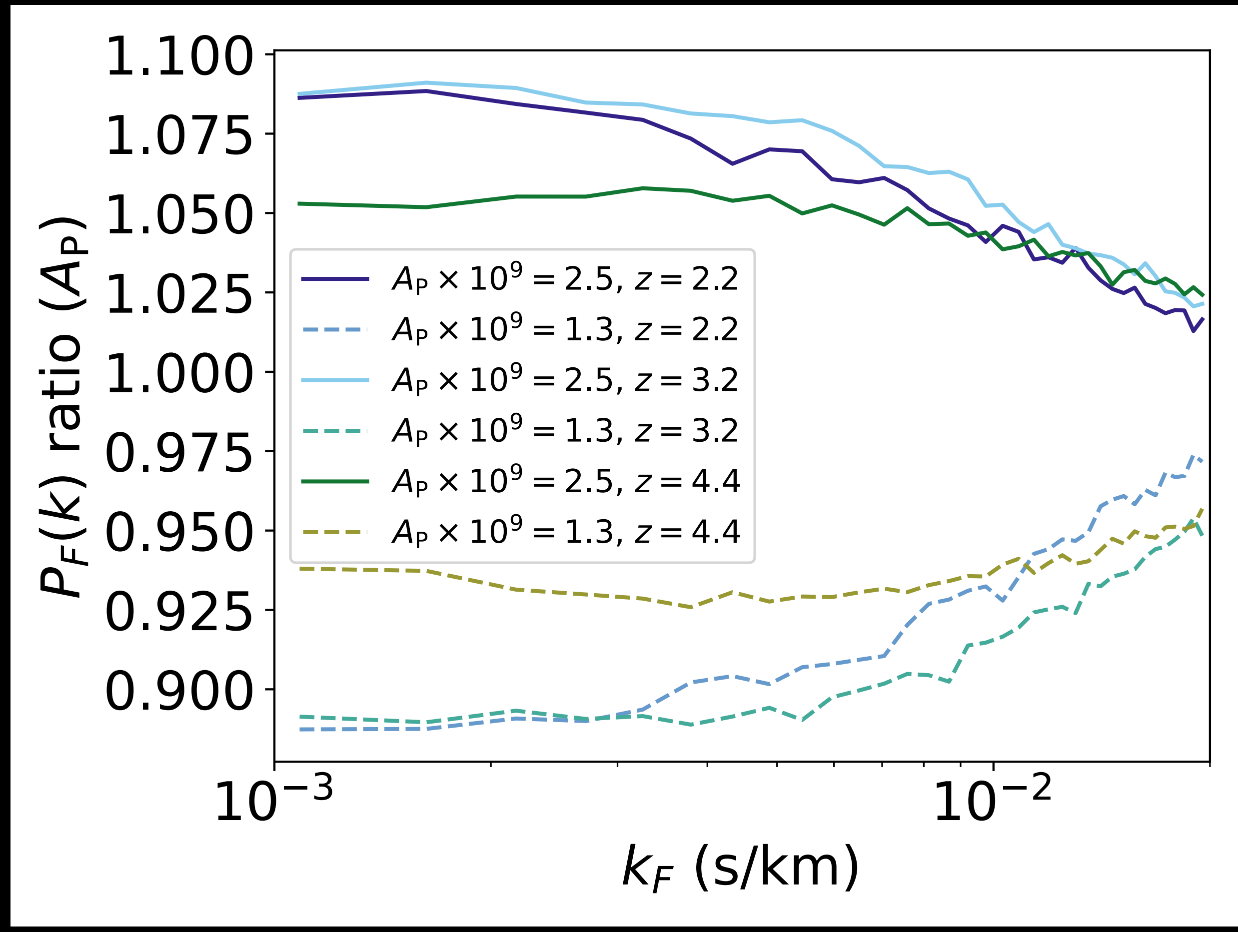




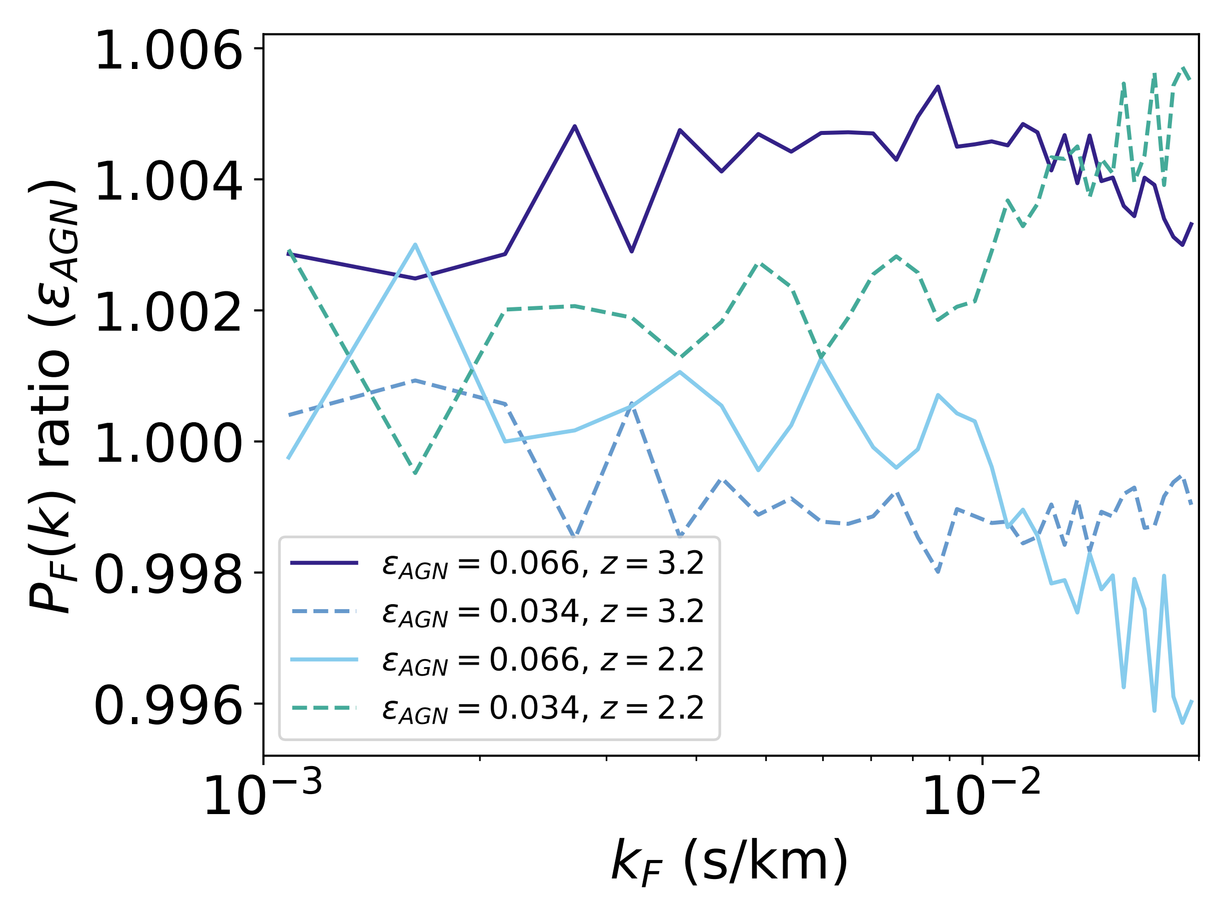
- Multi-fidelity method (without MF-Box) corrects the resolution up to 1% on 1D flux P(θ) (Bird 2023, arXiv:2306.05471)

Predicted 1D flux $P(\theta)$ from MFEmulator

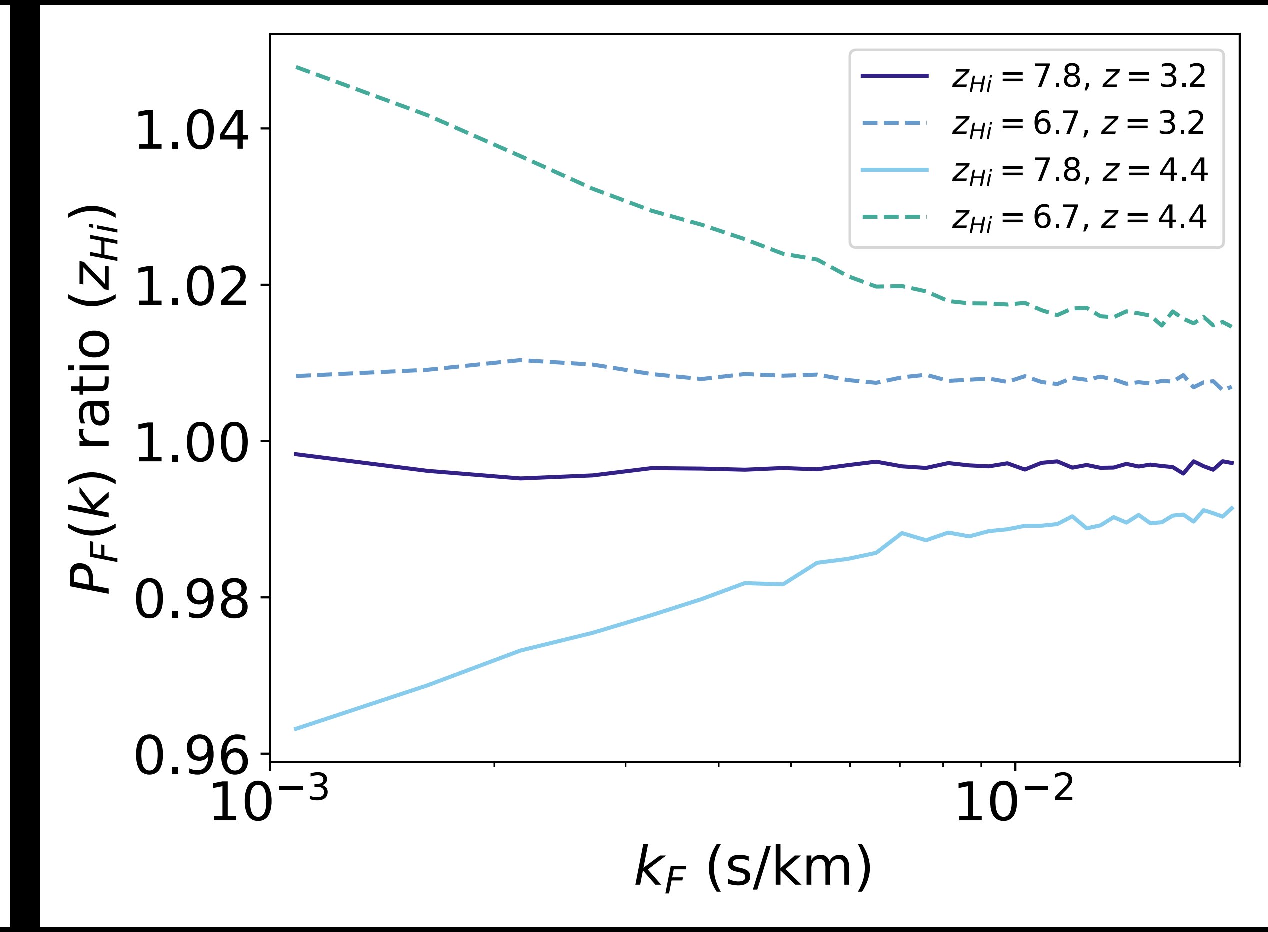
Include variations of astrophysical feedback/reionization parameters



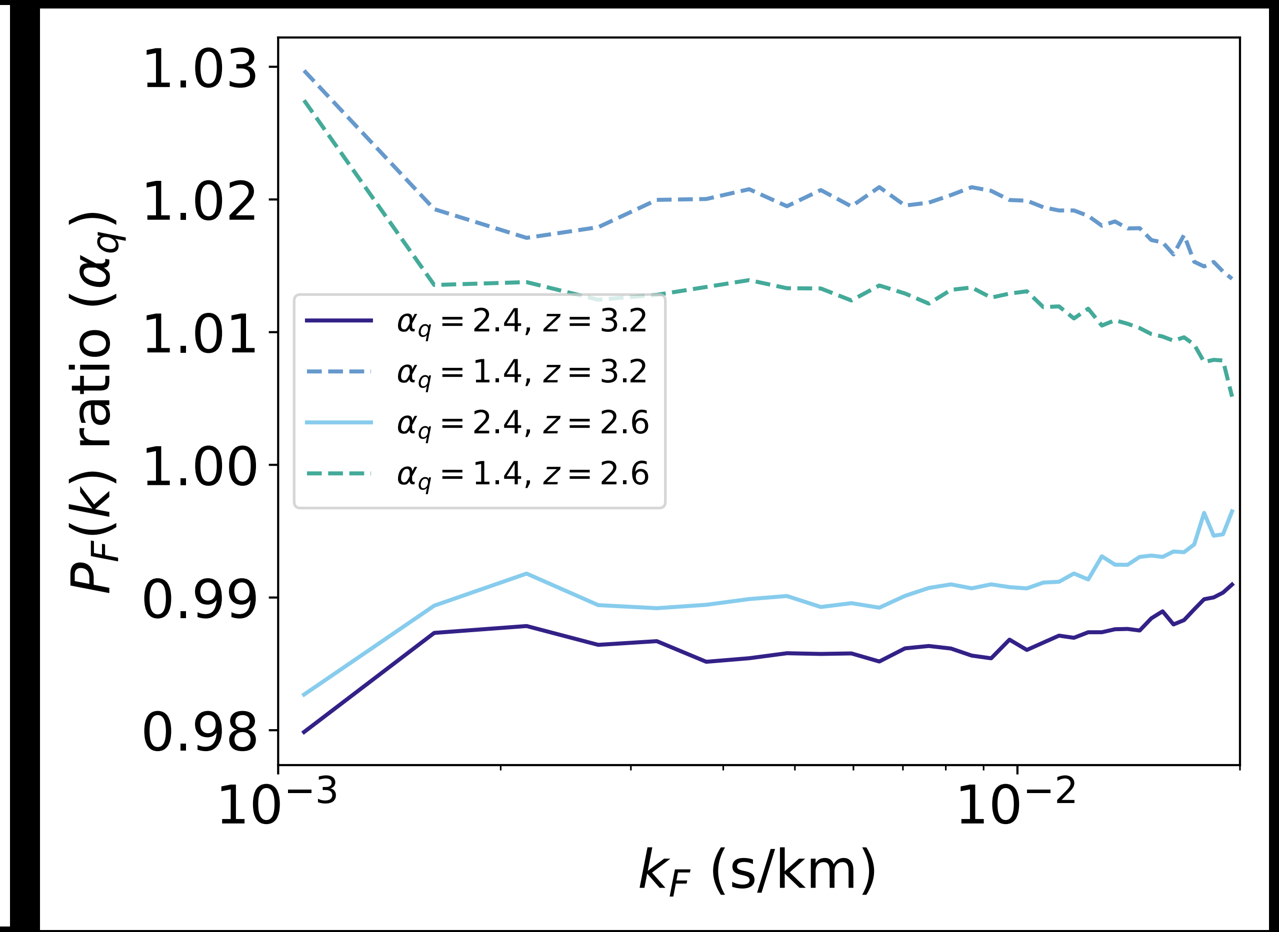
AGN Feedback
(Ni (2022))



HI Patchy Reionization
(Battaglia (2013))



Inhomogeneous HeII Reionization
(Upton Sanderbeck & Bird (2020))



$$P(k) = A_P \left(\frac{k}{0.78\ \mathrm{h/Mpc}} \right)^{n_P - 1}$$

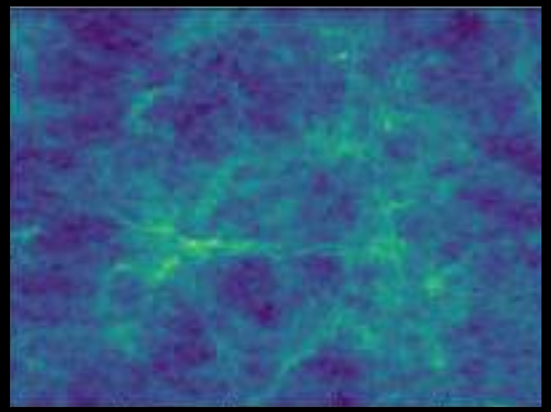
$$\dot{M}_{\mathrm{BH}} = \frac{4\pi\alpha G^2 M_{\mathrm{BH}}^2 \rho}{(c_s^2 + v_{\mathrm{rel}}^2)^{3/2}}$$

$$E_{\mathrm{AGN}} = \frac{\epsilon_{AGN}}{10} \dot{M}_{\mathrm{BH}} c^2$$

z_{Hi} : redshift of HI reionization

α_q : Spectral index of quasar radiation field inside a bubble

IGM temperature boosted to 15,000 K, observational constraint at $z \sim 6$ (Gaikwad 2020)



- Check Simeon Bird's talk, "***New Cosmological Analysis of the eBOSS Lyman-α Forest***," (July 6, late session) for the cosmological inference results using PRIYA simulation suite.

Conclusion

MF-Box and PRIYA emulator

- MF-Box: A multi-fidelity multi-scale tool, combining f(resolution, scale, θ) using Bayesian approach.
- MF-Box: Optimizing budget allocation, a decision tool for predicting budgets and writing proposals.
- PRIYA: A multi-resolution galaxy formation simulation suite for Lyα forest
- PRIYA: A big emulator varies astrophysical effects- AGN feedback, reionization, and cosmology
- MFEmulator/MF-Box: Expanding beyond cosmic emulation, serving as surrogates for galaxy formation simulations
- Chat with us to use MF-Box or PRIYA simulations!

MF-Box (arXiv:2306.03144)
PRIYA (arXiv:2306.05471)

Ming-Feng Ho (jibancat.github.io)
Simeon Bird (sbird.github.io)
Astrid simulation code (MP-Gadget)

19