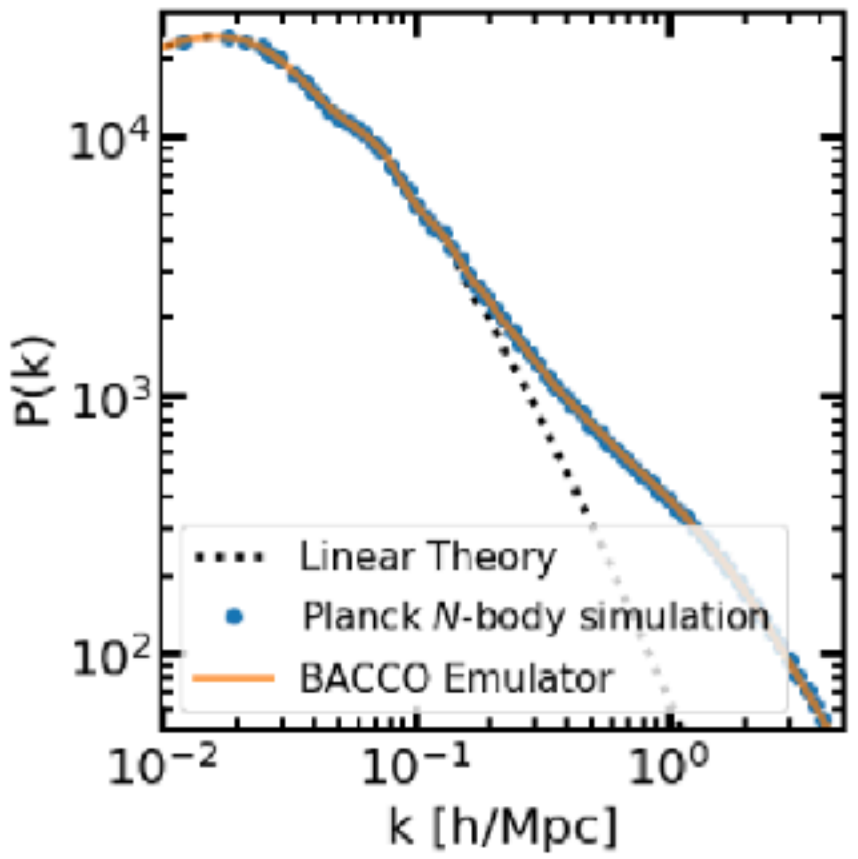
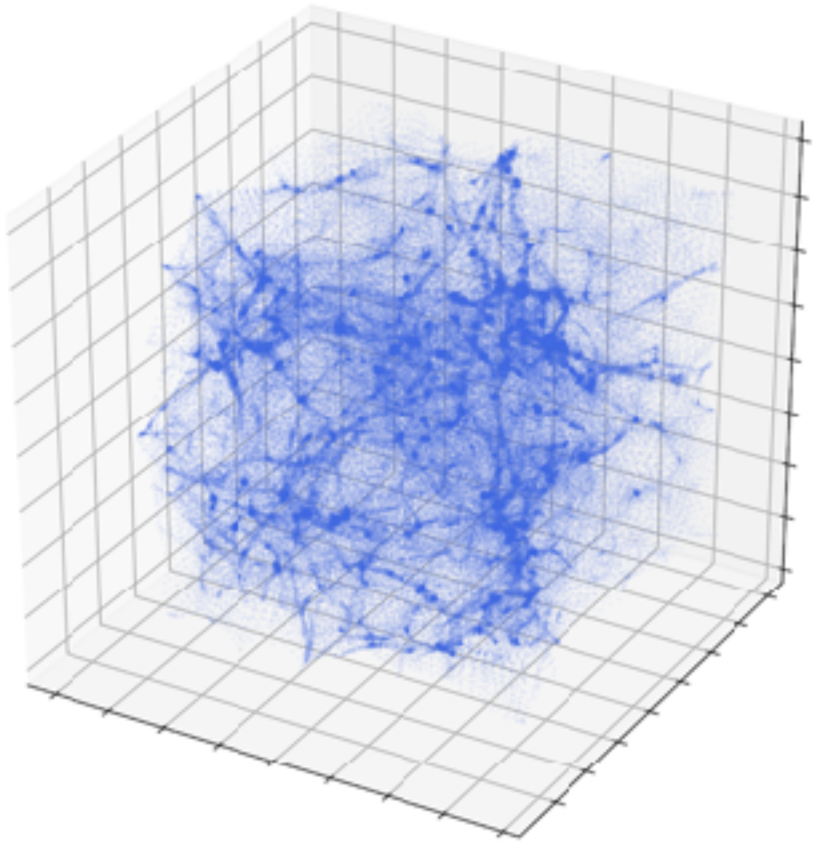
ML in Cosmology

1st FAIRS-Japan, Dec 3 2024

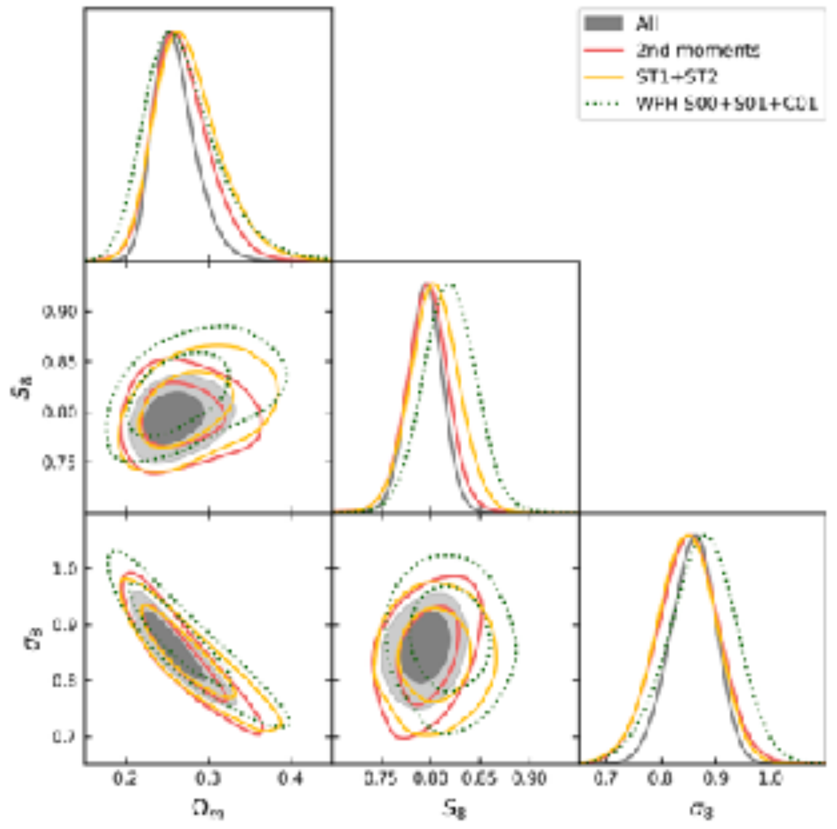
Leander Thiele (Kavli IPMU)



Interpolation



Emulation



Inference

generality

power

simplicity

robustness

Interpolation

Interpolation

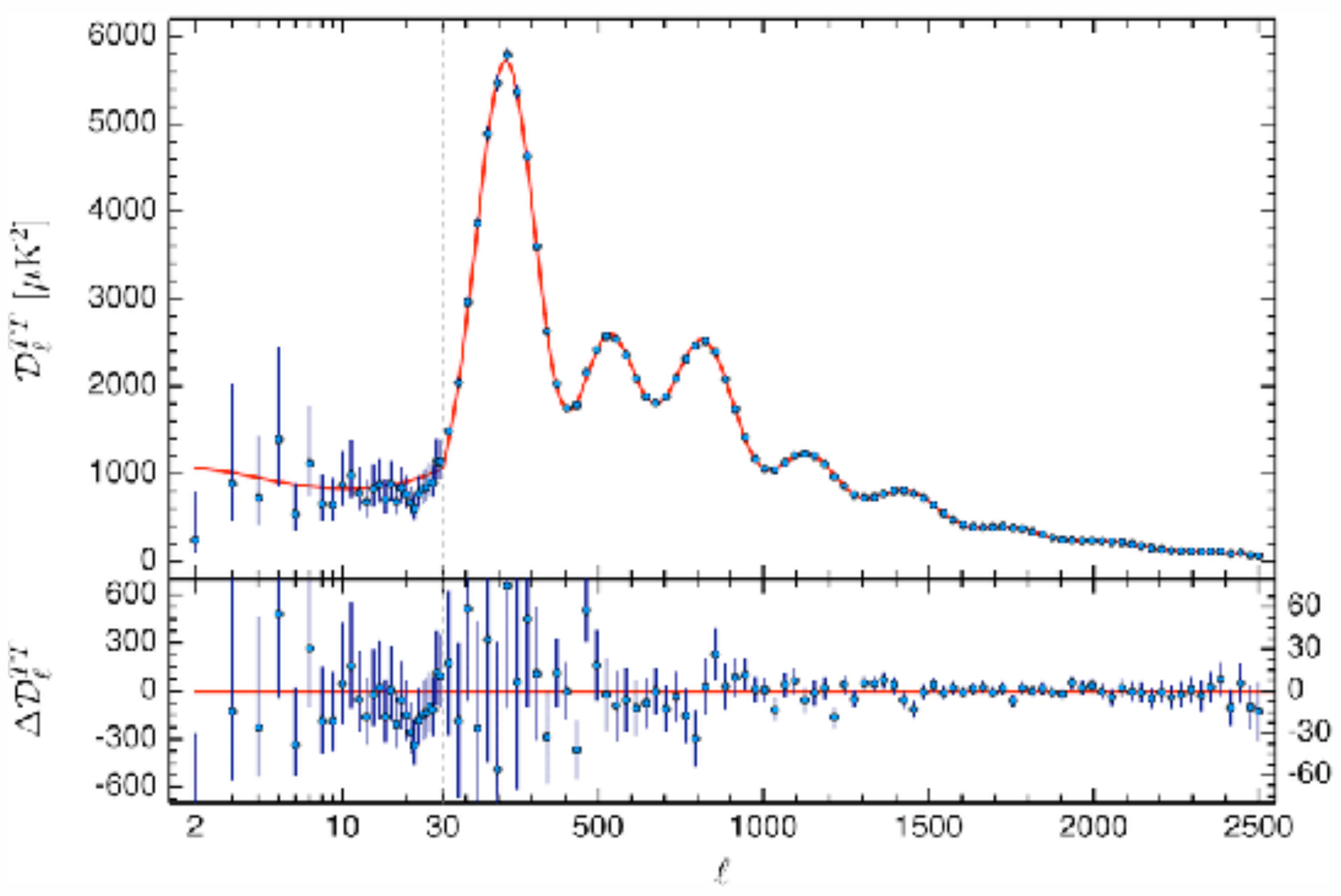
traditional Gaussian likelihood situation:

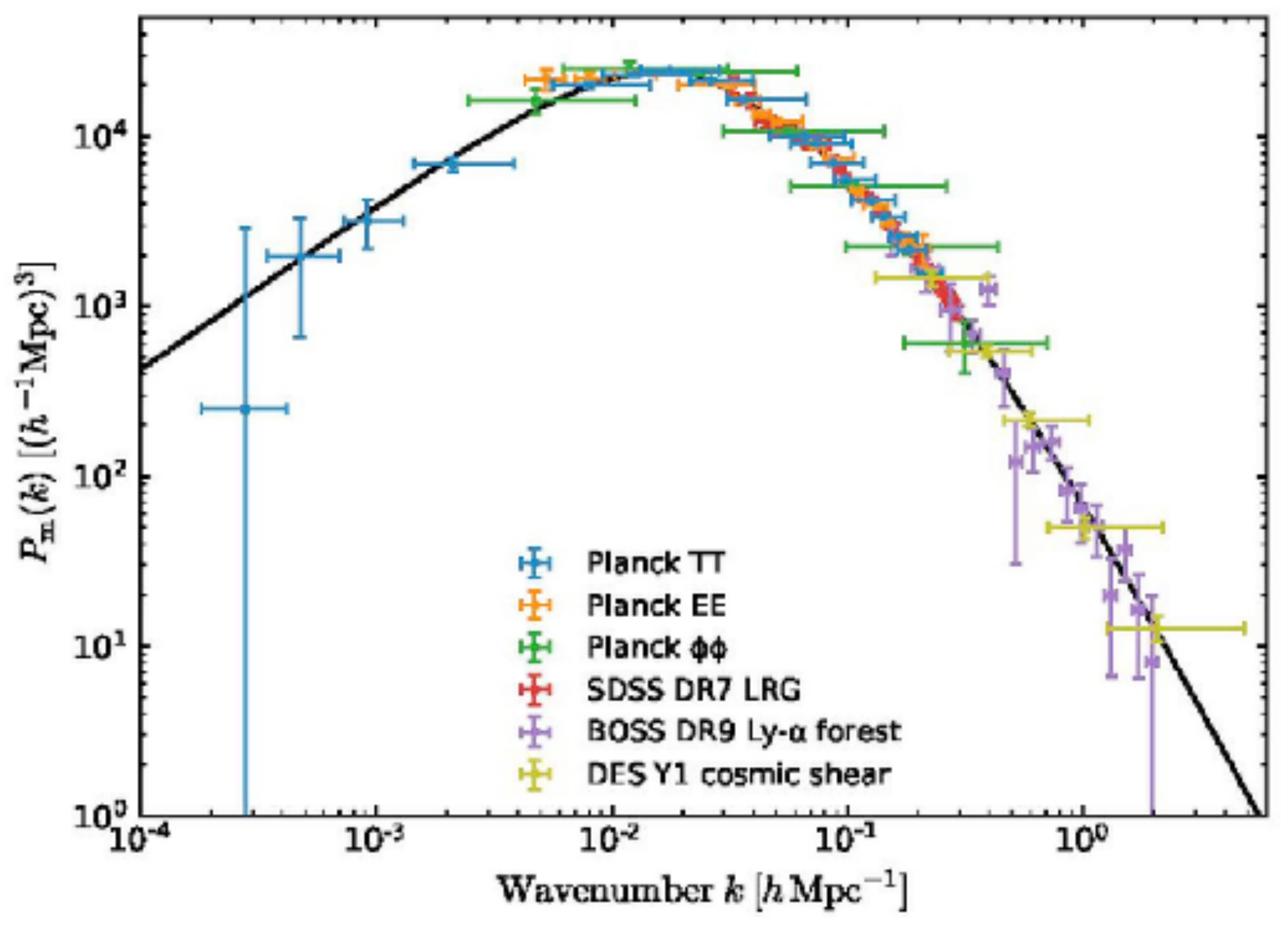
$$-2\log P(x|\theta) = [x - S(\mu(\theta)]^T \Sigma^{-1} [x - S(\mu(\theta)]$$

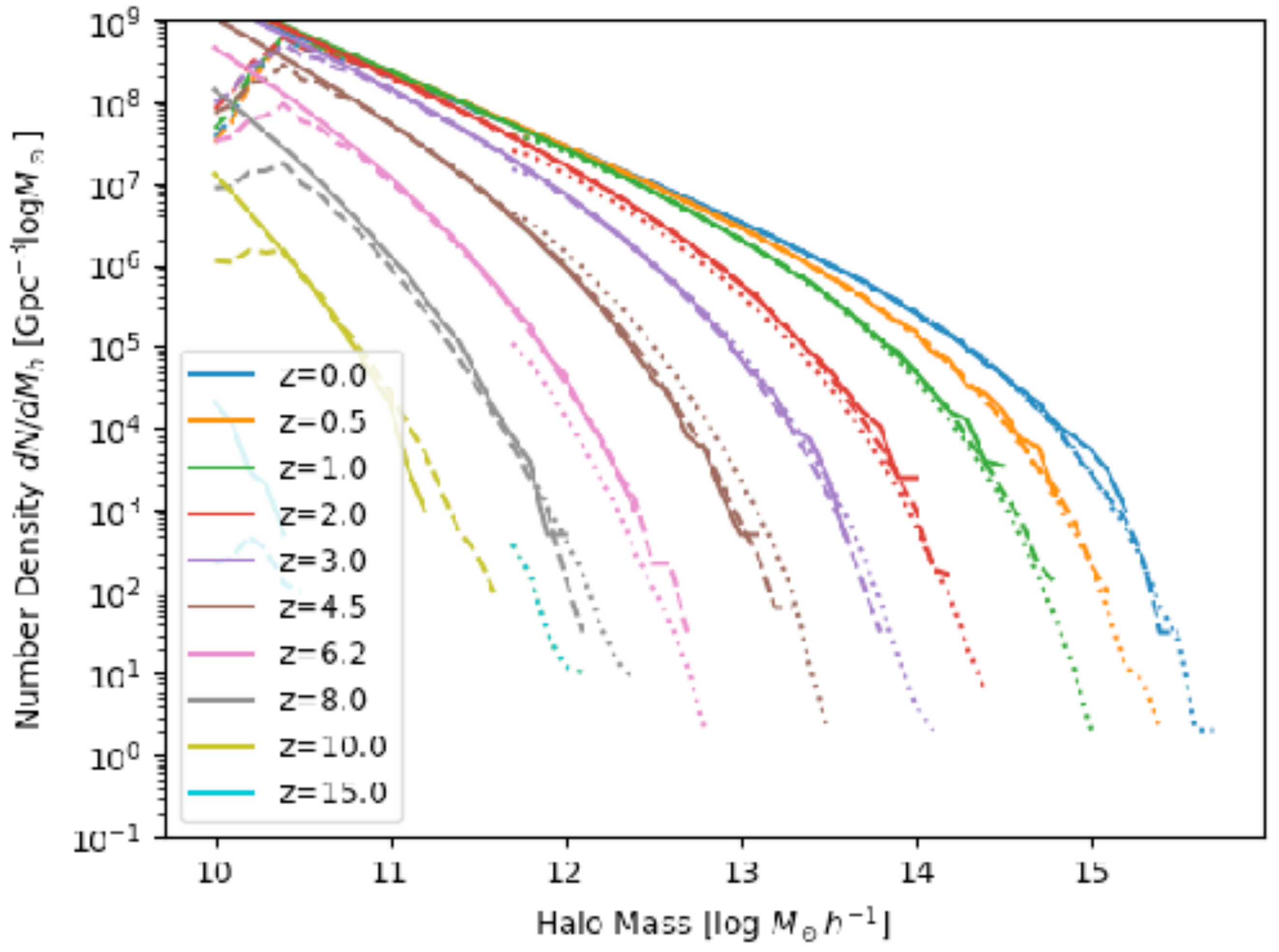
cheap, specific to experiment

- CMB beam
- lensing kernel
- selection function
- …

"expensive", universal

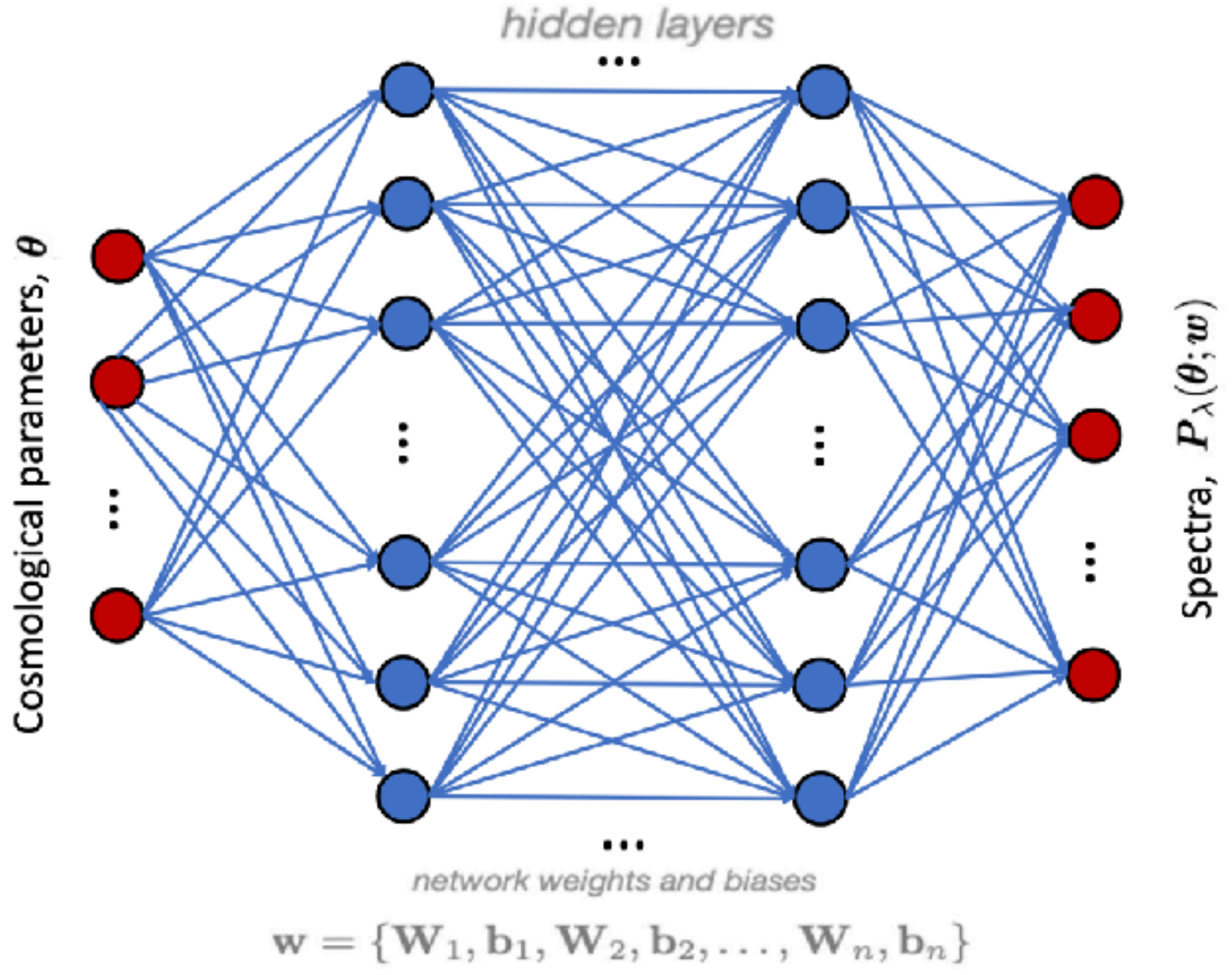




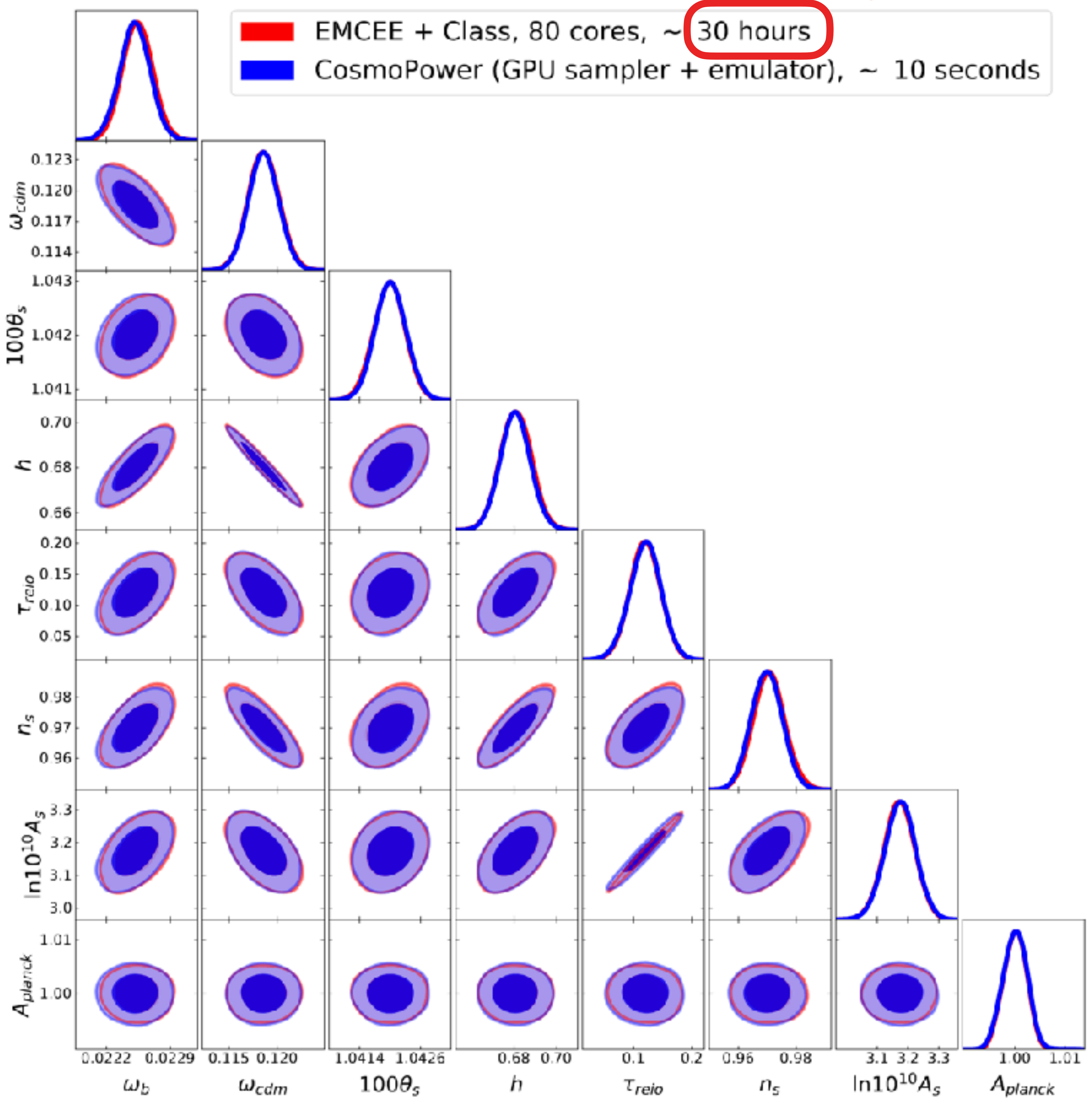


Interpolating "analytic" codes

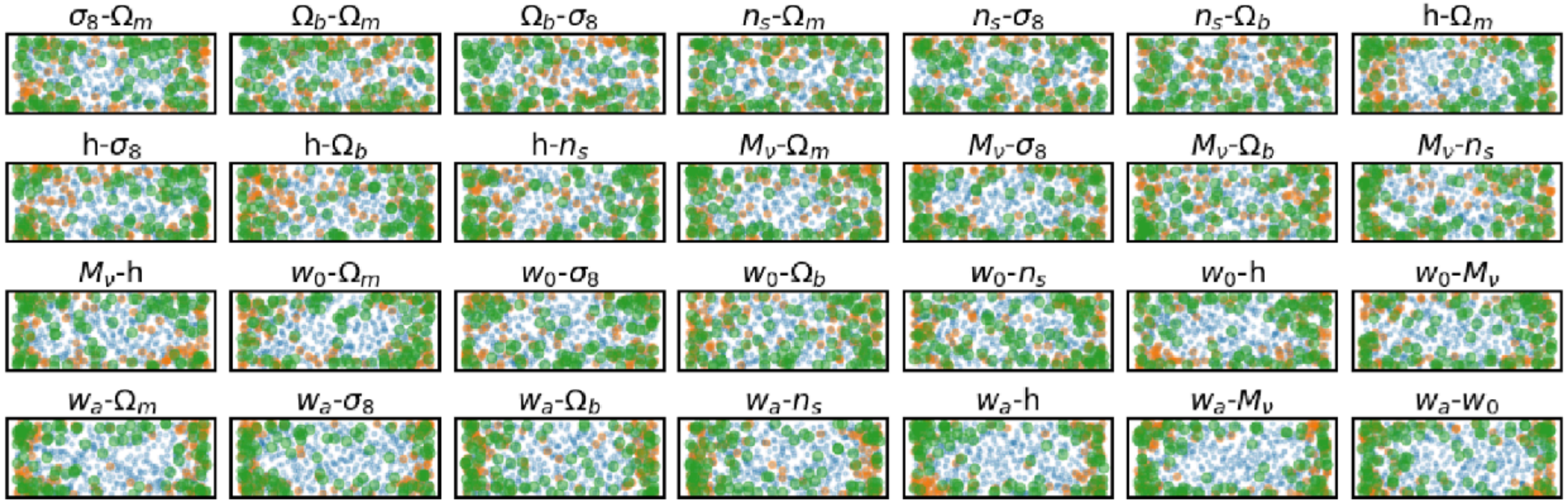
Cosmopower
[A. Spurio Mancini et al 2022]

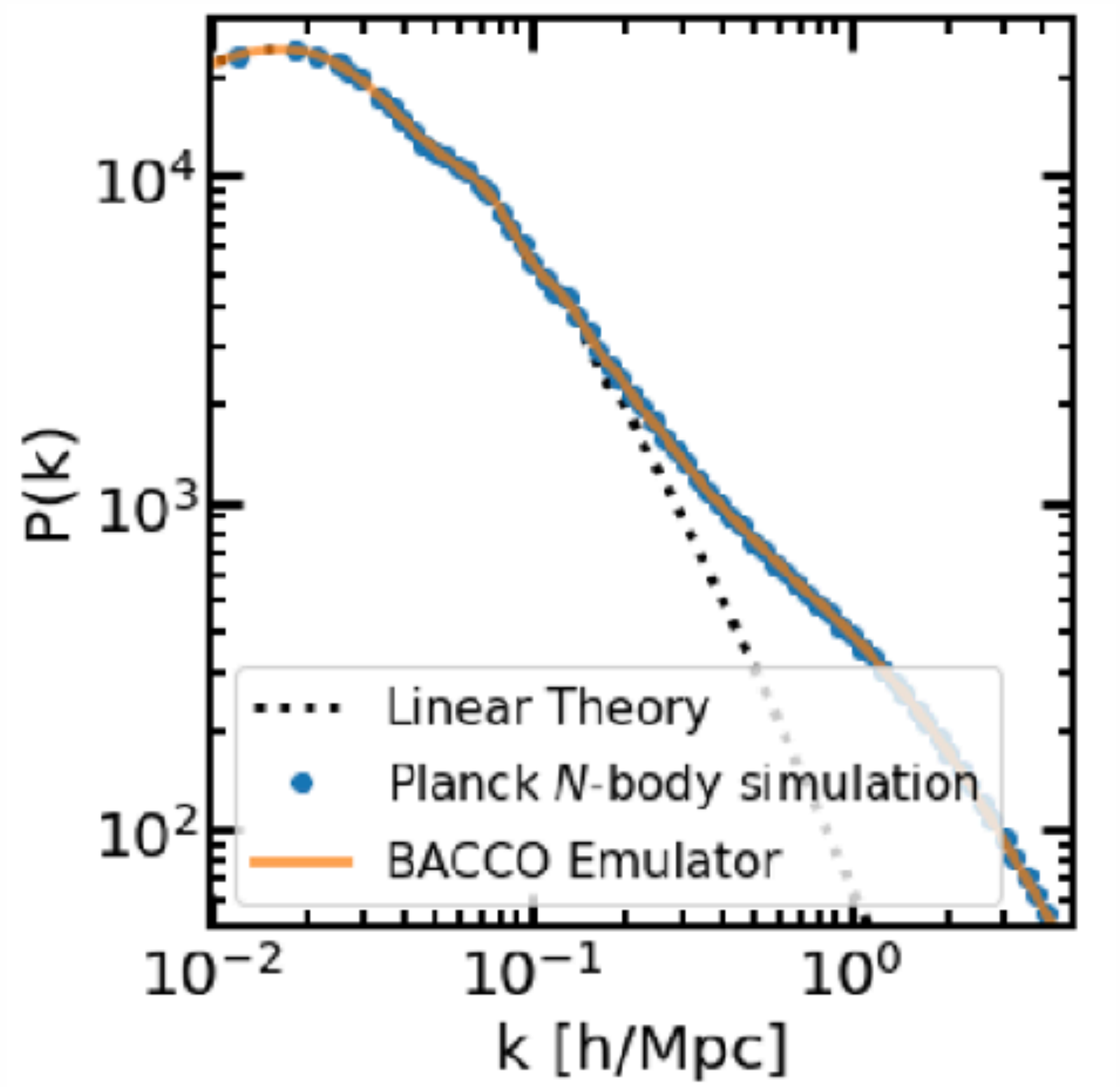


usually weeks

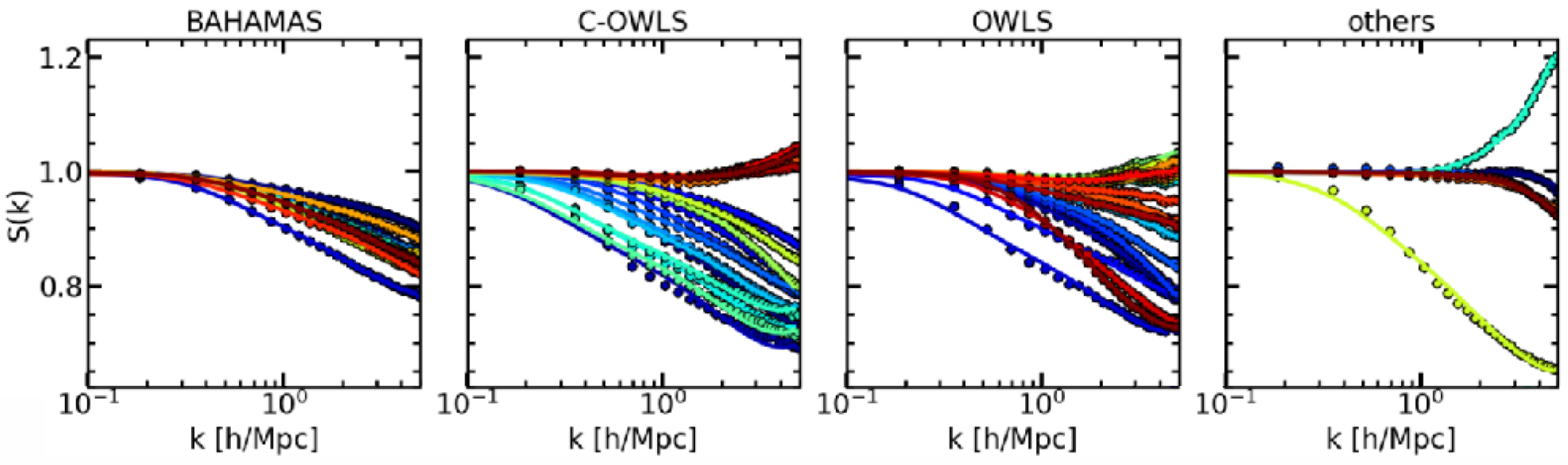


Interpolating simulations: matter

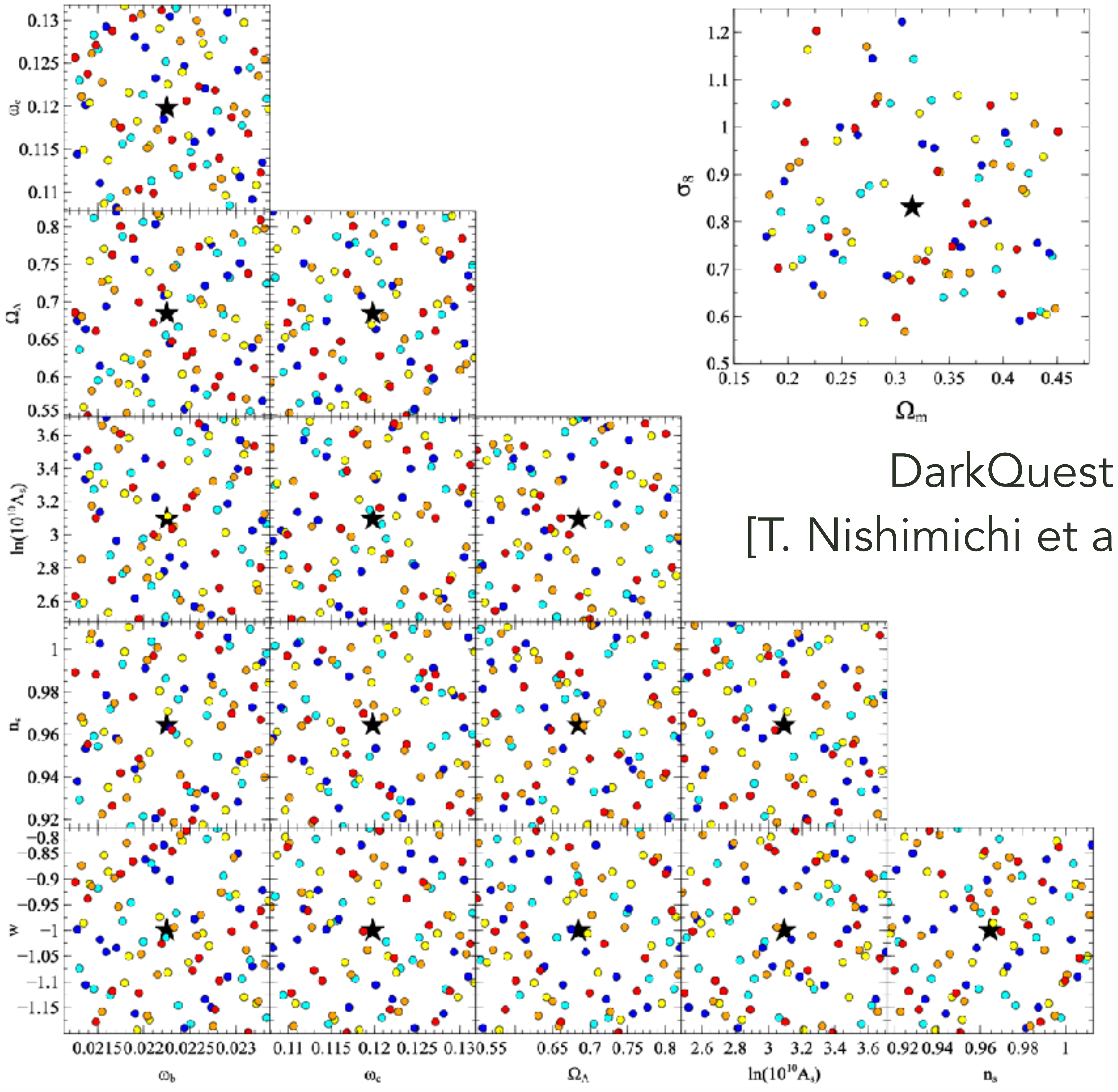




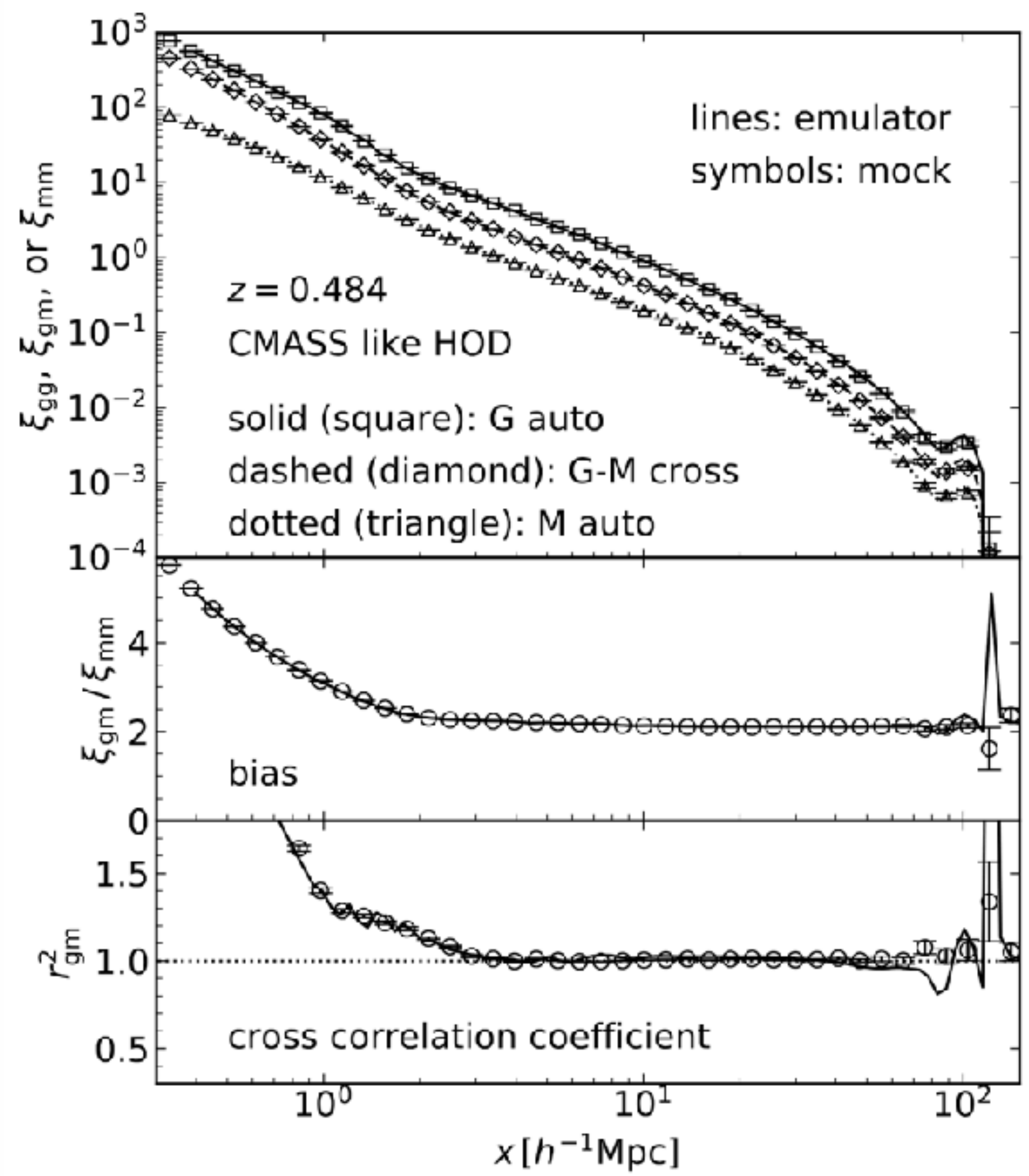
BACCO emulator
[R. Angulo et al 2020]



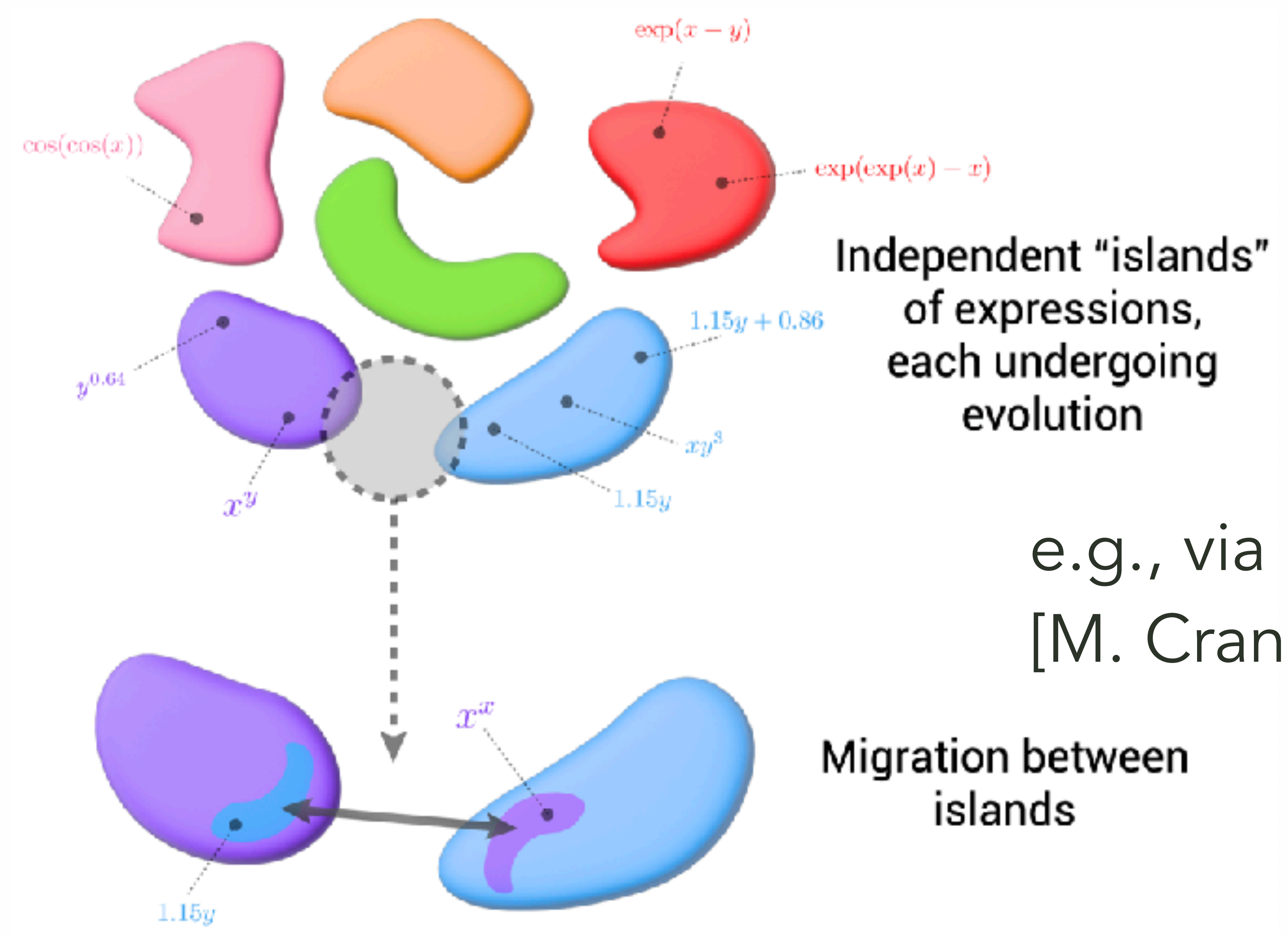
Interpolating simulations: halos & galaxies

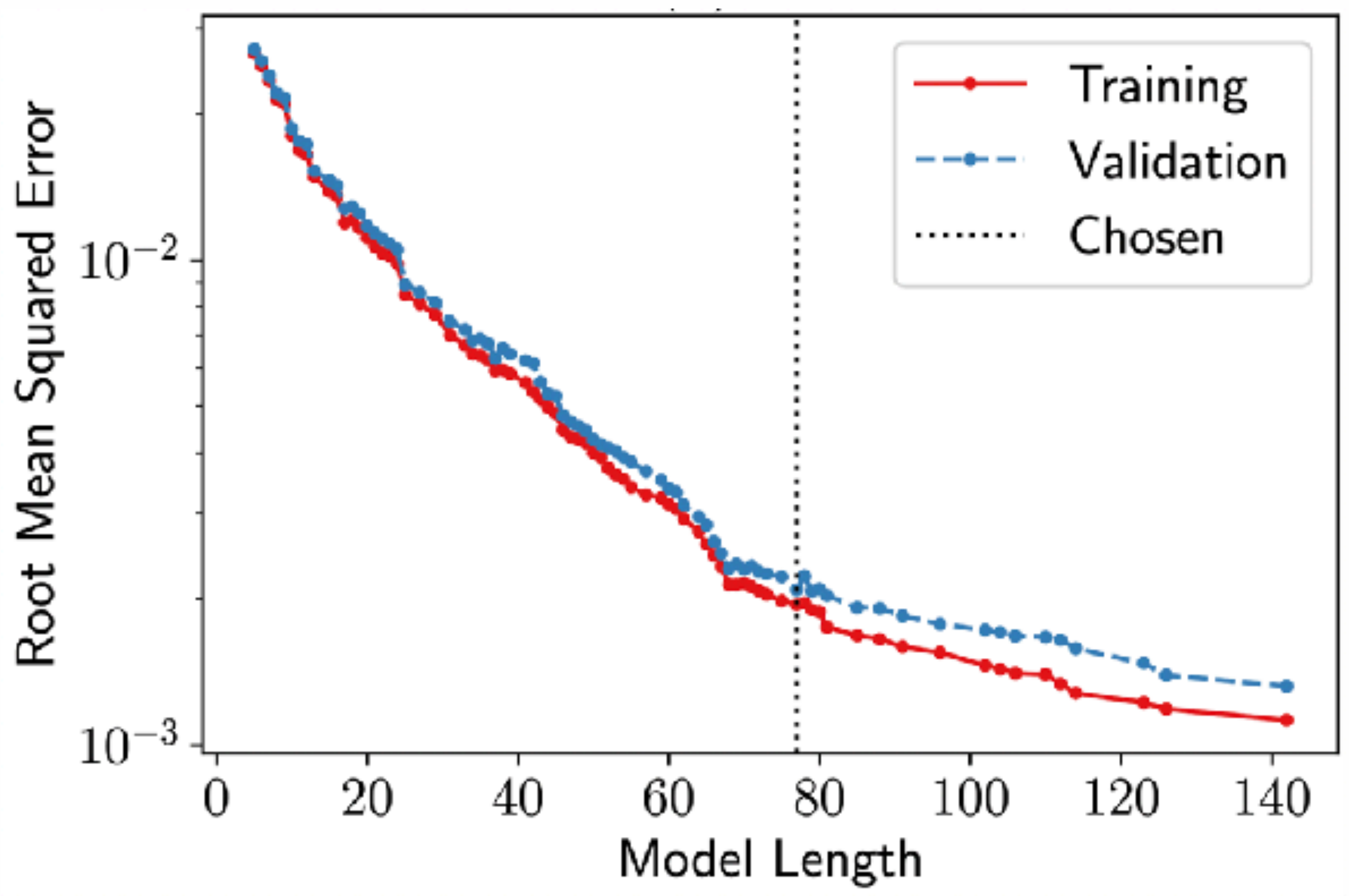


DarkQuest
[T. Nishimichi et al 2019]



Symbolic Regression

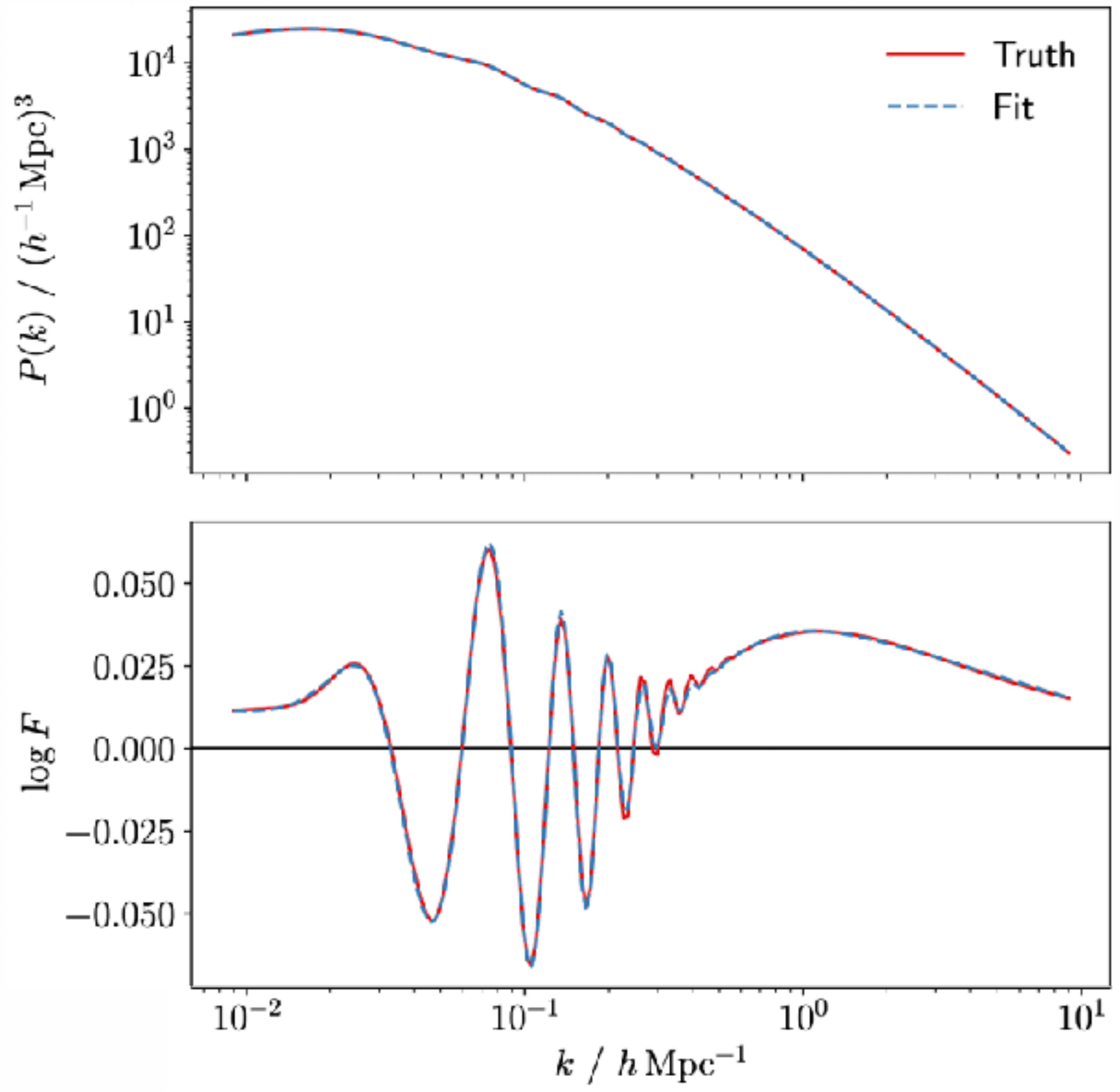




e.g., via genetic programming
[M. Cranmer 2023]

$$\log F \approx b_0 h - b_1 + \left(\frac{b_2\Omega_b}{\sqrt{h^2+b_3}}\right)^{b_4\Omega_m}\left[\frac{b_5 k - \Omega_b}{\sqrt{b_6+(\Omega_b - b_7 k)^2}}\, b_8 (b_9 k)^{-b_{10}k}\cos\left(b_{11}\Omega_m - \frac{b_{12}k}{\sqrt{b_{13}+\Omega_b^2}}\right) - b_{14}\left(\frac{b_{15}k}{\sqrt{1+b_{16}k^2}} - \Omega_m\right)\cos\left(\frac{b_{17}h}{\sqrt{1+b_{18}k^2}}\right)\right]$$
$$+ b_{19}\left(b_{20}\Omega_m + b_{21}h - \log(b_{22}k) + (b_{23}k)^{-b_{24}k}\right)\cos\left(\frac{b_{25}}{\sqrt{1+b_{26}k^2}}\right)$$
$$+ (b_{27}k)^{-b_{28}k}\left(b_{29}k - \frac{b_{30}\log(b_{31}k)}{\sqrt{b_{32}+(\Omega_m - b_{33}h)^2}}\right)\cos\left(b_{34}\Omega_m - \frac{b_{35}k}{\sqrt{b_{36}+\Omega_b^2}}\right),$$

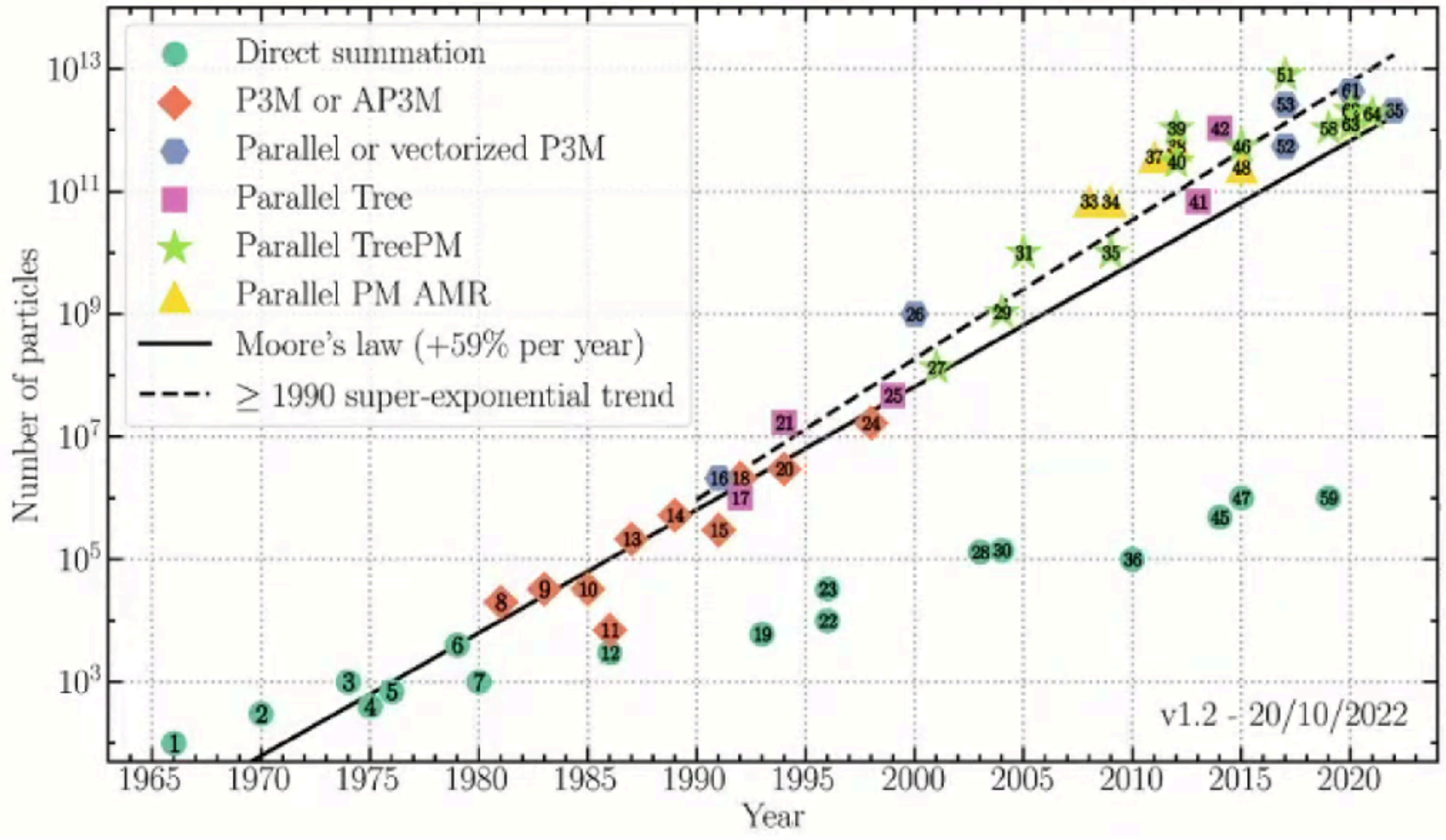
linear matter power spectrum
[D. Bartlett et al 2024]



Emulation

Emulation

- for upcoming surveys, need to simulate ~ 10^4 x (10 Gpc)^3
- currently largest "full physics" simulations: ~ (1 Gpc)^3 [and this is not truly full physics]



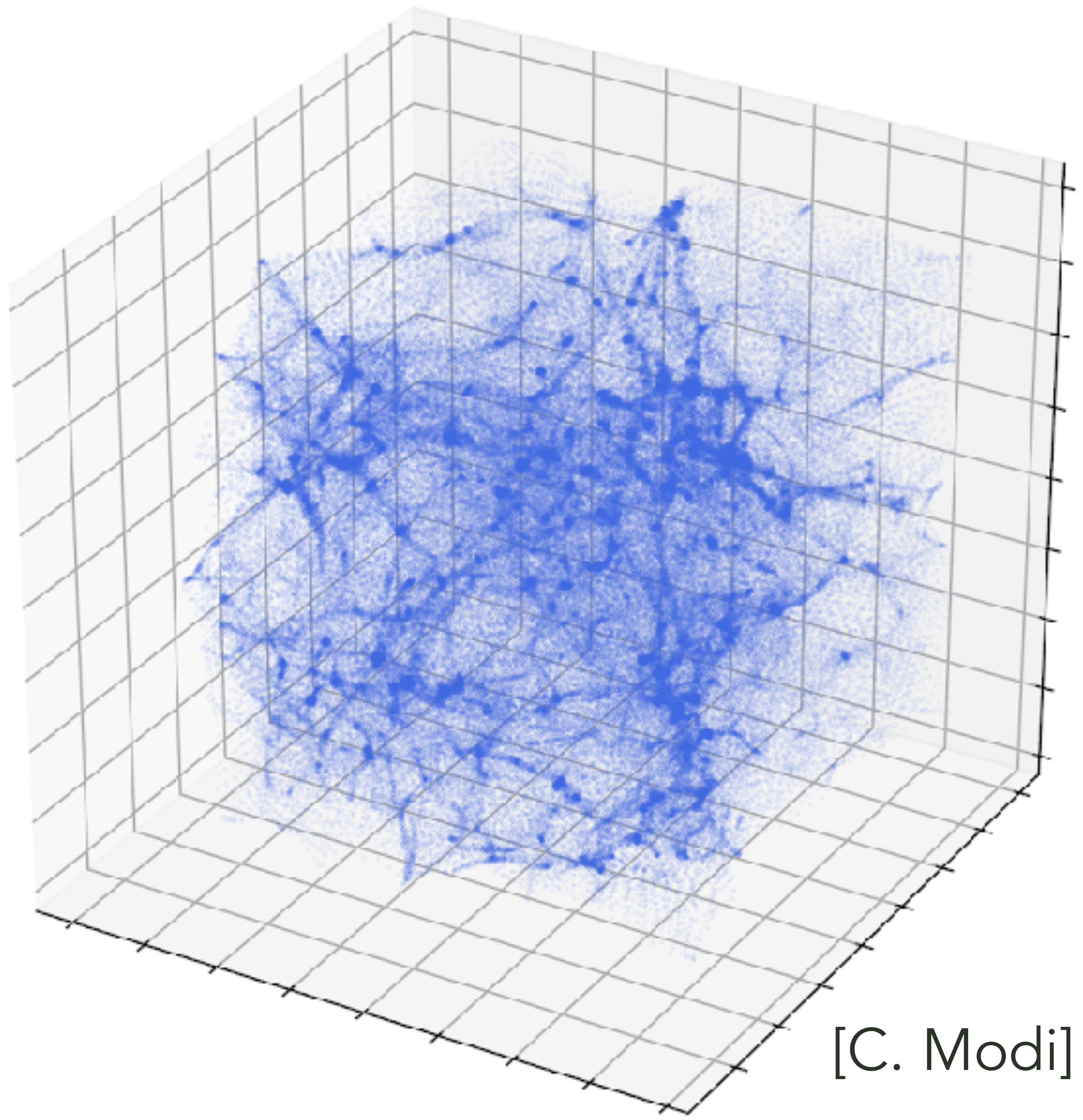
Moore's law won't help us

[F. Leclerc]

Emulation

- for upcoming surveys, need to simulate ~ 10^4 x (10 Gpc)^3
- currently largest "full physics" simulations: ~ (1 Gpc)^3 [and this is not truly full physics]

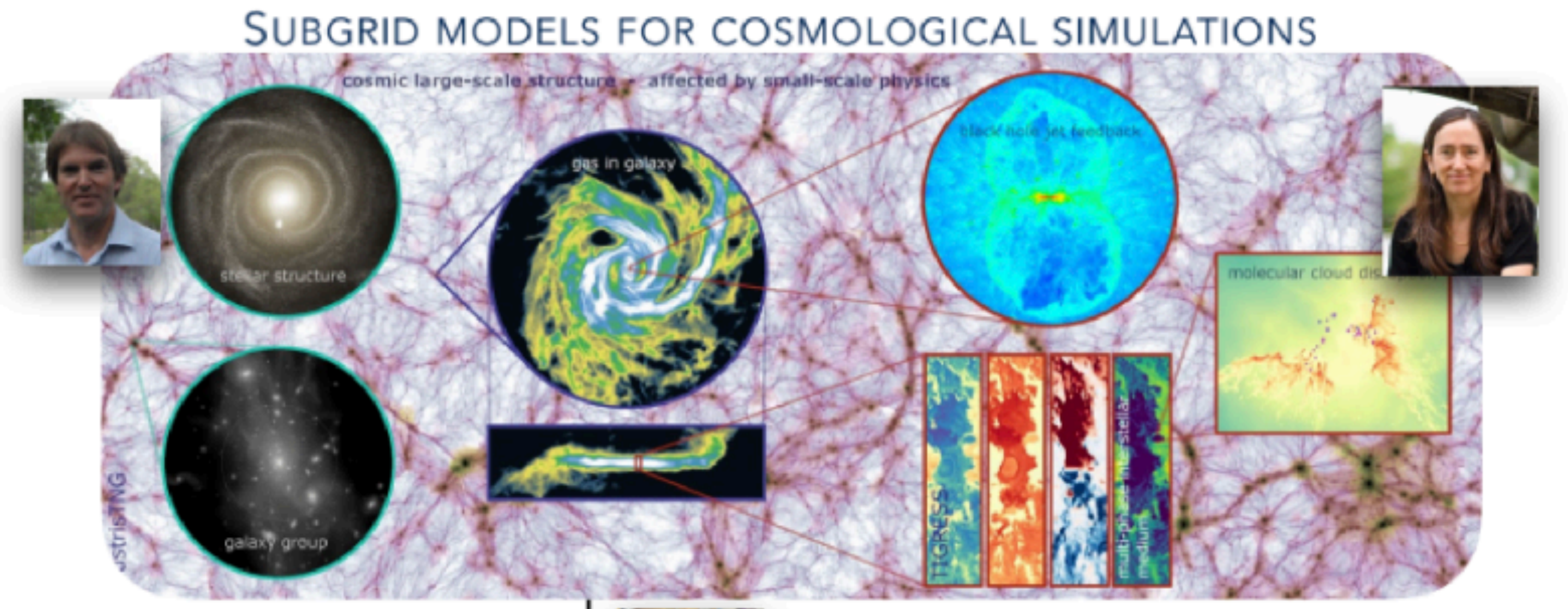
- large scales (> 10 Mpc) cheap gravity
- complicated physics local and on small scales
- we can hope to "paint in" the small-scales using emulators



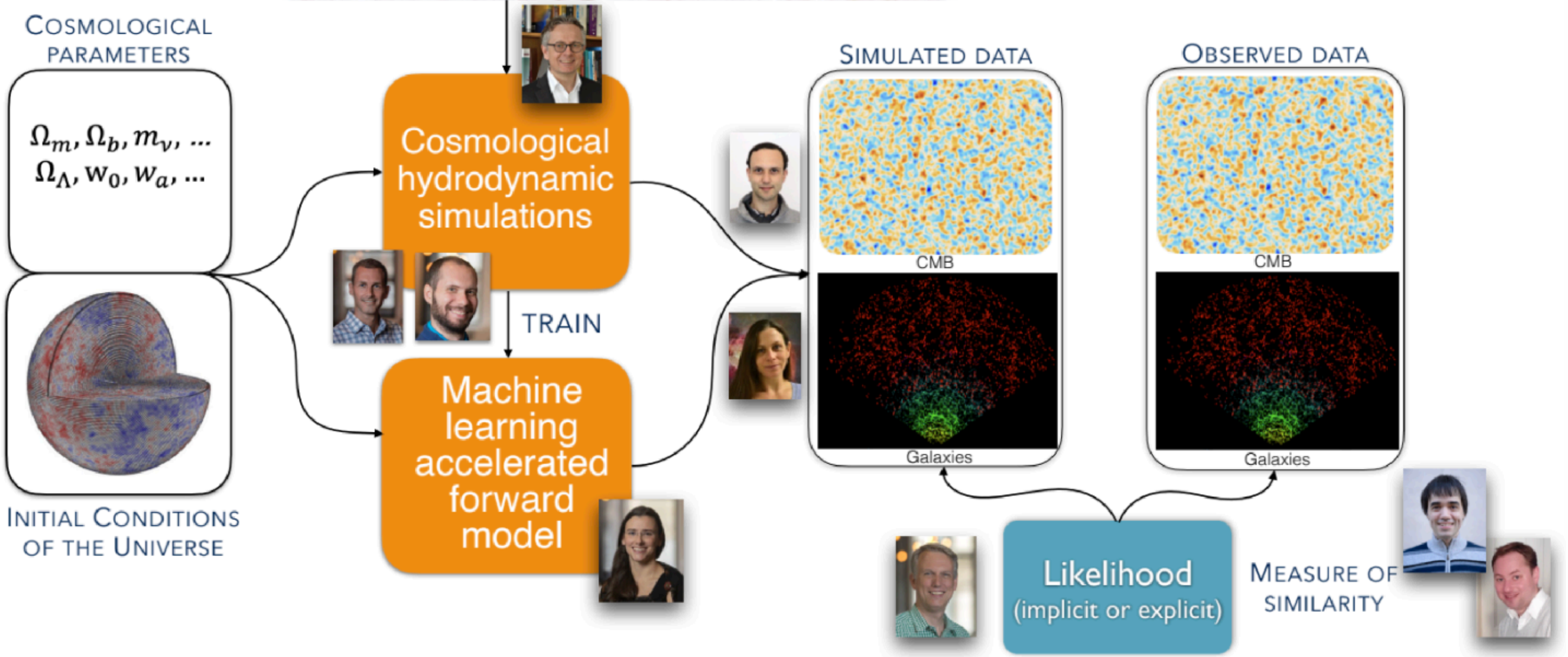
[C. Modi]

Emulation

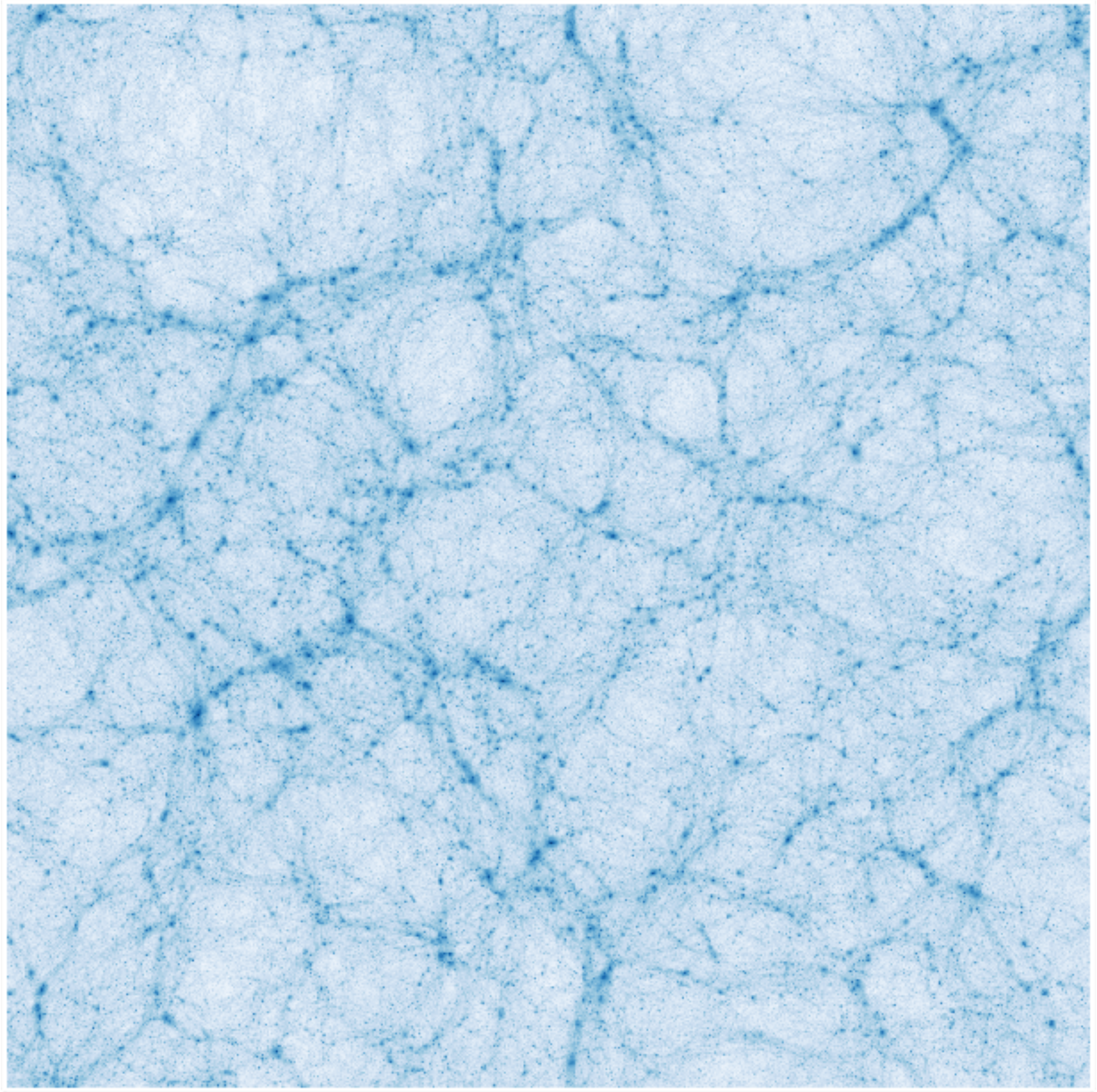
- large scales (> 10 Mpc) cheap gravity
- complicated physics local and on small scales
- we can hope to "paint in" the small-scales using emulators



e.g., Learning the Universe Simons Collaboration

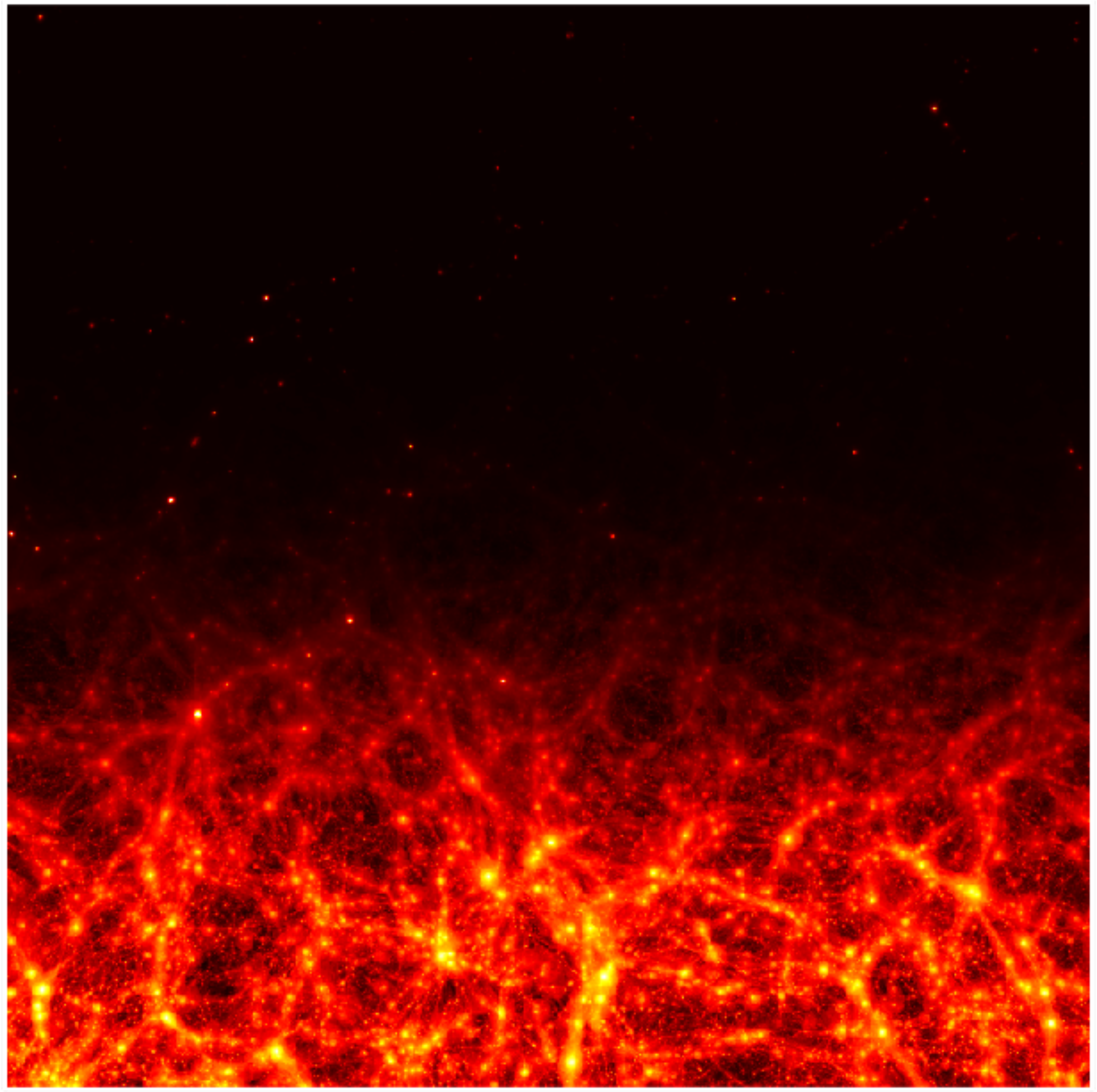


Emulating gas physics



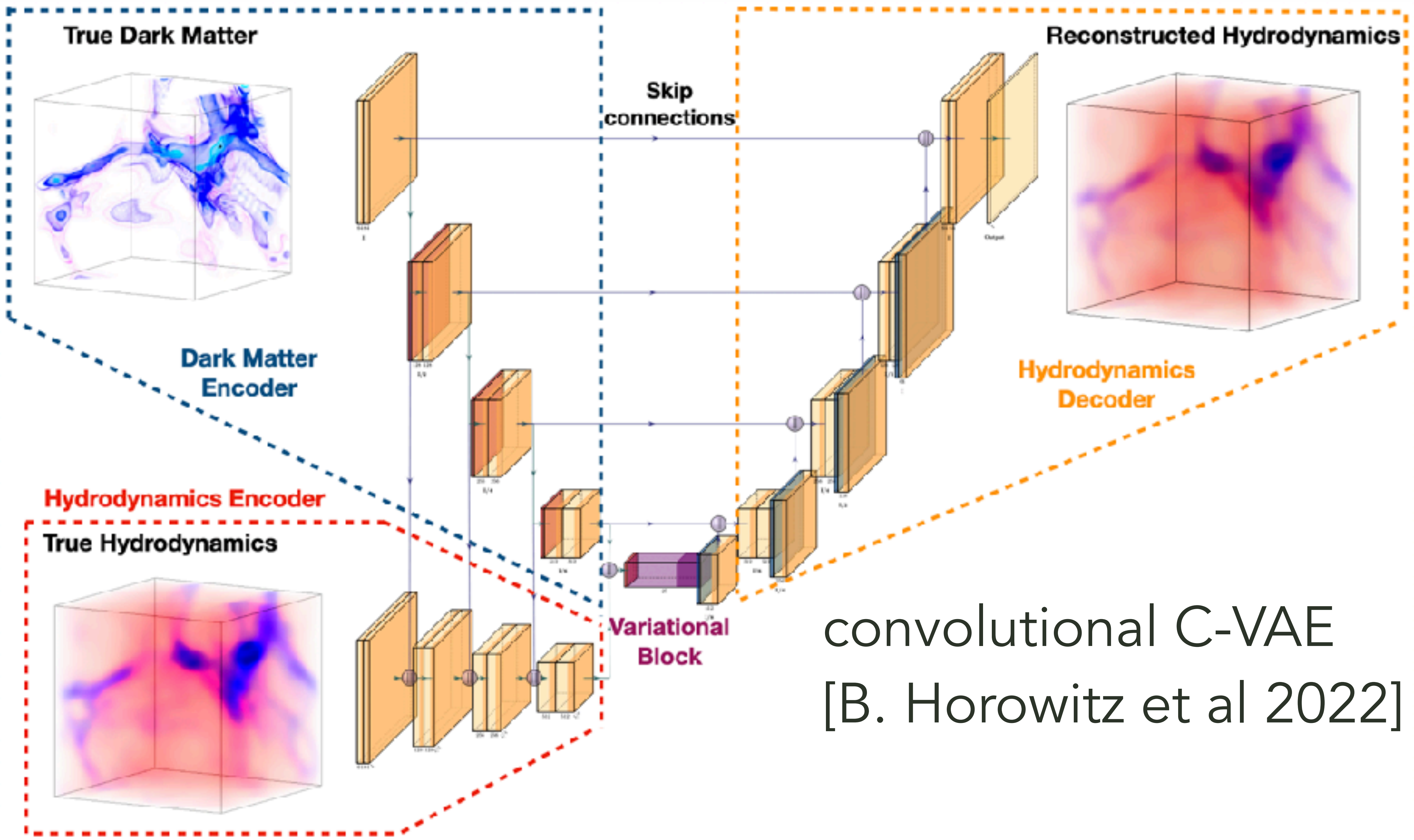
gravity-only simulation (cheap)

neural net
usually include
stochasticity



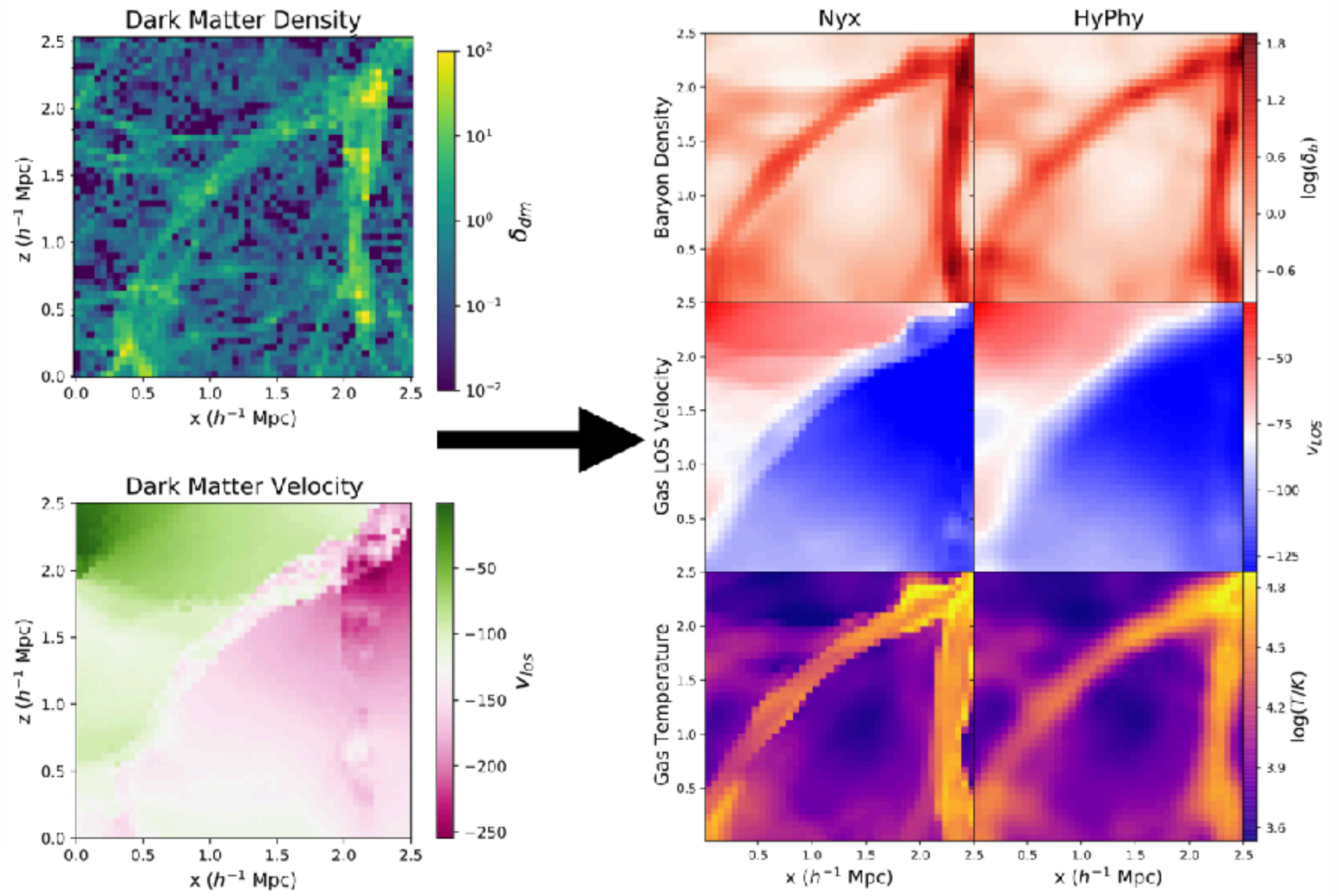
gas properties (expensive)

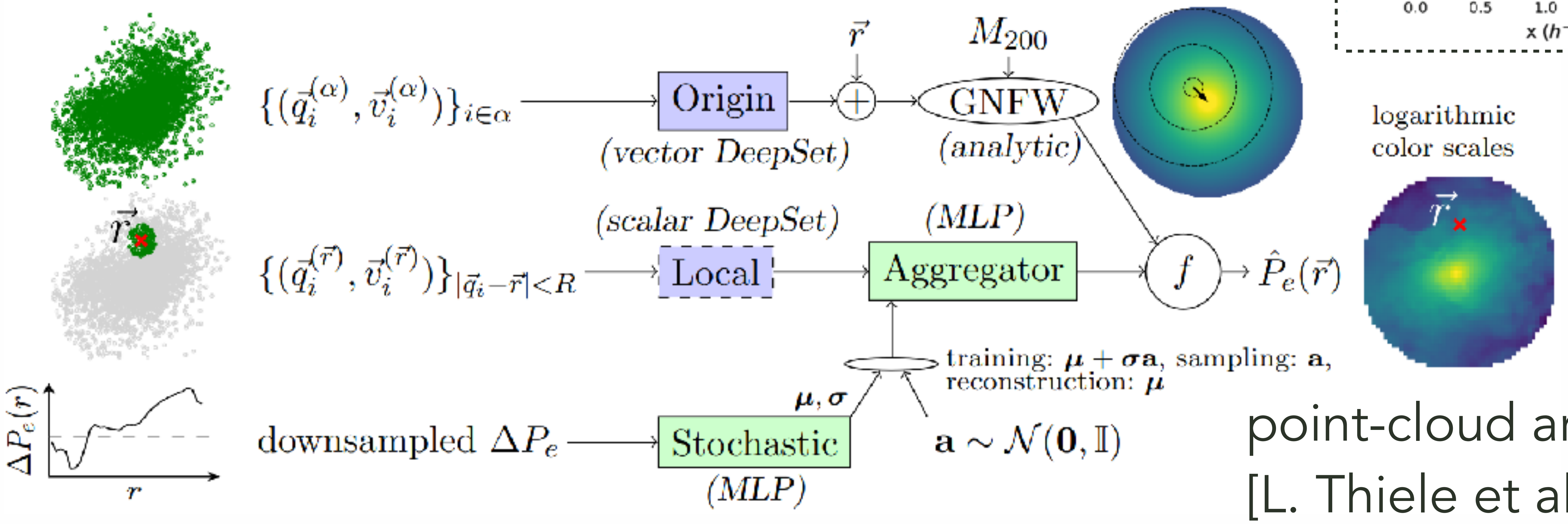
Emulating gas physics



convolutional C-VAE
[B. Horowitz et al 2022]

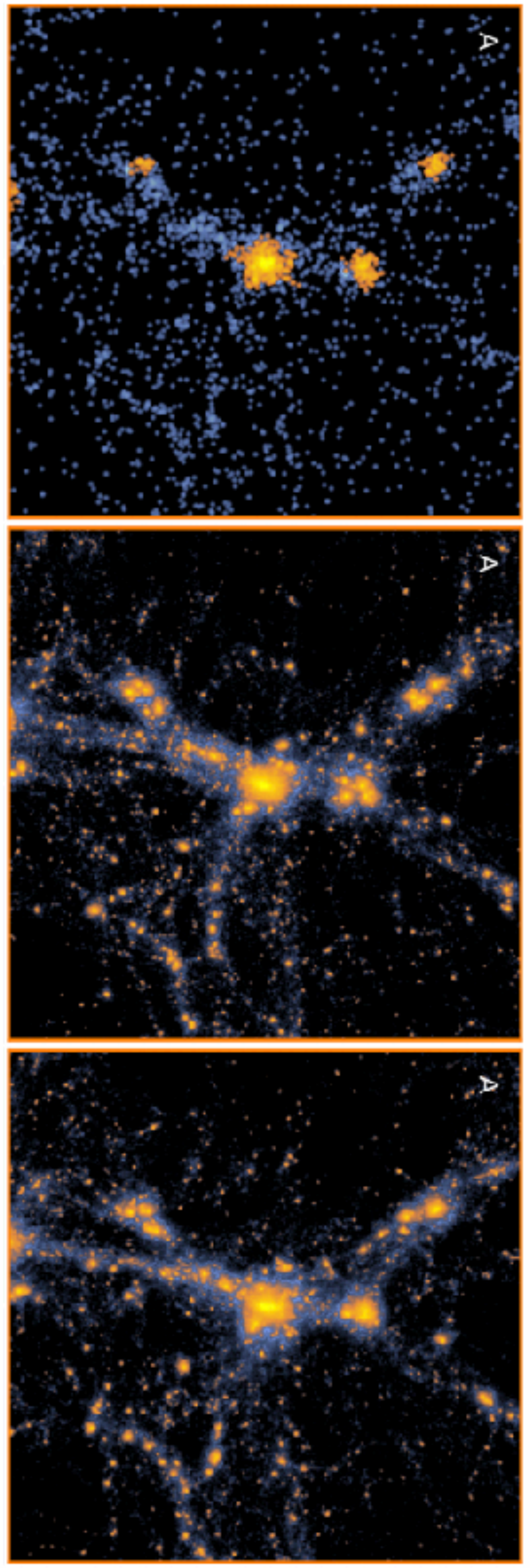
"truth" emulated





point-cloud architecture
[L. Thiele et al 2022]

Emulating small-scale clustering

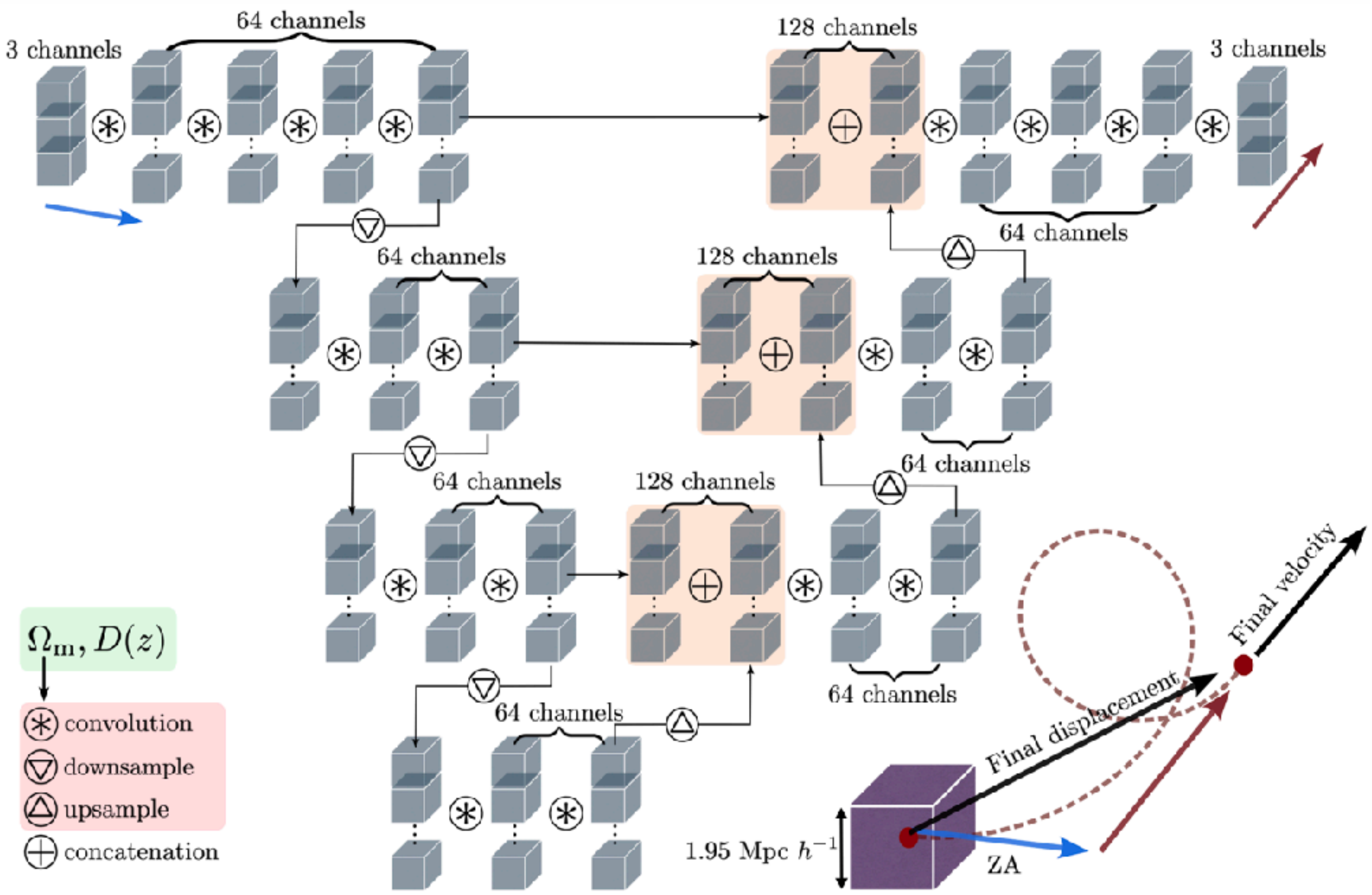


low-resolution

high-resolution

super-resolution
with neural net (GAN)

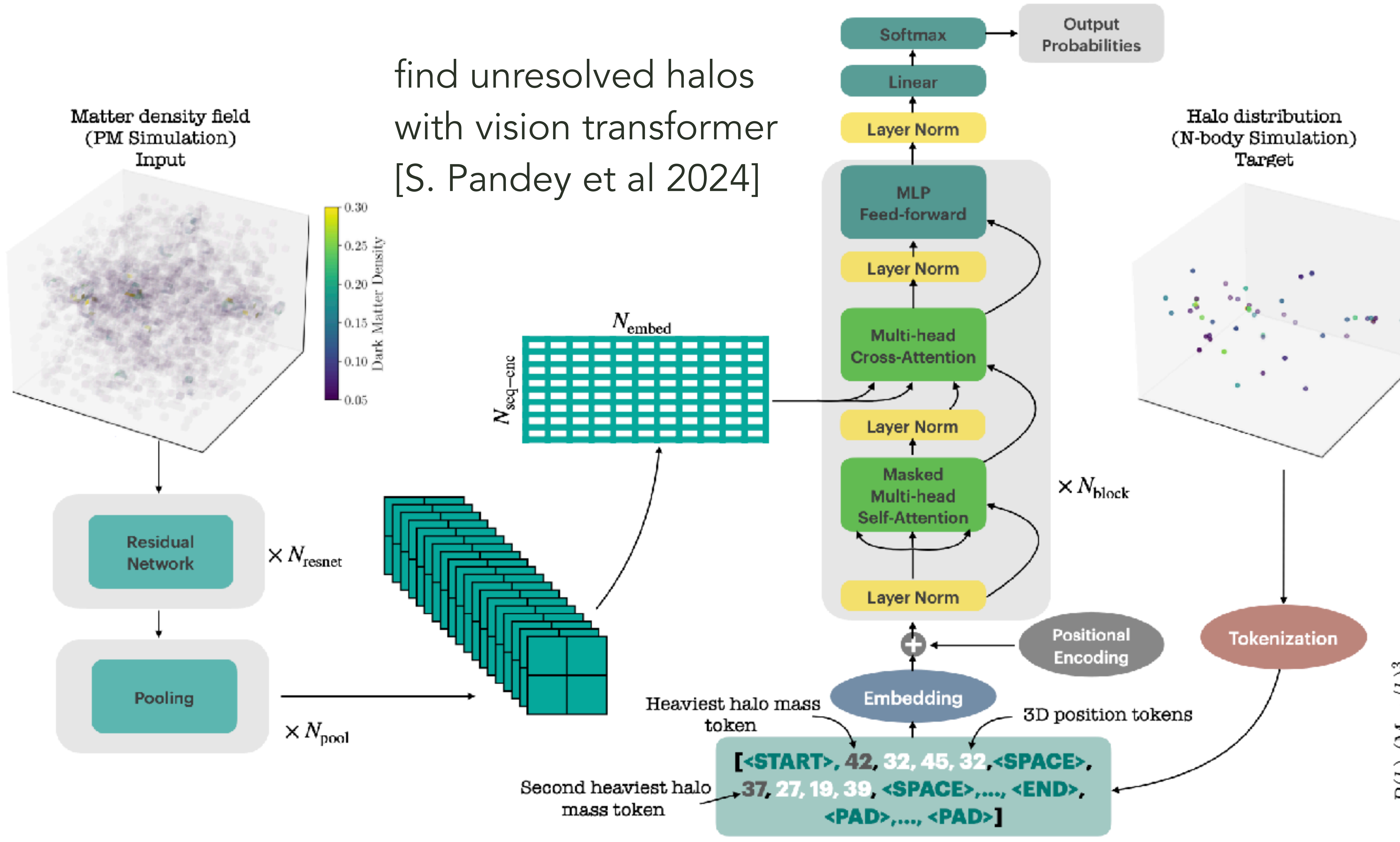
[Y. Li et al 2021]

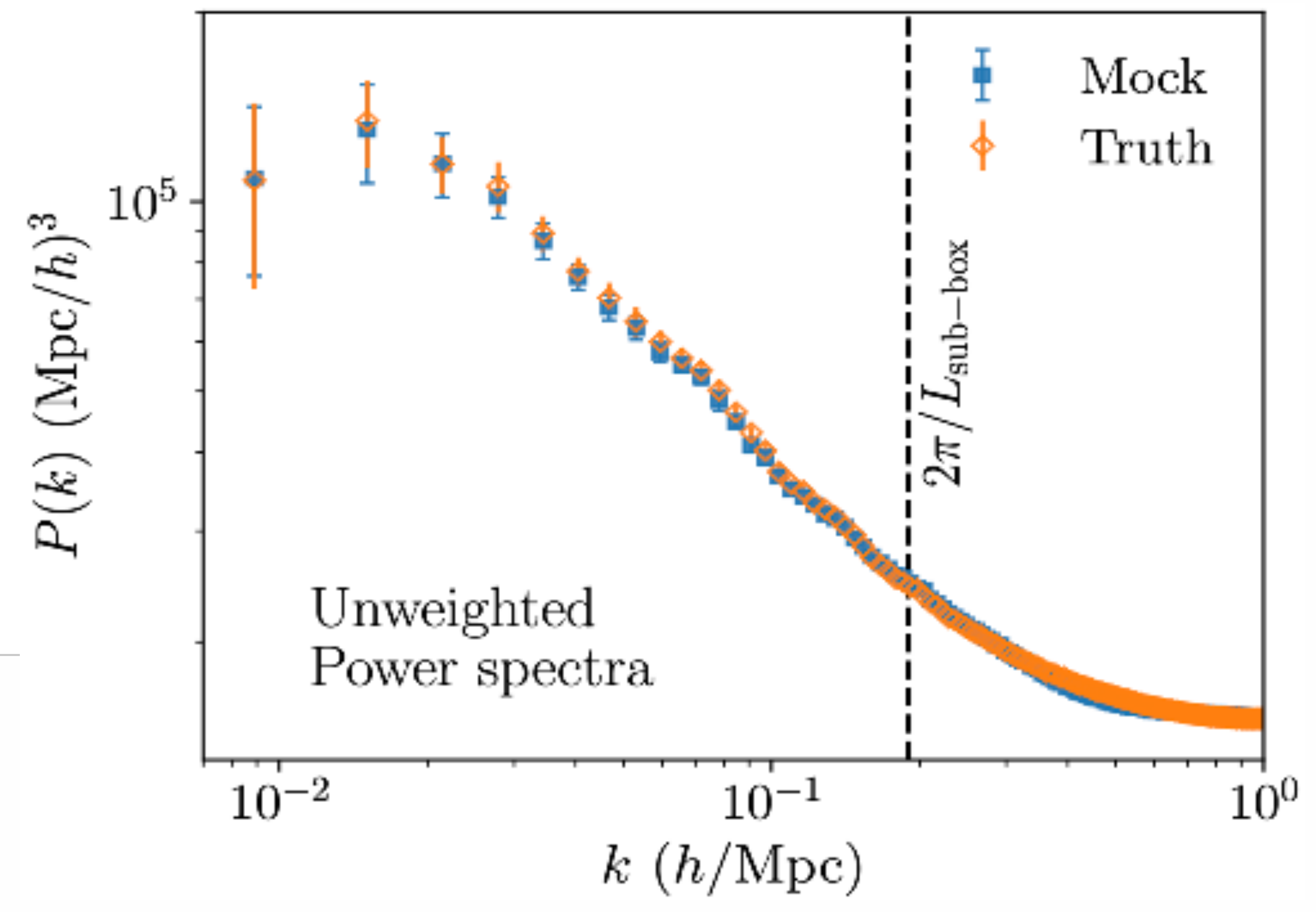


correct small-scale forces
[D. Jamieson et al 2024]

Emulating small-scale clustering

find unresolved halos
with vision transformer
[S. Pandey et al 2024]

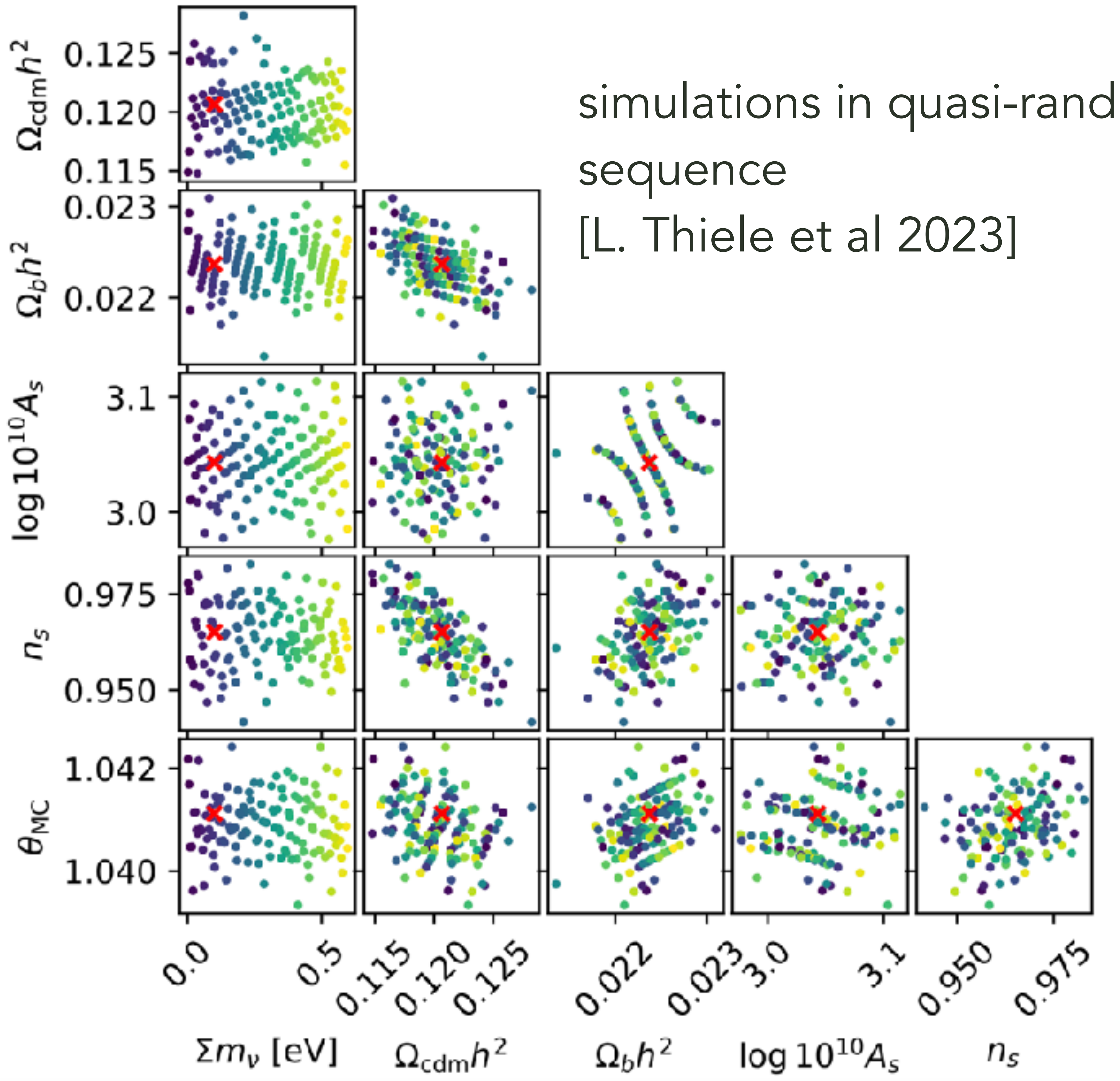




Inference

Implicit-likelihood inference

$$-2 \log P(x\,|\,\theta) = [x - S(\mu(\theta)]^T \Sigma^{-1} [x - S(\mu(\theta)]$$



simulations in quasi-random sequence
[L. Thiele et al 2023]

Learn parameter dependence and statistical distribution from simulations.

All observational effects directly folded in.

neural likelihood estimation (NLE)   neural posterior estimation (NPE)

$$P(\text{parameters}\,|\,\text{data}) = \frac{P(\text{data}\,|\,\text{parameters}) \qquad P(\text{parameters})}{P(\text{data})}$$

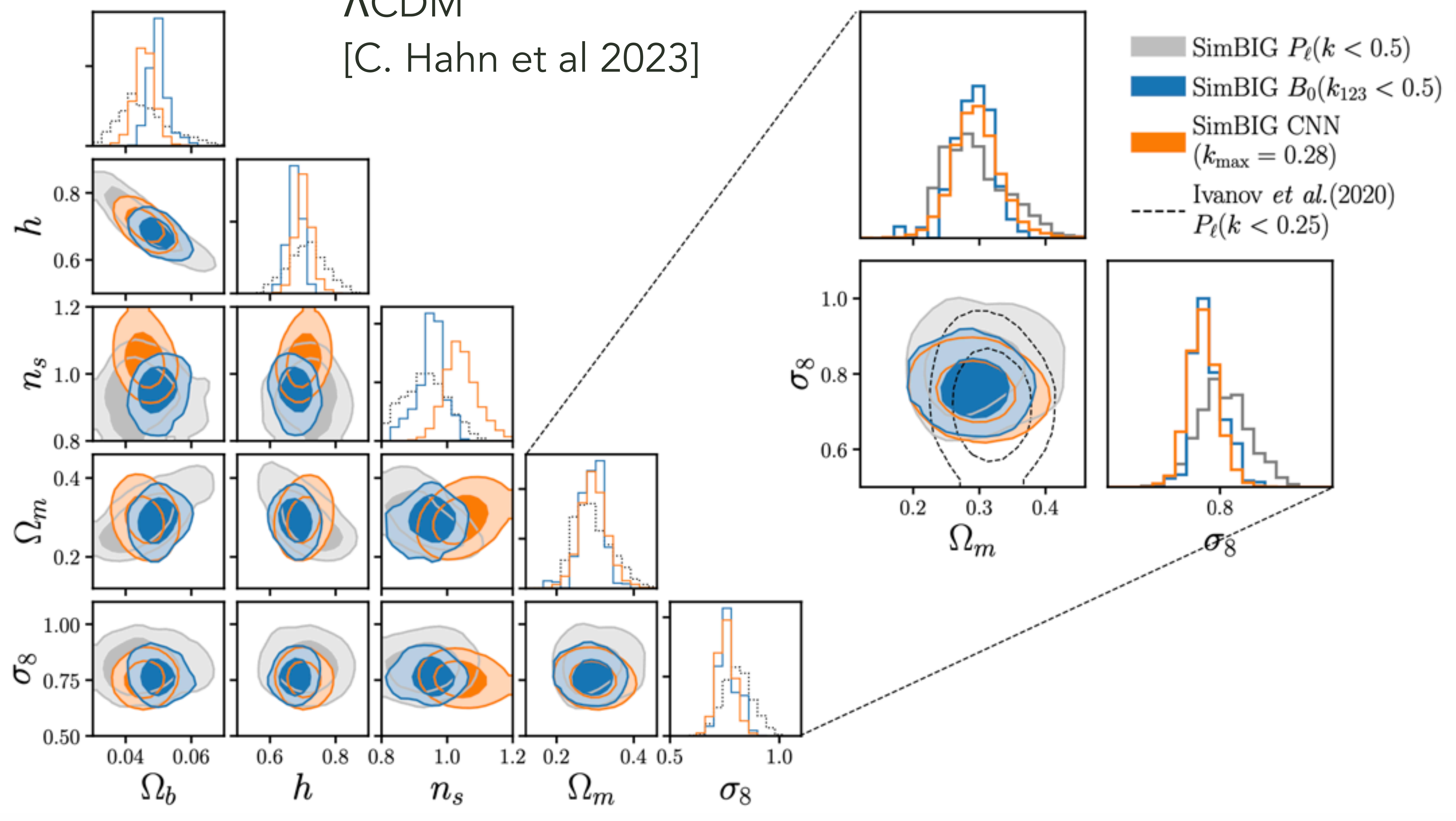
neural ratio estimation (NRE)

Choose between flow and classifier

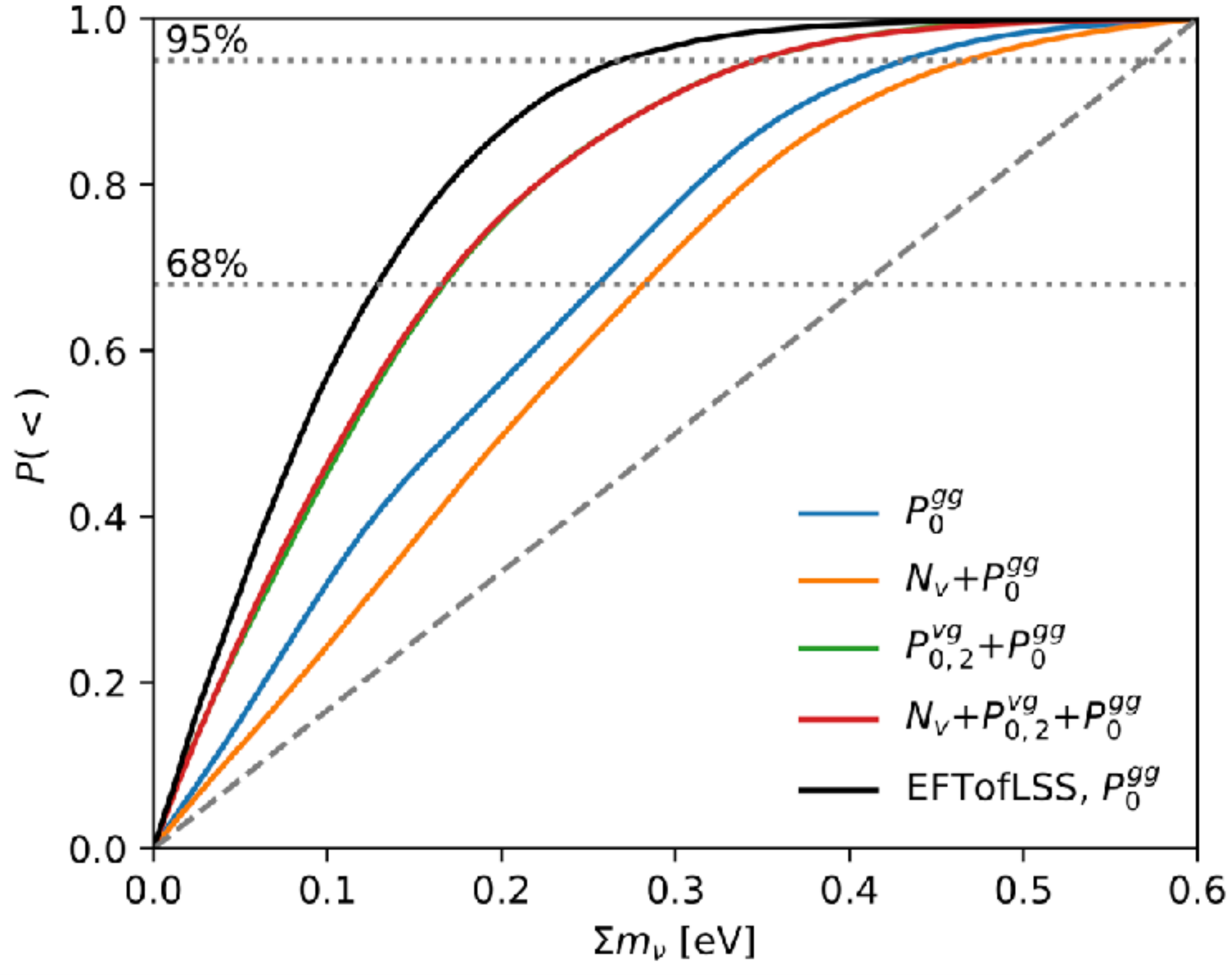
Implicit-likelihood inference: galaxy clustering

usually: field → summary statistics → compression → posterior

BOSS CMASS SGC
ΛCDM
[C. Hahn et al 2023]

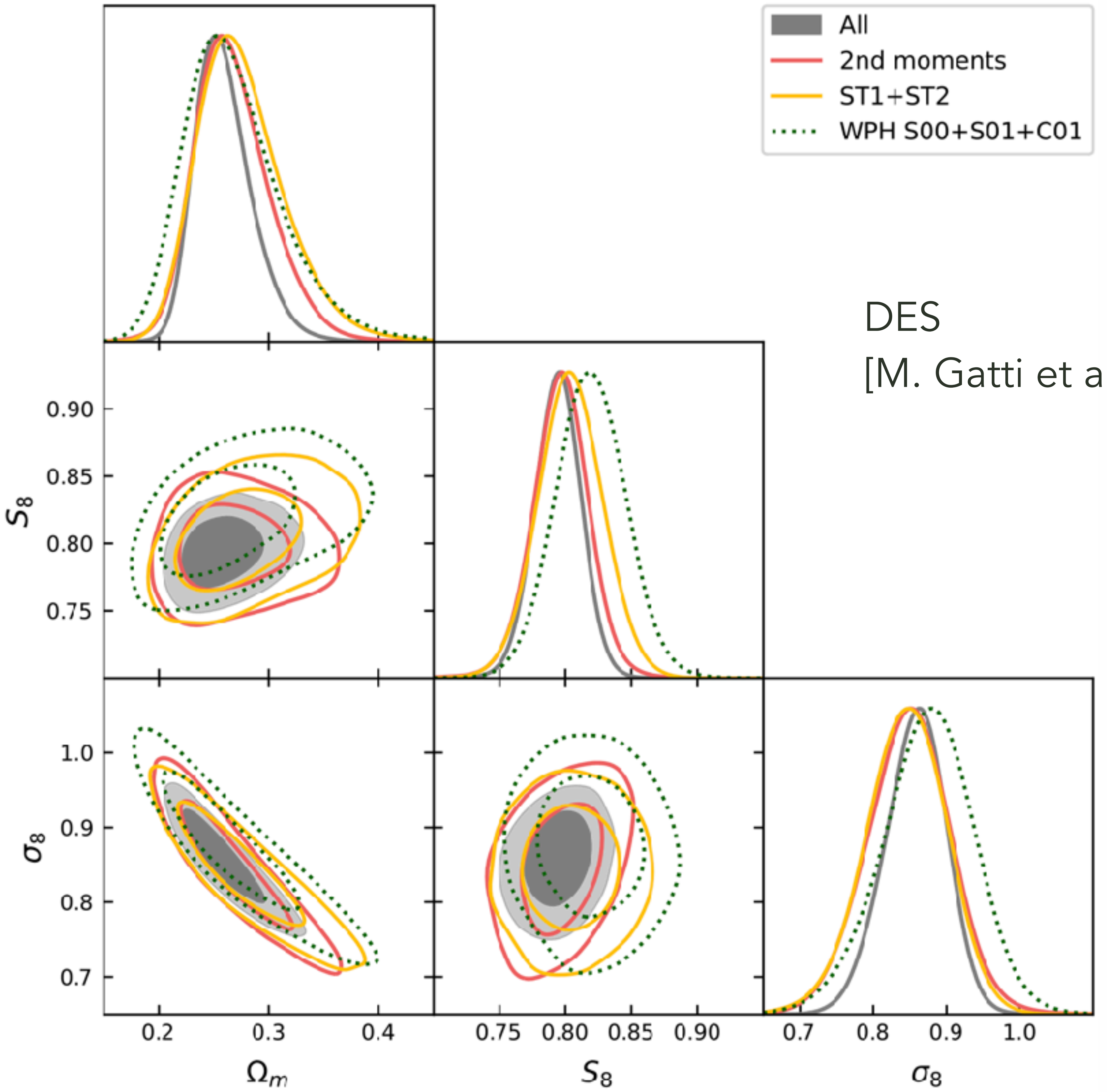


BOSS CMASS NGC
$\sum m_\nu$
[L. Thiele et al 2023]



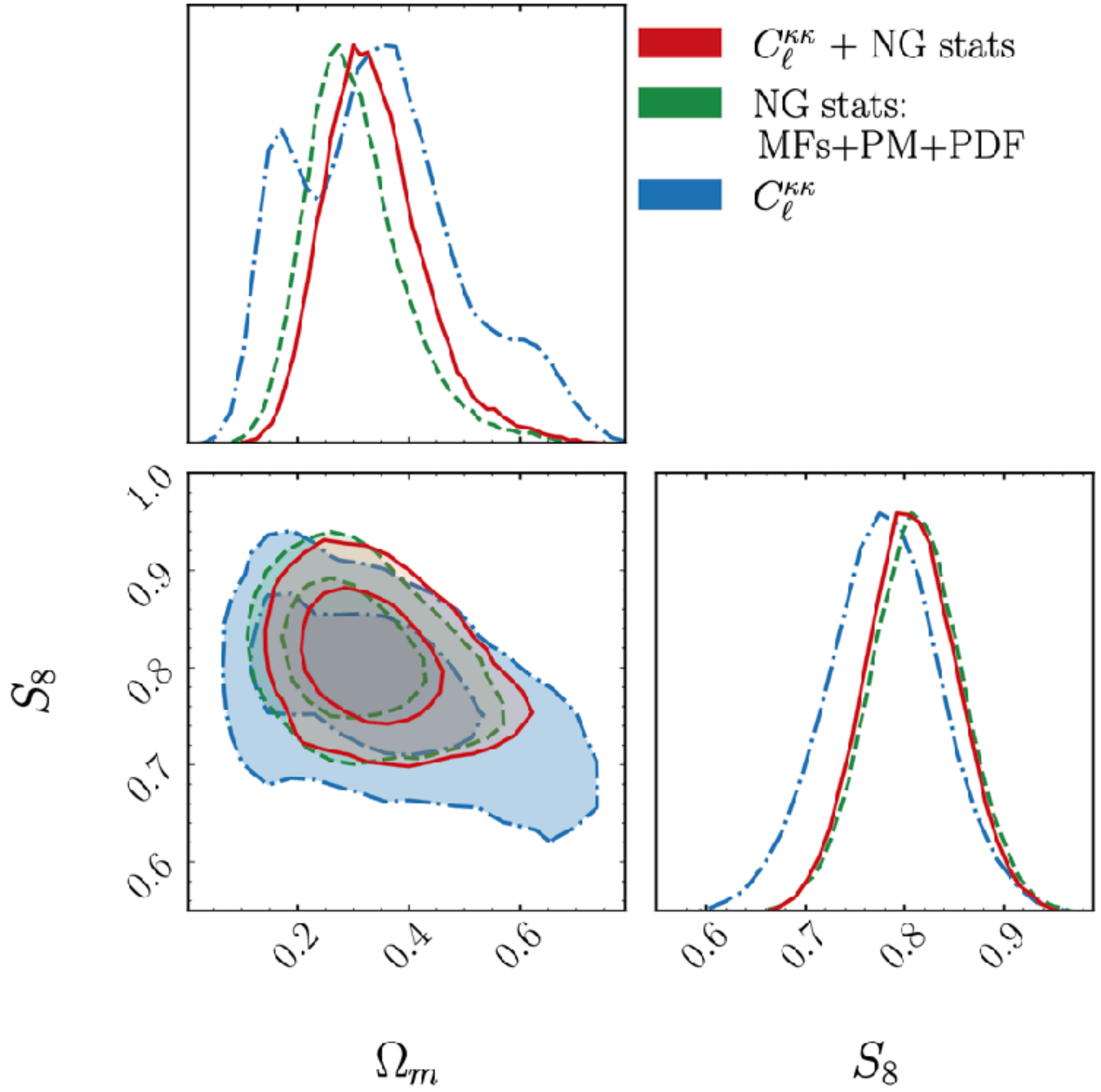
Implicit-likelihood inference: weak lensing

usually: field ➞ summary statistics ➞ compression ➞ posterior



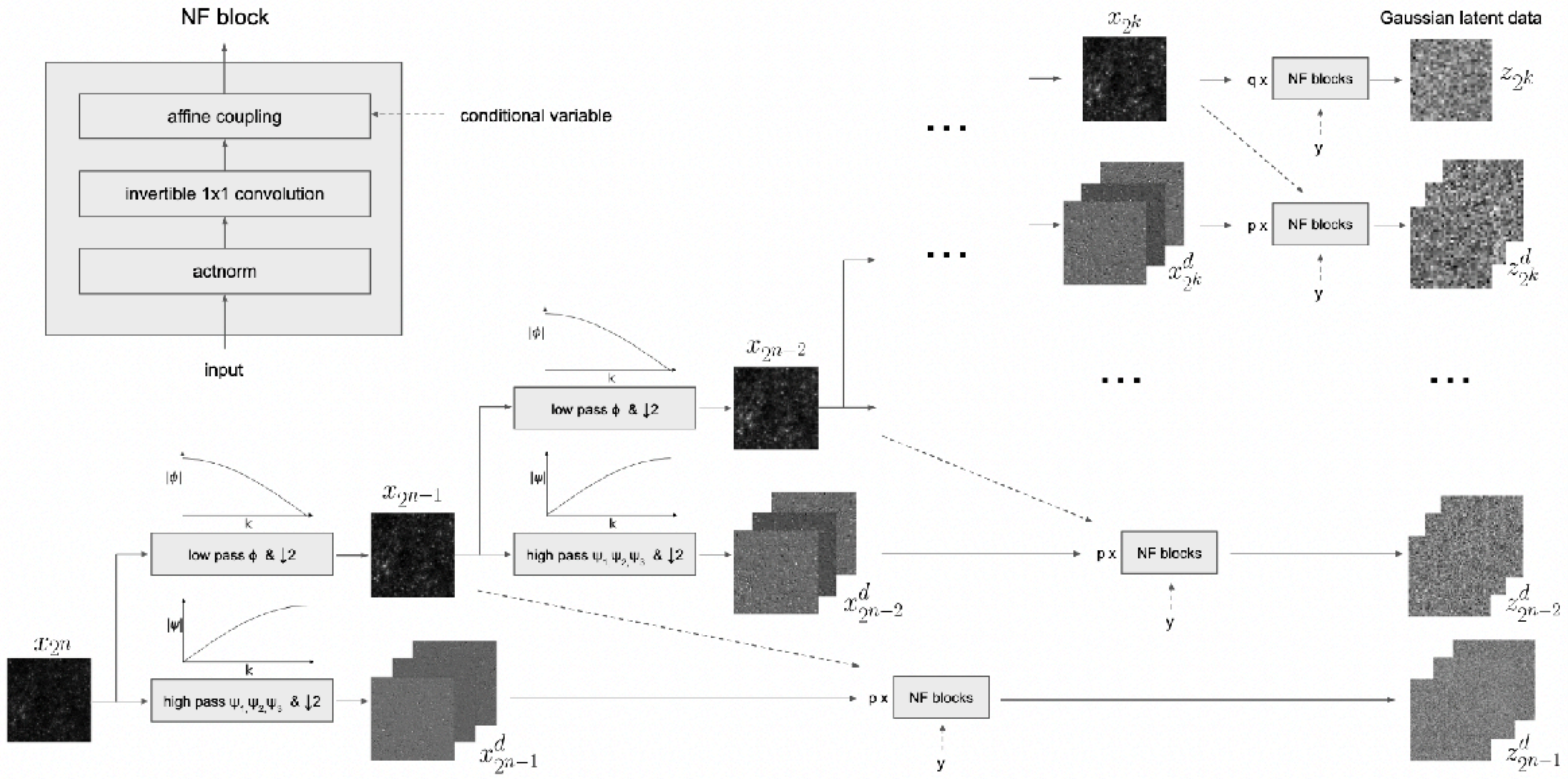
DES
[M. Gatti et al 2024]

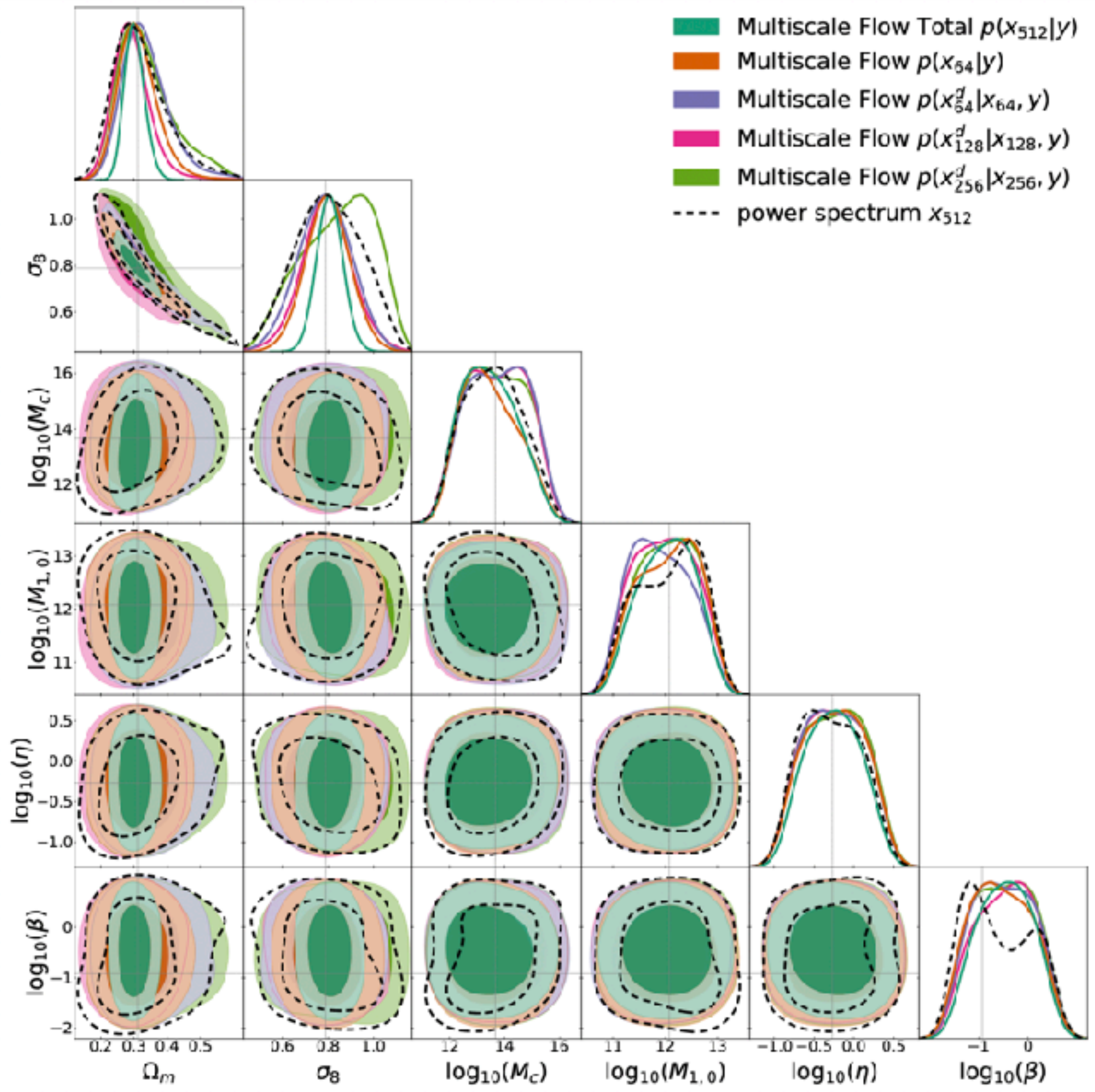
HSC
[C. Novaes et al 2024]



Implicit-likelihood inference: weak lensing

can also do: field-level with multi-scale density estimator (normalizing flow)

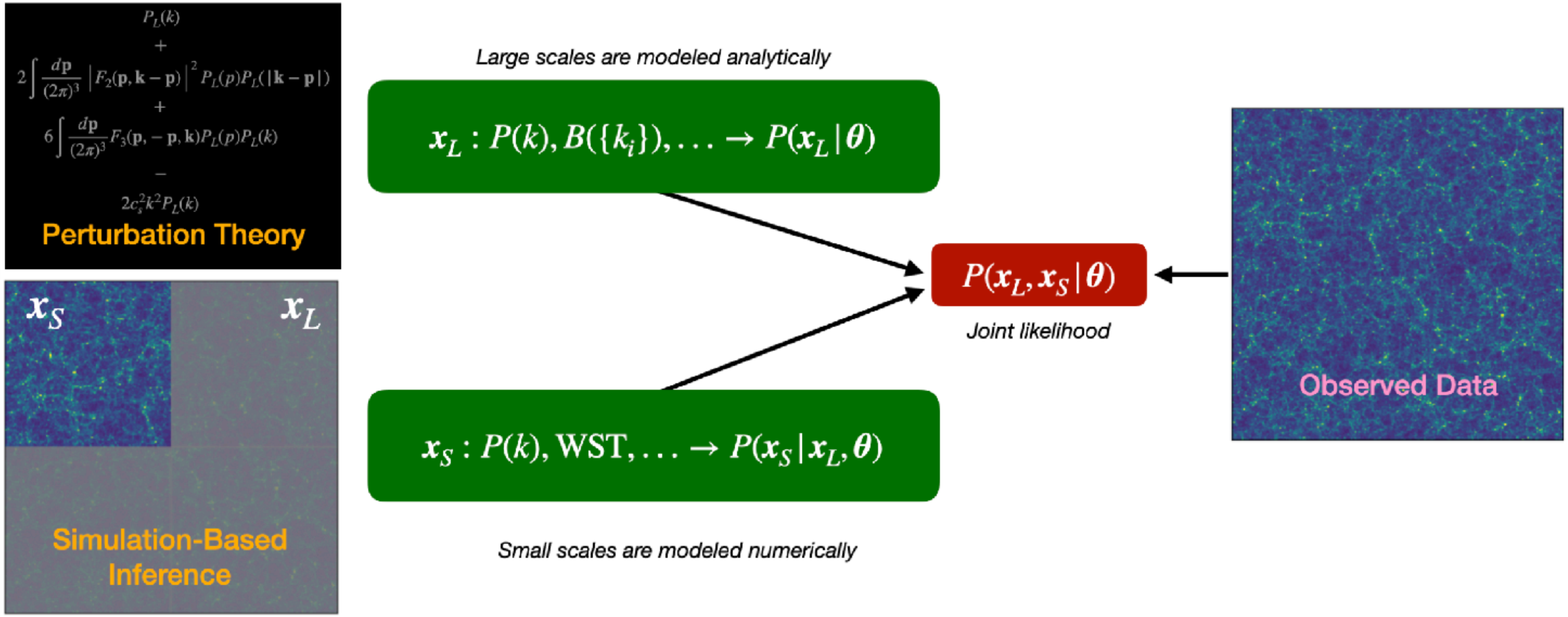




[B. Dai & U. Seljak 2024]

Hybrid Implicit Likelihood Inference

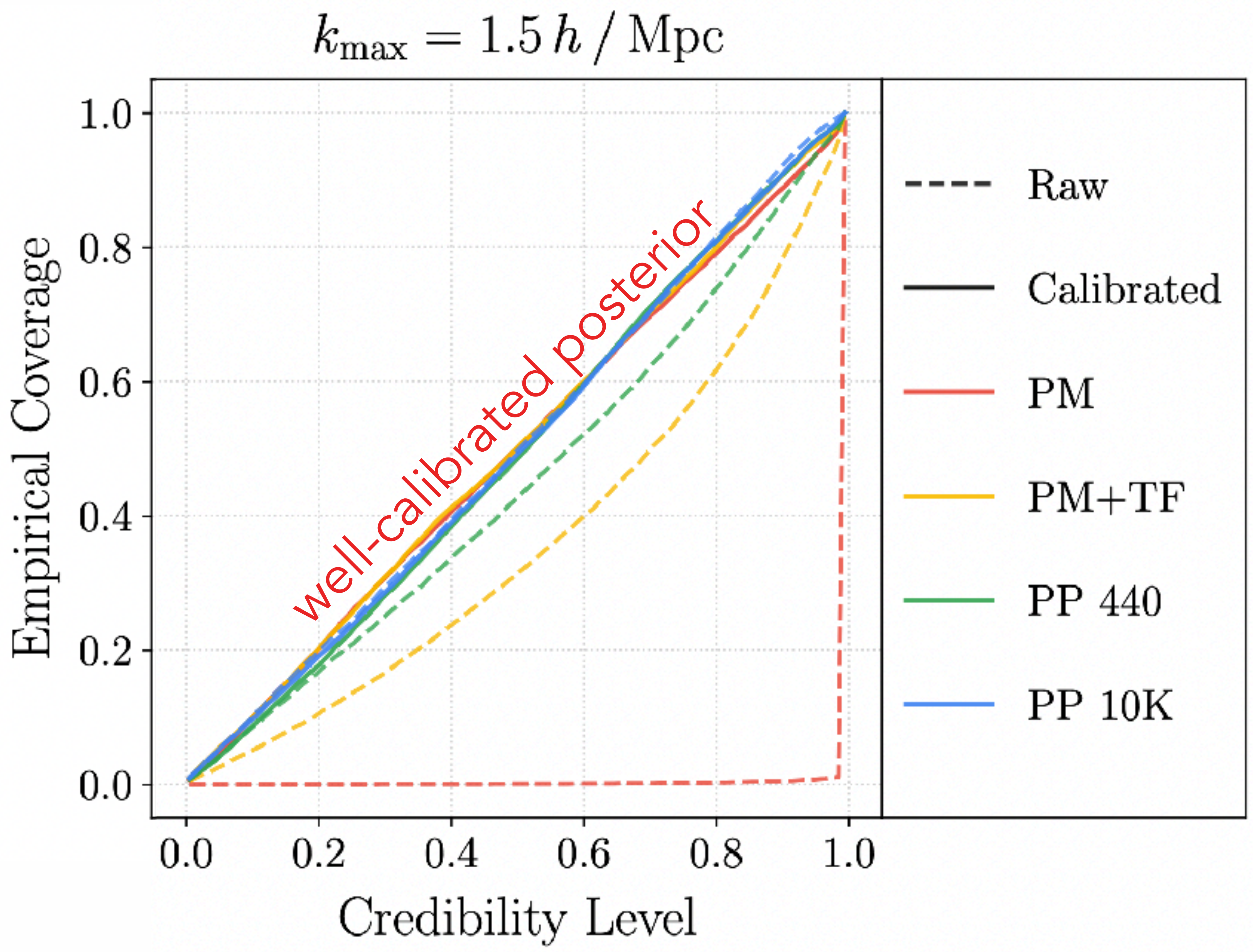
we know how to treat the large scales → maybe don't need to simulate them!

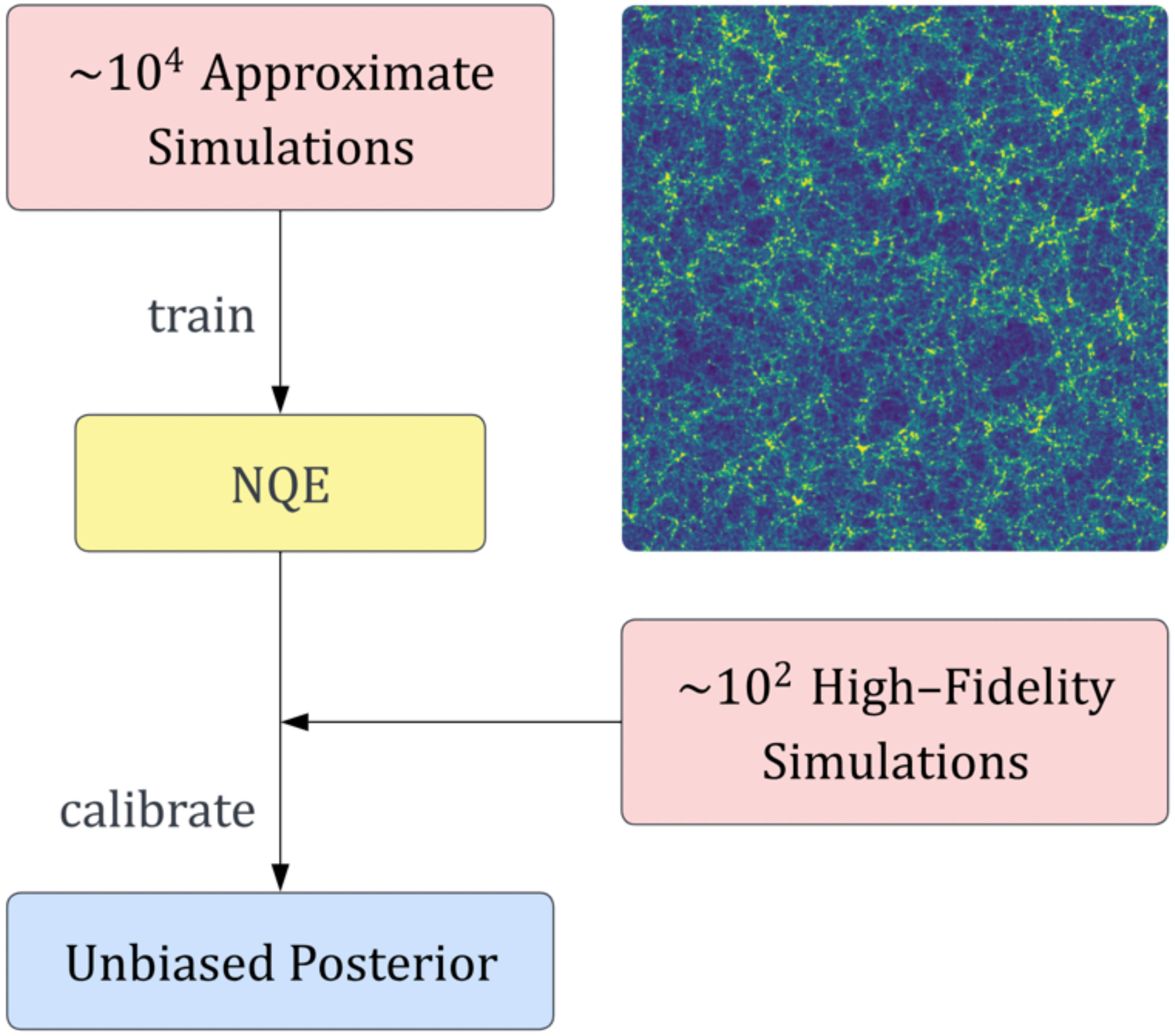


$$p(\boldsymbol{x}|\boldsymbol{\theta}) = p(\boldsymbol{x}_L, \boldsymbol{x}_S|\boldsymbol{\theta}) = p(\boldsymbol{x}_L|\boldsymbol{\theta})p(\boldsymbol{x}_S|\boldsymbol{x}_L, \boldsymbol{\theta})$$

[C. Modi & O. Philcox 2023]

Neural Quantile Estimation & post-training calibration





[He Jia 2024]

Open / active problems

- accounting for interpolation error
- modeling of small scales
- robust propagation of small-scale simulations into large-scale simulations (via emulators)
- in hybrid method, survey systematics & nuisance parameter correspondence
- combining low- and high-quality simulations
- smart simulation strategies while keeping calibration in check
- implicit priors of neural networks
- …