Implicit Likelihood Inference for Late-Time Cosmology

Leander Thiele (Kavli IPMU)

21st Rencontres du Vietnam, 8/14/2025

Outline

Why should nature have conspired to put all information into the (quasi-)linear modes we happen to be able to describe with perturbation theory?

Accessing the information in non-linear regime typically requires simulation-based methods.

Outline

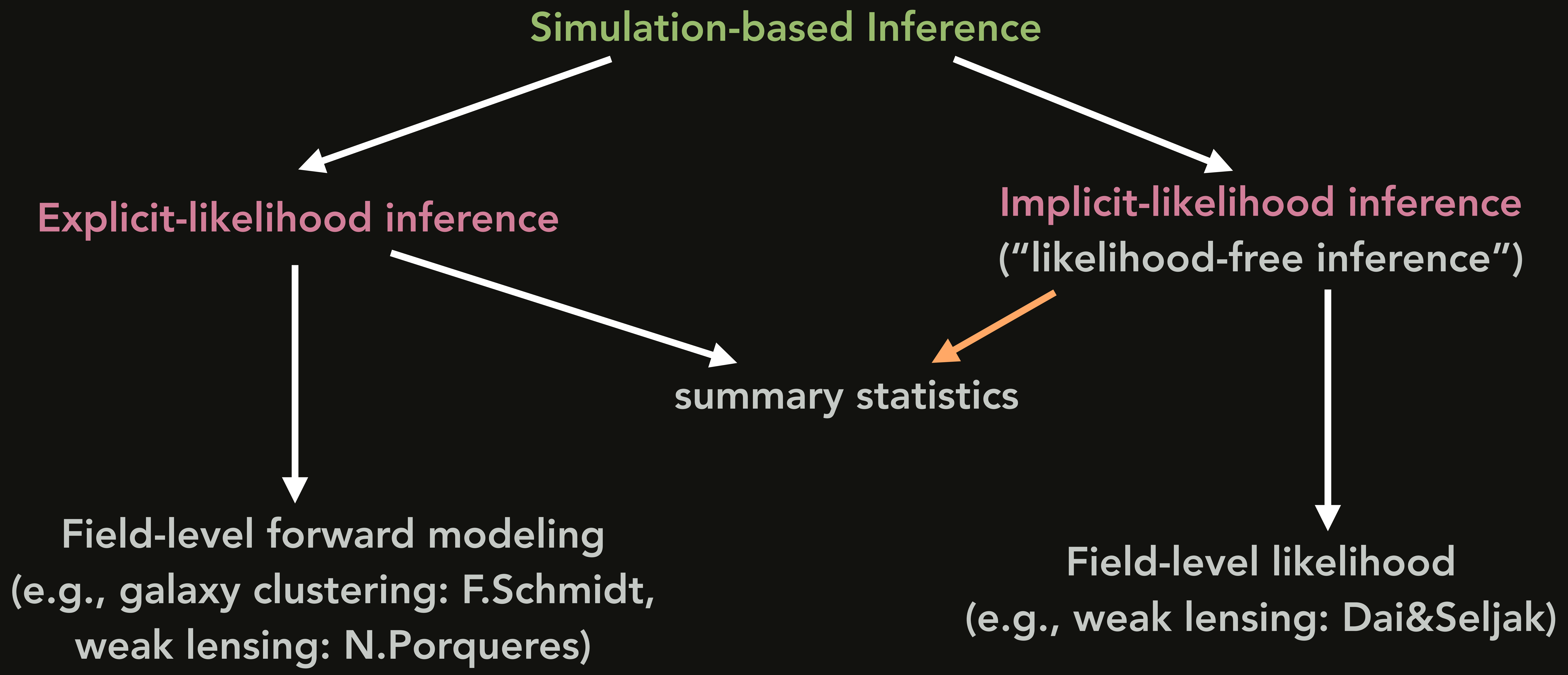
Implicit Likelihood Inference: A tool to solve inverse problems which are implicitly defined through simulations, typically using deep neural networks

1. Motivation & Overview of Methods

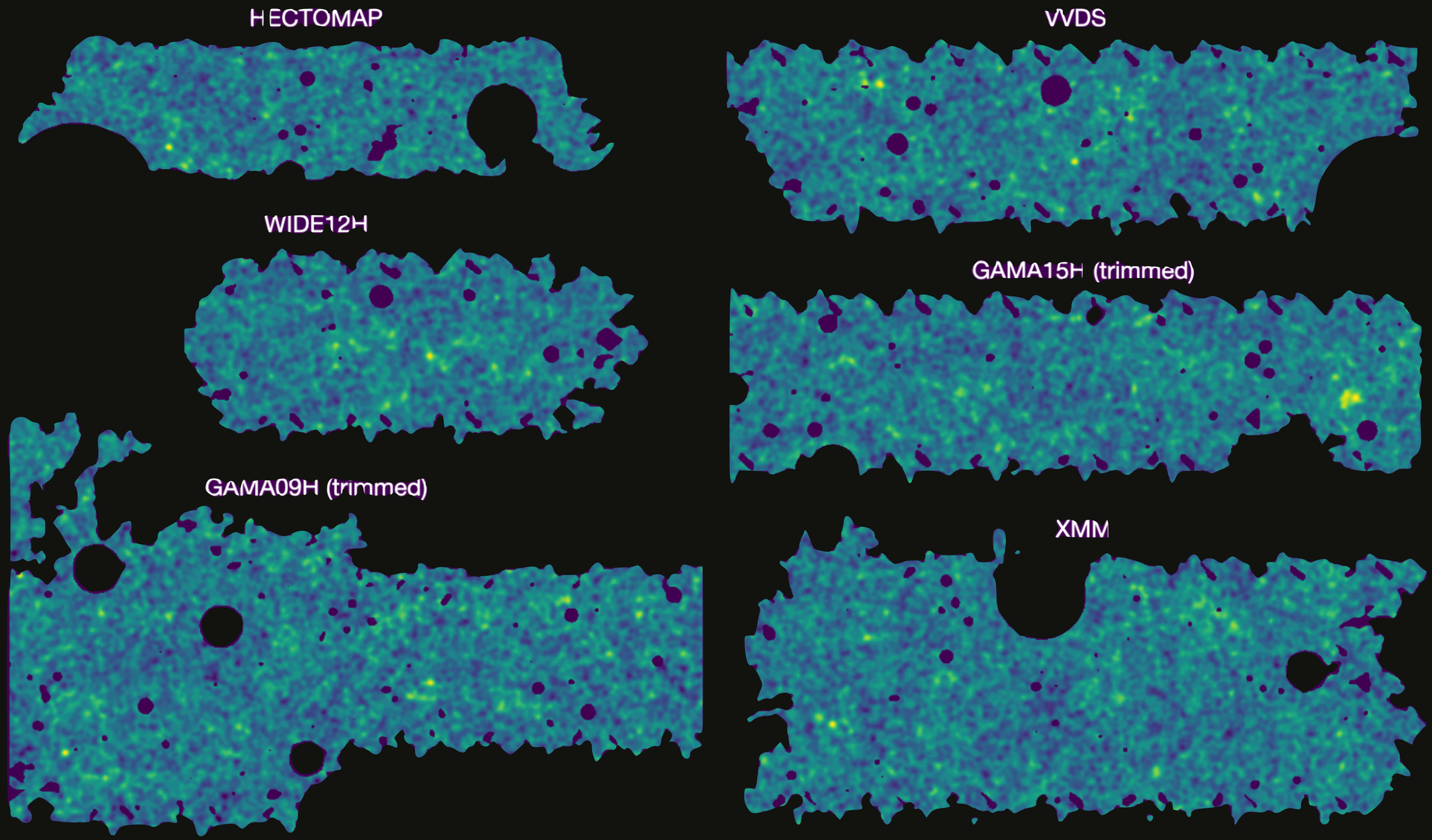
2. Example applications

3. The “elephant in the room”: How to make it reliable and computationally feasible

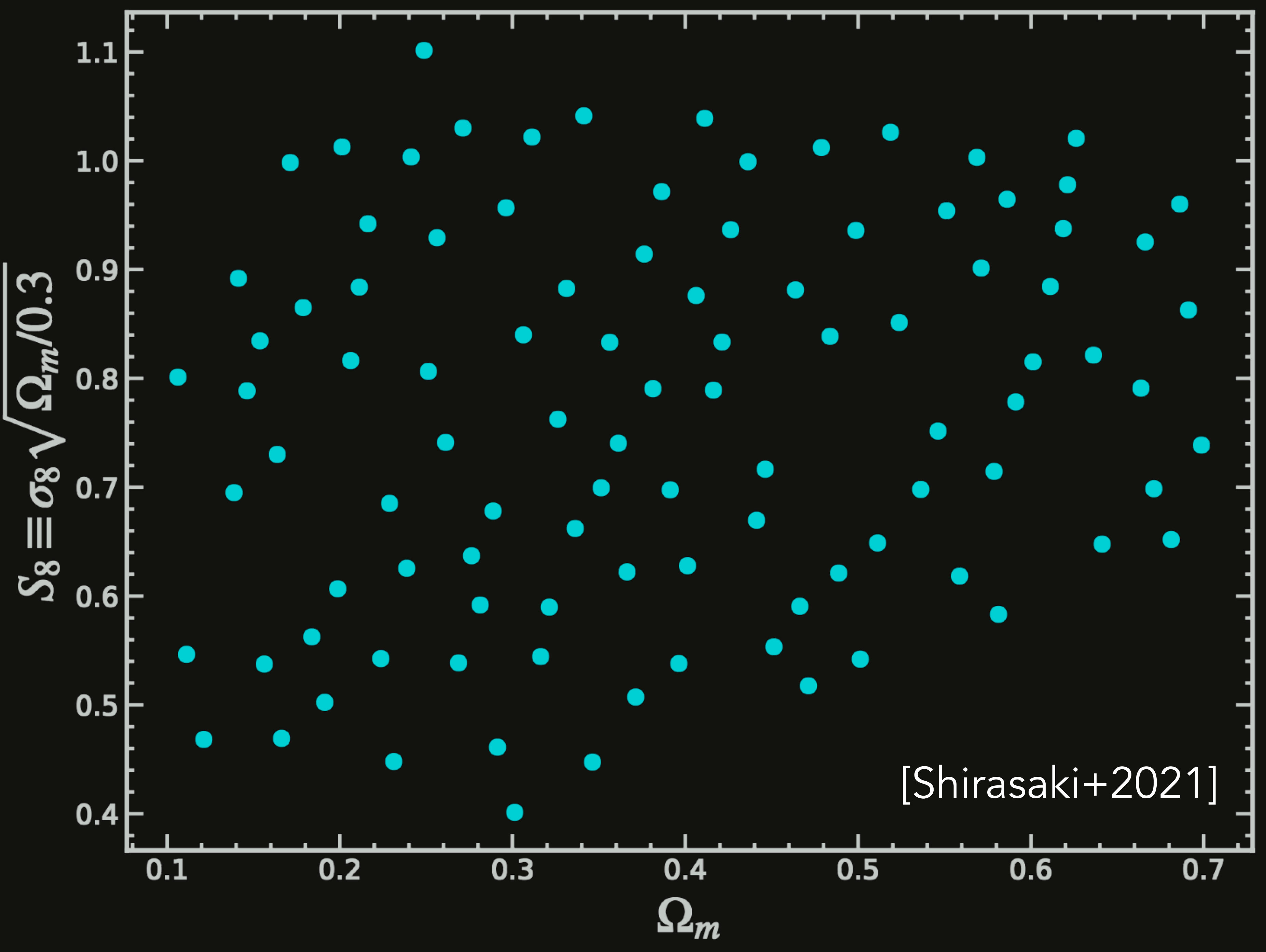
A word about words

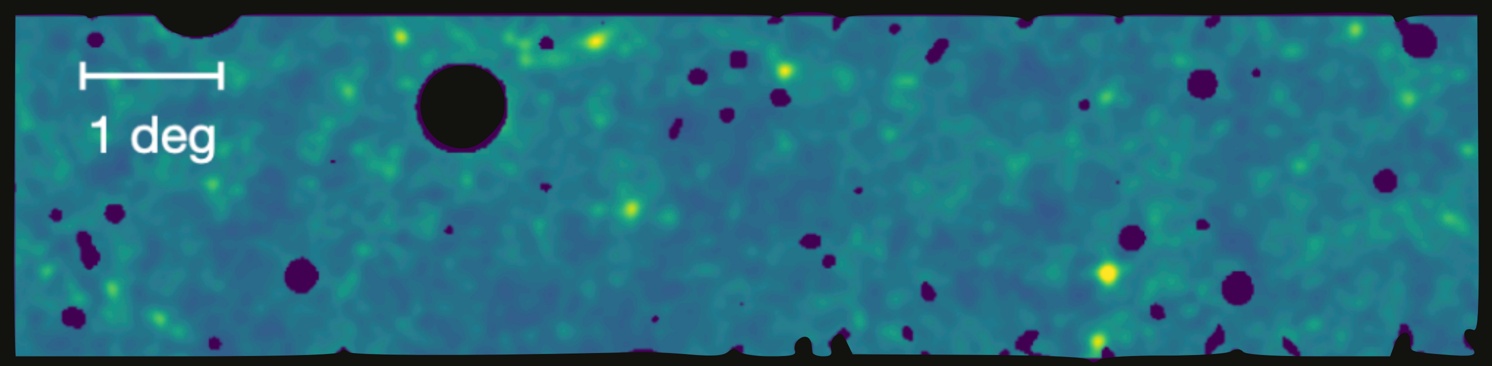


Cosmology from HSC year-1 higher order statistics



training simulations

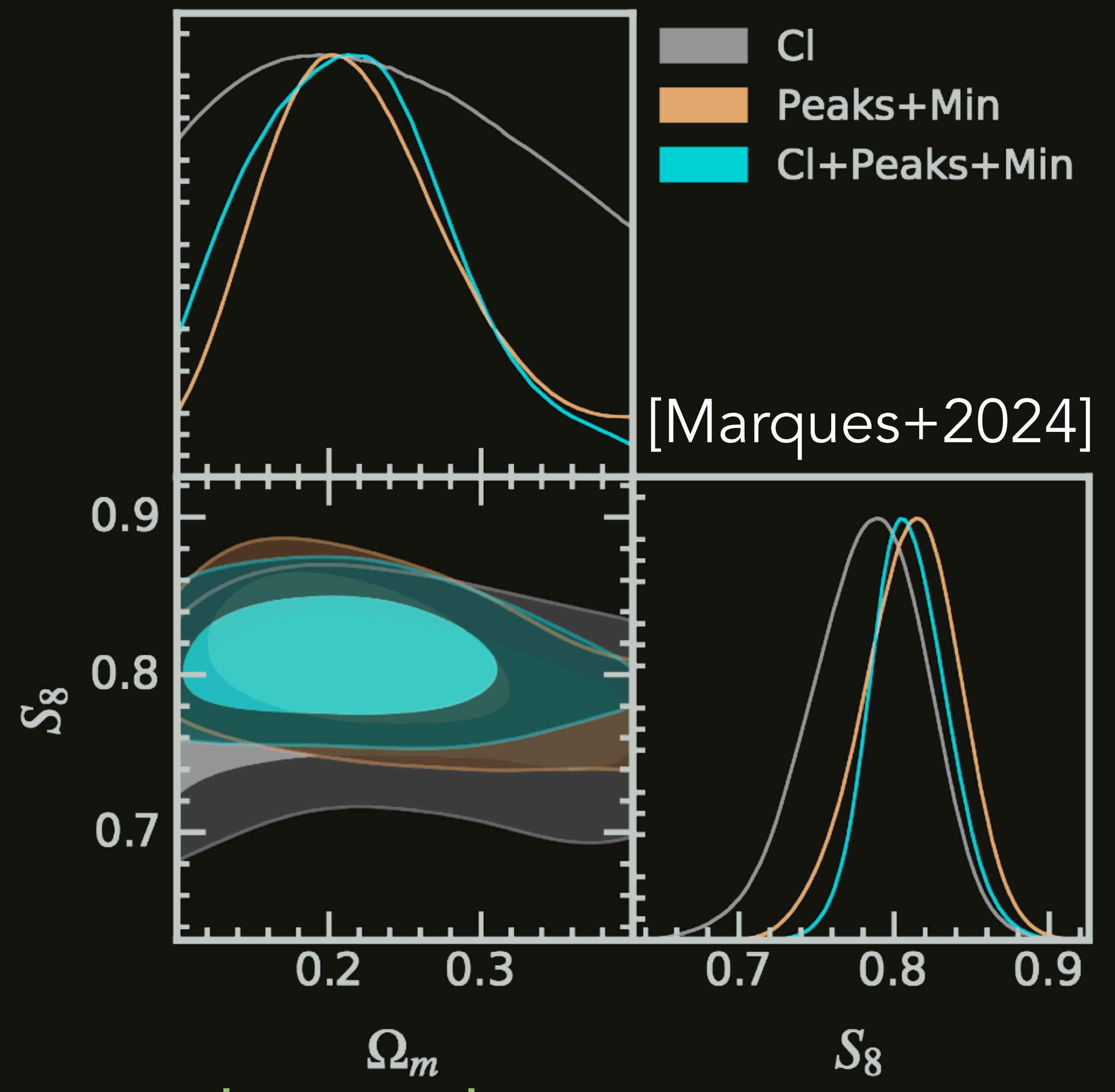




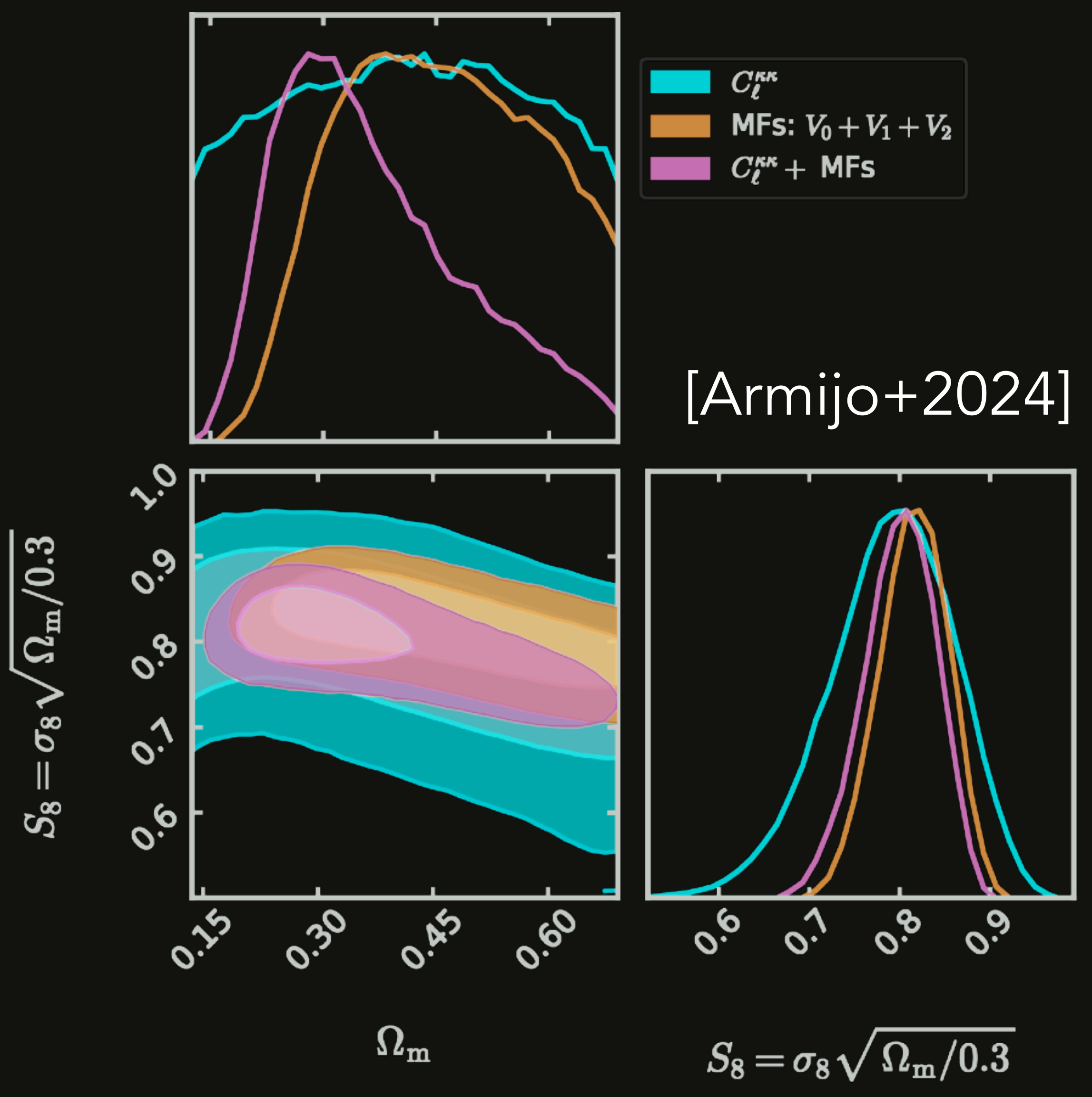
noiseless simulated map for comparison: signal down to very small scales

with Jessica Cowell, Daniela Grandon, Joaquin Armijo, Camila Novaes, Sihao Cheng, Gabriela Marques, Masato Shirasaki, Jia Liu

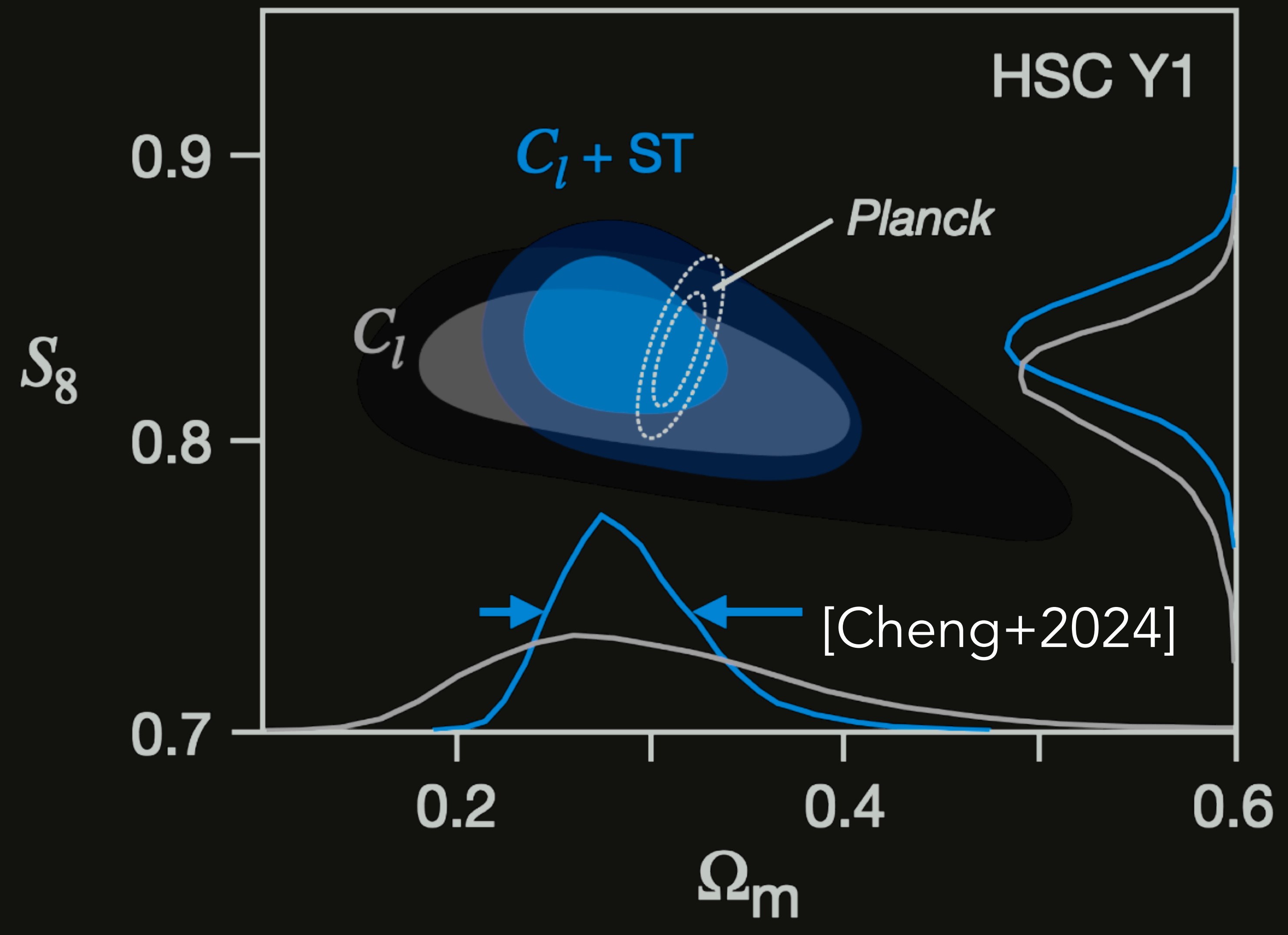
Cosmology from HSC year-1 higher order statistics



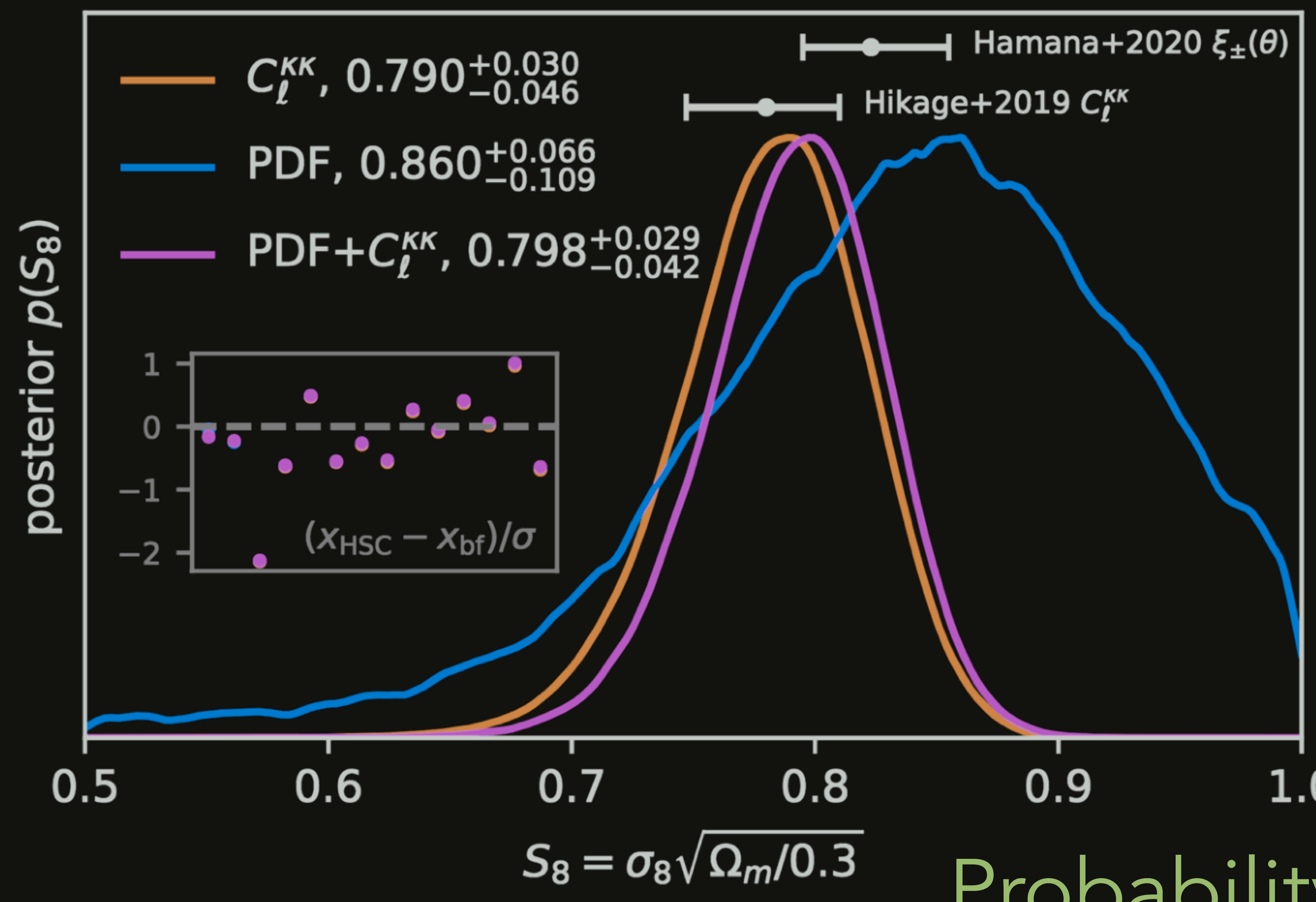
peaks and minima



Minkowski functionals



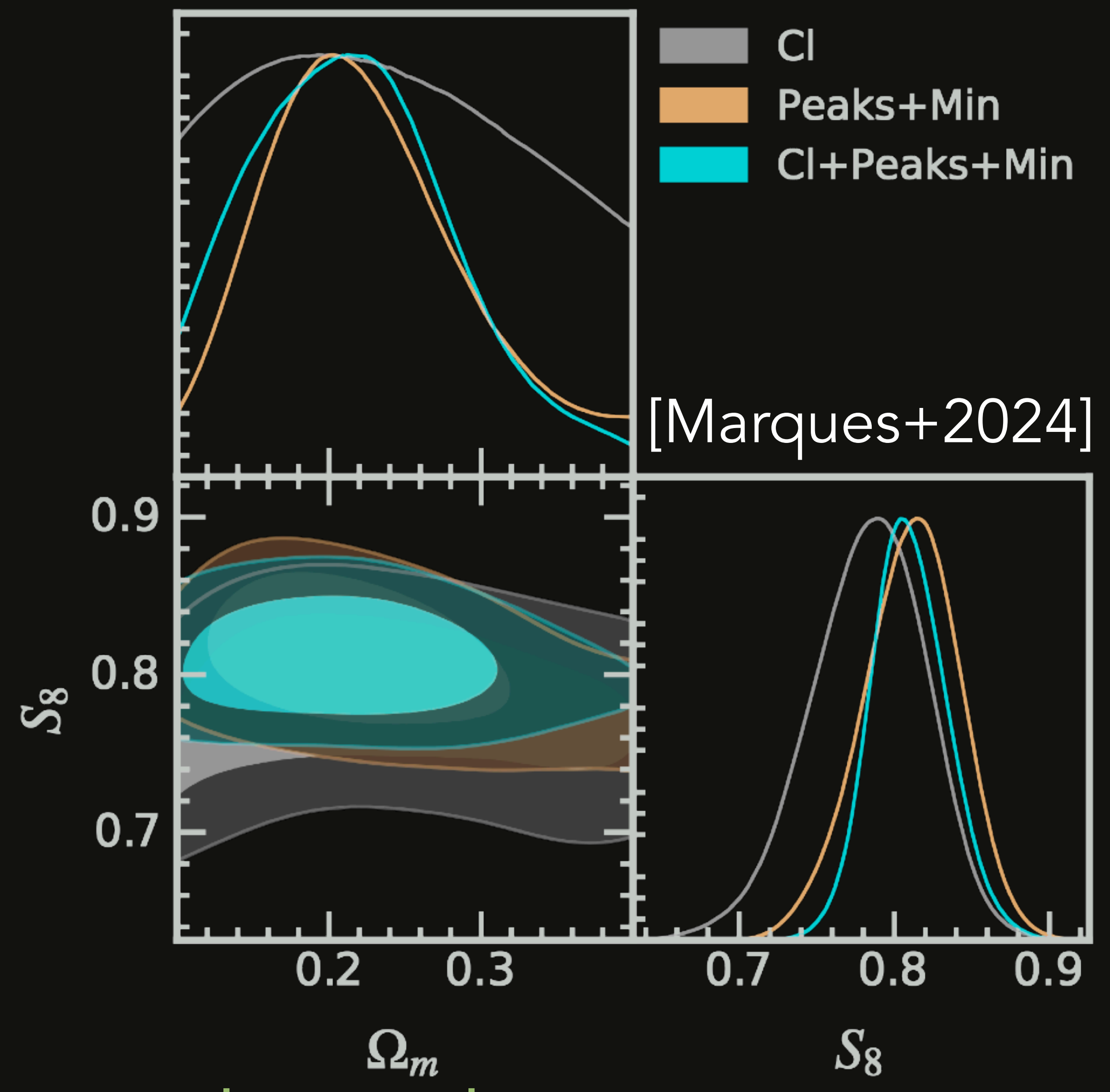
scattering transform



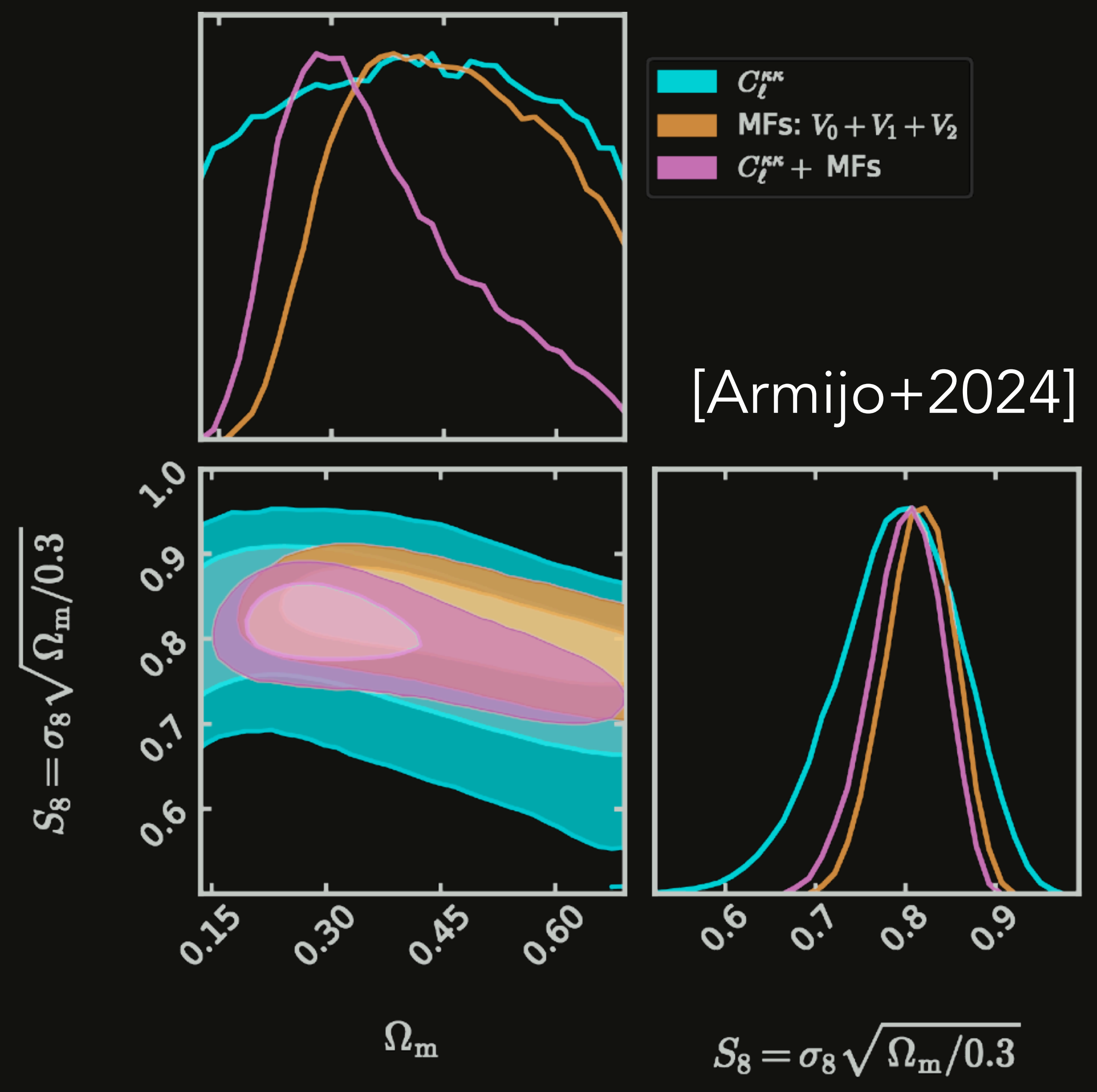
[Thiele+2023]

Probability distribution function

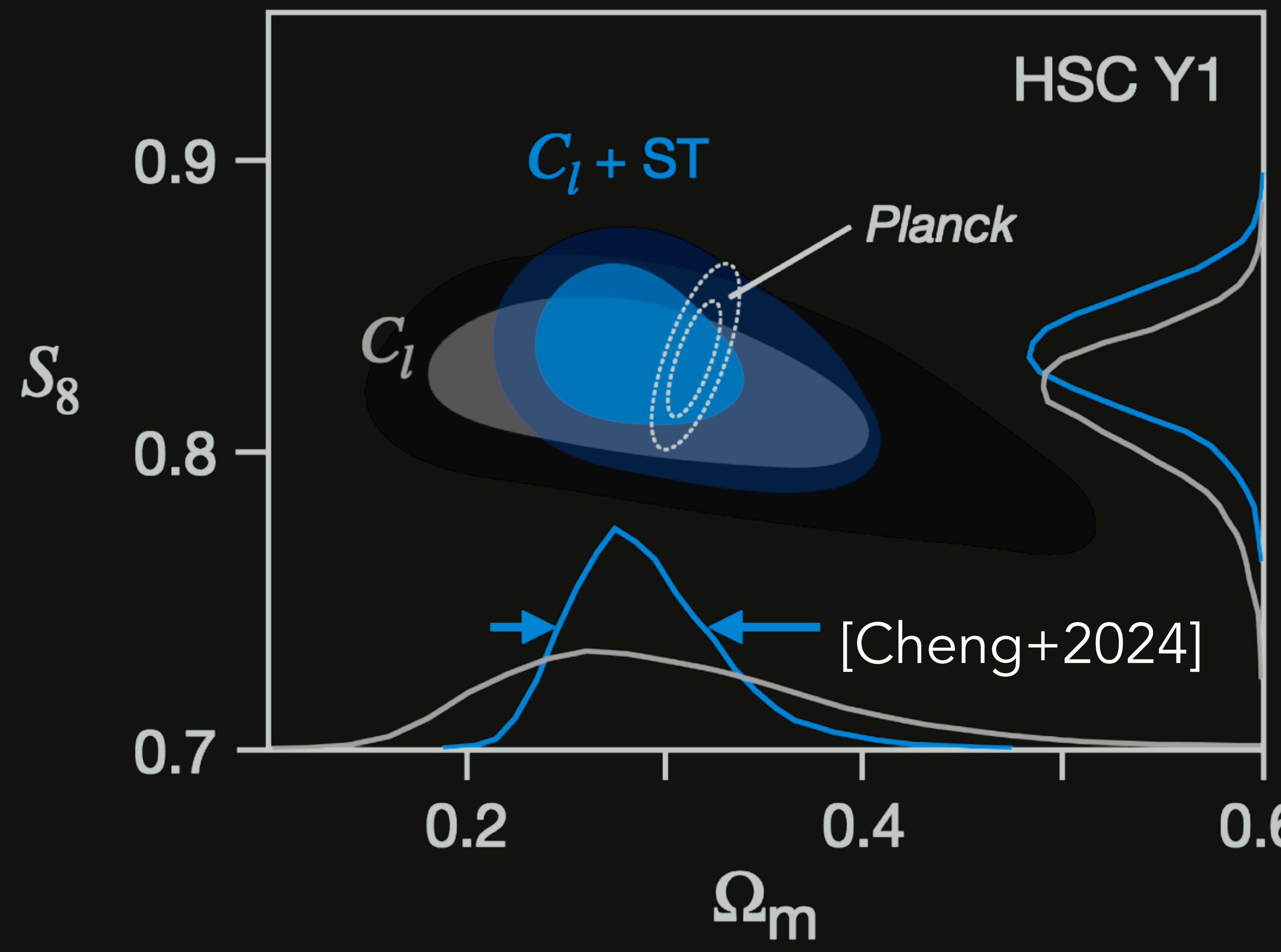
Cosmology from HSC year-1 higher order statistics



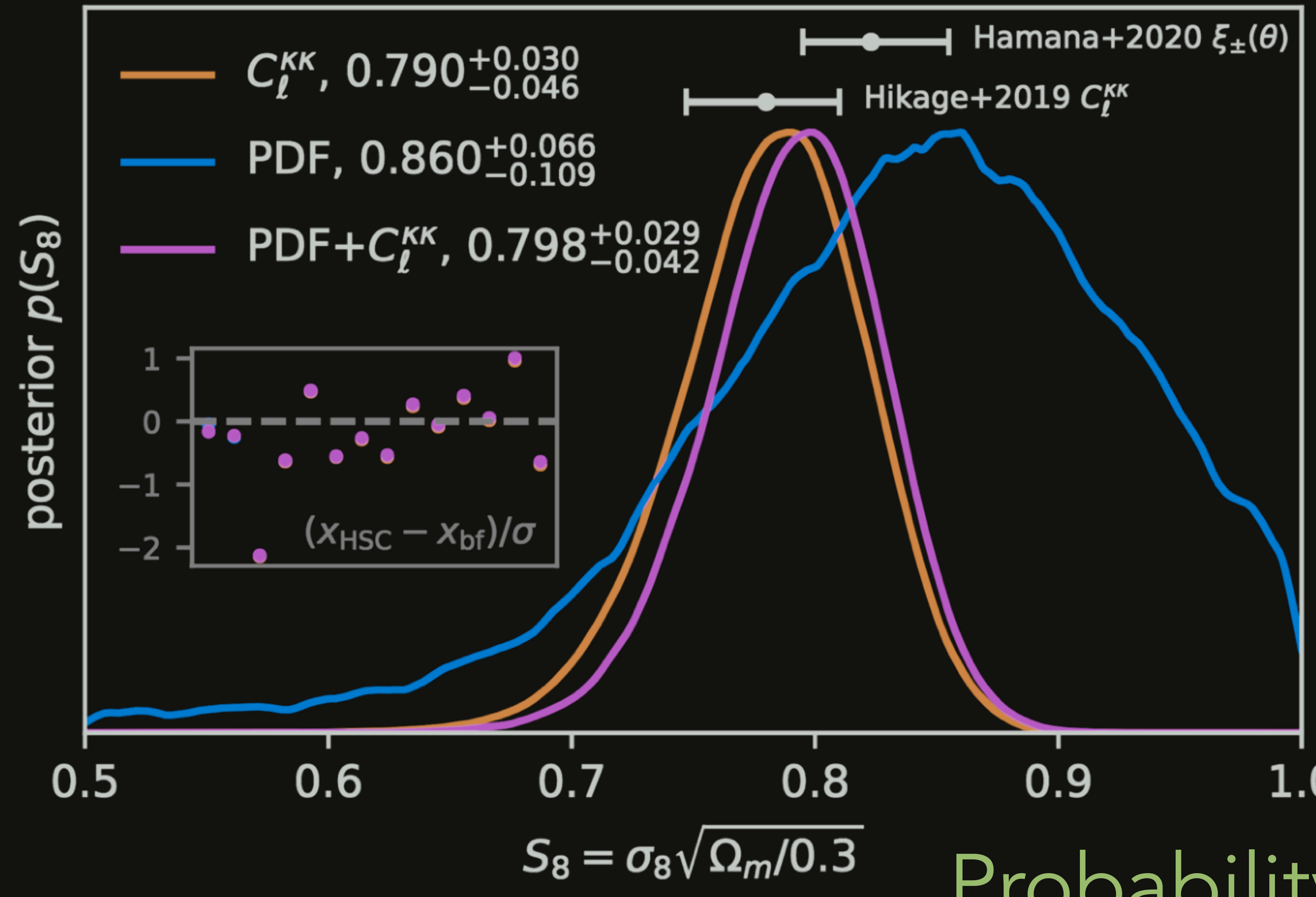
peaks and minima



Minkowski functionals



scattering transform



[Thiele+2023]

Probability distribution function

Can we avoid Gaussian likelihood approximation?

To get the best constraints from our data!

Perhaps even neural summary statistics?

Bayesian inference

likelihood

prior

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

posterior

evidence

Bayesian inference

More concretely:

θ=interesting parameters, η=nuisance parameters, ζ=initial conditions,

x=data, m=model:

$$P(x \mid \theta) = \int D\eta \, D\zeta \, \delta[\, x - m(\theta, \eta, \zeta) \,]$$

$$\underbrace{P(\text{parameters} \mid \text{data})}_{\text{posterior}} = \frac{\overbrace{P(\text{data} \mid \text{parameters})}^{\text{likelihood}} \; \overbrace{P(\text{parameters})}^{\text{prior}}}{\underbrace{P(\text{data})}_{\text{evidence}}}$$

Bayesian inference

More concretely:

θ=interesting parameters, η=nuisance parameters, ζ=initial conditions,

x=data, m=model:

$$P(x \mid \theta) = \int D\eta\, D\zeta\, \delta[\, x - m(\theta, \eta, \zeta)\,]$$

Traditional case:

$$P(x \mid \theta) = \int D\eta\ \text{Gaussian}[\, x - \mu(\theta, \eta), \Sigma\,]$$

[do remaining low-dimensional η-integral with Monte Carlo]

Bayesian inference

More concretely:

θ=interesting parameters, η=nuisance parameters, ζ=initial conditions,

x=data, m=model:

$$P(x \mid \theta) = \int D\eta\, D\zeta\, \delta[\, x - m(\theta, \eta, \zeta)\,]$$

Traditional case:

$$P(x \mid \theta) = \int D\eta\, \text{Gaussian}[\, x - \mu(\theta, \eta), \Sigma\,]$$

But what do we do if the Gaussian approximation doesn't hold?

→ assume we have simulator that evaluates m(θ, η, ζ) accurately

Neural Implicit Likelihood Inference (ILI)

More concretely:

θ=interesting parameters, η=nuisance parameters, ζ=initial conditions, x=data, m=model:

$$P(x \mid \theta) = \int D\eta\, D\zeta\, \delta[\, x - m(\theta, \eta, \zeta)\,]$$

But what do we do if the Gaussian approximation doesn't hold?

→ assume we have simulator that evaluates $m(\theta, \eta, \zeta)$ accurately

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

$$P(\text{parameters} \mid \text{data}) = \frac{P(\text{data} \mid \text{parameters})}{P(\text{data})} P(\text{parameters})$$

neural ratio estimation (NRE)

Neural Implicit Likelihood Inference (ILI)

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

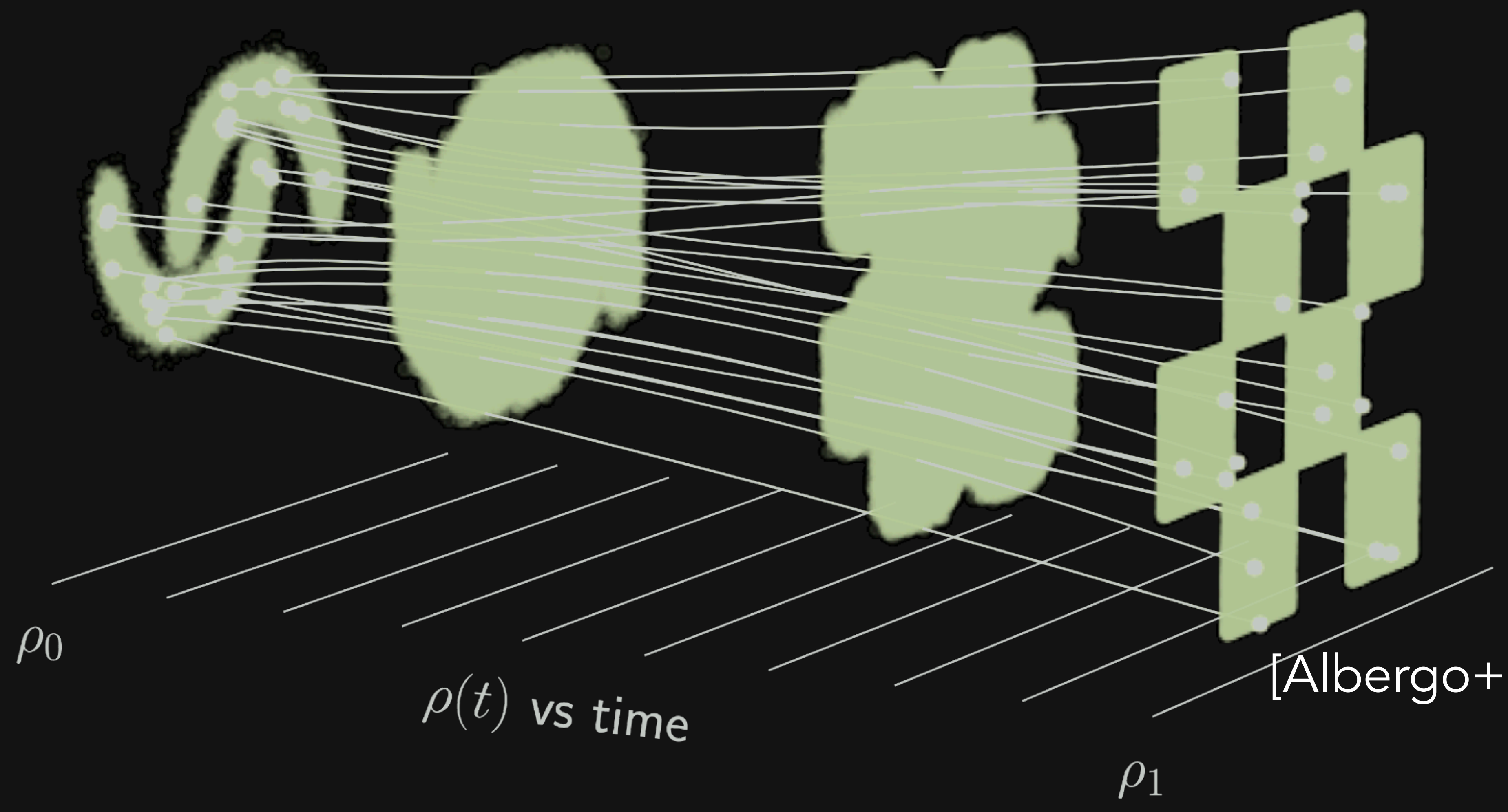
$$P(\text{parameters}\,|\,\text{data}) = \frac{P(\text{data}\,|\,\text{parameters})\;P(\text{parameters})}{P(\text{data})}$$

neural ratio estimation (NRE)

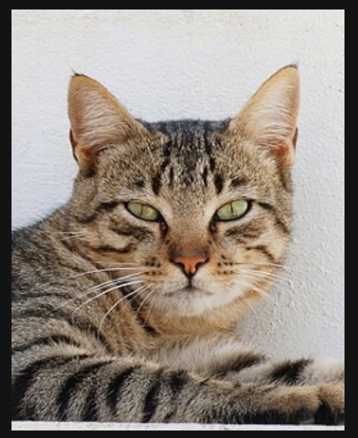
neural likelihood estimation (NLE):
conditioned flow q(x|θ)

neural posterior estimation (NPE):
conditioned flow q(θ|x)

neural ratio estimation (NRE):
classifier between (x, θ) ~p(x, θ) and ~p(x)p(θ)



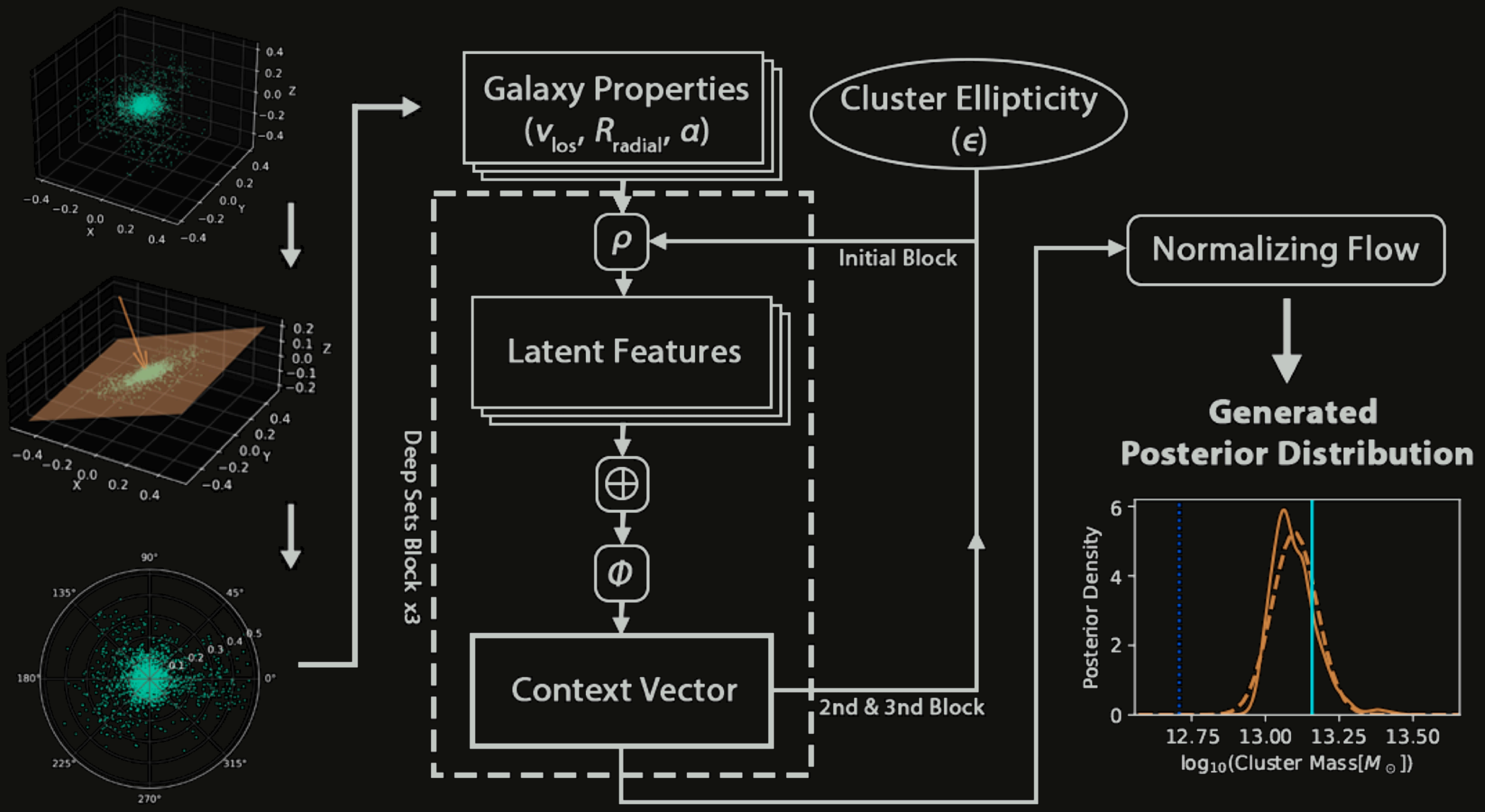
[Albergo+2023]



?

Practicality: data size & structure

Usually, data vectors too high dimensional or live in awkward spaces
→ We deal with this by constructing a useful latent representation through *embedding networks*



[B.Y.Wang & LT 2025]

Neural Implicit Likelihood Inference (ILI)

- *in the limit*, any likelihood learnable
- any simulate-able effect can be incorporated
- no formal difference between nuisance parameters and initial conditions
- primary choice at the moment:
  - NPE: empirically good performance, need to deal with flow
  - NRE: classification → super flexible, empirically more tuning required

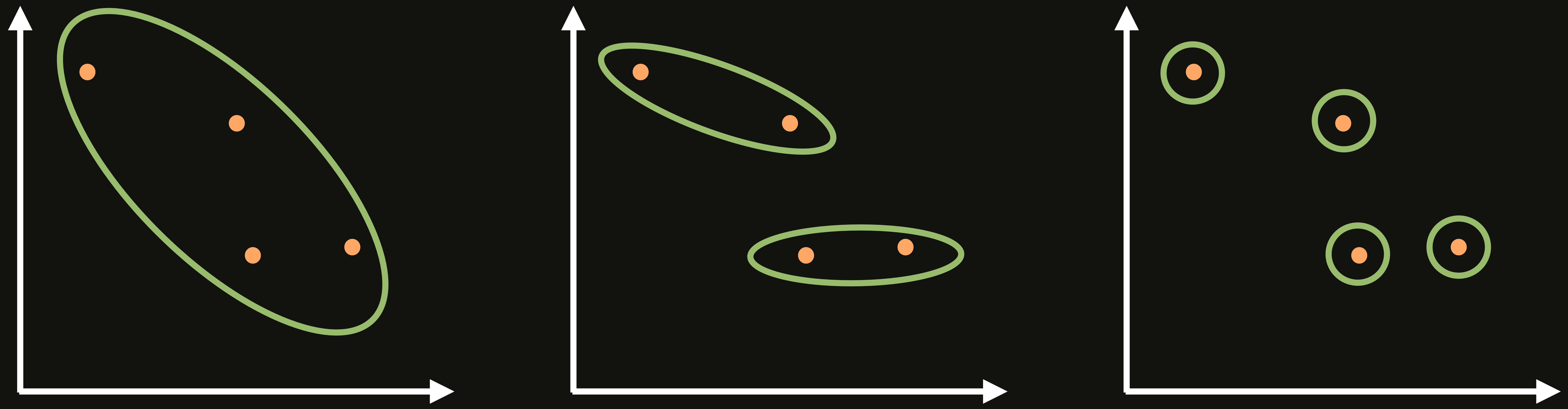
Curse of dimensionality

Let's say: 200,000 simulations, 10 model parameters, 10 data points
→ this is optimistic!

We're trying to learn a density in this 10+10 dimensional space

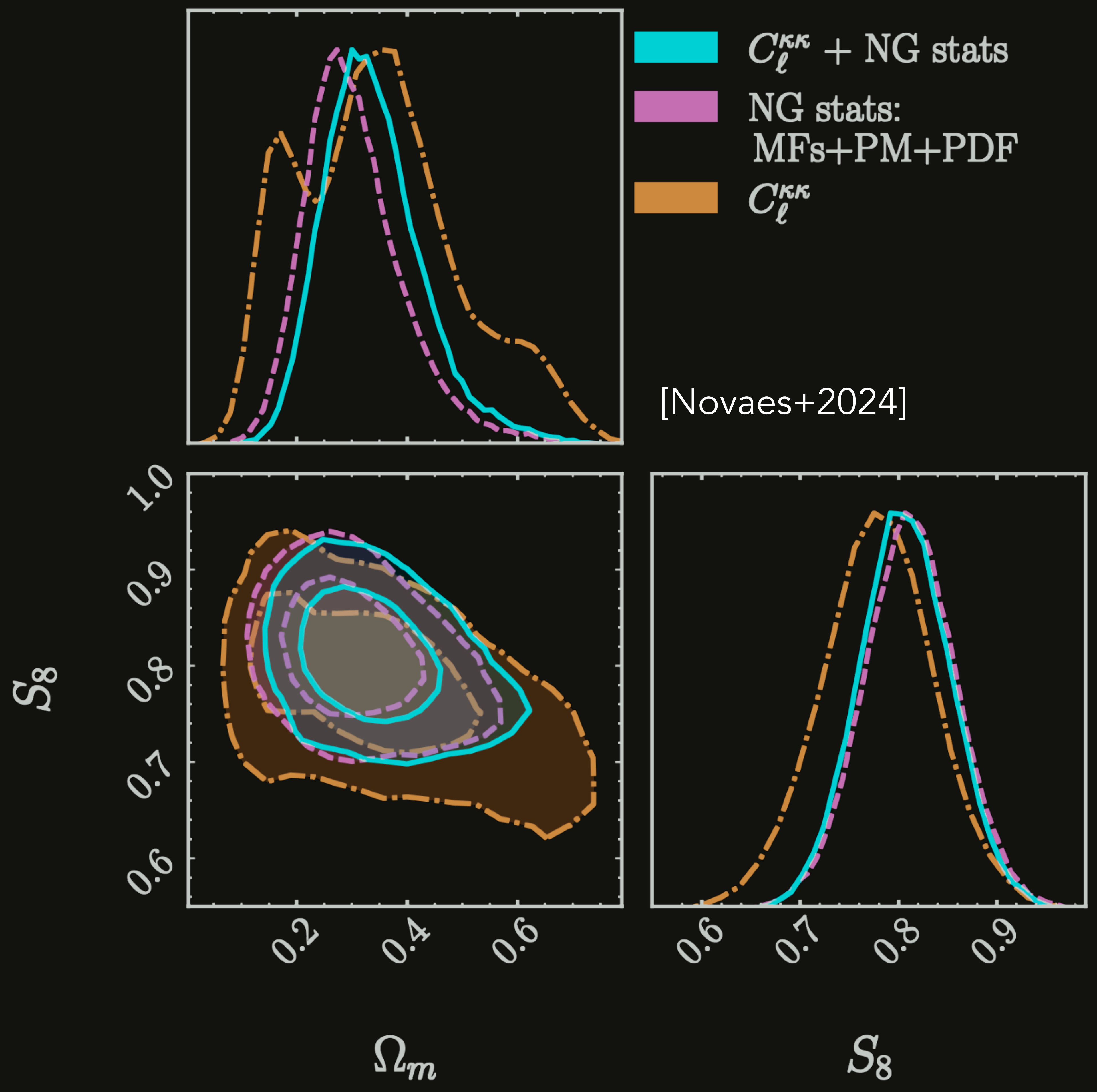
$$(200{,}000)^{1\,/\,(10\,+\,10)} = 1.8, \text{ let's say } 2$$

What is the implicit prior?

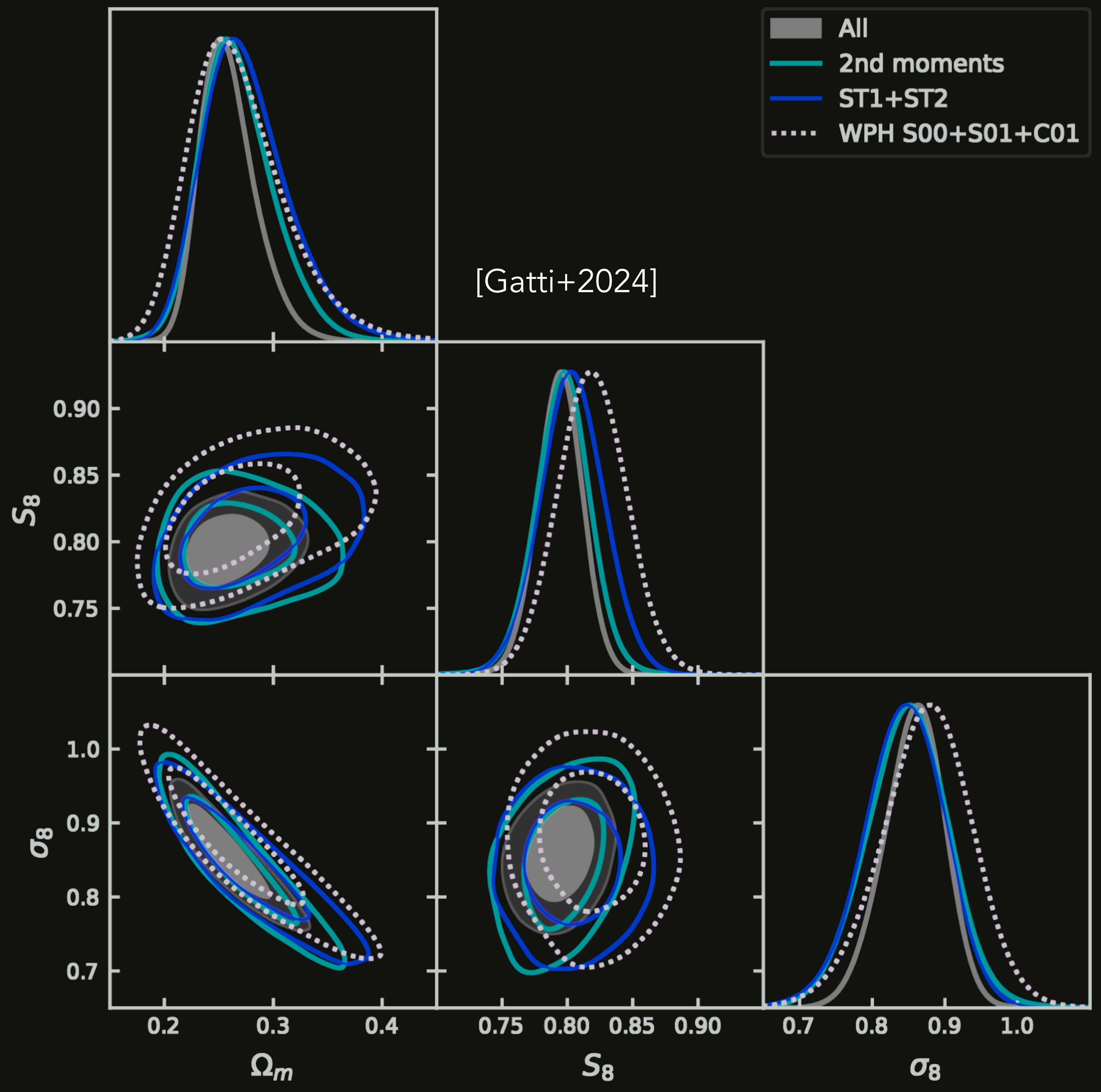


ILI applied in weak lensing

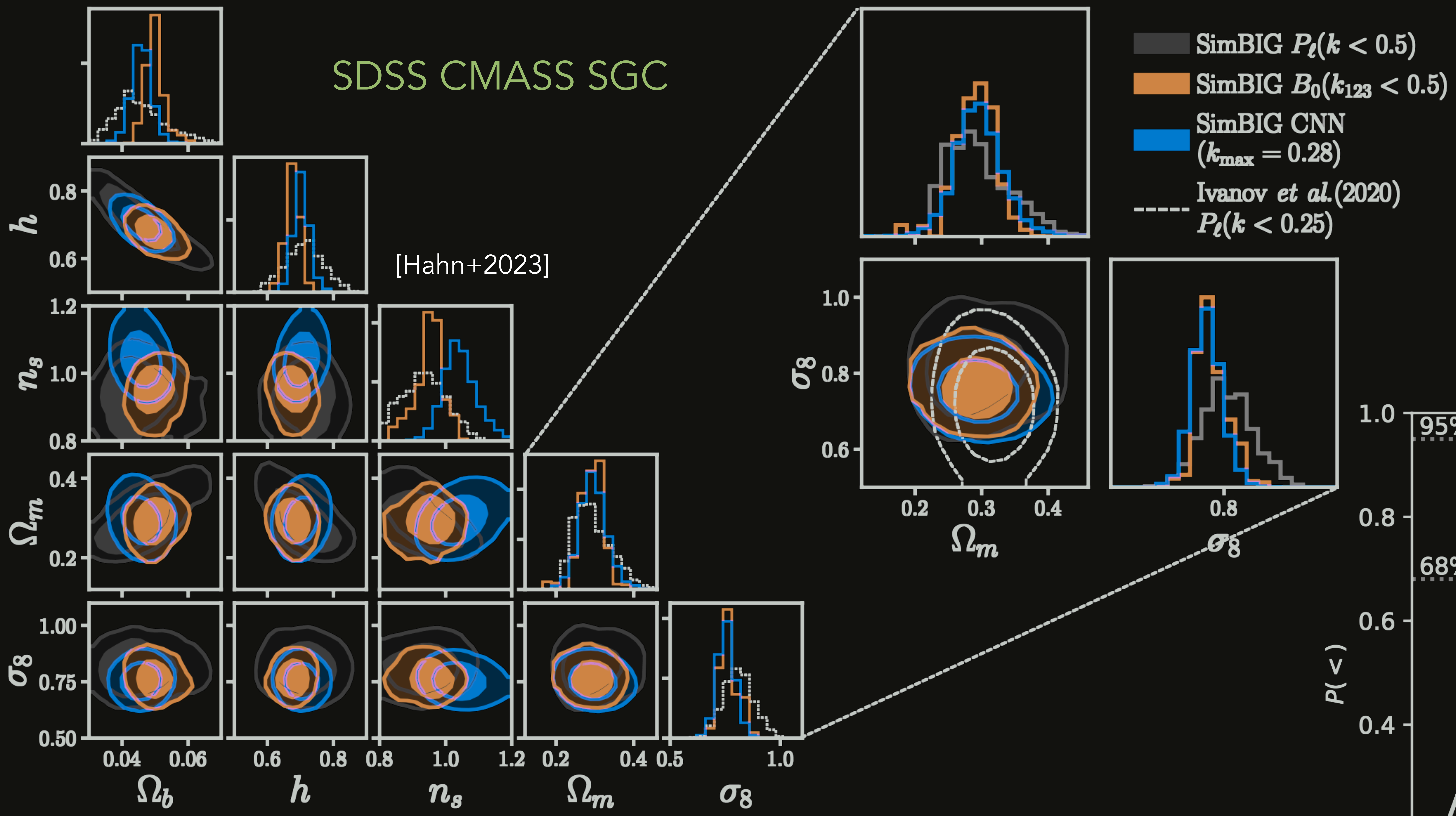
HSC Y1

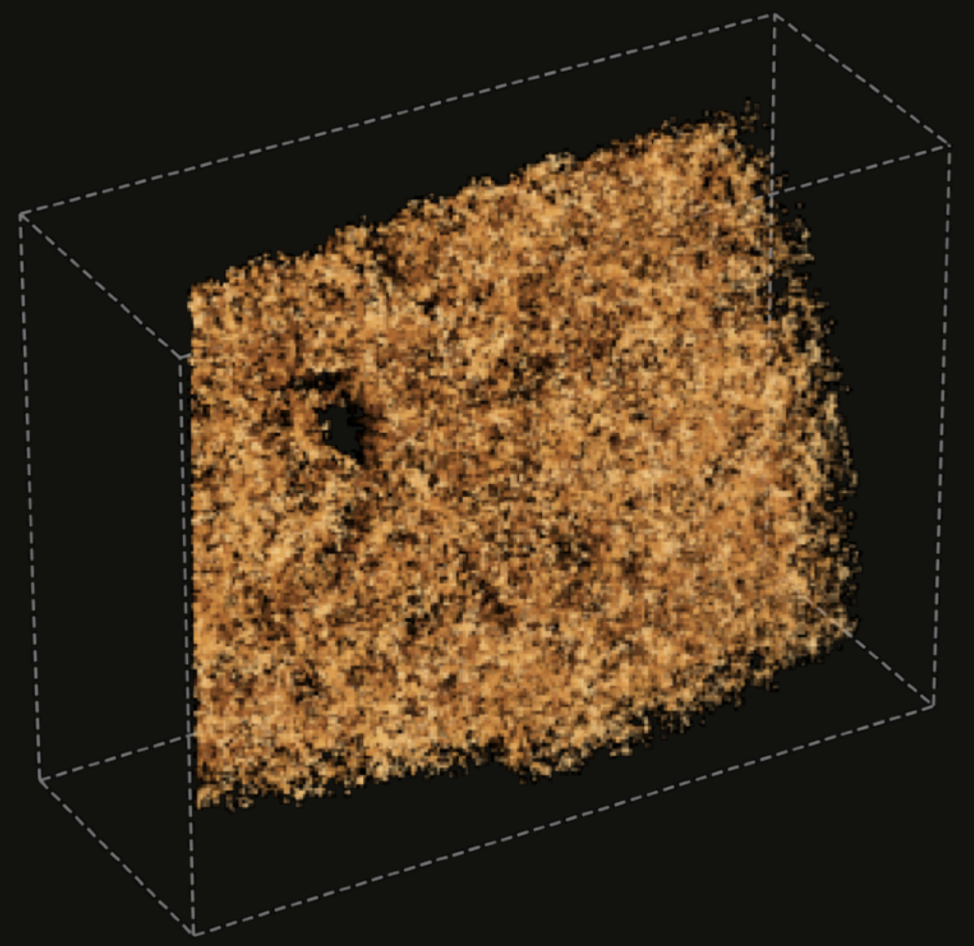


DES Y3

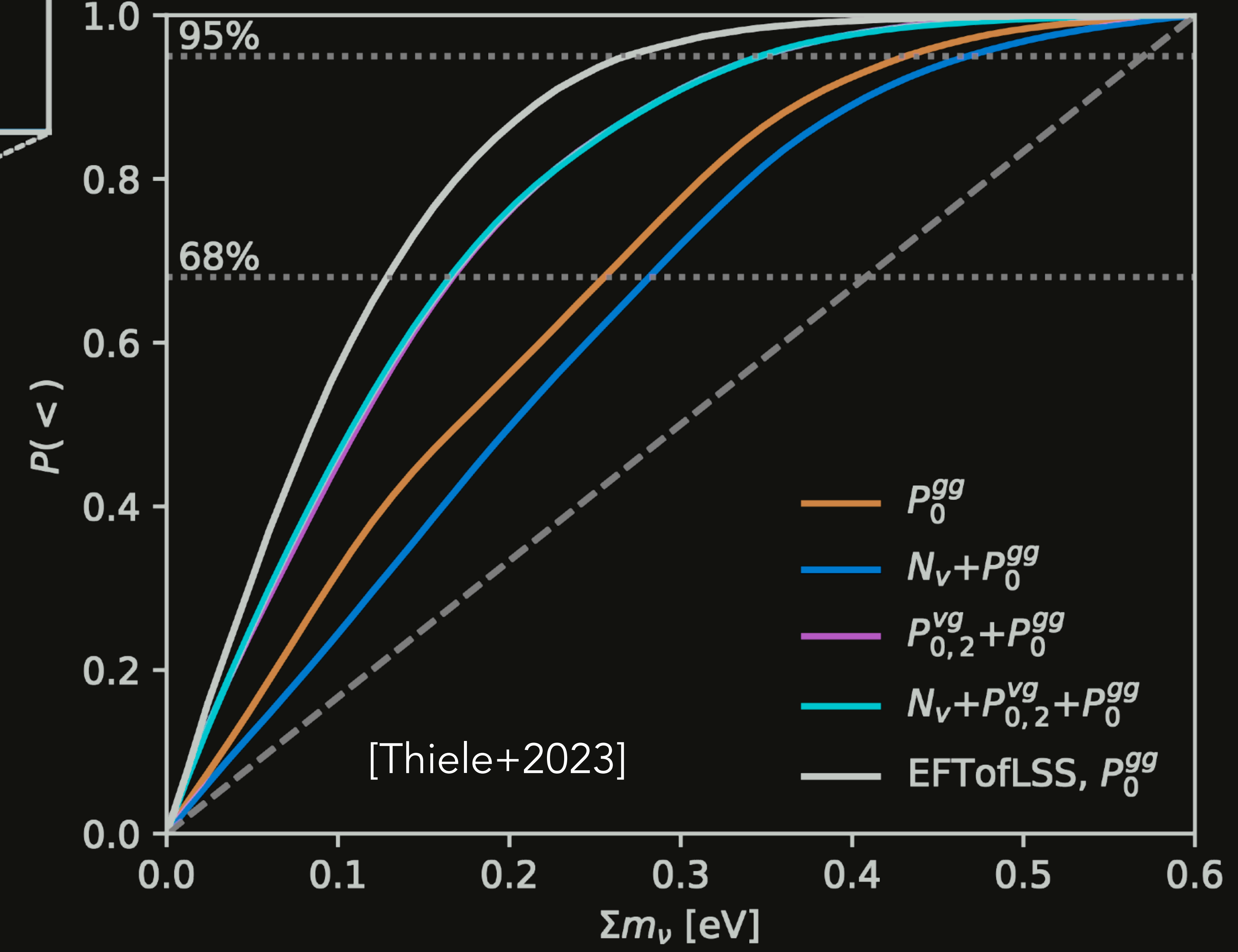


ILI applied in galaxy clustering

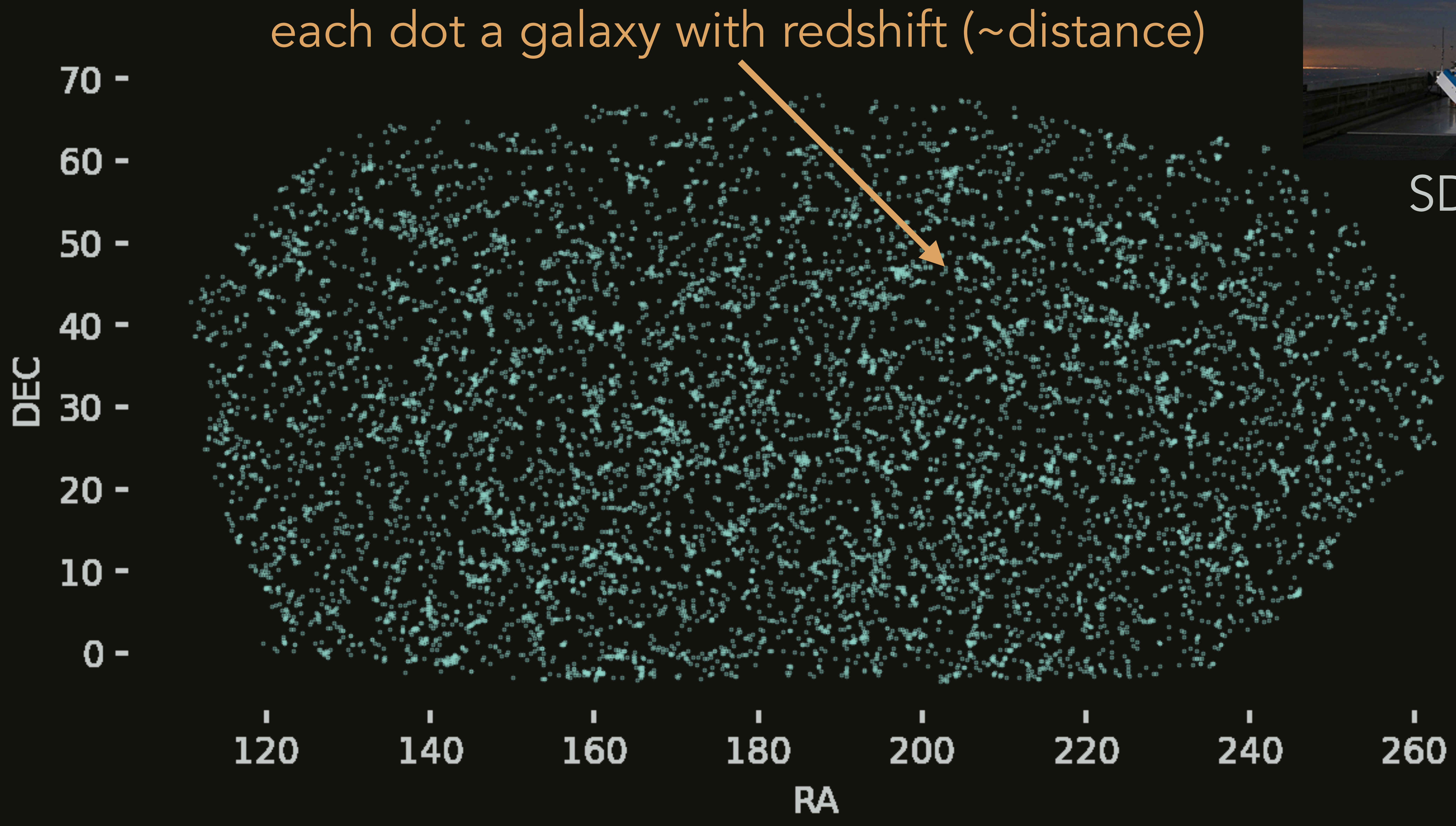


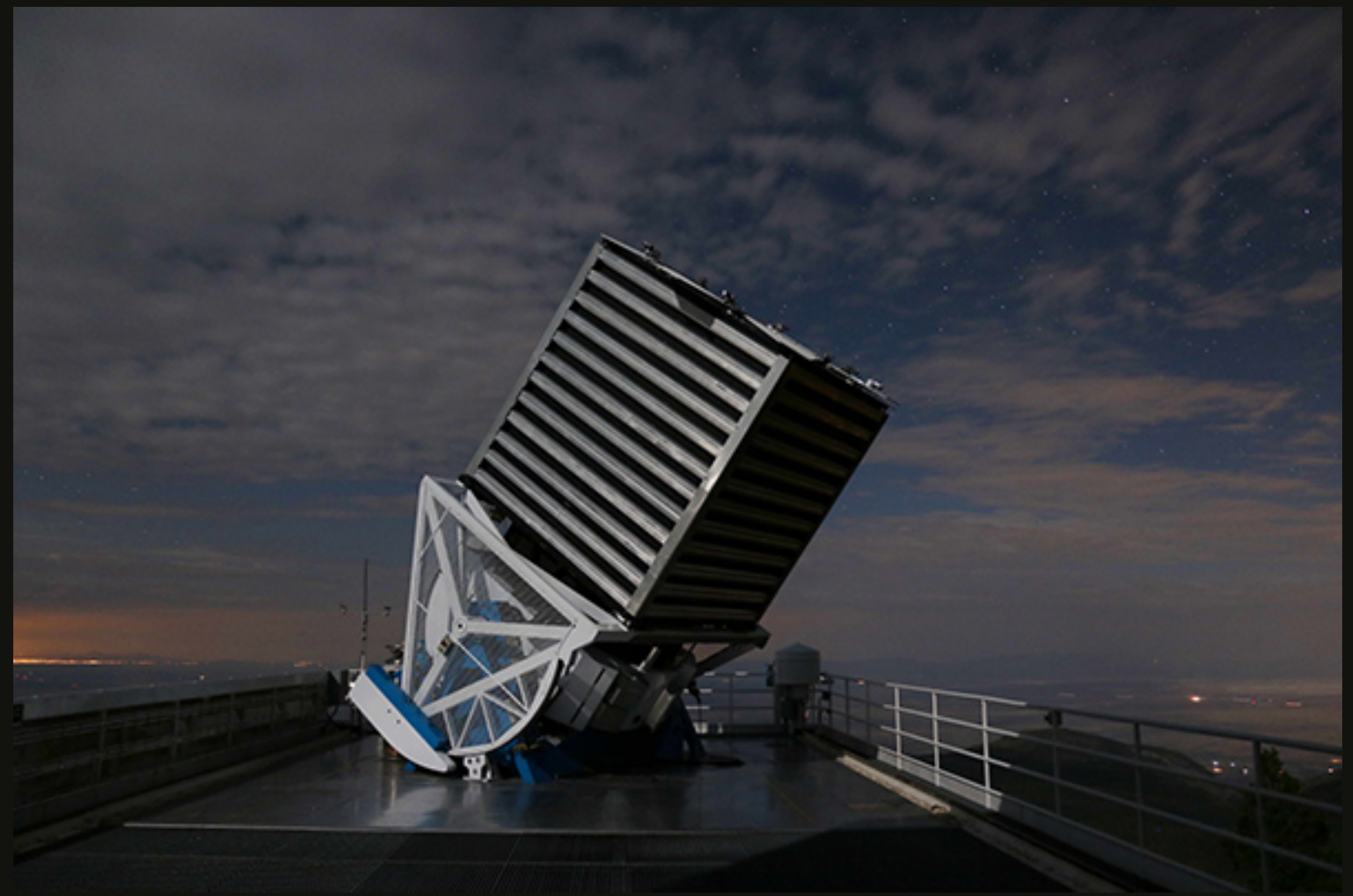


SDSS CMASS NGC voids



A 3-D Map of the Universe





SDSS/BOSS survey

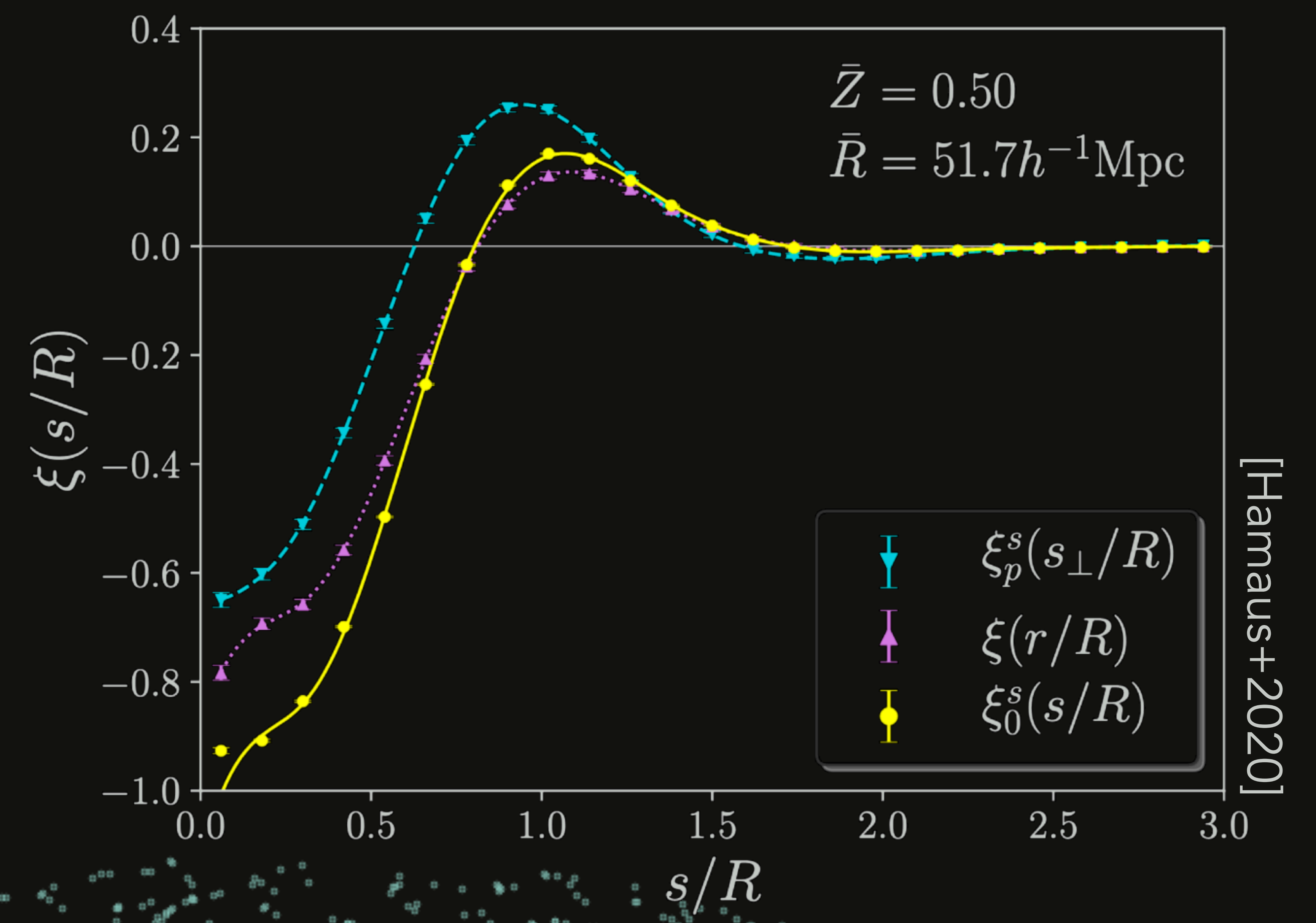
How to summarize this map?

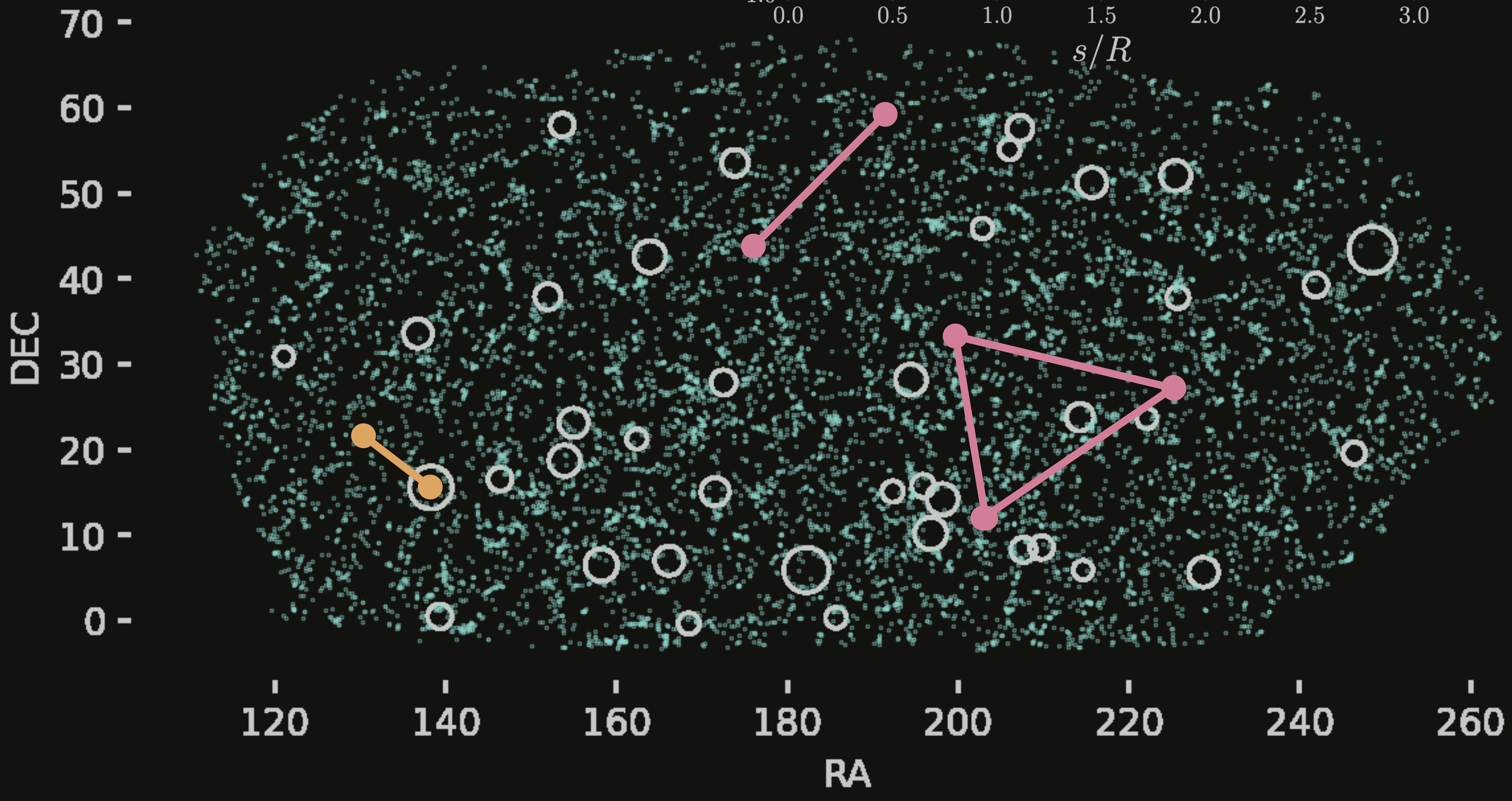
1) pairs of galaxies (power spectrum)

2) triangles of galaxies (bispectrum)

3) …

4) "empty regions": *cosmic voids*
   - size distribution
   - void-galaxy pairs
   - …

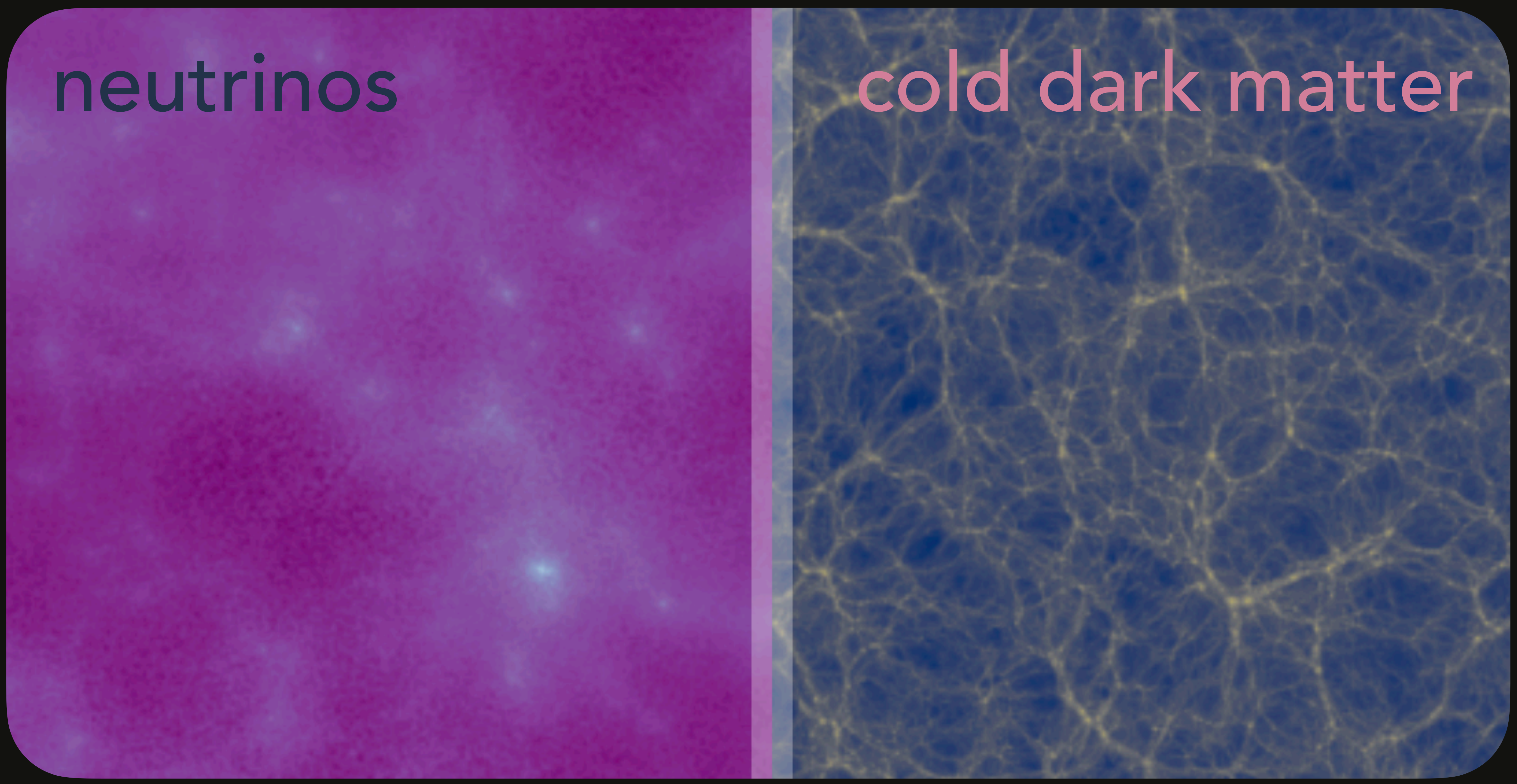


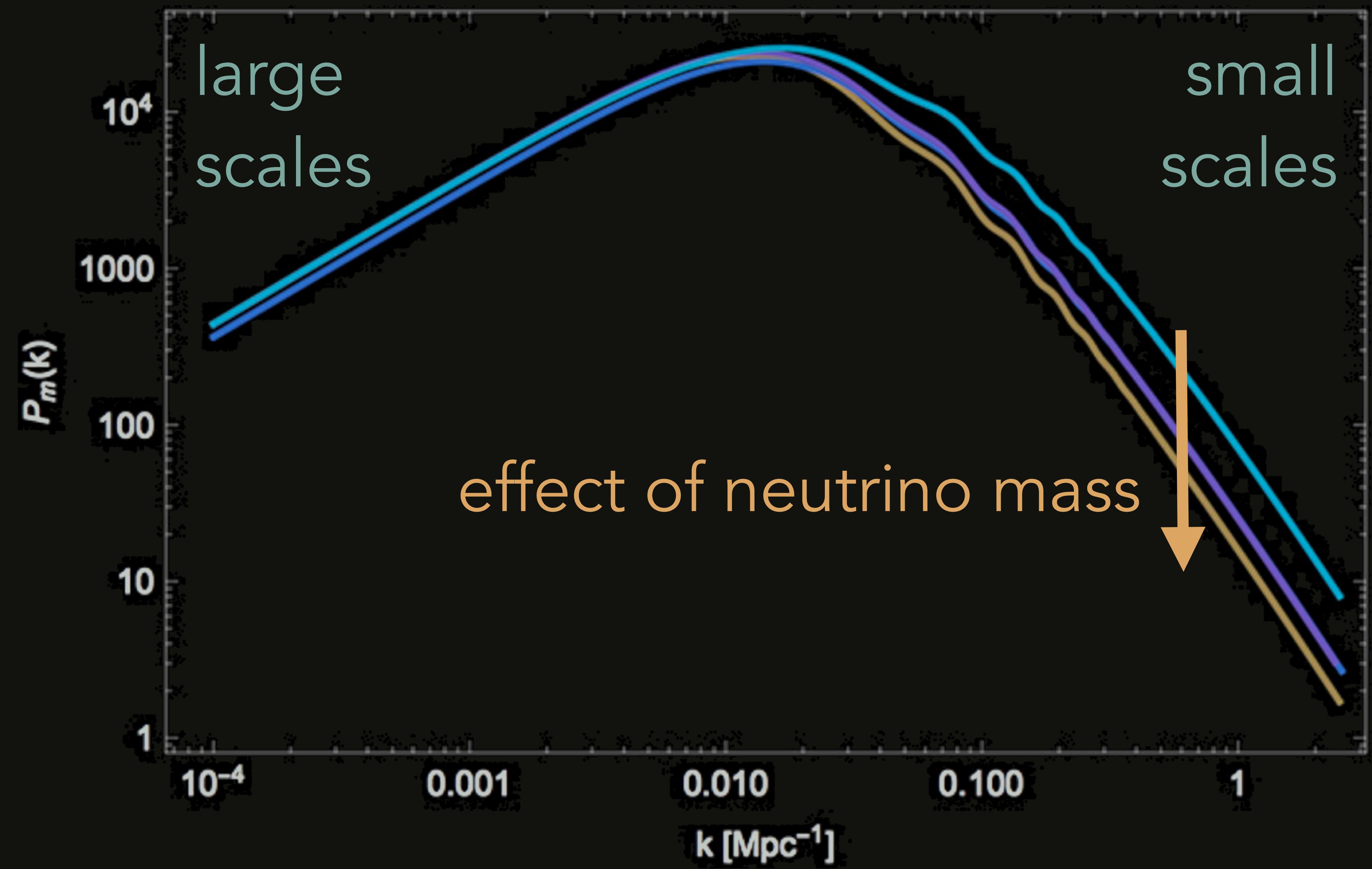


What can voids do for us?

- upweight underdensities → complementary to correlation functions
  - corrections to general relativity
  - dark energy
  - neutrino mass

[FLAMINGO]

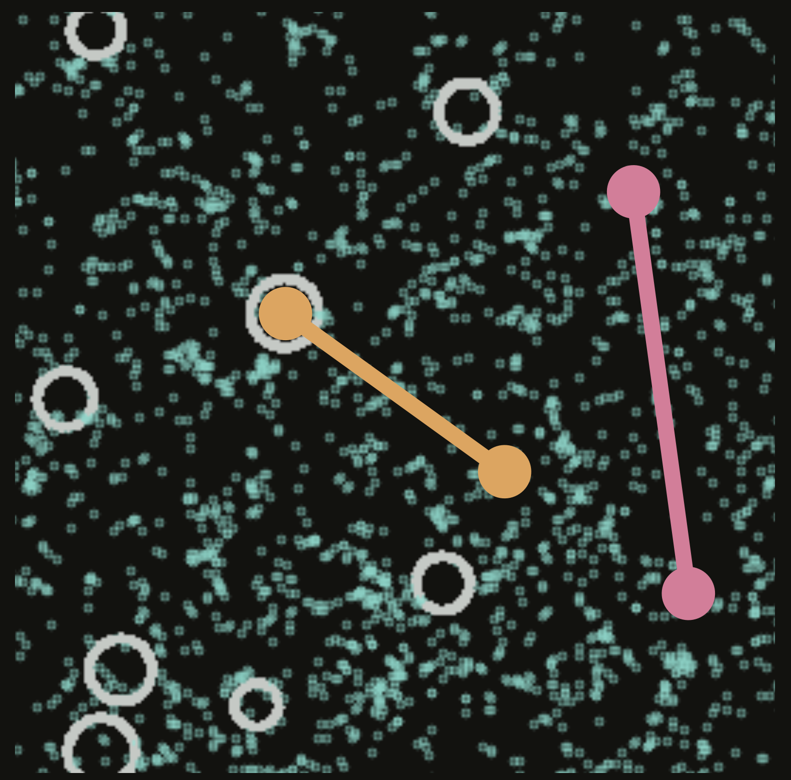




[Gerbino&Lattanzi2018]

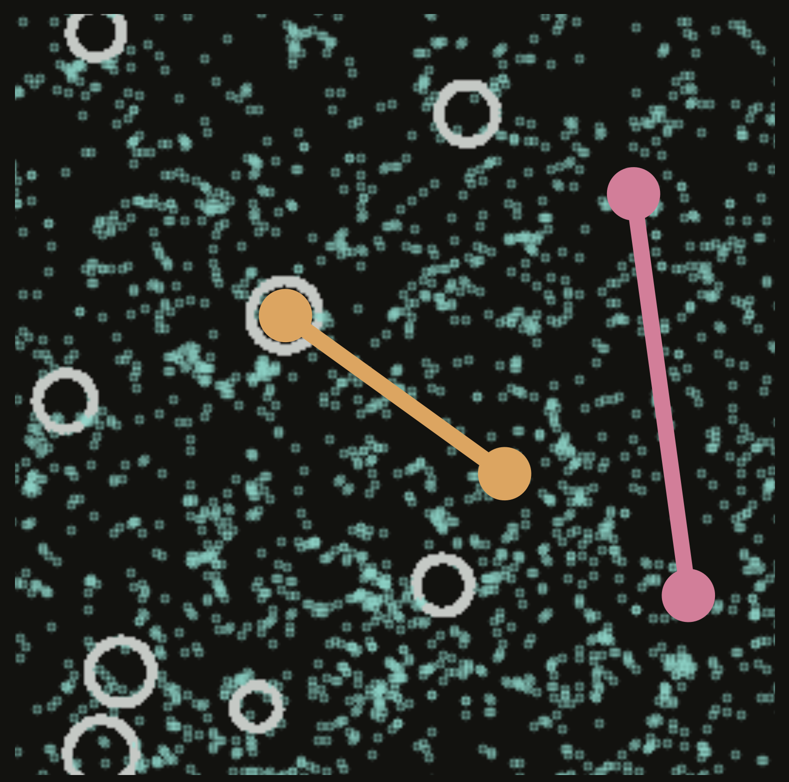
Simulation-based inference

- Want to constrain neutrino mass sum, $\sum m_\nu$, with BOSS data:
  - galaxy auto power spectrum
  - void size function (histogram of void sizes)
  - void galaxy cross power spectrum (void profile)

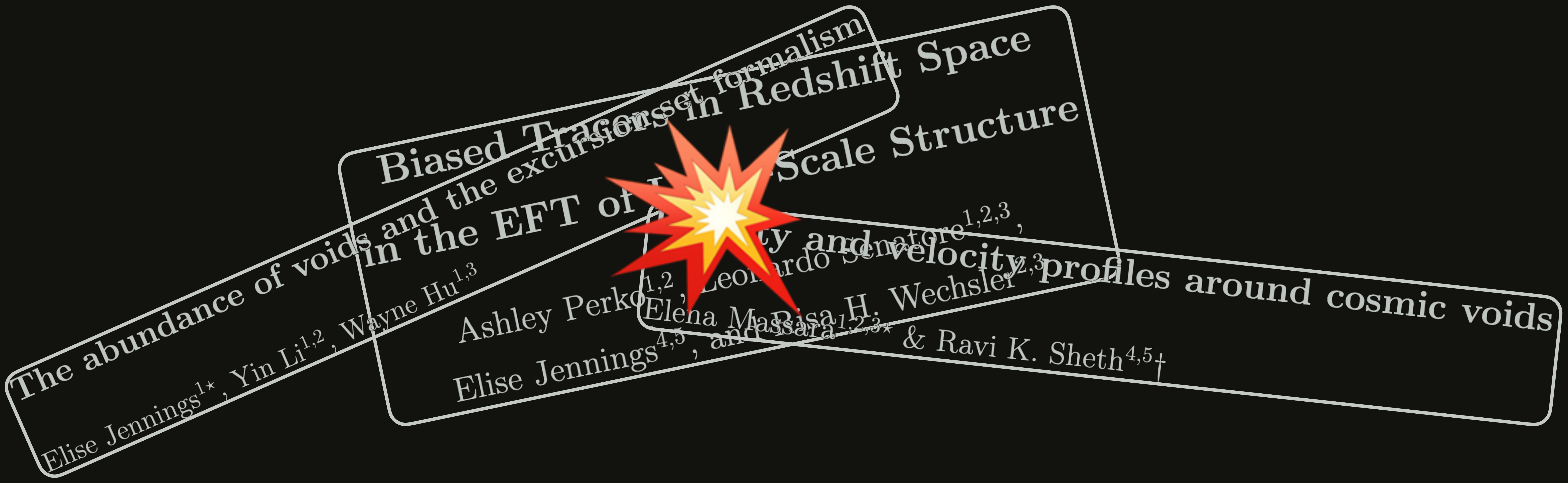


Simulation-based inference

- Want to constrain neutrino mass sum, $\sum m_\nu$, with BOSS data:
  - galaxy auto power spectrum
  - void size function (histogram of void sizes)
  - void galaxy cross power spectrum (void profile)

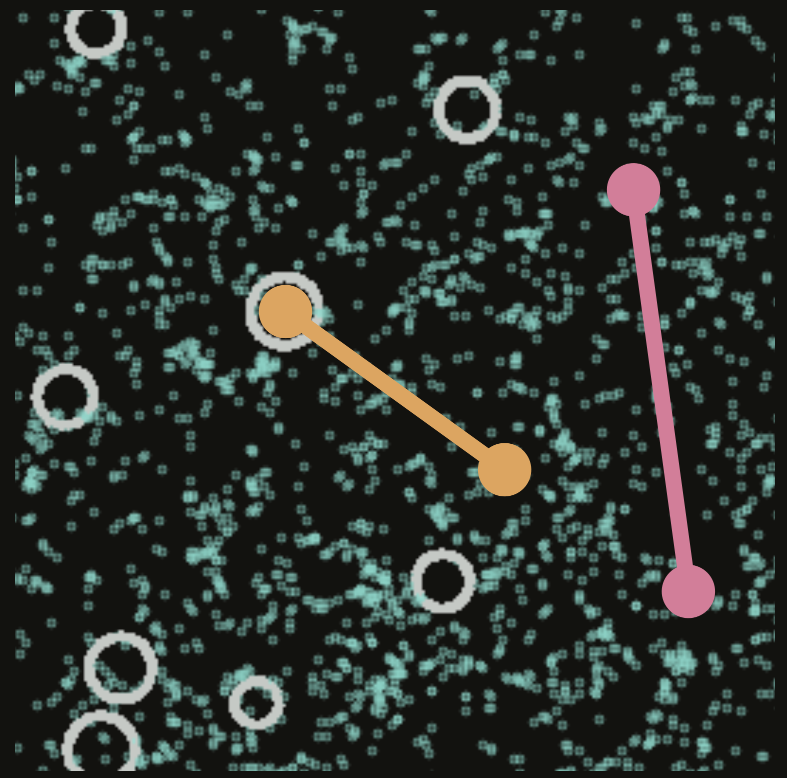


- Joint modeling of these statistics difficult with analytic methods



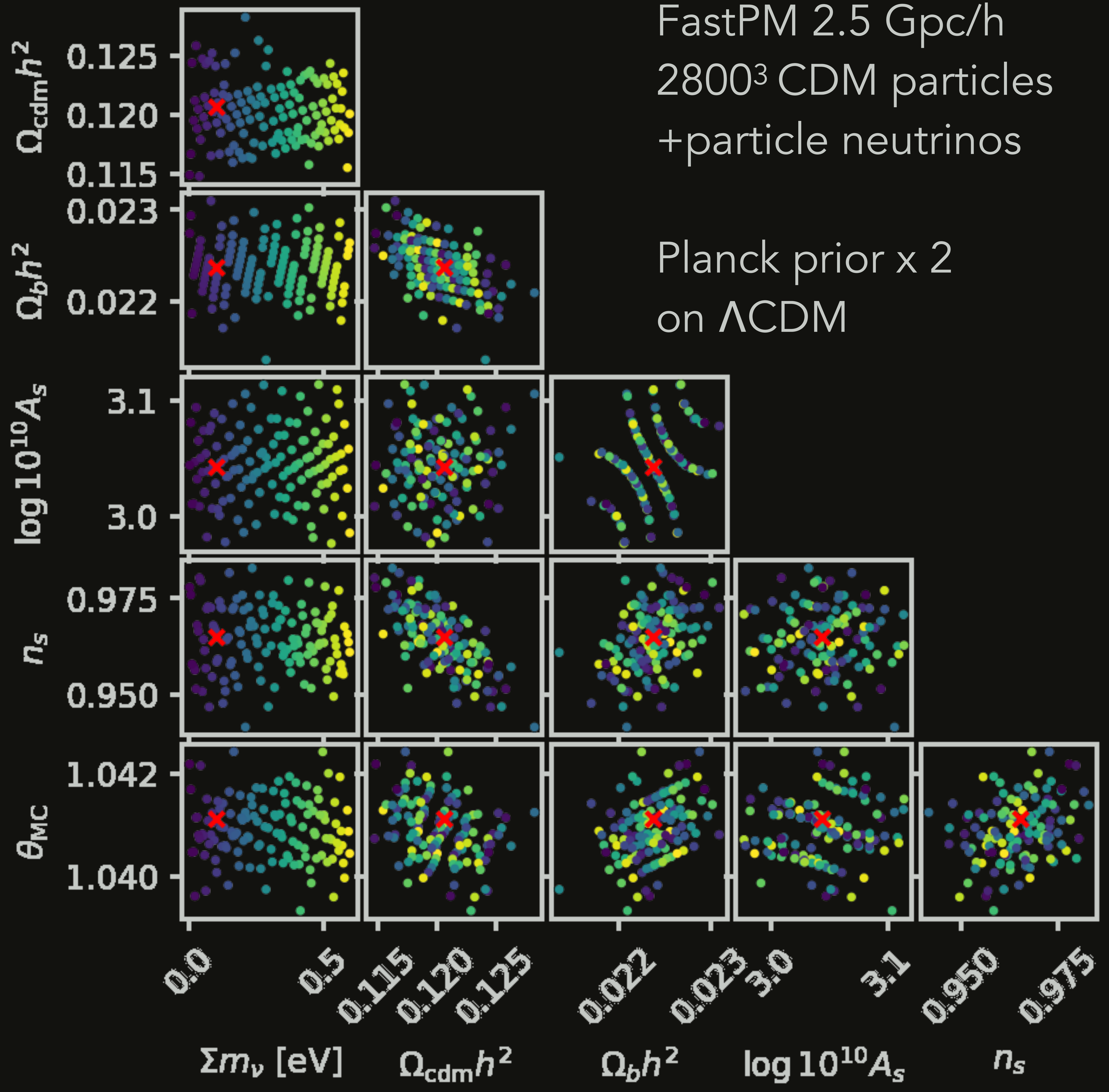
Simulation-based inference

- Want to constrain neutrino mass sum, $\sum m_\nu$, with BOSS data:
  - galaxy auto power spectrum
  - void size function (histogram of void sizes)
  - void galaxy cross power spectrum (void profile)



- Joint modeling of these statistics difficult with analytic methods
- Thus, resort to simulations (PM + HOD)
- Likelihood unknown → Implicit-Likelihood Inference

Simulations



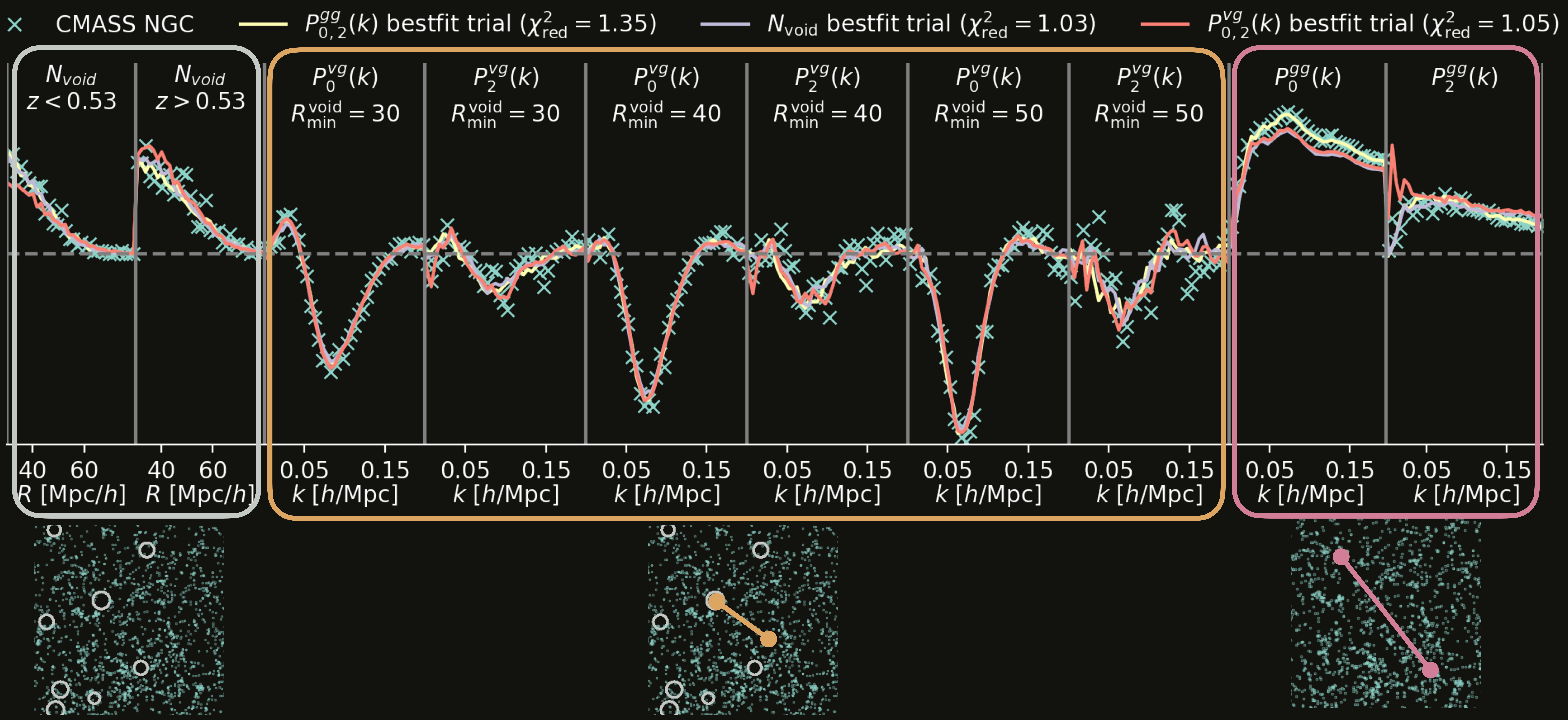
FastPM 2.5 Gpc/h
2800^3 CDM particles
+particle neutrinos

Planck prior x 2
on ΛCDM

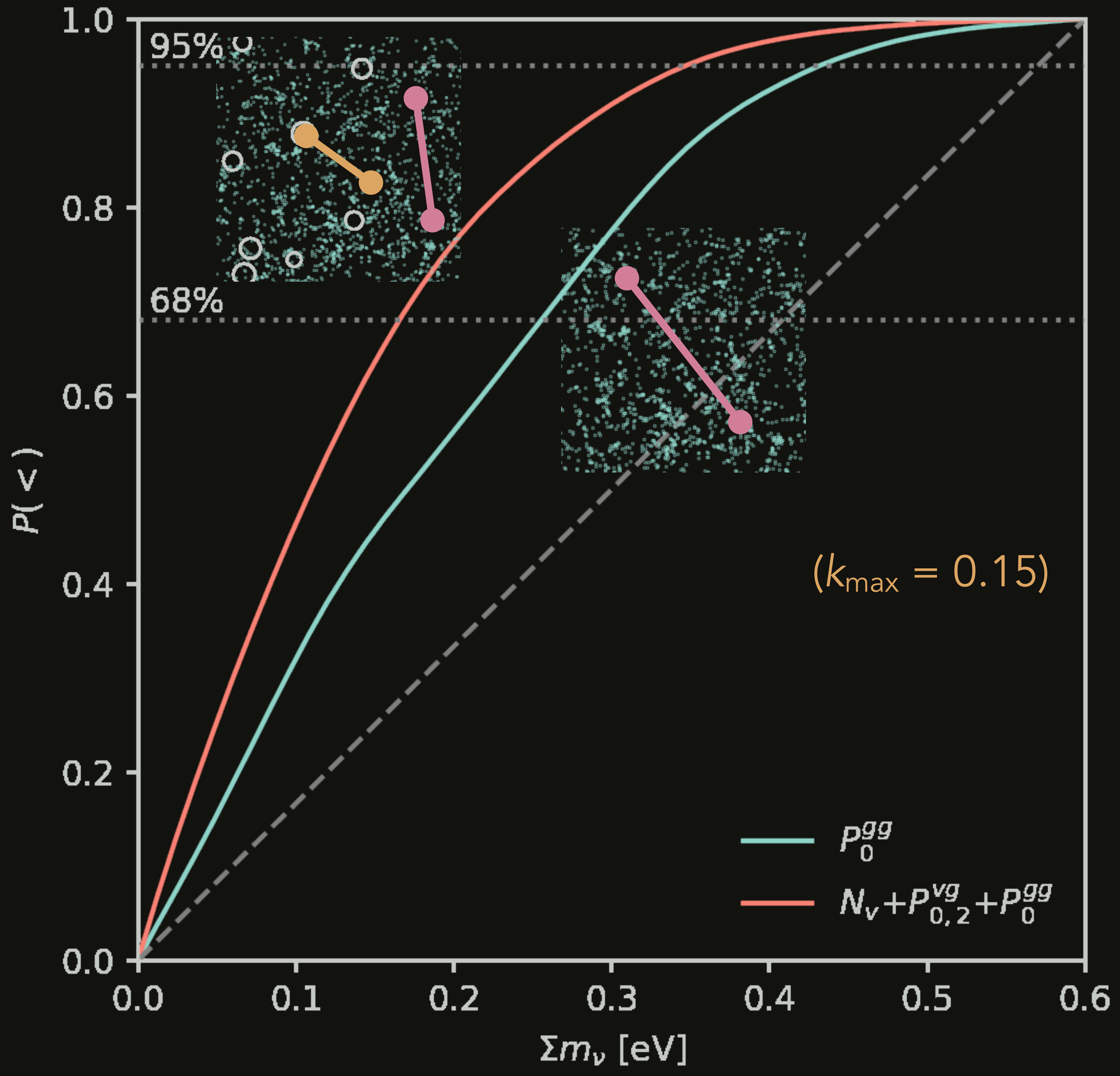
- populate gravity-only simulations with galaxies using HOD
- project on lightcone and add survey realism

Data Vector

Use MOPED compression to reduce dimensionality.



Main posterior



With conservative scale cut of k_{max}=0.15 hMpc^{-1}, voids tighten upper bound on neutrino mass.

So far, more or less “toy examples”.

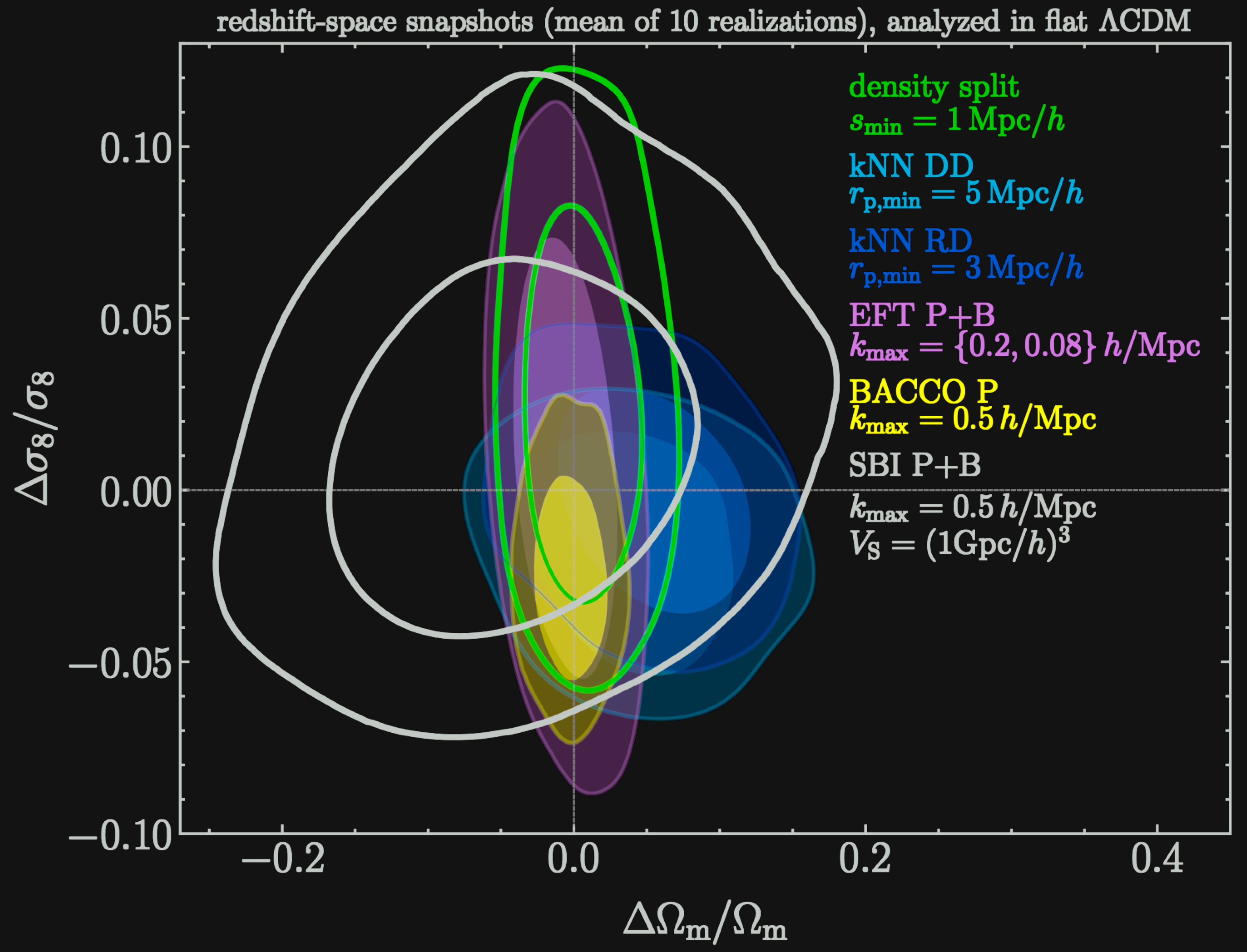
Not the same level of trust as traditional analyses.

A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering Statistics*

THE BEYOND-2PT COLLABORATION

ELISABETH KRAUSE,[1] YOSUKE KOBAYASHI,[1,2] ANDRÉS N. SALCEDO,[1] MIKHAIL M. IVANOV,[3] TOM ABEL,[4,5,6] KAZUYUKI AKITSU,[7] RAUL E. ANGULO,[8,9] GIOVANNI CABASS,[10] SOFIA CONTARINI,[11,12,13] CAROLINA CUESTA-LAZARO,[14,15,16] CHANGHOON HAHN,[17] NICO HAMAUS,[18,19] DONGHUI JEONG,[20,21] CHIRAG MODI,[22,23] NHAT-MINH NGUYEN,[24,25] TAKAHIRO NISHIMICHI,[2,26,27] ENRIQUE PAILLAS,[28,29] MARCOS PELLEJERO IBAÑEZ,[30] OLIVER H. E. PHILCOX,[31,32] ALICE PISANI,[33,22,34,17] FABIAN SCHMIDT,[35] SATOSHI TANAKA,[26] GIOVANNI VERZA,[36,22] SIHAN YUAN,[4,6] MATTEO ZENNARO,[37]

building trust



Issues away from the limit

- *in the limit*, any likelihood learnable
- what is the limit?
  - infinite model expressivity (usually ok in cosmology)
  - ability to find good global optimum (usually ok)
  - infinite training set size / fast & accurate simulation codes

$$L = \int_{p(x,\theta)} \mathscr{L} \approx \sum_{\text{training set}} \mathscr{L}_{\text{approx}}$$

Implicit Likelihood Inference in Crisis?

A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful

Joeri Hermans*
Unaffiliated

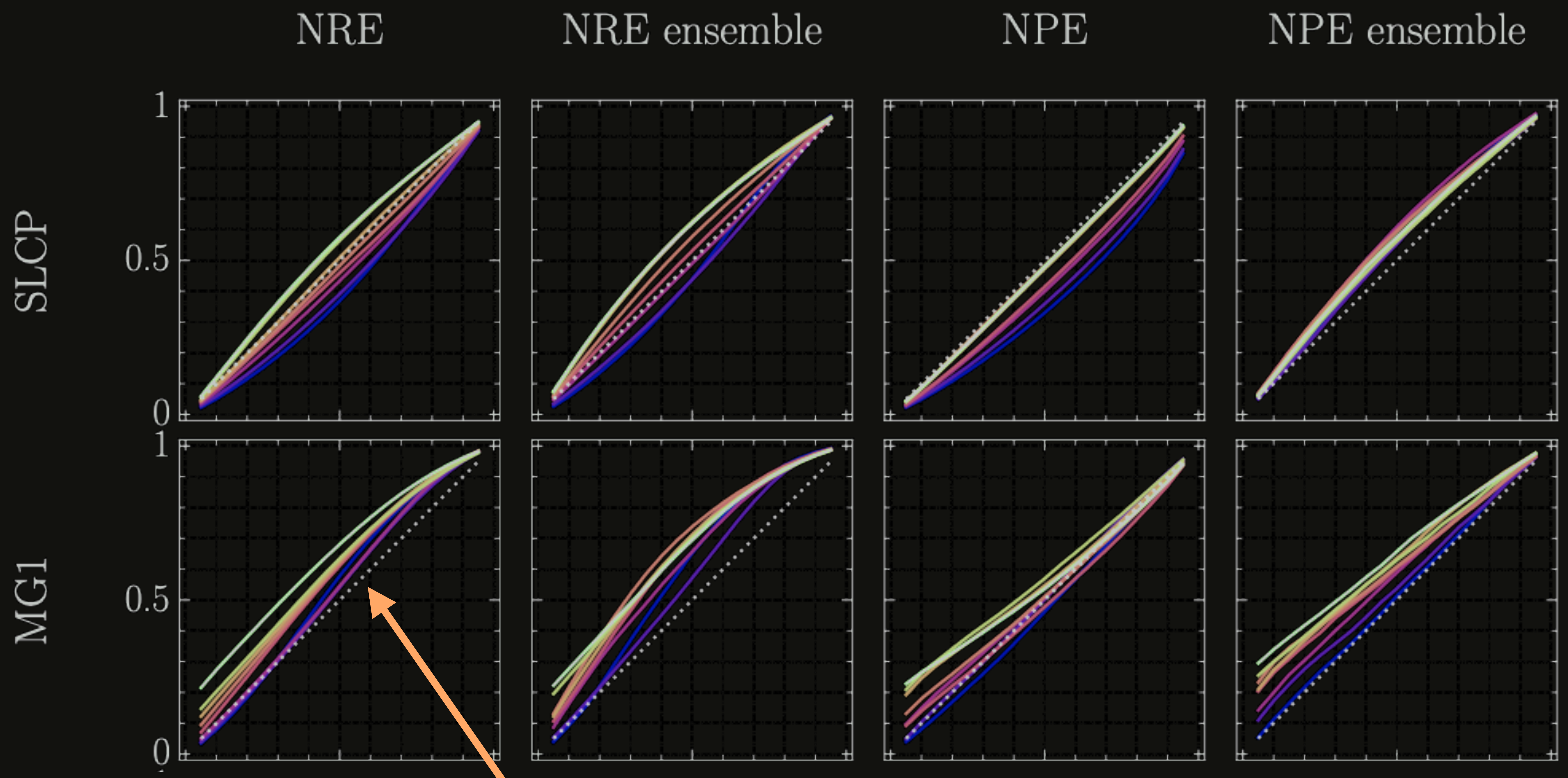
Arnaud Delaunoy*
University of Liège

François Rozet
University of Liège

Antoine Wehenkel
University of Liège

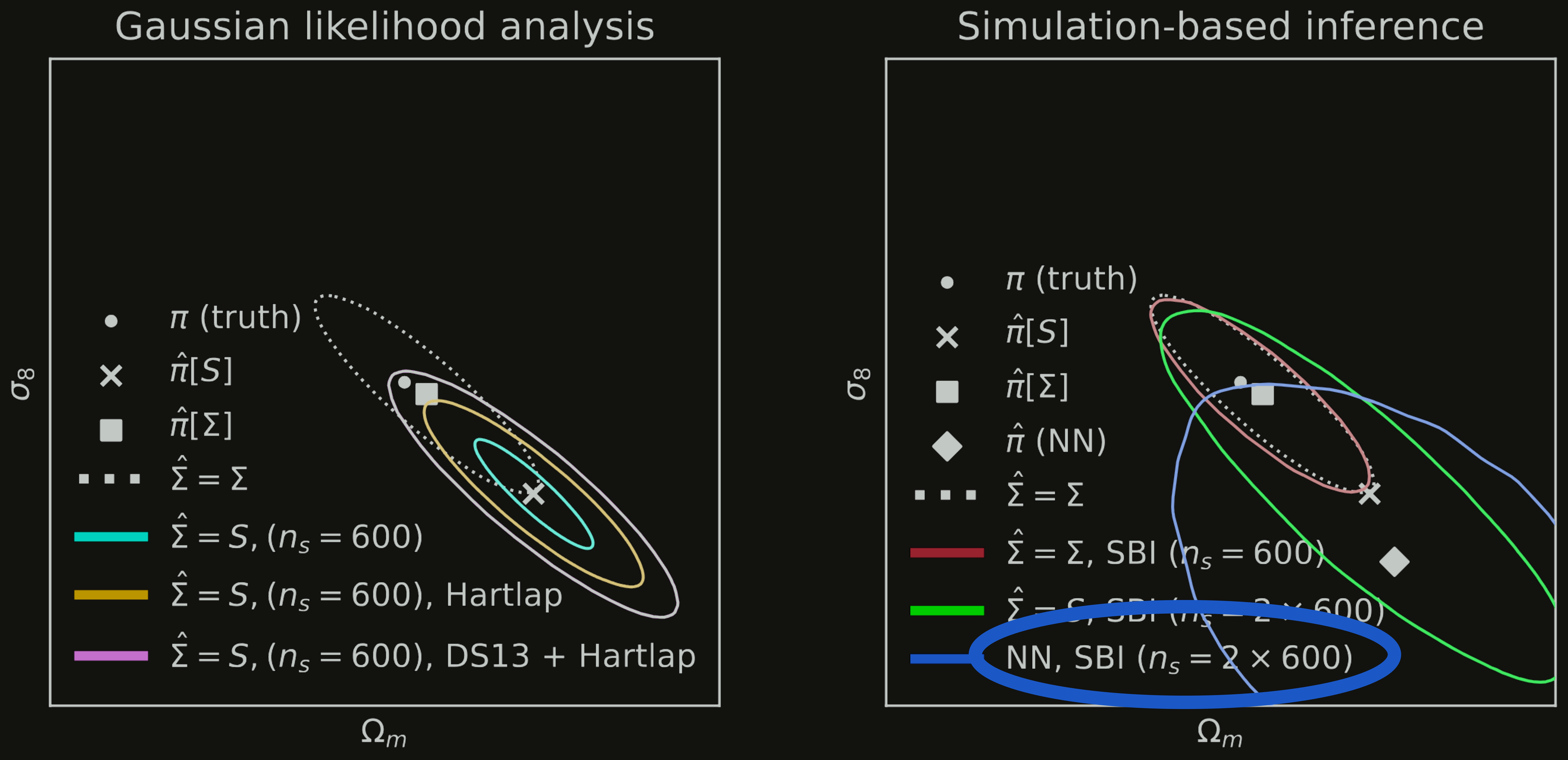
Volodimir Begy
University of Vienna

Gilles Louppe
University of Liège



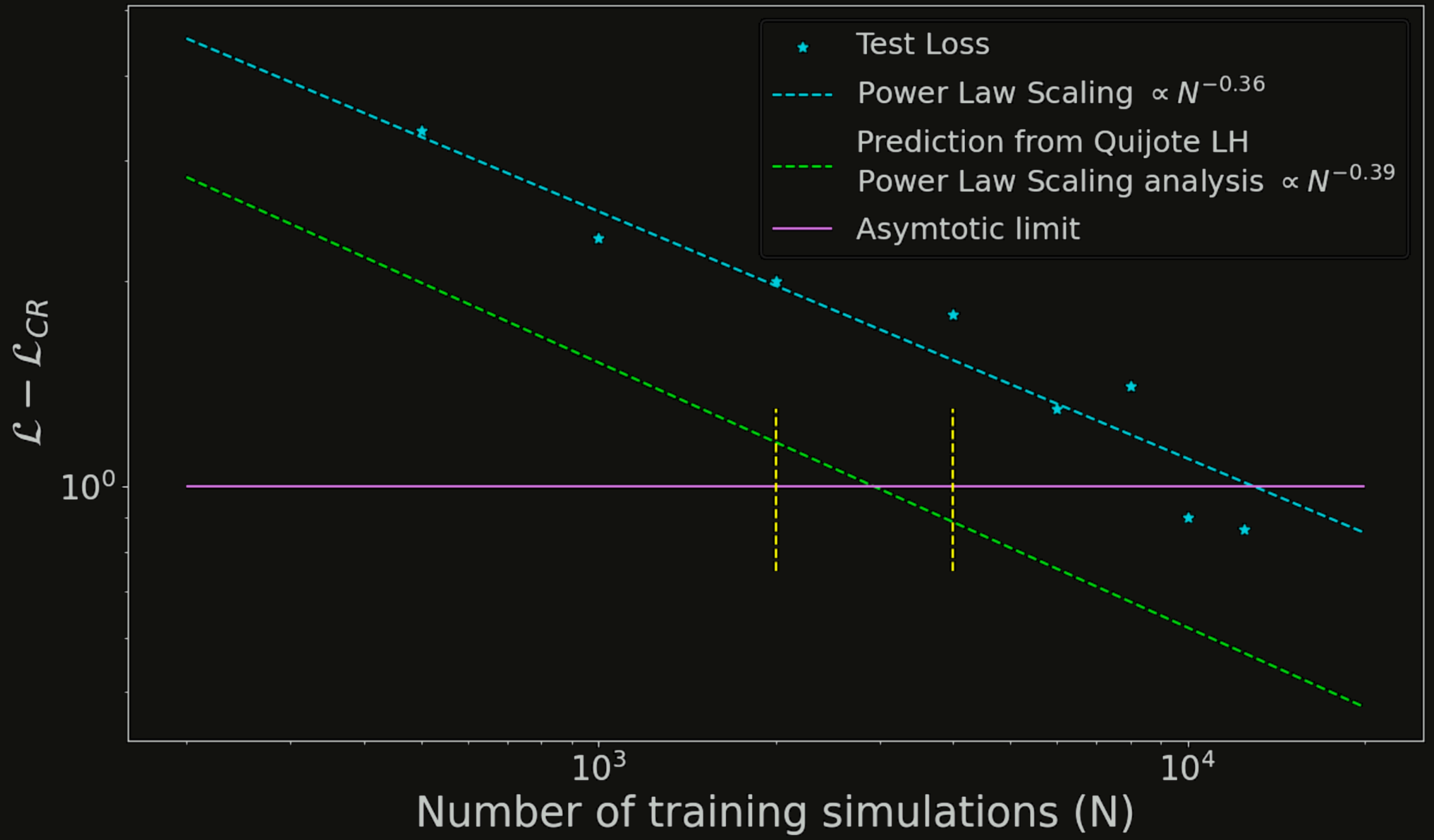
Want to be along diagonal

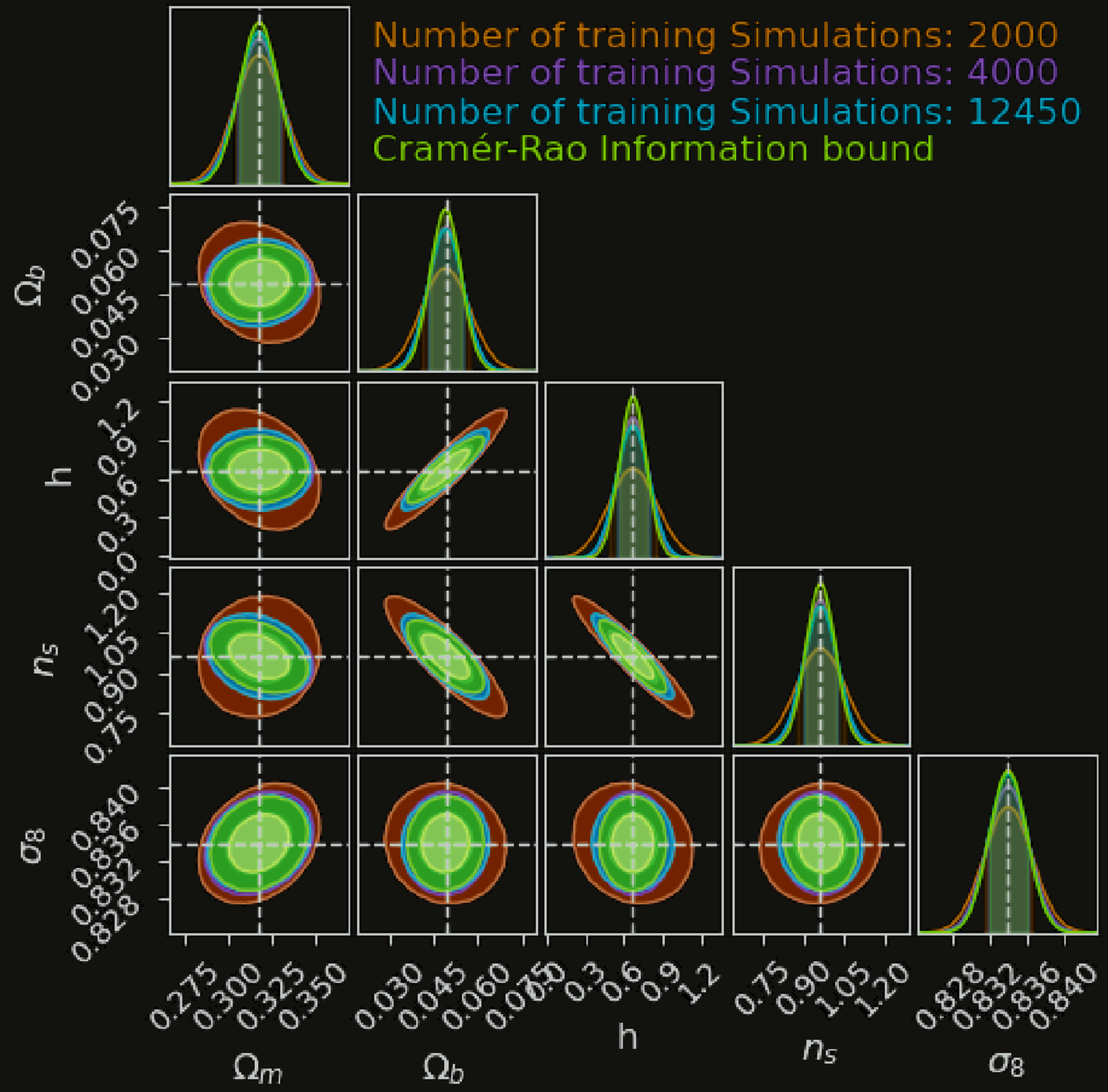
Implicit Likelihood Inference in Crisis?



[Homer, Friedrich, Gruen 2024]

Implicit Likelihood Inference in Crisis?





[Bairagi, Wandelt, Villaescusa-Navarro 2025]

So far, more or less “toy examples”.

Not the same level of trust as traditional analyses.

So far, more or less “toy examples”.

Not the same level of trust as traditional analyses. **For good reason!**

Stage-IV deluge of data will make the problem much more challenging…

Now is the time to become clever:

— run simulations for cheaper

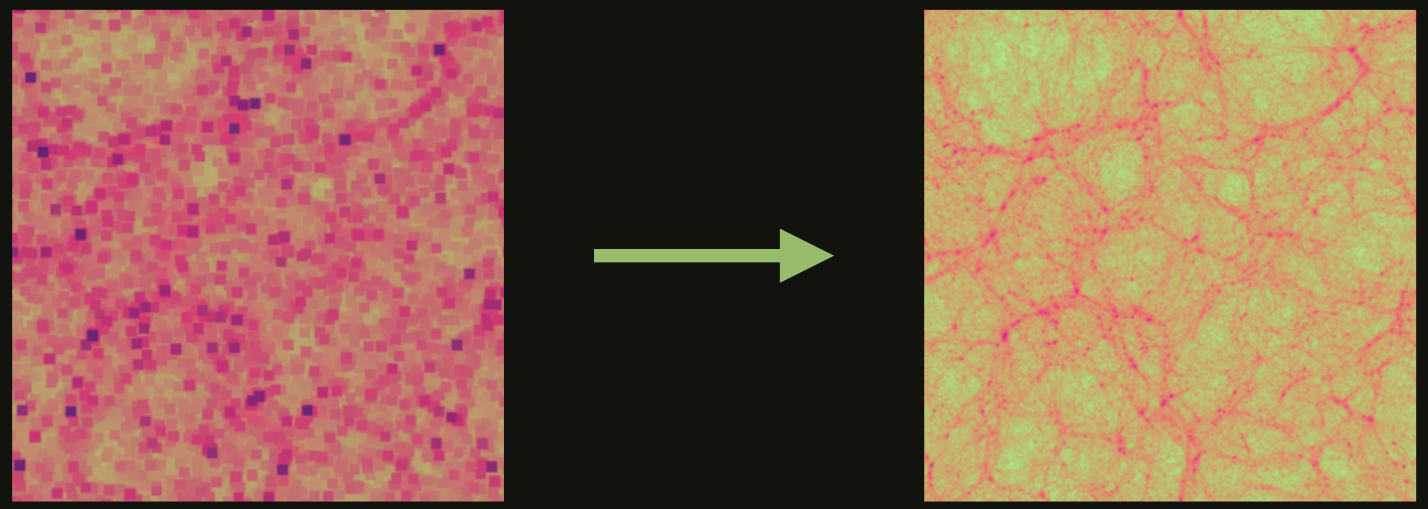
— run them where it counts

— combine simulations of different qualities

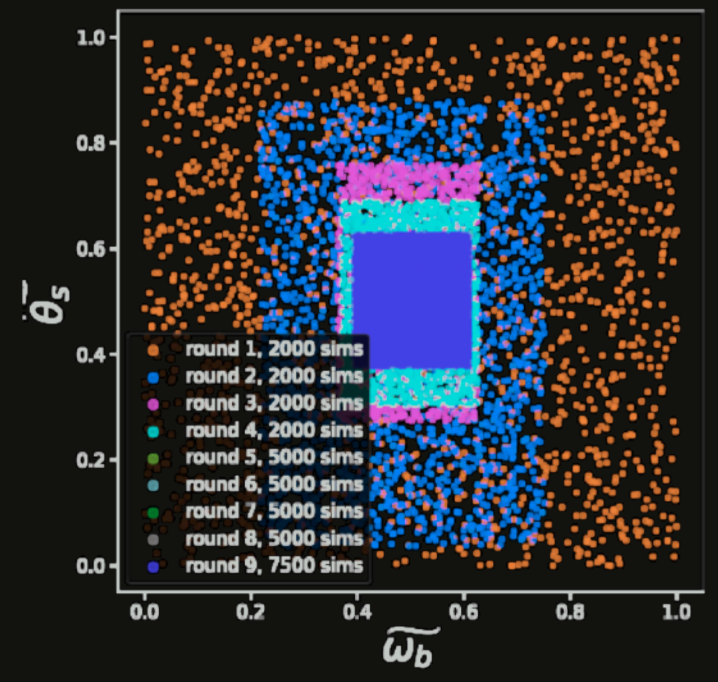
Some promising approaches developed already!

Scaling up

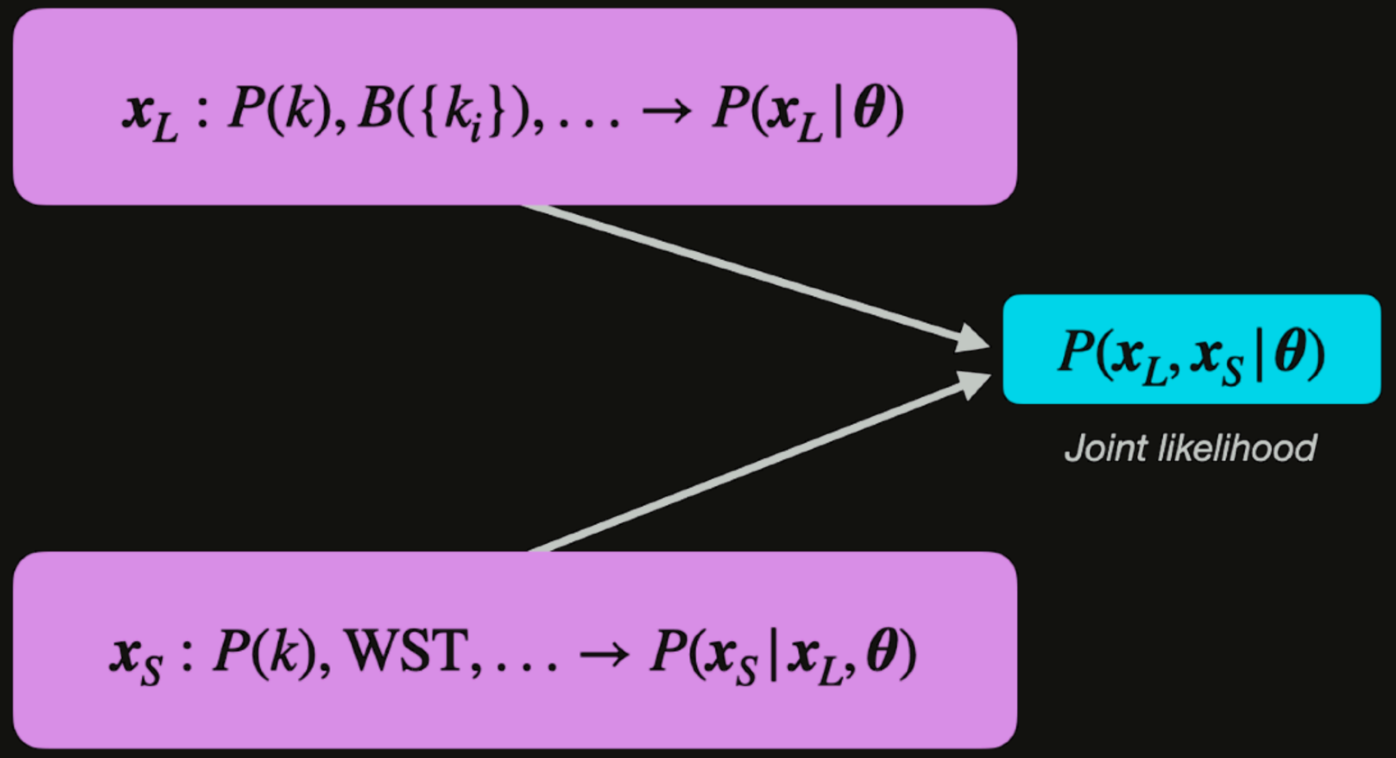
- (ML) accelerated simulations [e.g., Jamieson]



- “painting” into simulations [gas, galaxies, …]
- sequential inference [e.g., Cole]

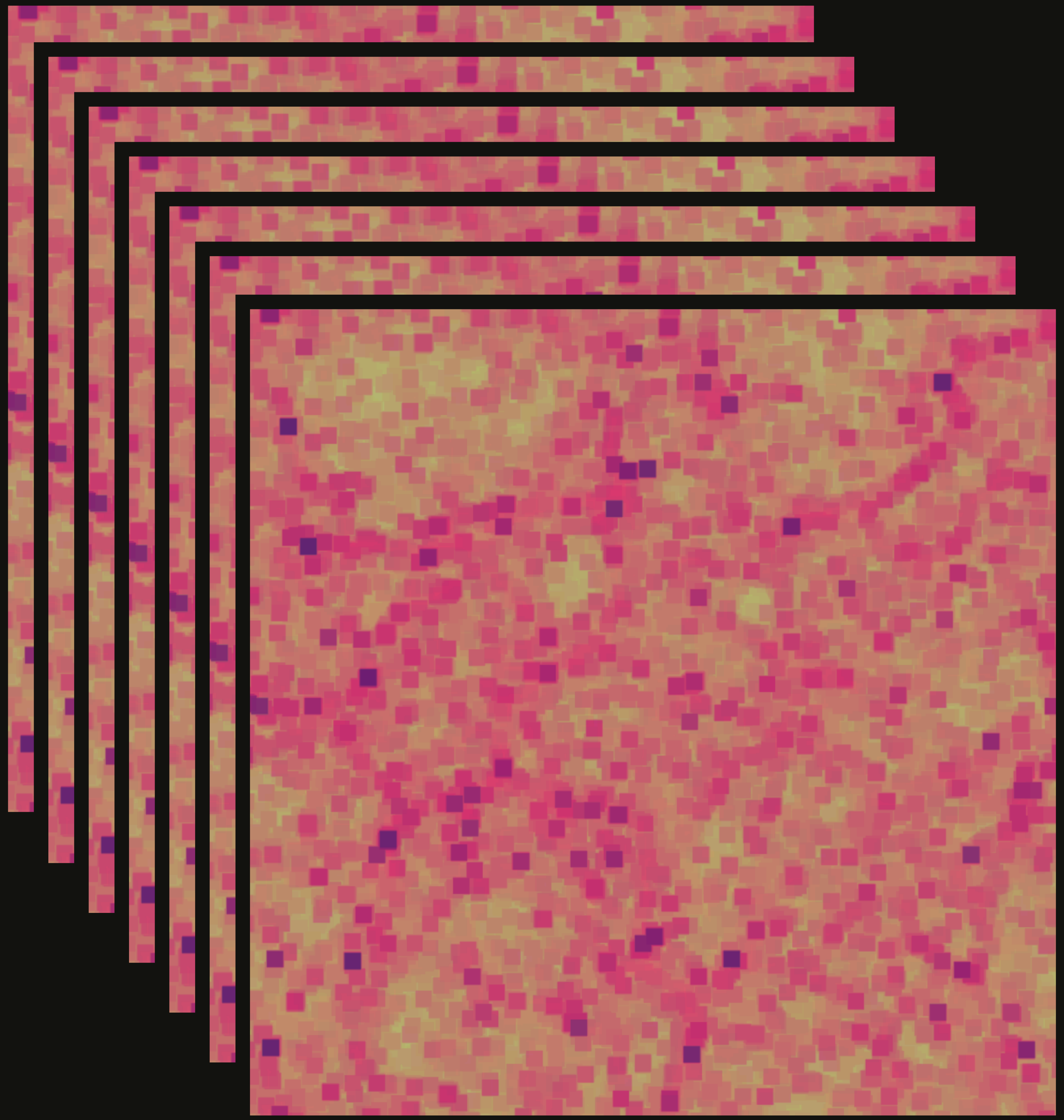


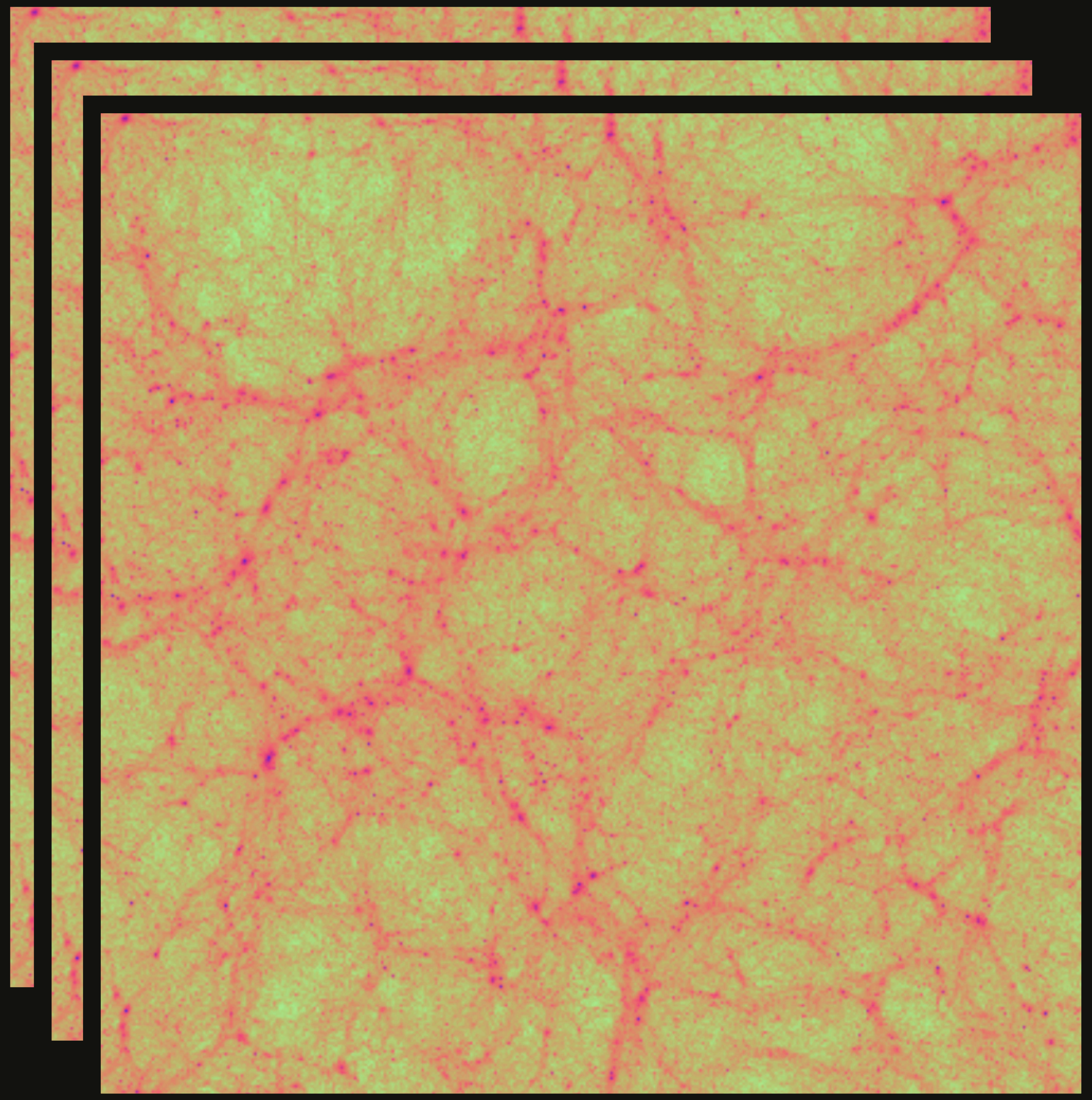
- hybrid analytic & SBI [Modi&Philcox]



- *multi-fidelity*

Multi-fidelity





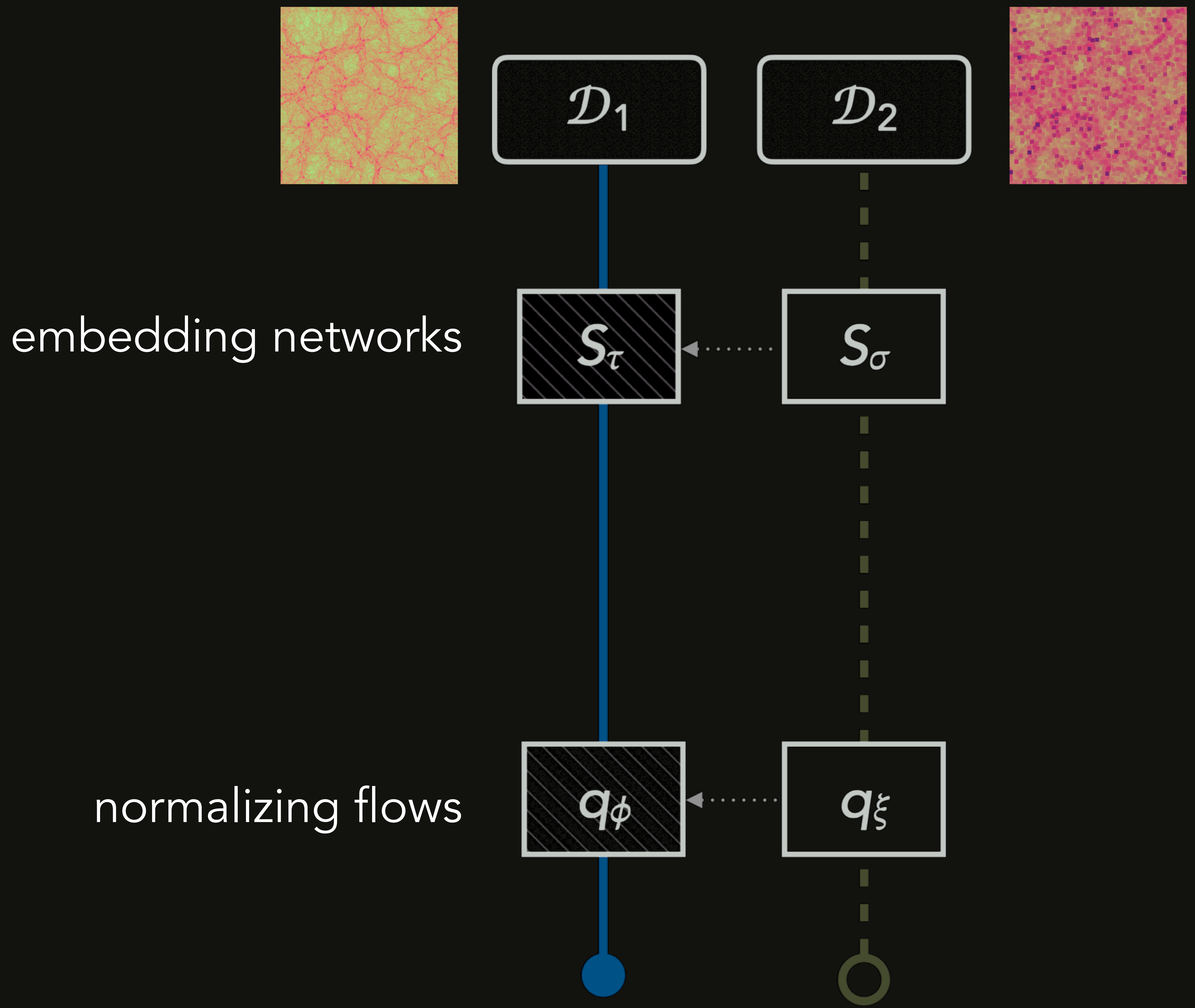
tiny low-fidelity set(s)
e.g., hydrodynamic

large low-fidelity set(s)
e.g., linear theory + 2LPT + particle-mesh + tree + HOD + SAM

[LT, A.Bayer, N.Takeishi 2025]

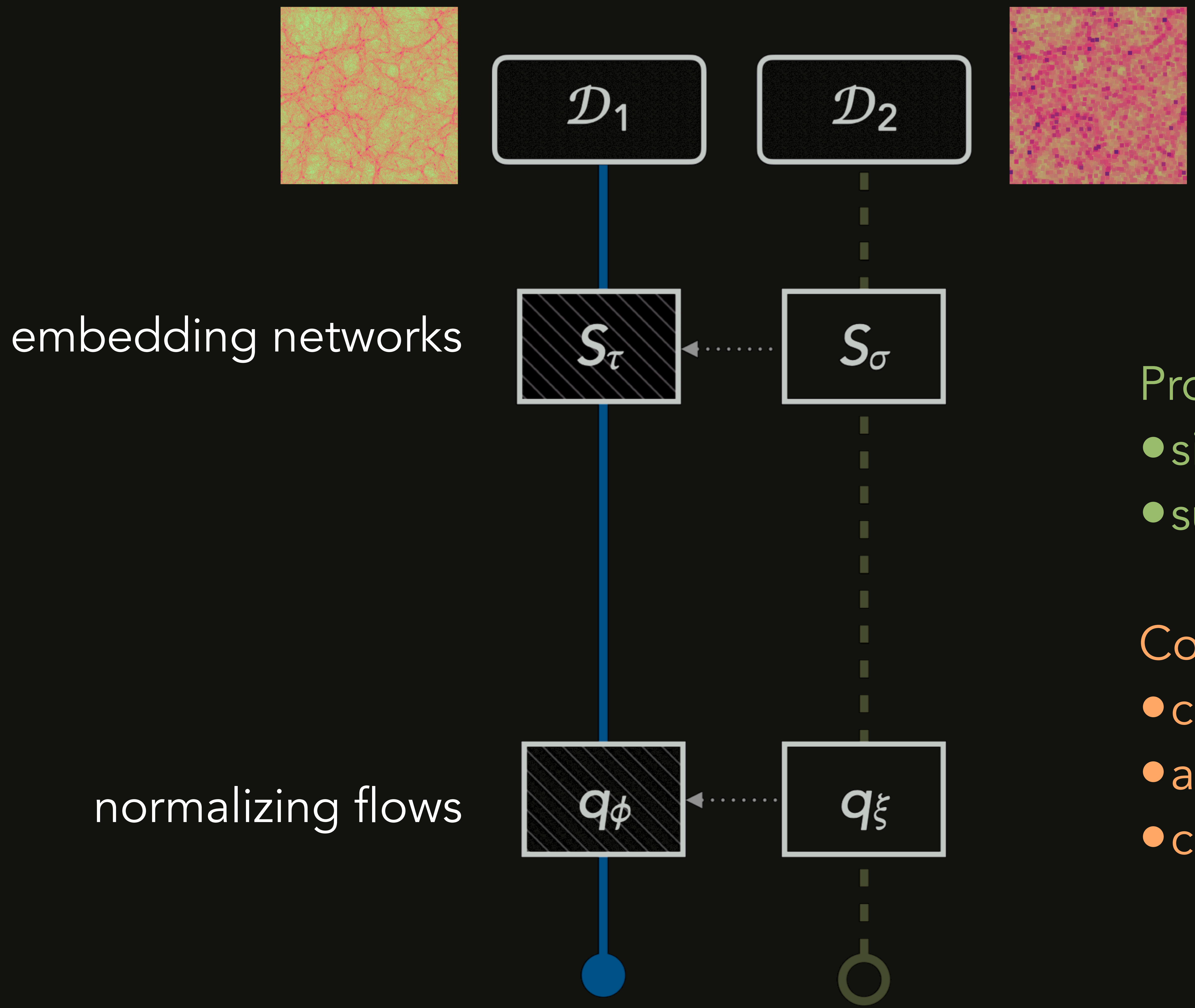
Multi-fidelity

simplest idea: transfer learning via weight initialization



Multi-fidelity

simplest idea: transfer learning via weight initialization



Pros:

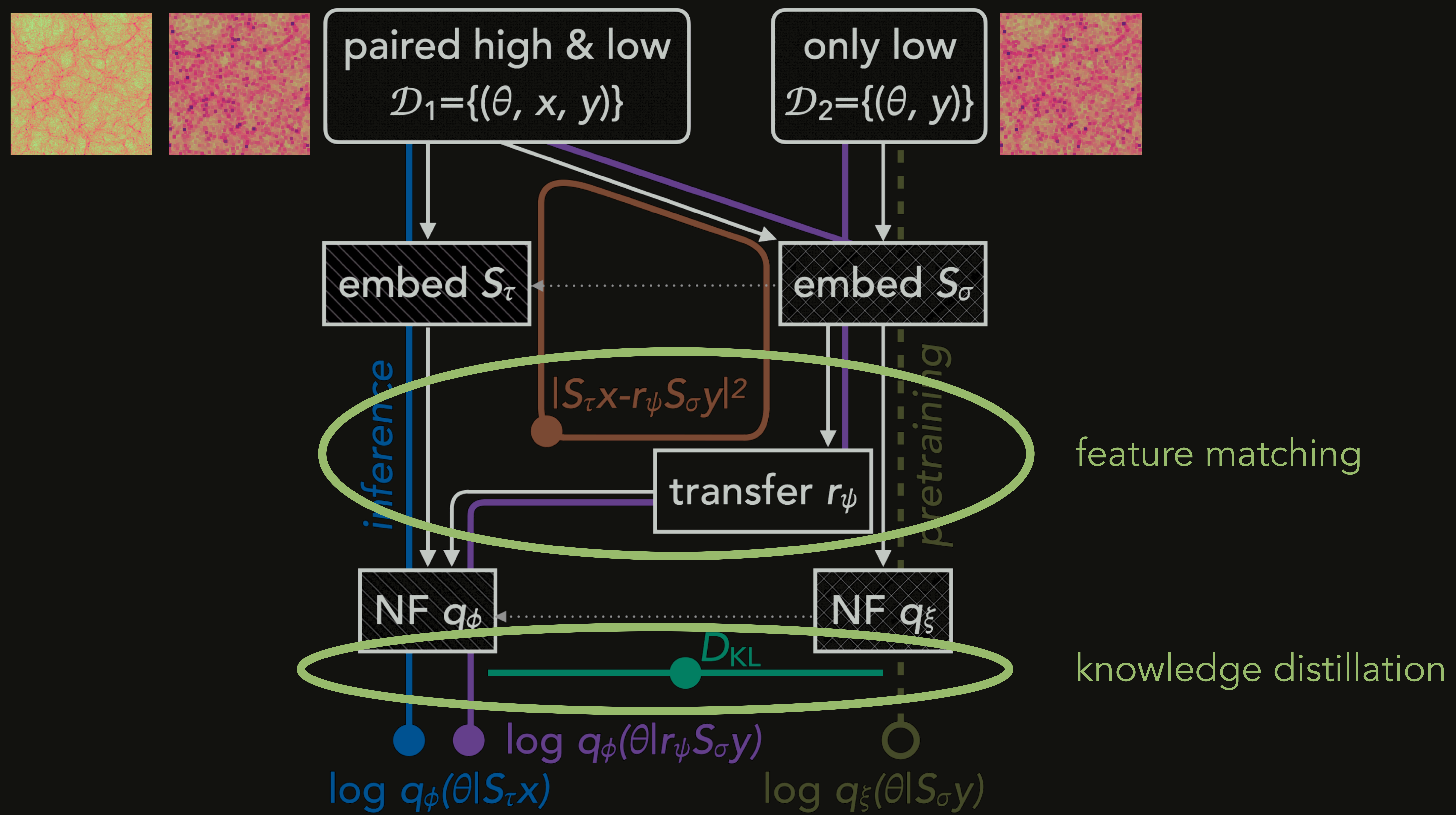
- simple
- surprisingly effective

Cons:

- clearly suboptimal
- architecture constrained
- can be unstable

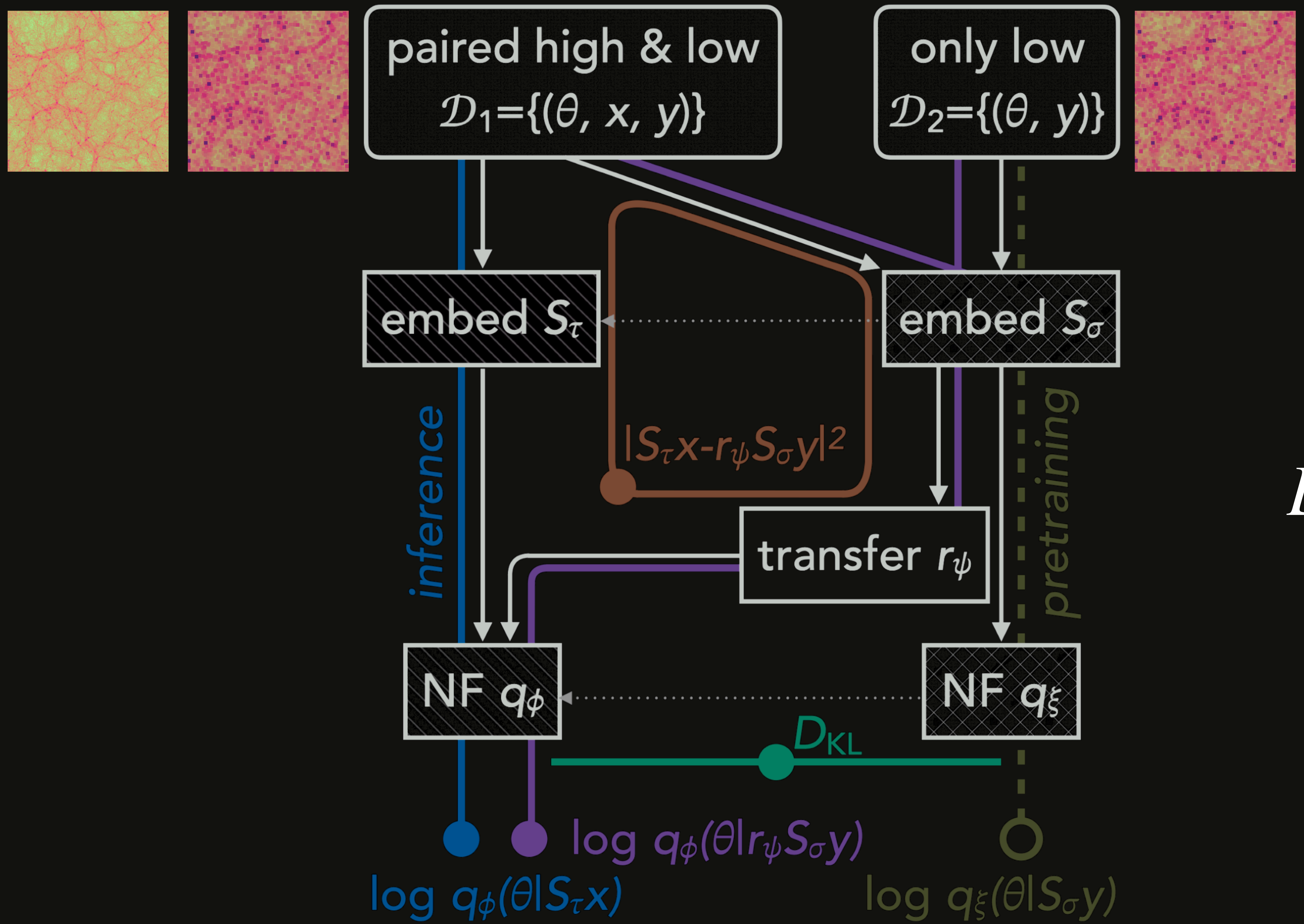
Multi-fidelity

improvement: feature matching & knowledge distillation



Multi-fidelity

improvement: feature matching & knowledge distillation



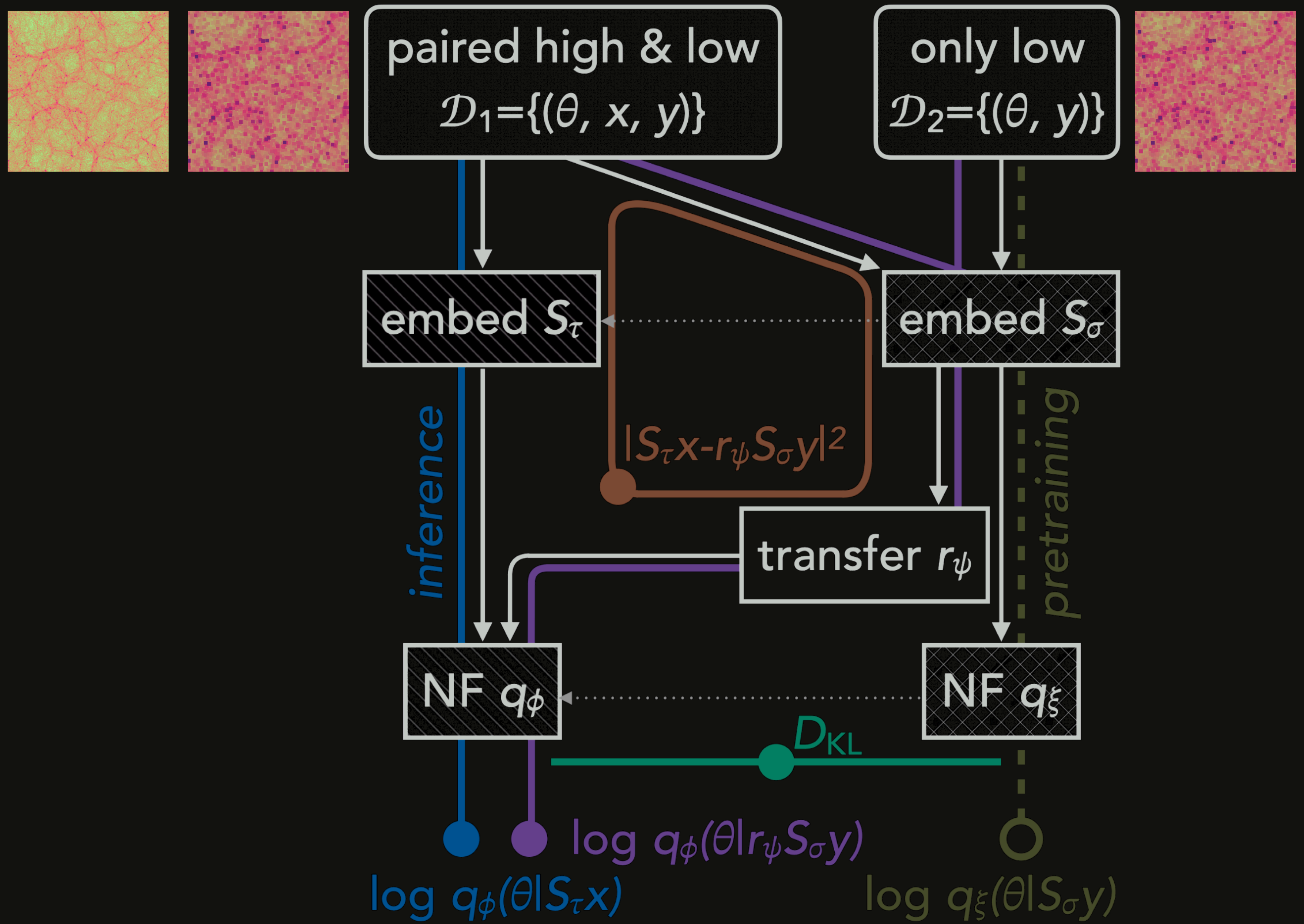
typical variance reduction at cost of introducing a bias

$$L = \int_{p(x,\theta)} \mathscr{L}$$

$$\approx \sum_{\text{training set}} \mathscr{L} + \sum \lambda_a R_a$$

Multi-fidelity

improvement: feature matching & knowledge distillation



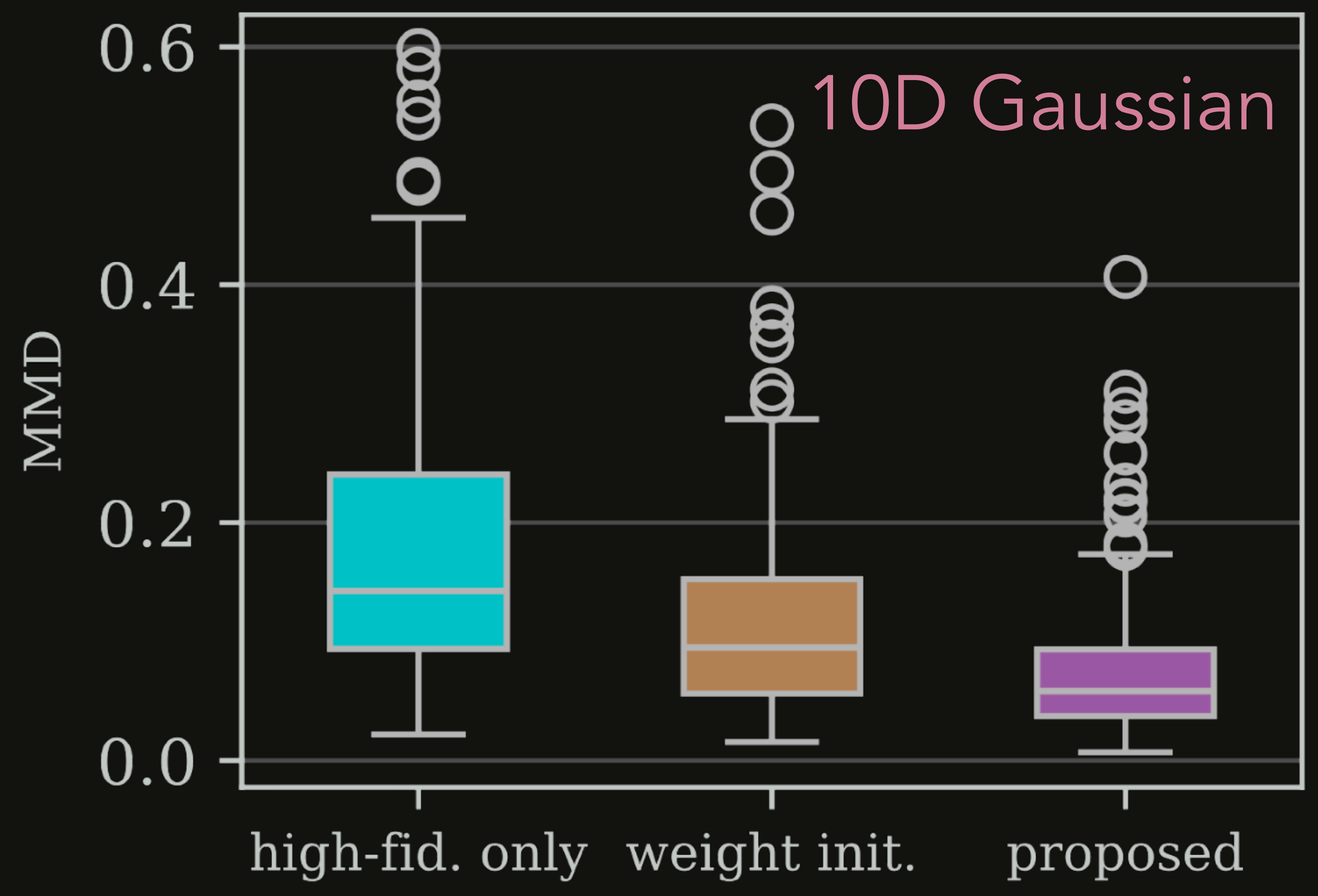
Pros:
- very flexible (structure, levels)
- "soft": often better

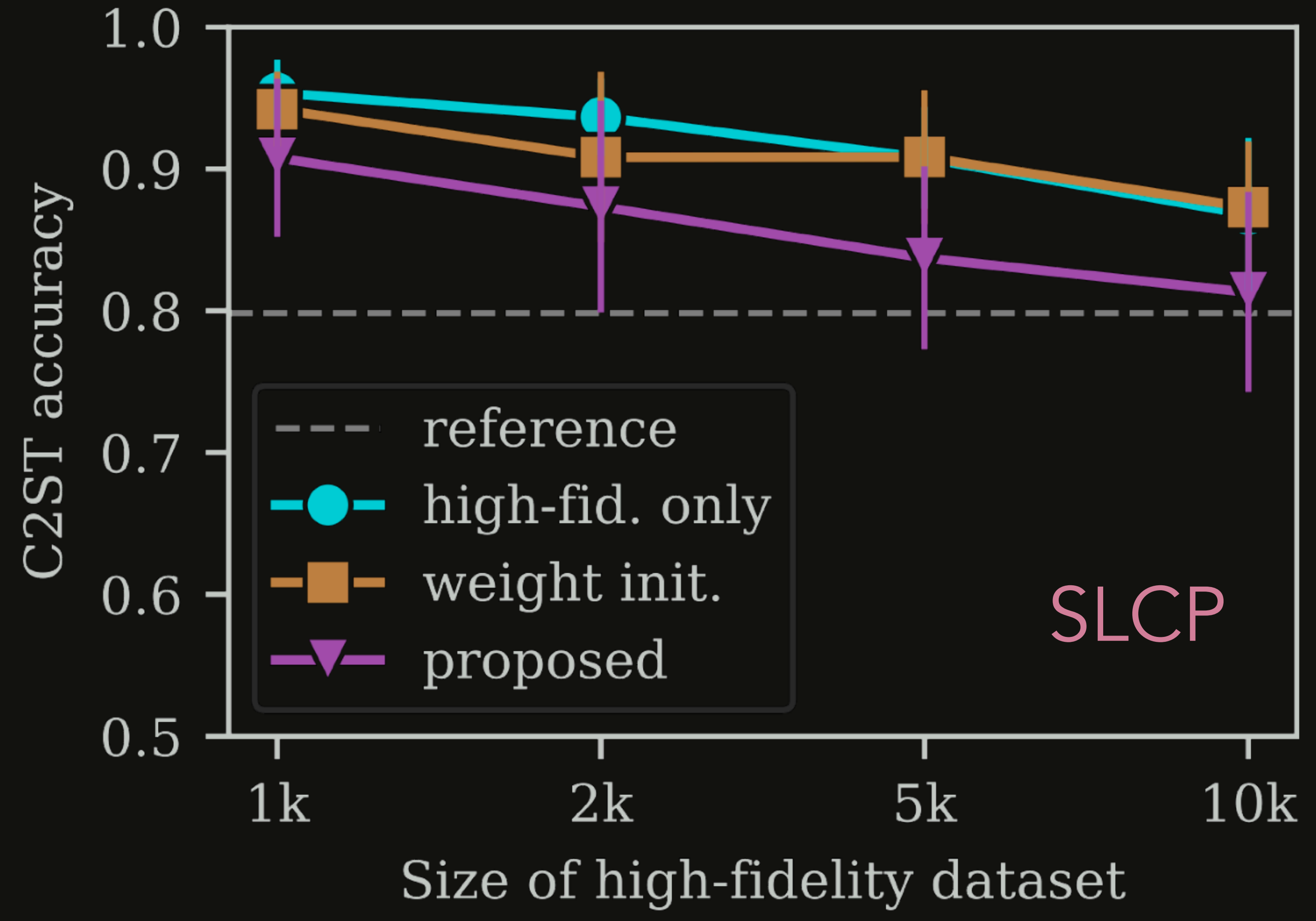
Cons:
- multiple hyperparameters
- heuristics

→ improvement work in progress

Multi-fidelity: results

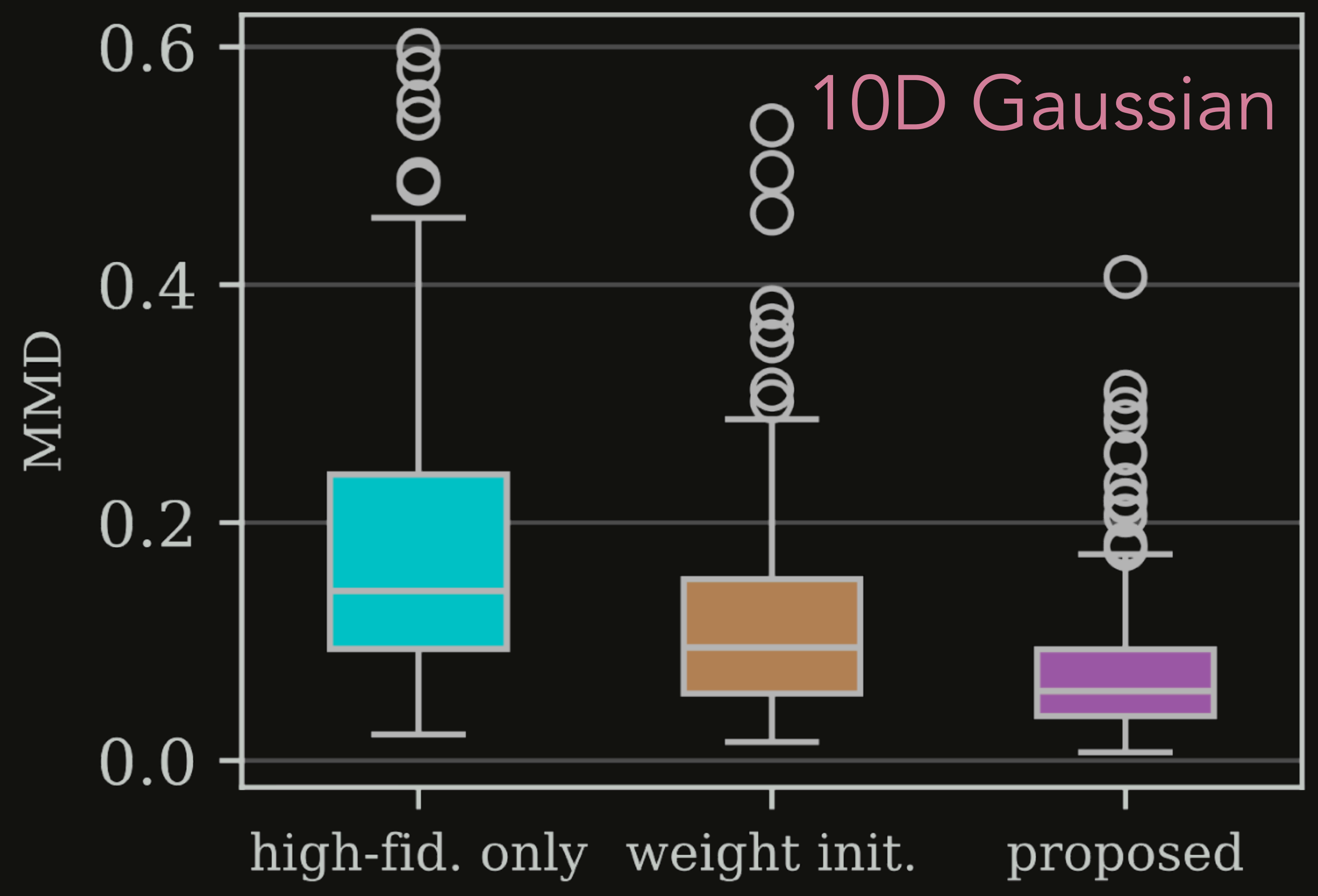
synthetic examples
(lower is better for all)

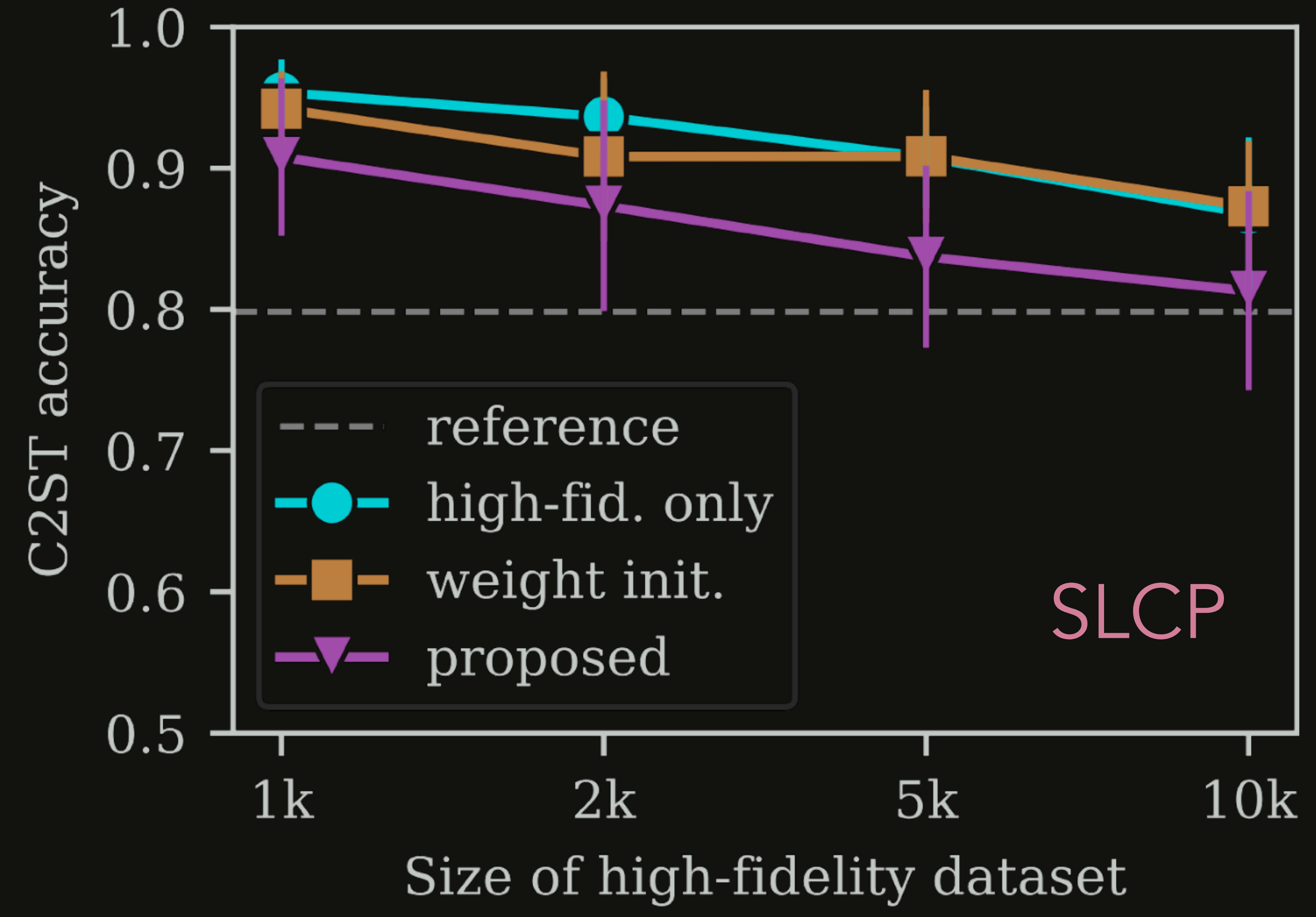




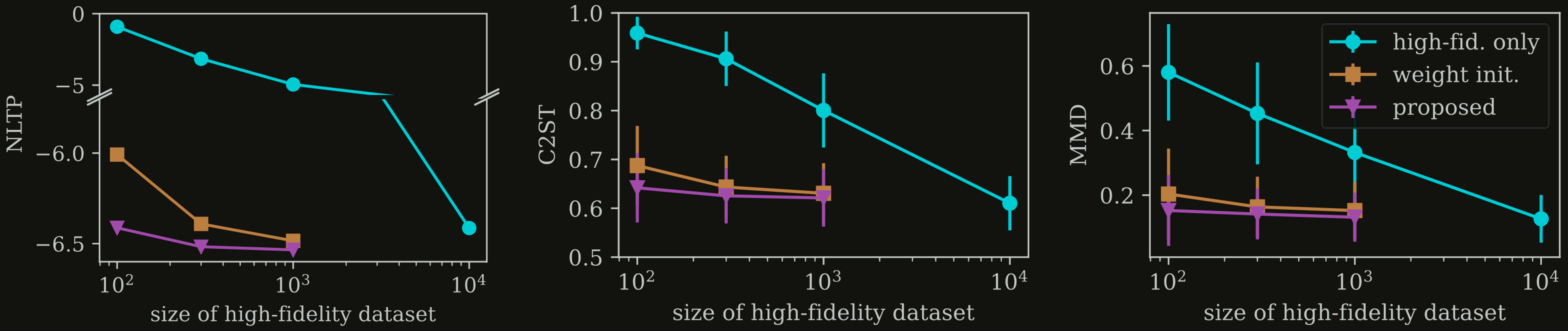
Multi-fidelity: results

synthetic examples
(lower is better for all)





cosmology example (lower is better for all)



Summary

- Simulation-based inference = machine learning method to solve inverse problems defined implicitly through a simulator
- Simulation-based inference has proven viability in simple examples:
  - weak lensing (e.g., HSC Y1)
  - galaxy clustering (e.g., SDSS CMASS)
- In order to make it a standard tool, need to increase simulation quality while reducing training cost
  - sequential methods
  - learning corrections
  - combine with traditional approaches for large scales
  - multi-fidelity training
- Have demonstrated a regularization method towards multi-fidelity training