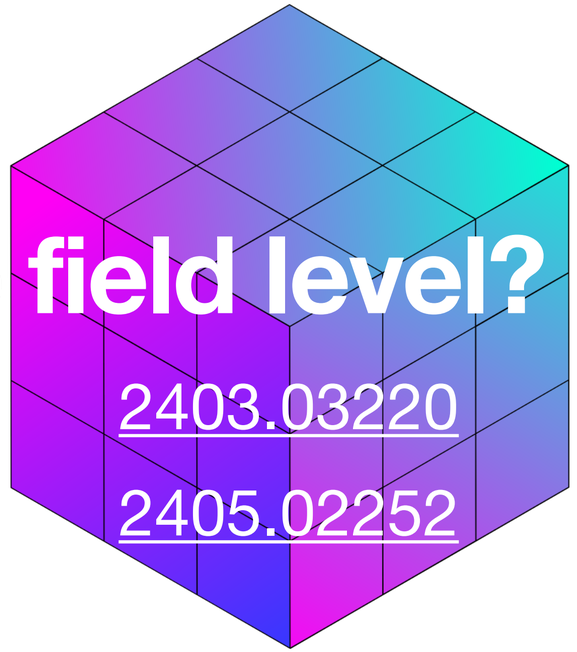
How much information can we extract from galaxy clustering at the field level?



Minh Nguyen (University of Michigan)
nguyenmn@umich.edu

2403.03220

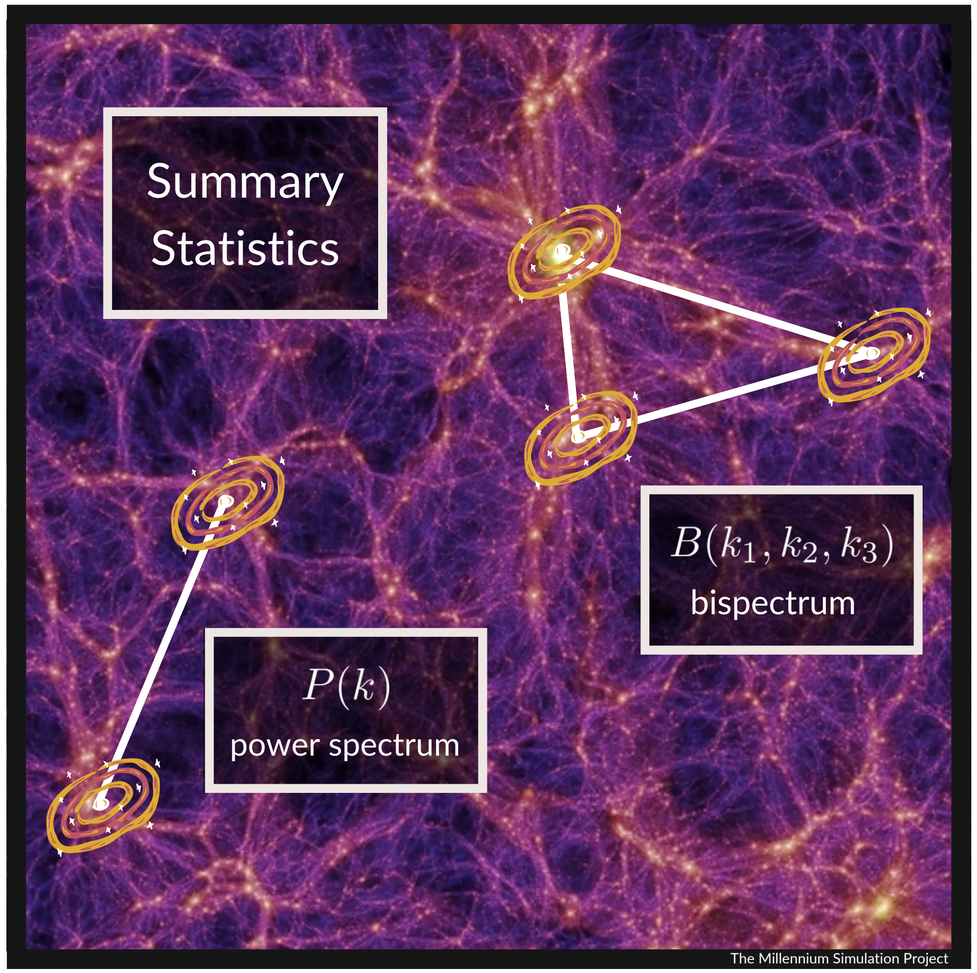
2405.02252

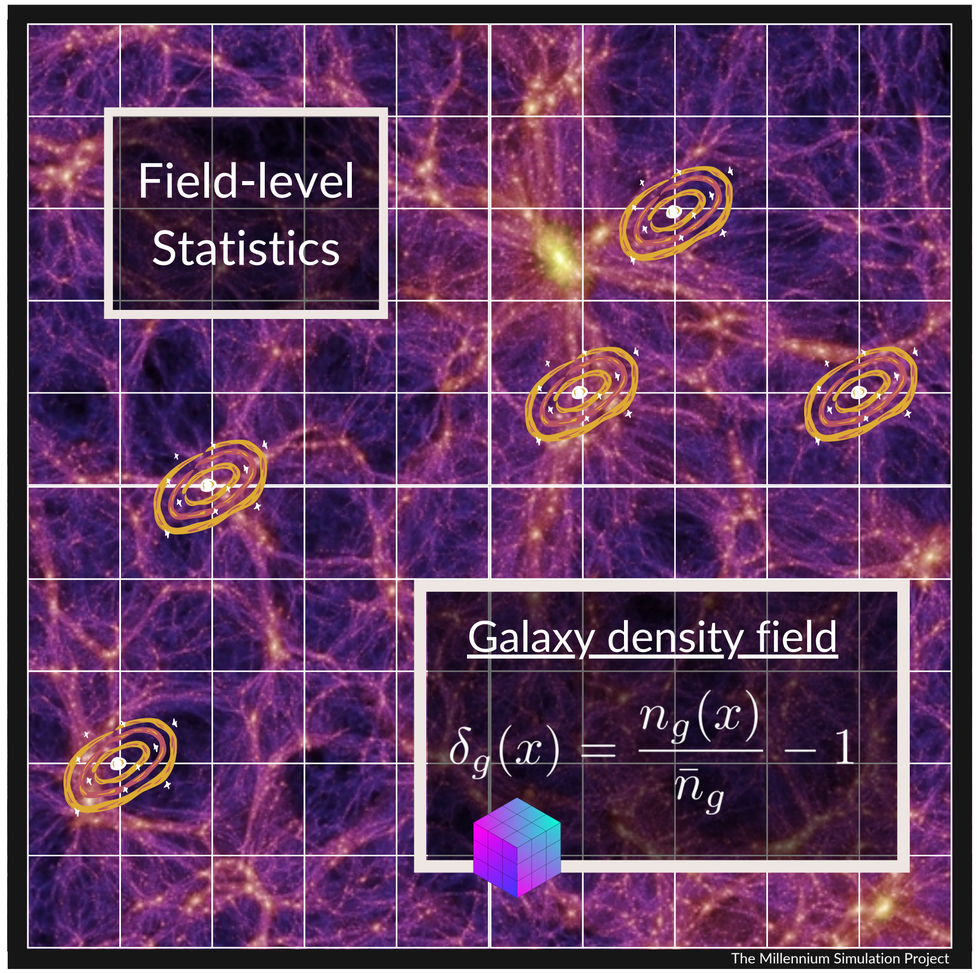
KIPAC Cosmology Seminar
Stanford, May 20 2024

⚡ Summary

Question: How much can we gain?

Summary versus Field-level Statistics of Galaxy Clustering



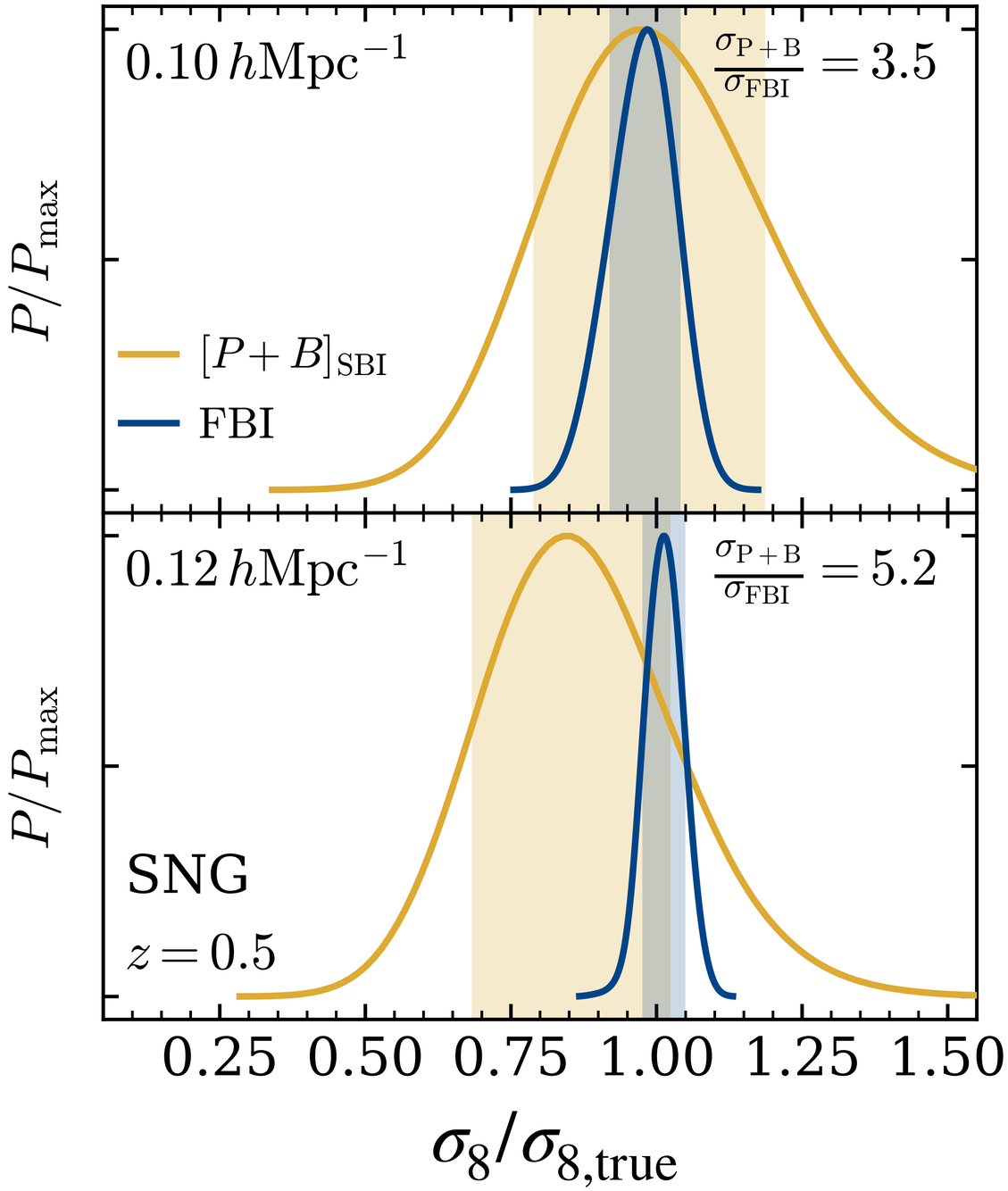


Answer: A lot more!

Constraints from N-body halos

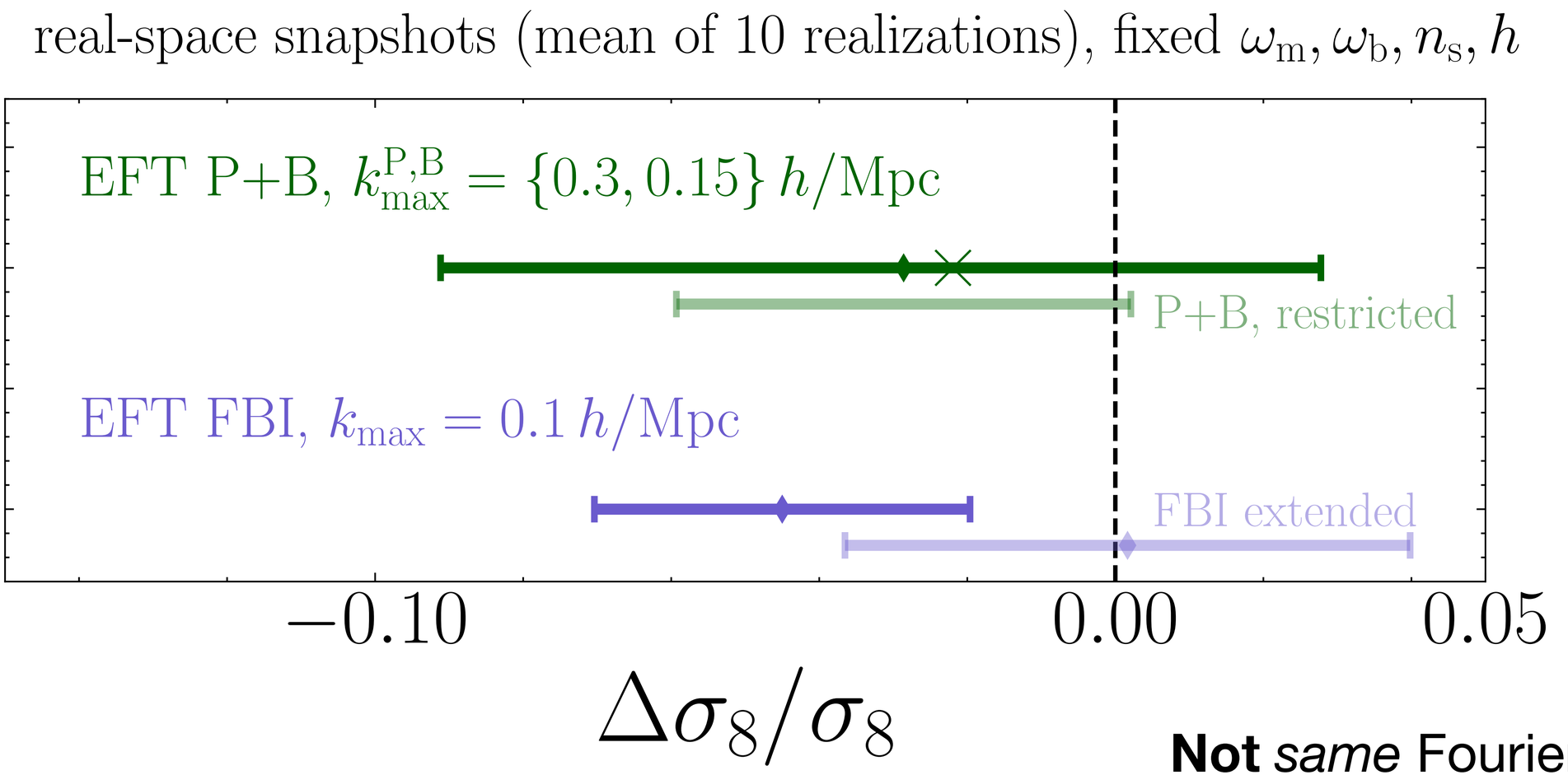
Exact *same* Fourier modes

Pure gain from *nonlinear* clustering (no RSD)

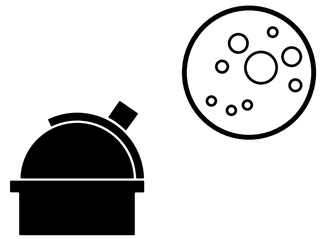


Answer: A lot more, really!

Constraints from HOD galaxies

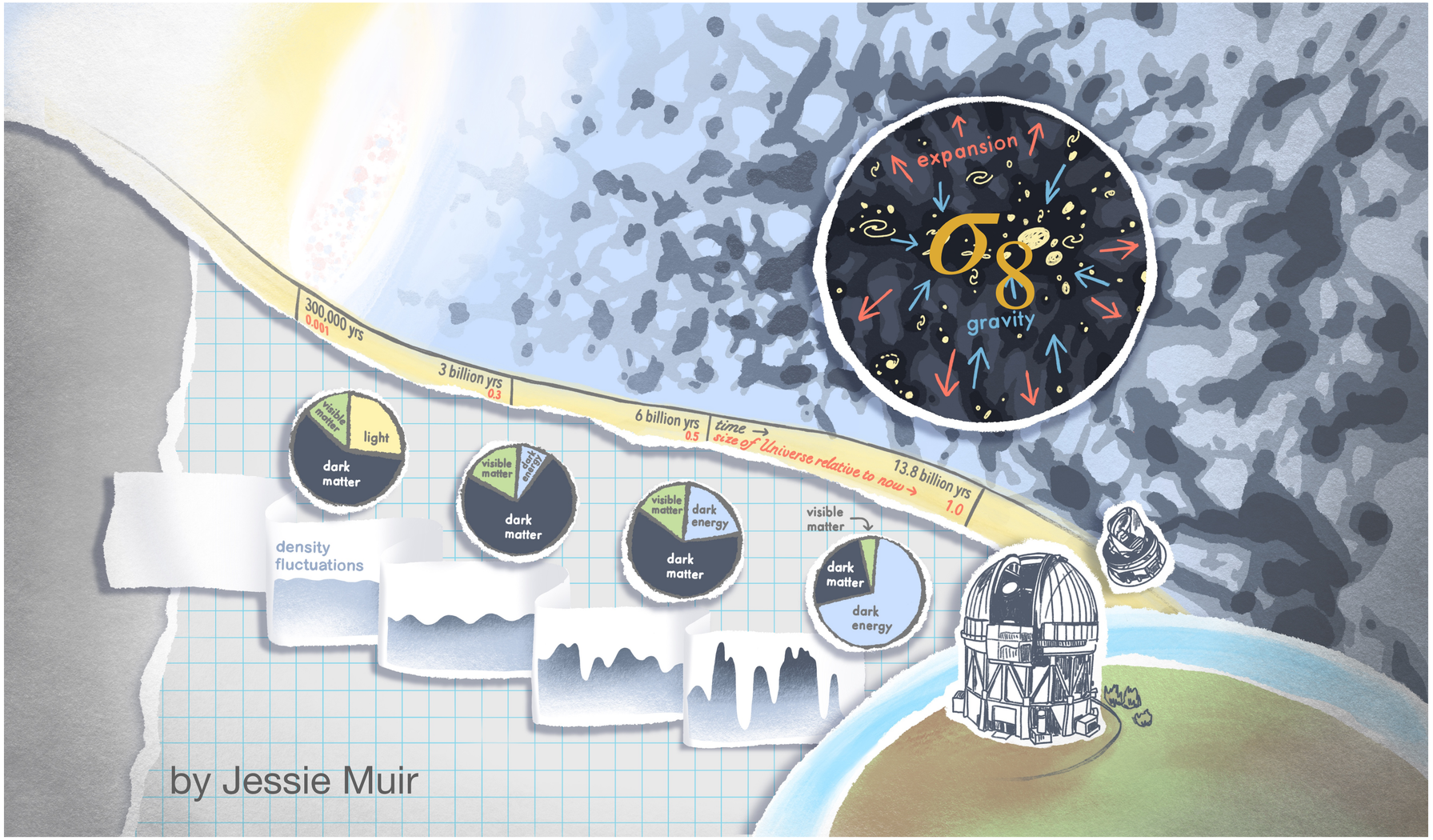


Not *same* Fourier modes

Galaxy Clustering 

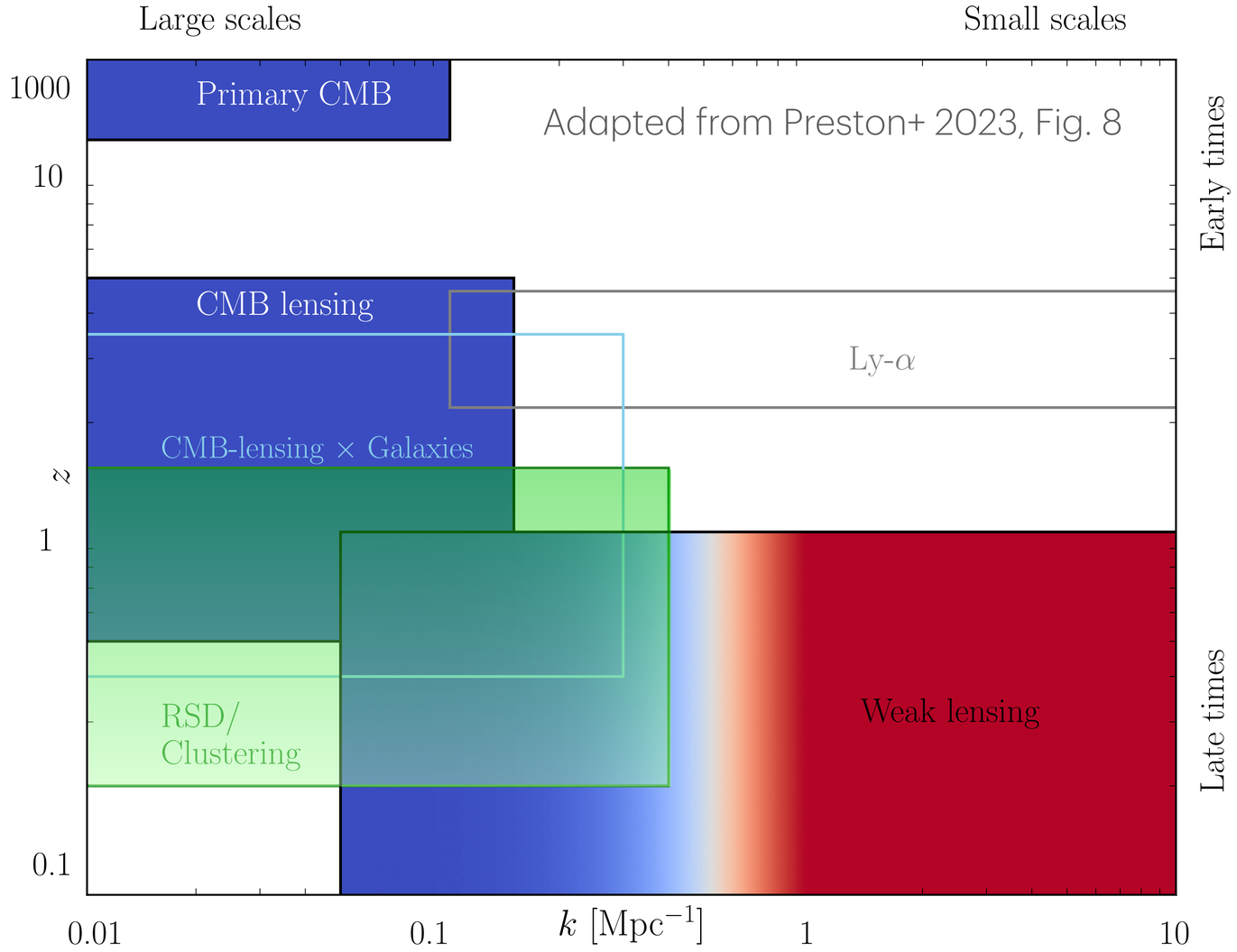
Growth of Structure from Galaxy Clustering

Dark Energy, Gravity and Dark Matter

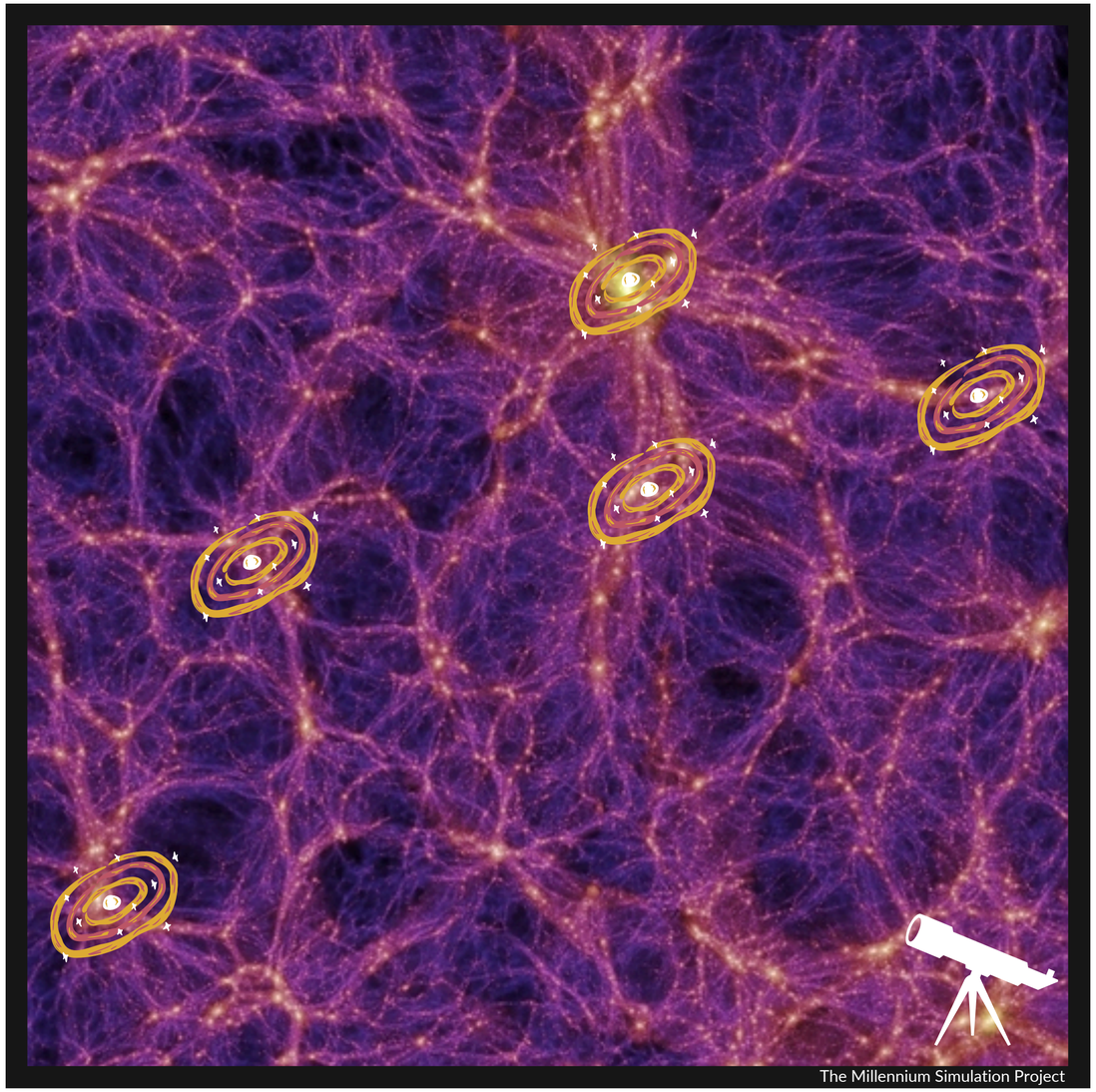


Growth of Structure from Galaxy Clustering

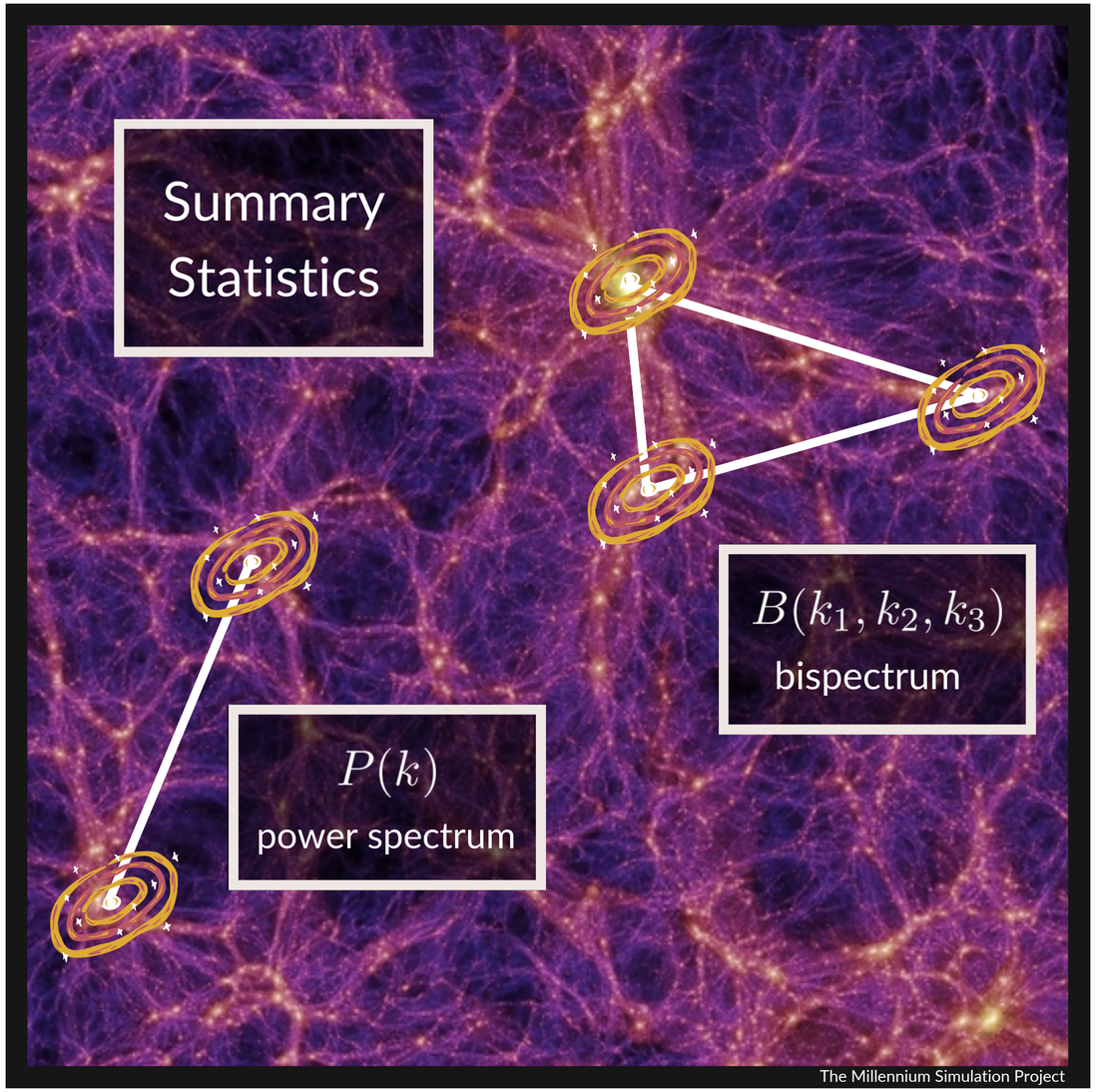
New Physics from late- vs early-time and small- vs large-scale



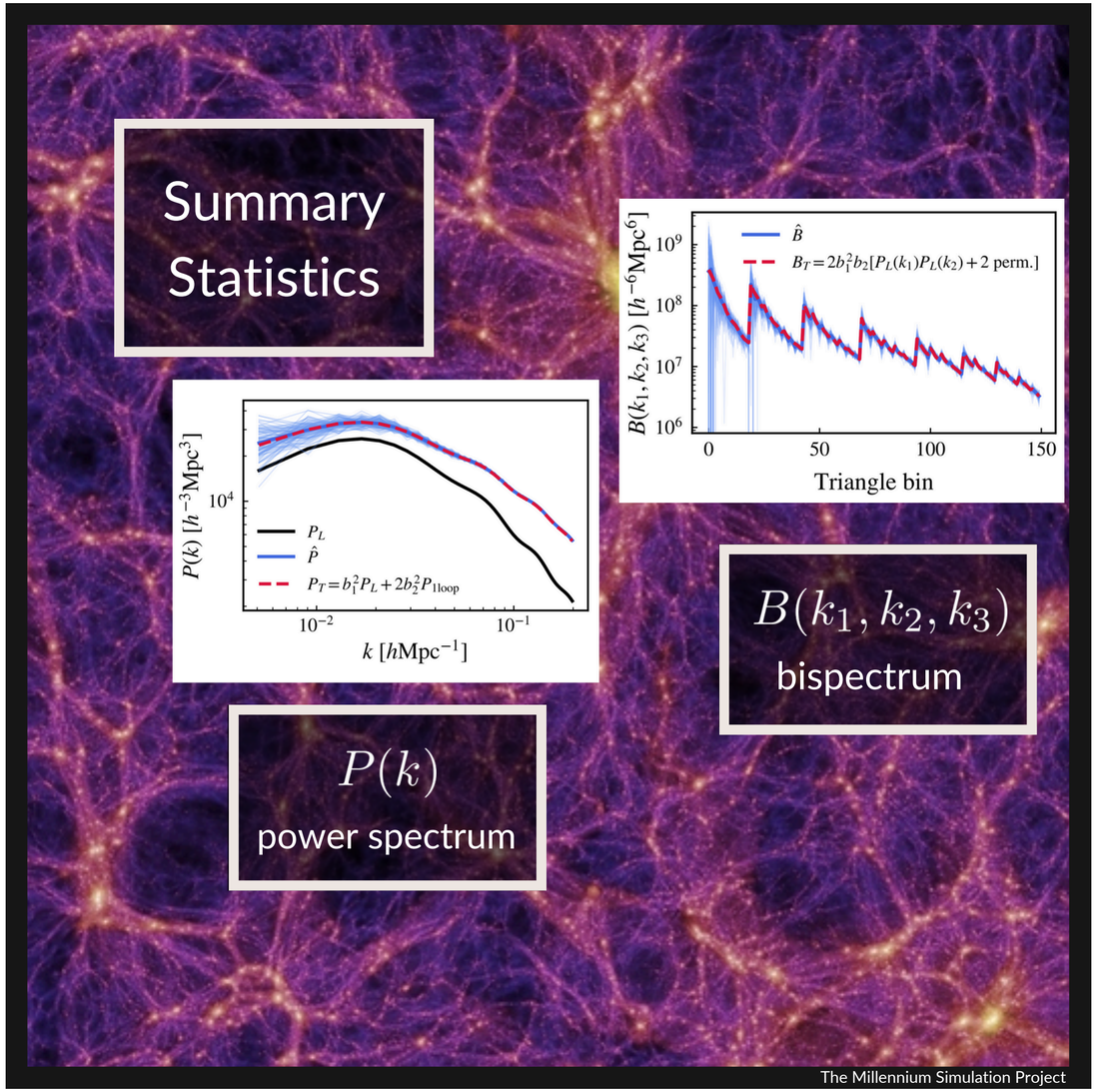
Galaxies are biased tracers of LSS



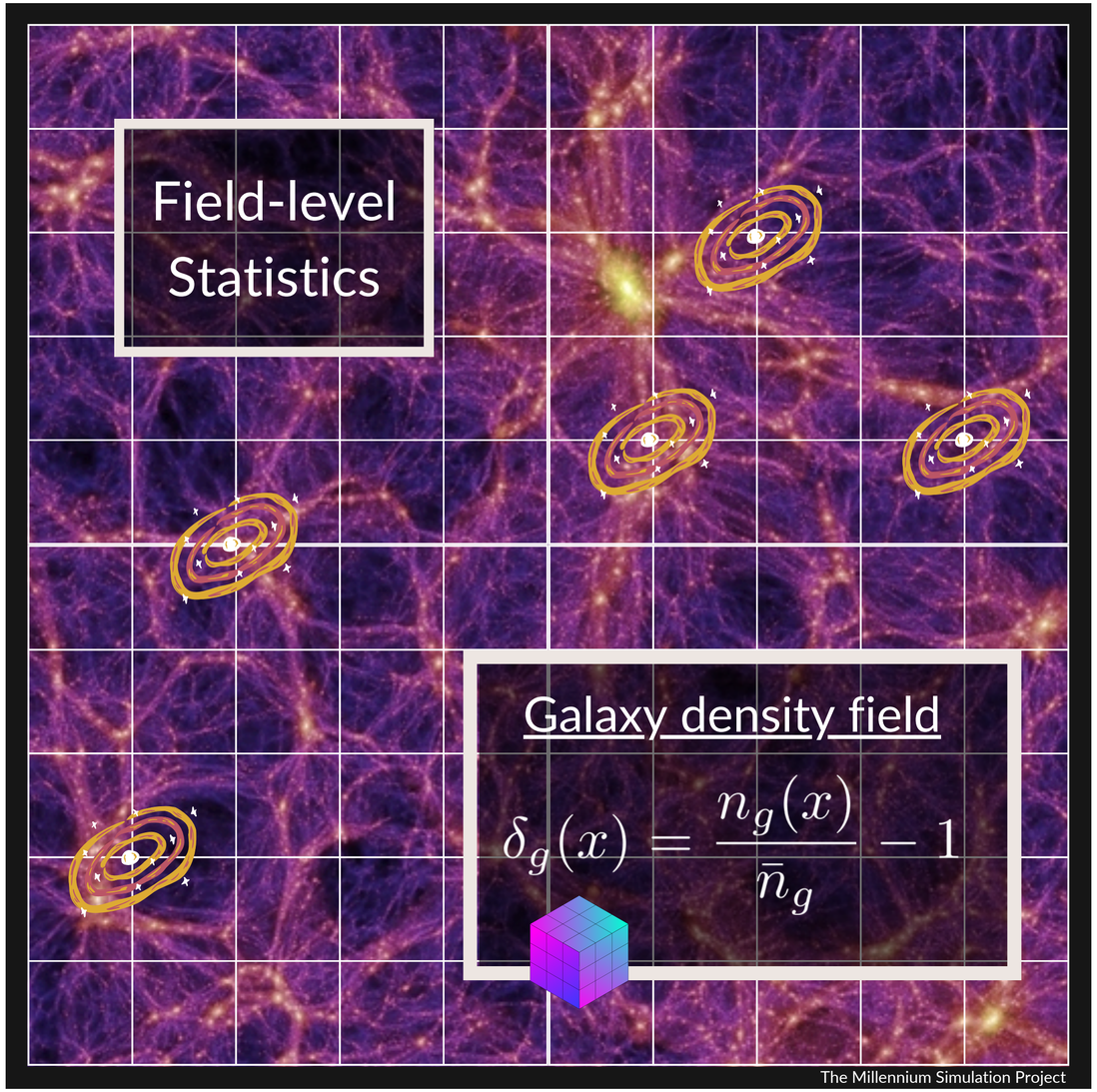
Galaxy Clustering - Summary Statistics



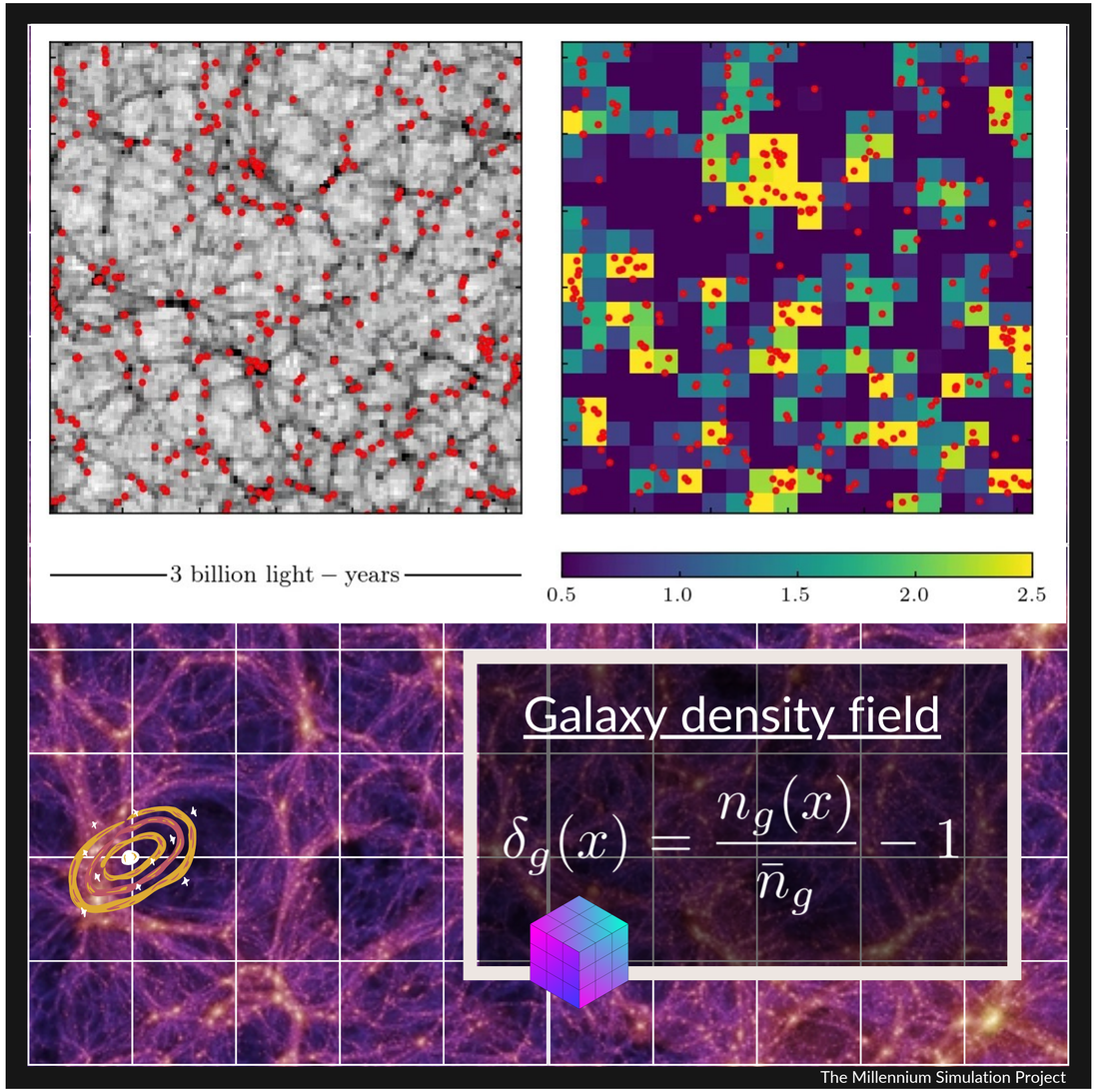
Galaxy Clustering - Summary Statistics

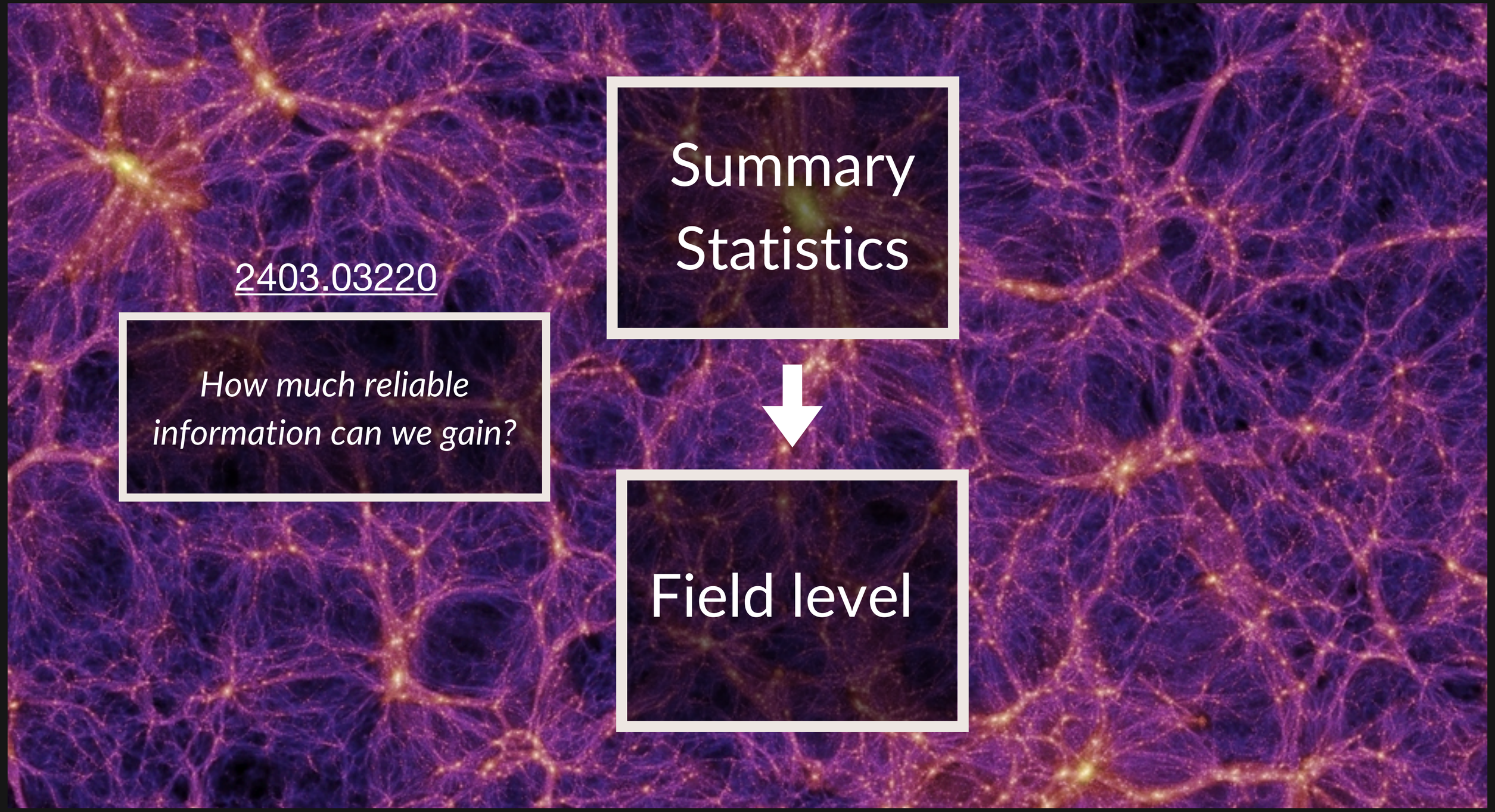


Galaxy Clustering - Field-level statistics



Galaxy Clustering - Field-level statistics



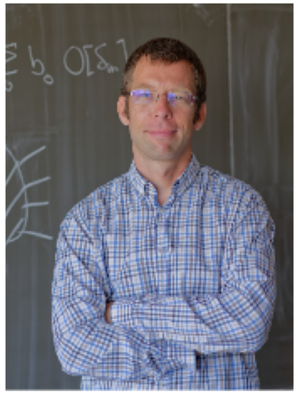

Summary Statistics
2403.03220
How much reliable information can we gain?
Field level
The Millennium Simulation Project

Summary Statistics

2403.03220

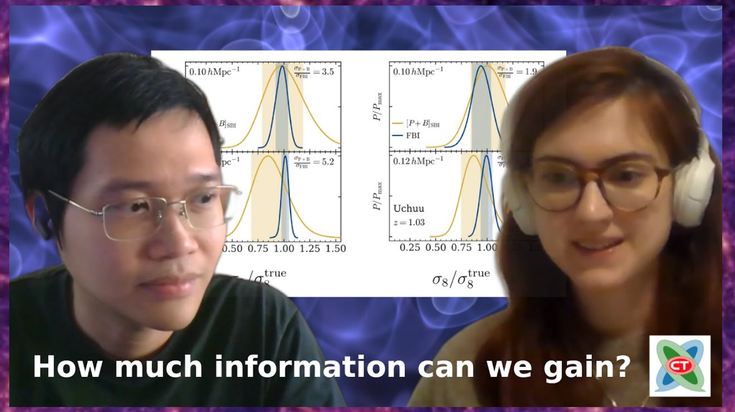
How much reliable information can we gain?

Fabian Schmidt Beatriz Tucci



Field level

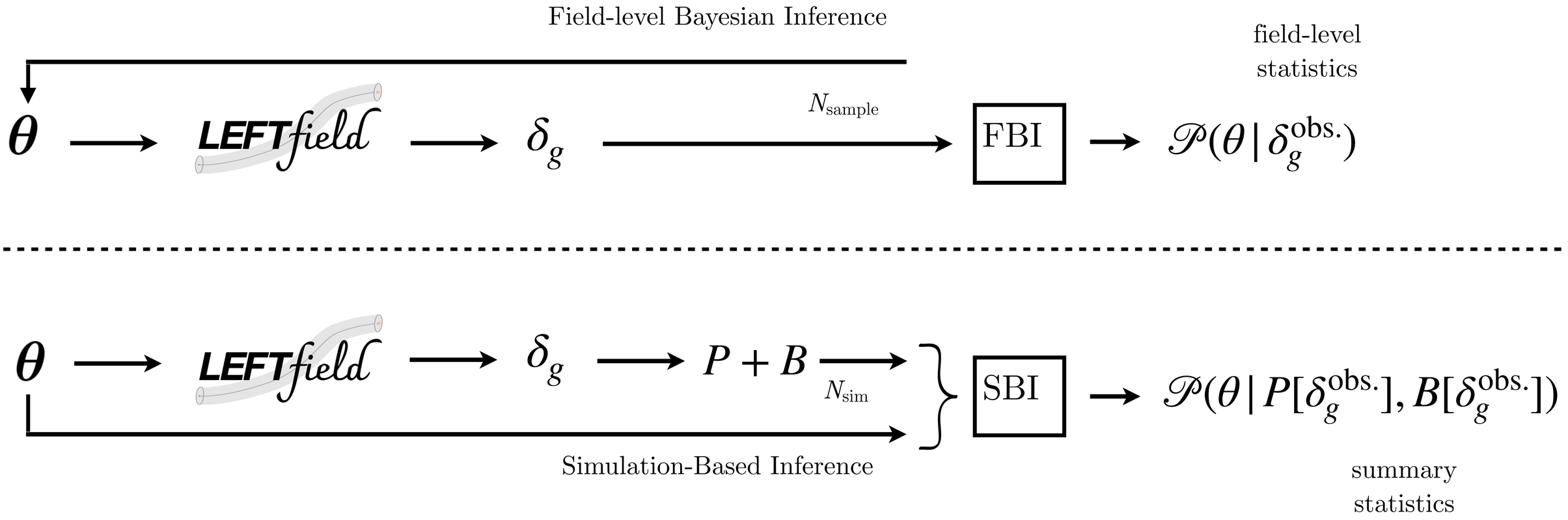
youtu.be/Pxfsf89jZwI



Cosmology Talks

An apple-to-apple comparison

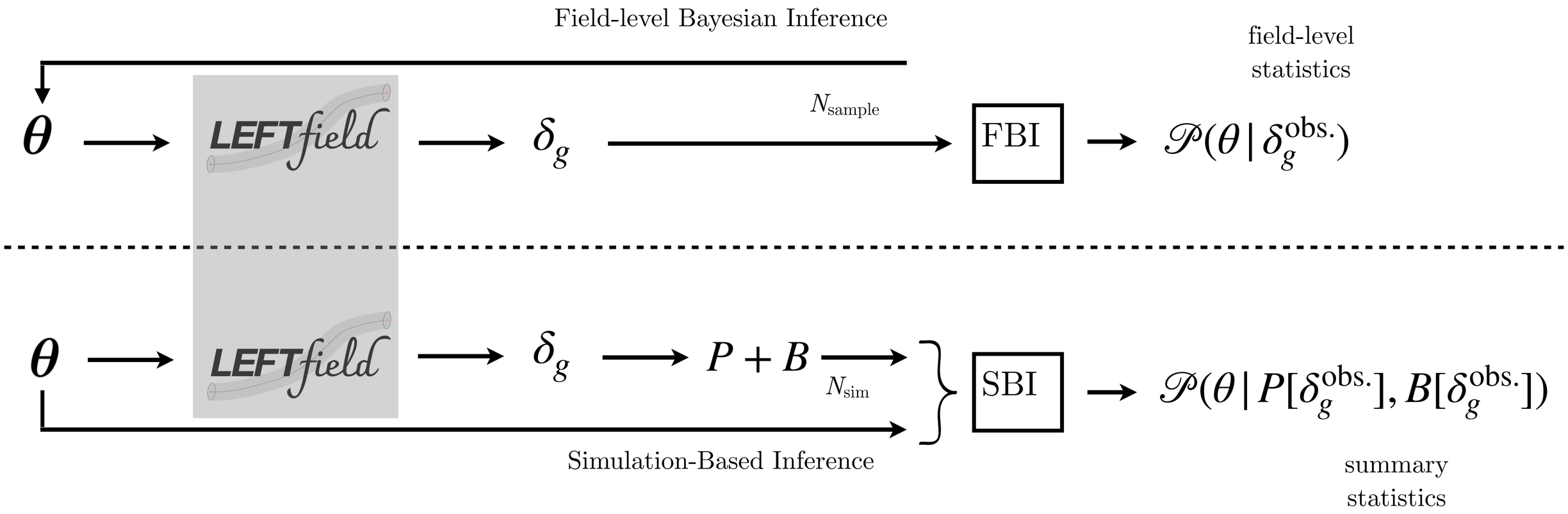
between field-level and summary statistics inference



MN+ 2024, Fig. 1

An apple-to-apple comparison

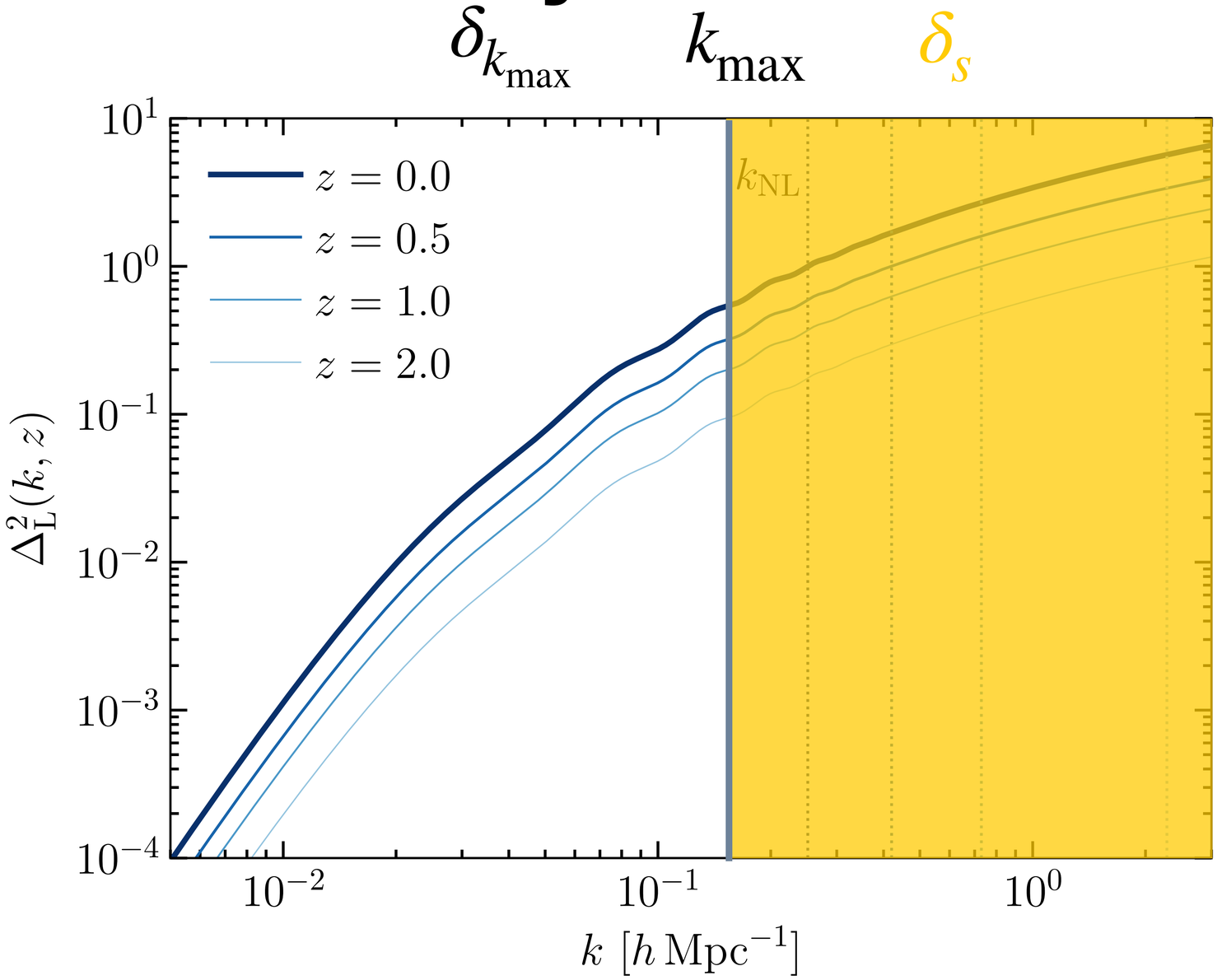
between field-level and summary statistics inference



MN+ 2024, Fig. 1

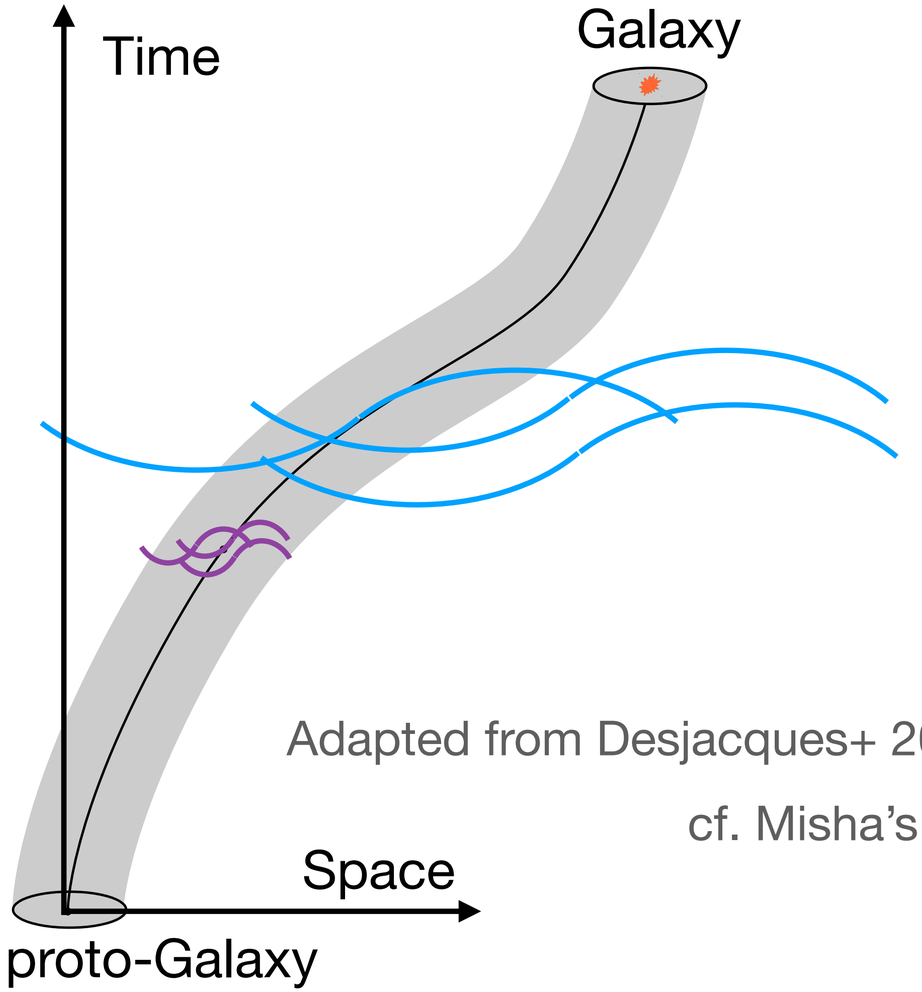
Forward Model

Effective Field Theory of LSS



Field-level EFT Model for Galaxy Clustering

Lagrangian, *EFT*-based forward model (LEFTfield)



Adapted from Desjacques+ 2016, Fig. 6

cf. Misha's talk

Field-level EFT bias expansion

$$\delta_g(\theta, \hat{s}) = \sum_O b_O O(\theta, \hat{s}) + \epsilon$$

+

n^{th}-order **Lagrangian** displacement field

=

LEFTfield

Field-level EFT Model for Galaxy Clustering

Lagrangian, EFT-based *forward* model (LEFTfield)

$$\delta_\Lambda^{(1)}(\boldsymbol{k}, z) = W_\Lambda(k)\sqrt{\alpha^2 P_L(k, z)}\,\hat{s}(\boldsymbol{k})$$

$$\alpha \equiv \sigma_8/\sigma_8^{\rm fid}$$

$$\hat{s}(\boldsymbol{x}) \sim \mathcal{N}(0, 1)$$

$$1 + \delta(\boldsymbol{x}, \tau) = |\mathbf{1} + \boldsymbol{M}(\boldsymbol{q}, \tau)|^{-1} \qquad M_{ij} \equiv \partial_i s_j$$

$$\mathrm{tr}[\boldsymbol{M}_\Lambda^{(1)}] = -\delta_\Lambda^{(1)}$$

LPT recursion relations

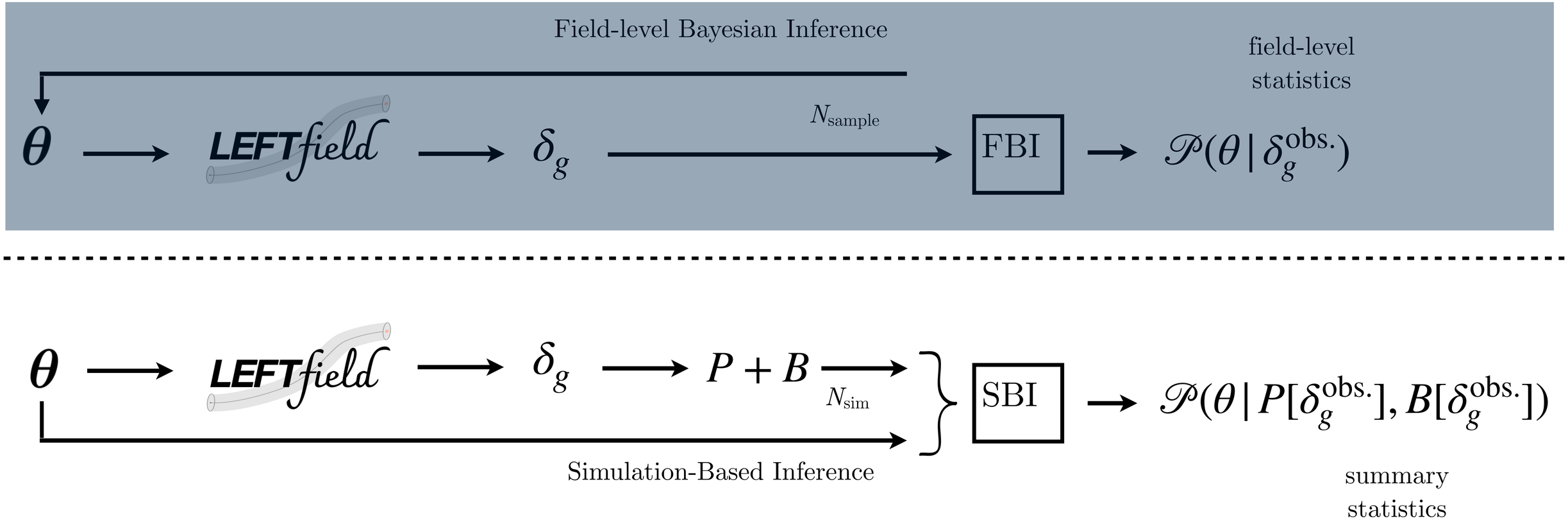
$\boldsymbol{s}^{(n)}$

n^{th}-order **Lagrangian** displacement field

2012.09837

An apple-to-apple comparison

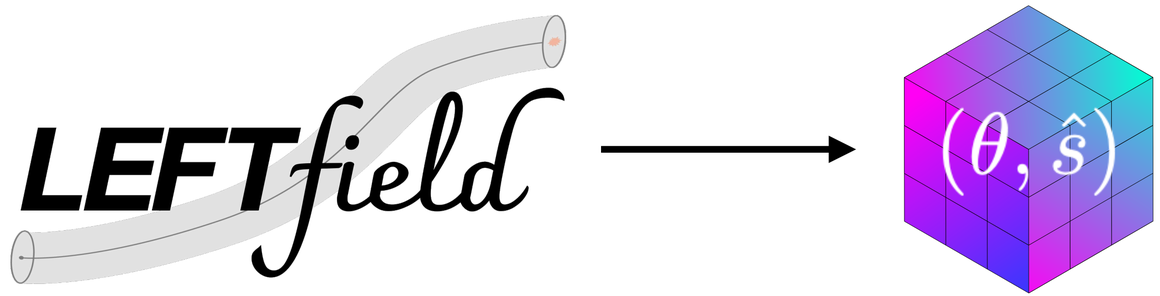
Field-level and summary statistics inference



MN+ 2024, Fig. 1

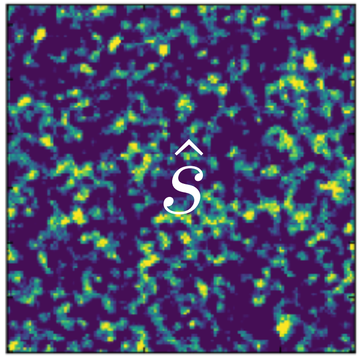
Field-level Bayesian Inference

Constrain both cosmology and initial conditions with LEFTfield



gradient-based sampling

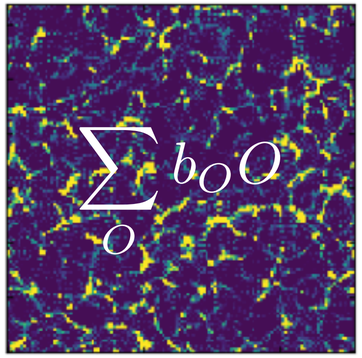
Gaussian



θ

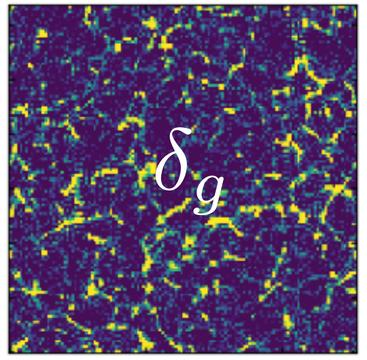
forward model

2012.09837



EFT likelihood

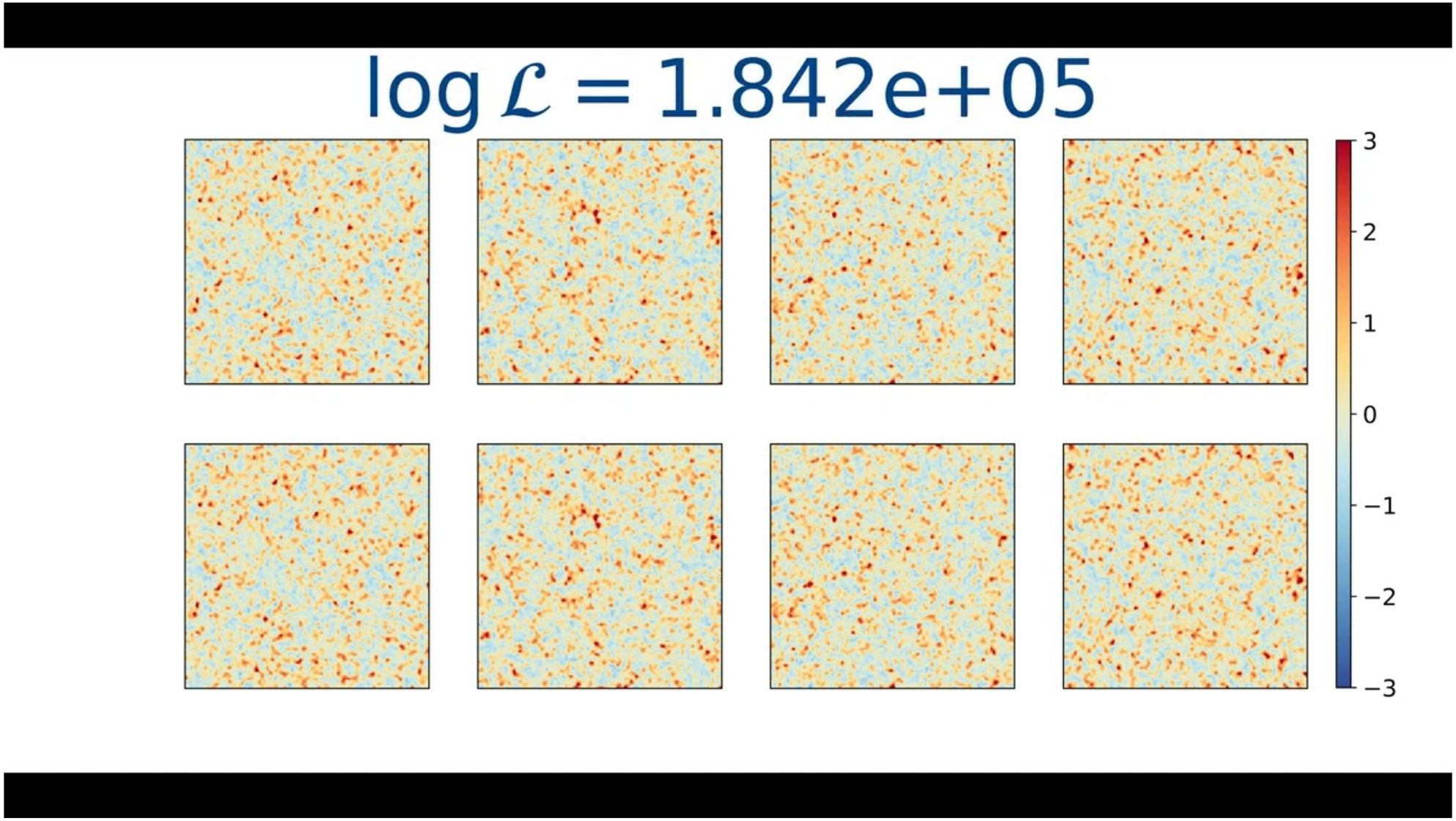
1909.04022



non-Gaussian

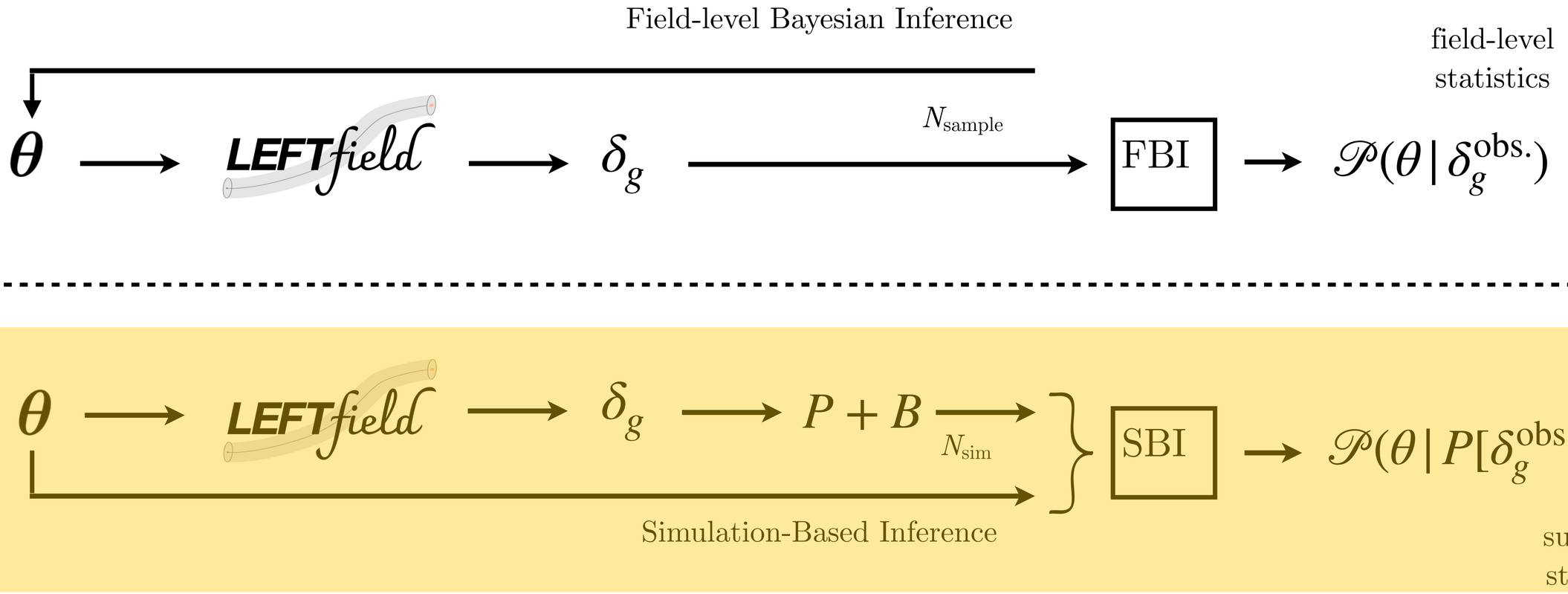
Adapted from Julia Stadler

Field-level Bayesian Inference in action



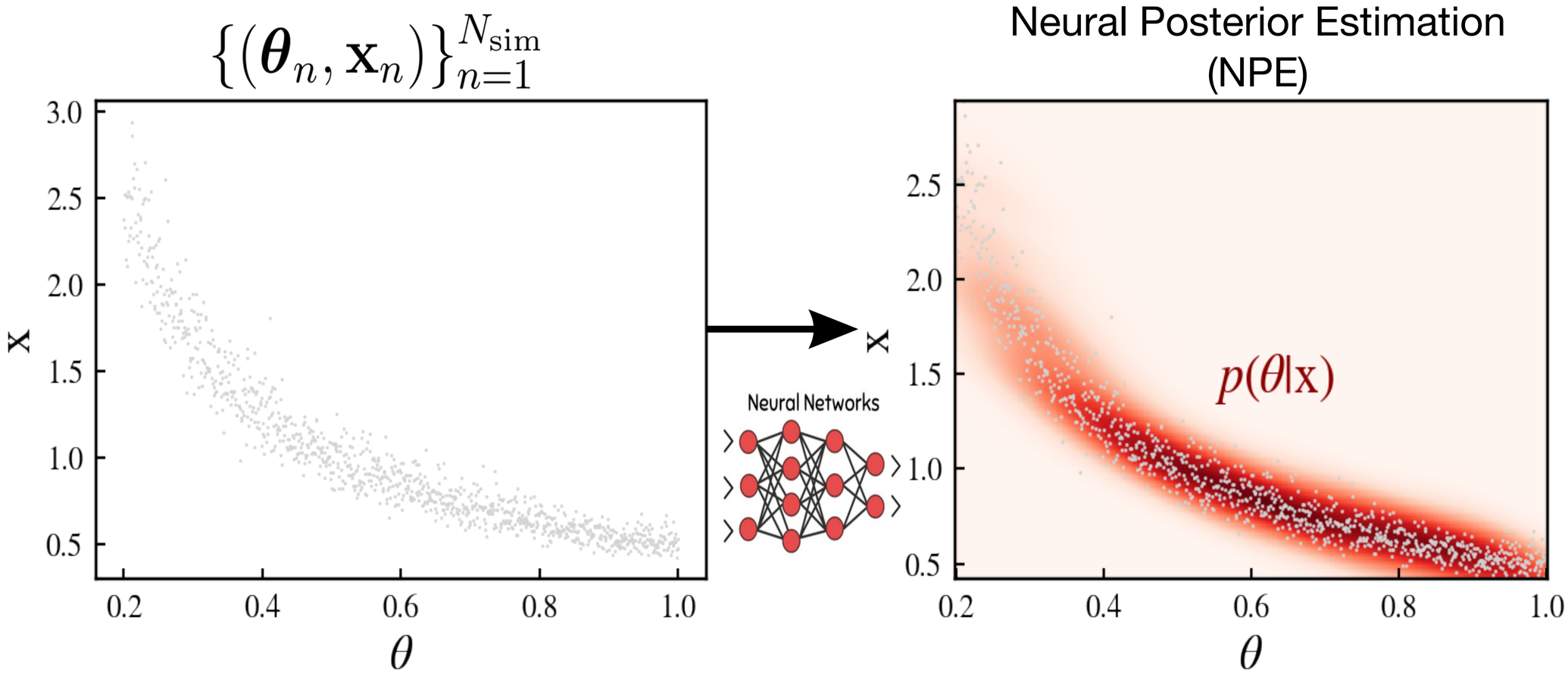
An apple-to-apple comparison

Field-level and summary statistics inference

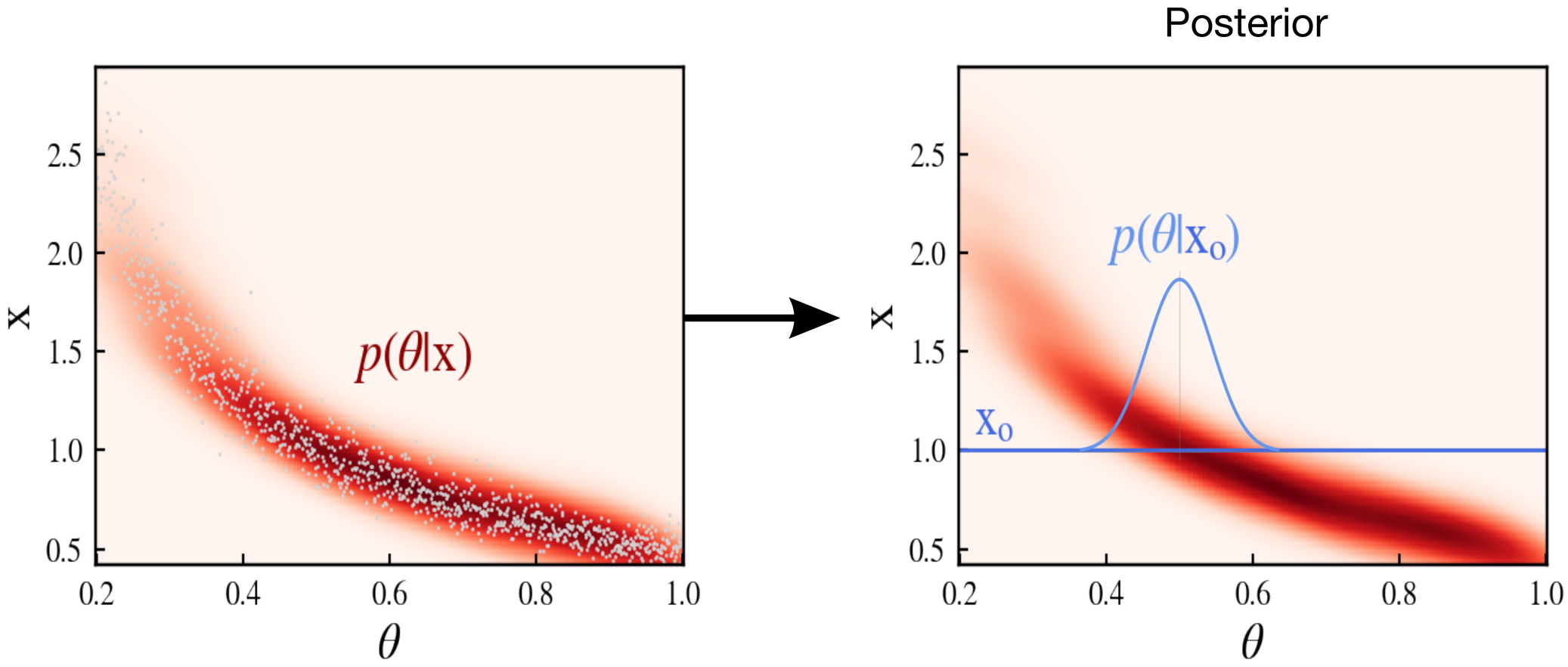


MN+ 2024, Fig. 1

Simulation-based Inference - step 1

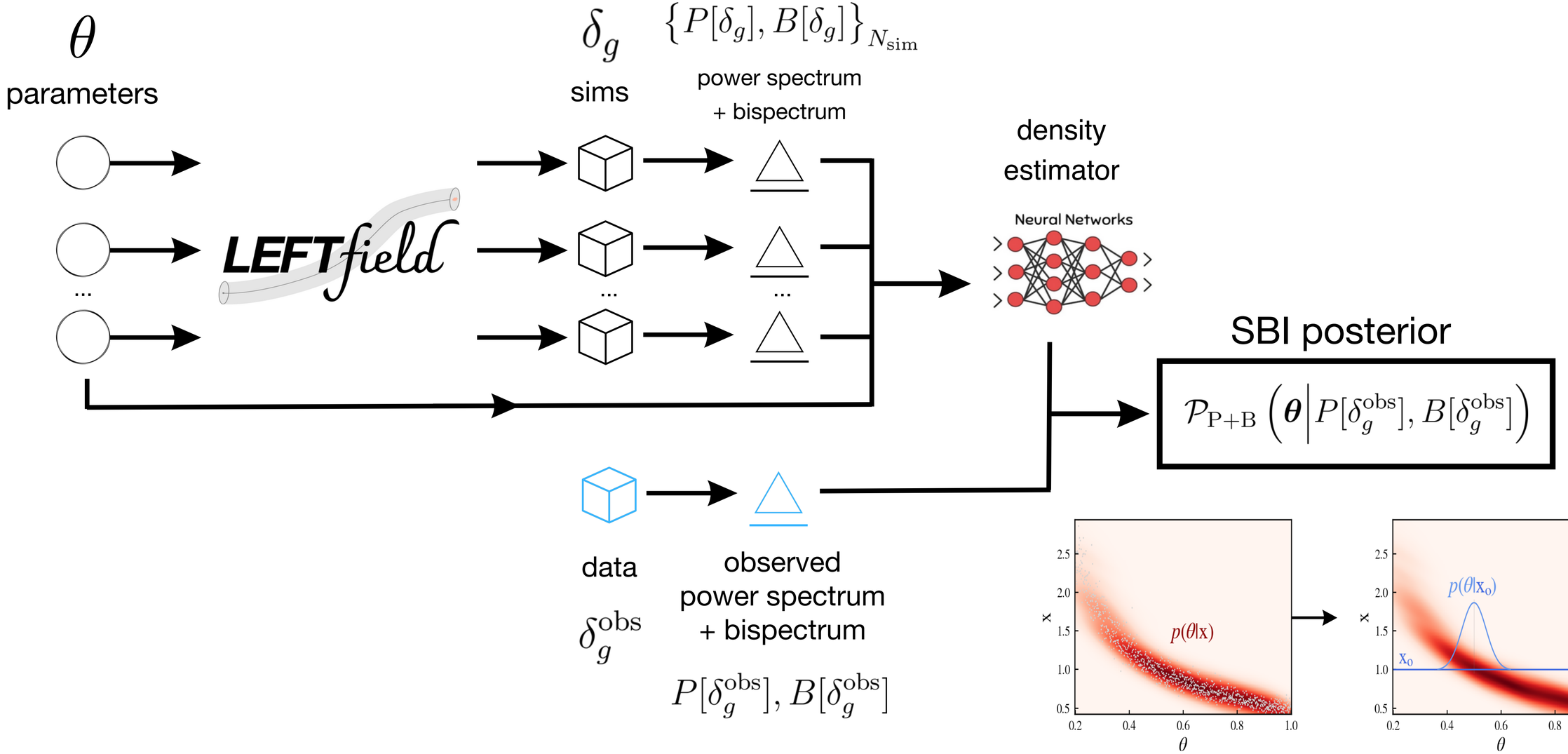


Simulation-based Inference - step 2



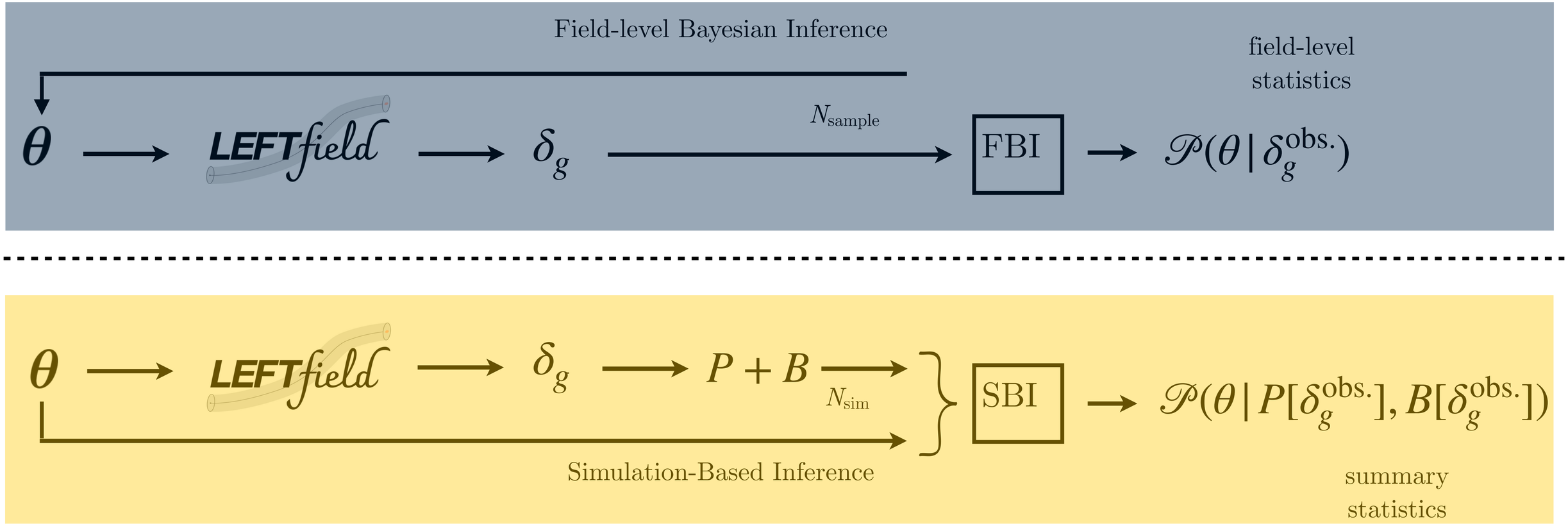
Simulation-based Inference of P+B

LEFTfield as the simulator



An apple-to-apple comparison

Field-level and summary statistics inference

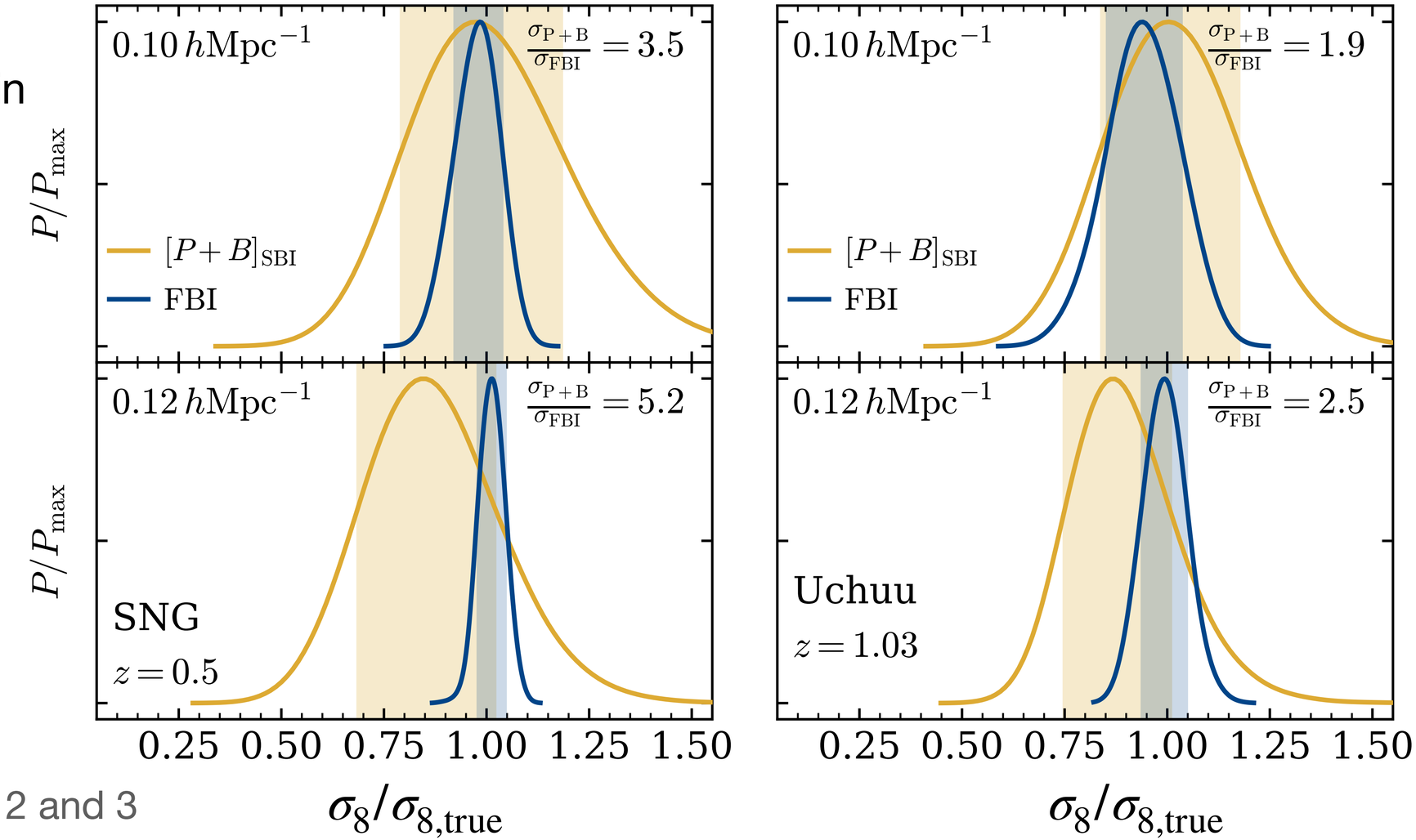


MN+ 2024, Fig. 1

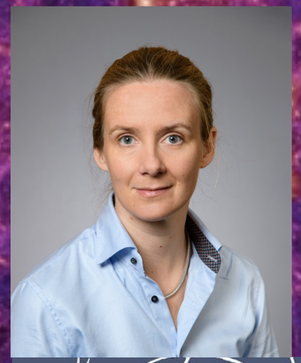
How much reliable information can we gain?

Several-factor improvement from P+B to field-level statistics

- Different samples of halos in different cosmologies [2403.03220]
- Exact *same* modes
- Constraints do not include information from RSD

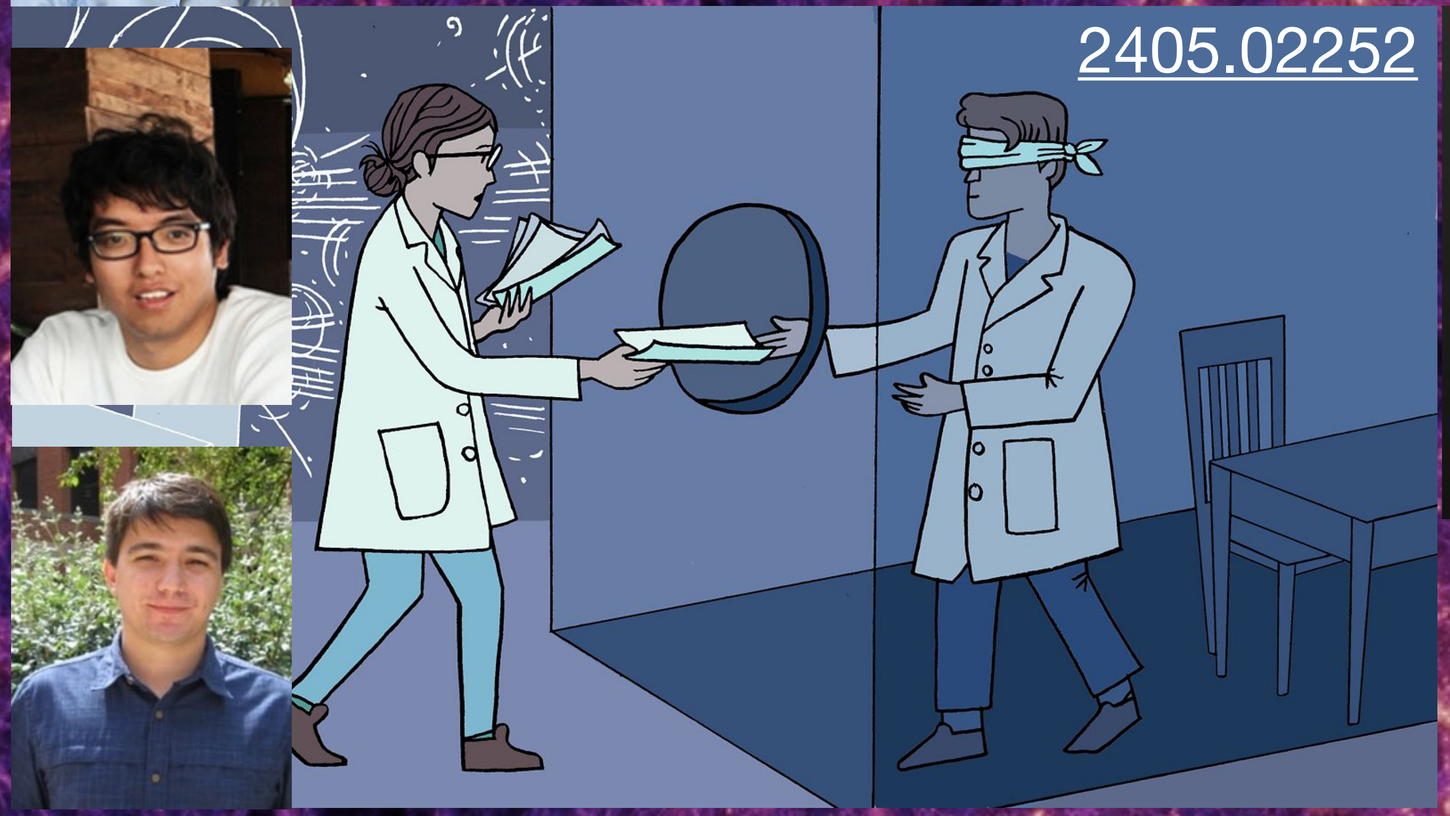


MN+ 2024, Fig. 2 and 3



Beyond-2pt parameter-masked challenge

Beyond-2pt Collaboration



This week on Cosmology Talks

A mock challenge for *beyond*-2pt statistics

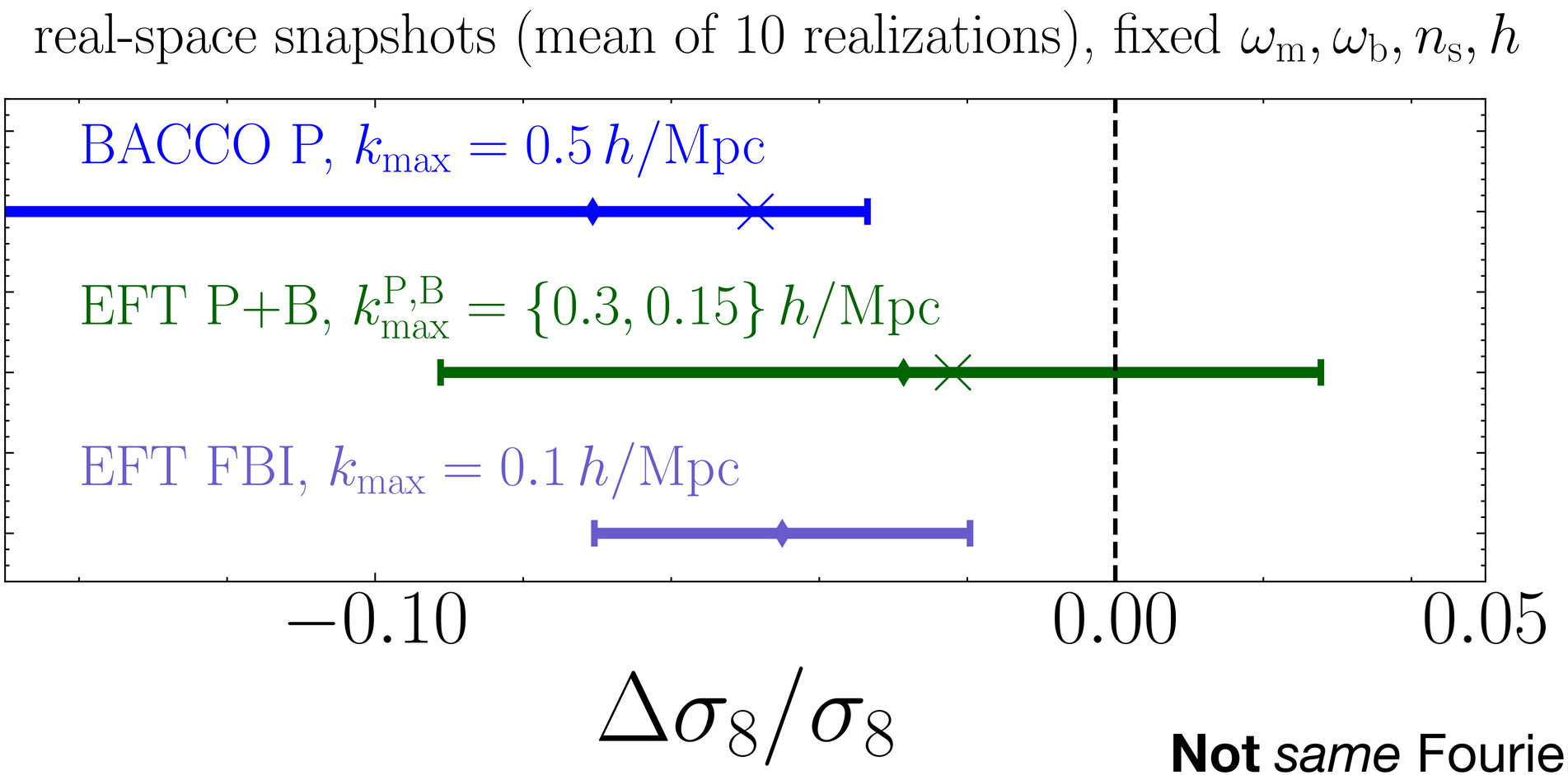
- N-body halos populated with HOD galaxies to produce galaxy catalogs
- Different setups: real-space snapshots, redshift-space snapshots, light cone
- Cosmology and HOD models kept masked
- Real-space and redshift-space snapshots:
  - $z = 1.0$
  - $V = 8\,(\mathrm{Gpc}/h)^3$
  - 10 realizations

N-body simulations

HOD catalogs

Beyond-2pt challenge - real-space mocks

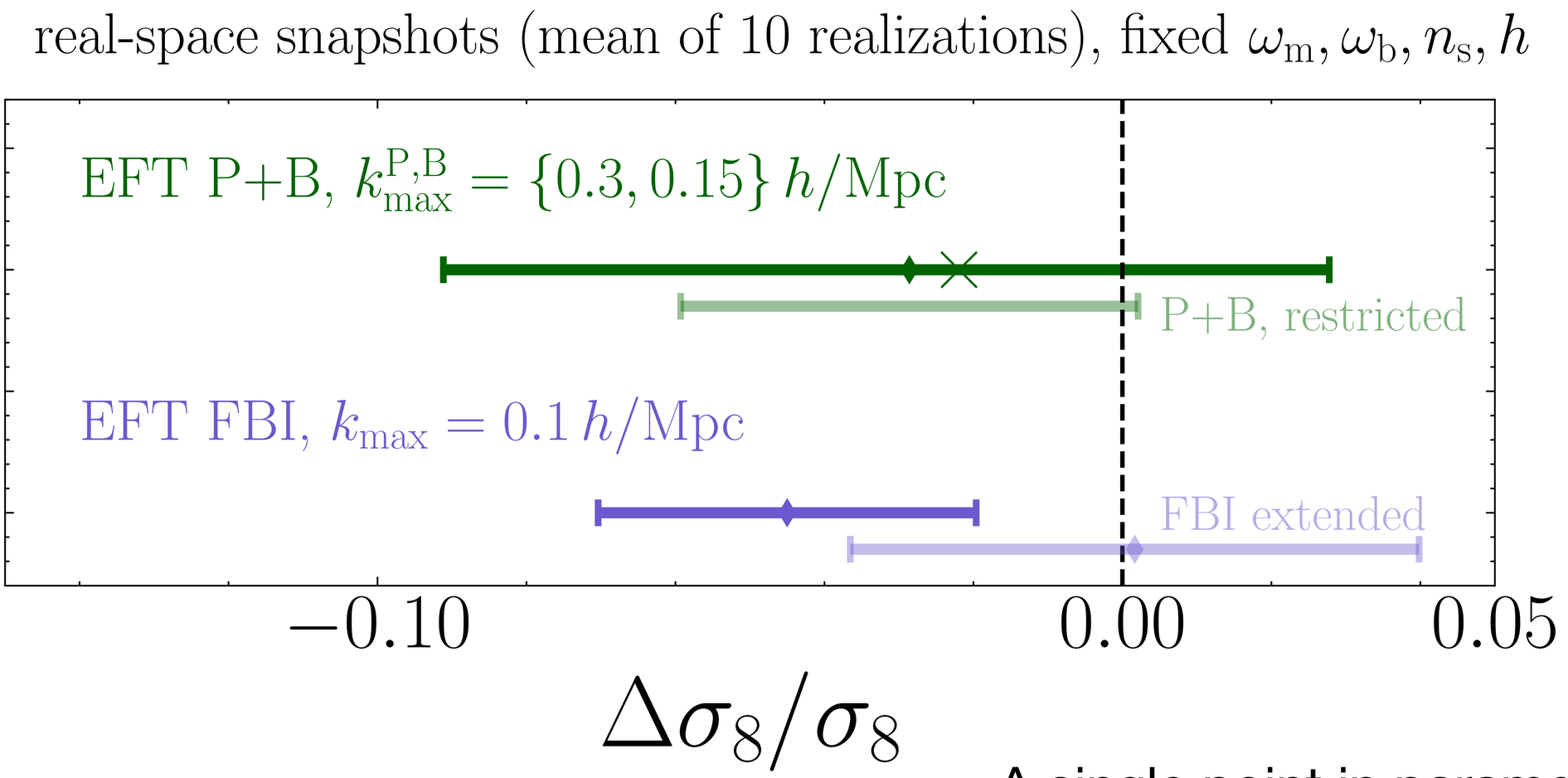
Pinning down σ_8 *without* RSD information



Not *same* Fourier modes

Beyond-2pt challenge - Post-unmasking

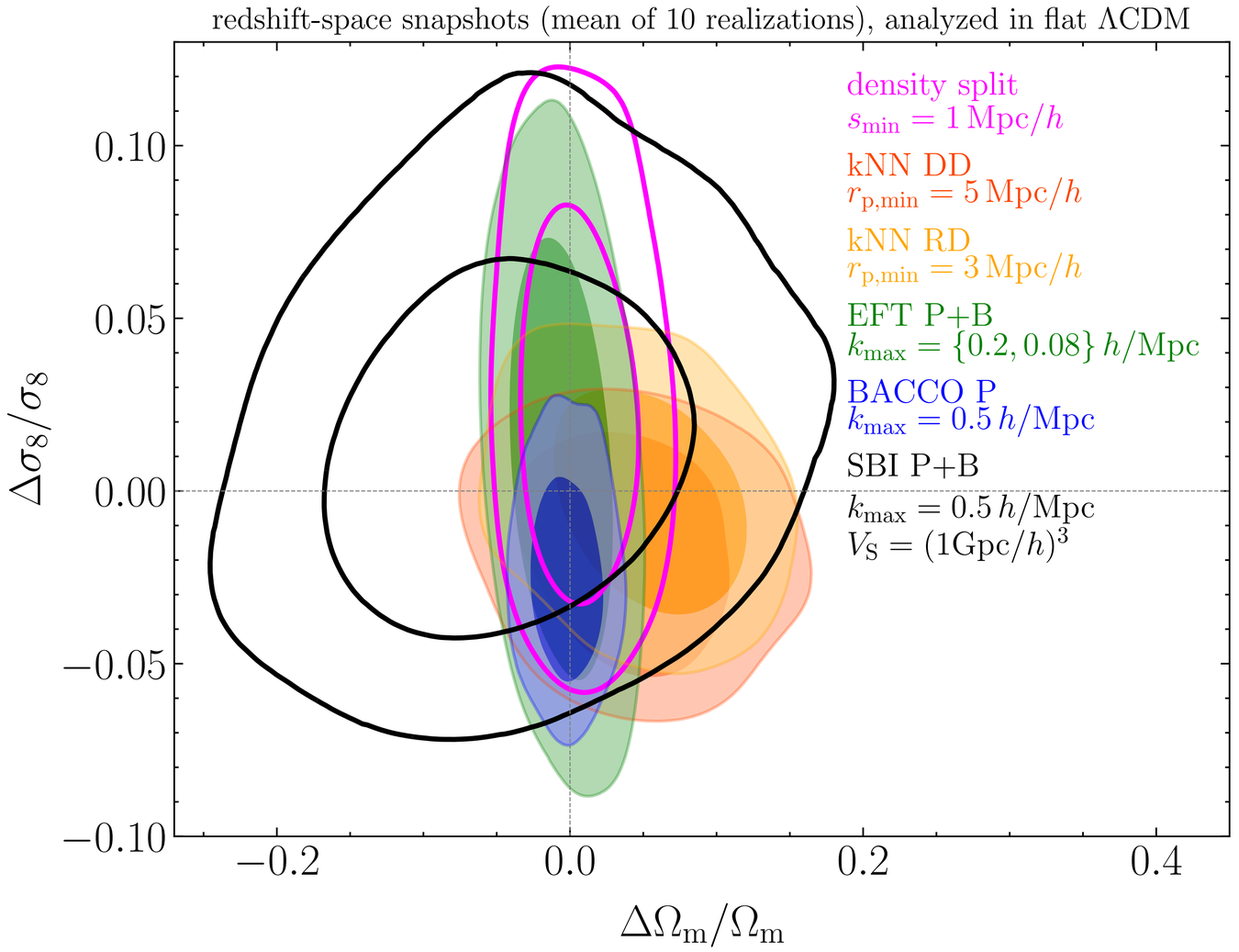
P+B and FBI in practice (*not* a comparison)



A single point in parameter space

Beyond-2pt challenge - redshift-space mocks

All consistent within 68% CL



Outlook for future survey application of FBI

Connect EFT FBI and beyond-2pt methods to spectroscopic surveys

- A lot of information can be (robustly) extracted from galaxy spectroscopic surveys, even at quasi-linear regime
- Apple-to-apple comparison show field-level statistics can improve several factors over P+B, implying order of magnitude increase of effective volume
- **FBI** to-do #1: Extend parameter and observable space (e.g. shape, velocity)
- **FBI** to-do #2: Include observational effects and survey systematics
- **Community** to-do: coordinate efforts between simulation and analysis methods

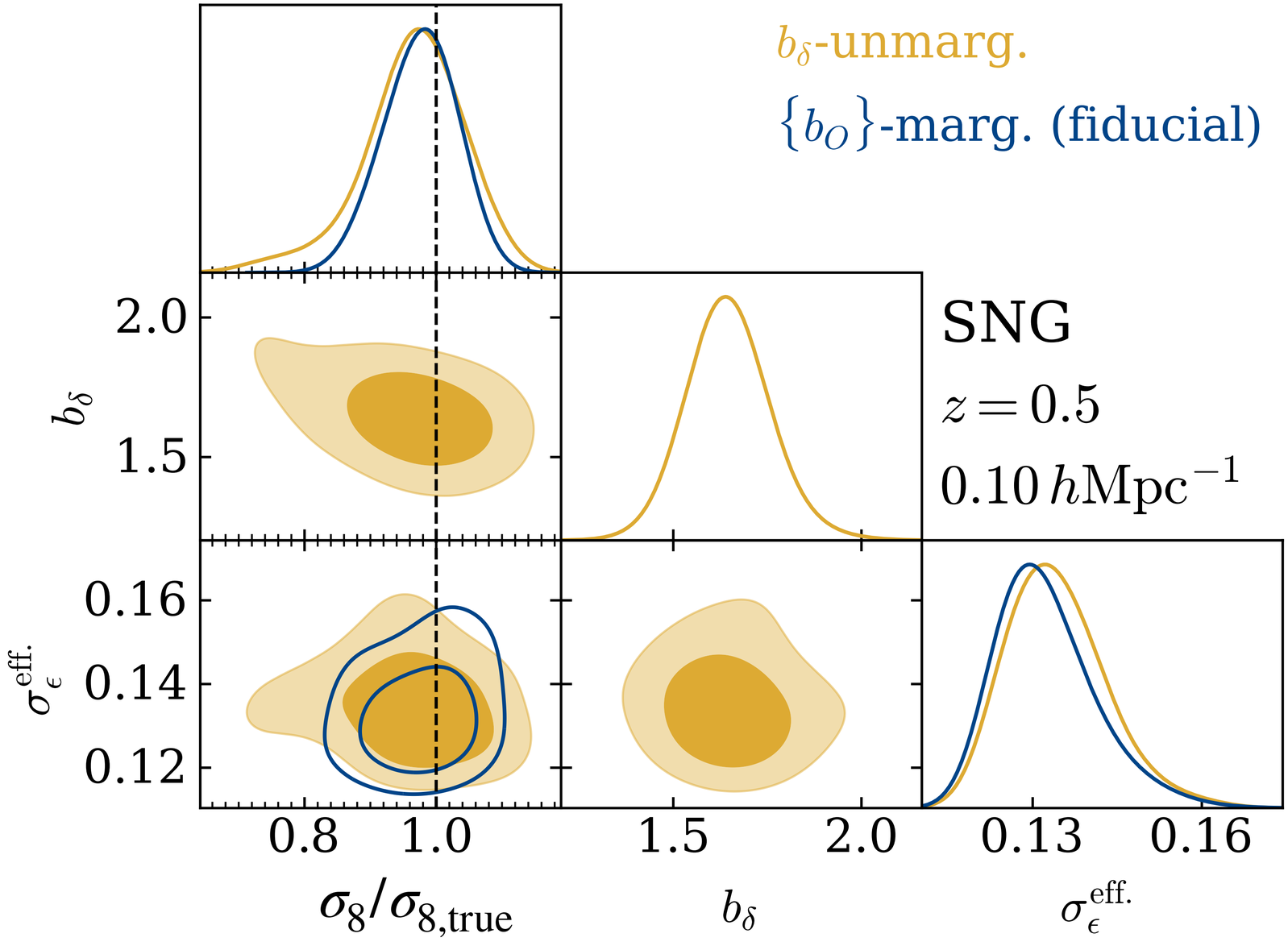
Thank you!

Details? Please check out the Cosmology Talks interviews!

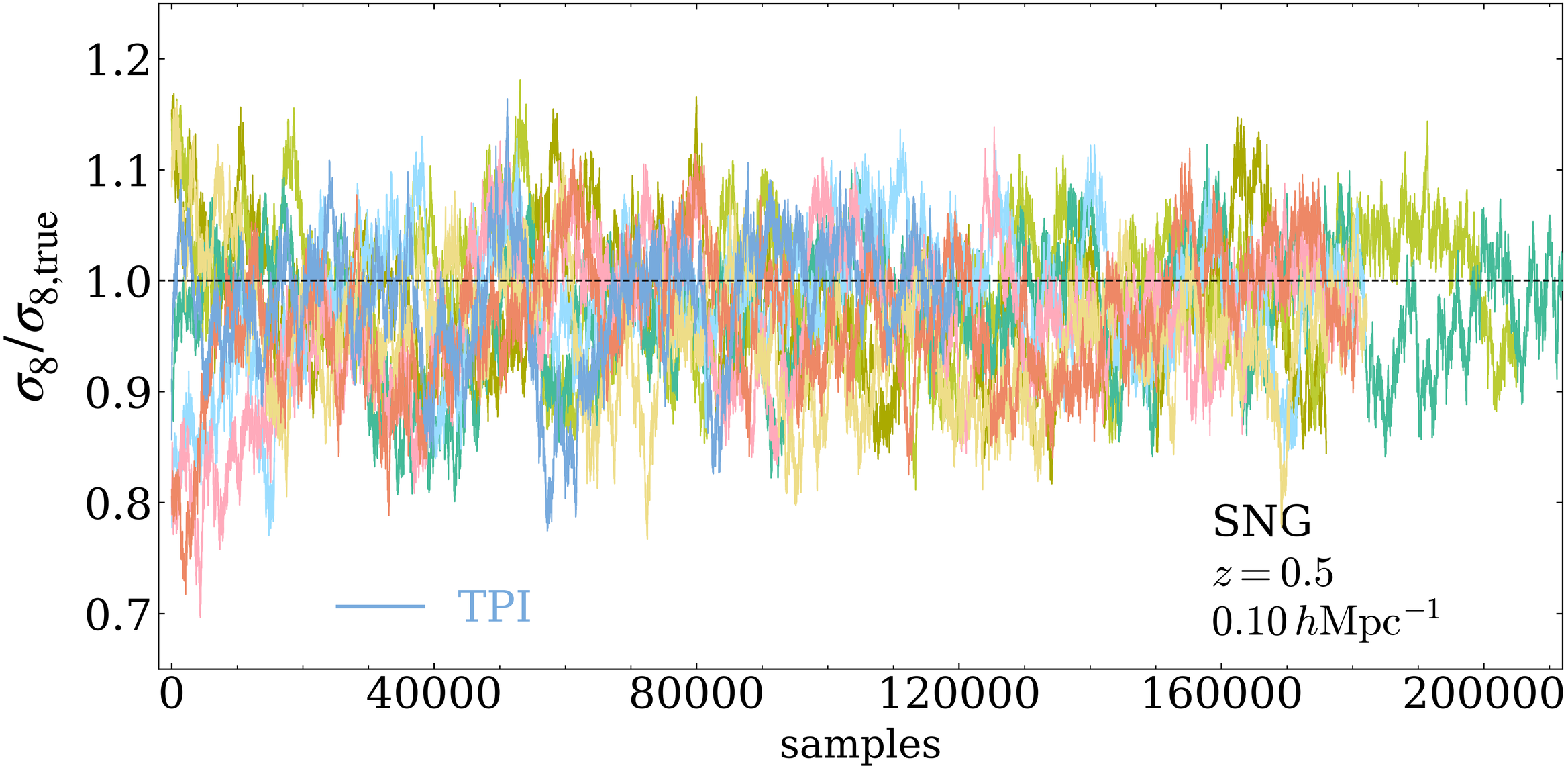
Questions? nguyenmn@umich.edu

Details for Experts

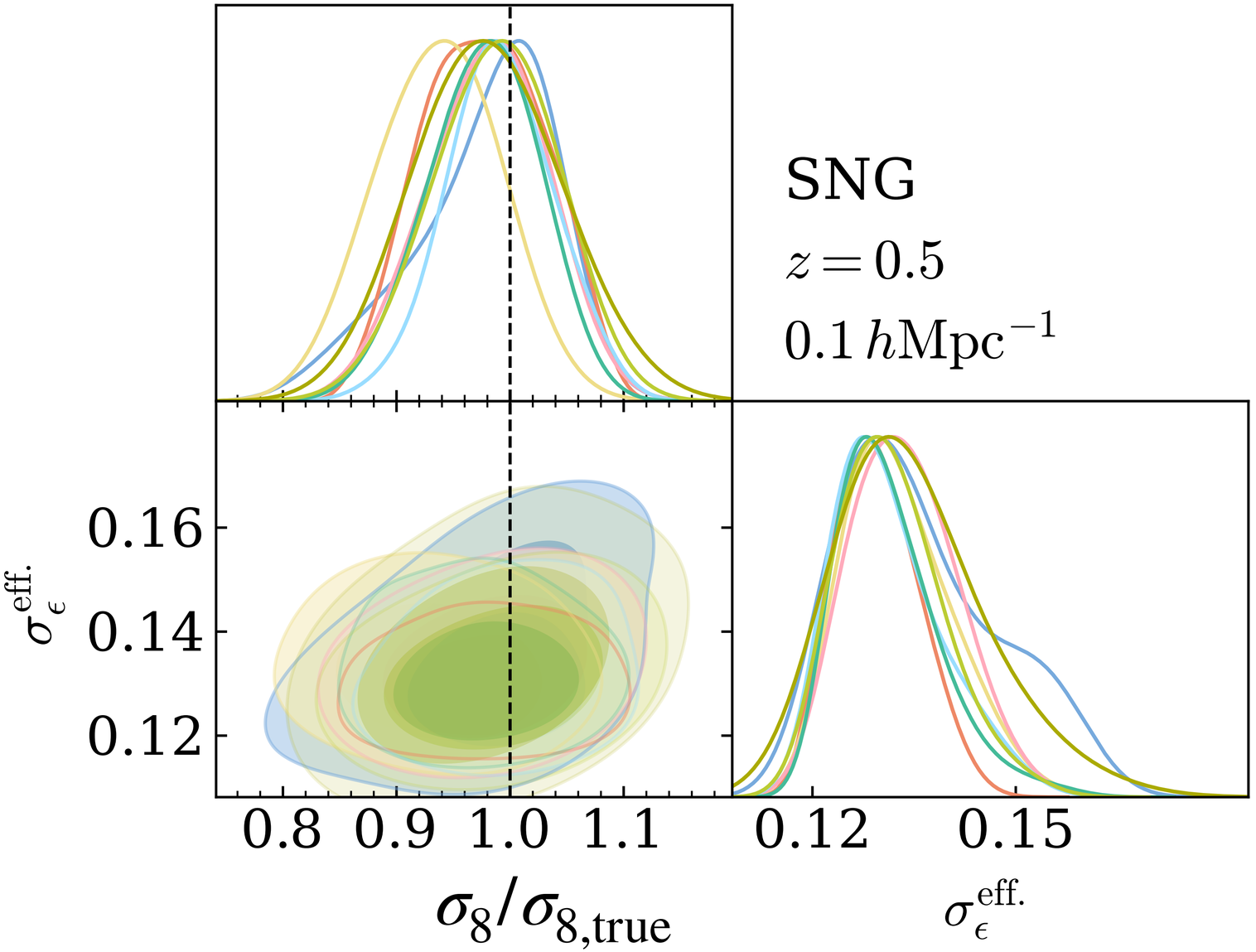
FBI - Breaking bias—σ_8 degeneracy



FBI - MCMC convergence with raw samples



FBI - Posterior consistency between MCMC chains



SBI P+B - Including non-Gaussian noise

Bispectrum stochasticity contains non-Gaussian contributions

Perturbation Theory

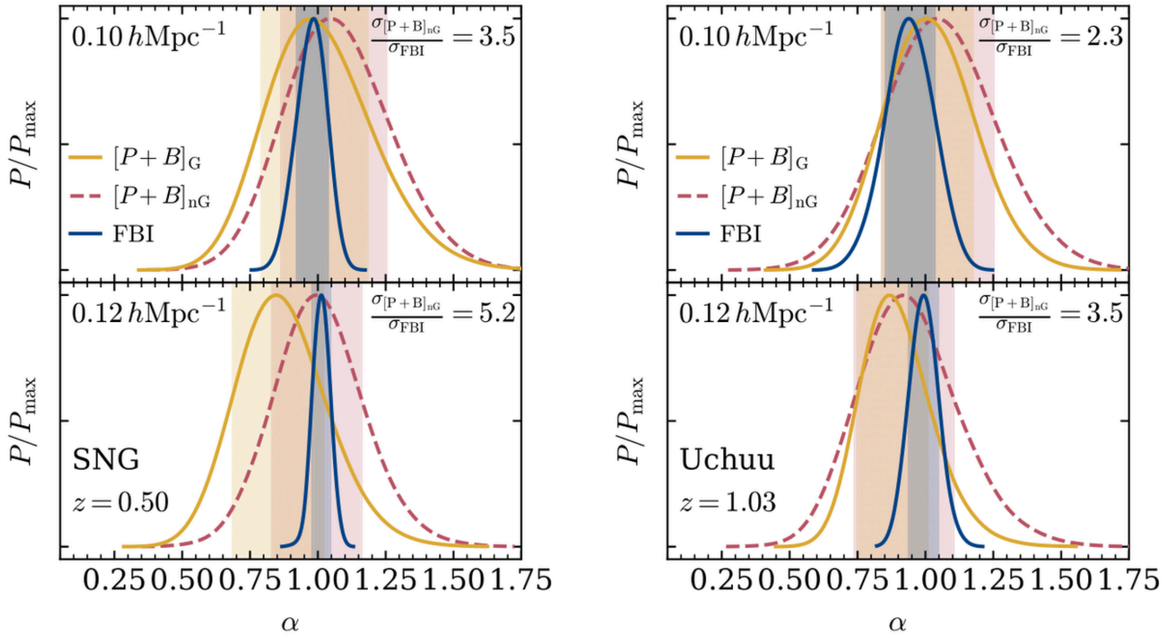
$$\langle \delta_g(k_1)\delta_g(k_2)\delta_g(k_3)\rangle'^{\rm LO}_{\rm stoch} = B_\varepsilon + 2b_1 P_{\varepsilon\varepsilon_\delta}(P_{\rm m}(k_1) + 2 \text{ perm.})$$

Forward Model

$$\langle \delta_g(k_1)\delta_g(k_2)\delta_g(k_3)\rangle'^{\rm LO}_{\rm stoch} = 6c_\varepsilon^{\rm NG} P_\varepsilon^2 + 2b_1 P_\varepsilon \sigma_{\varepsilon\delta}(P_{\rm m}(k_1) + 2 \text{ perm.})$$

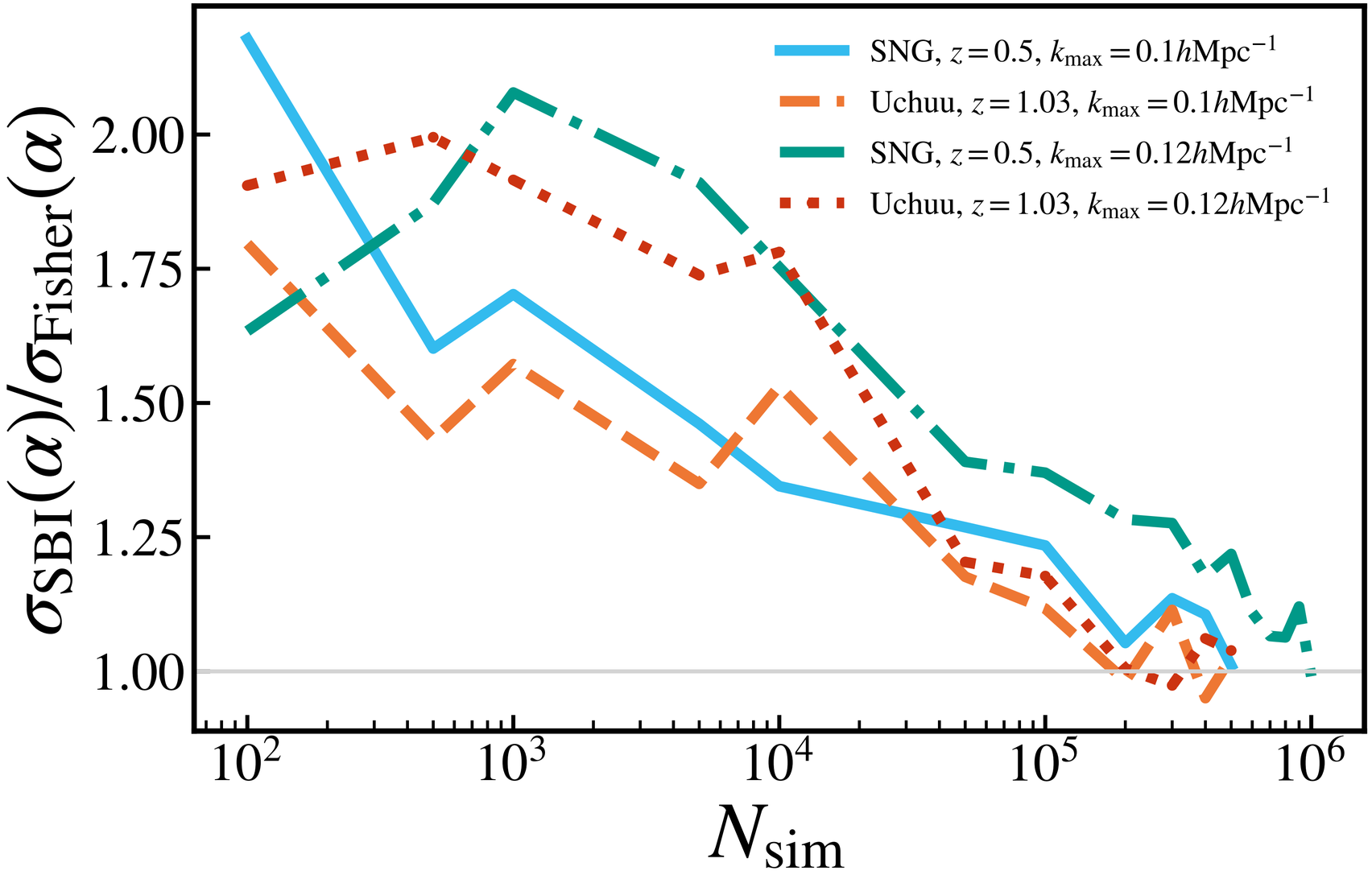
$$\delta_g(\mathbf{x},\tau) = \delta_{g,\rm det}(\mathbf{x},\tau) + \varepsilon(\mathbf{x},\tau) + \sigma_{\varepsilon\delta}(\tau)\varepsilon(\mathbf{x},\tau)\delta(\mathbf{x},\tau) + c_\varepsilon^{\rm NG}(\tau)\varepsilon^2(\mathbf{x},\tau)$$

What if we account for that in our SBI P+B analysis?



See also Misha's talk, Ivanov+, Philcox+Ivanov

SBI P+B - Convergence with $N_{\rm sim}$



SBI P+B - Coverage tests (Simulation-based calibration)

$$\alpha = \sigma_8/\sigma_{8,\mathrm{true}}$$

No sign of over- or under-estimation of parameter uncertainties