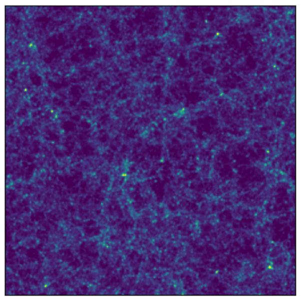
How much information is left in galaxy clustering beyond the power spectrum?

Kazuyuki Akitsu (High Energy Accelerator Research Organization (KEK))

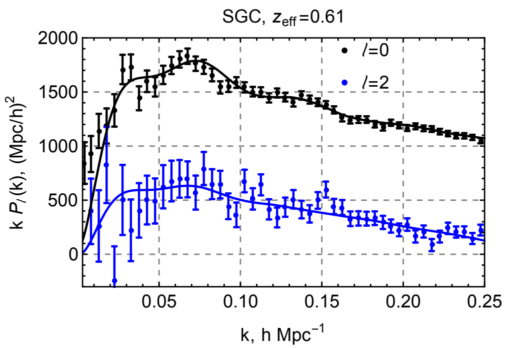
with Shi-Fan Chen, Giovanni Cabass, Marko Simonović and Matias Zaldarriaga

Standard cosmology analysis

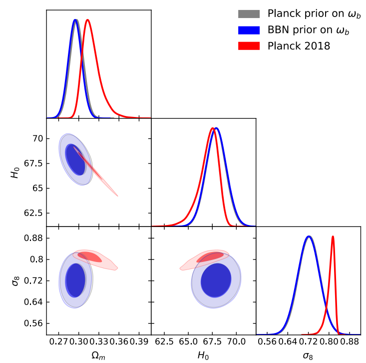
1. Measure the summary statistics
    - e.g. 2-pt function/power spectrum
2. Develop a theoretical prediction for that summary statistics
    - perturbation theory and/or N-body simulation
3. Construct an analytic likelihood with covariance
    - usually assumed to be Gaussian $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}\left[\mathbf{d}-\mathbf{m}(\boldsymbol{\theta})\right]^{\mathrm{T}}\mathbf{Cov}^{-1}\left[\mathbf{d}-\mathbf{m}(\boldsymbol{\theta})\right]\right]$
4. Run MCMC to infer the posterior distribution $\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$



Compress

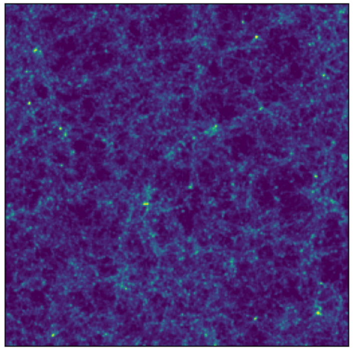


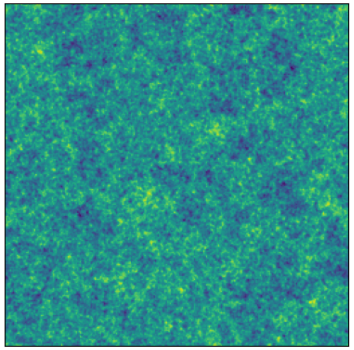
MCMC

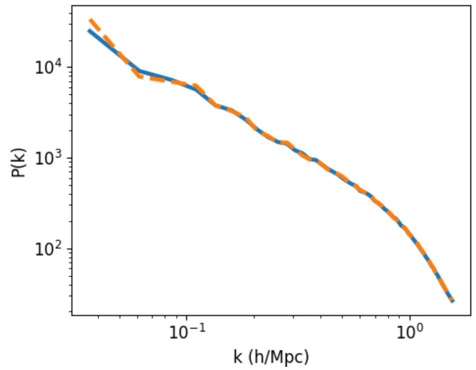


But... which summary statistics?

- The power spectrum is **no longer** sufficient for galaxy clustering.







- Higher-order statistics? (or any other summary statistics?)
  - Hard to measure, and evaluating their covariance is challenging

$$\mathrm{Cov}[\hat{P}_g(\mathbf{k}), \hat{P}_g(\mathbf{k}')] \supset \langle \delta_g(\mathbf{k})\delta_g(-\mathbf{k})\delta_g(\mathbf{k}')\delta_g(-\mathbf{k}')\rangle_c$$

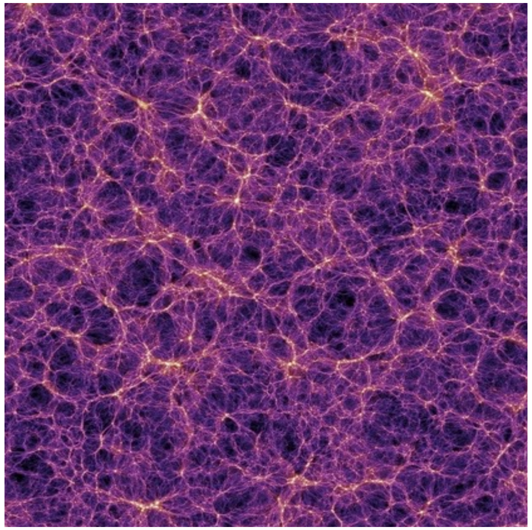
  - In principle, there are infinite series of n-pt functions.

- What is the optimal way to analyze non-Gaussian random field?
  - Field-level inference

Field-level inference

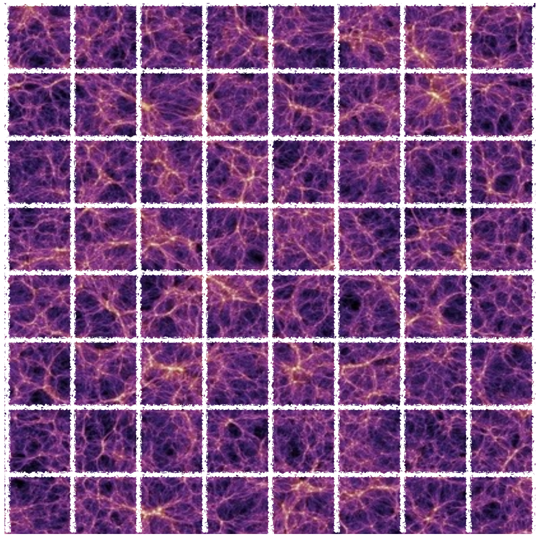
Jasche&Wandelt13, Wang++14, Schmidt++18, Kitaura++21, …

- Cosmology inference without data compression



Observed distribution

Compress →

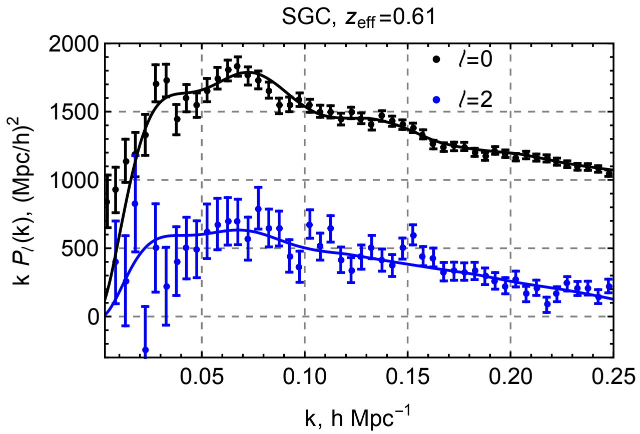


Number density field

$$\delta_g(\mathbf{x}) \equiv \frac{n_g(\mathbf{x}) - \bar{n}_g}{\bar{n}_g}$$

Compress →

$$\langle \delta_g(\mathbf{k}) \delta_g(\mathbf{k}') \rangle = (2\pi)^3 \delta_\mathrm{D}(\mathbf{k}+\mathbf{k}') P_{gg}(k)$$



Power spectra

- Compare the observation with theory **at each voxel (or Fourier mode)**
  - In principle, all available information can be extracted.

The posterior of the field-level inference

- The field-level posterior of cosmological parameters $\boldsymbol{\theta}$ given the data $\hat{\delta}_g$:

$$\mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \propto \int \mathcal{D}\delta_{\rm IC}\ \mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC},\boldsymbol{\theta})] \times \pi[\delta_{\rm IC},\boldsymbol{\theta}]$$

- The prior of the initial conditions: Gaussian with diagonal covariance in Fourier space

$$\pi[\delta_{\rm IC},\boldsymbol{\theta}] \propto \exp\left(-\frac{1}{2}\int\frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\delta_{\rm IC}(\mathbf{k})|^2}{P_{\rm lin}(\boldsymbol{\theta};k)}\right) \qquad \langle\delta(\mathbf{k})\delta^*(\mathbf{k}')\rangle = (2\pi)^3\delta_{\rm D}(\mathbf{k}-\mathbf{k}')P_{\rm lin}(k)$$

- Assume the likelihood form to be Gaussian

$$\mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC},\boldsymbol{\theta})] \propto \exp\left(-\frac{1}{2}\int\frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\hat{\delta}_g(\mathbf{k})-\delta_{\rm model}(\boldsymbol{\theta},\delta_{\rm IC};\mathbf{k})|^2}{P_\epsilon}\right)$$

- We need a forward-model of galaxy clustering at field level
  - Lagrangian EFT of LSS Schmittfull+ 19, Schmittfull+ 21

$$\delta_{\rm model}(\mathbf{k}) = \sum_i b_i\tilde{\mathcal{O}}_i(\mathbf{k}) \qquad \tilde{\mathcal{O}}_i(\mathbf{k}) = \int {\rm d}^3\mathbf{q}\ \mathcal{O}(\mathbf{q})e^{i\mathbf{k}\cdot(\mathbf{q}+\boldsymbol{\Psi}_Z(\mathbf{q}))}$$

Challenge: Marginalization over the ICs

- Explicit (or numerical) marginalization over the initial conditions

$$\mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \propto \int \mathcal{D}\delta_{\rm IC}\, \mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC}, \boldsymbol{\theta})] \times \pi[\delta_{\rm IC}, \boldsymbol{\theta}]$$

  - We have to evaluate the extremely high dimensional integral w.r.t. the ICs
  - The dims of the integral $\propto N_{\rm pixel} > 10^6$

- Common practice: compress the data to summary statistics

$$\langle \cdots \rangle \propto \int \mathcal{D}\delta_{\rm IC}\, (\cdots) \exp\left(-\frac{1}{2}\int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^3} \frac{|\delta_{\rm IC}(\mathbf{k})|^2}{P_{\rm lin}(\boldsymbol{\theta}; k)}\right)$$

Hamiltonian Monte Carlo (HMC)

$$\mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \propto \int \mathcal{D}\delta_{\rm IC}\ \mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC},\boldsymbol{\theta})] \times \pi[\delta_{\rm IC},\boldsymbol{\theta}]$$

$$\propto \int \mathcal{D}\delta_{\rm IC}\ \ \exp\left(-\frac{1}{2}\int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\hat{\delta}_g(\mathbf{k}) - \delta_{\rm model}(\boldsymbol{\theta},\delta_{\rm IC};\mathbf{k})|^2}{P_\epsilon}\right) \times \exp\left(-\frac{1}{2}\int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\delta_{\rm IC}(\mathbf{k})|^2}{P_{\rm lin}(\boldsymbol{\theta};k)}\right)$$

cf. $Z \propto \int \mathcal{D}\phi\ e^{iS[\phi]}$

- The technique initially developed in the lattice QCD Duane+ 87
  - use the gradient information to sample efficiently
  - need the gradient of likelihood w.r.t. parameters

$$\frac{\partial \mathcal{L}[\hat{\delta}_g|\boldsymbol{\theta},\delta_{\rm IC}(\mathbf{k})]}{\partial \delta_{\rm IC}(\mathbf{k}')} \quad \frac{\partial \mathcal{L}[\hat{\delta}_g|\boldsymbol{\theta},\delta_{\rm IC}(\mathbf{k})]}{\partial \boldsymbol{\theta}}$$

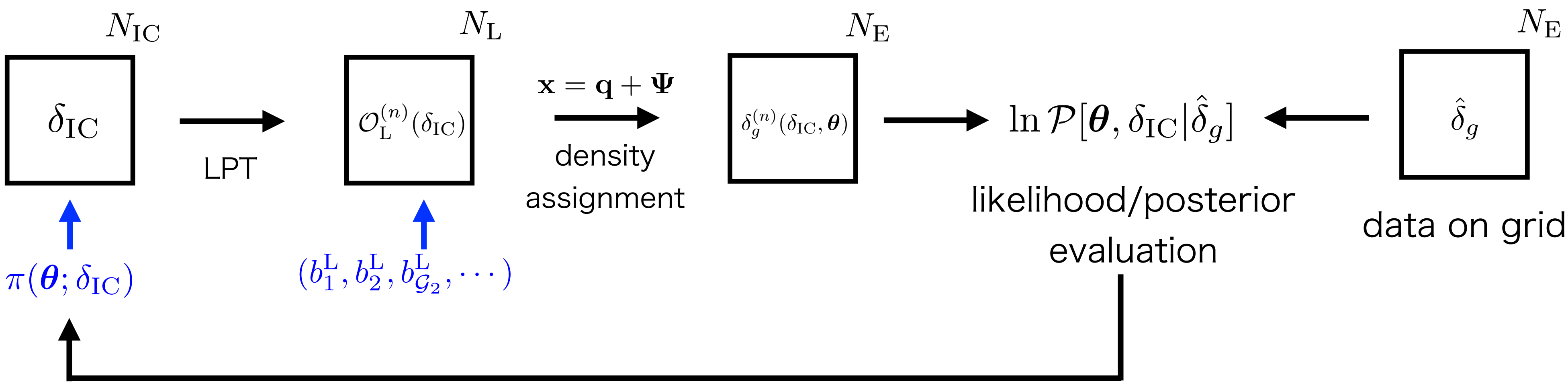
Field-level inference in practice

- We want to numerically evaluate

$$\mathcal{P}[\boldsymbol{\theta}|\hat{\delta}_g] \propto \int \mathcal{D}\delta_{\rm IC}\ \mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC},\boldsymbol{\theta})] \times \pi[\delta_{\rm IC},\boldsymbol{\theta}]$$

$$\propto \int \mathcal{D}\delta_{\rm IC}\ \exp\left(-\frac{1}{2}\int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\hat{\delta}_g(\mathbf{k})-\delta_{\rm model}(\boldsymbol{\theta},\delta_{\rm IC};\mathbf{k})|^2}{P_\epsilon}\right) \times \exp\left(-\frac{1}{2}\int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\delta_{\rm IC}(\mathbf{k})|^2}{P_{\rm lin}(\boldsymbol{\theta};k)}\right)$$

  - Infer the cosmological parameters and the initial conditions simultaneously



Implementation details

- Lagrangian EFT of LSS: a differentiable simulator Schmittfull+ 19, Schmittfull+ 21

$$\delta_{\rm model}[\delta_{\rm IC}, \boldsymbol{\theta}] = b_1 \tilde{\delta}_{\rm IC} + \frac{1}{2} b_2 \tilde{\delta^2}_{\rm IC} + b_{\mathcal{G}_2} \tilde{\mathcal{G}}_2[\delta_{\rm IC}] + \cdots \qquad \tilde{\mathcal{O}}_i(\mathbf{k}) = \int \mathrm{d}^3\mathbf{q}\ \mathcal{O}(\mathbf{q}) e^{i\mathbf{k}\cdot(\mathbf{q}+\boldsymbol{\Psi}_Z(\mathbf{q}))}$$

  - Not so computationally expensive (accelerated by GPUs)
  - Easy to take a derivatives w.r.t. initial random fields as well as biases
    - Auto-grad library (e.g. JAX)
    - Possible to be extended to use differentiable N-body (Hybrid-EFT)

      e.g. FlowPM(Modi++20), pmwd(Li++22) Modi++20
- Sampling method: No-U-Turn Sampler(NUTS)
  - adaptive mass matrix & step sizes
- Written in JAX (python), running on GPUs (Nvidia A100)

A sanity check: Gaussian random field

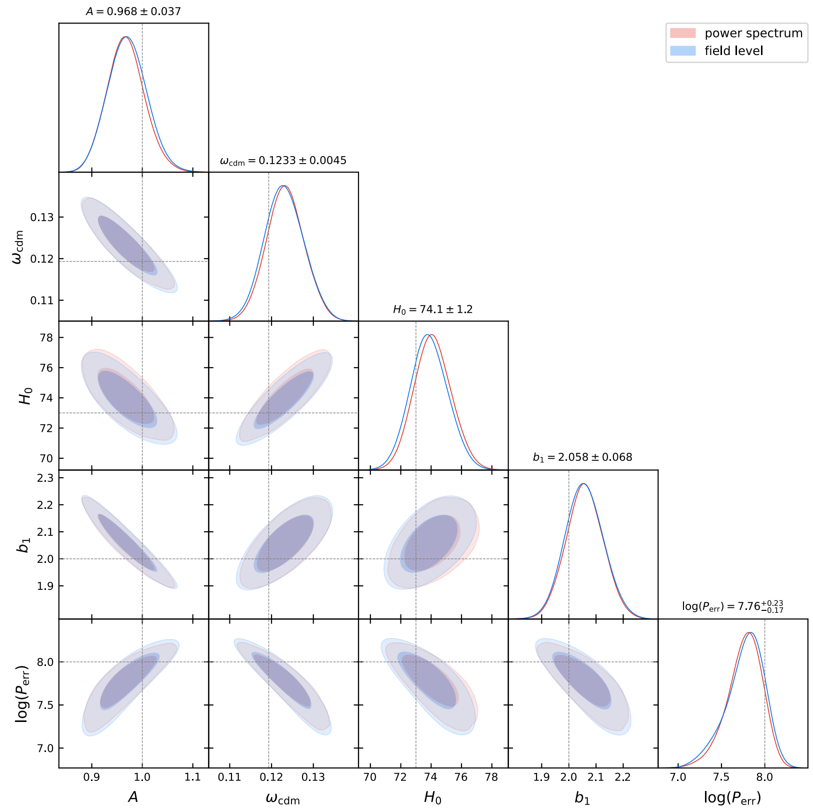
- Model: $\delta_g = \mathcal{M}(\boldsymbol{\theta})\delta_{\rm IC}$
  Gaussian
  - parameters: $(A, \omega_{\rm cdm}, H_0, b_1, P_\epsilon; \{\delta_{\rm IC}\})$

- $\mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC}, \boldsymbol{\theta})]$

$$\propto \exp\left(-\frac{1}{2}\int \frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\hat{\delta}_g(\mathbf{k}) - \delta_{\rm model}(\boldsymbol{\theta}, \delta_{\rm IC}; \mathbf{k})|^2}{P_\epsilon}\right)$$

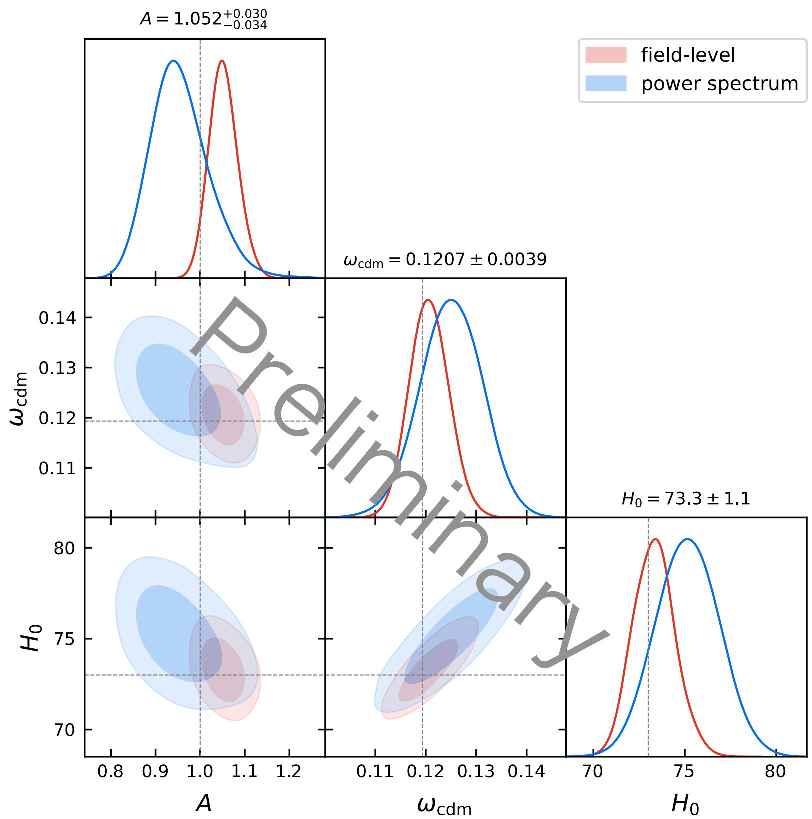
- $L = 2\ {\rm Gpc}/h,\ N_{\rm IC} = 128^3,\ k_{\rm max} = 0.12\ h/{\rm Mpc},\ z = 0.5$

- In the Gaussian case, the power spectrum should be lossless.



Non-linear (non-Gaussian) example

- Model: Zeldovich + quadratic Lagrangian bias in redshift space
  - $\delta_g(\delta_{\rm IC}, \boldsymbol{\theta}; \mathbf{k}) = \int \mathrm{d}^3\mathbf{q} \left[1 + b_1\delta_{\rm IC}(\mathbf{q}) + \frac{1}{2}b_2\delta_{\rm IC}^2(\mathbf{q}) + b_{\mathcal{G}_2}\mathcal{G}_2(\mathbf{q})\right] e^{i\mathbf{k}\cdot\mathbf{q}+\boldsymbol{\Psi}_Z(\mathbf{q})}$
  - params: $(A, \omega_{\rm cdm}, H_0, b_1, b_2, b_{\mathcal{G}_2}, c_s^2, c_1, c_2, P_\epsilon; \{\delta_{\rm IC}\})$
- $L = 2\ \mathrm{Gpc}/h,\ N_{\rm IC} = 128^3,\ k_{\rm max} = 0.12\ h/\mathrm{Mpc},\ z = 0.5$
- The mock is generated with the same model but with high resolution $N_{\rm IC} = 512^3$.
- Sampling takes about ~ a week.
- $\sigma_A^{\rm field} \sim 0.03\,, \qquad \sigma_A^{\rm power} \sim 0.07$

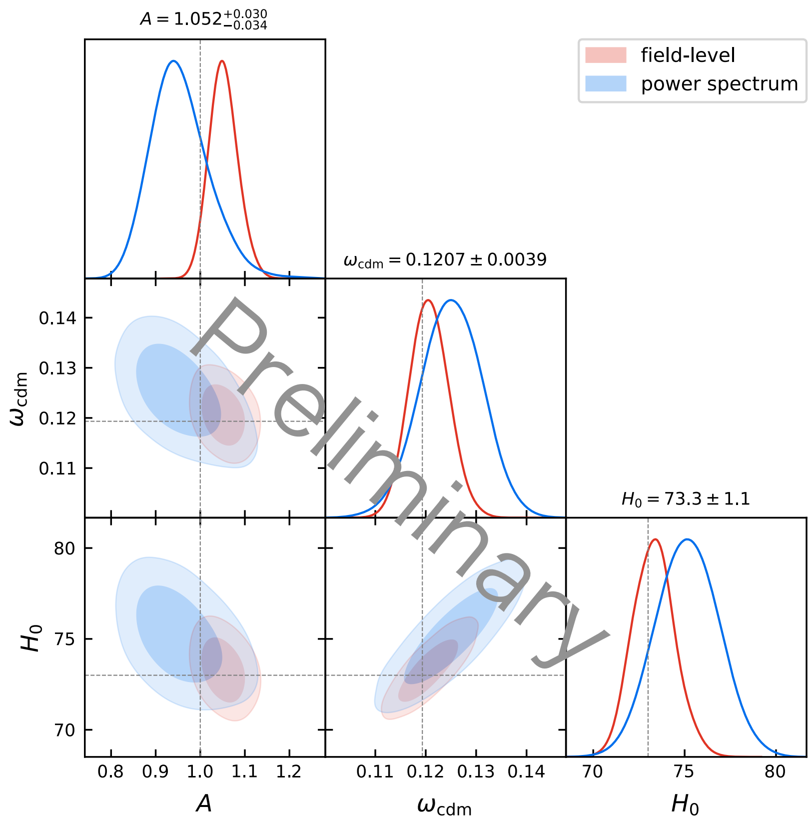


Where does the information come from?

- The mock is not “real”; The cubic and higher order bias params are all set to zero

- The linear and quadratic bias params -> P + B
  - The higher order correlation functions can be solely used for cosmology

- Not sure if this is the case in the “real” world. c.f. Nguyen+ 24



Effective shot noise from large b2

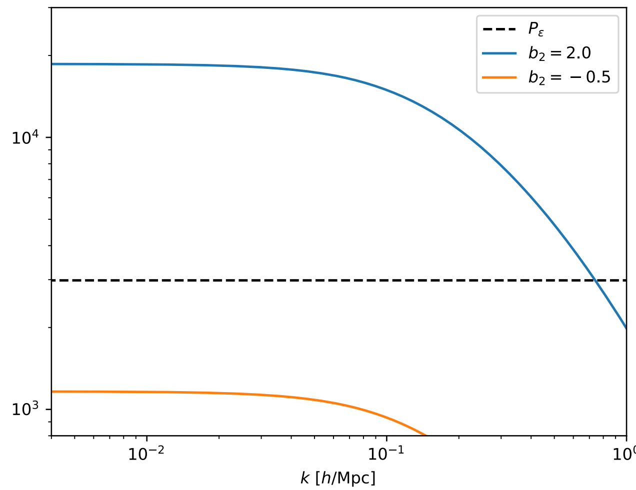
- "Renormalization of the shot noise"

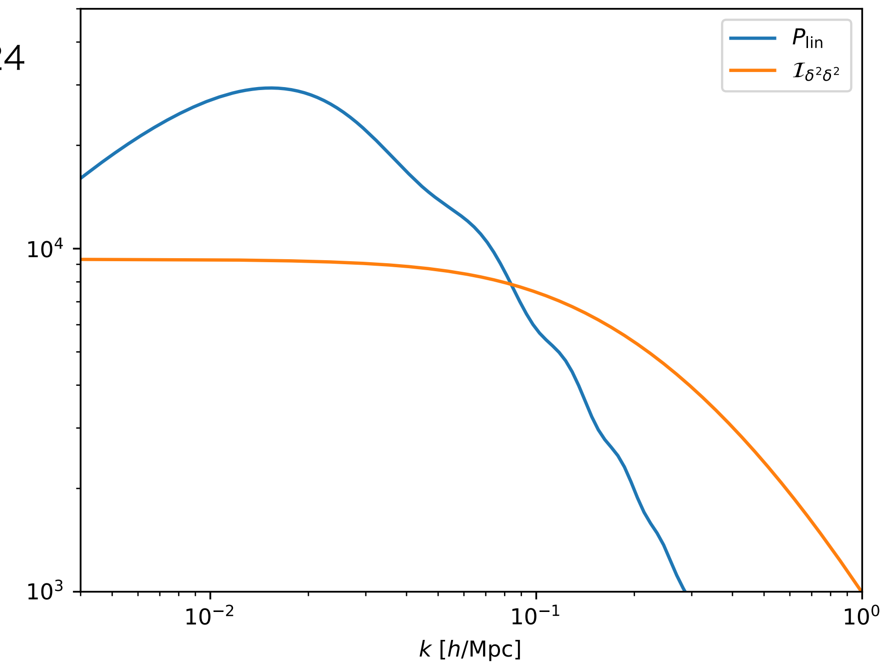
$$P_\epsilon^{\mathrm{eff}} = P_\epsilon + \frac{b_2^2}{4} \lim_{k\to 0} \mathcal{I}_{\delta^2\delta^2}(k) \qquad \mathcal{I}_{\delta^2\delta^2}(k) = \langle \delta^2 \delta^2 \rangle' = \int_{\mathbf{q}} P_{\mathrm{lin}}(\mathbf{q}) P_{\mathrm{lin}}(\mathbf{k}-\mathbf{q})$$

- Large b2 -> apparently large shot noise

Obuljen+ 23, Cabass+ 23, Foreman+ 24

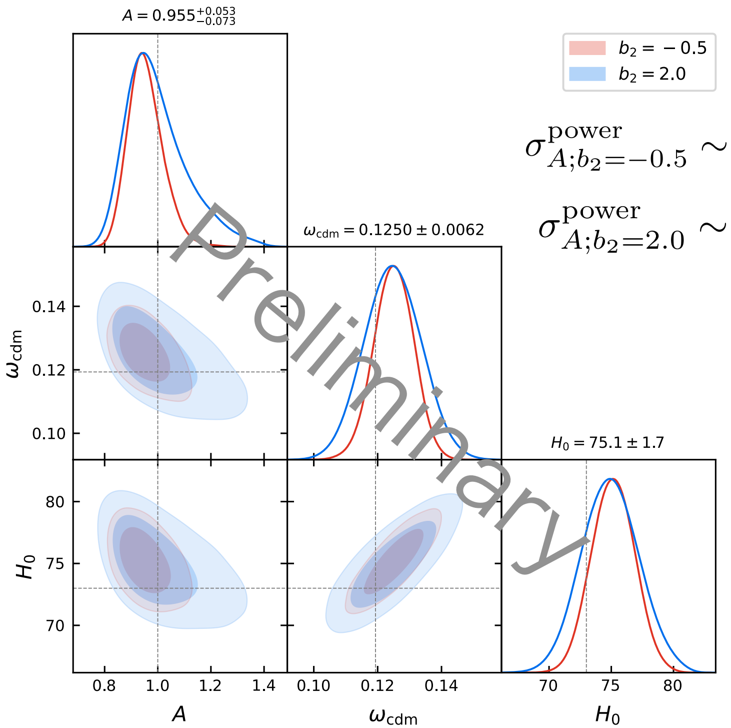
- Two mocks: $b_2 = -0.5, \quad b_2 = 2.0$





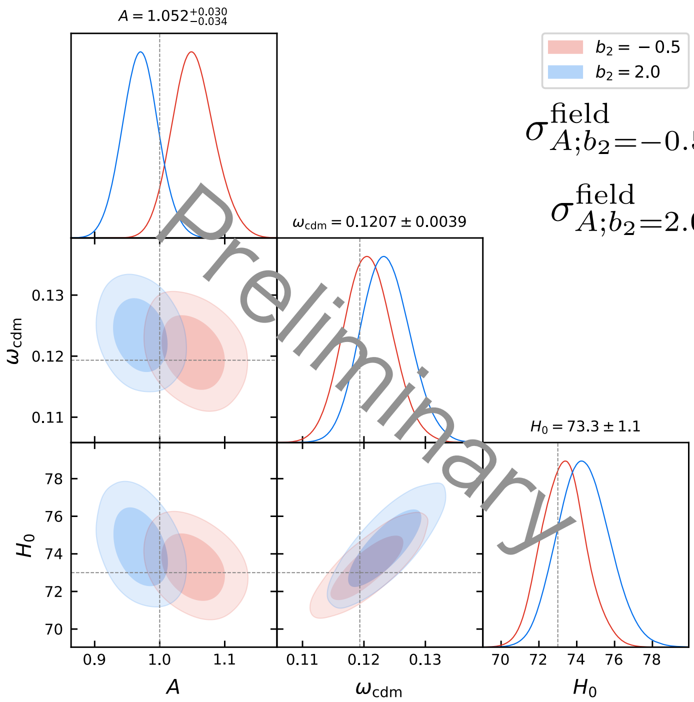
Large b2 vs small b2

- The constant piece in $\mathcal{I}_{\delta^2\delta^2}(k)$
  - Power spectrum: absorbed into the shot noise (i.e. treat it as a noise)
  - Field level: treat it as a signal (up to $k_{\rm NL}$ or $k_{\rm Ny}^{\rm IC}$)



$\sigma^{\rm power}_{A;b_2=-0.5} \sim 0.07$

$\sigma^{\rm power}_{A;b_2=2.0} \sim 0.12$

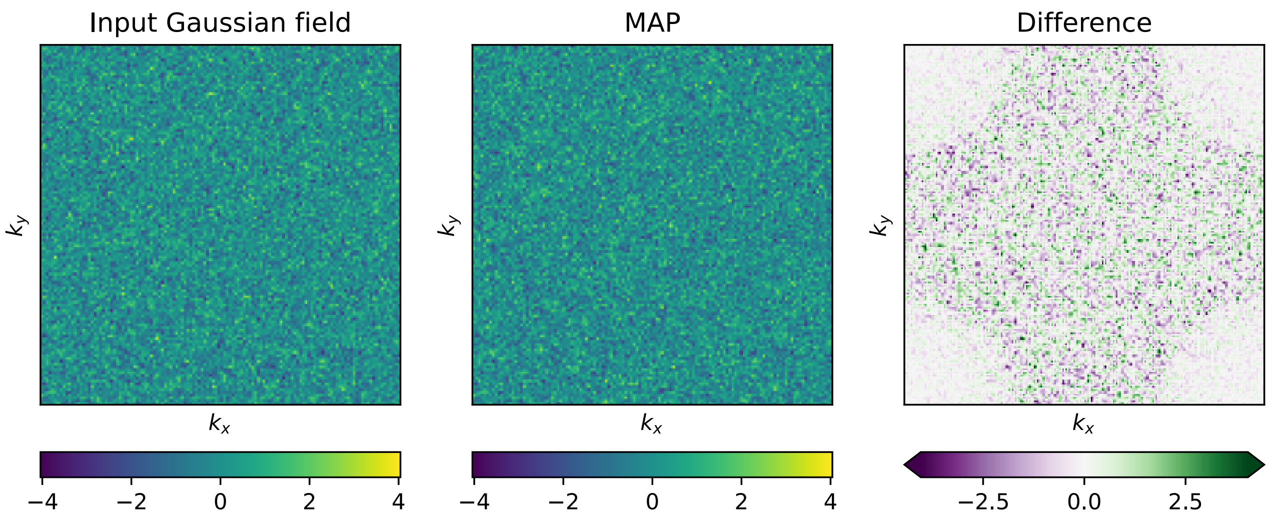


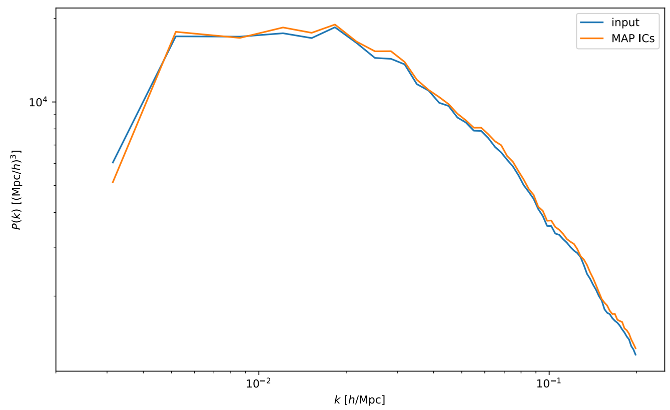
$\sigma^{\rm field}_{A;b_2=-0.5} \sim 0.03$

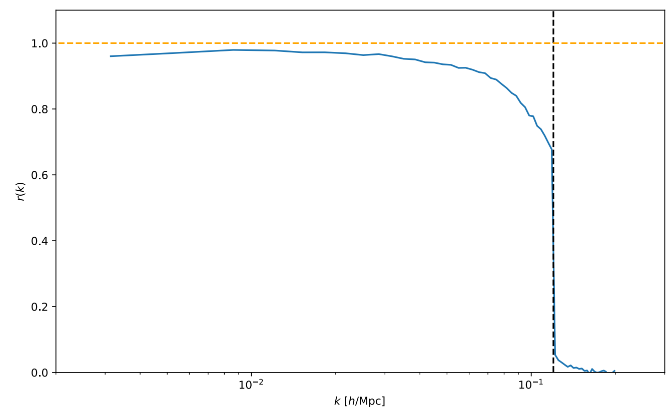
$\sigma^{\rm field}_{A;b_2=2.0} \sim 0.03$

Byproduct: the reconstruction of ICs

- The real part of the initial Gaussian field $\mathrm{Re}[\delta_{\mathrm{IC}}(\mathbf{k})]$ at $k_z = k_{\mathrm{F}}$







- Can be useful for cross-correlations

Questions

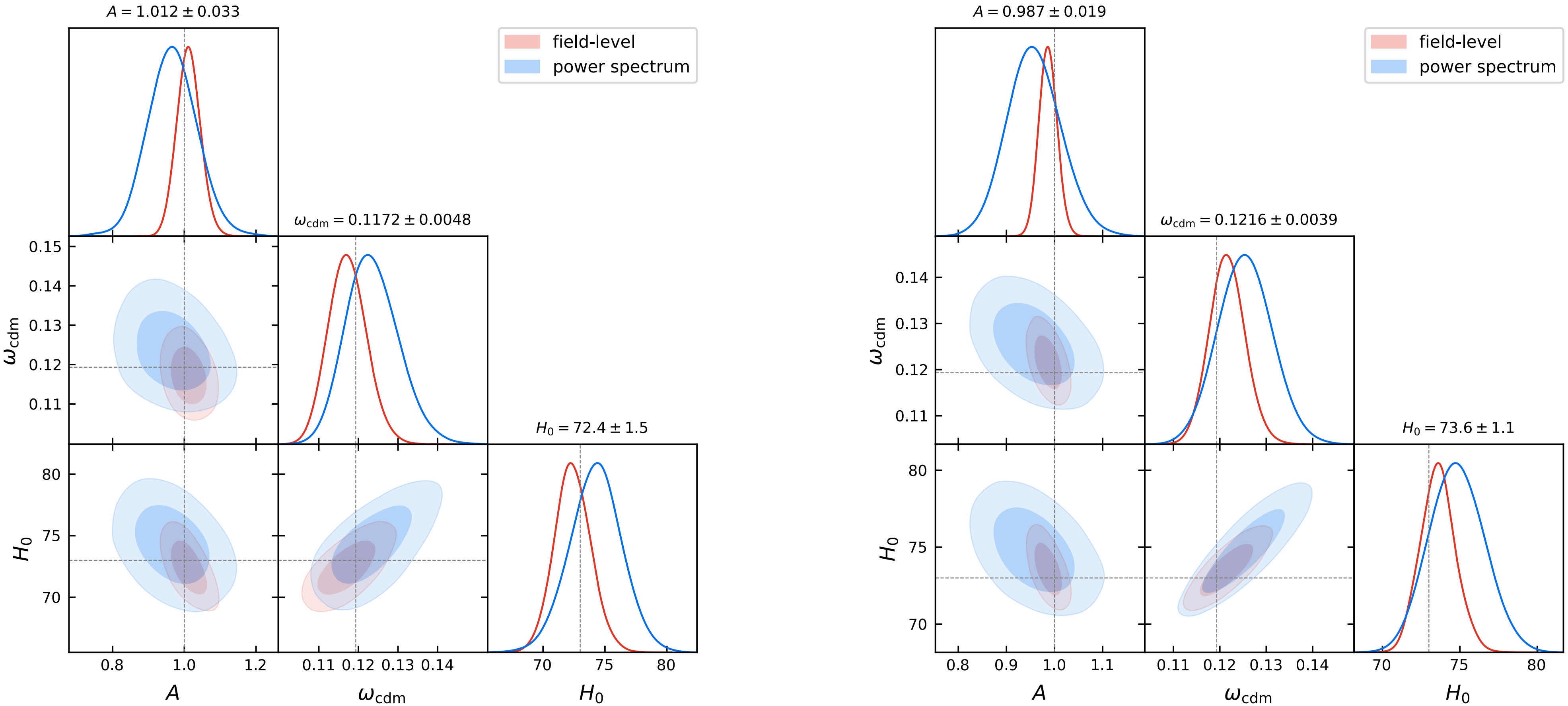
- Noise property at field-level: $\epsilon = \hat{\delta}_g - \delta_{g,\mathrm{PT\ model}}$
  - No successful example of the **full** field-level inference to the “real” N-body (without putting a prior on the noise parameter by hand)
- Likelihood for the field-level inference
  - Gaussian likelihood would not be a good approximation for realistic mock
- In general, does really the field-level outperform the n-pt function analysis? (in the perturbative regime) Cabass+ 23

Summary

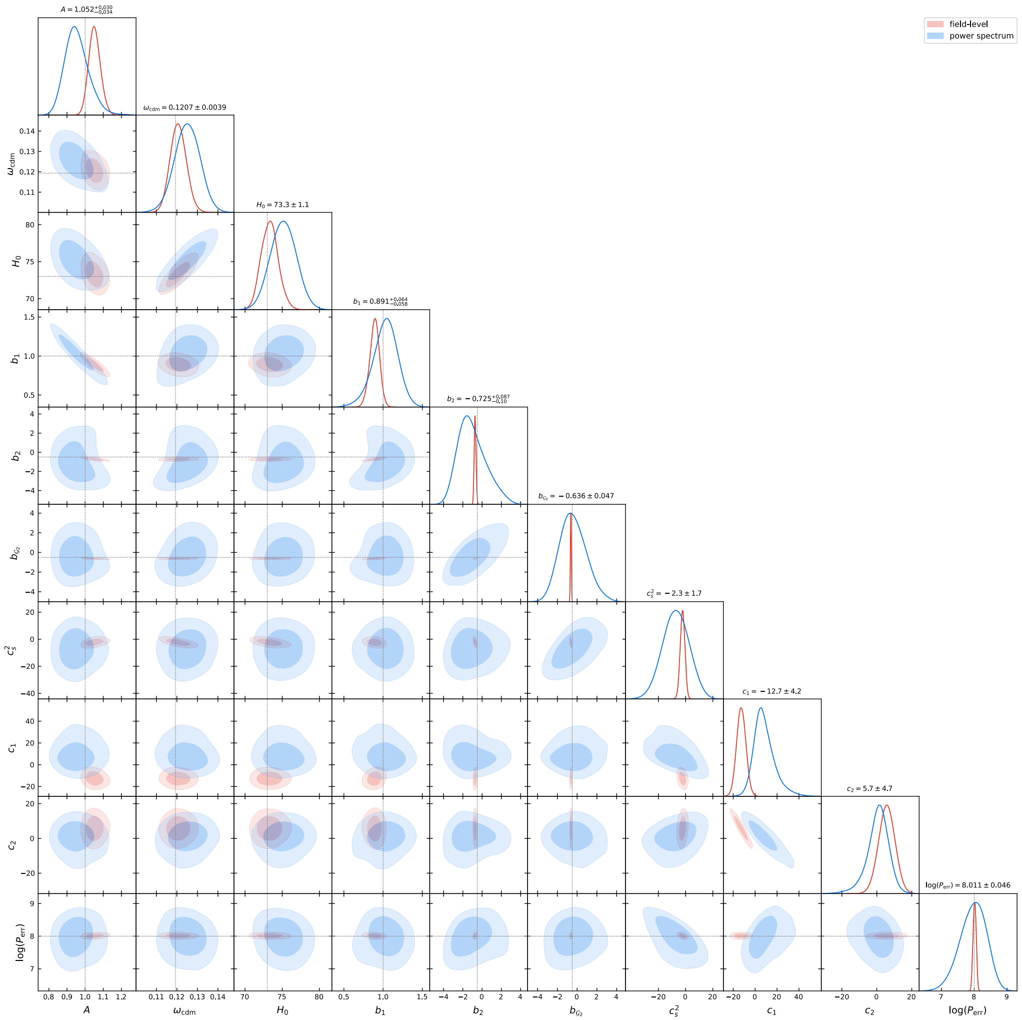
- Galaxy clustering offers a lot of cosmological information, although information is diluted to the non-Gaussian part of the galaxy density field.
  - need to go beyond the power spectrum
- Field-level inference is an optimal (i.e. lossless) analysis method to extract cosmology from non-Gaussian random field
  - Differentiable forward model at field-level & Hamiltonian Monte Carlo
  - If we can use the information in higher-order terms/correlation function only for cosmology, the field-level gives better results (as expected)
  - For the large b2 case, the field-level can be better than the power spectrum, since the part of the constant piece in $\langle \delta^2 \delta^2 \rangle$ can be inferred.

Backups

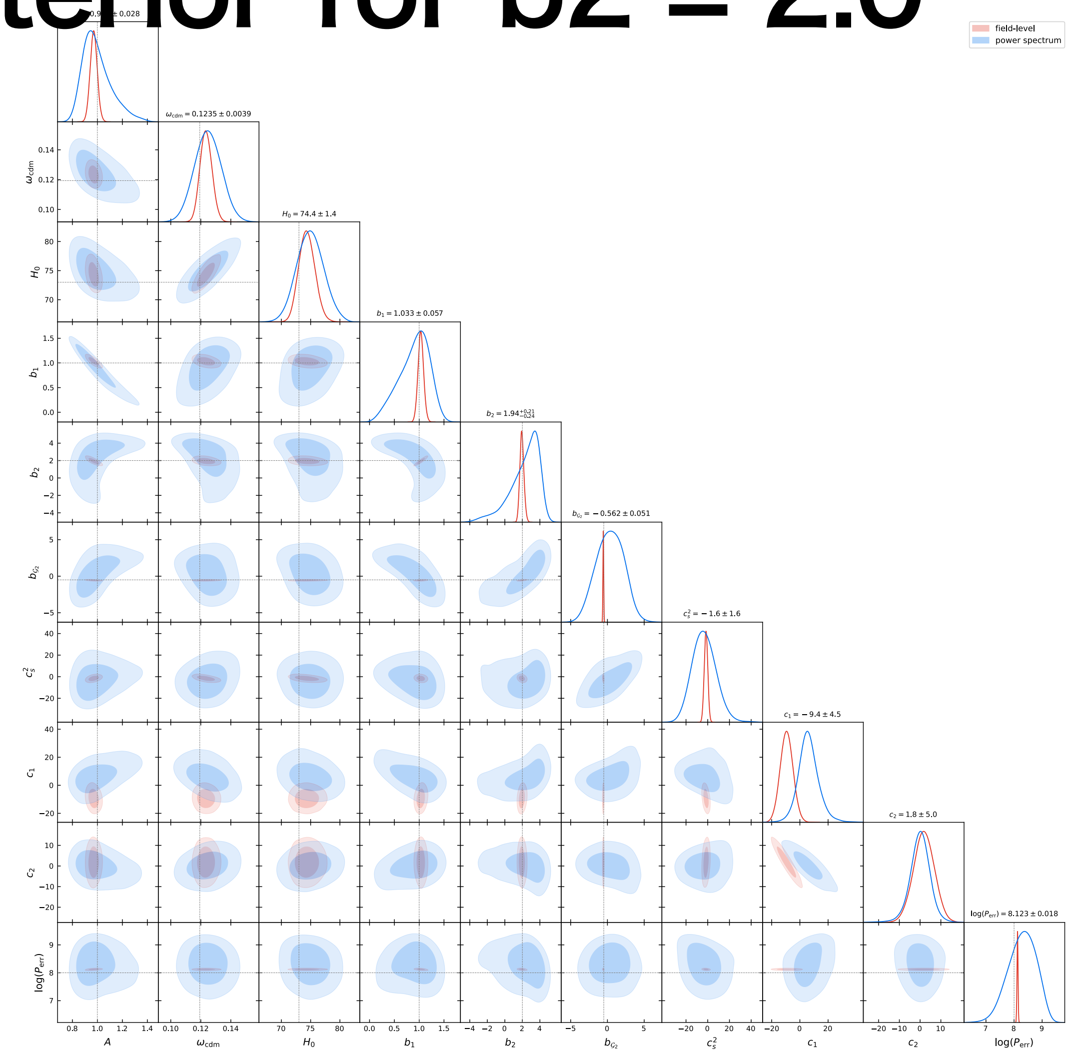
Zeldovich + linear bias



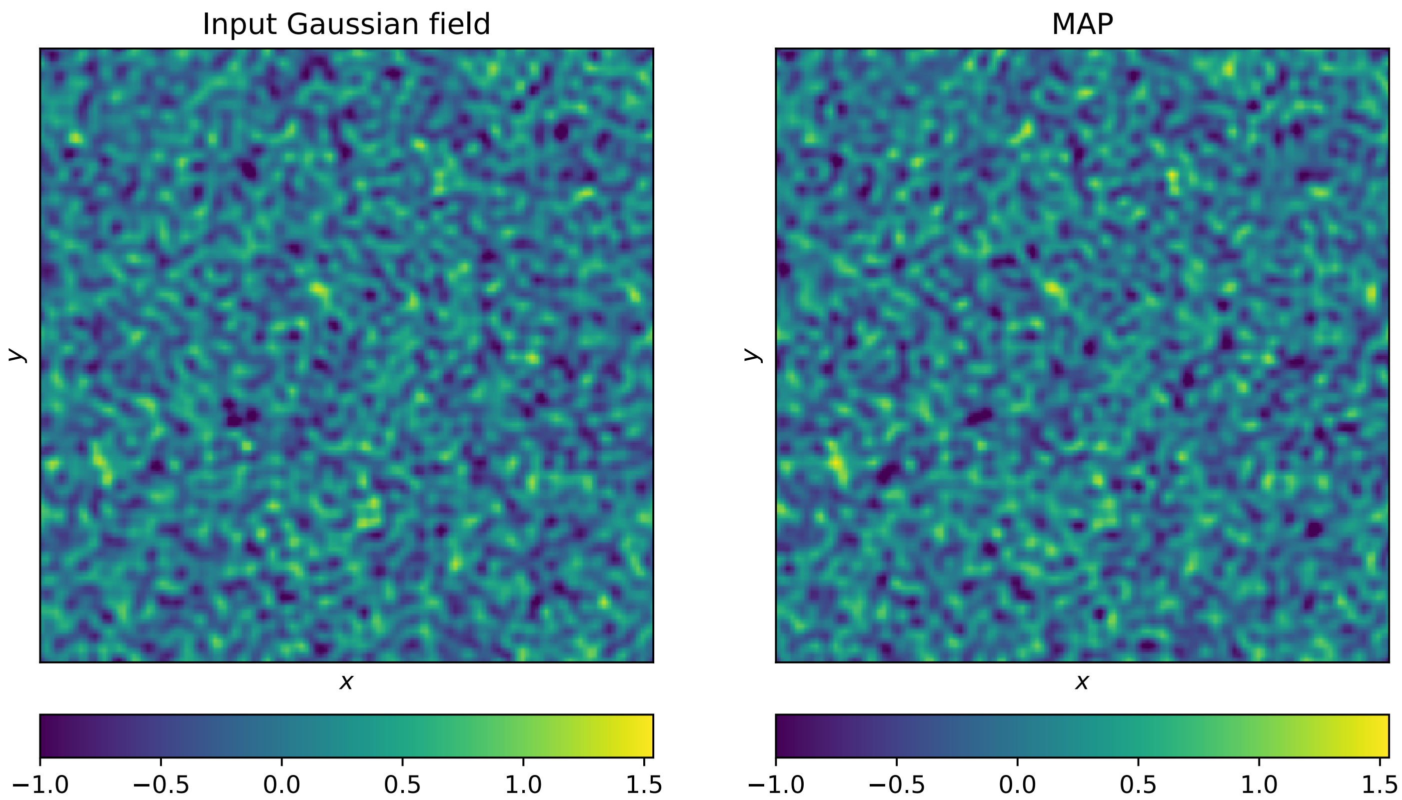
Full posterior for b2 = -0.5



Full posterior for b2 = 2.0



Reconstructed ICs in position space



Equivalence of Fourier and configuration space likelihood

- Parseval's theorem:
  - $\mathcal{L}[\hat{\delta}_g|\delta_{\rm model}(\delta_{\rm IC},\boldsymbol{\theta})]$

$$\propto \exp\left(-\frac{1}{2}\int\frac{{\rm d}^3\mathbf{k}}{(2\pi)^3}\frac{|\hat{\delta}_g(\mathbf{k})-\delta_{\rm model}(\boldsymbol{\theta},\delta_{\rm IC};\mathbf{k})|^2}{P_\epsilon}\right)$$

$$=\exp\left(-\frac{1}{2}\int{\rm d}^3\mathbf{x}\,\frac{|\hat{\delta}_g(\mathbf{x})-\delta_{\rm model}(\boldsymbol{\theta},\delta_{\rm IC};\mathbf{x})|^2}{P_\epsilon}\right)$$

- Configuration-space likelihood would be useful for e.g. the density-dependent noise