Bayesian inference of the initial conditions from large-scale structure surveys

Florent Leclercq

Institut d'Astrophysique de Paris
Institut Lagrange de Paris
École polytechnique ParisTech

March 14th, 2014

In collaboration with:

Jacopo Chevallard (U. São Paulo), Héctor Gil-Marín (U. Portsmouth/U. Barcelona), Nico Hamaus (IAP), Jens Jasche (IAP), Alice Pisani (IAP), Emilio Romano-Díaz (U. Bonn), Paul M. Sutter (IAP/Ohio State U.), Svetlin Tassev (U. Princeton), Benjamin Wandelt (IAP/U. Illinois), Matías Zaldarriaga (IAS Princeton)

Some specificities of cosmology

- **Unicity**. The experience is unique and irreproducible by physical experimentation. There is no exteriority nor anteriority. The properties of the Universe cannot be determined statistically on a set.
- **Energy**. The energy scales at stake in the Early Universe are orders of magnitude higher than anything we can reach on Earth.
- **Arrow of time**. Reasoning in cosmology is "bottom-up". The final state is known and the initial state has to be inferred.

➡ The **initial conditions** of the Universe have a **particular status** with respect to other physical phenomena.

Cosmostatistics of the initial conditions

- **"Initial conditions"**: ICs for *gravitational evolution*…
  - AFTER inflation
  - AFTER Hot Big Bang phenomena

(*primordial nucleosynthesis, decoupling, recombination, free-streaming of neutrinos, acoustic oscillations of the photon-baryon plasma, transition from radiation to matter dominated universe…)*

- **Cosmostatistics**: discipline dealing with stochastic quantities as seeds of structure in the Universe
  - prediction of cosmological observables from random inputs

    (from theory to data)
  - use of the departures from homogeneity in astronomical surveys to distinguish between cosmological models

    (from data to theory)

see also FL, Pisani & Wandelt, arXiv:1403.1260

Bayesian inference of the ICs

- Why do we need Bayesian inference?

Inference of signals = ill-posed problem

- Noise
- Incomplete observations: survey geometry, selection effects
- Systematic uncertainties, biases
- Cosmic variance

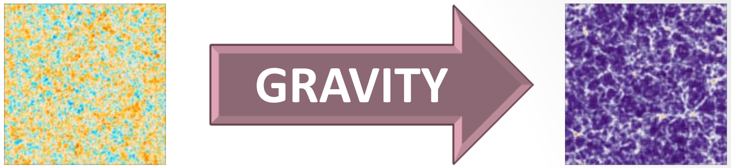
➡ **No unique recovery is possible!**

"What are the initial conditions of the Universe?" ➡ "What is the probability distribution of possible initial conditions (signals) compatible with the observations?"

$$p(s|d)p(d) = p(d|s)p(s)$$

Bayesian inference of the ICs

- Physical motivation:
  - Complex final state, simple initial state
  - A "forward only" problem (we have a generative model for the final state)



Initial state

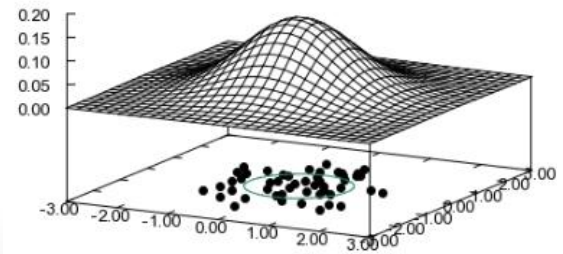
Final state

- Problems:
  - Highly dimensional inference (10^7 parameters)
  - A large number of correlated parameters

→ No reduction of the problem size is possible!

  - Potentially complex posterior distribution

- Numerical approximation: sampling the posterior



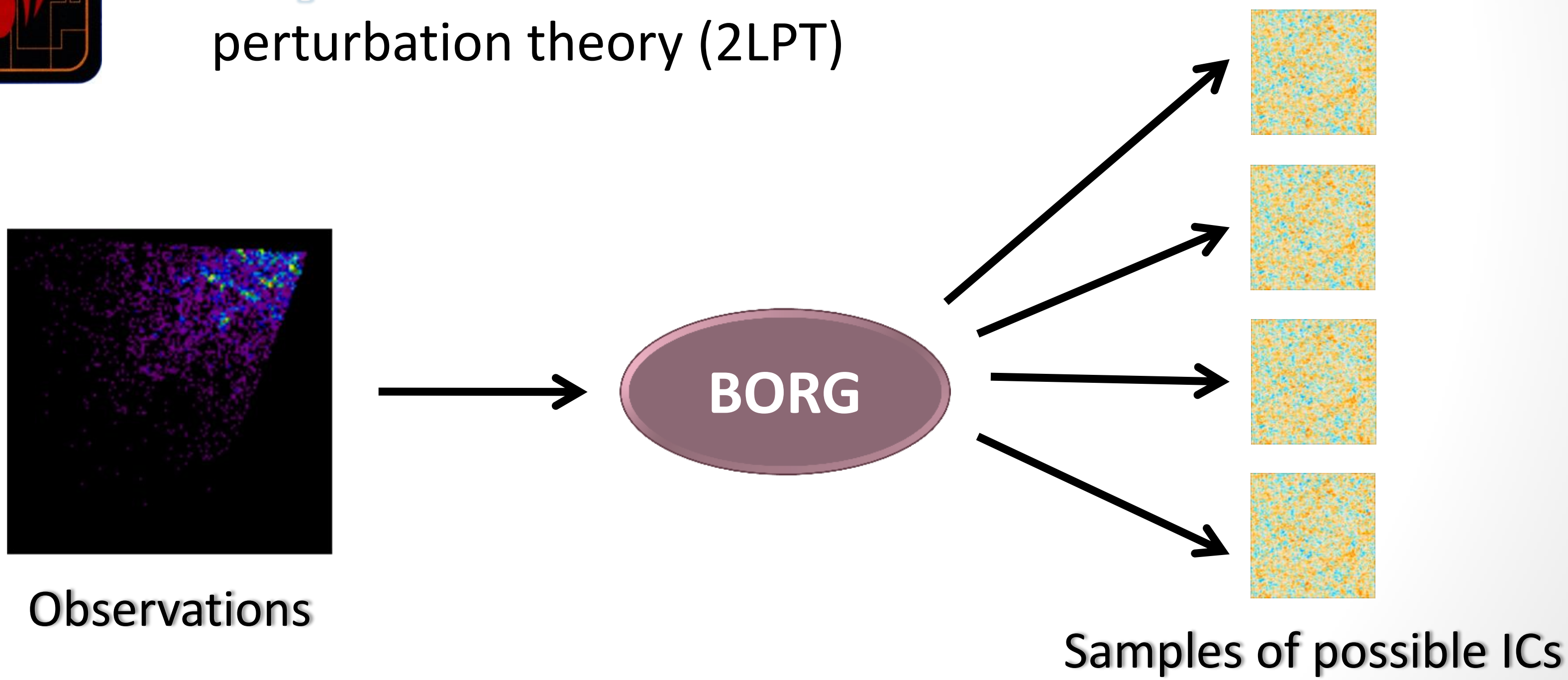
$$p(s|d) \to p_N(s|d) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\mathrm{D}}(s - s_i)$$

- But how to "get the dots" ?

BORG: *Bayesian Origin Reconstruction from Galaxies*

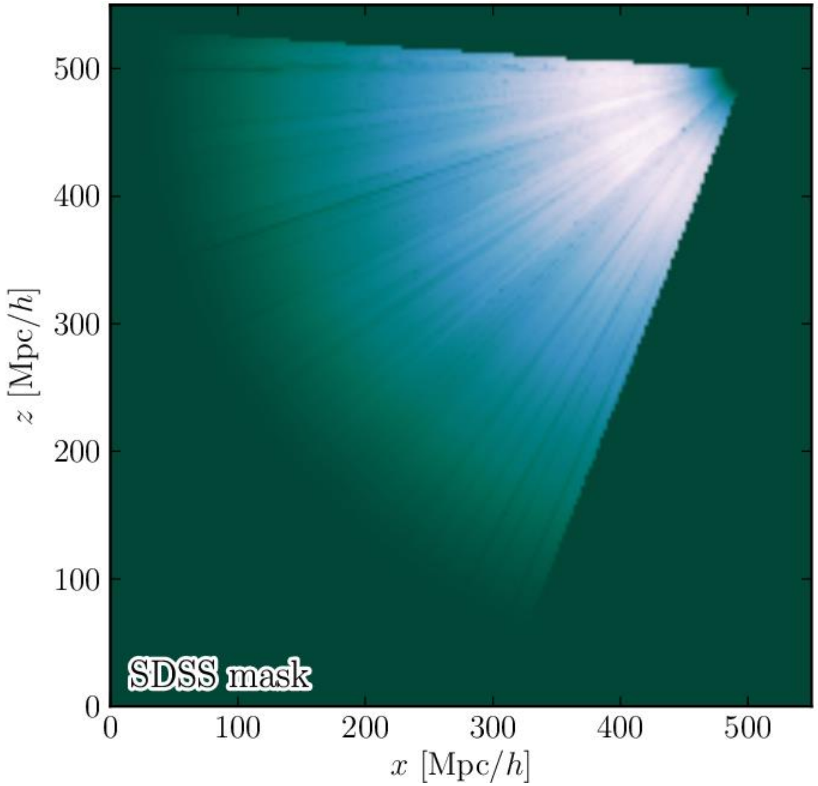
What makes the problem tractable:

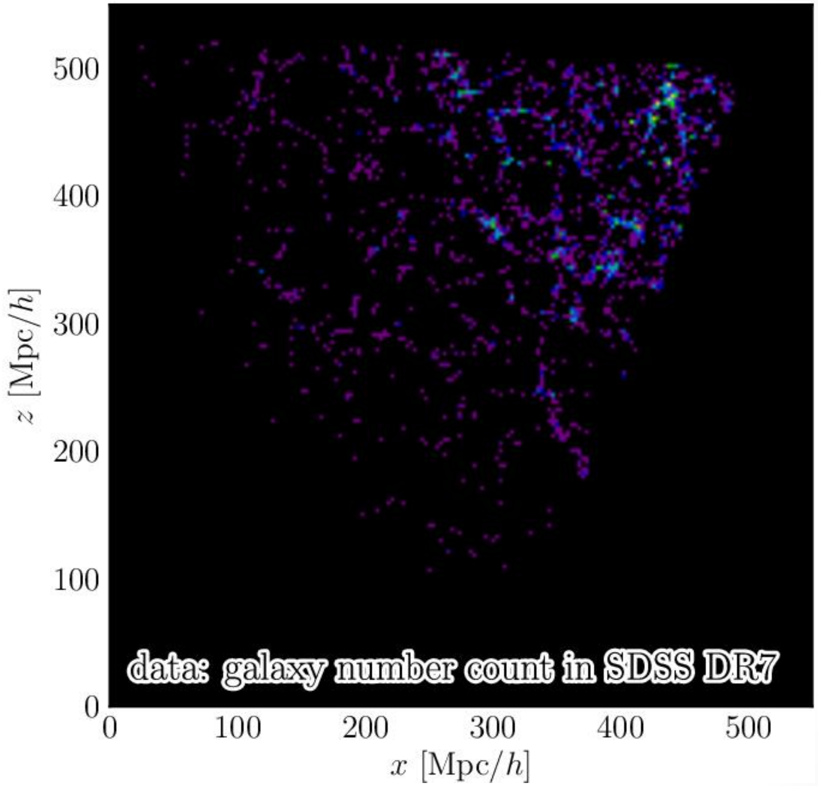
- **Sampler**: Hamiltonian Markov Chain Monte Carlo method
- **Physical model**: Second-order Lagrangian perturbation theory (2LPT)



Jasche & Wandelt 2012, arXiv:1203.3639

BORG: reconstructions from SDSS DR7

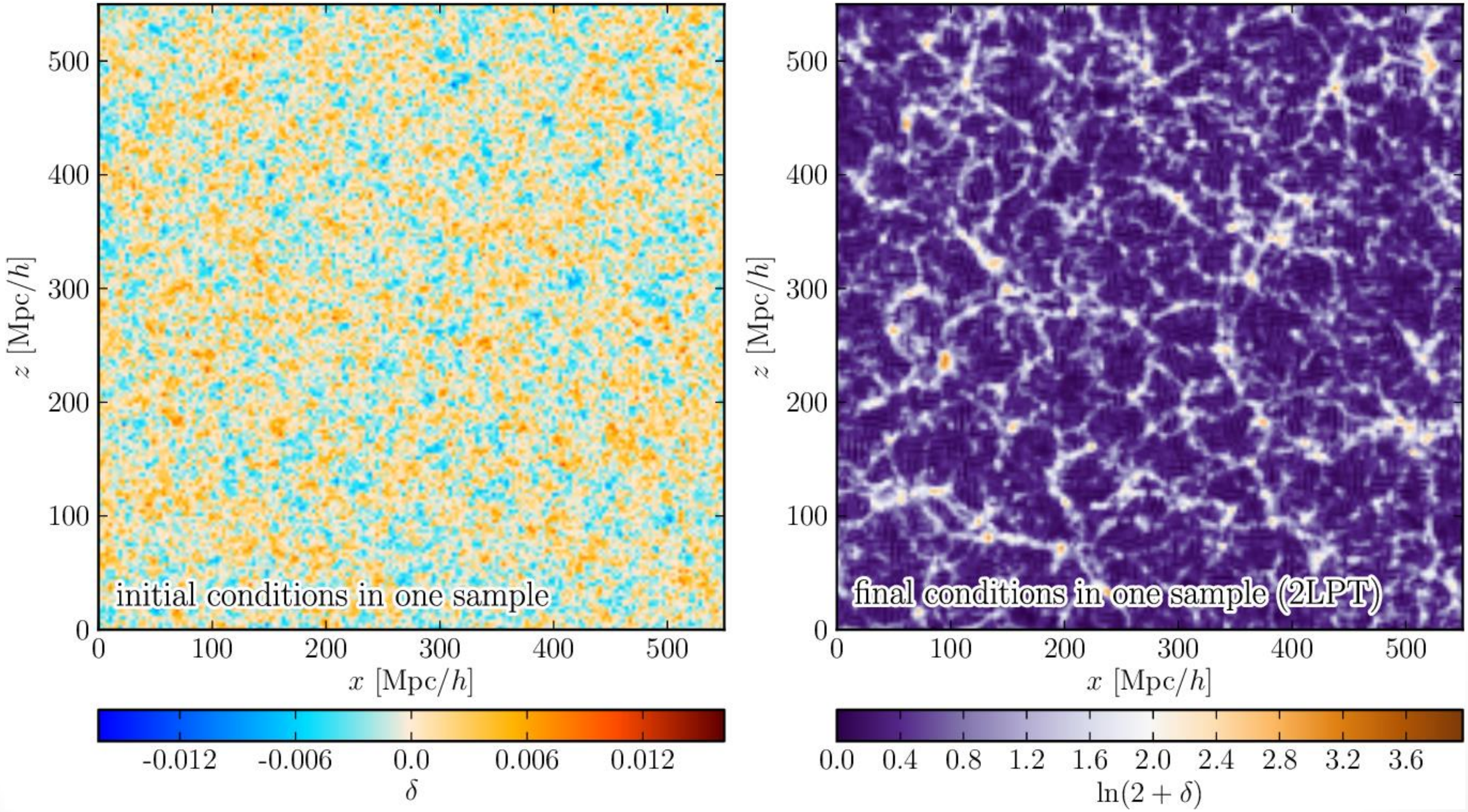




Data

Jasche, FL & Wandelt, in prep.

BORG: reconstructions from SDSS DR7



One sample

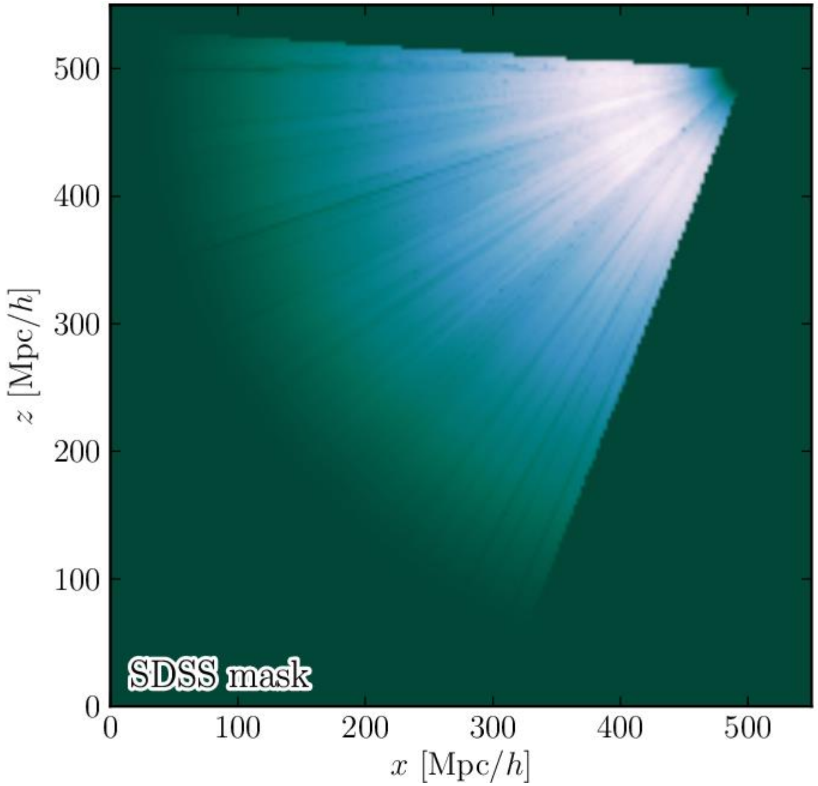
Jasche, FL & Wandelt, in prep.

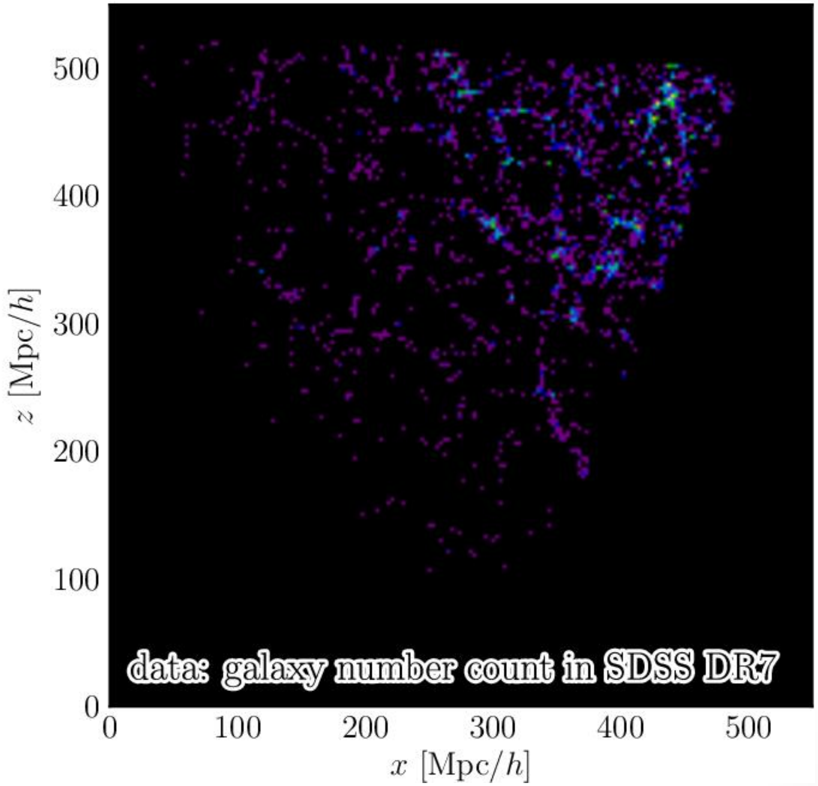
Samples of the posterior density

- Each sample: a possible version of the truth
- The variation between samples quantifies the uncertainty that results from having
  - only one Universe (a more precise version of “cosmic variance”)
  - incomplete observations (mask, finite volume and number of galaxies, selection effects)
  - imperfect data (noise, biases, photometric redshifts…)

see also FL, Pisani & Wandelt, arXiv:1403.1260

BORG: reconstructions from SDSS DR7

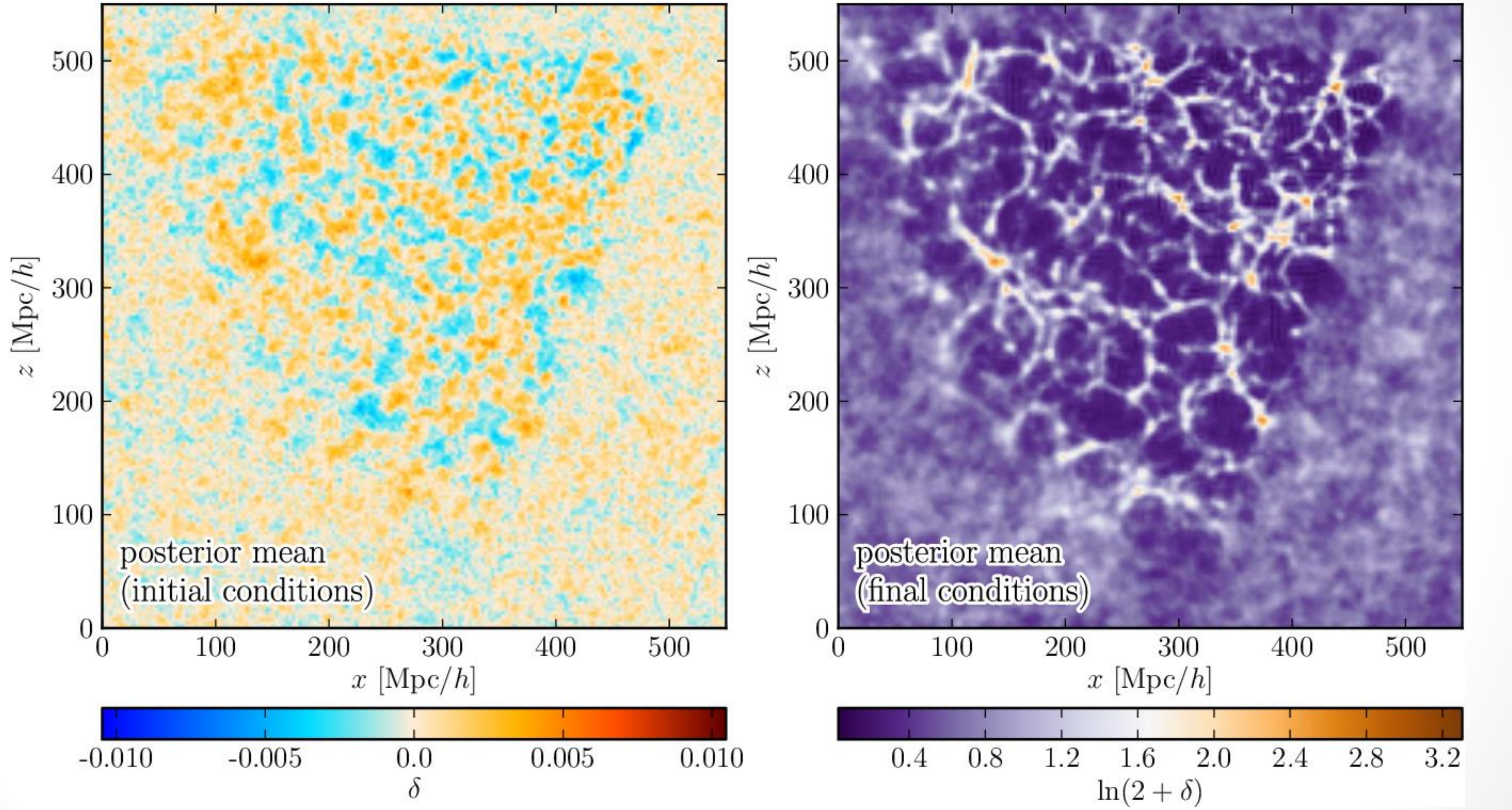




Data

Jasche, FL & Wandelt, in prep.

BORG: reconstructions from SDSS DR7



Posterior mean

Jasche, FL & Wandelt, in prep.

Beyond 2LPT

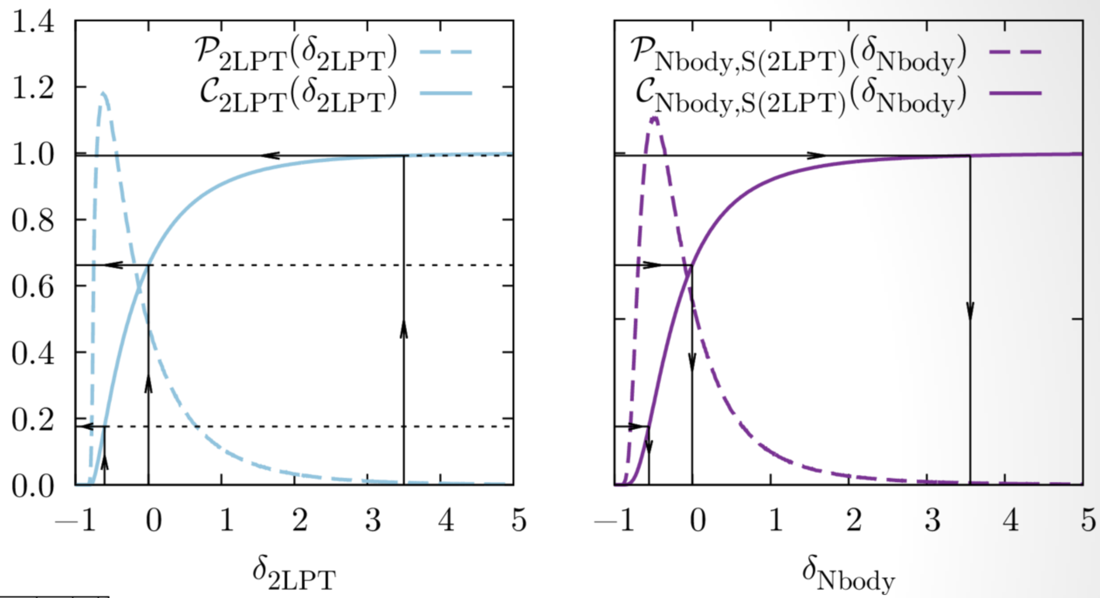
- 2LPT breaks down at small scales.
- The number of usable modes goes like k^3.
- Even small improvements yield a wealth of yet unexploited cosmological information (in existing surveys!).
- We need numerically efficient and flexible tools to model cosmic structure formation in the non-linear regime.

FL, Jasche, Gil-Marín & Wandelt 2013, arXiv:1305.4642
Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322

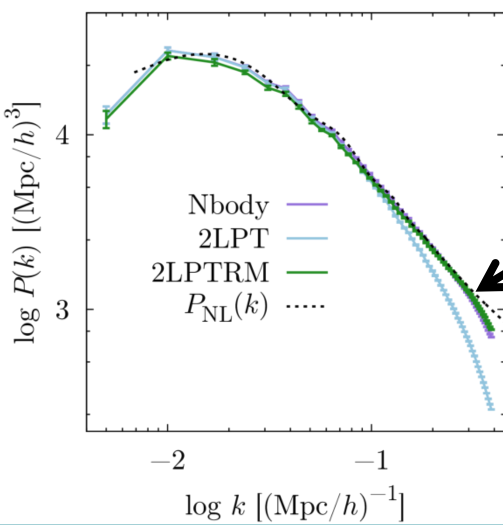
Remapping 2LPT in the mildly non-linear regime

FL, Jasche, Gil-Marín & Wandelt 2013, arXiv:1305.4642

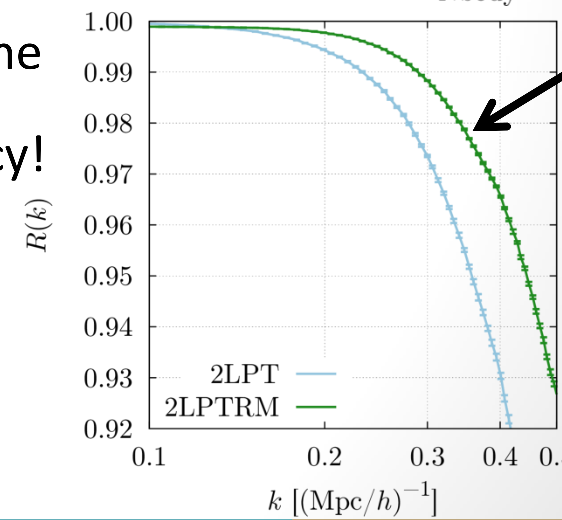
- Replacing the one-point distribution of 2LPT by one which accounts for the full non-linear system…



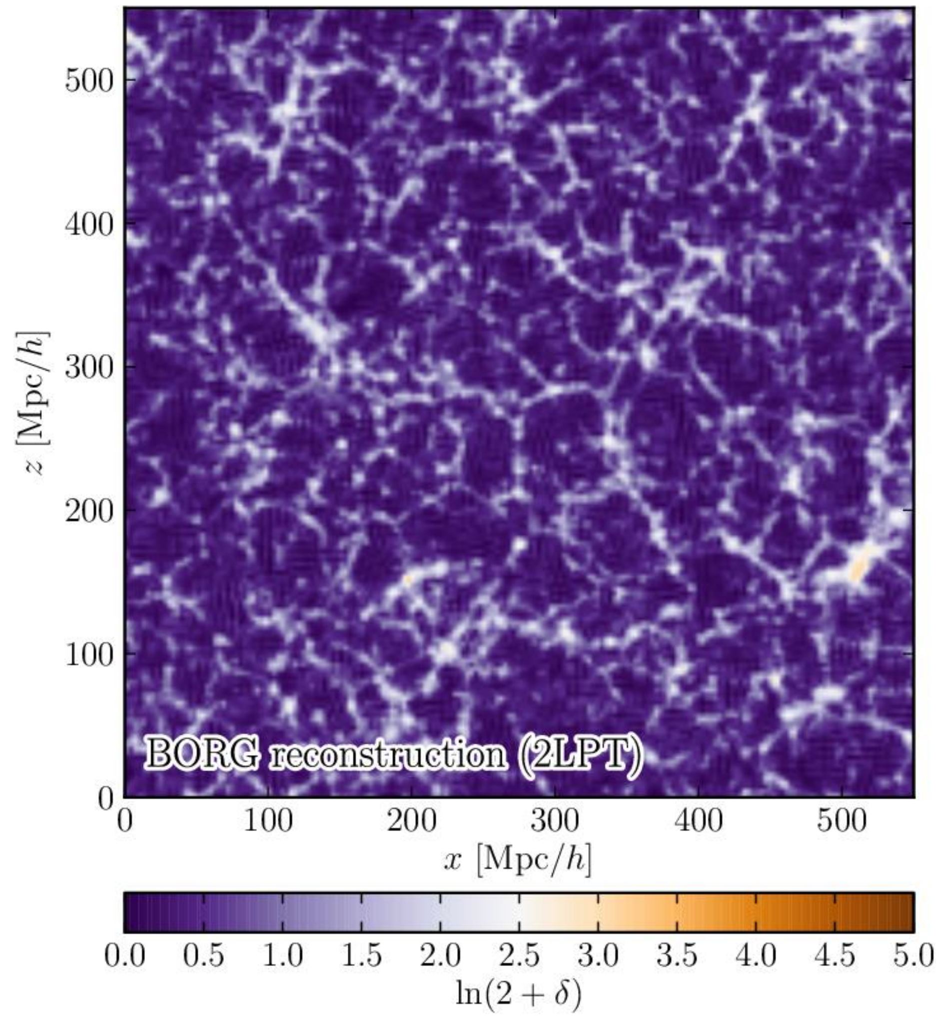
- …also improves the higher-order correlators…



- …and the phase accuracy!

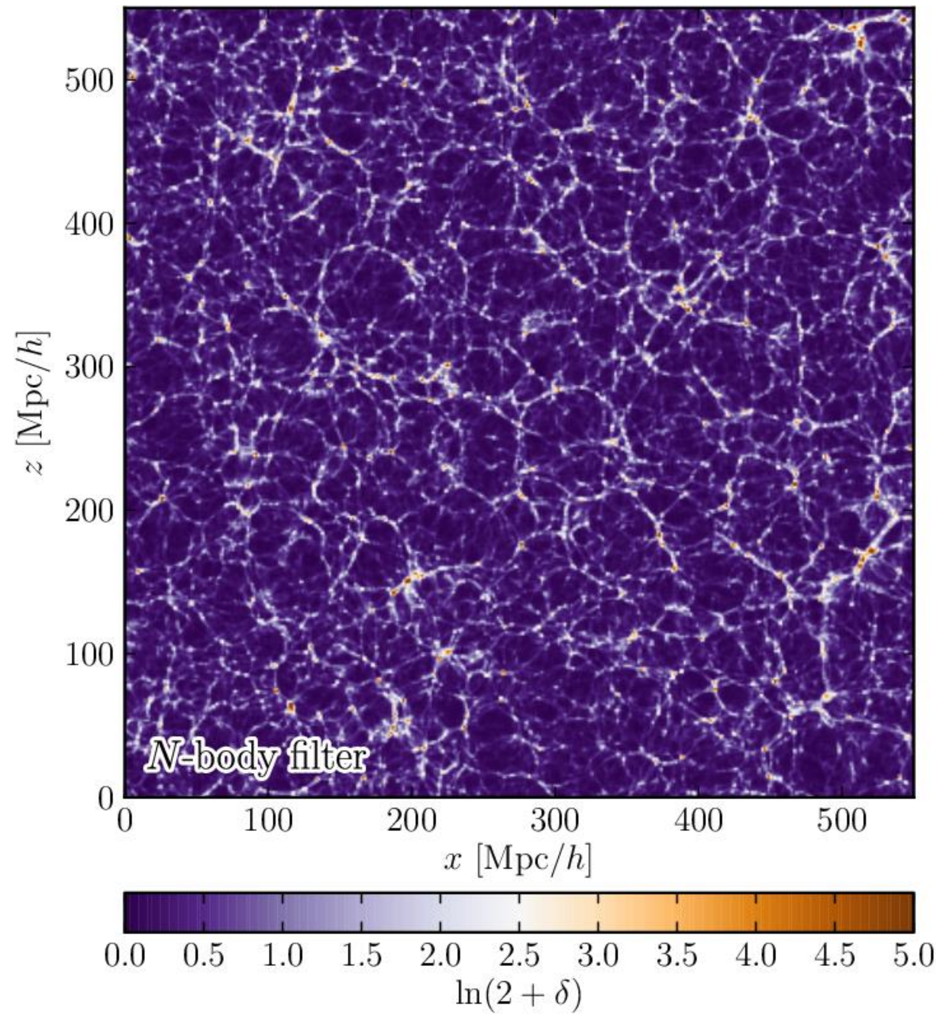


Non-linear constrained realizations



Jasche, FL, Romano-Diaz & Wandelt, in prep.

Non-linear constrained realizations

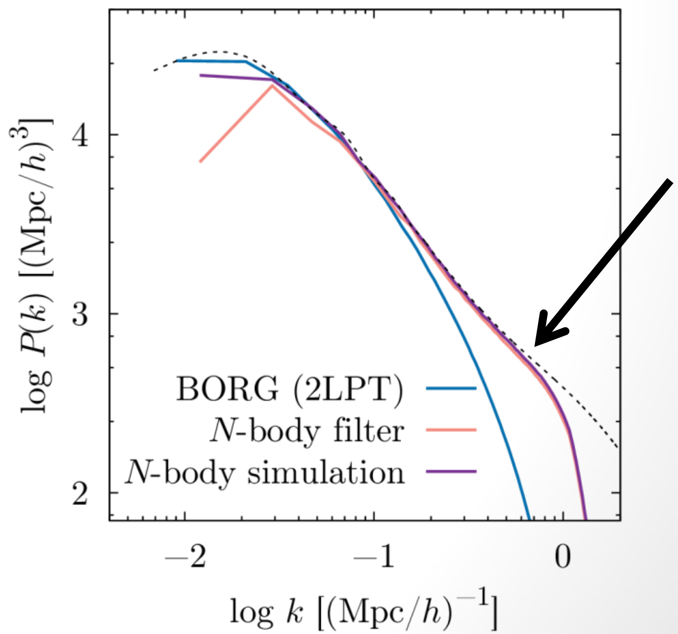


Jasche, FL, Romano-Diaz & Wandelt, in prep.

Non-linear constrained realizations

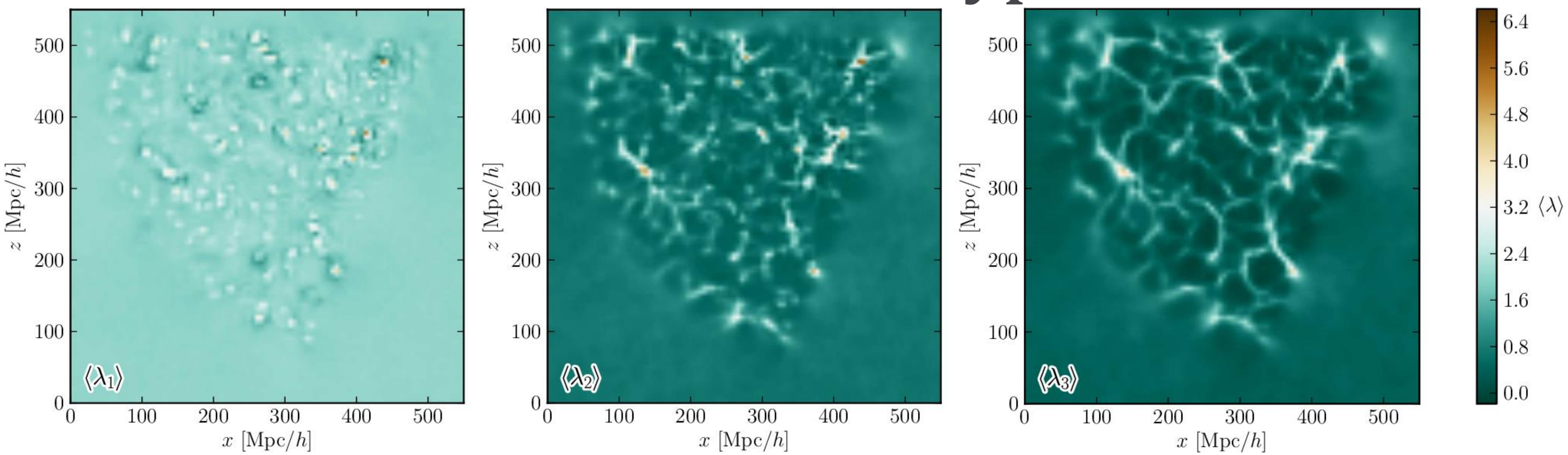
- A dynamic physical model naturally introduces some **correlations** between the constrained and unconstrained parts
- Constrained resimulations act as **hypothesis generating machines**, whose predictions can be tested with complementary observations in the actual sky.

- With a full N-body simulation, we address the **non-linear regime** of structure formation!



Jasche, FL, Romano-Diaz & Wandelt, in prep.

Identification of structure types



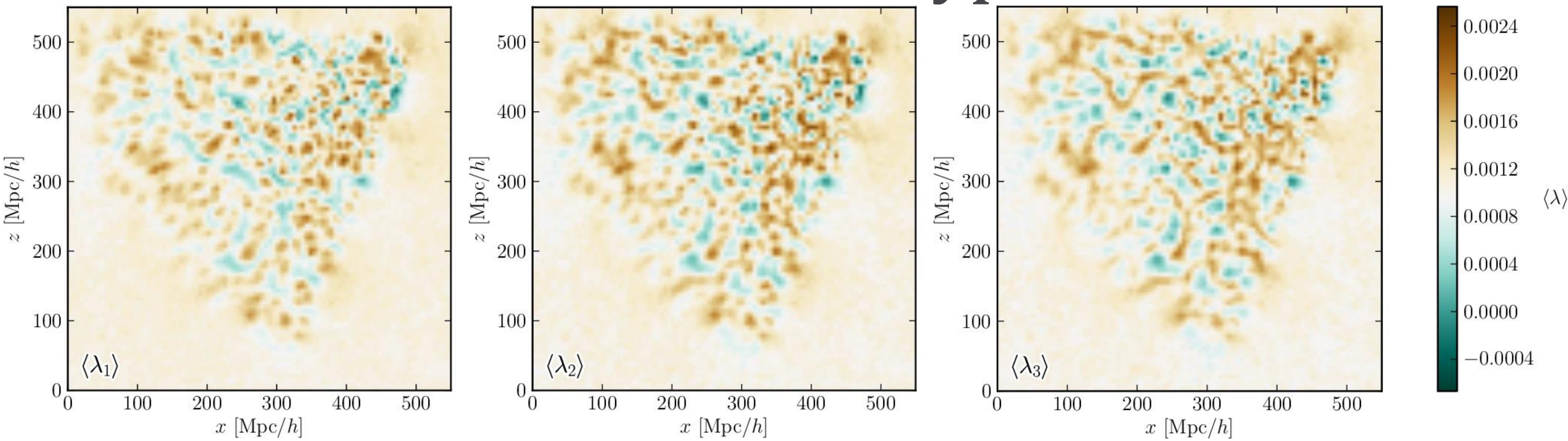
FL, Jasche, Chevallard & Wandelt, in prep.

Final conditions

- Classification of structures based on tidal shear analysis
- $\lambda_1, \lambda_2, \lambda_3$: eigenvalues of the Hessian of the gravitational potential
  - Voids: $\lambda_1, \lambda_2, \lambda_3 < 0$
  - Sheets: $\lambda_1 > 0 \text{ and } \lambda_2, \lambda_3 < 0$
  - Filaments: $\lambda_1, \lambda_2 > 0 \text{ and } \lambda_3 < 0$
  - Clusters: $\lambda_1, \lambda_2, \lambda_3 > 0$

Hahn, Porciani, Carollo & Dekel, 2006, arXiv:astro-ph/0610280

Identification of structure types



FL, Jasche, Chevallard & Wandelt, in prep.

Initial conditions

Preliminary

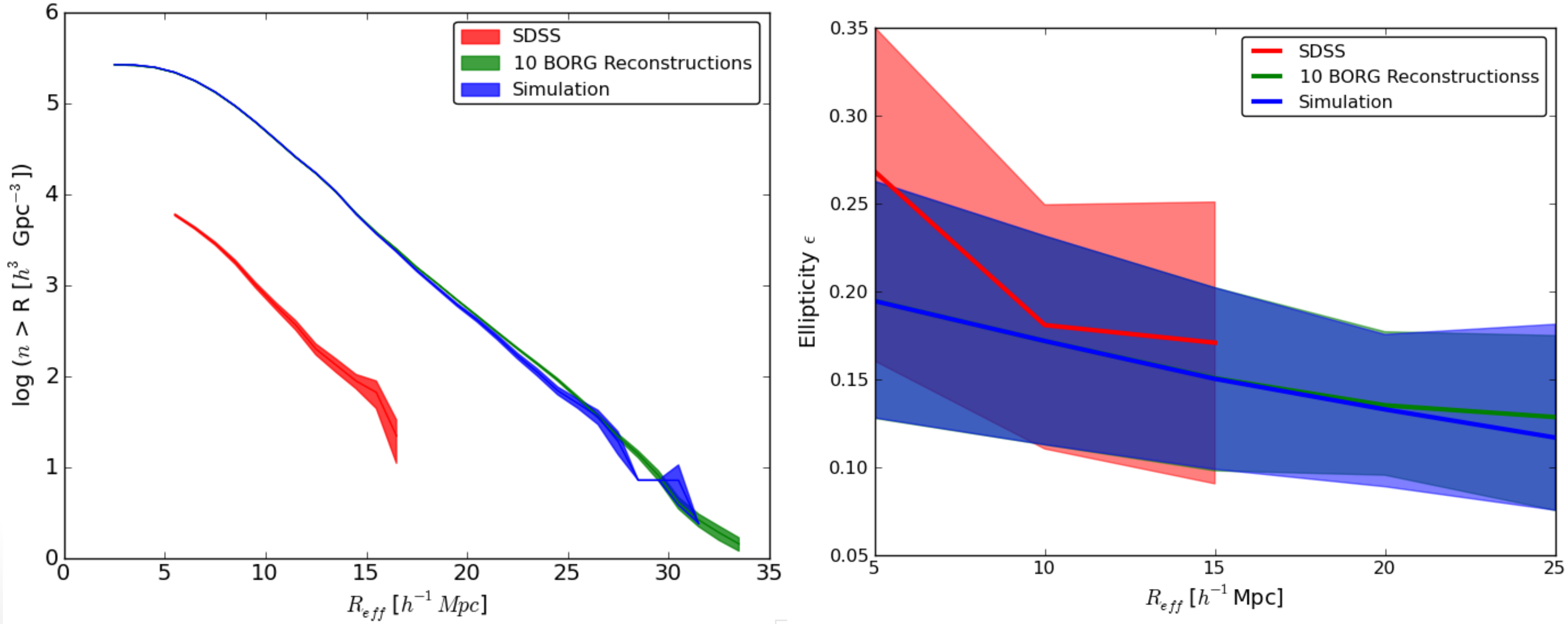
- Correlation of the large-scale environment (density field, tidal shear field) with galaxy properties reveals interesting correlations.

Dark matter voids in the SDSS galaxies

FL, Jasche, Sutter, Hamaus & Wandelt, in prep.

Preliminary

- Made possible by our technology:
  - Bias. Voids are defined in the dark matter distribution, not in galaxies.
  - Noise. Galaxy voids are Poisson noise-dominated. We get 10x more.
- Properties consistent with *N*-body simulations.



Concluding thoughts

- BORG: A non-linear time machine using Bayesian posterior exploration to infer primordial quantities from late-time observations.
- Cosmological physical reconstruction of the initial conditions of the Universe is becoming feasible.
- Need for efficient tools to model cosmic structure formation the non-linear regime.
- Great science is waiting behind the door.