Cosmostatistics: the initial conditions and the large-scale structure of the Universe

Florent Leclercq

Institut d'Astrophysique de Paris & École polytechnique ParisTech

November 20th, 2013

In collaboration with:

Héctor Gil-Marín (U. Portsmouth/U. Barcelona), Jens Jasche (IAP),
Emilio Romano-Díaz (U. Bonn), 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 infered.

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

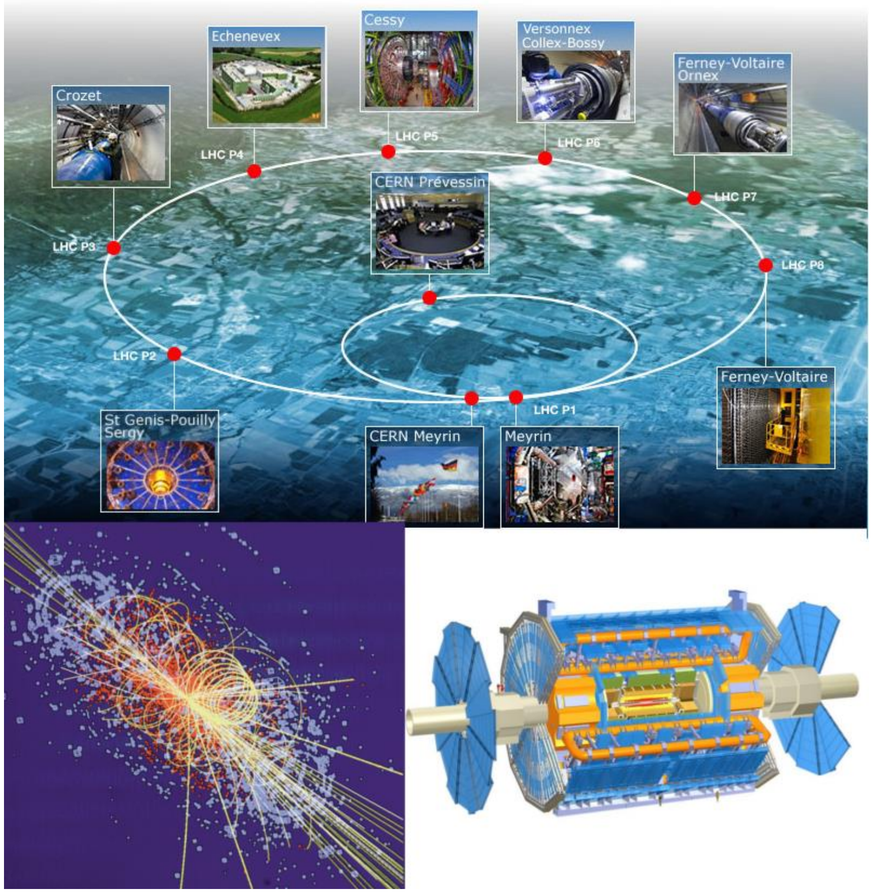
Cosmostatistics of the initial conditions

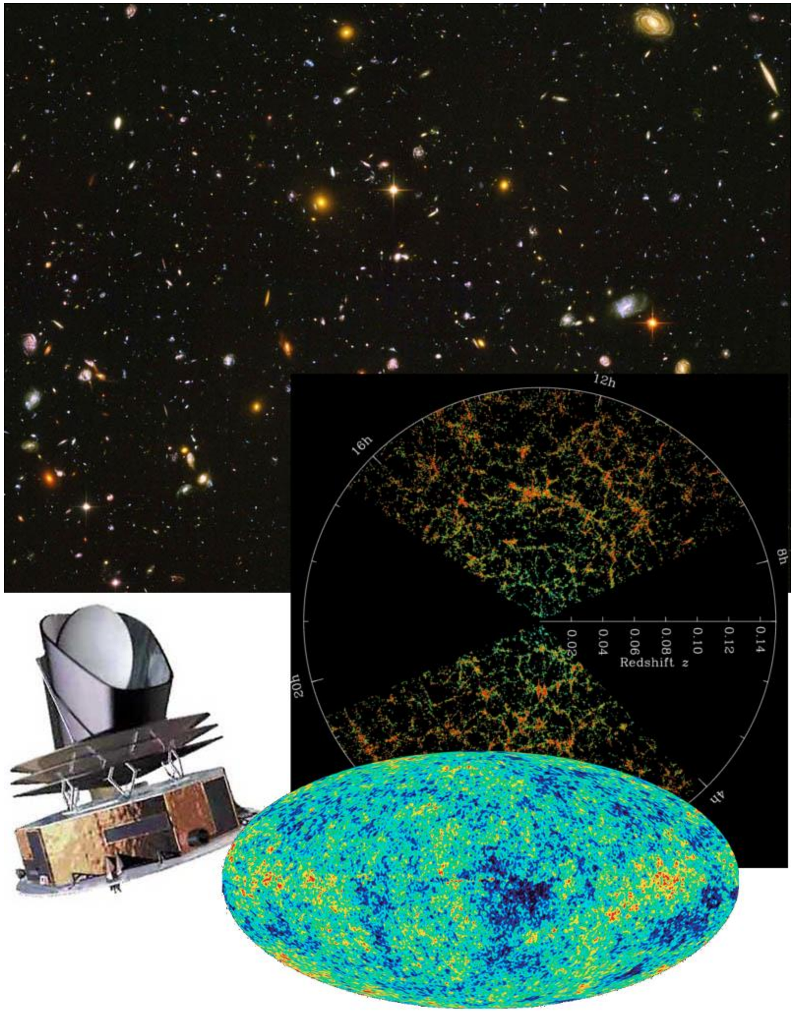
- **"Initial conditions"**: ICs for *gravitational evolution*…
  AFTER inflation
  AFTER Hot Big Bang phenomena

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

- **Cosmostatistics**: discipline of **using the departures from homogeneity** observed in astronomical surveys to **distinguish between cosmological models**.
- Huge data sets, but fundamental limits to information:
  - on large scales: **causality**
  - on small scales: **non-linearity**

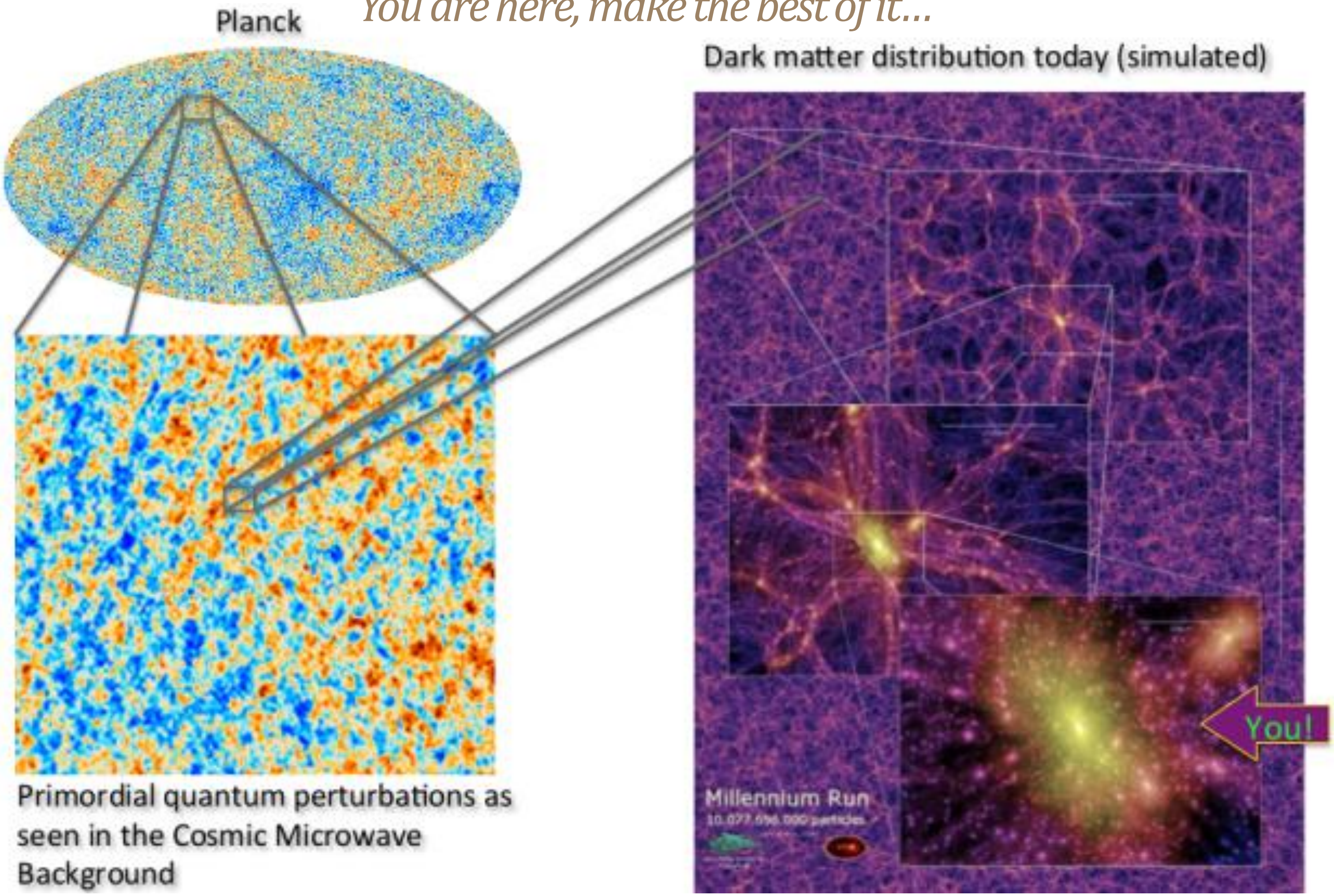
High-energy physics experiments





The inhomogeneous Universe:
the big picture

You are here, make the best of it…

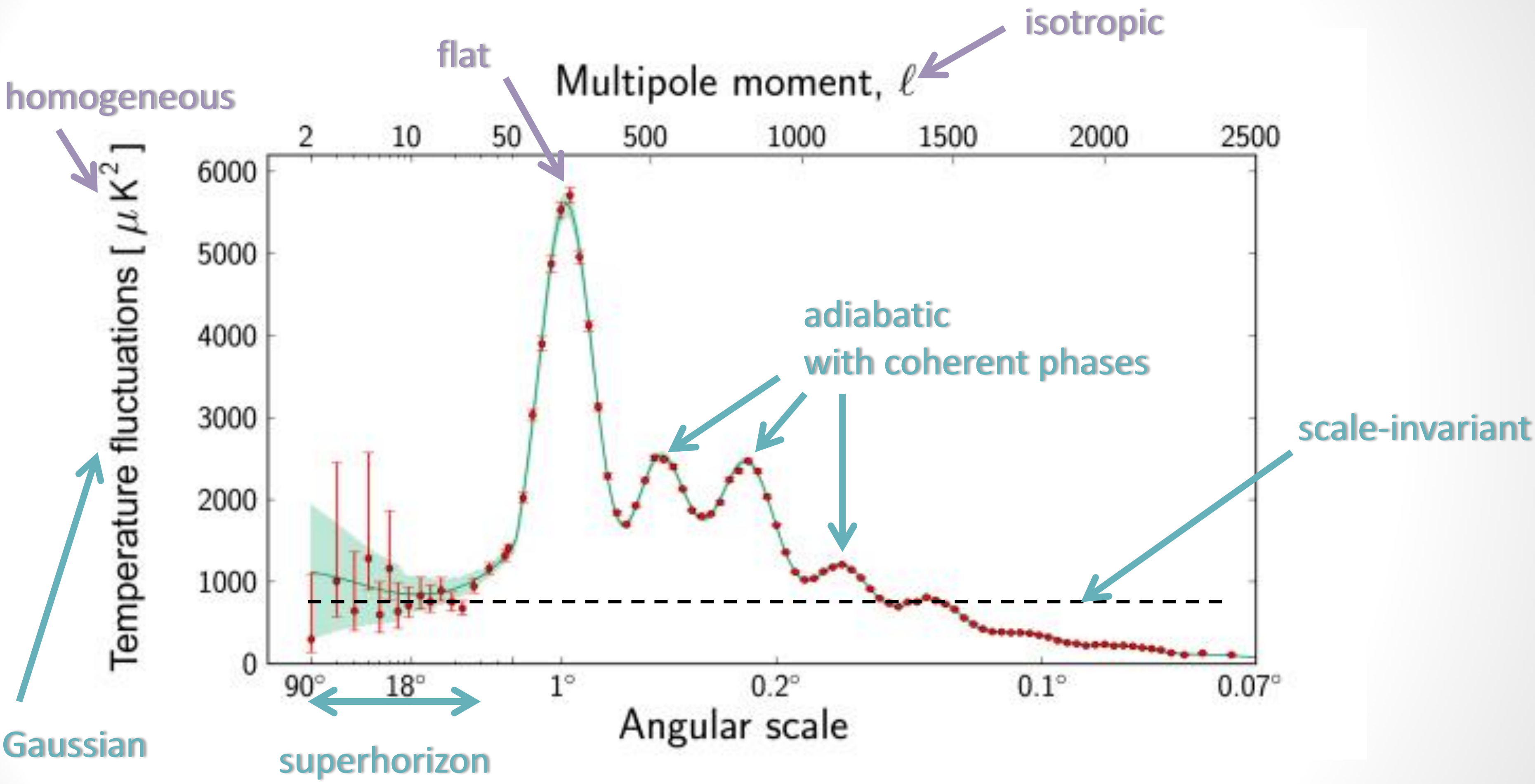


Planck

Dark matter distribution today (simulated)

Primordial quantum perturbations as seen in the Cosmic Microwave Background

The most boring Universe?



Planck collaboration 2013, arXiv:1303.5062

- **Phenomenologically**, inflation is a great success…

The case for physical reconstruction of the ICs

- … but what is the microphysics of inflation?

- Some challenges:
  - The eta problem: scale-invariant, superhorizon fluctuations require

$$\eta = M_{\mathrm{Pl}}^2 \frac{V''}{V} = \frac{m_\phi^2}{3H^2} \ll 1$$

    How to achieve and stabilize this mass hierachy?

  - Large-field inflation: observational gravitational waves require

$$r \geq 0.01 \quad \longleftrightarrow \quad \Delta\phi \gg M_{\mathrm{Pl}}$$

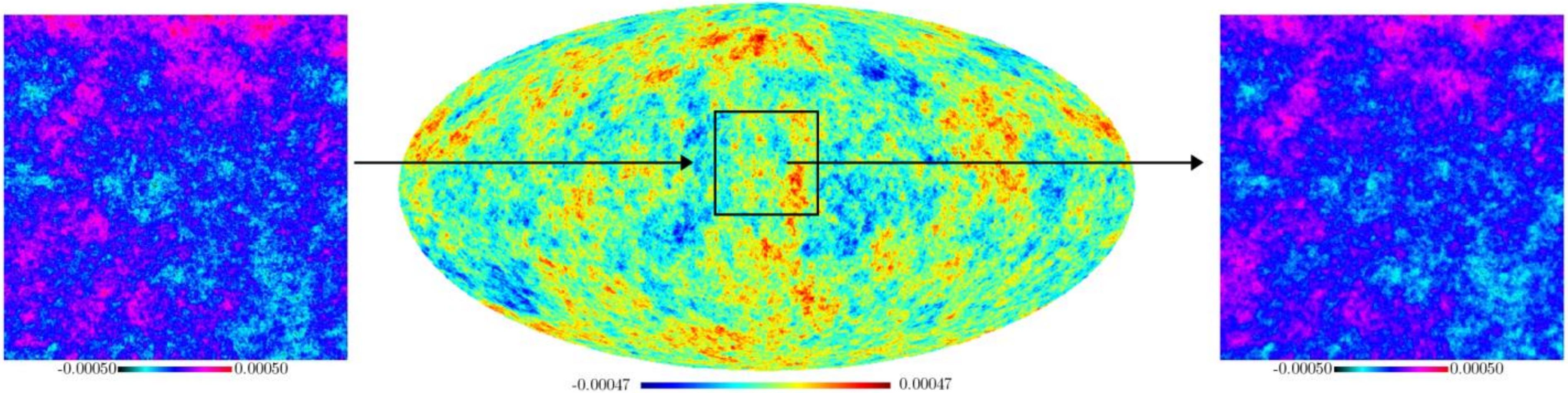
    (Lyth bound)

    Astrophysics **Quantum gravity**

- Some open questions: *multi-field inflation? non-standard kinetic term? periods of fast-roll? non-trivial pre-inflationary state? non-Bunch-Davies vacuum?*

The CMB time-machine

- A time-machine (380,000 yrs ⇨ 10^{-35} s): linear perturbation theory



Input gravitational potential

CMB

Reconstruction of the gravitational potential

adapted from Elsner & Wandelt 2009, arXiv:0909.0009

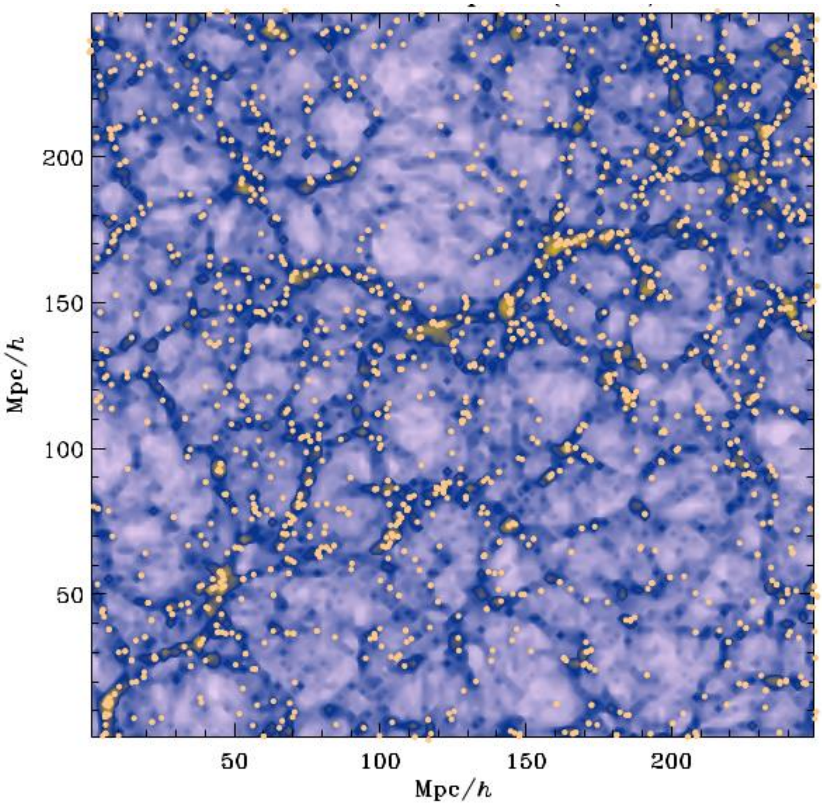
Komatsu, Spergel & Wandelt 2005, arXiv:astro-ph/0305189

Yadav & Wandelt 2005, arXiv:astro-ph/0505386

- Relies on:
  - Gaussian random fields
  - Linear transfer
  - Optimal inference of a GRF from a GRF: Wiener filtering

see also FL, Pisani & Wandelt, proceedings to appear soon

A large-scale structure in the Universe



Blue: matter distribution
Orange: dark matter halos / galaxies

- Halos trace mass distribution (of *dark matter*).
- Halos are NOT randomly distributed: there exists a Large Scale Structure of the Universe
- How do we analyze this structure quantitatively?

Correlation functions and Fourier analysis

A call to modesty…

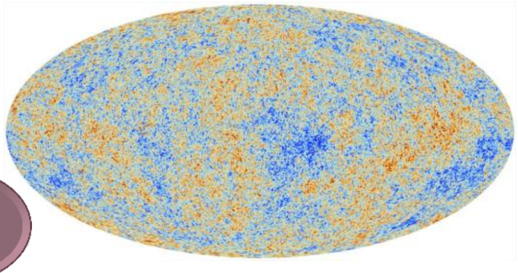
"*Hominem te esse*"

"*Memento mori*"

Reconstruction of the initial conditions…

- … a solved problem !
- And…

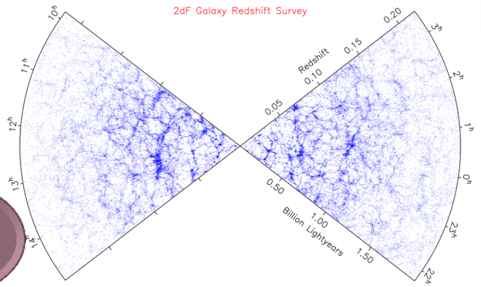
CMB:



2D

$$N_{\rm mode}^{\rm CMB} \propto l_{\rm max}^2$$

LSS:



3D

$$N_{\rm mode}^{\rm LSS} \propto k_{\rm max}^3$$

- The challenges : non-linearity and non-Gaussianity
  - Non-linear transfer functions in the Hot Big Bang phenomena
  - Gravitational evolution
  - Primordial non-Gaussianity (…?)
  - Data imperfection and systematics…

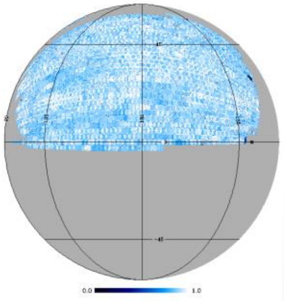
Can we go from the linear to the non-linear problem?

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!**



- A good question: “What is the probability distribution of possible signals compatible with the observations?”

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

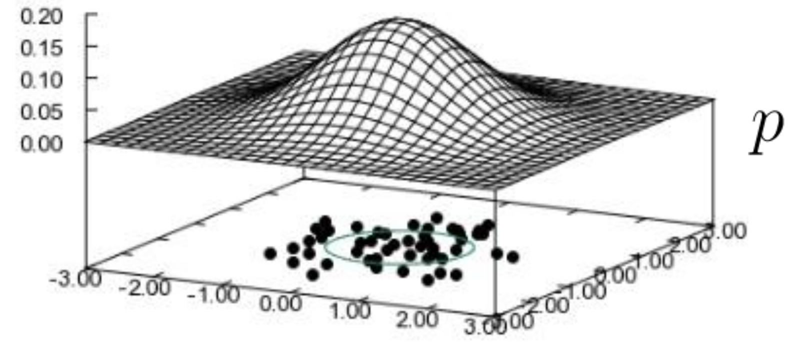
from J. Jasche

Bayesian inference of the ICs

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

➡ **No reduction of the problem size is possible!**

  - Complex posterior distribution
- Numerical approximation: for dim > 4: **sampling** the posterior distribution



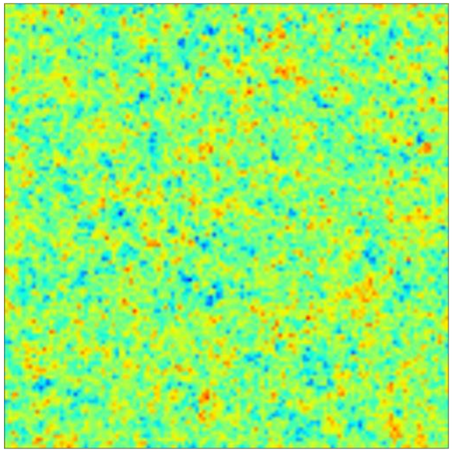
$$p(s|d) \rightarrow p_N(s|d) = \frac{1}{N}\sum_{i=1}^{N}\delta_\mathrm{D}(s - s_i)$$

- But how to "get the dots" ?

from J. Jasche

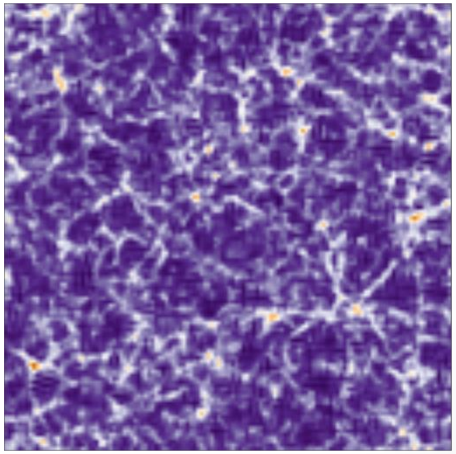
4D physical inference of the ICs

- Physical motivation:
  - Complex final state
  - Simple initial state
  - A “forward only” problem (we have a generative model for the final state)



Initial state

GRAVITY



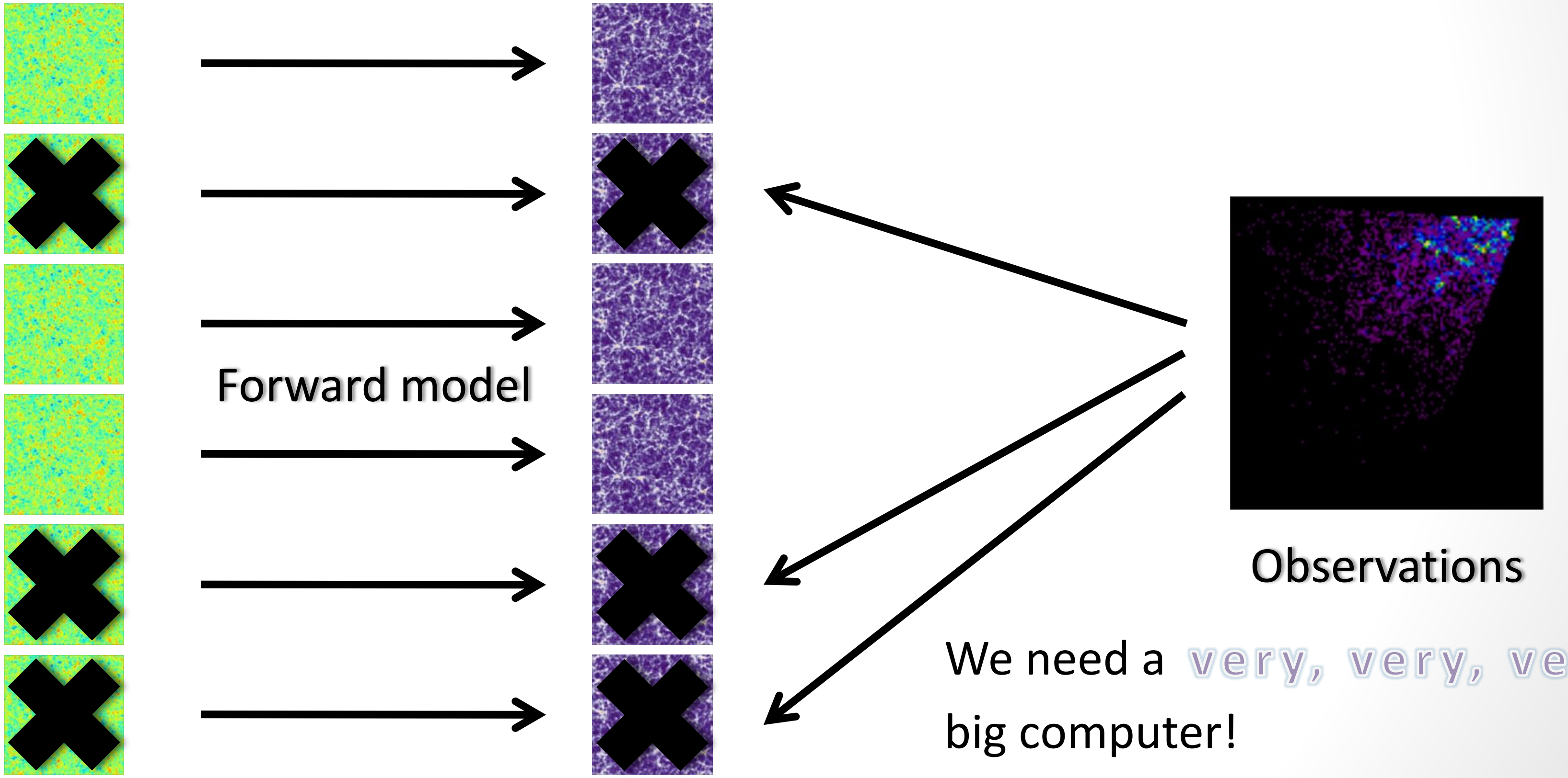
Final state

from J. Jasche

4D physical inference of the ICs

- The ideal scenario:

Forward model = N-body simulation + Halo occupation + Galaxy formation + Feedback + …



Forward model

Observations

We need a very, very, very big computer!

Every possible ICs

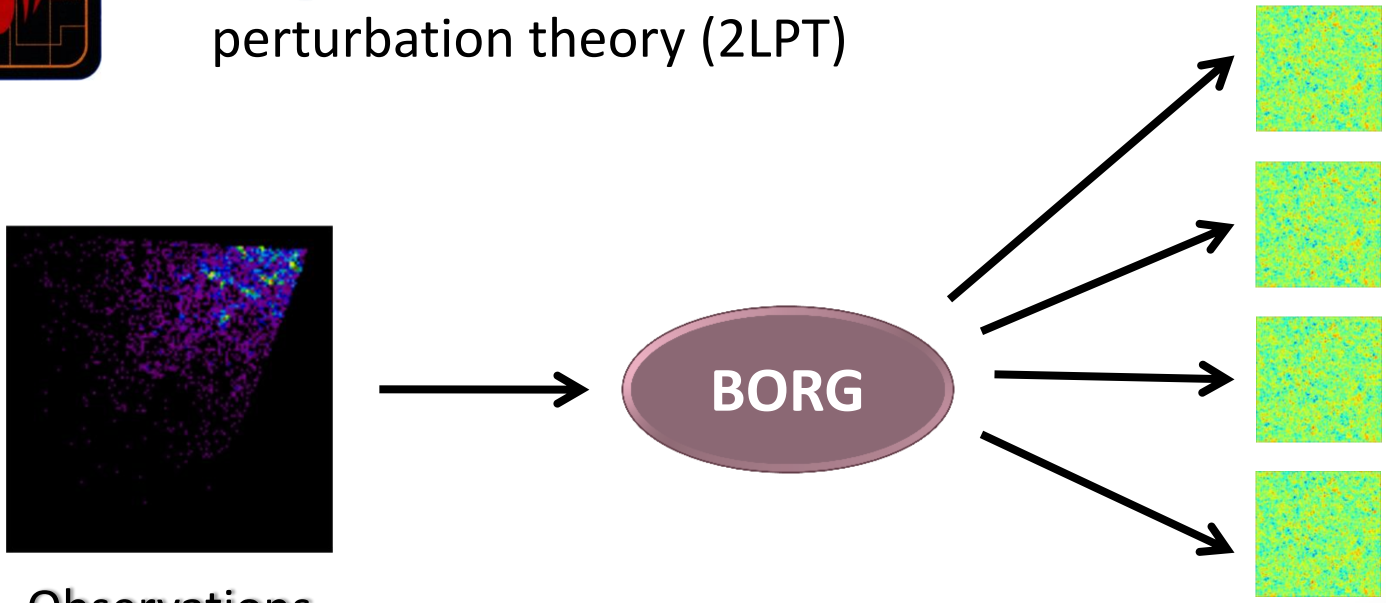
Every possible FCs

from J. Jasche

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)



Observations

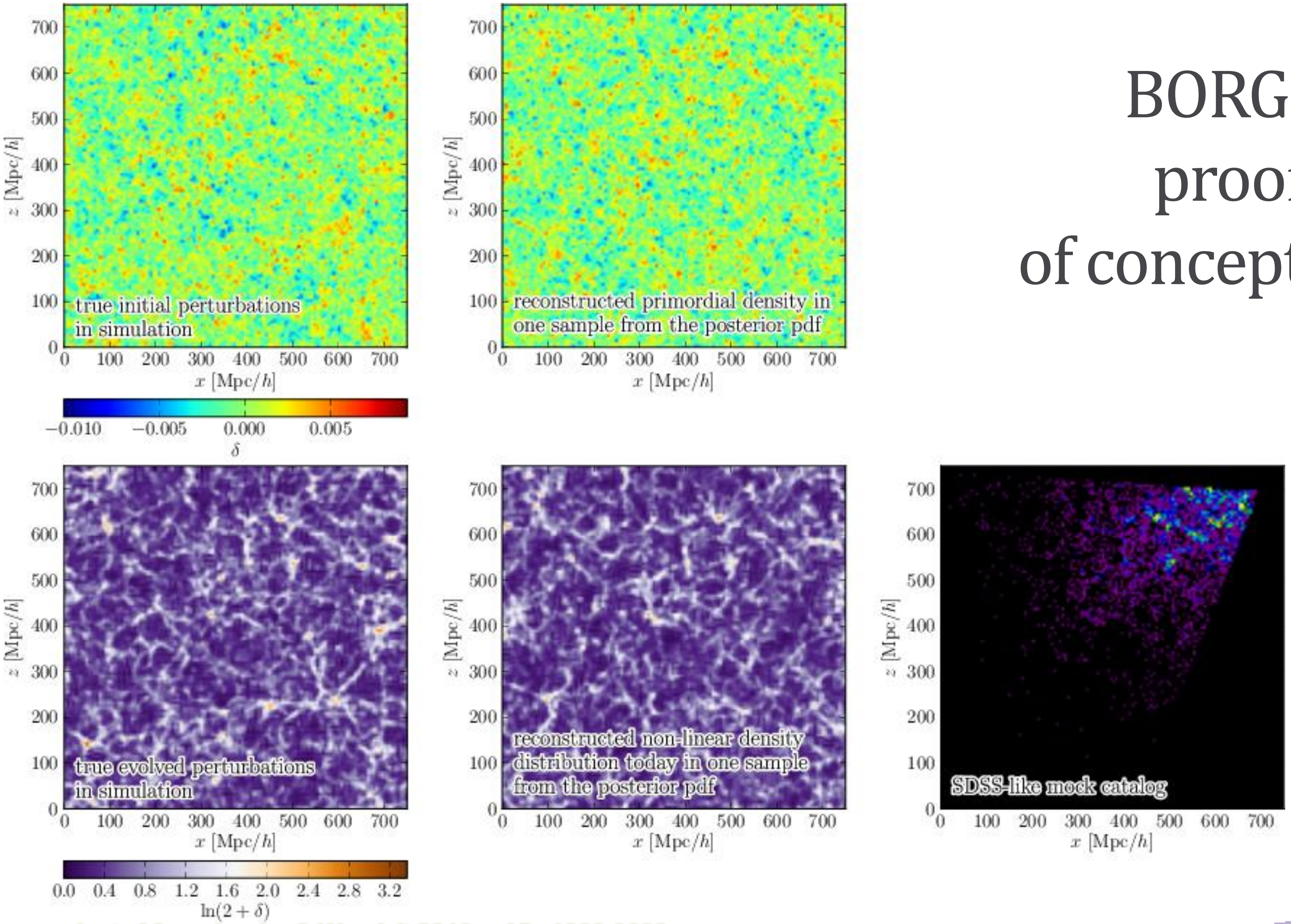
BORG

Samples of possible ICs

Jasche & Wandelt 2012, arXiv:1203.3639

from J. Jasche

BORG: proof of concept



adapted from Jasche & Wandelt 2012, arXiv:1203.3639

Samples of the posterior density

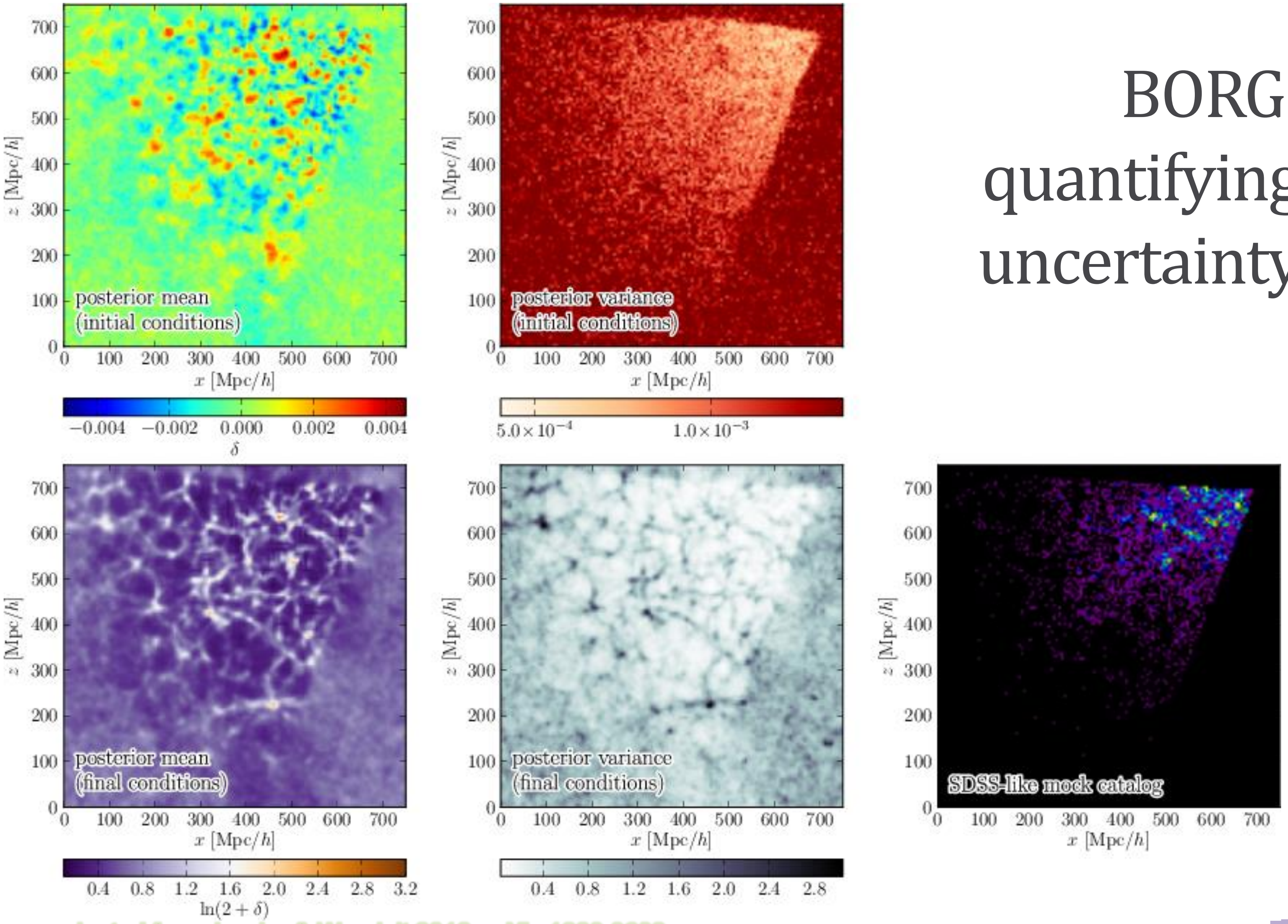
- Each sample: a possible version of the truth

- 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…)

- By the way… Bayesian probability theory deals with uncertainty independently of its origin

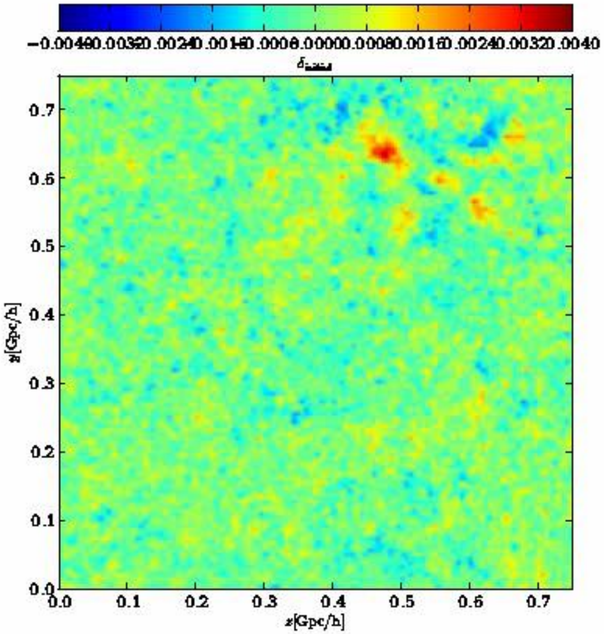
see also FL, Pisani & Wandelt, proceedings to appear soon

BORG: quantifying uncertainty

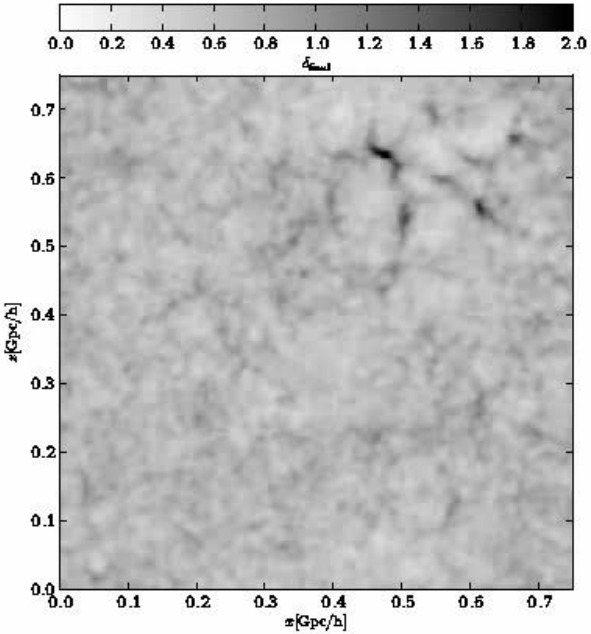


adapted from Jasche & Wandelt 2012, arXiv:1203.3639

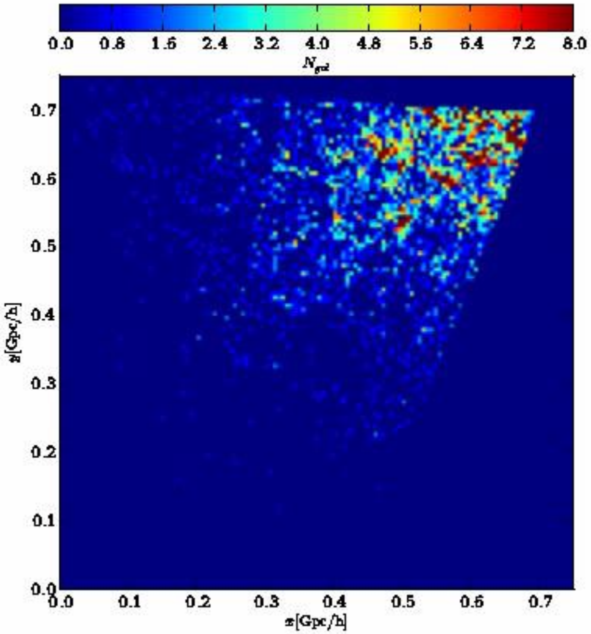
BORG at work



Initial conditions



Final conditions



Observations

Jasche & Wandelt 2012, arXiv:1203.3639

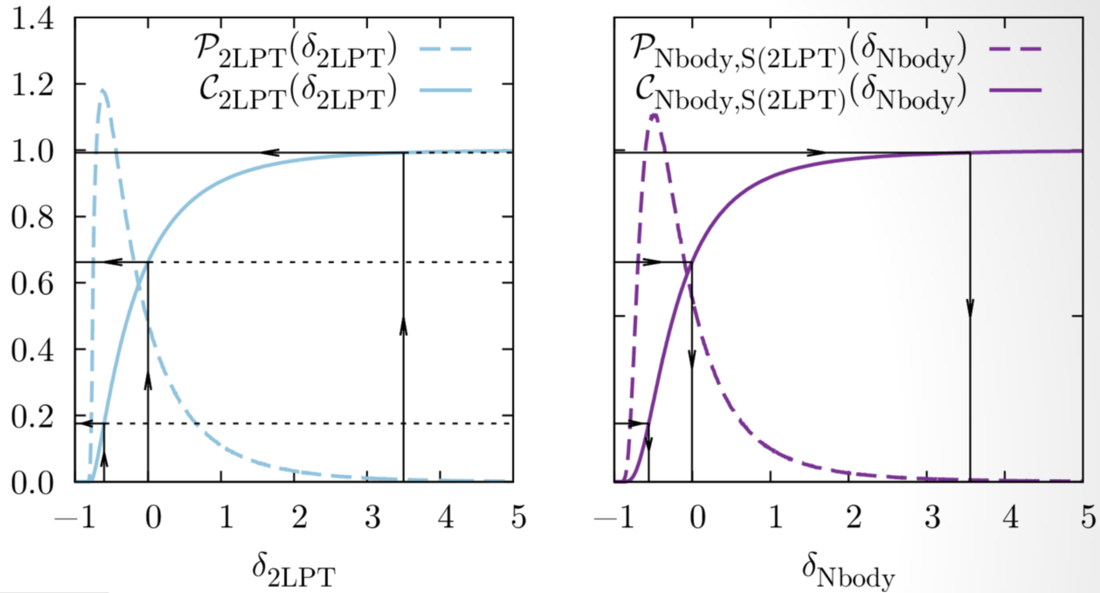
Aside: beyond 2LPT?

- 2LPT breaks down at small scales.
- Recall 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.

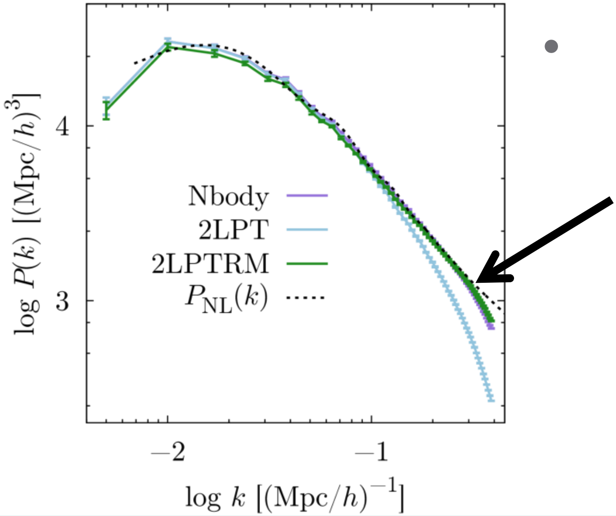
Remapping 2LPT in the mildly non-linear regime

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

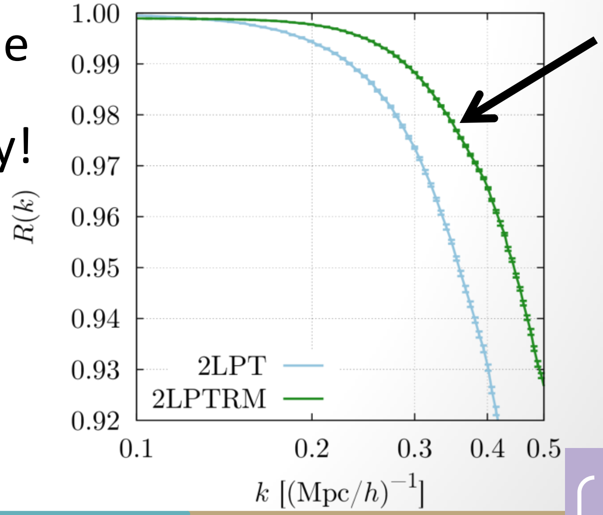
- Replacing the one-point distribution of 2LPT by one which accounts for the full non-linear system…



- …also improves the higher-order correlators…



- …and the phase accuracy!



COLA: *COmoving Lagrangian Acceleration*

- Write the display vector as: $\mathbf{s} = \mathbf{s}_{\mathrm{LPT}} + \mathbf{s}_{\mathrm{MC}}$

Tassev & Zaldarriaga 2012, arXiv:1203.5785

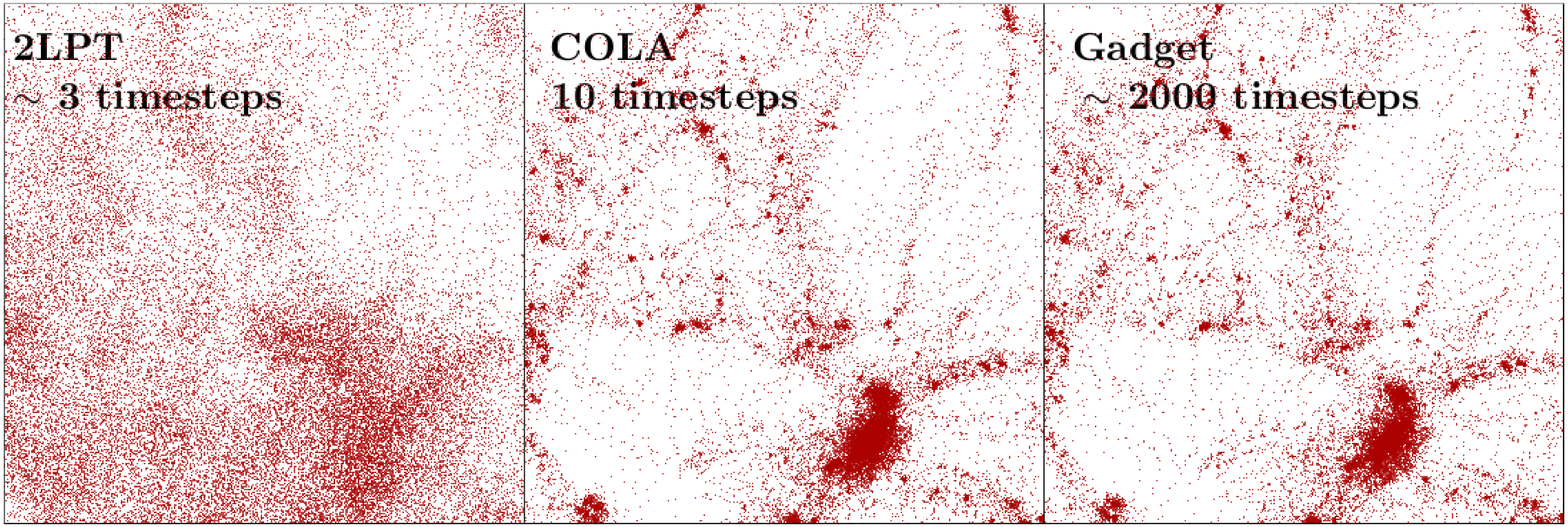
- Time-stepping (omitted constants and Hubble expansion):

Standard:

$$\partial_\tau^2 \mathbf{s} = -\nabla\Phi$$

Modified:

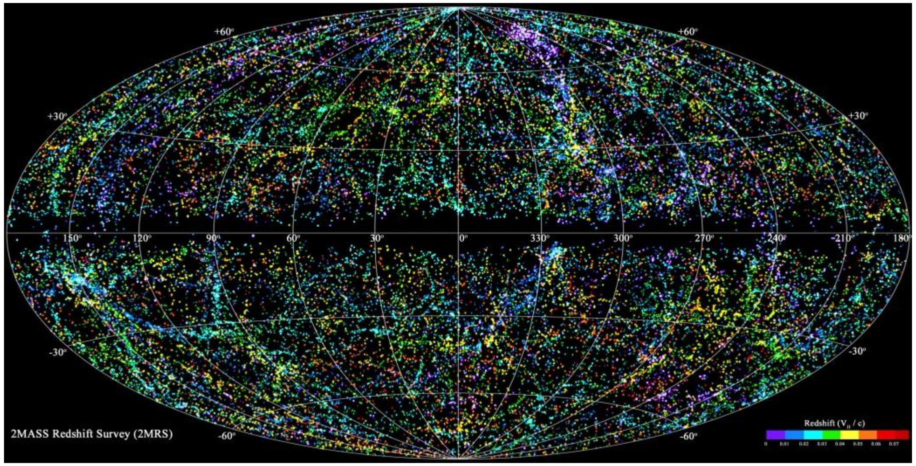
$$\partial_\tau^2 \mathbf{s}_{\mathrm{MC}} = \partial_\tau^2 (\mathbf{s} - \mathbf{s}_{\mathrm{LPT}}) = -\nabla\Phi - \partial_\tau^2 \mathbf{s}_{\mathrm{LPT}}$$



Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322

Back to BORG

- What about the real Universe?



  - Reconstruction of the initial conditions from SDSS DR7

    Jasche, FL & Wandelt, in prep.

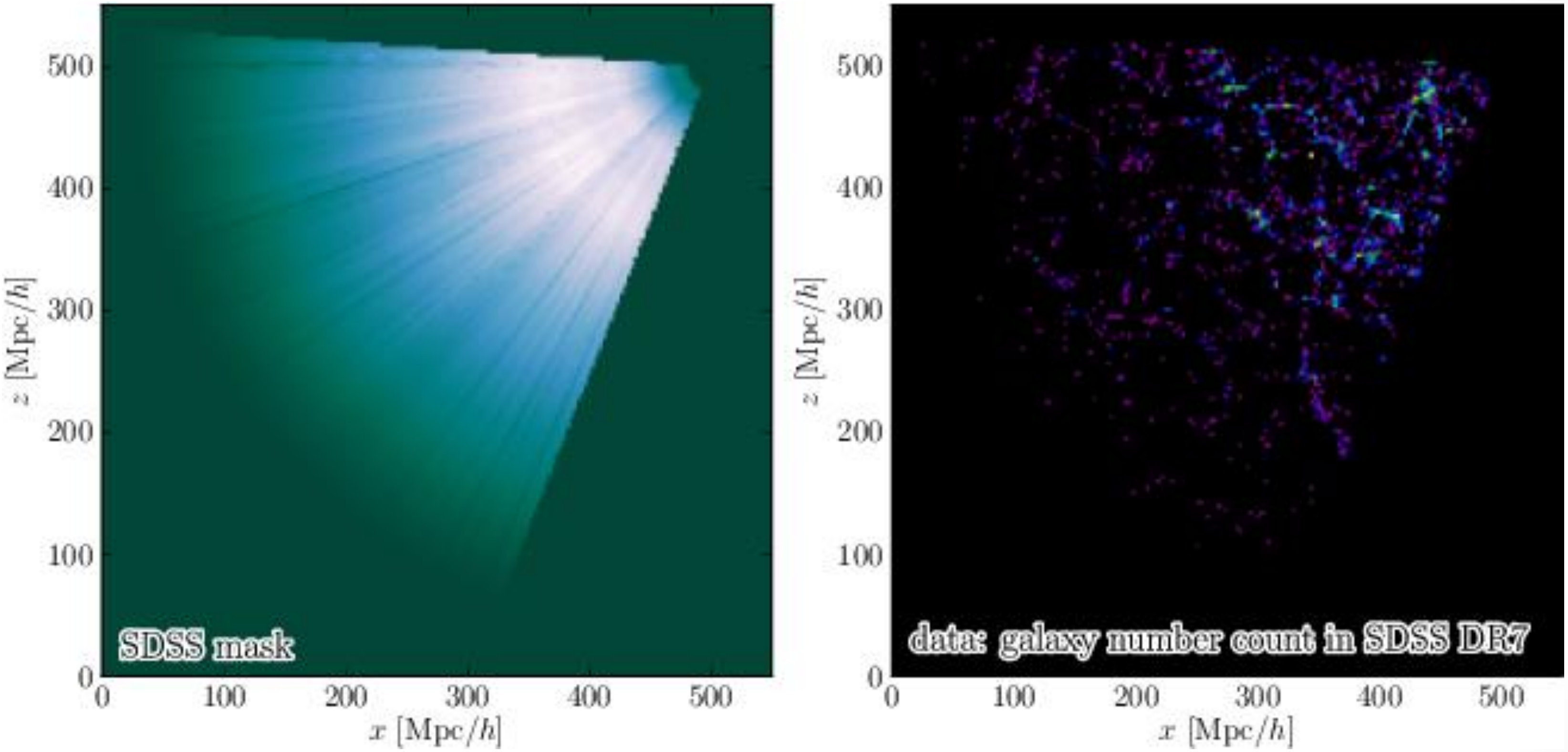
  - Resimulation of the late-time Universe from these ICs

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

  - Optimizing an ensemble of constrained simulations to probe the non-linear regime

    FL, Jasche & Wandelt, in prep.

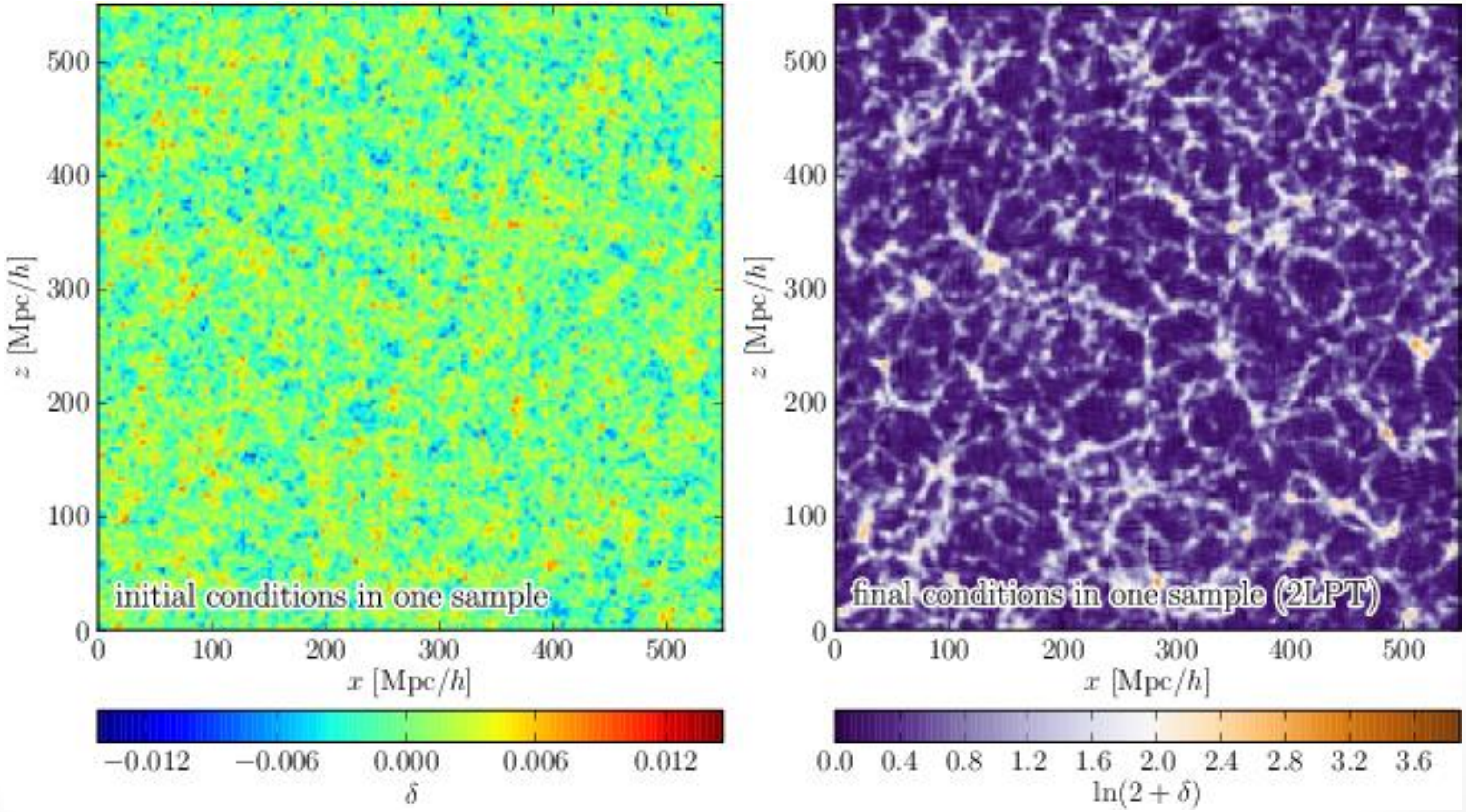
BORG: reconstructions from SDSS DR7



Data

Jasche, FL & Wandelt, in prep.

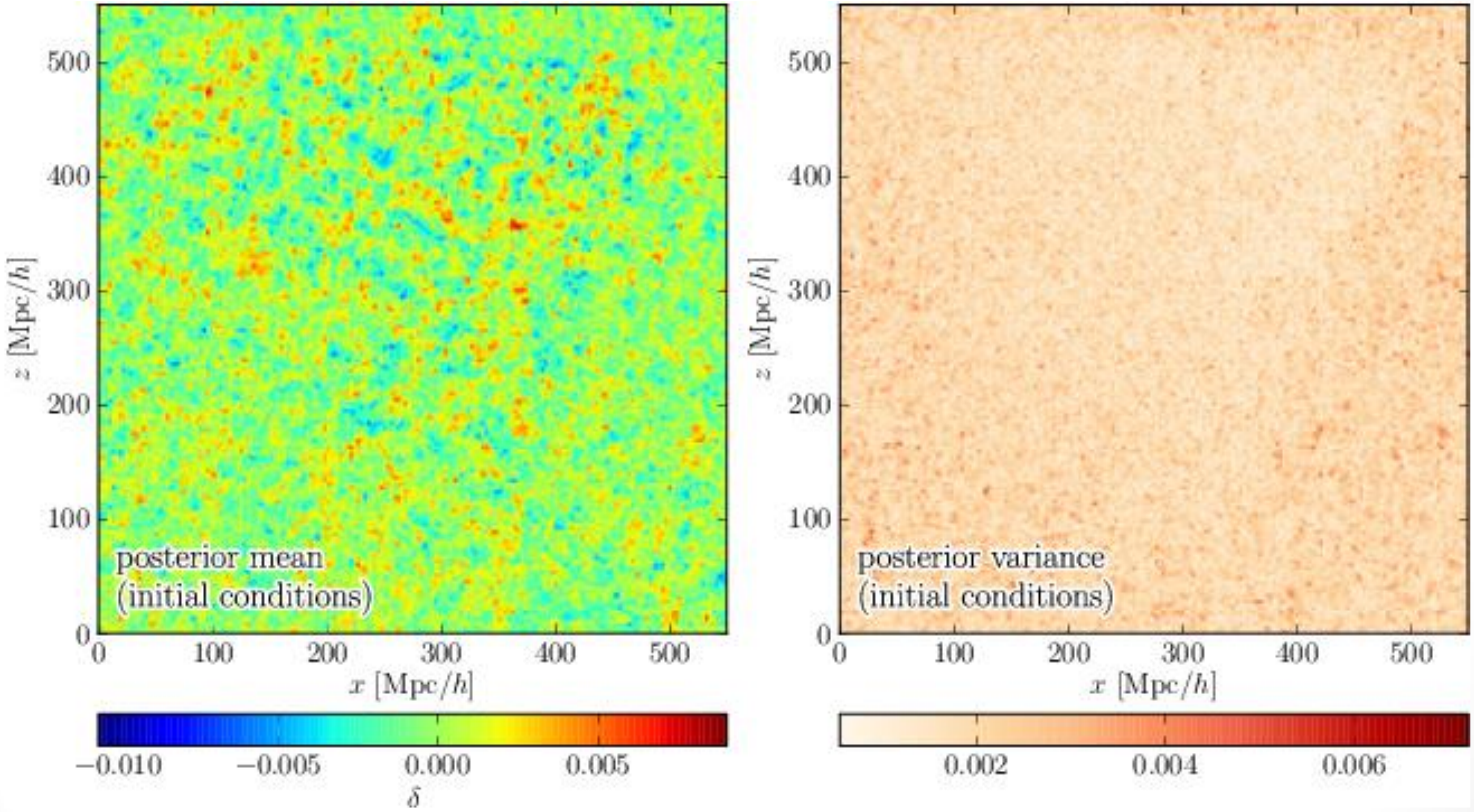
BORG: reconstructions from SDSS DR7



One sample

Jasche, FL & Wandelt, in prep.

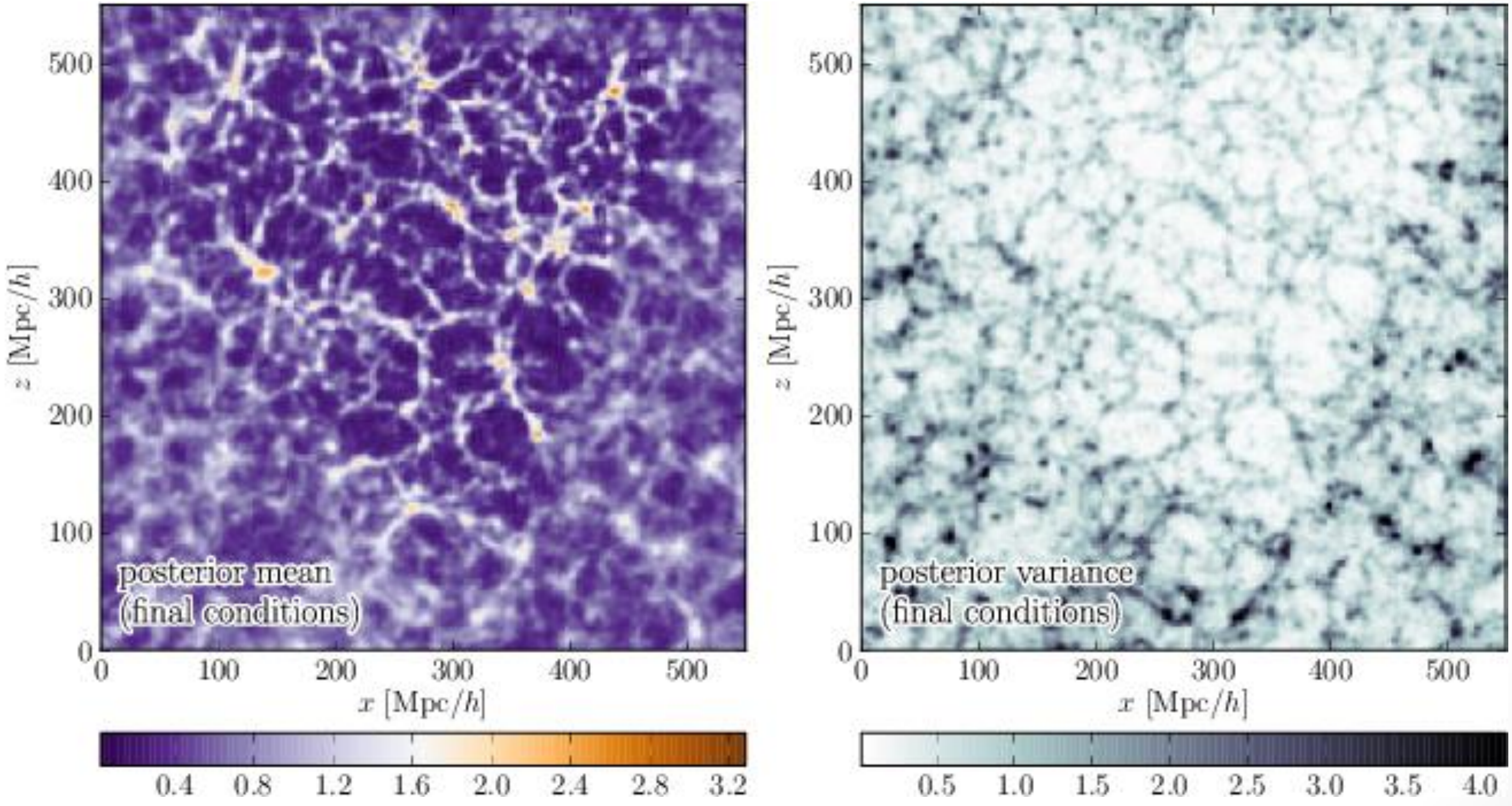
BORG: reconstructions from SDSS DR7



Quantifying uncertainty in the ICs

Jasche, FL & Wandelt, in prep.

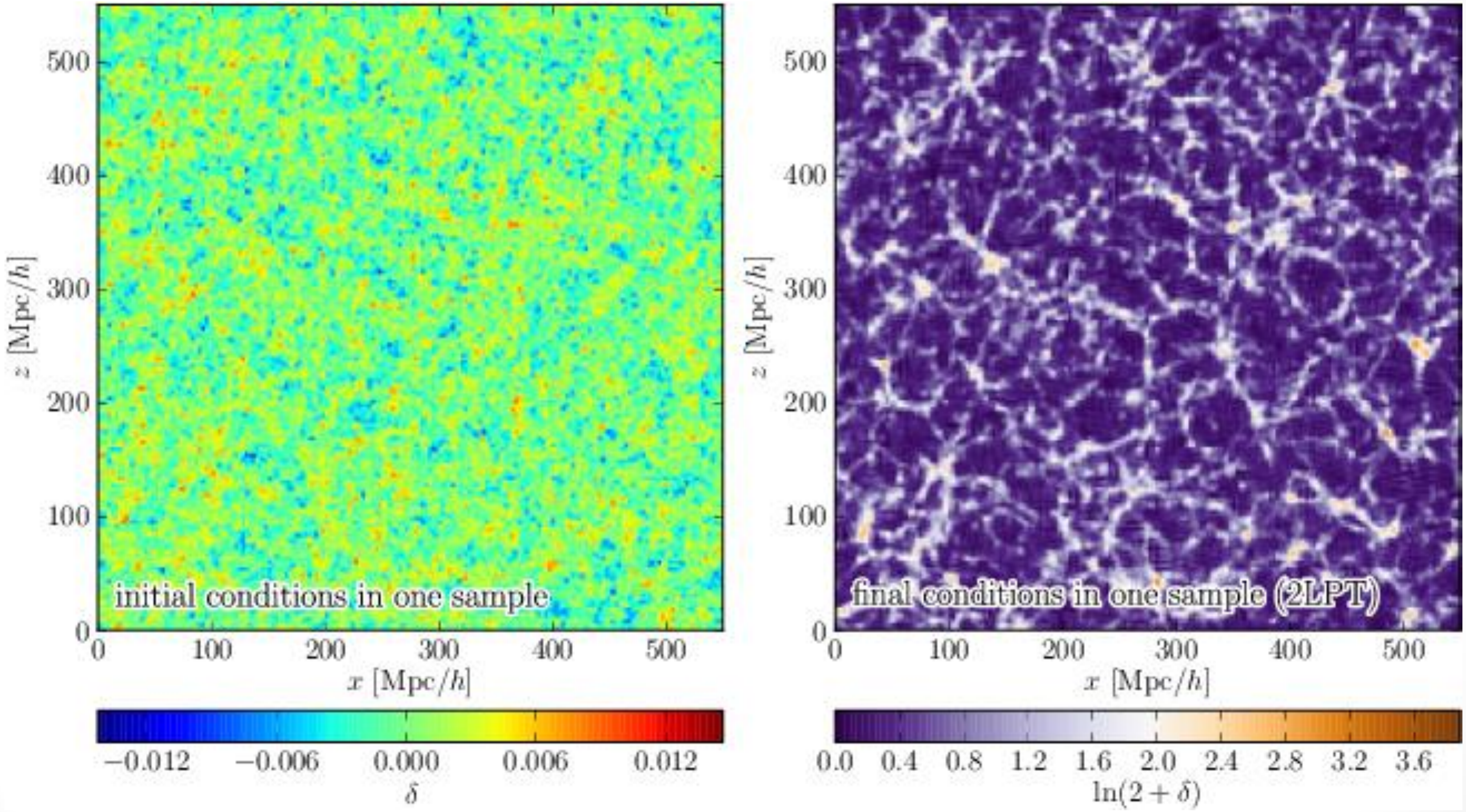
BORG: reconstructions from SDSS DR7



Quantifying uncertainty in the FCs

Jasche, FL & Wandelt, in prep.

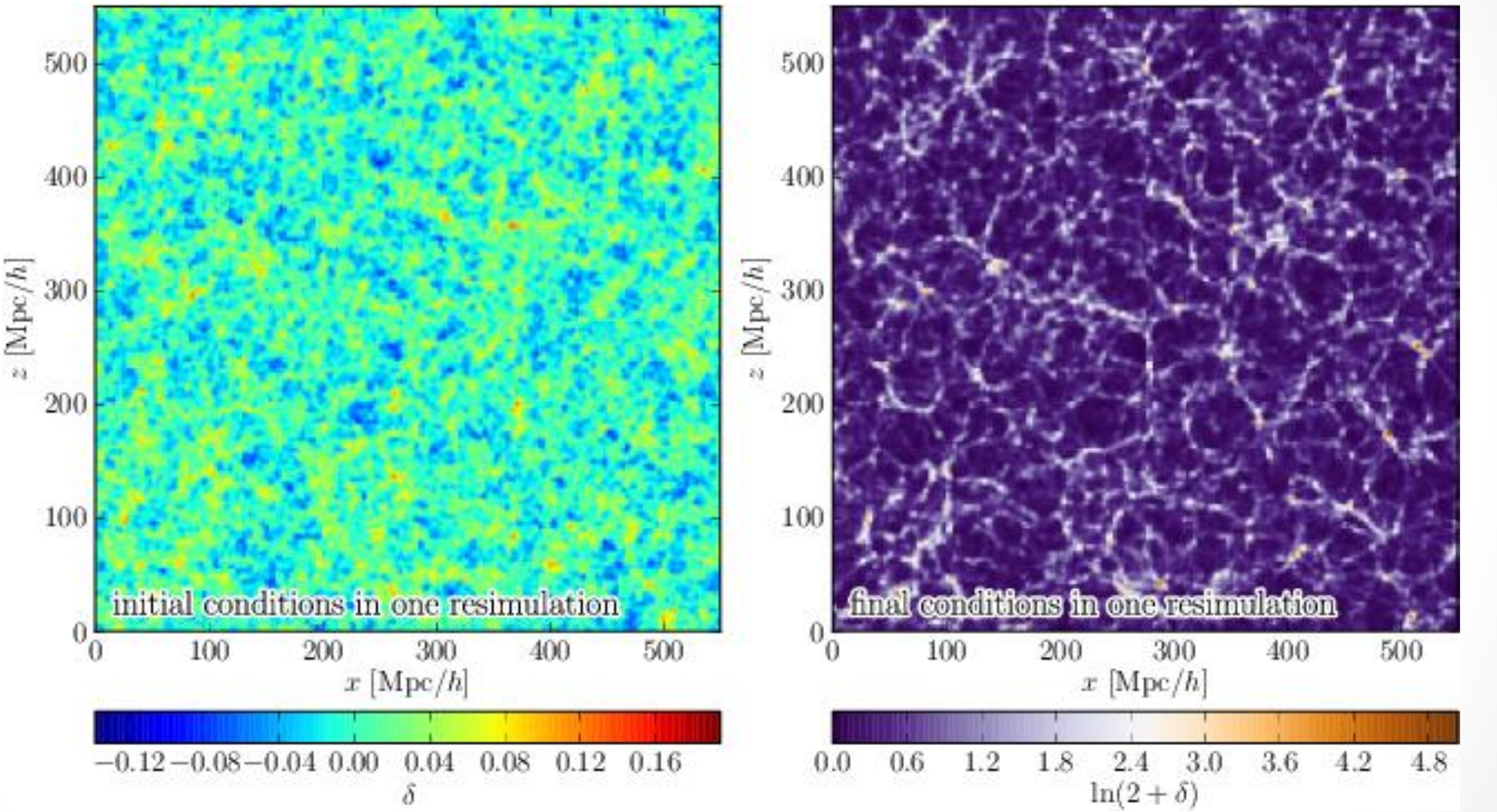
BORG: reconstructions from SDSS DR7



One sample

Jasche, FL & Wandelt, in prep.

BORG: resimulating the Universe



Resimulation of one sample

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

BORG: resimulating the Universe

- 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.

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

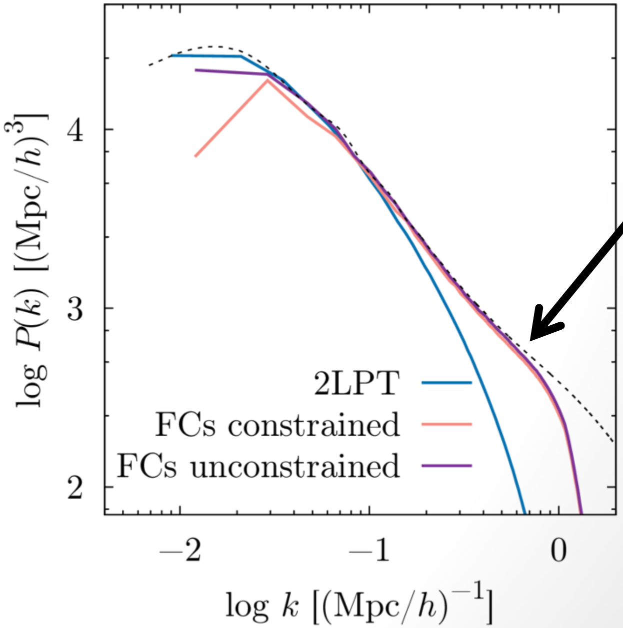
- With a full N-body simulation, we address the non-linear regime of structure formation!

more on non-linear reconstruction of the ICs in FL, Jasche & Wandelt, in prep.

see also:

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

Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322



Concluding thoughts

- BORG: A non-linear time machine using Bayesian posterior exploration to infer primordial quantities from late-time observations
- Need for efficient tools to model cosmic structure formation the non-linear regime
- Cosmological physical reconstruction of the initial conditions of the Universe is becoming feasible. Great science is waiting behind the door:
  - Baryon acoustic oscillations, clusters, voids
  - Non-Gaussianity
  - Isocurvature perturbations
  - Gravitational waves in the large-scale structure…

Don't fight non-linearity to get cosmological information – embrace it!