Bayesian large-scale structure inference and cosmic web analysis

Florent Leclercq

Institut d'Astrophysique de Paris
Institut Lagrange de Paris
École polytechnique ParisTech - Université Paris-Saclay

September 24th, 2015

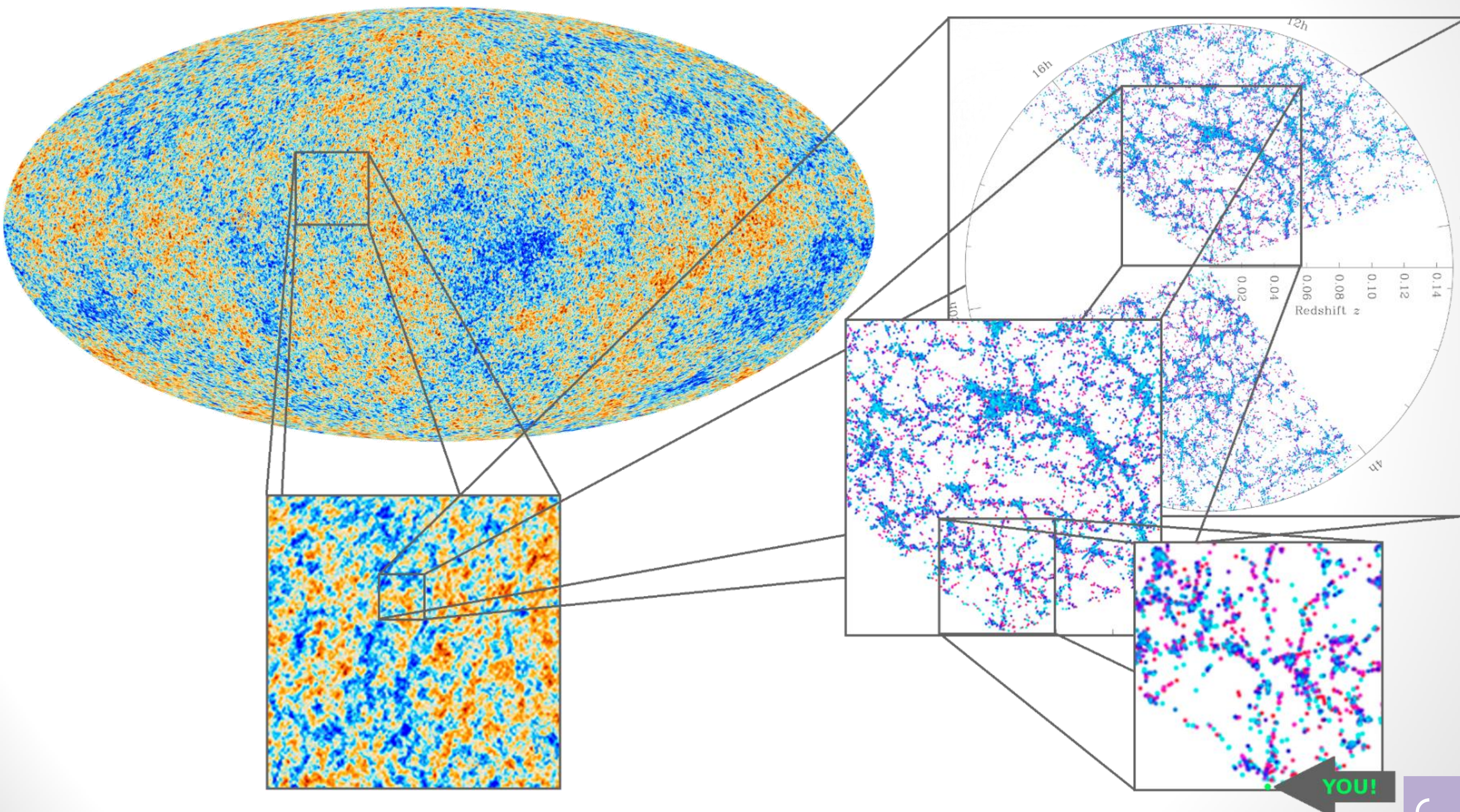
Supervised by Benjamin Wandelt (IAP/U. Illinois)

In collaboration with:

Héctor Gil-Marín (ICG Portsmouth), Nico Hamaus (LMU/IAP), Jens Jasche (ExC Universe, Garching/IAP), Guilhem Lavaux (IAP), Alice Pisani (LAM/IAP), Emilio Romano-Díaz (U. Bonn), Paul M. Sutter (Trieste/IAP/Ohio State U.)

The big picture: the Universe is highly structured

You are here. Make the best of it…



Planck collaboration (2013) M. Blanton and the Sloan Digital Sky Survey (2010-2013)

How did structure appear in the Universe?

A joint problem!

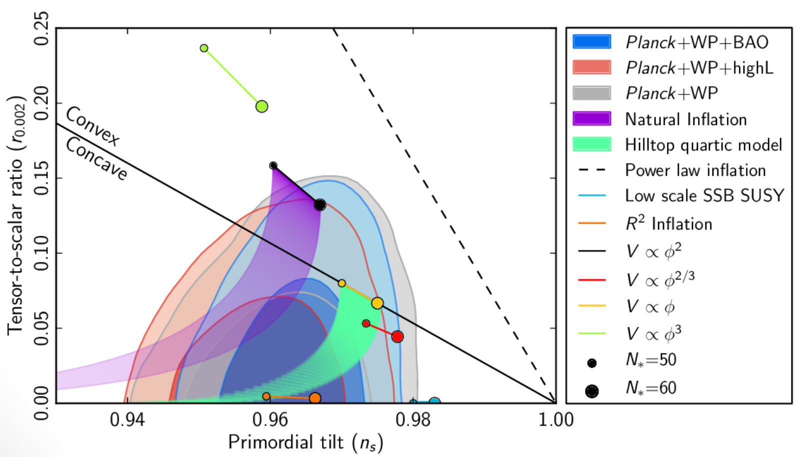
- **How did the Universe begin?**
  - What are the statistical properties of the initial conditions?
- **How did the large-scale structure take shape?**
  - What is the physics of dark matter and dark energy?

We have theoretical and computer models…

- **Initial conditions:**
  a Gaussian random field

$$\mathcal{P}(\delta^{\mathrm{i}}|S) = \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} \sum_{x,x'} \delta^{\mathrm{i}}_x S^{-1}_{xx'} \delta^{\mathrm{i}}_{x'} \right)$$

Everything seems consistent with the simplest inflationary scenario, as tested by Planck.

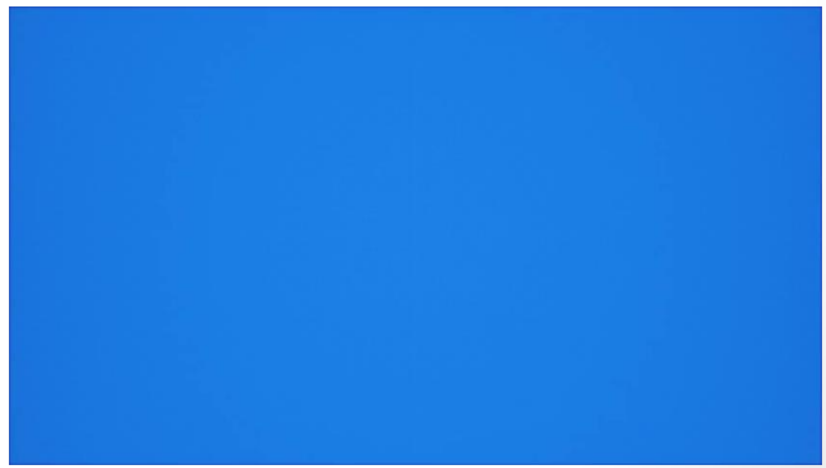


Planck 2015 XX, arXiv:1502.02114

- **Structure formation:**
  numerical solution of the Vlasov-Poisson system for dark matter dynamics

$$\frac{\partial f}{\partial \tau} + \frac{\mathbf{p}}{ma} \cdot \nabla f - ma \nabla \Phi \cdot \frac{\partial f}{\partial \mathbf{p}} = 0$$

$$\Delta \Phi = 4\pi \mathrm{G} a^2 \bar{\rho} \delta$$



Y. Dubois & S. Colombi (IAP)

But some questions remain

1. How do we **test** these frameworks?
   - Usually the two problems of initial conditions and structure formation are addressed in isolation.
   - Ideally, galaxy surveys should be analyzed in terms of the joint constraints that they place on these two questions.

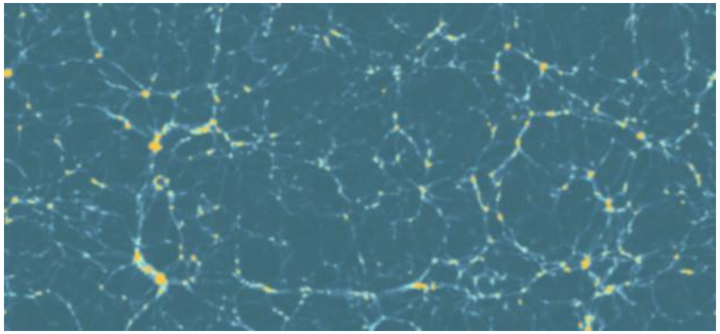
2. How did this happen in **our** Universe?

1. How do we test our models?

J. Cham – PhD comics

- Precise tests require many modes.

In 3D galaxy surveys, the number of modes usable scales as k_{max}^3 .

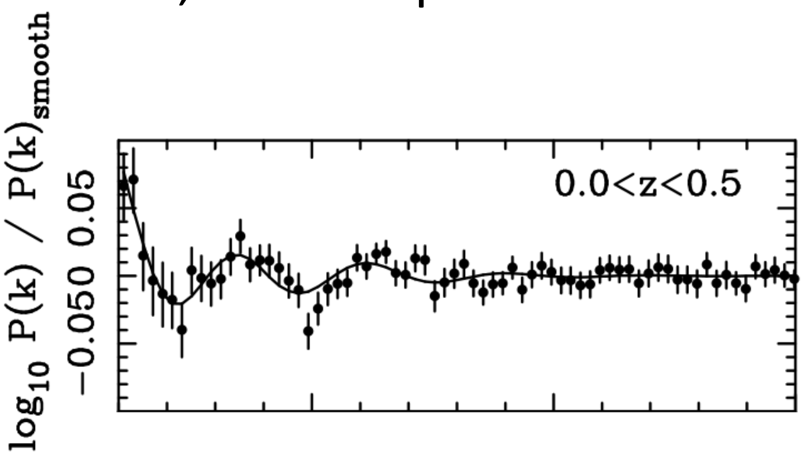


- The challenge: non-linear evolution at **small scales** and **late times**.
- The strategy:
  - Pushing down the smallest scale usable for cosmological analysis
  - Inferring the initial conditions from galaxy positions

In other words: go beyond the **linear** and **static** analysis of the LSS.

2. How did this happen in our Universe?

- This means that we cannot do, for example:



Percival *et al*. 2010, arXiv:0907.1660

- Standard analyses: reduce the data to some statistics, then fit some model parameters

- We have to do a **joint analysis** of all aspects, including **density reconstruction**
  - Provides powerful constraints
  - Propagates uncertainties between all parts of the analysis
  - Avoids using the data twice
- It is a process known as **data assimilation**

Can we just **fit the entire survey**?

Why Bayesian inference?

- What do we need to fit the entire survey?

Inference of signals = ill-posed problem

- Incomplete observations: finite resolution, survey geometry, selection effects
- Noise, biases, systematic effects
- Cosmic variance

➡ **No unique recovery is possible!**

"What is the formation history of the Universe?" ➡ "What is the probability distribution of possible formation histories (signals) compatible with the observations?"

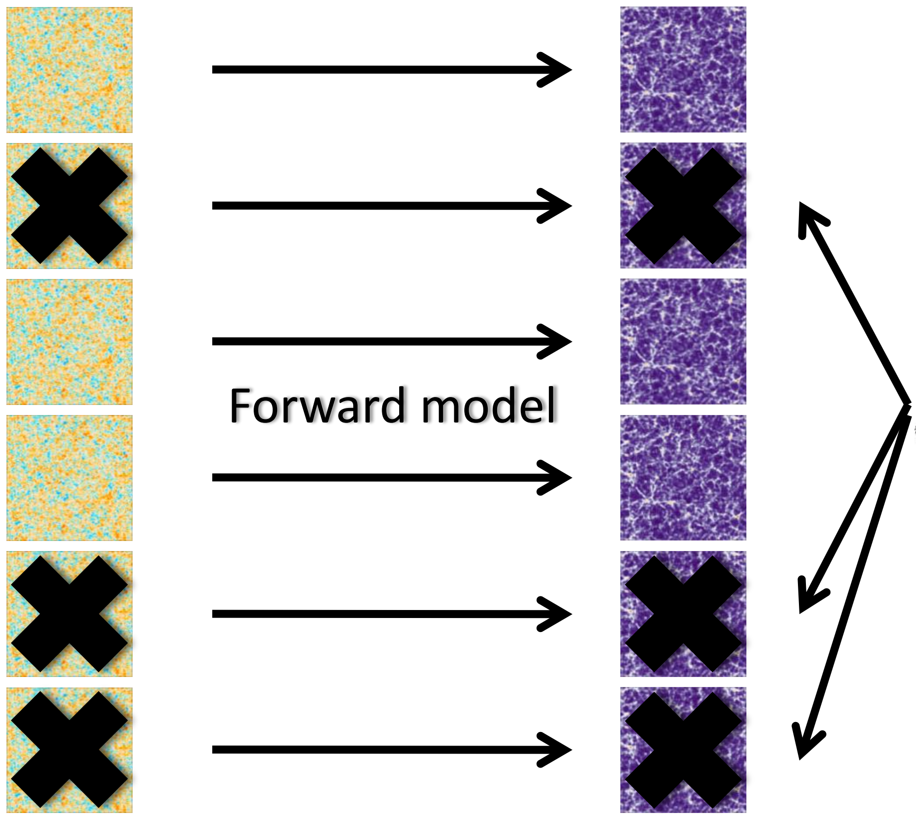
Bayes' theorem: $\mathcal{P}(s|d)\mathcal{P}(d) = \mathcal{P}(d|s)\mathcal{P}(s)$

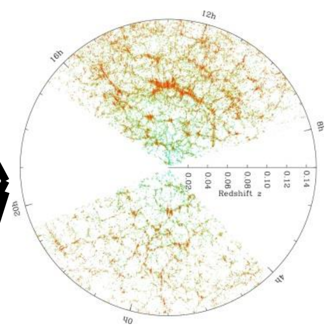
- Cox-Jaynes theorem: Any system to manipulate "*plausibilities*", consistent with Cox's desiderata, is isomorphic to **(Bayesian) probability theory**

➡ How to do that? Thesis chapter 3

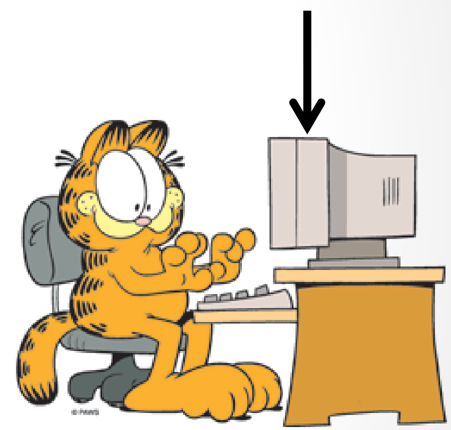
Bayesian forward modeling: the ideal scenario

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





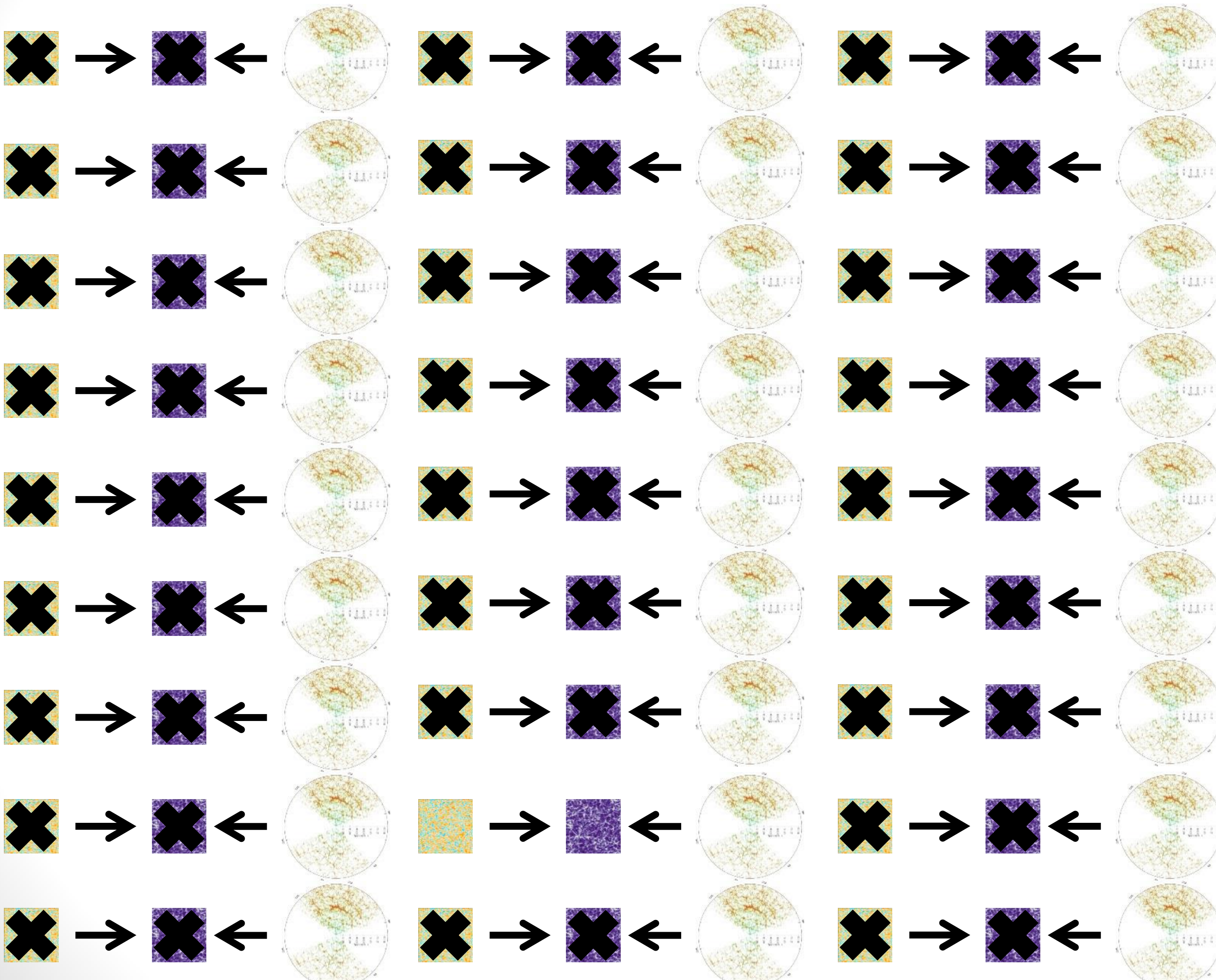
Observations



All possible ICs

All possible FCs

Bayesian forward modeling: the ideal scenario



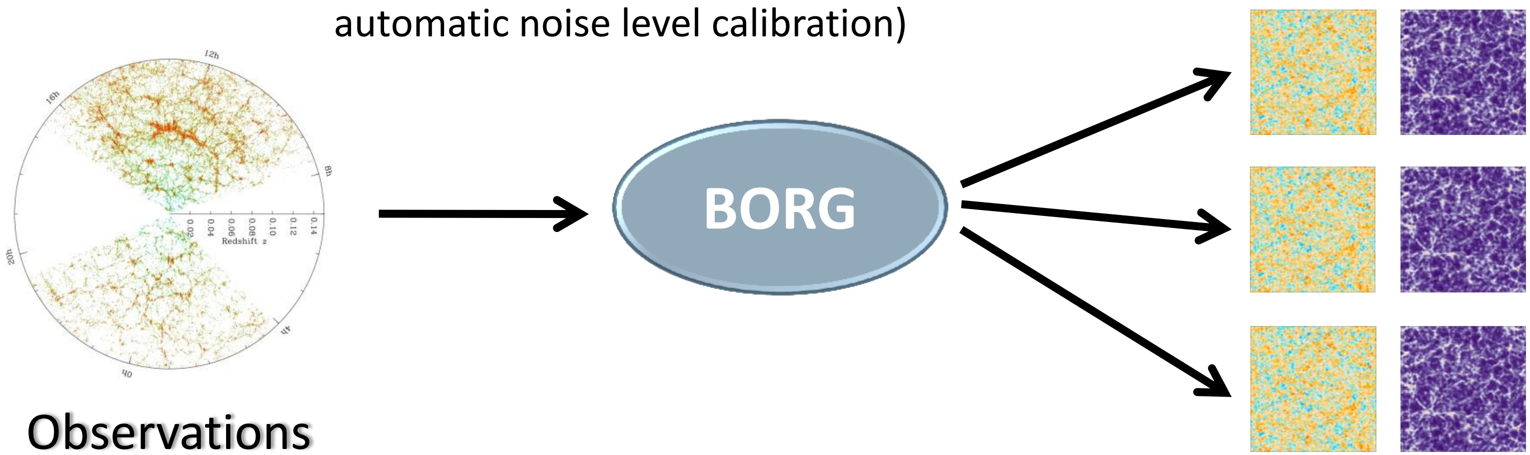
$d \approx 10^7$

BORG: *Bayesian Origin Reconstruction from Galaxies*

What makes the problem tractable:

- **Sampler**: Hamiltonian Markov Chain Monte Carlo method
- **Data model**: Gaussian prior – Second-order Lagrangian perturbation theory (2LPT) – Poisson likelihood

  (and also: luminosity-dependent galaxy bias, automatic noise level calibration)



Observations

(galaxy catalog + meta-data: selection functions, completeness...)

Samples of possible 4D states

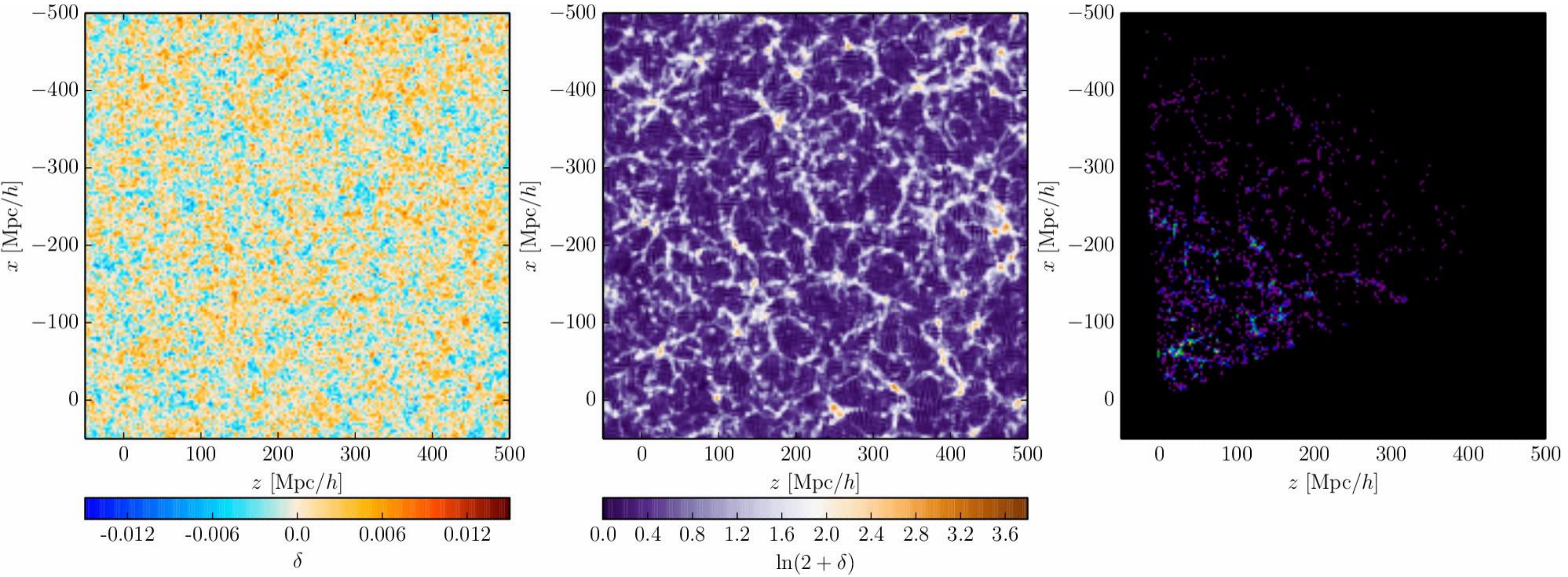
Jasche & Wandelt 2013, arXiv:1203.3639

Jasche, FL & Wandelt 2015, arXiv:1409.6308

Thesis chapter 4

Chrono-Cosmography

BORG at work: SDSS chrono-cosmography



Initial conditions

Final conditions

Observations

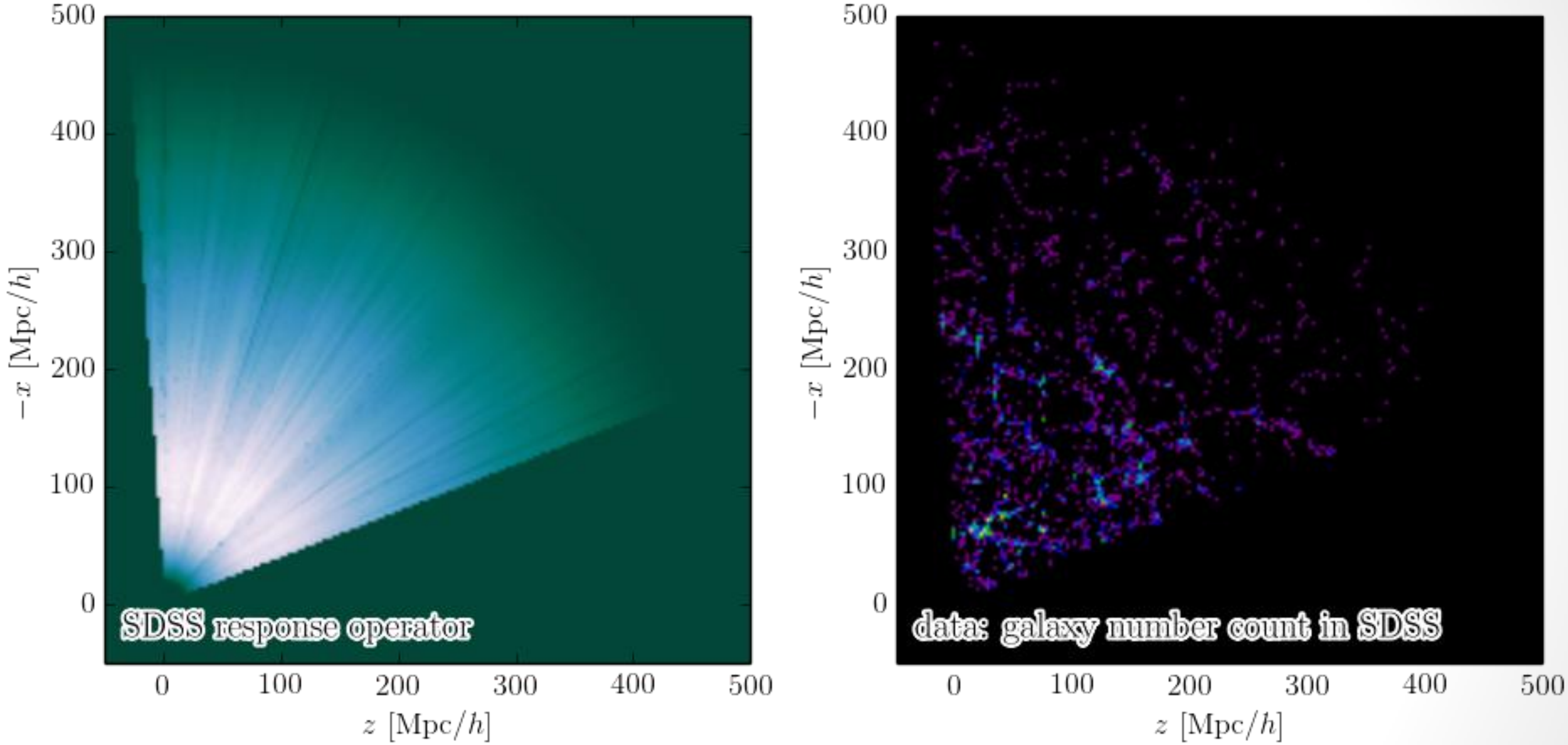
The BORG SDSS run:

334,074 galaxies, ≈ 17 millions parameters, 12,000 samples, 3 TB, 10 months on 32 cores

Jasche, FL & Wandelt 2015, arXiv:1409.6308

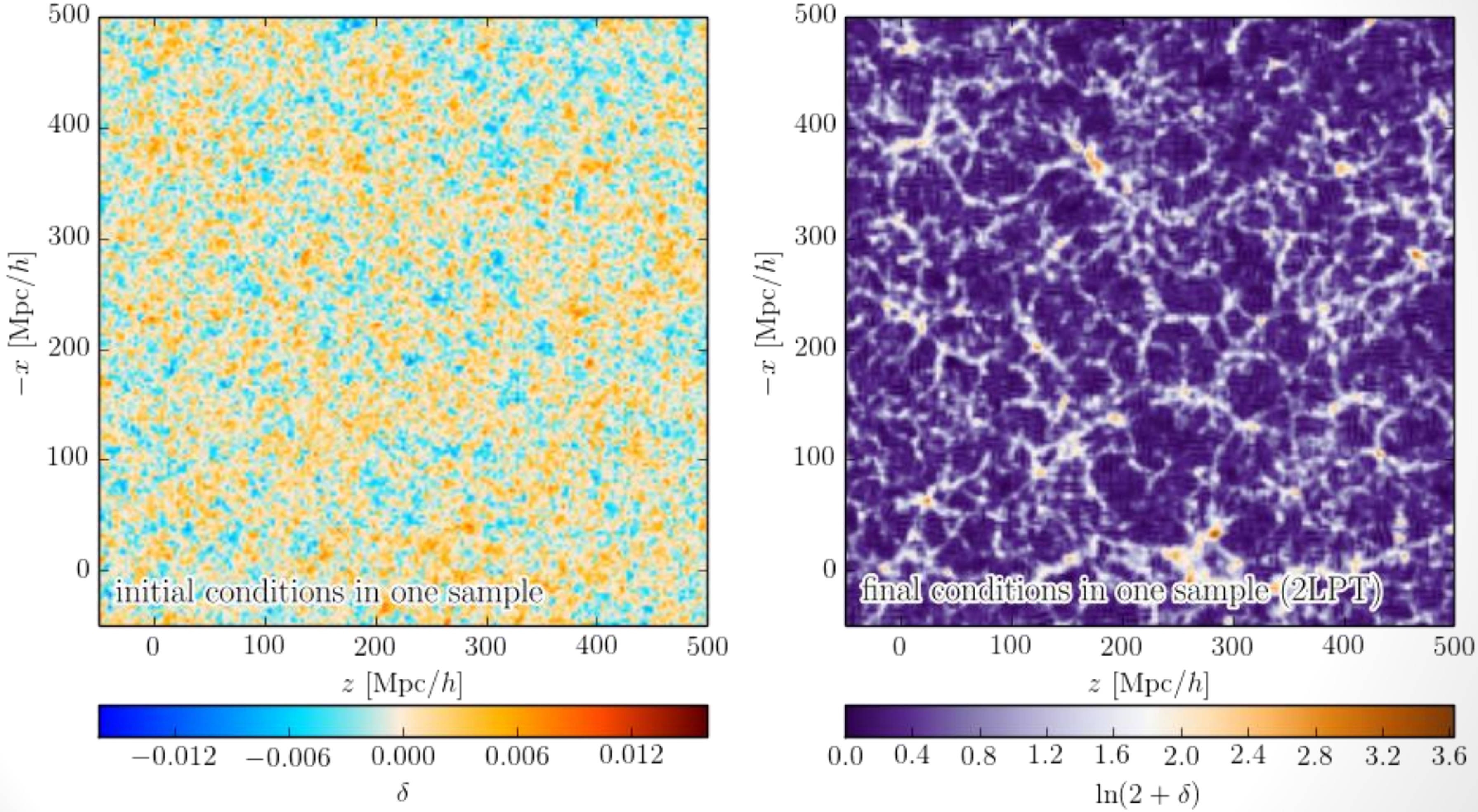
Thesis chapter 5

Bayesian chrono-cosmography from SDSS DR7



Data

Bayesian chrono-cosmography from SDSS DR7

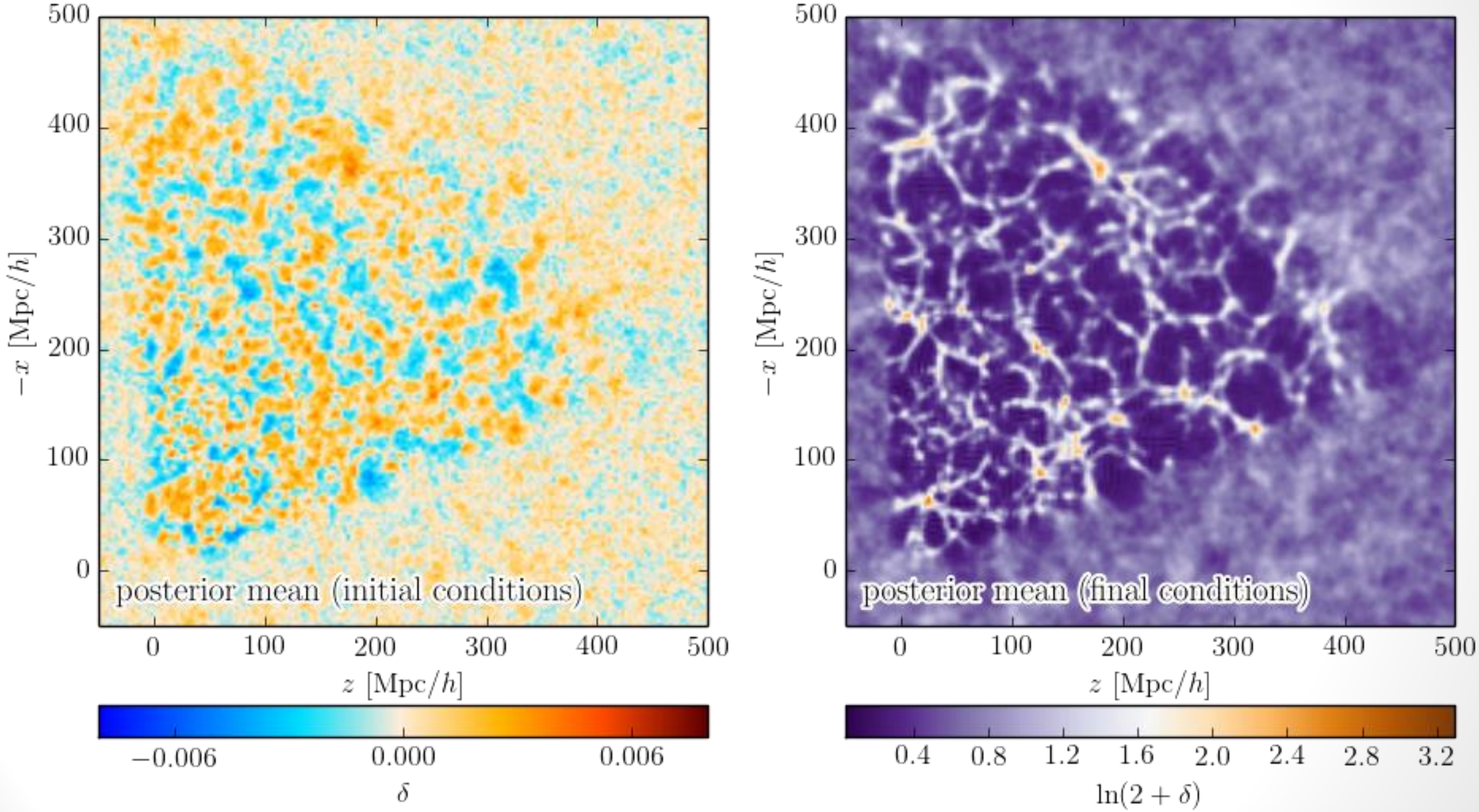


One sample

Jasche, FL & Wandelt 2015, arXiv:1409.6308

Thesis chapter 5

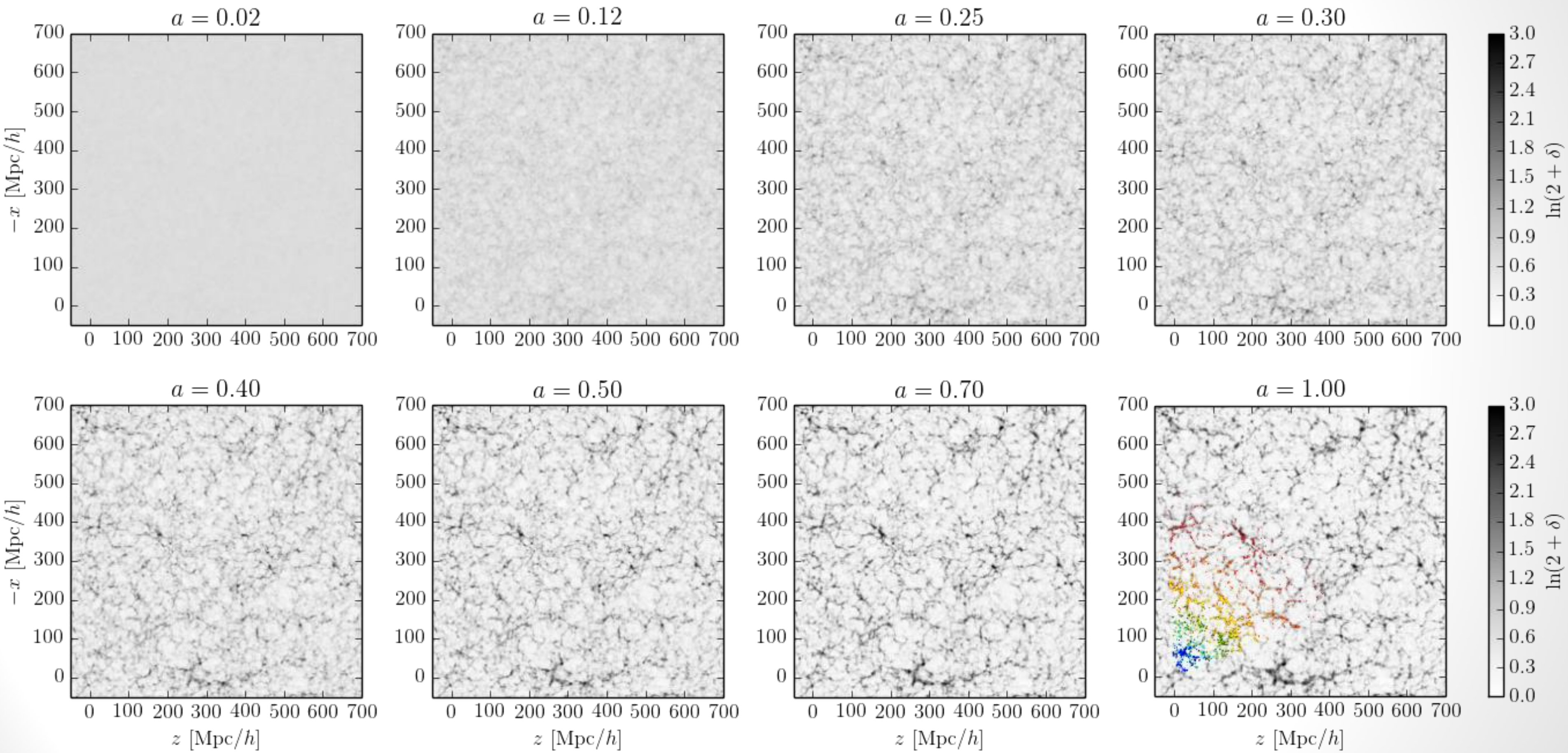
Bayesian chrono-cosmography from SDSS DR7



Posterior mean

Jasche, FL & Wandelt 2015, arXiv:1409.6308

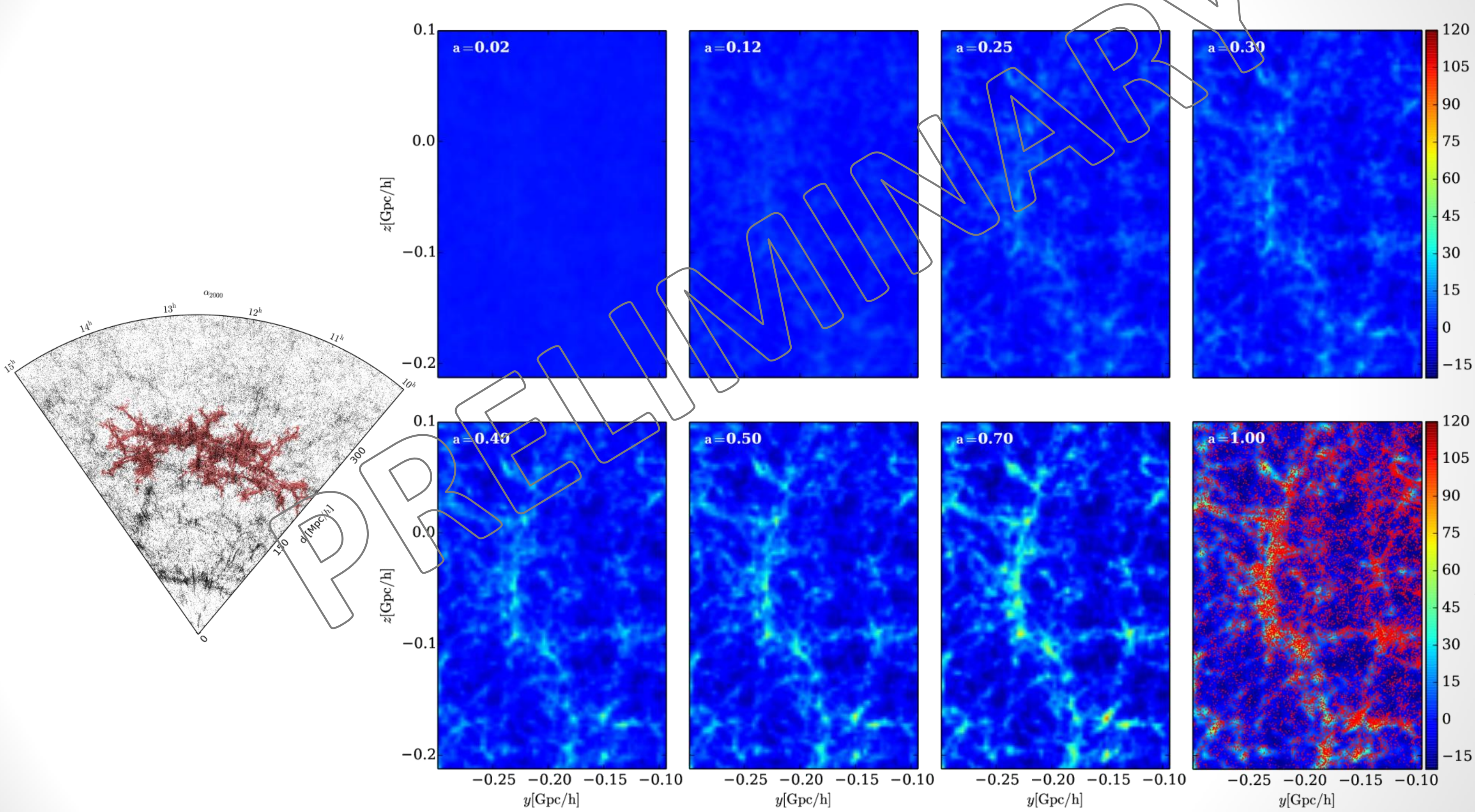
Evolution of cosmic structure



Jasche, FL & Wandelt 2015, arXiv:1409.6308

Thesis chapter 5

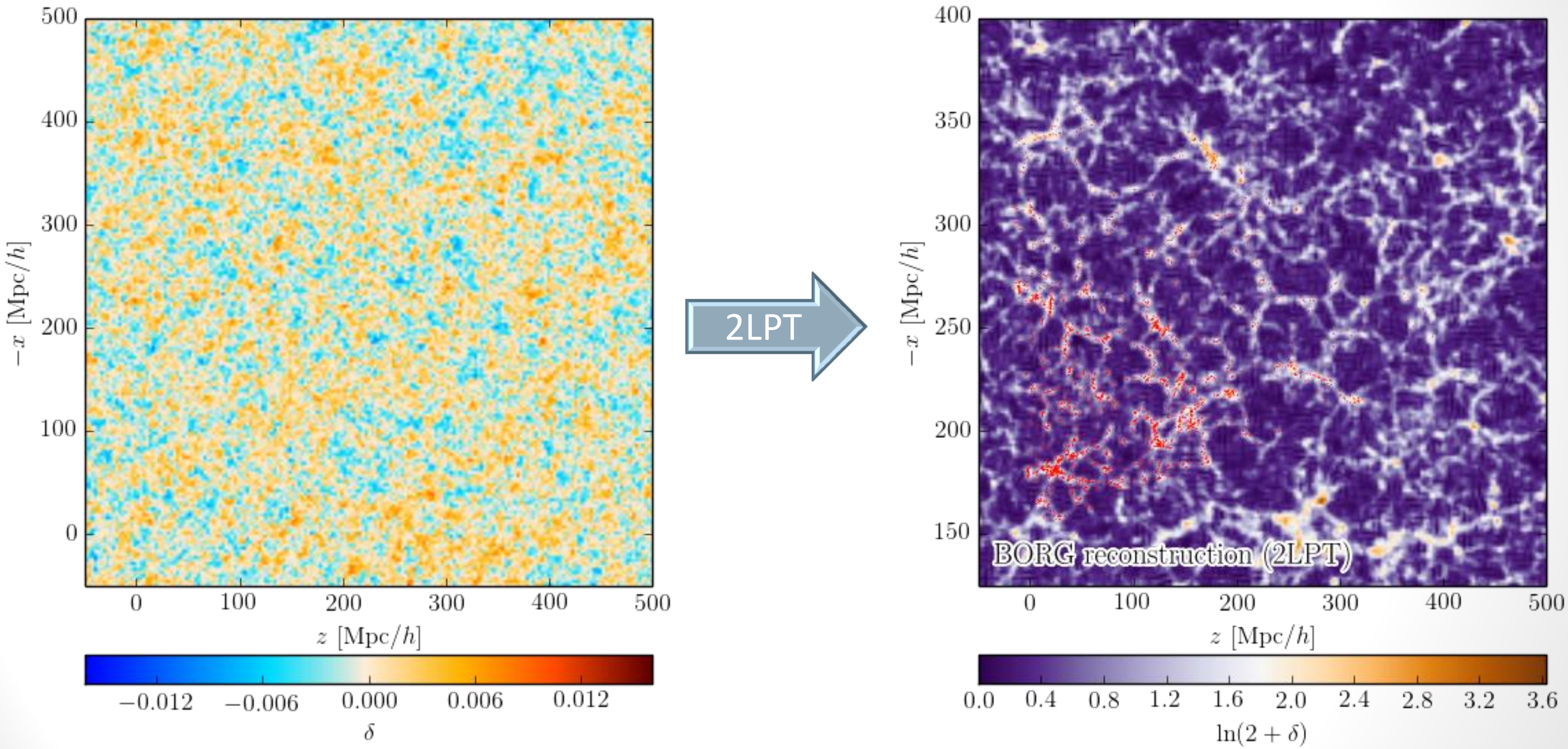
The formation history of the Sloan Great Wall



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

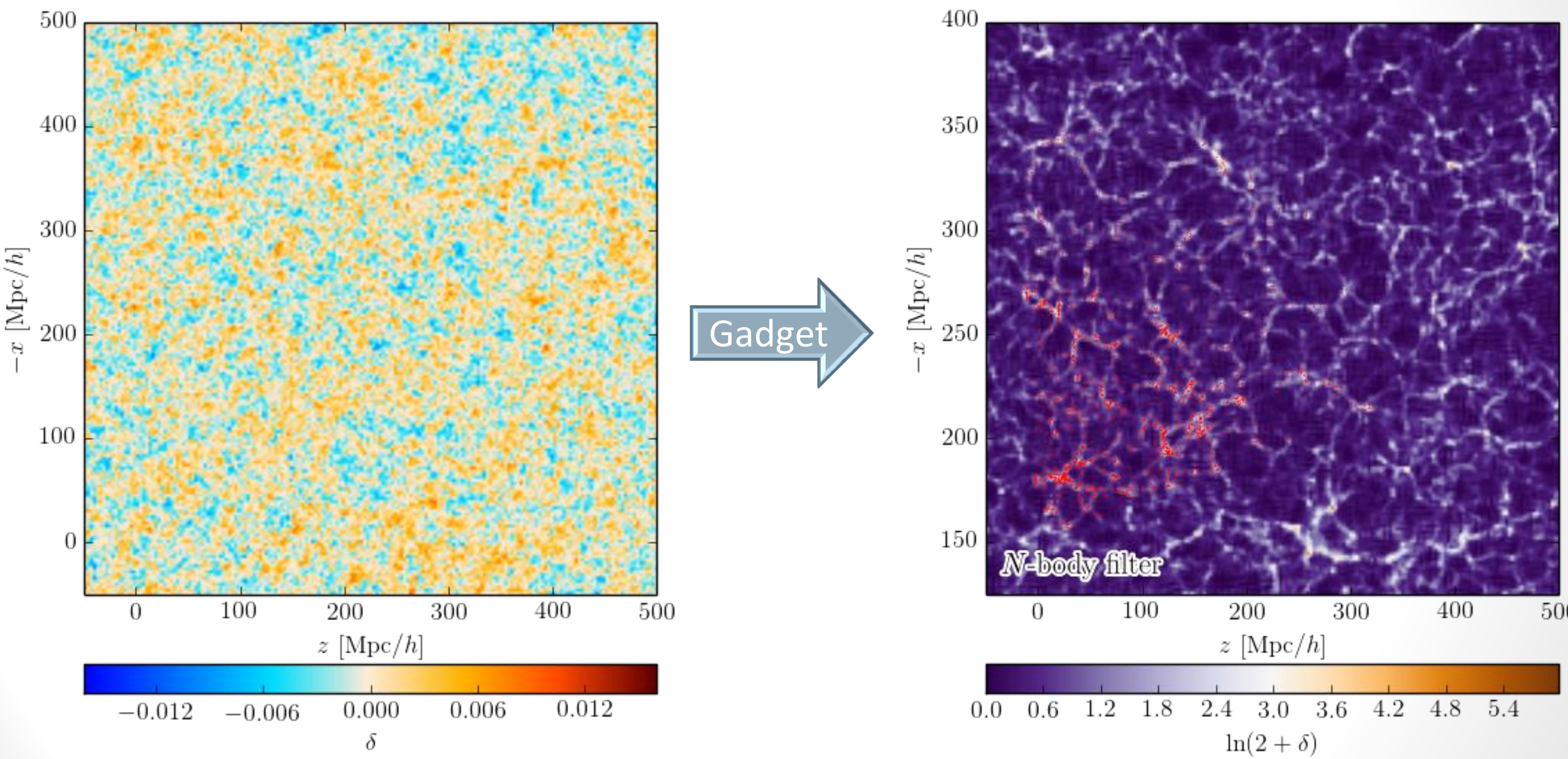
The Non-Linear Regime of Structure Formation

Non-linear filtering via constrained simulations



FL, Jasche, Sutter, Hamaus & Wandelt 2014, arXiv:1410.0355

Non-linear filtering via constrained simulations



FL, Jasche, Sutter, Hamaus & Wandelt 2014, arXiv:1410.0355

COLA: *COmoving Lagrangian Acceleration*

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

Tassev & Zaldarriaga 2012, arXiv:1203.5785

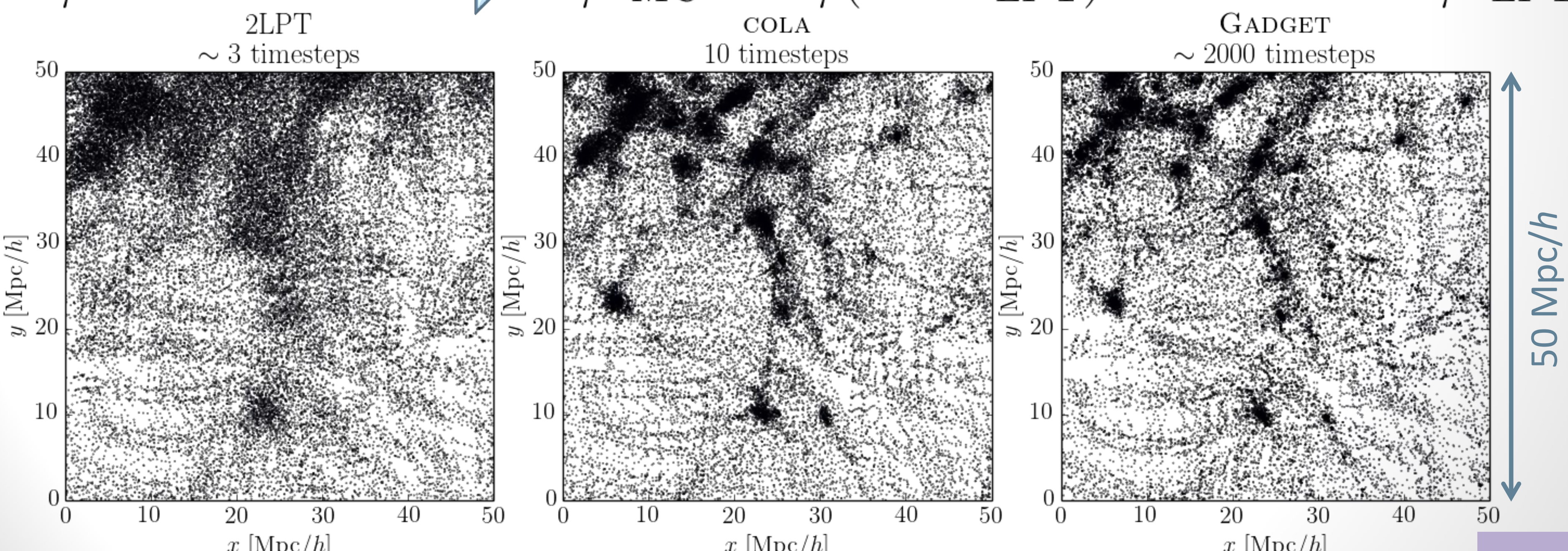
- Time-stepping (omitted constants and Hubble expansion):

Standard:

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

Modified:

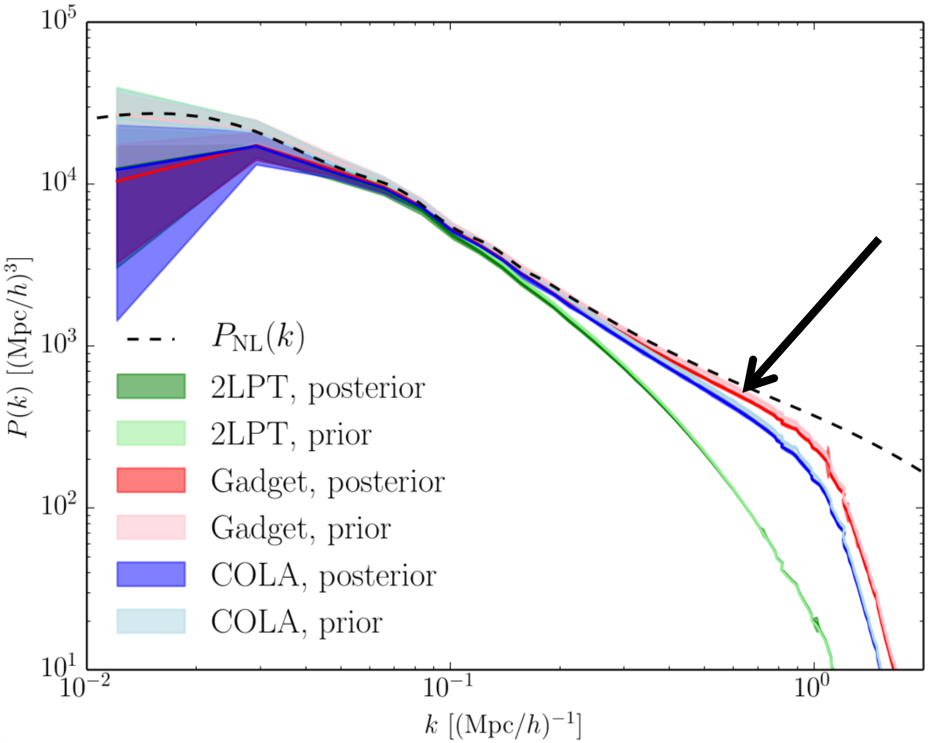
$$\partial_\tau^2 \mathbf{s}_{\mathrm{MC}} = \partial_\tau^2(\mathbf{s} - \mathbf{s}_{\mathrm{LPT}}) = -\nabla\Phi - \partial_\tau^2 \mathbf{s}_{\mathrm{LPT}}$$

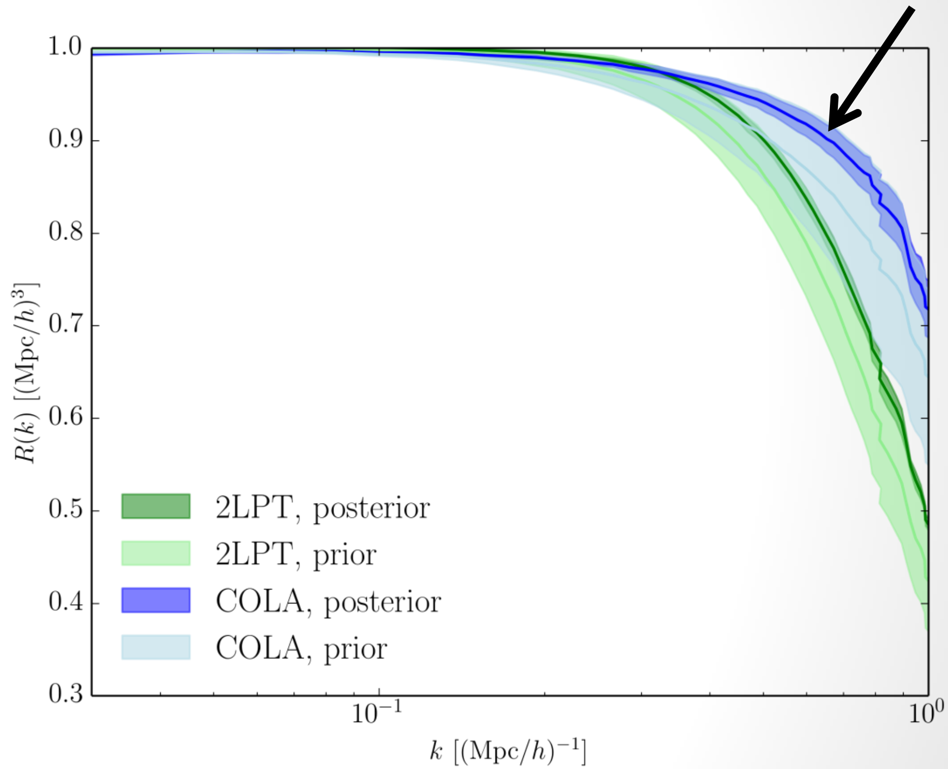


Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322

Thesis chapter 7

Non-linear filtering improves the fit

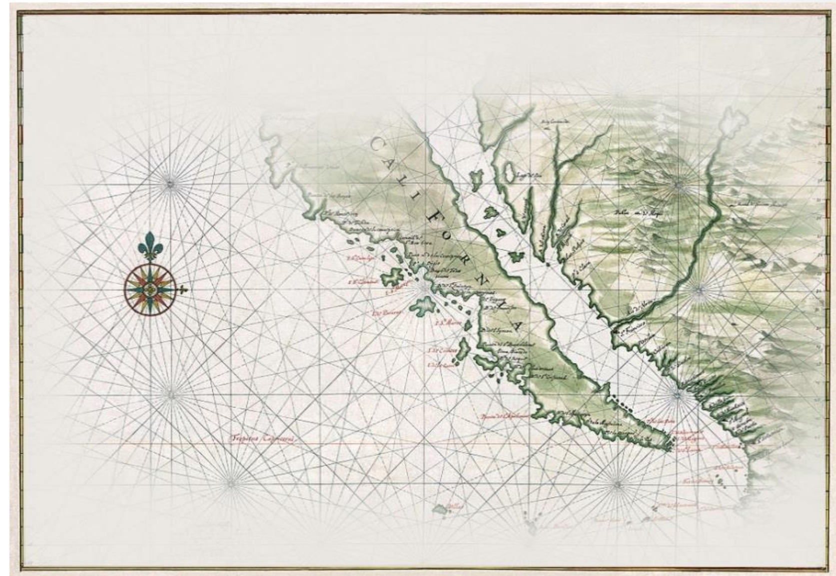




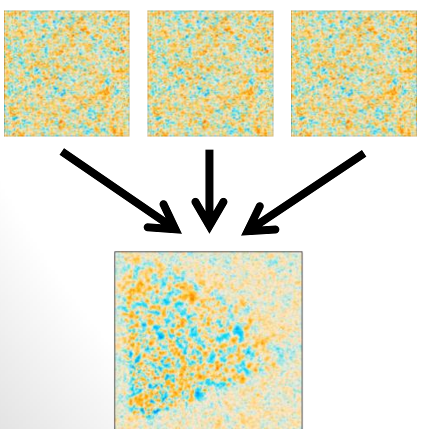
FL, Jasche, Sutter, Hamaus & Wandelt 2014, arXiv:1410.0355

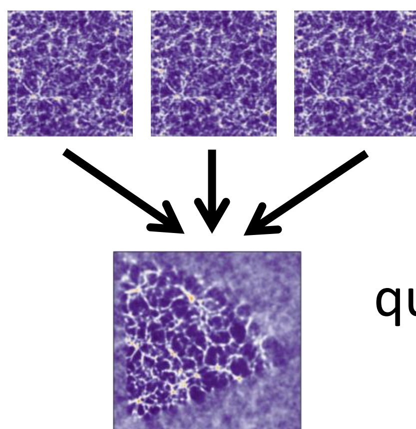
How is the Cosmic Web Woven?

Uncertainty quantification



Uncertainty quantification is crucial!





Can we **propagate** uncertainty quantification to **cosmic web analysis**?

Cosmic web classification procedures

void, sheet, filament, cluster?

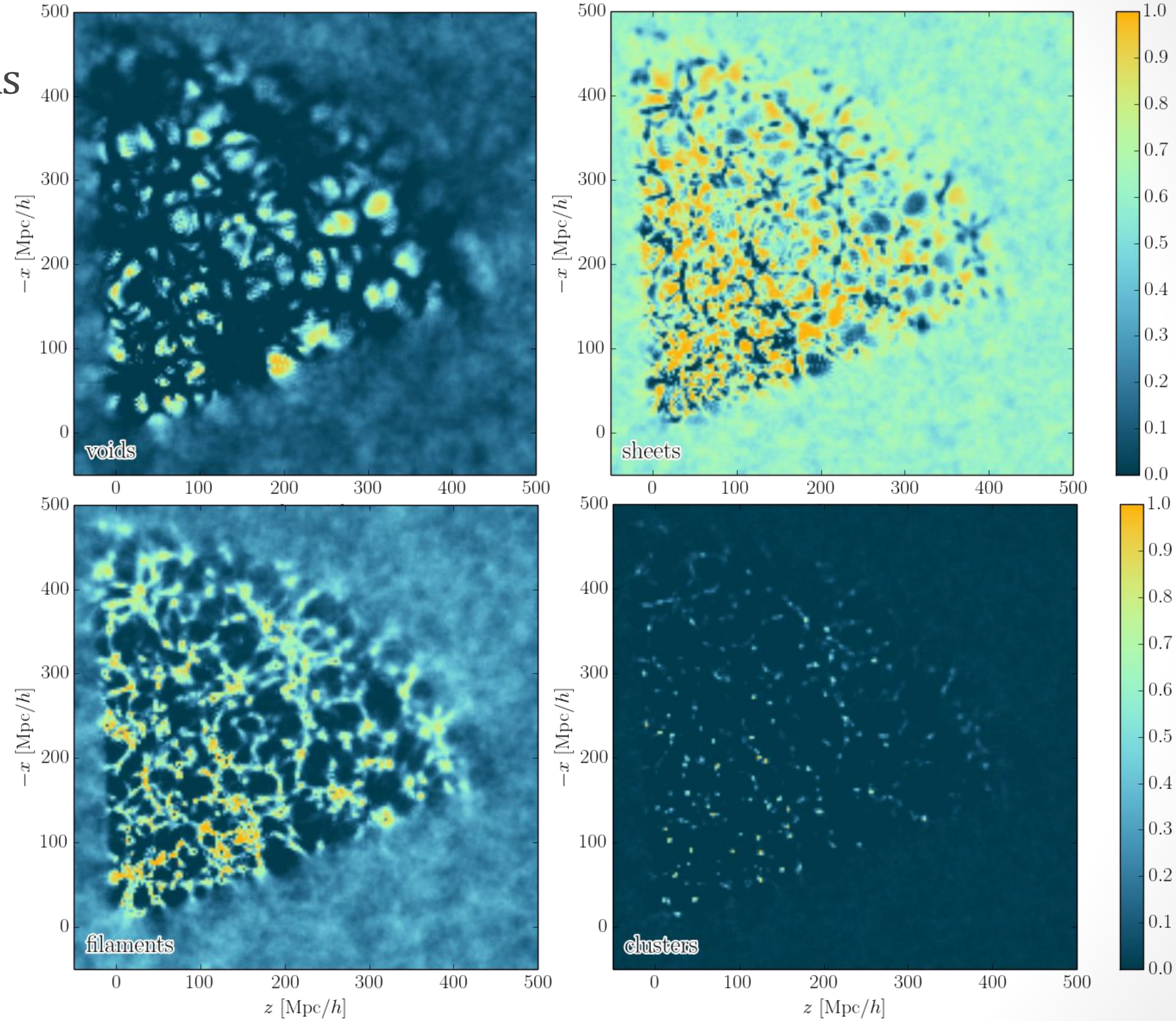
- The **T-web**:

uses the sign of μ_1, μ_2, μ_3: eigenvalues of the tidal field tensor, Hessian of the gravitational potential: $T_{ij}(\mathbf{x}) = \partial_i \partial_j \Phi(\mathbf{x})$

Hahn *et al*. 2007, arXiv:astro-ph/0610280

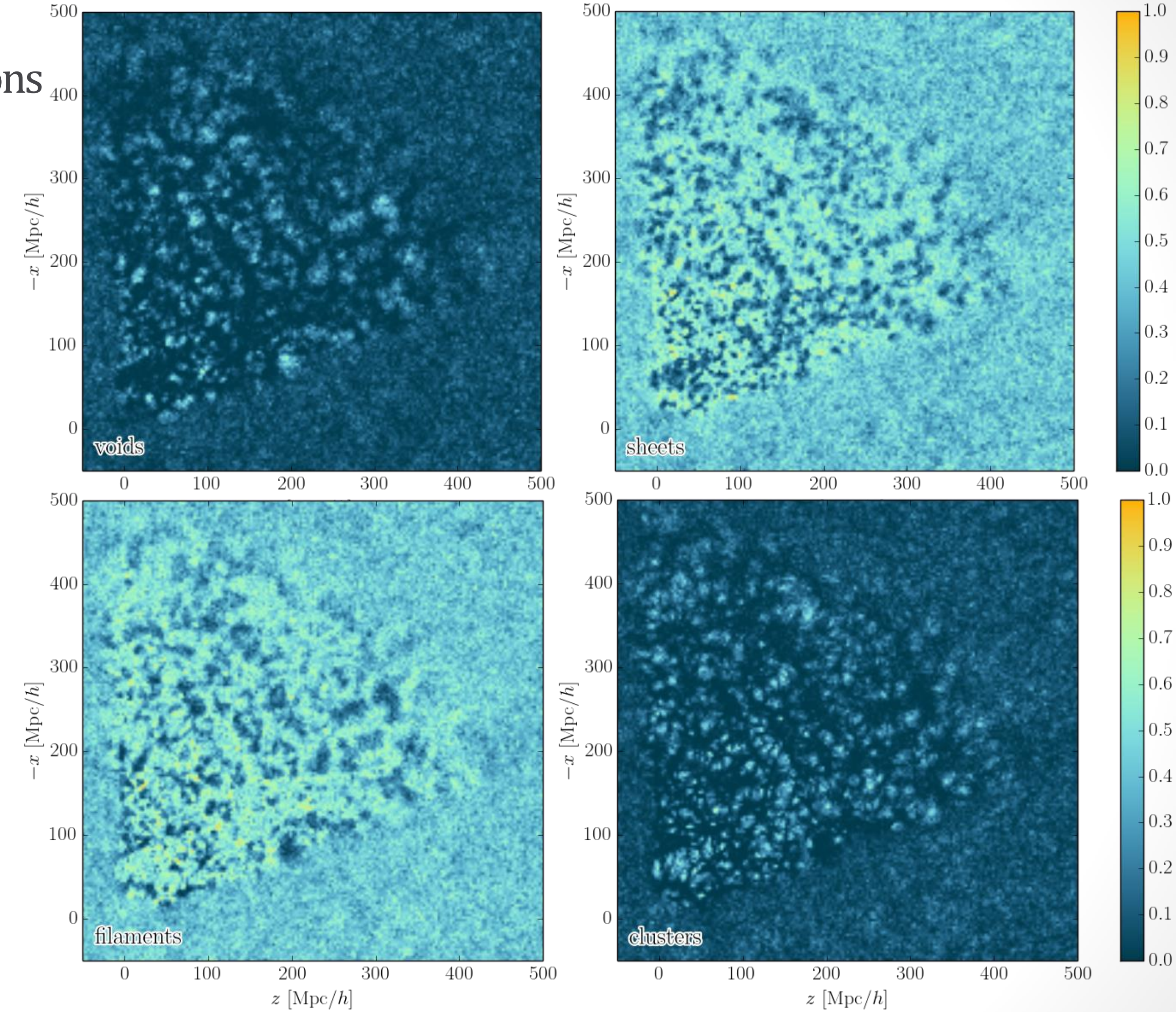
T-web structures inferred by BORG

Final conditions



T-web structures inferred by BORG

Initial conditions



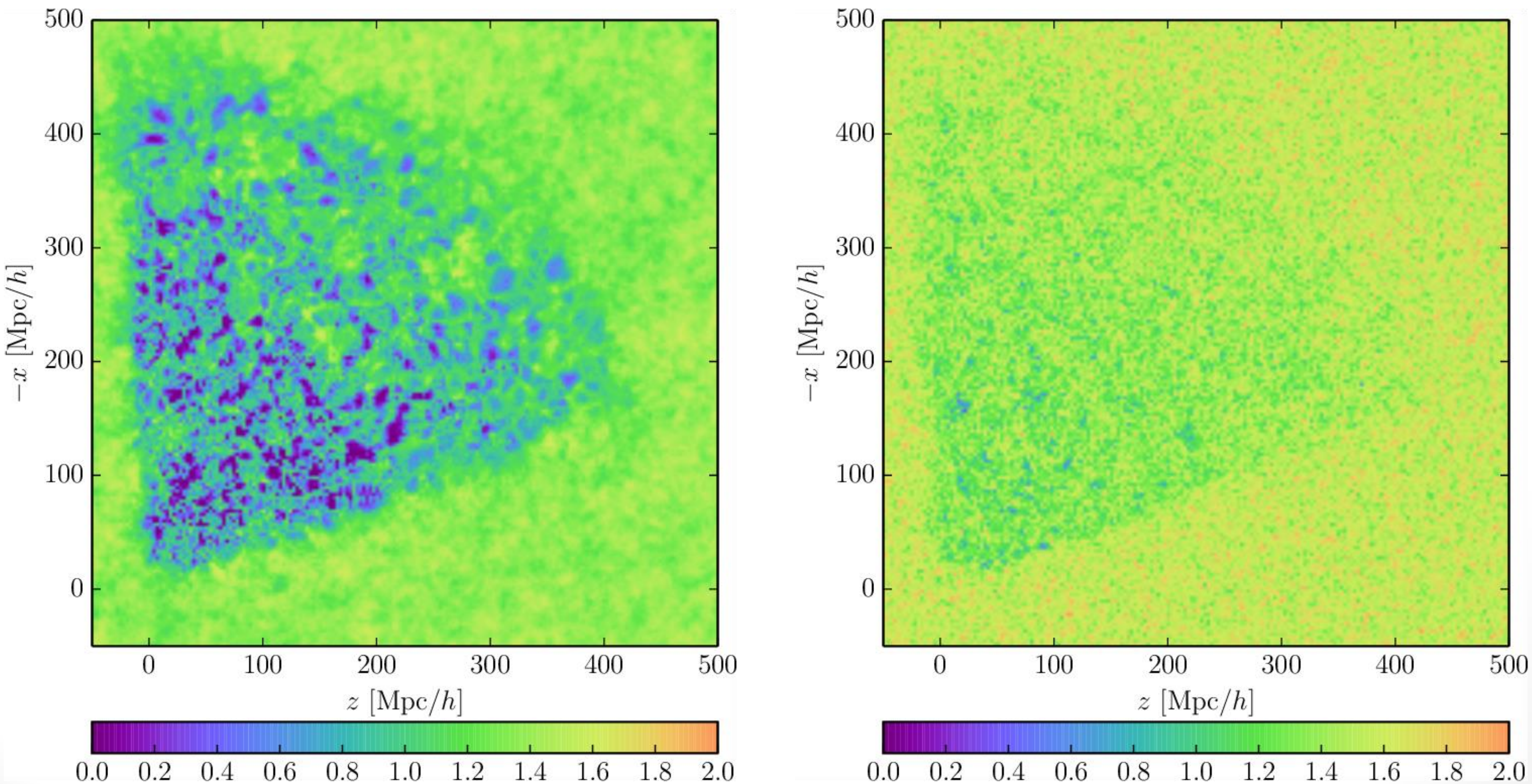
FL, Jasche & Wandelt 2015, arXiv:1502.02690

Thesis chapter 9

Entropy of the structure types posterior

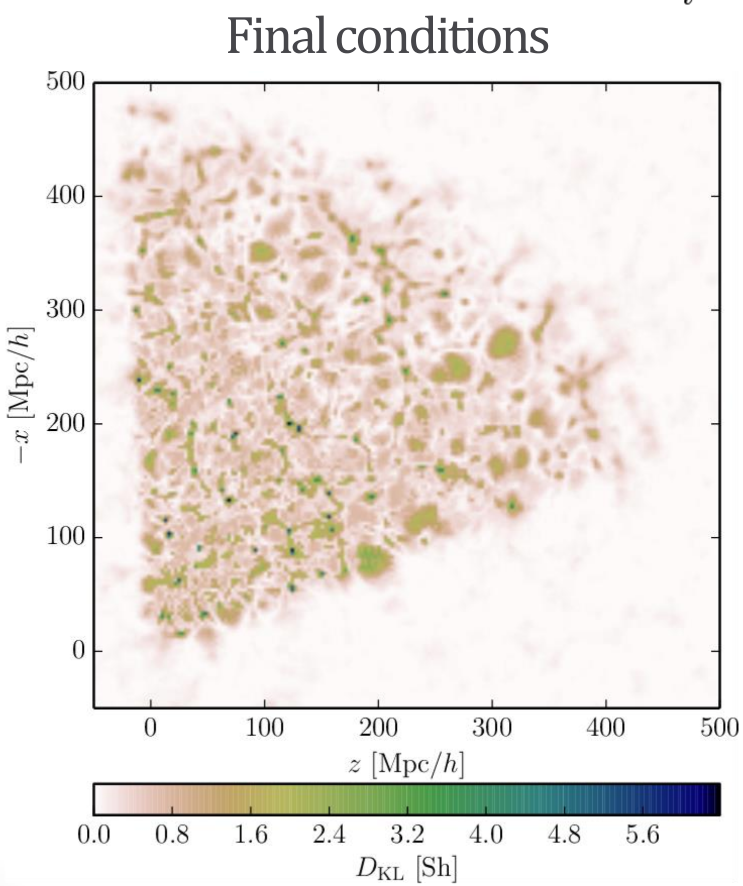
$$H\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)\right] \equiv - \sum_{i=0}^{3} \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d) \log_2(\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d))$$ in shannons (Sh)

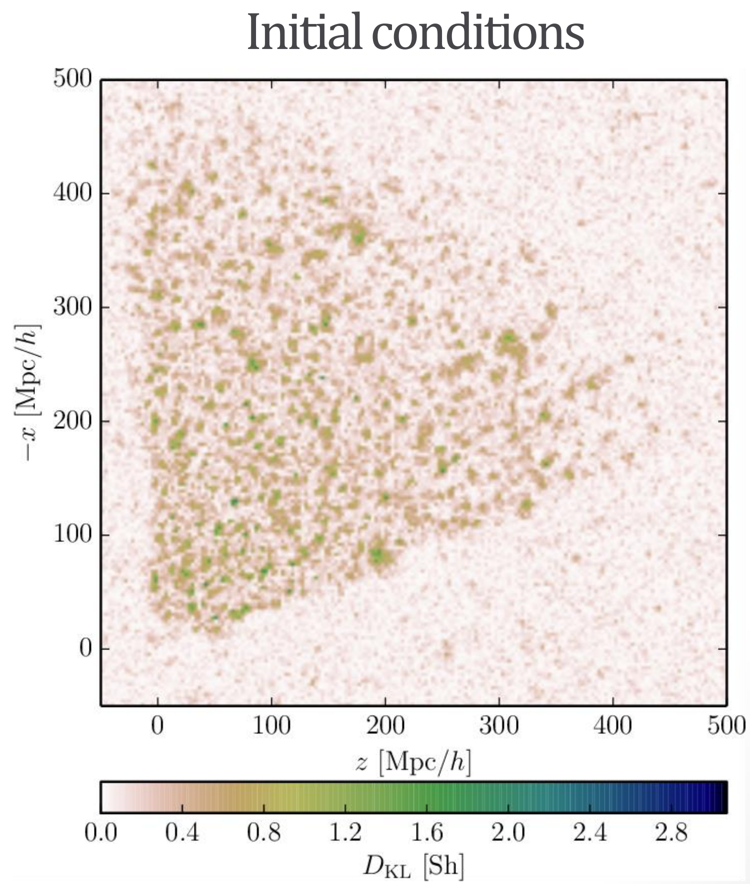
Final conditions | Initial conditions



How much did the data surprise us?

$$D_{\mathrm{KL}}(\mathcal{P}(\mathrm{T}(\vec{x_k})|d)||\mathcal{P}(\mathrm{T})) \equiv \sum_i \mathcal{P}(\mathrm{T}_i(\vec{x_k})|d) \log_2 \left(\frac{\mathcal{P}(\mathrm{T}_i(\vec{x_k})|d)}{\mathcal{P}(\mathrm{T}_i)} \right)$$ in Sh





FL, Jasche & Wandelt 2015, arXiv:1502.02690

A decision rule for structure classification

- Space of "input features":

$$\{\mathrm{T}_0 = \text{void}, \mathrm{T}_1 = \text{sheet}, \mathrm{T}_2 = \text{filament}, \mathrm{T}_3 = \text{cluster}\}$$

- Space of "actions":

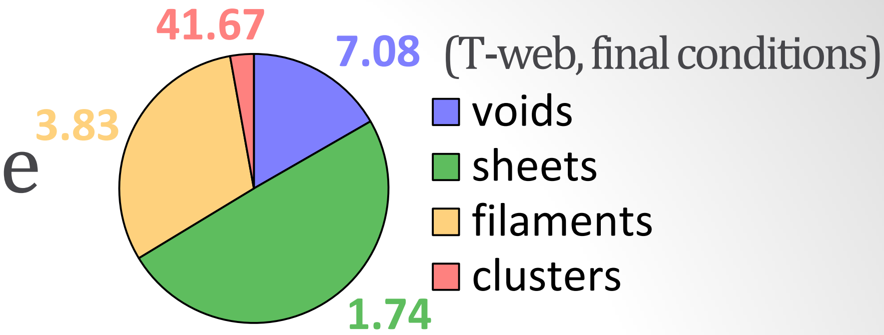
$$\{a_0 = \text{``decide void''}, a_1 = \text{``decide sheet''}, a_2 = \text{``decide filament''}, a_3 = \text{``decide cluster''}, a_{-1} = \text{``do not decide''}\}$$

A problem of **Bayesian decision theory**:
one should take the action which maximizes the utility

$$U(a_j(\vec{x}_k)|d) = \sum_{i=0}^{3} G(a_j|\mathrm{T}_i)\, \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)$$

- How to write down the gain functions?

Gambling with the Universe



- One proposal:

$$G(a_j|\mathrm{T}_i) = \begin{cases} \frac{1}{\mathcal{P}(\mathrm{T}_i)} - \alpha & \text{if } j \in [\![0,3]\!] \text{ and } i = j \\ -\alpha & \text{if } j \in [\![0,3]\!] \text{ and } i \neq j \\ 0 & \text{if } j = -1. \end{cases}$$

"Winning"

"Loosing"

"Not playing"

- Without data, the expected utility is

$$U(a_j) = 1 - \alpha \quad \text{if } j \neq 1$$

"Playing the game"

$$U(a_{-1}) = 0$$

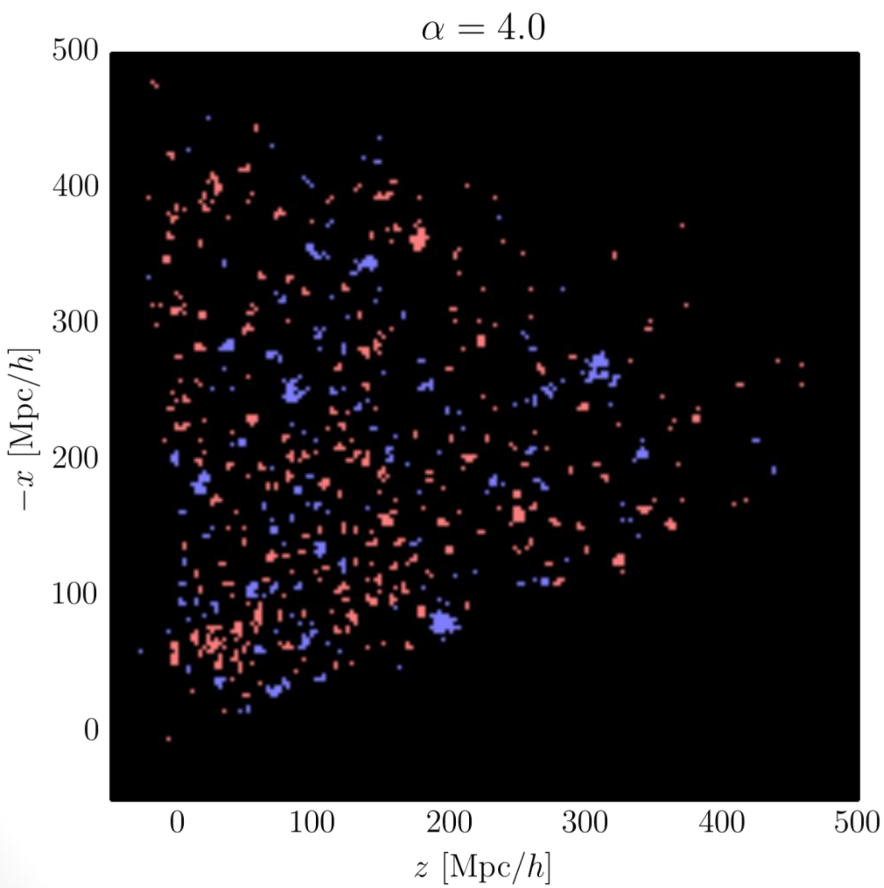
"Not playing the game"

- With $\alpha = 1$, it's a *fair game* ⇨ always play
  ⇨ "speculative map" of the LSS
- Values $\alpha > 1$ represent an *aversion for risk*
  ⇨ increasingly "conservative maps" of the LSS

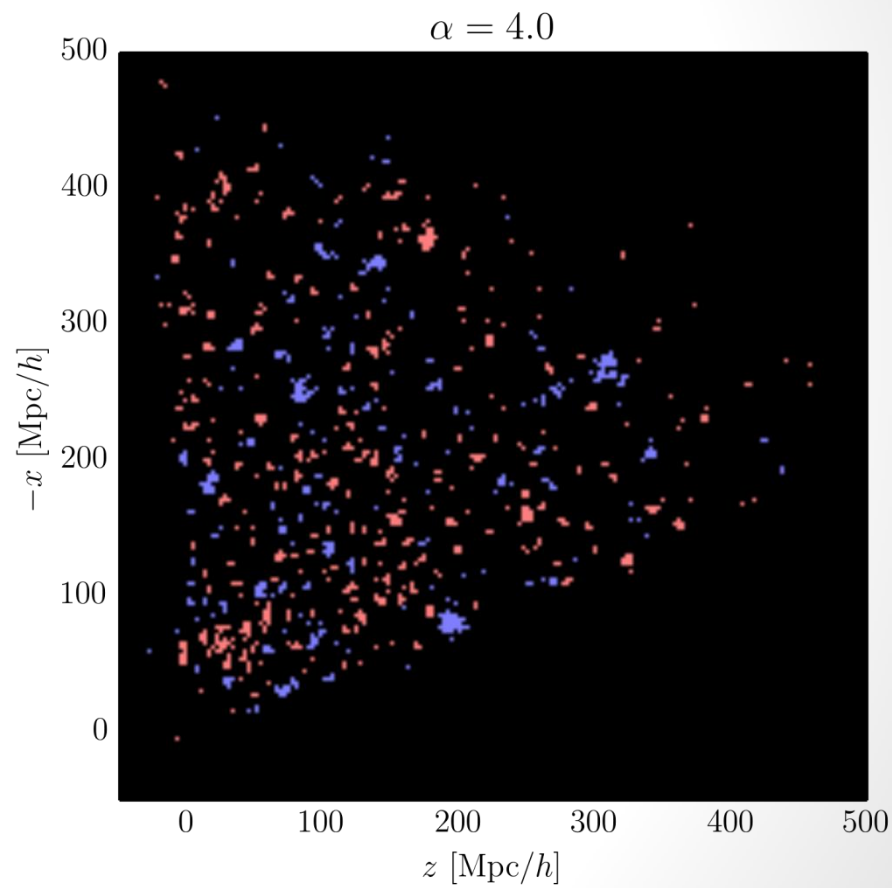
Playing the game…

voids
filaments
undecided
sheets
clusters

Final conditions



Initial conditions



FL, Jasche & Wandelt 2015, arXiv:1503.00730

Cosmic web classification procedures

void, sheet, filament, cluster?

- The **T-web**:

uses the sign of μ_1, μ_2, μ_3: eigenvalues of the tidal field tensor, Hessian of the gravitational potential: $T_{ij}(\mathbf{x}) = \partial_i \partial_j \Phi(\mathbf{x})$

Hahn *et al*. 2007, arXiv:astro-ph/0610280

- **DIVA**:

uses the sign of $\lambda_1, \lambda_2, \lambda_3$: eigenvalues of the shear of the Lagrangian displacement field: $R_{\ell m}(\mathbf{q}) = \partial_m \Psi_\ell(\mathbf{q})$

Lavaux & Wandelt 2010, arXiv:0906.4101

- **ORIGAMI** :

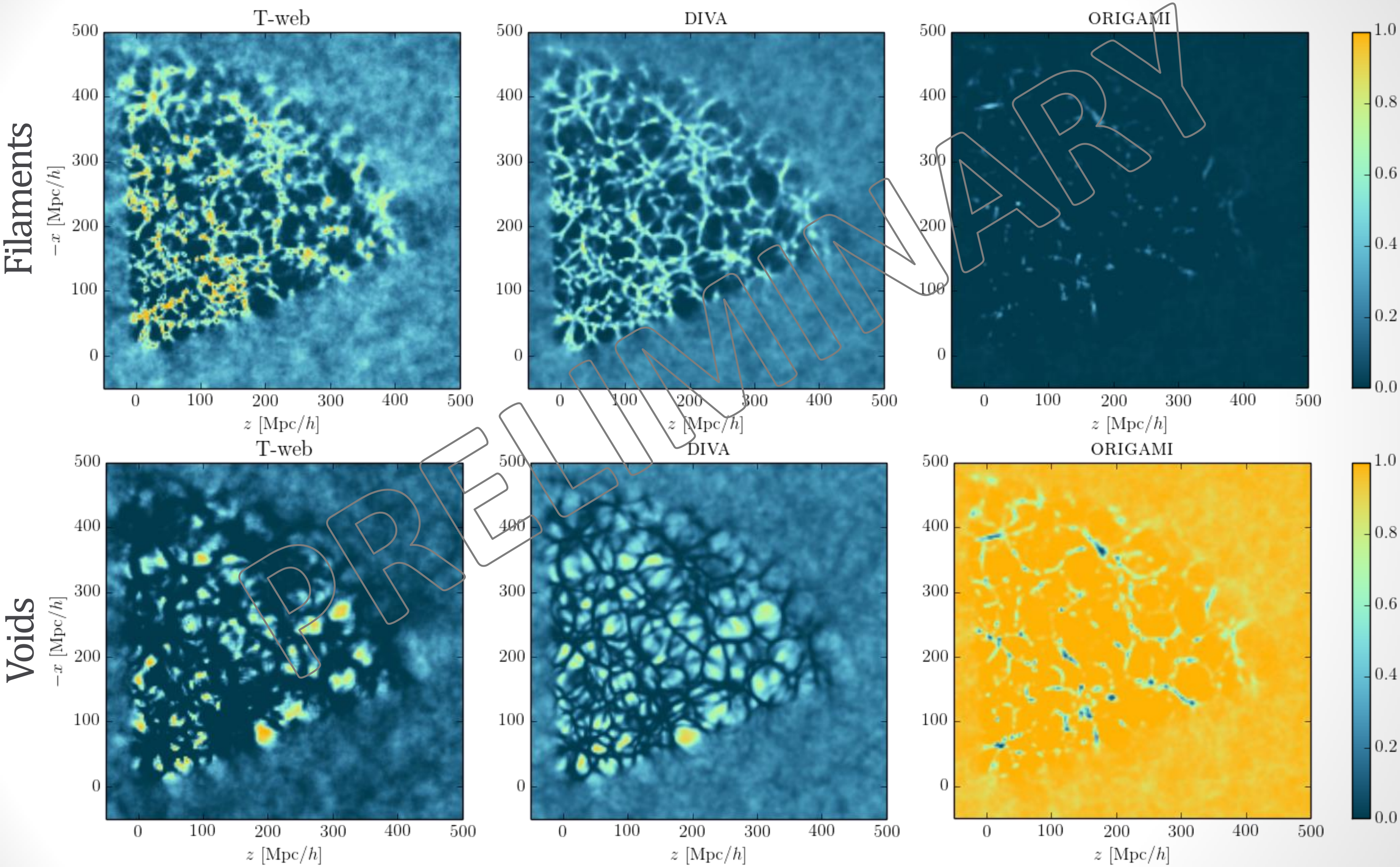
uses the dark matter "phase-space sheet" (number of orthogonal axes along which there is shell-crossing)

Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

Lagrangian classifiers

now usable in real data!

Comparing classifiers

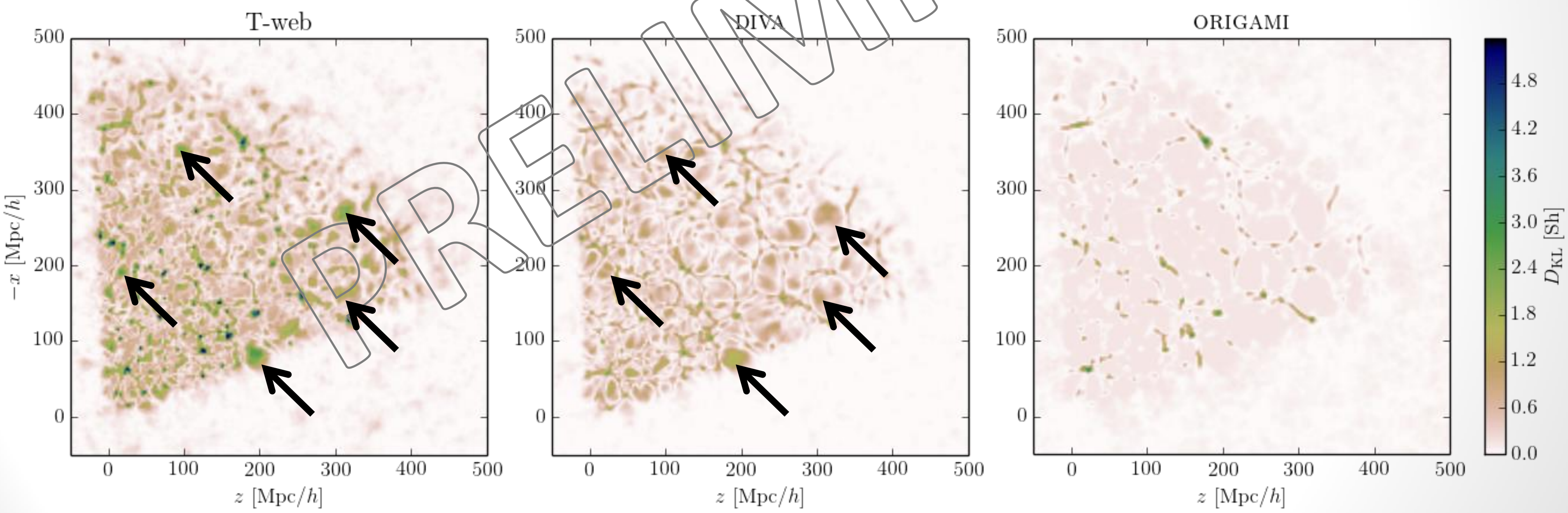


FL, Jasche, Lavaux & Wandelt, in prep.

What is the best classifier?

- One possible criterion, in analogy with Bayesian experimental design: **maximize the expected information gain**, $U(\mathrm{T})$

$$U(d,\mathrm{T}) = D_{\mathrm{KL}}(\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)||\mathcal{P}(\mathrm{T}))$$



Cosmic voids carry large **information gain**

FL, Jasche, Lavaux & Wandelt, in prep.

Hints from the Dark

Dark matter voids: pipeline

Why BORG?

Sparsity & Bias

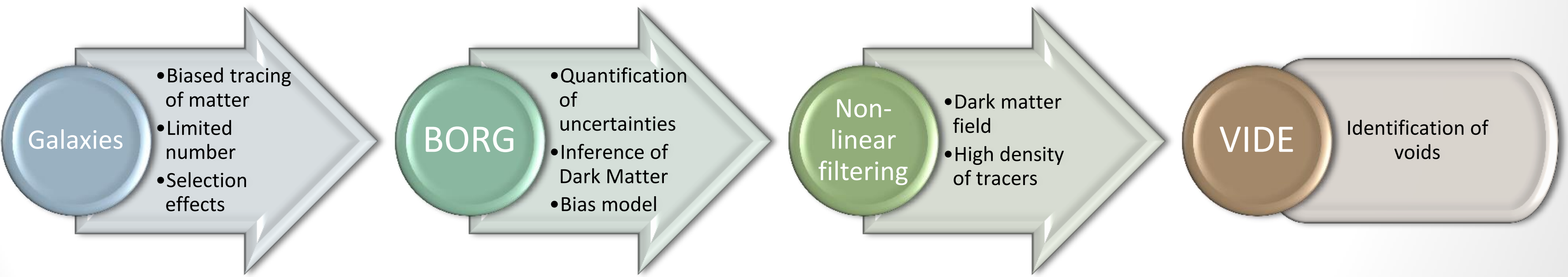
Sutter *et al*. 2013, arXiv:1309.5087

Sutter *et al*. 2013, arXiv:1311.3301

How?

VIDE toolkit: Sutter *et al*. 2015, arXiv:1406.1191
www.cosmicvoids.net

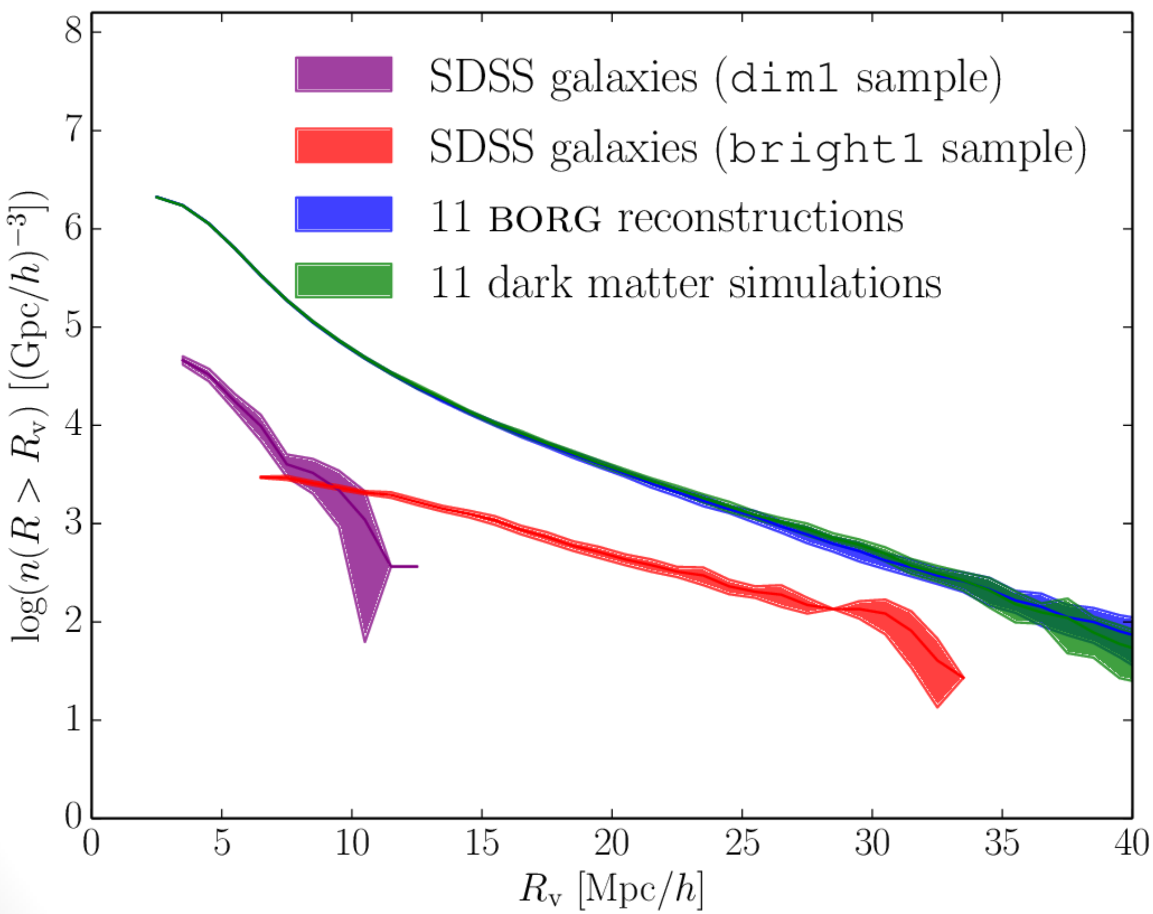
based on ZOBOV: Neyrinck 2007, arXiv:0712.3049

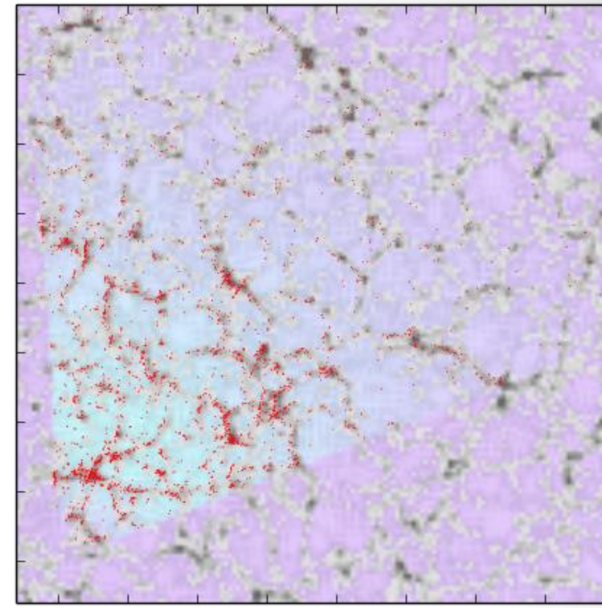


FL, Jasche, Sutter, Hamaus & Wandelt 2015, arXiv:1410.0355

BORG unveils many more voids

Void number function





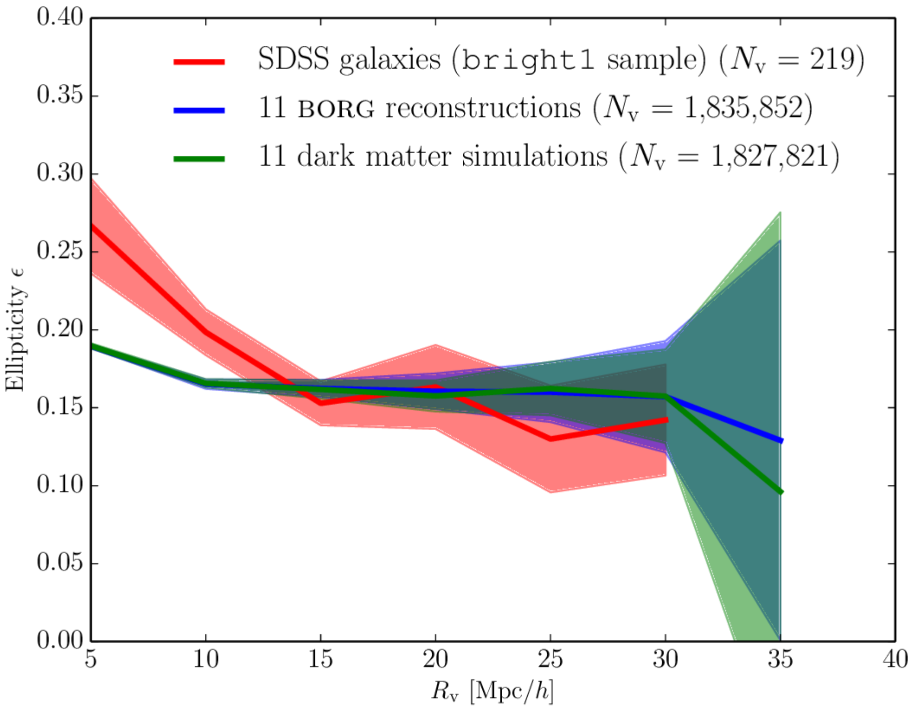
Voids in constrained regions only

Voids are **Poisson-dominated** objects:
10x more voids require 100x more galaxies!

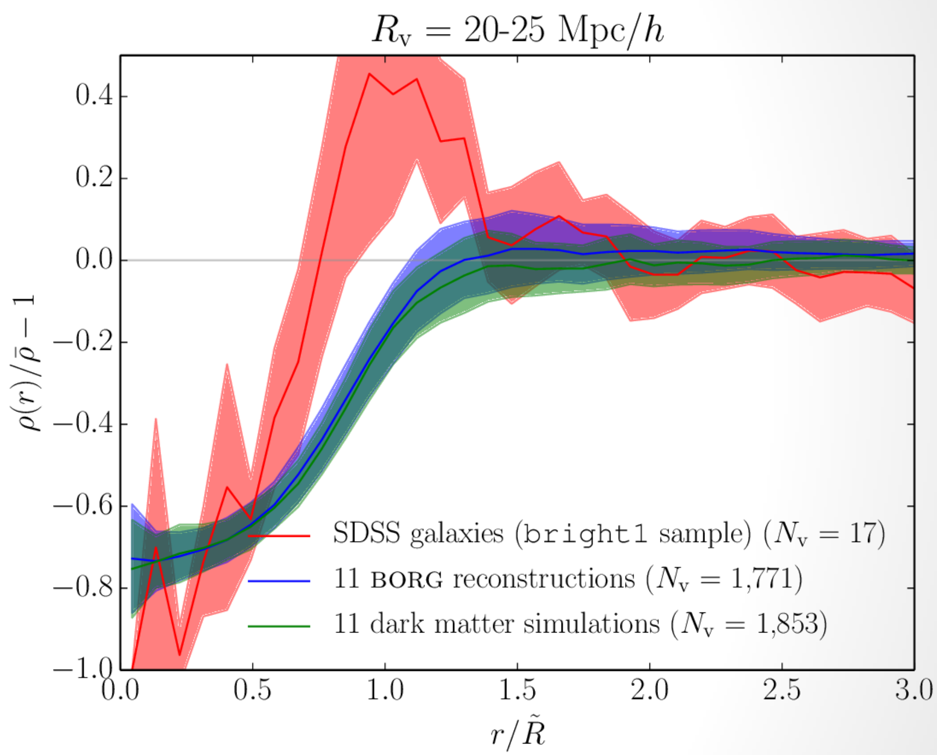
FL, Jasche, Sutter, Hamaus & Wandelt 2015, arXiv:1410.0355

Reduction of statistical uncertainty in voids catalogs

Ellipticity distribution



Radial density profile



All catalogs are publicly available at www.cosmicvoids.net for follow-up projects.

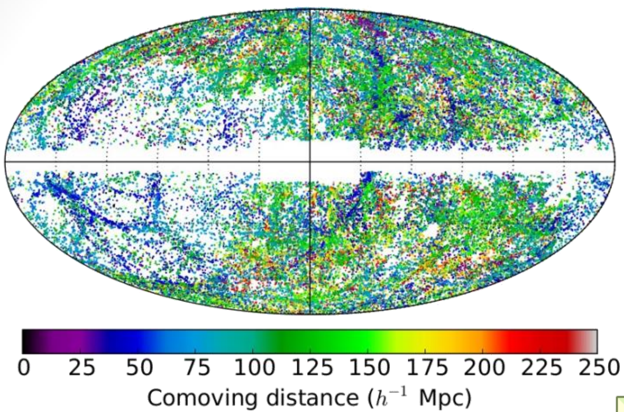
For example, these voids should have an **effect on CMB photons**…

FL, Jasche, Sutter, Hamaus & Wandelt 2015, arXiv:1410.0355

Thesis chapter 8

How to detect secondary effects in the Cosmic Microwave Background?

Producing LSS-CMB observables



2M++ catalog — Lavaux & Hudson 2011, arXiv:1105.6107

Initial conditions from BORG

Lavaux & Jasche 2015, arXiv:1509.05040

⇩

Filtering with COLA
Tassev, Zaldarriaga & Eisenstein, arXiv:1301.0322

Non-linear dynamics

Gravitational potential ⇩ Integrated Sachs-Wolfe (iSW) and Rees-Sciama (RS) effects	Momentum field ⇩ kinetic Sunyaev-Zel'dovich (kSZ) effect	Gas profiles in clusters ⇩ thermal Sunyaev-Zel'dovich (tSZ) effect

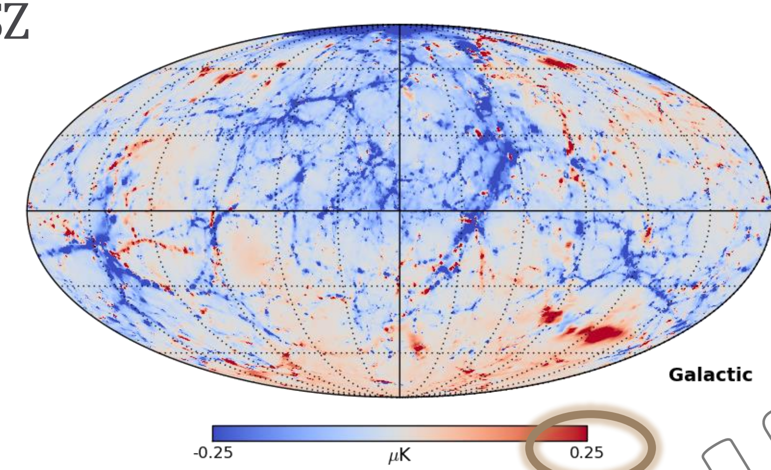
Raytracing algorithm
Cai *et al*. 2010, arXiv:1003.0974

kSZ/tSZ model
Lavaux, Afshordi & Hudson 2012, arXiv:1207.1721

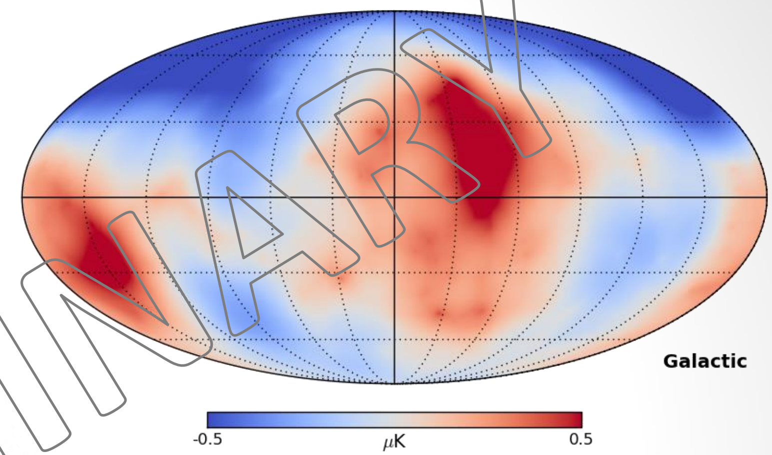
Better modeling yields higher Signal/Noise ratio

Templates for secondary effects in the CMB

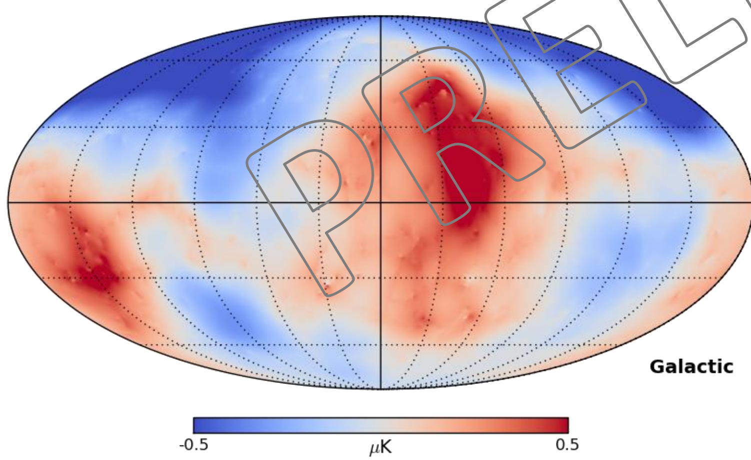
kSZ



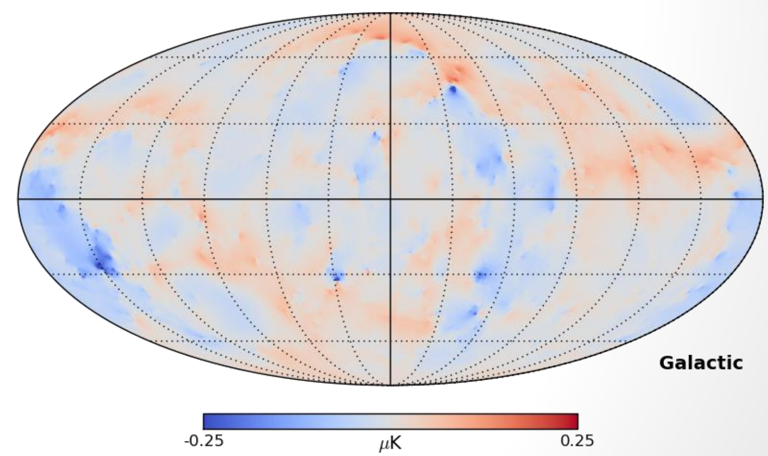
iSW



iSWRS



Only non-linear effects (iSWRS – iSW)



- Simulations in **one** BORG sample, raytraced from 0 to 100 Mpc/*h*
- The full posterior is available for Hierarchical Bayesian analysis

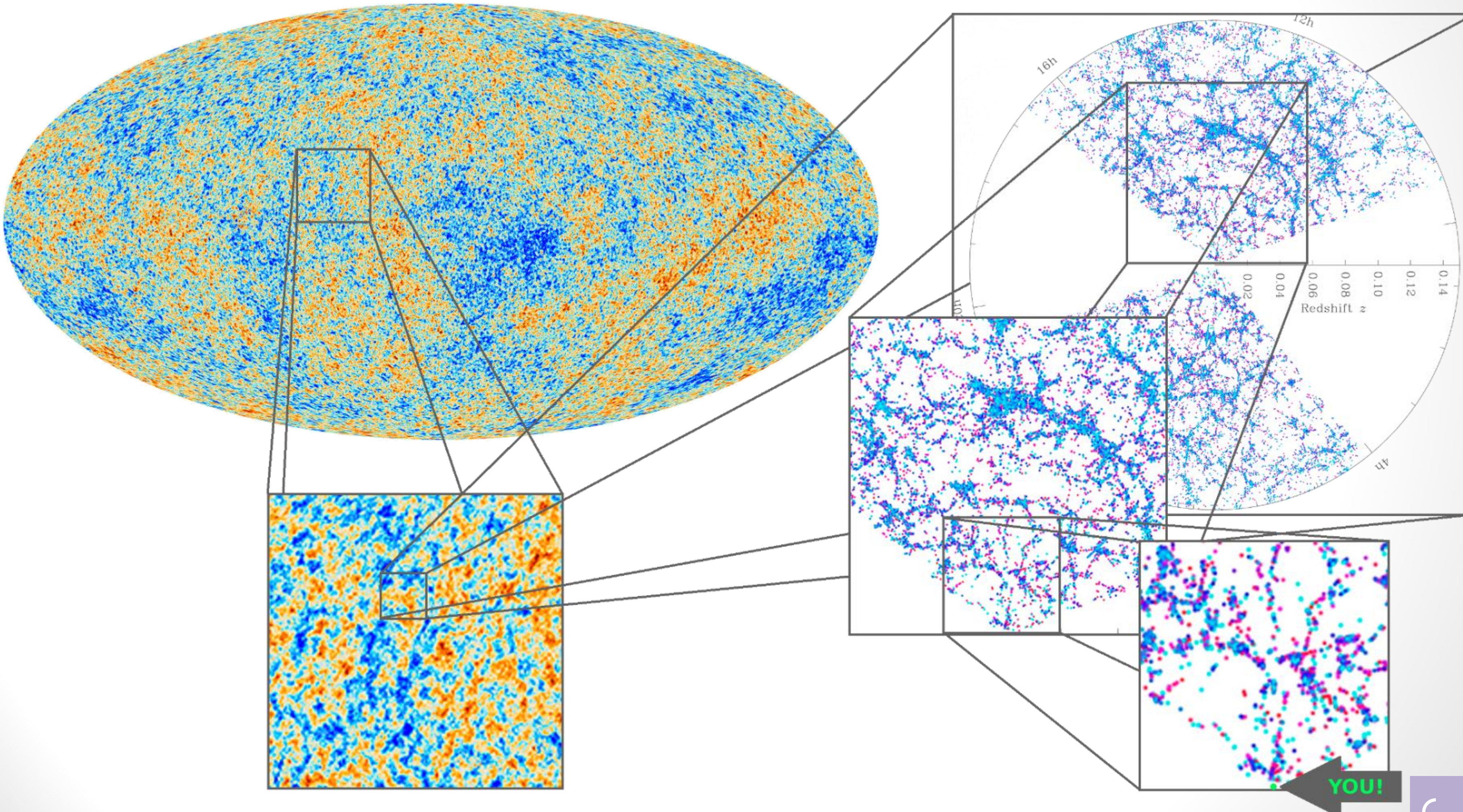
with G. Lavaux, J. Jasche, B. Wandelt

Summary & concluding thoughts

- A new method for principled analysis of galaxy surveys: **Bayesian large-scale structure inference**
  - Uncertainty quantification (noise, survey geometry, selection effects and biases)
  - Non-linear and non-Gaussian inference, with improving techniques
- Application to data: four-dimensional **chrono-cosmography**
  - Simultaneous analysis of the morphology and formation history of the large-scale structure
  - Physical reconstruction of the initial conditions
  - Characterization of the dynamic cosmic web underlying galaxies
  - Inference of cosmic voids at the level of the dark matter field
  - Cross-correlation of galaxy surveys and CMB data through kSZ/iSW/RS effects

Back to the big picture…

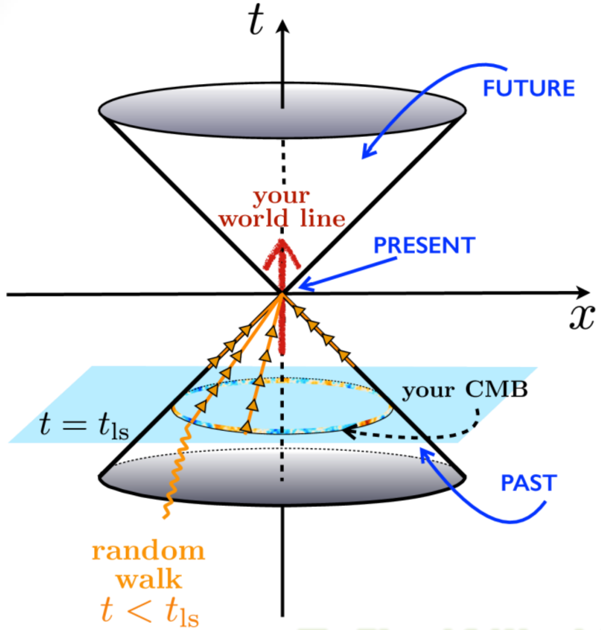
The large-scale structure is highly informative, but how informative is it?



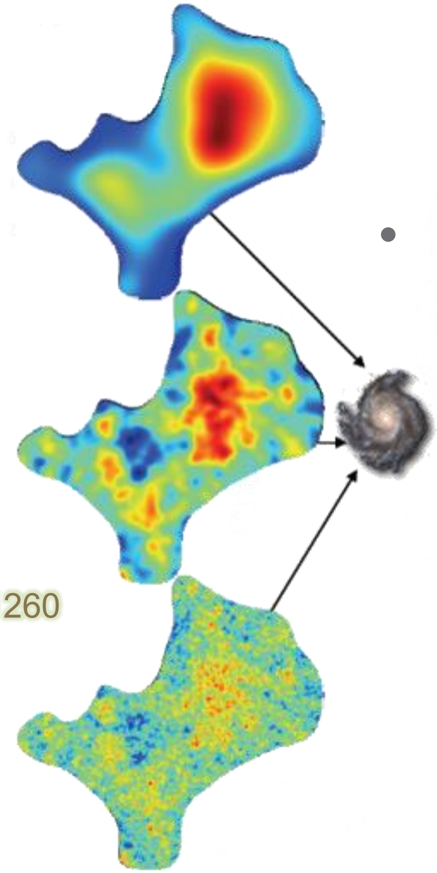
Planck collaboration (2013)

M. Blanton and the Sloan Digital Sky Survey (2010-2013)

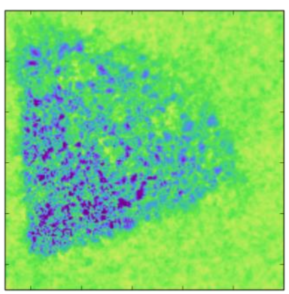
What can ultimately be learned from the LSS?



FL, Pisani & Wandelt 2014, arXiv:1403.1260



- Quantification of the information content of Lagrangian patches that collapse to form structures



$$H\left[S\right] = -\sum_i p_i \log_2 p_i$$

FL, Jasche & Wandelt 2015, arXiv:1502.02690

Neyrinck 2015, arXiv:1409.0057

- Link between information-theoretic and physical entropy
- Scale-dependent test of the degree of determinism in structure formation

Mapping the Universe: epilogue?

J. Cham – PhD comics