Bayesian large-scale structure inference and cosmic web analysis

Florent Leclercq

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

March 2nd, 2015

In collaboration with:

Héctor Gil-Marín (U. Portsmouth), Nico Hamaus (IAP), Jens Jasche (Excellence Cluster Universe Garching), Alice Pisani (LAM/IAP), Emilio Romano-Díaz (U. Bonn), Paul M. Sutter (Trieste/IAP/Ohio State U.), Benjamin Wandelt (IAP/U. Illinois), Matías Zaldarriaga (IAS Princeton)

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?
- Usually these problems are addressed in isolation.
- This talk:
  - A case for physical inference of four-dimensional dynamic states
  - A description of methodology and progress towards enriching the standard for analysis of galaxy surveys
  - A round trip: from theory to data, from data to theory

FL, Pisani & Wandelt 2014, arXiv:1403.1260

(Lectures Varenna 2013 and Paris École Doctorale for Astronomy and Astrophysics)

Outline

1. Bayesian Inference
2. Chrono-Cosmography
3. The Non-Linear Regime of Structure Formation
4. Cosmic Web Classification

1. Bayesian Inference

- Data assimilation with BORG
- The BORG SDSS run

J. Jasche, B. Wandelt, arXiv:1203.3639.
Bayesian physical reconstruction of initial conditions from large scale structure surveys

J. Jasche, B. Wandelt, arXiv:1306.1821.
Methods for Bayesian power spectrum inference with galaxy surveys

J. Jasche, F. Leclercq, B. Wandelt, arXiv:1409.6308.
Past and present cosmic structure in the SDSS DR7 main sample

Why Bayesian inference?

- Why do we need Bayesian inference?

Inference of signals = ill-posed problem

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

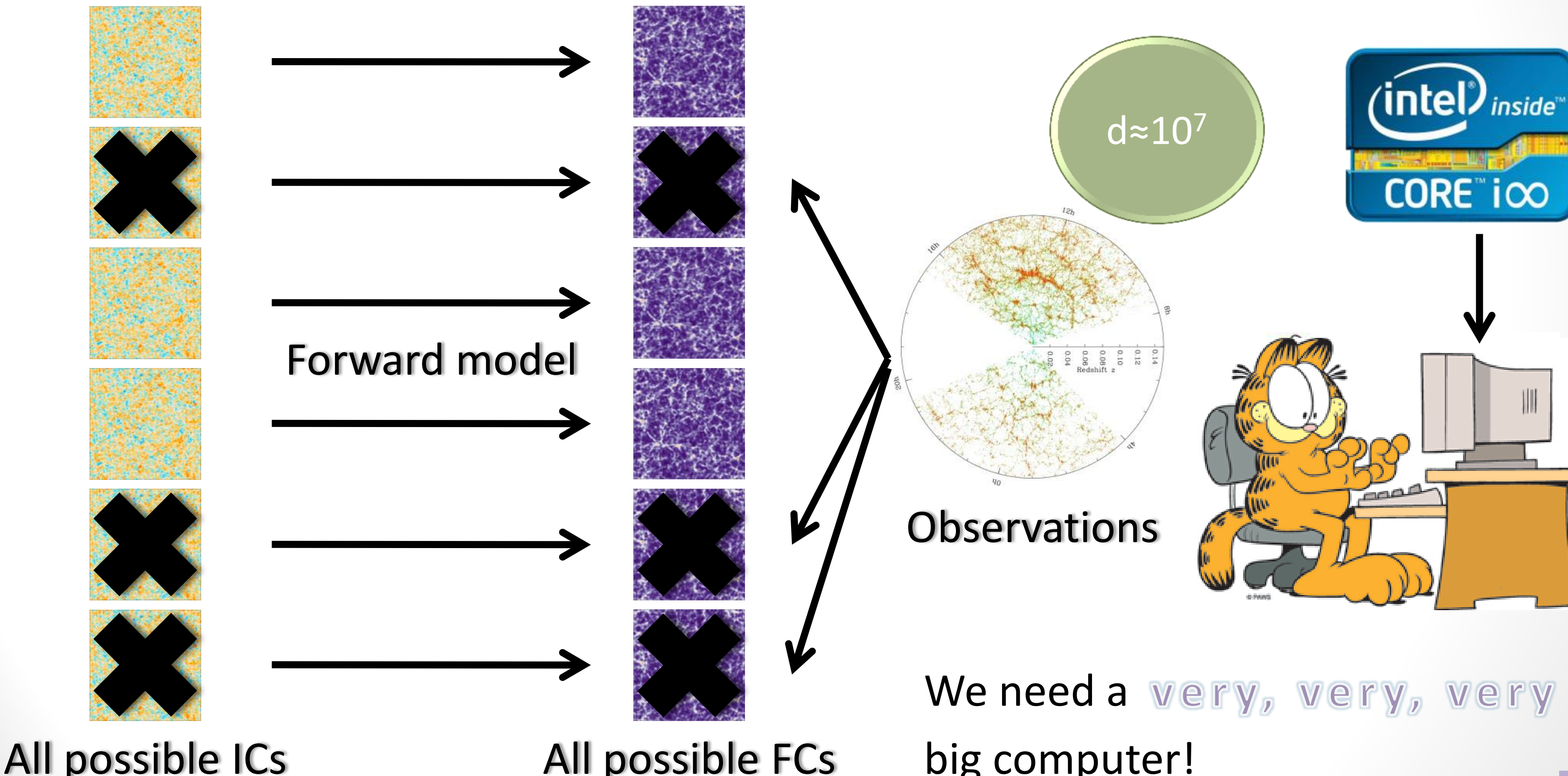
⇒ **No unique recovery is possible!**

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

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

Bayesian forward modeling: the ideal scenario

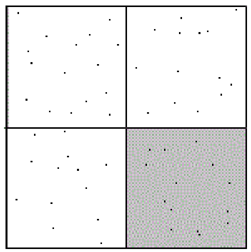
Forward model = N-body simulation + Halo occupation + Galaxy formation + Feedback + …



We need a very, very, very big computer!

(Parameter) Space: the final frontier

- The “curse of dimensionality” Bellman 1961

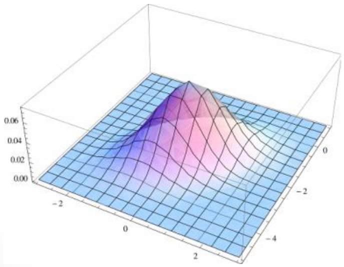


dimension	fraction of particles in quadrant of hypercube
1	$2^{-1} = 0.5$
10	$2^{-10} = 9.7 \times 10^{-4}$
100	$2^{-100} = 7.8 \times 10^{-31}$
1000	$2^{-1000} = 9.3 \times 10^{-302}$

Adding extra dimensions…

- Exponential increase of the number of particles needed for uniform sampling
- Exponential increase of sparsity given a fixed amount of particles

- High-dimensional probability distribution functions



⇒ Traditional sampling methods will fail

but gradients carry capital information

Hamiltonian Monte Carlo

- Use classical mechanics to solve statistical problems!
  - The potential: $\psi(\mathbf{x}) \equiv -\ln(\mathcal{P}(\mathbf{x}))$
  - The Hamiltonian: $H \equiv \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + \psi(\mathbf{x})$

$$(\mathbf{x},\mathbf{p}) \Rightarrow \left\{ \begin{array}{l} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{M}^{-1}\mathbf{p} \\ \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathbf{x}} = -\frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}} \end{array} \right\} \Rightarrow (\mathbf{x}',\mathbf{p}')$$

gradients

$$a(\mathbf{x}',\mathbf{x}) = \mathrm{e}^{-(H'-H)} = 1$$

acceptance ratio unity

- HMC beats the curse of dimensionality by:
  - Exploiting gradients
  - Using conservation of Hamiltonian

Duane *et al.* 1987

The LSS data model

- Gaussian prior for the initial density field: $\mathcal{P}(\delta^i|S) = \frac{\exp\left(\frac{1}{2}\sum_{kk'}\hat{\delta}^i_k S^{-1}_{kk'}\hat{\delta}^i_{k'}\right)}{|2\pi S|}$

- Poisson likelihood: $\mathcal{P}(N^\ell|\delta^i) = \prod_x \frac{\exp\left(-\lambda^\ell_x(\delta^i)\right)\left(\lambda^\ell_x(\delta^i)\right)^{N^\ell_x}}{N^\ell_x!}$

Poisson "intensity field": $\lambda^\ell_x(\delta^i) = R^\ell_x \bar{N}^\ell(1+\delta^\ell_{x,g}) = R^\ell_x \bar{N}^\ell \beta^\ell (1+\delta_{\mathrm{NL}}[\delta^i]_x)^{\alpha^\ell}$

$$\tilde{N}^\ell \equiv \bar{N}^\ell \beta^\ell$$

- The posterior distribution:

$$\mathcal{P}(\delta^i|\{N^\ell\},S) = \frac{\exp\left(\frac{1}{2}\sum_{kk'}\hat{\delta}^i_k S^{-1}_{kk'}\hat{\delta}^i_{k'}\right)}{|2\pi S|} \times \prod_{x,\ell} \frac{\exp\left(-R^\ell_x \tilde{N}^\ell(1+\delta_{\mathrm{NL}}[\delta^i]_x)^{\alpha^\ell}\right)\left(R^\ell_x \tilde{N}^\ell(1+\delta_{\mathrm{NL}}[\delta^i]_x)^{\alpha^\ell}\right)^{N^\ell_x}}{N^\ell_x!}$$

NGP-binned data

Covariance matrix

Survey response operator (mask, radial selection)

Noise level

Structure formation model (2LPT+CiC)

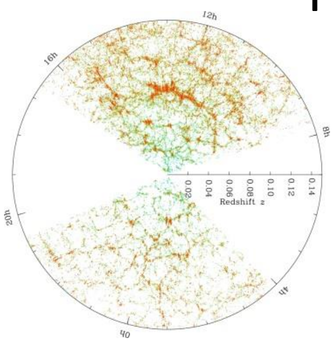
Bias parameters

Jasche & Wandelt 2013, arXiv:1203.3639
Jasche & Wandelt 2013, arXiv:1306.1821
Jasche, FL & Wandelt 2014, arXiv:1409.6308

BORG: *Bayesian Origin Reconstruction from Galaxies*

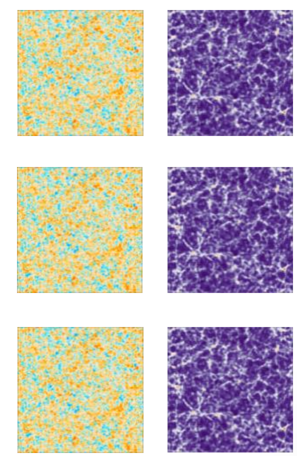
What makes the problem tractable:

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



Observations

BORG



Samples of possible 4D states

Jasche & Wandelt 2013, arXiv:1203.3639

see also:
Kitaura 2013, arXiv:1203.4184
Wang, Mo, Yang & van den Bosch 2013, arXiv:1301.1348

The BORG SDSS run

- 463,230 galaxies from the NYU-VAGC based on SDSS DR7
- Comoving cubic box of side length 750 Mpc/h, with periodic boundary conditions
- 256^3 grid, resolution 3 Mpc/h ⇒ ≈ 17 millions parameters
- 12,000 samples, four-dimensional maps
- ≈ 3 TB disk space
- 10 months wallclock time on 16-32 cores

Jasche, FL & Wandelt 2014, arXiv:1409.6308

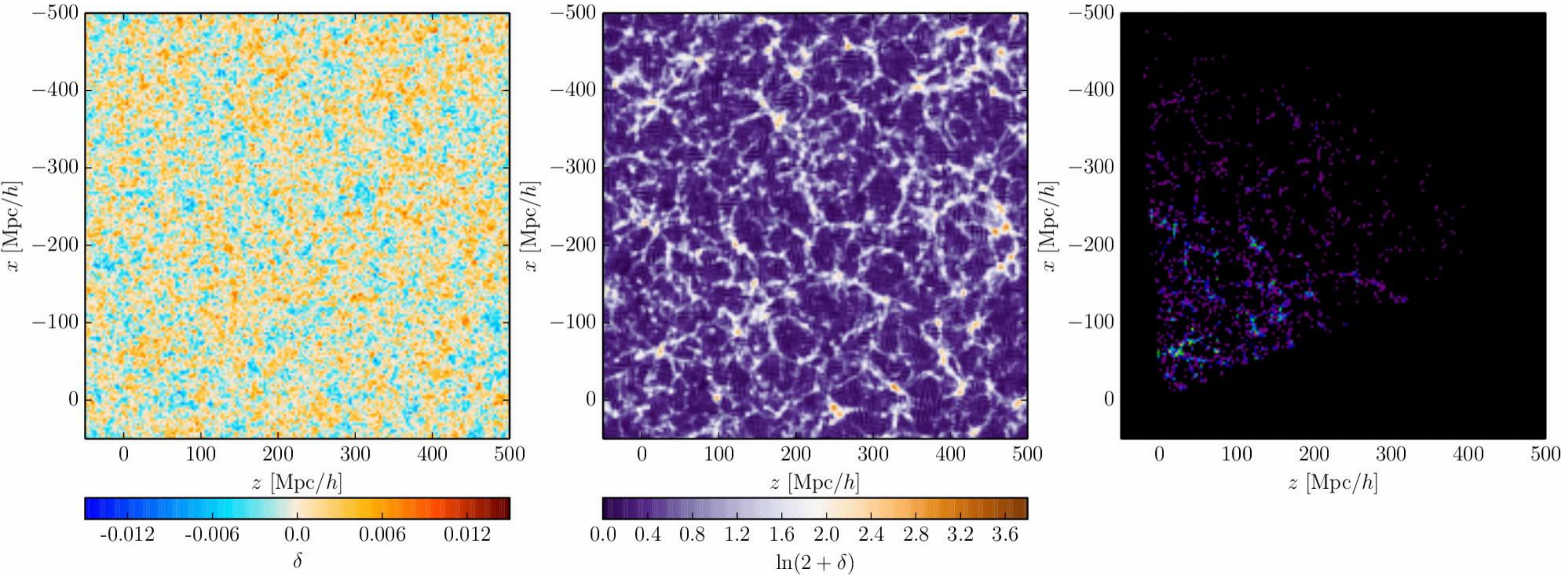
2. Chrono-Cosmography

- Past and present cosmic structure in the Sloan volume
- The Sloan Great Wall

J. Jasche, F. Leclercq, B. Wandelt, arXiv:1409.6308.
Past and present cosmic structure in the SDSS DR7 main sample

J. Jasche, E. Romano-Díaz, F. Leclercq, B. Wandelt, in prep.
The Sloan Great Wall through Bayesian eyes

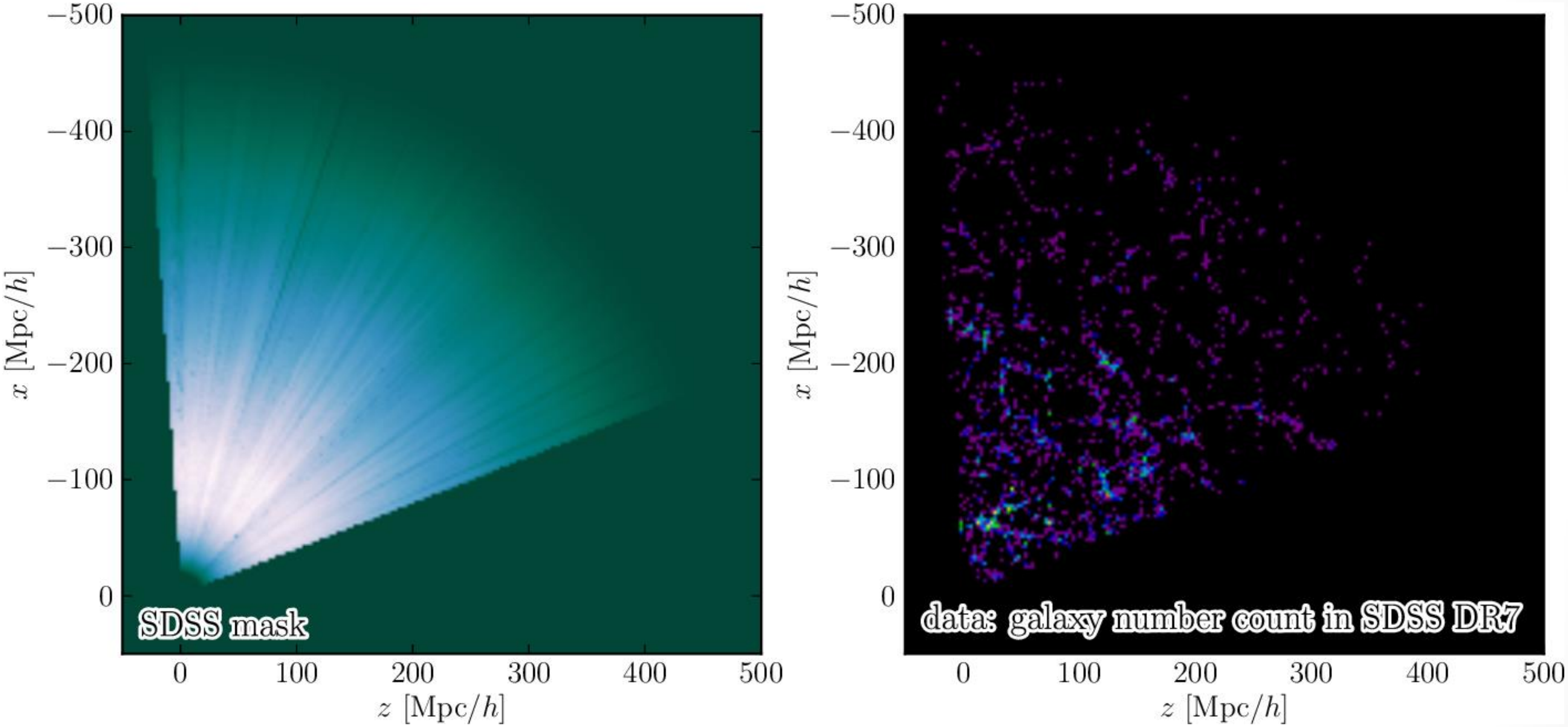
BORG at work – chronocosmography



Initial conditions | Final conditions | Observations

Jasche, FL & Wandelt 2014, arXiv:1409.6308

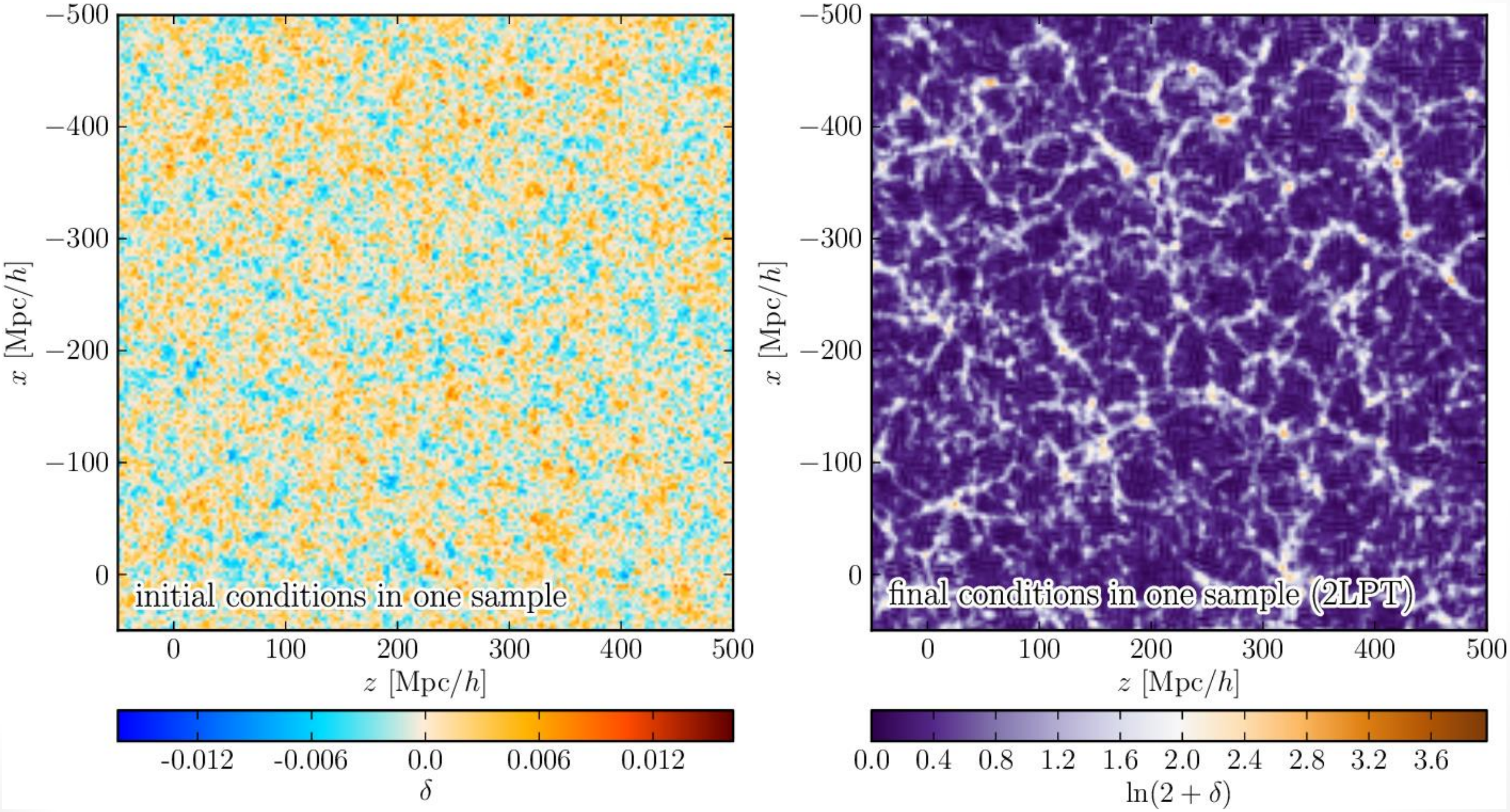
Bayesian chronocosmography from SDSS DR7



Data

Jasche, FL & Wandelt 2014, arXiv:1409.6308

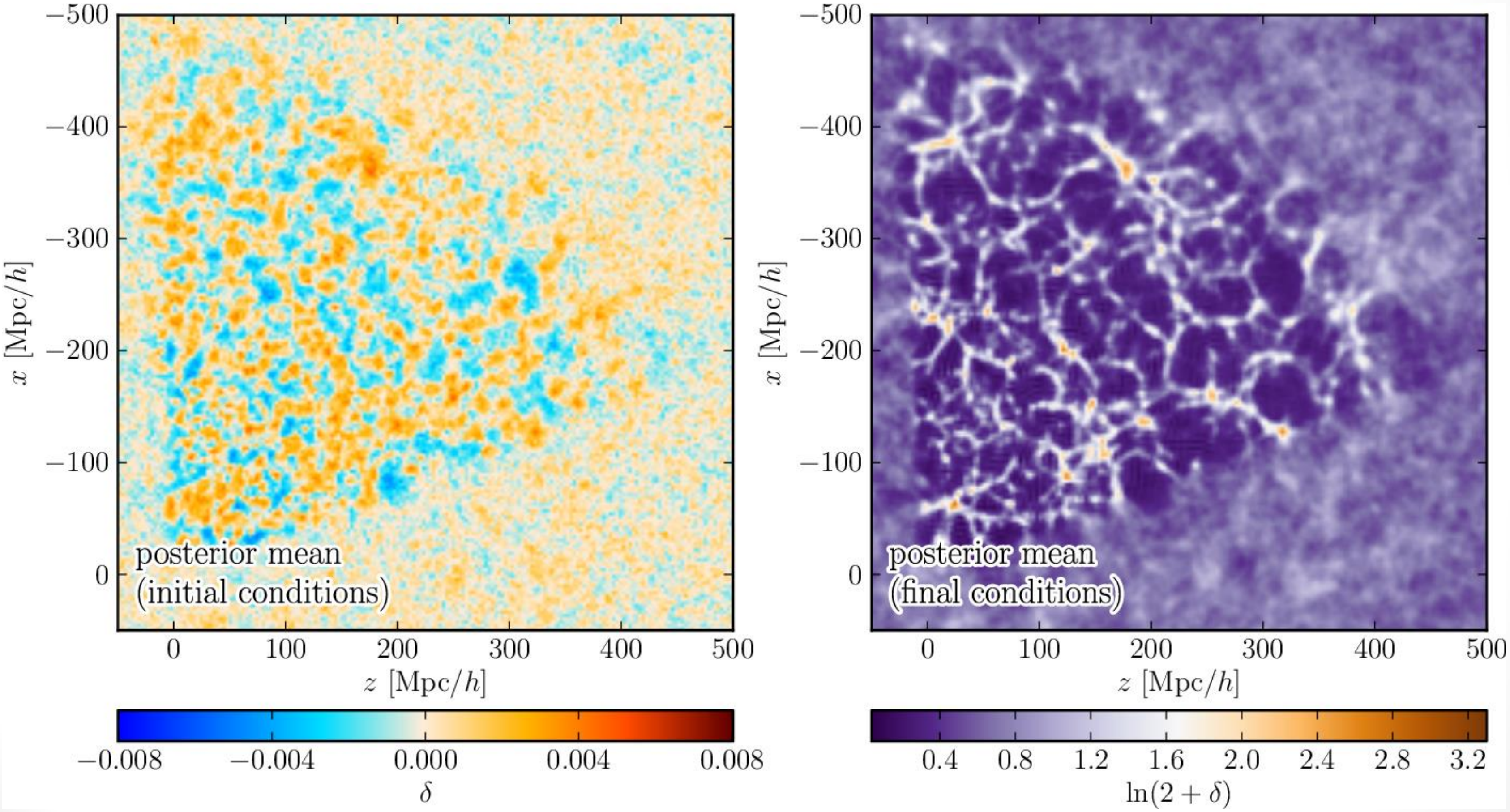
Bayesian chronocosmography from SDSS DR7



One sample

Jasche, FL & Wandelt 2014, arXiv:1409.6308

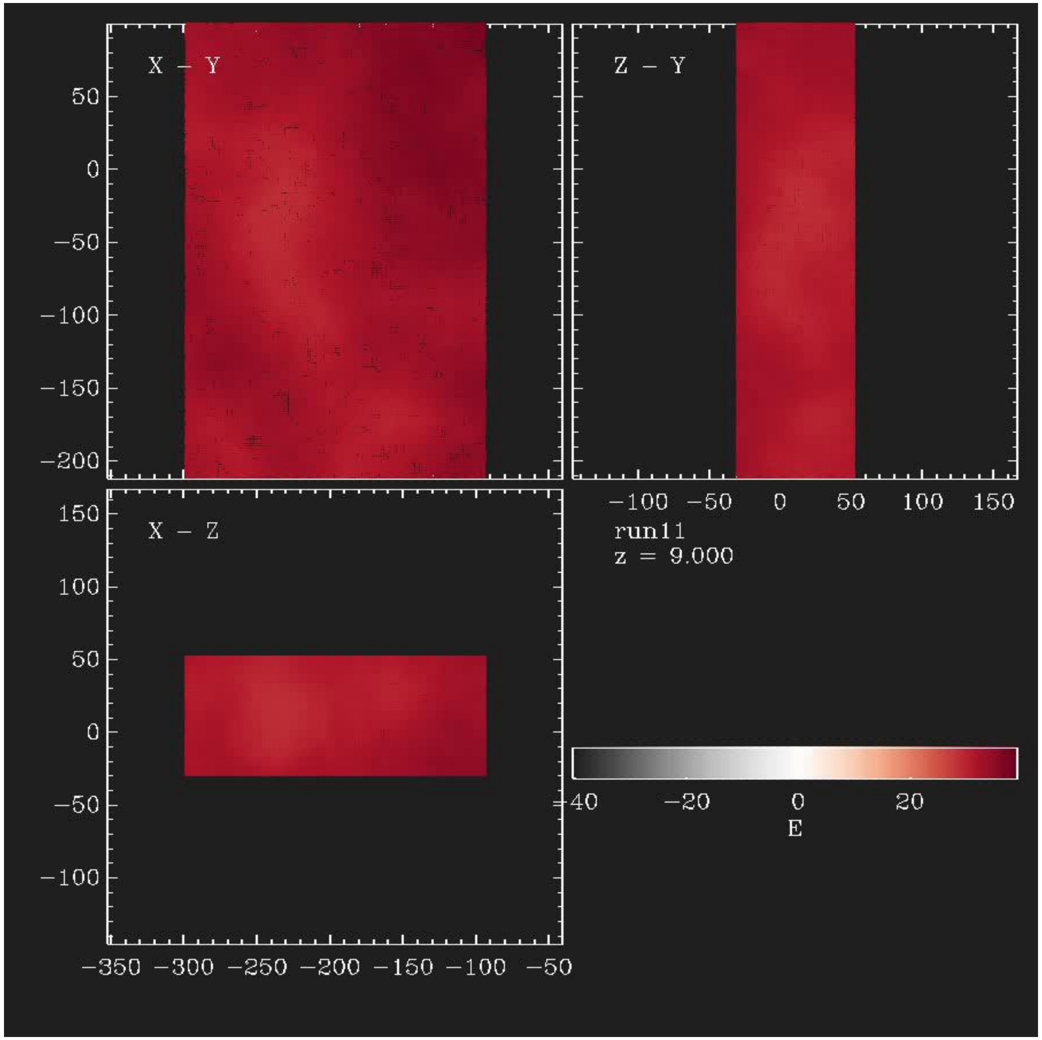
Bayesian chronocosmography from SDSS DR7



Posterior mean

Jasche, FL & Wandelt 2014, arXiv:1409.6308

The Sloan Great Wall through Bayesian eyes



The Sloan Great Wall:

Gott *et al*. 2005, arXiv:astro-ph/0310571

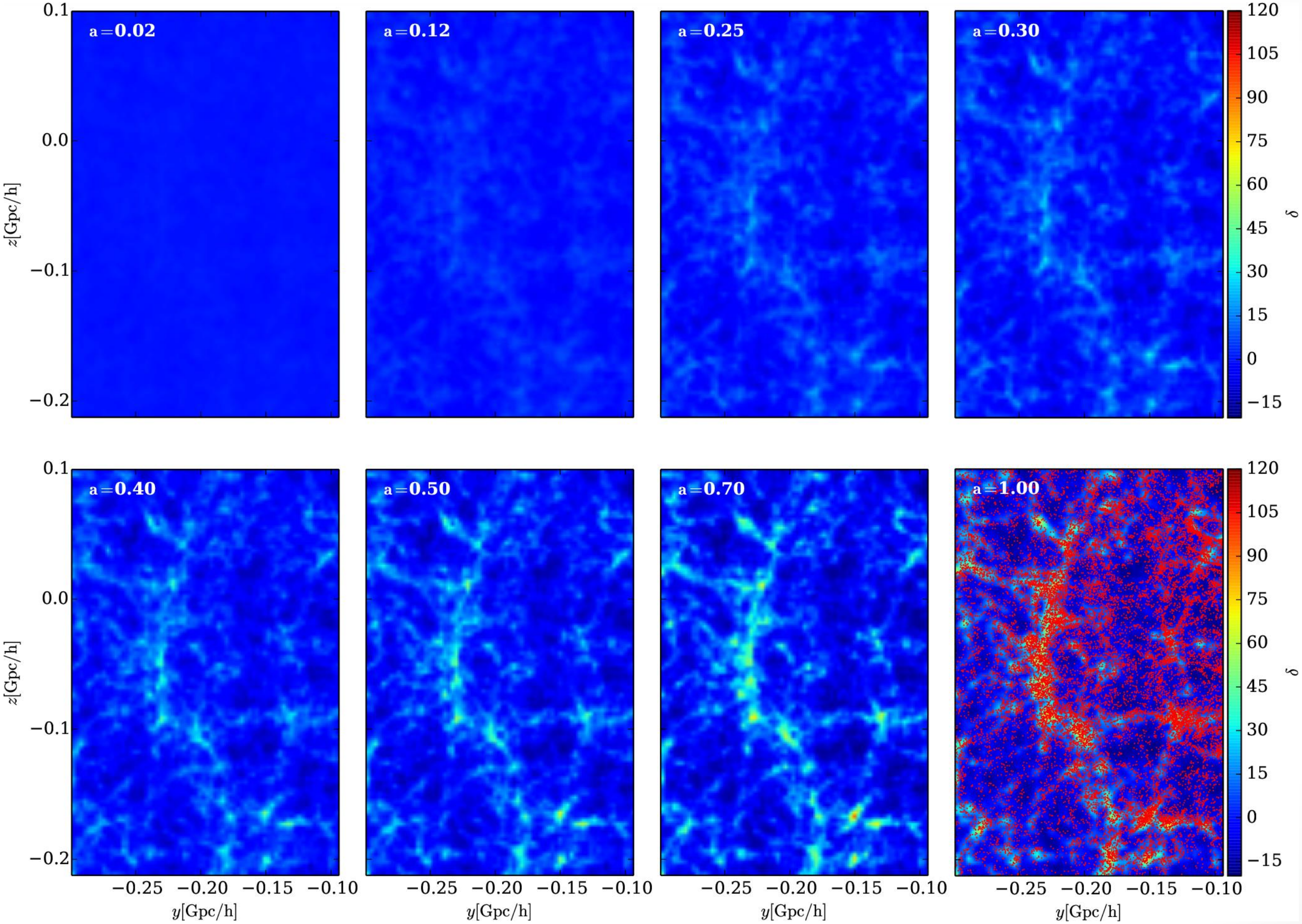
Nichol et al. 2006, arXiv:astro-ph/0602548

Einasto *et al*. 2010, arXiv:1007.4492

Einasto *et al*. 2011, arXiv:1105.1632

Jasche, Romano-Díaz, FL & Wandelt 2015, in prep.

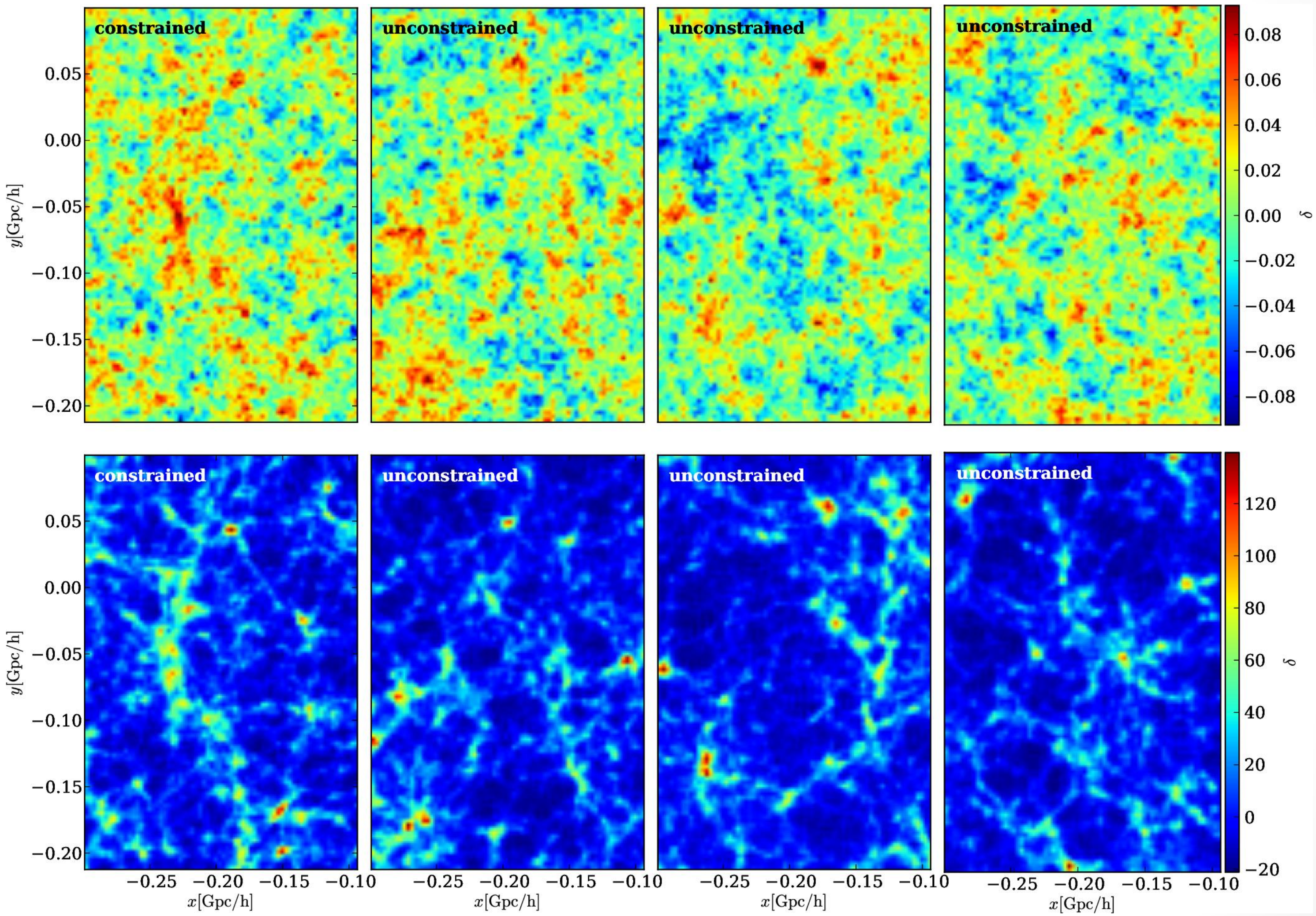
The formation history of the SGW



Jasche, Romano-Díaz, FL & Wandelt 2015, in prep.

Is the SGW special in ΛCDM?

Sylos Labini *et al*. 2009, arXiv:0909.0132



Jasche, Romano-Díaz, FL & Wandelt 2015, in prep.

3. The Non-Linear Regime of Structure Formation

- Non-linear filtering of BORG results
- Remapping Lagrangian Perturbation Theory
- The COLA method

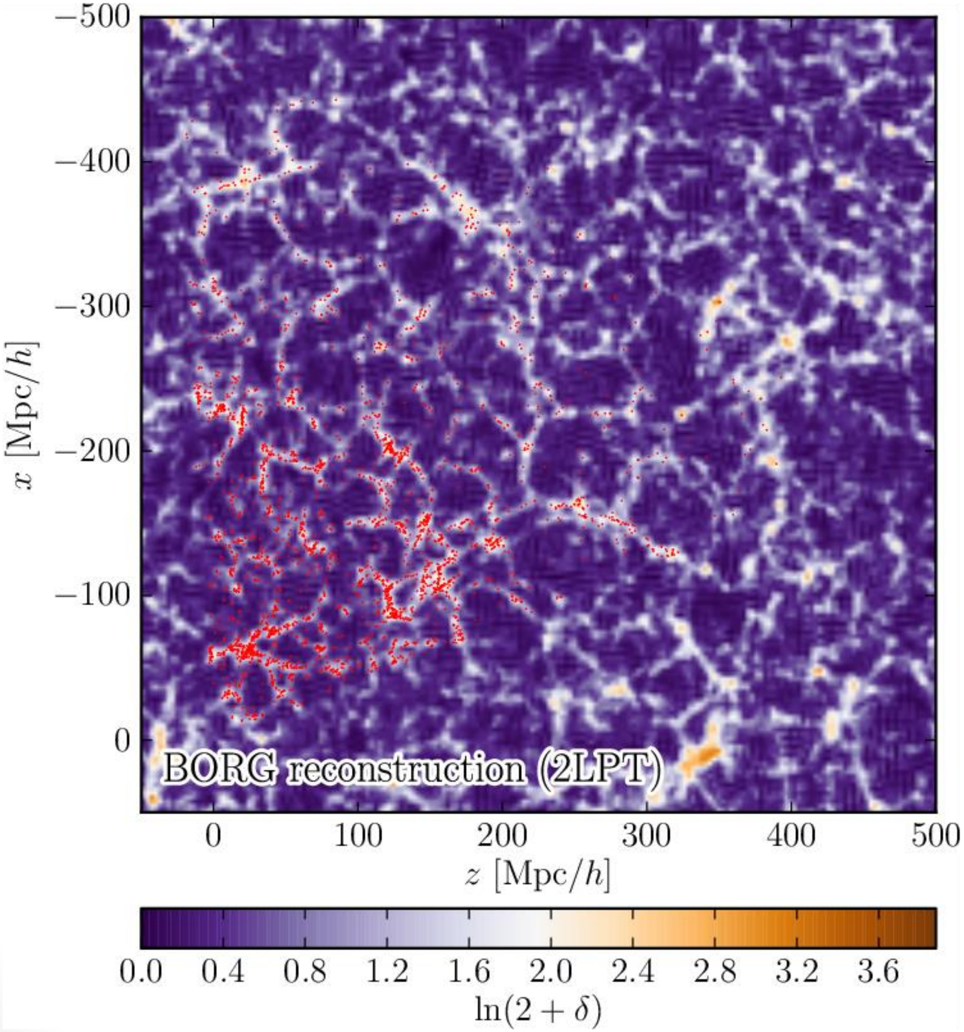
F. Leclercq, J. Jasche, P. M. Sutter, N. Hamaus, B. Wandelt, arXiv:1410.0355.
Dark matter voids in the SDSS galaxy survey

F. Leclercq, J. Jasche, H. Gil-Marín, B. Wandelt, arXiv:1305.4642.
One-point remapping of Lagrangian perturbation theory in the mildly non-linear regime of cosmic structure formation

S. Tassev, M. Zaldarriaga, D. Eisenstein, arXiv:1301.0322.
Solving Large Scale Structure in Ten Easy Steps with COLA

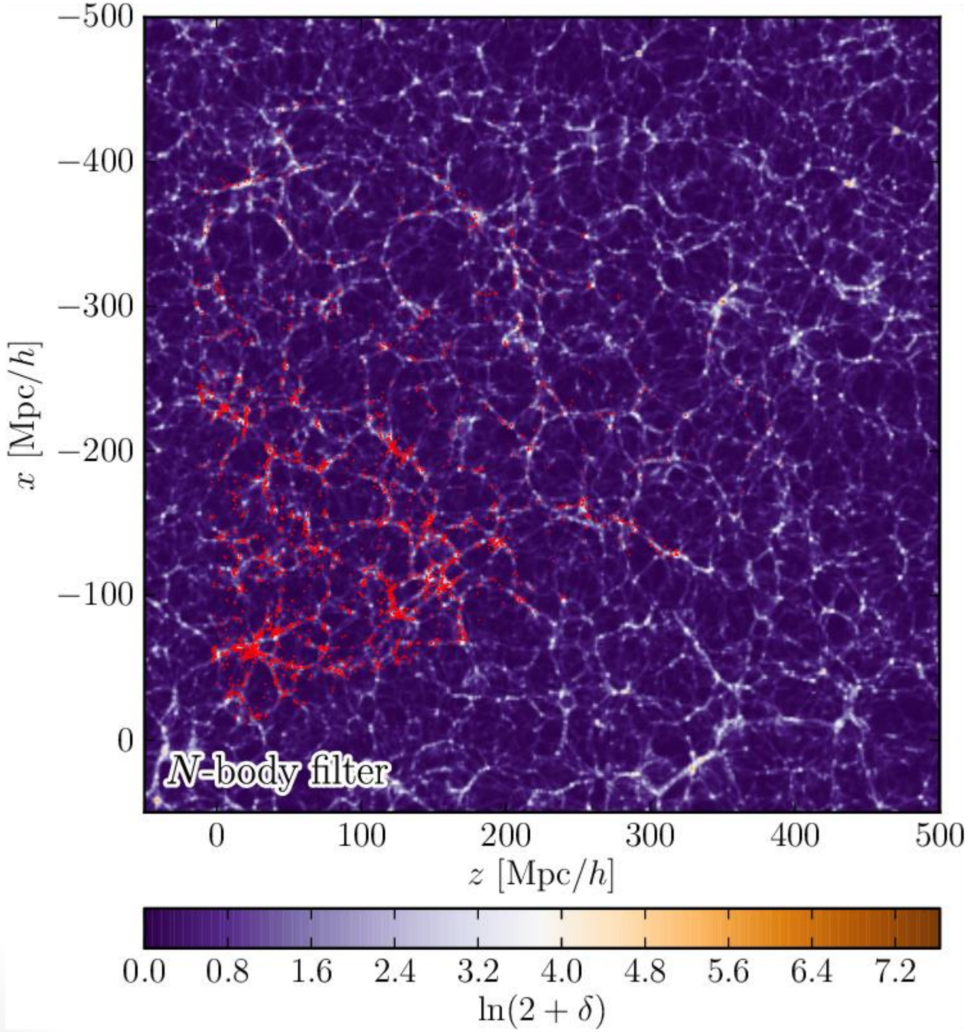
S. Tassev, D. Eisenstein, B. Wandelt, M. Zaldarriaga, in prep. + F. Leclercq, B. Wandelt, *et al.*, in prep.
sCOLA: The N-body COLA method extended to the Spatial Domain

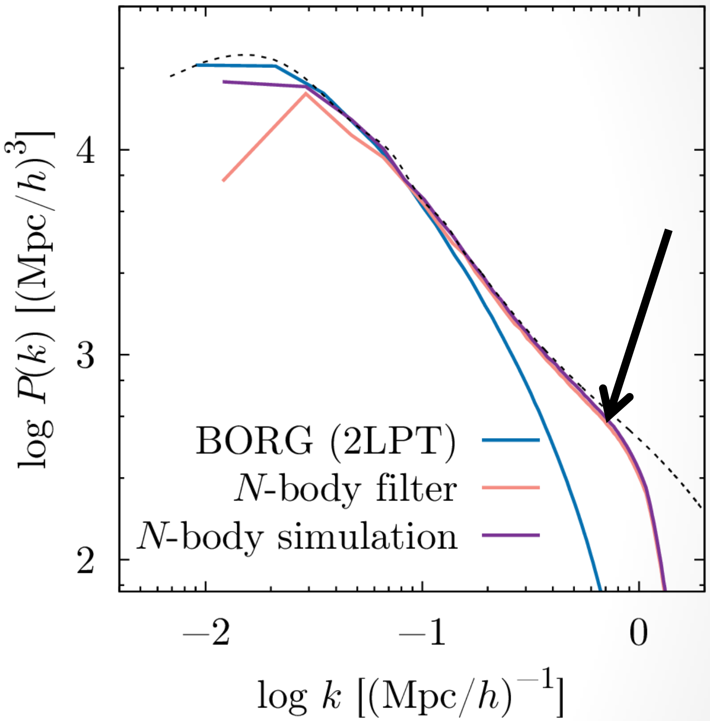
Non-linear filtering



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

Non-linear filtering



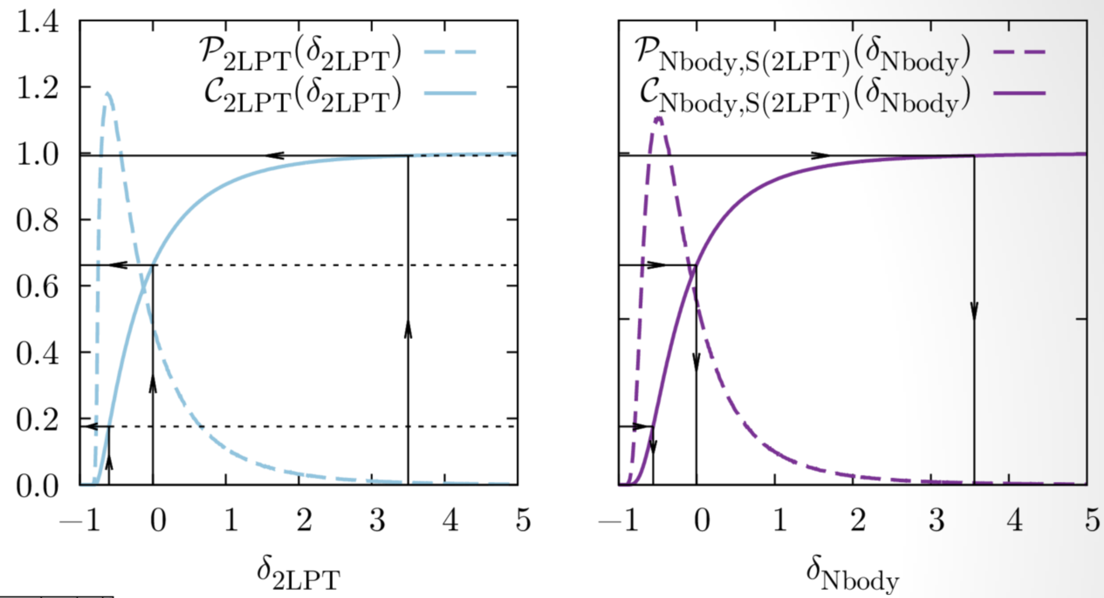


FL, Jasche, Sutter, Hamaus & Wandelt 2014, arXiv:1410.0355

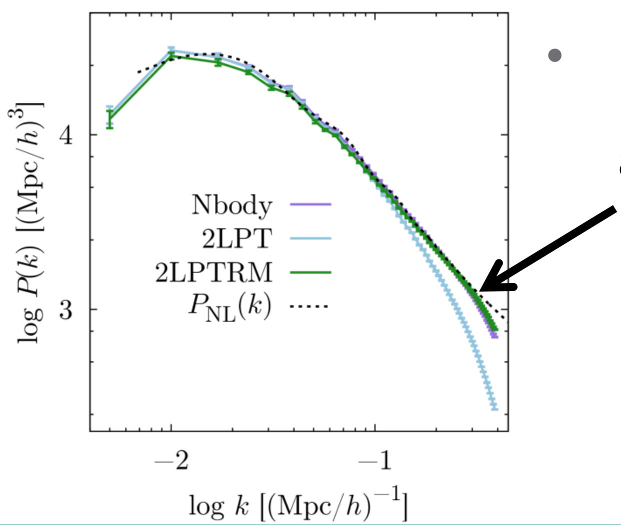
Remapping 2LPT in the mildly non-linear regime

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

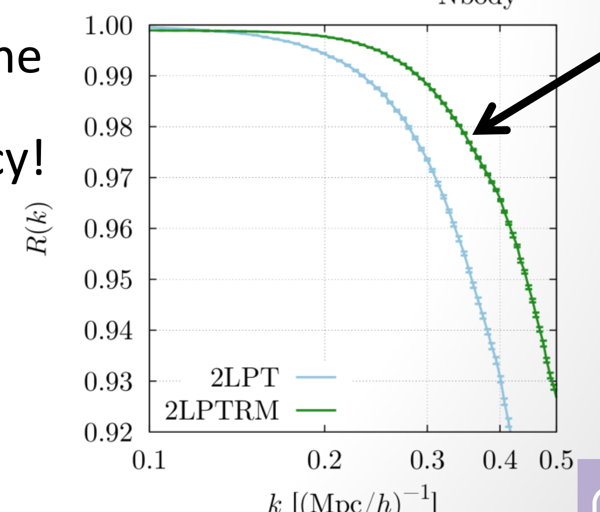
- Replacing the one-point distribution of 2LPT by one which accounts for the full non-linear system…



- …also improves the higher-order correlators…



- …and the phase accuracy!



COLA: *COmoving Lagrangian Acceleration*

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

Tassev & Zaldarriaga 2012, arXiv:1203.5785

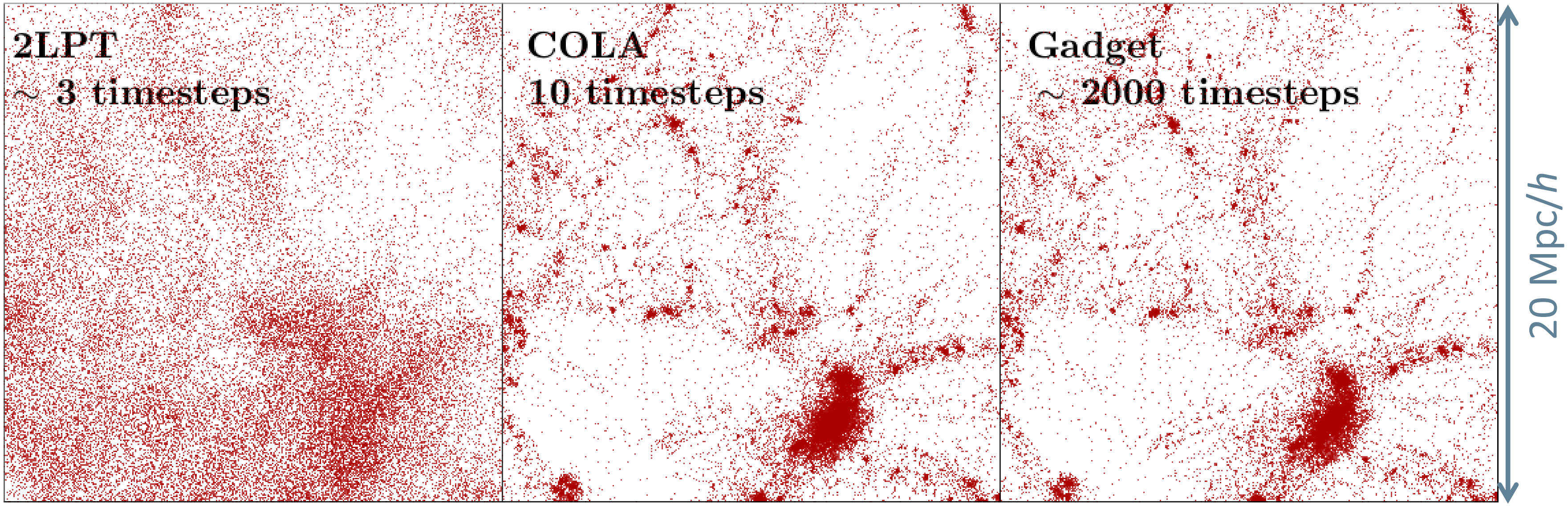
- Time-stepping (omitted constants and Hubble expansion):

Standard:

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

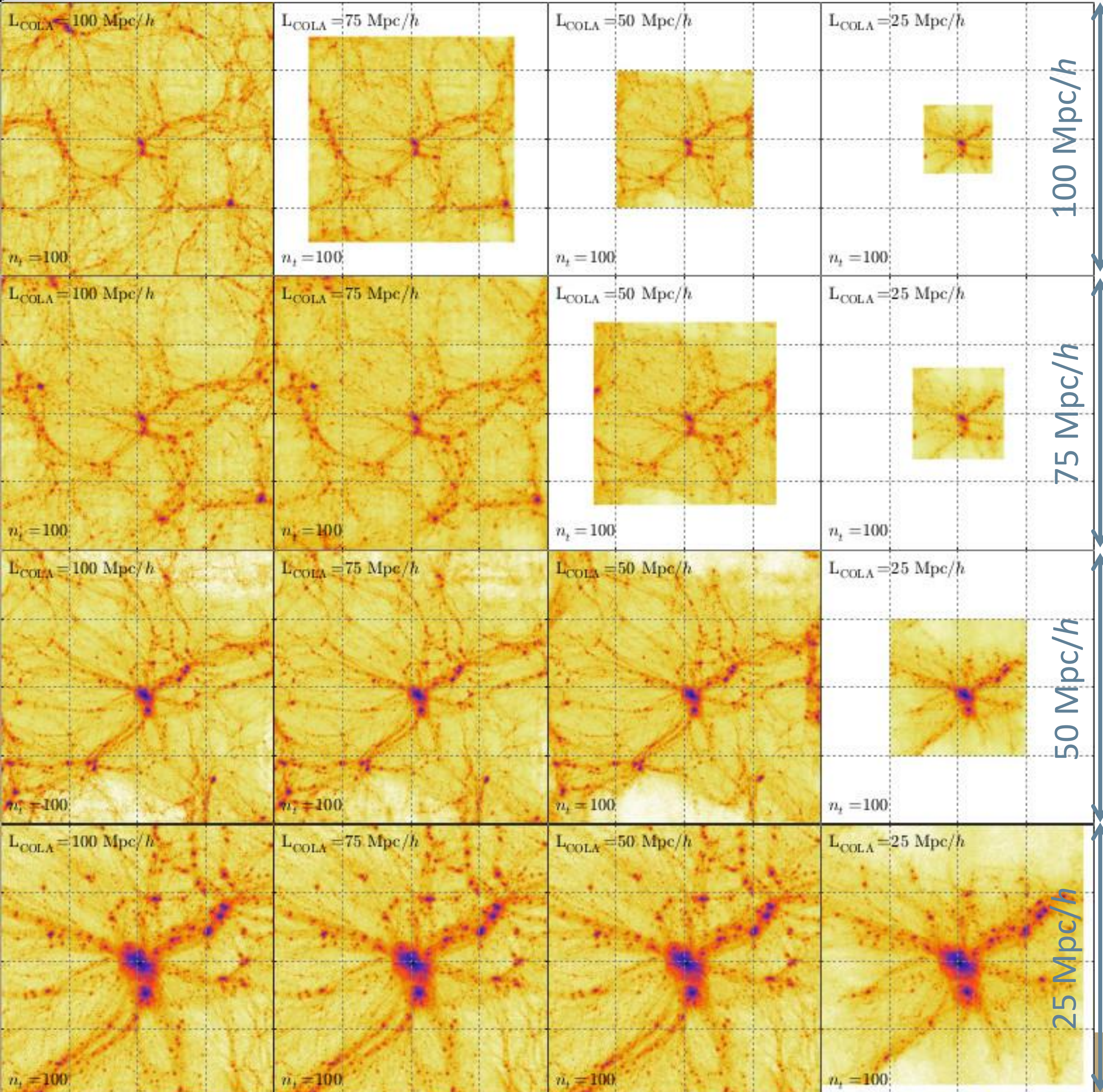
Modified:

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



Original COLA "in time"

Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322



Extending COLA

New COLA “in space”

$$n_t = 100$$

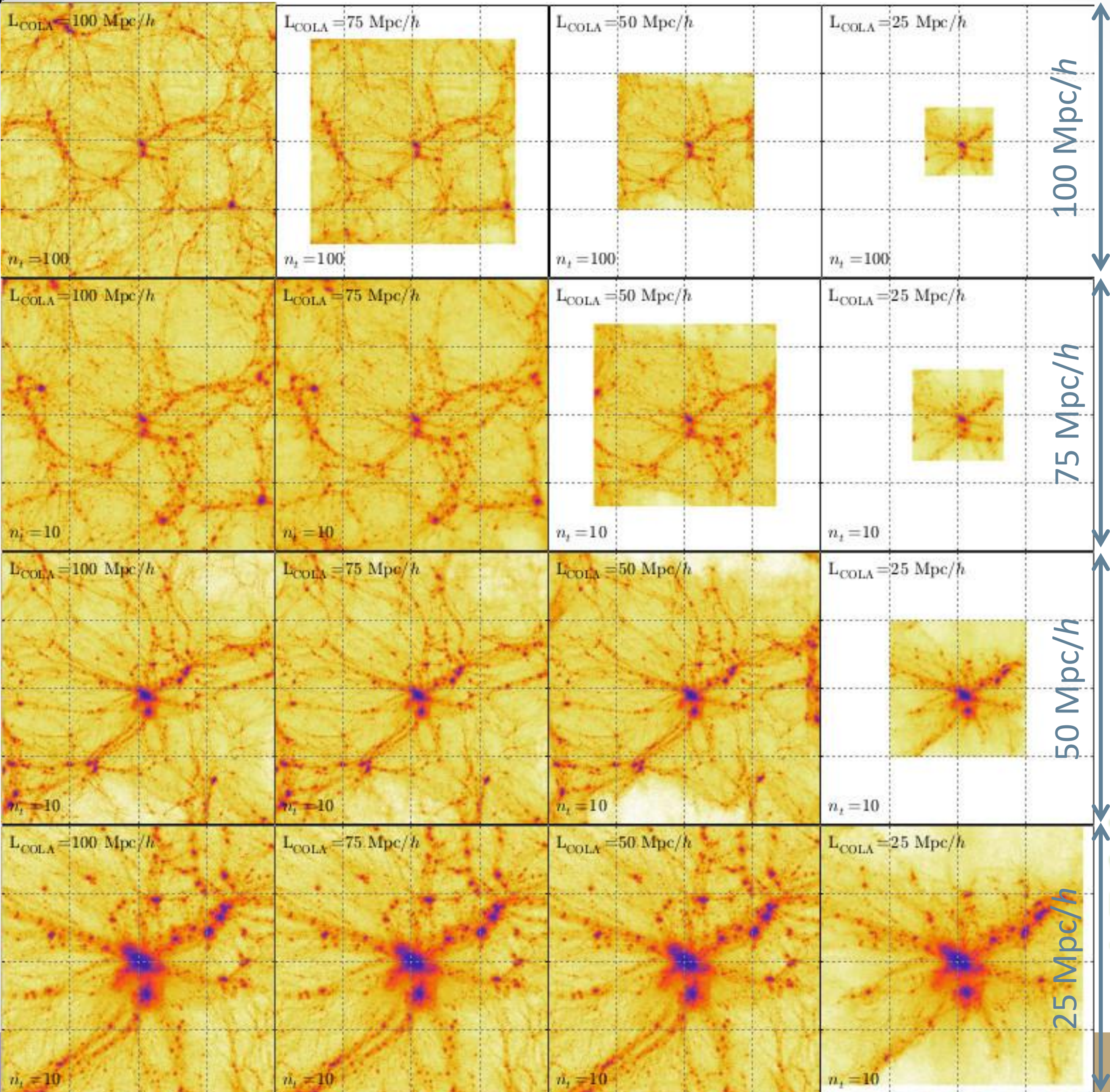
Tassev, Eisenstein, Wandelt & Zaldarriaga, in prep.

+ FL, Wandelt, *et al.*, in prep.

Extending COLA

New COLA "in space and time"

$$n_t = 10$$



Tassev, Eisenstein, Wandelt & Zaldarriaga, in prep.
+ FL, Wandelt, *et al.*, in prep.

4. Cosmic Web Classification

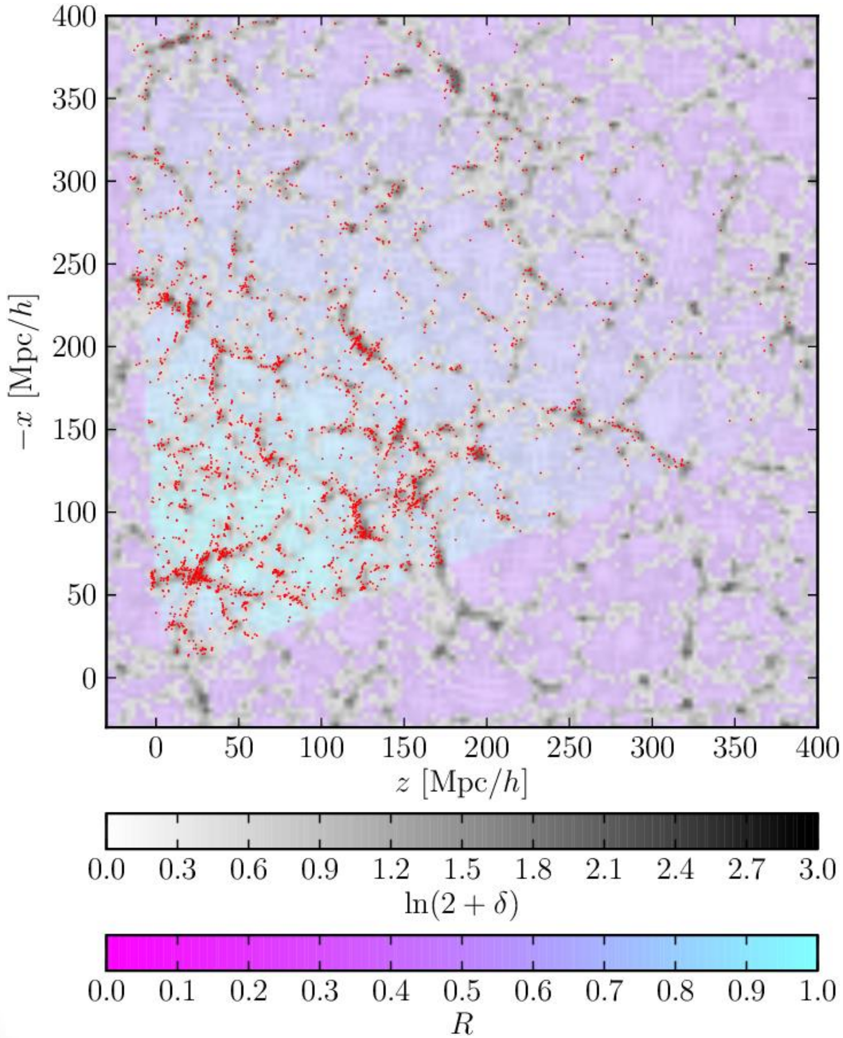
- Dark matter voids in the SDSS
- Tidal shear analysis in the SDSS, dynamic structure type classification
- Cosmic-web classification and Bayesian decision theory

F. Leclercq, J. Jasche, P. M. Sutter, N. Hamaus, B. Wandelt, arXiv:1410.0355.
Dark matter voids in the SDSS galaxy survey

F. Leclercq, J. Jasche, B. Wandelt, arXiv:1502.02690.
Bayesian analysis of the dynamic cosmic web in the SDSS galaxy survey

F. Leclercq, J. Jasche, B. Wandelt, in prep.
Cosmic web-type classification using decision theory

Dark matter voids in the SDSS



- Why?

Sparsity & **Bias**

Sutter *et al.* 2013, arXiv:1309.5087

Sutter *et al.* 2013, arXiv:1311.3301

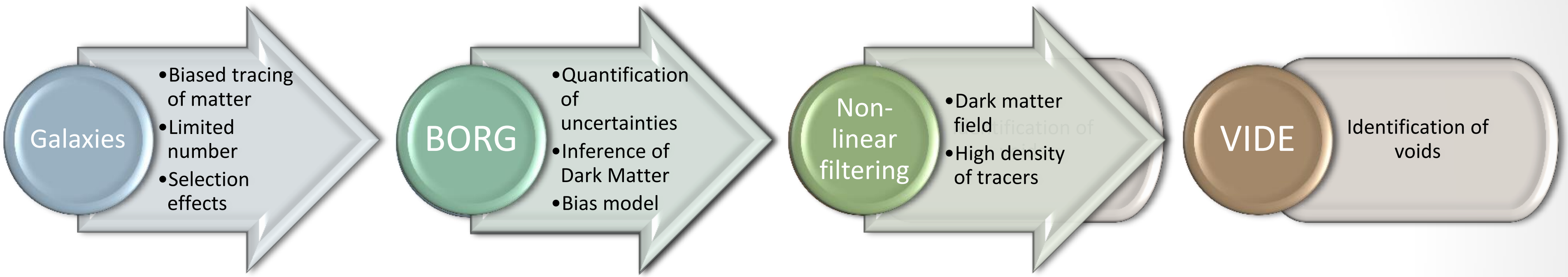
- How?

VIDE toolkit: Sutter *et al.* 2014, arXiv:1406.1191
www.cosmicvoids.net

based on ZOBOV: Neyrinck 2007, arXiv:0712.3049

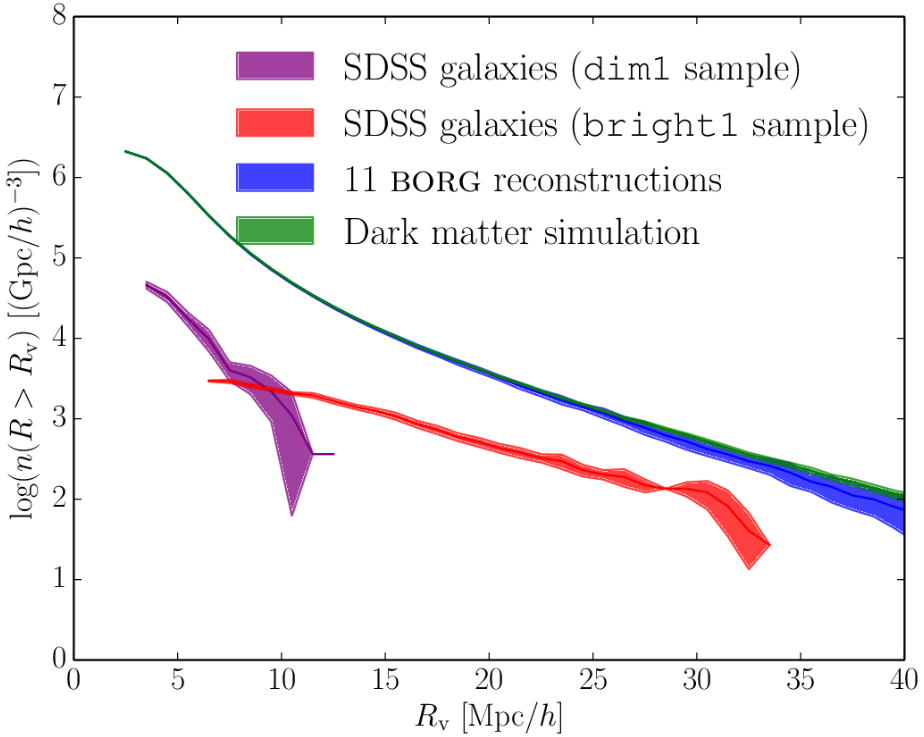
FL, Jasche, Sutter, Hamaus & Wandelt 2014, arXiv:1410.0355

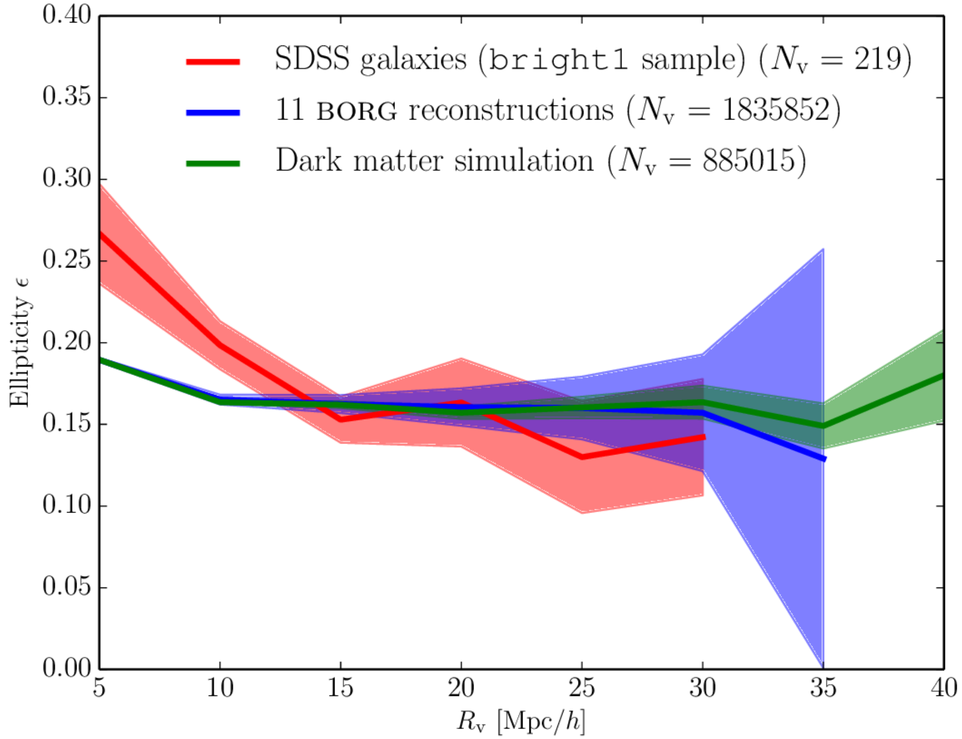
Dark matter voids: pipeline



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

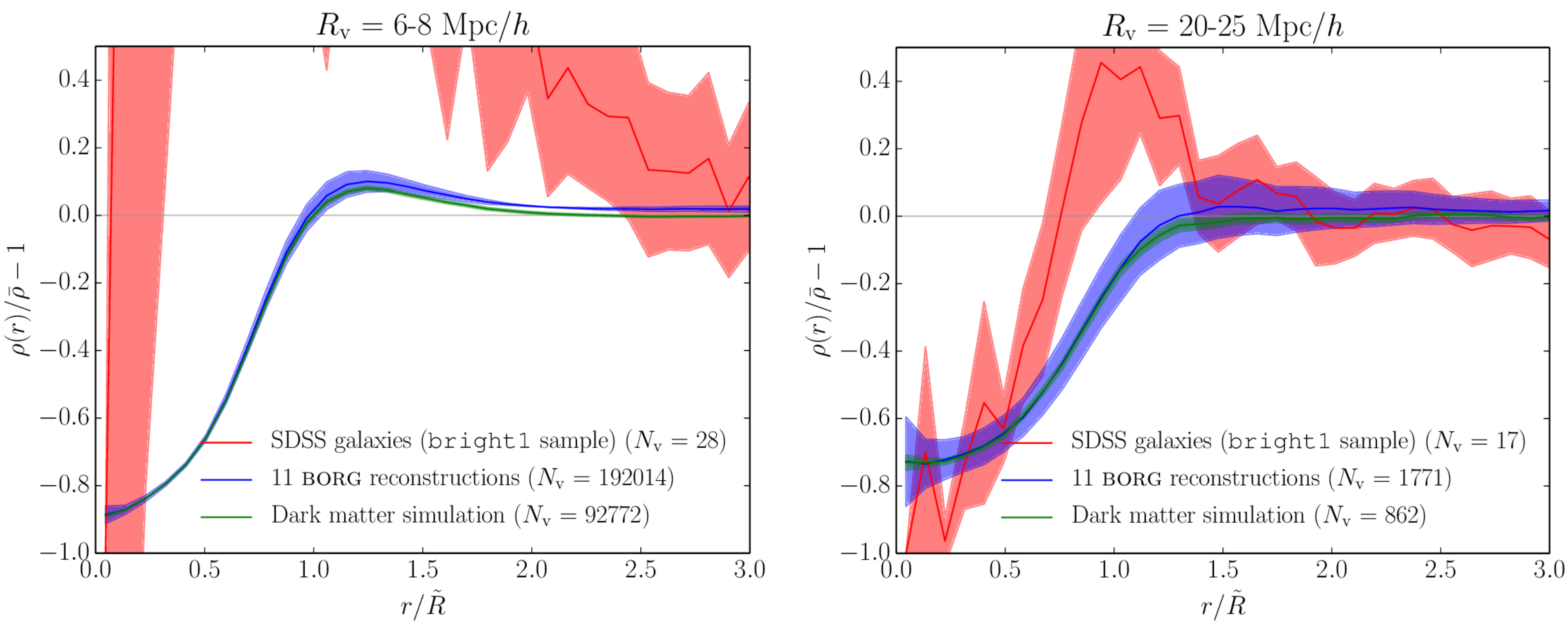
Dark matter void properties





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

Dark matter void properties



All catalogs will be made publicly available at
www.cosmicvoids.net

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

Tidal shear analysis

- $\lambda_1, \lambda_2, \lambda_3$: eigenvalues of the tidal field tensor, the Hessian of the gravitational potential: $T_{ij} = \partial_i \partial_j \Phi$ $\quad \lambda_1 + \lambda_2 + \lambda_3 = \delta$
  - 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 *et al*. 2007, arXiv:astro-ph/0610280

see also:

- Extensions:

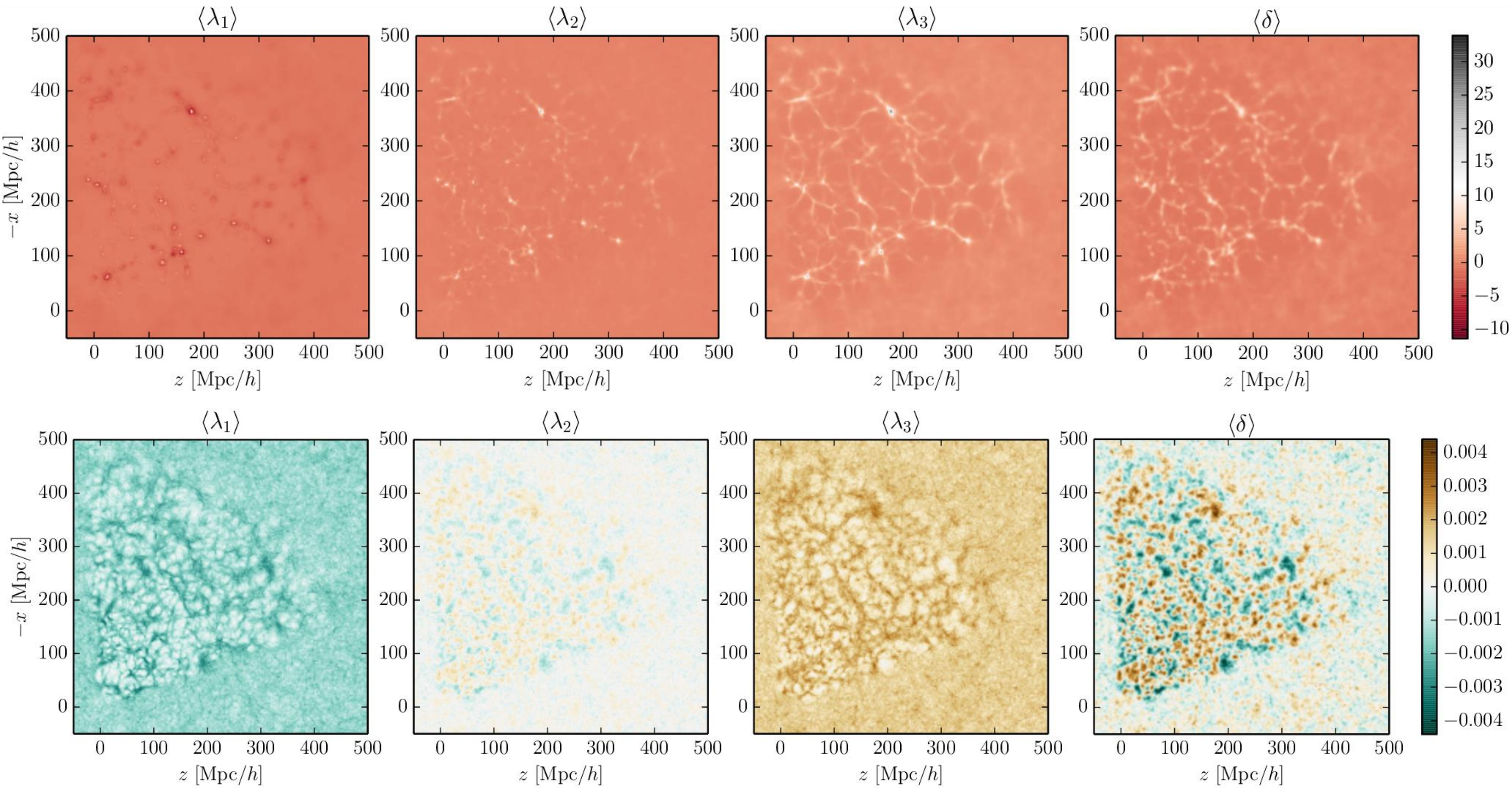
Forero-Romero *et al*. 2008, arXiv:0809.4135
Hoffman *et al*. 2012, arXiv:1201.3367

- Similar web classifiers:

DIVA, Lavaux & Wandelt 2010, arXiv:0906.4101
ORIGAMI, Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

Tidal shear analysis

Final conditions

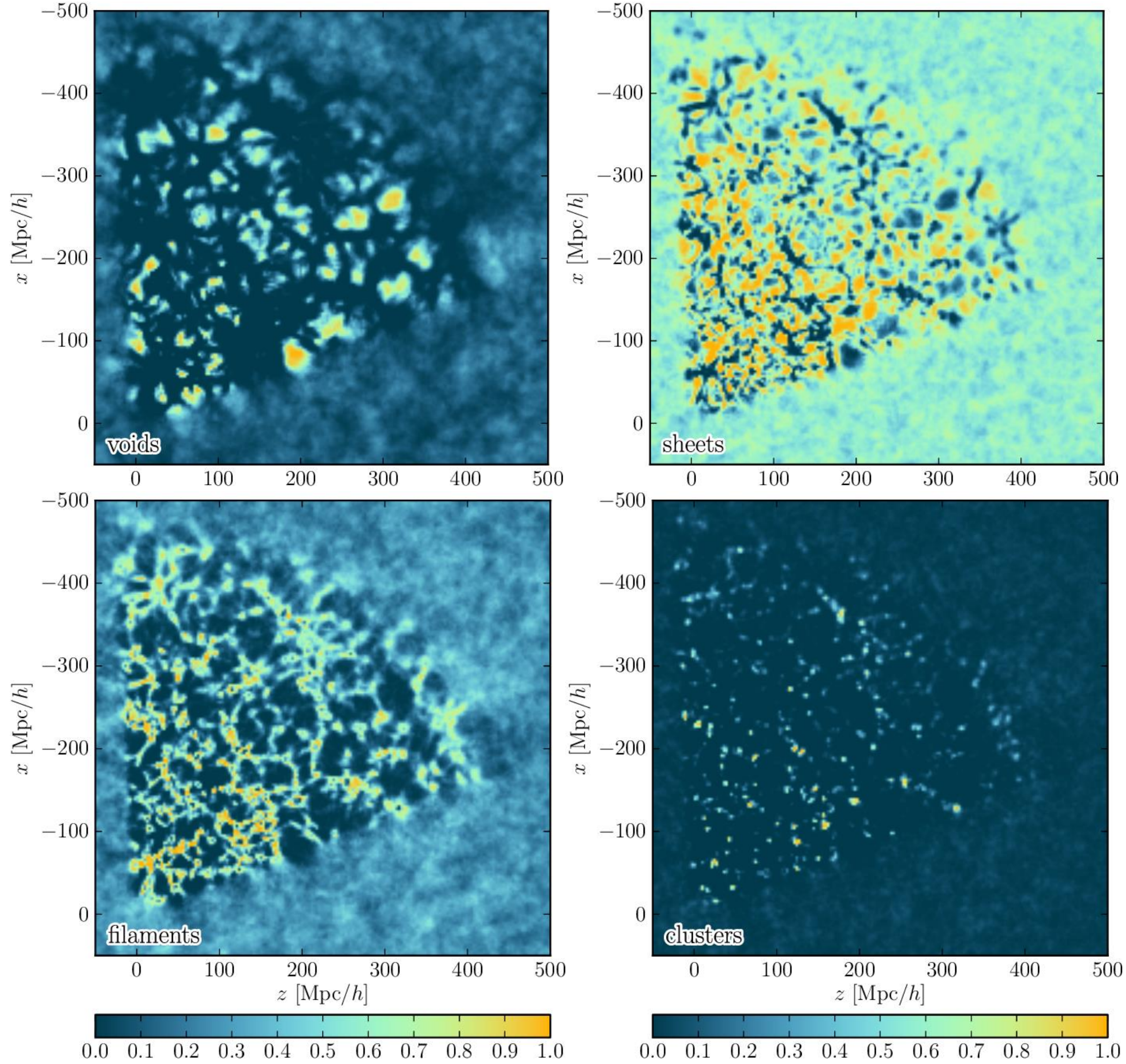


Initial conditions

FL, Jasche & Wandelt 2015, arXiv:1502.02690

Dynamic structures inferred by BORG

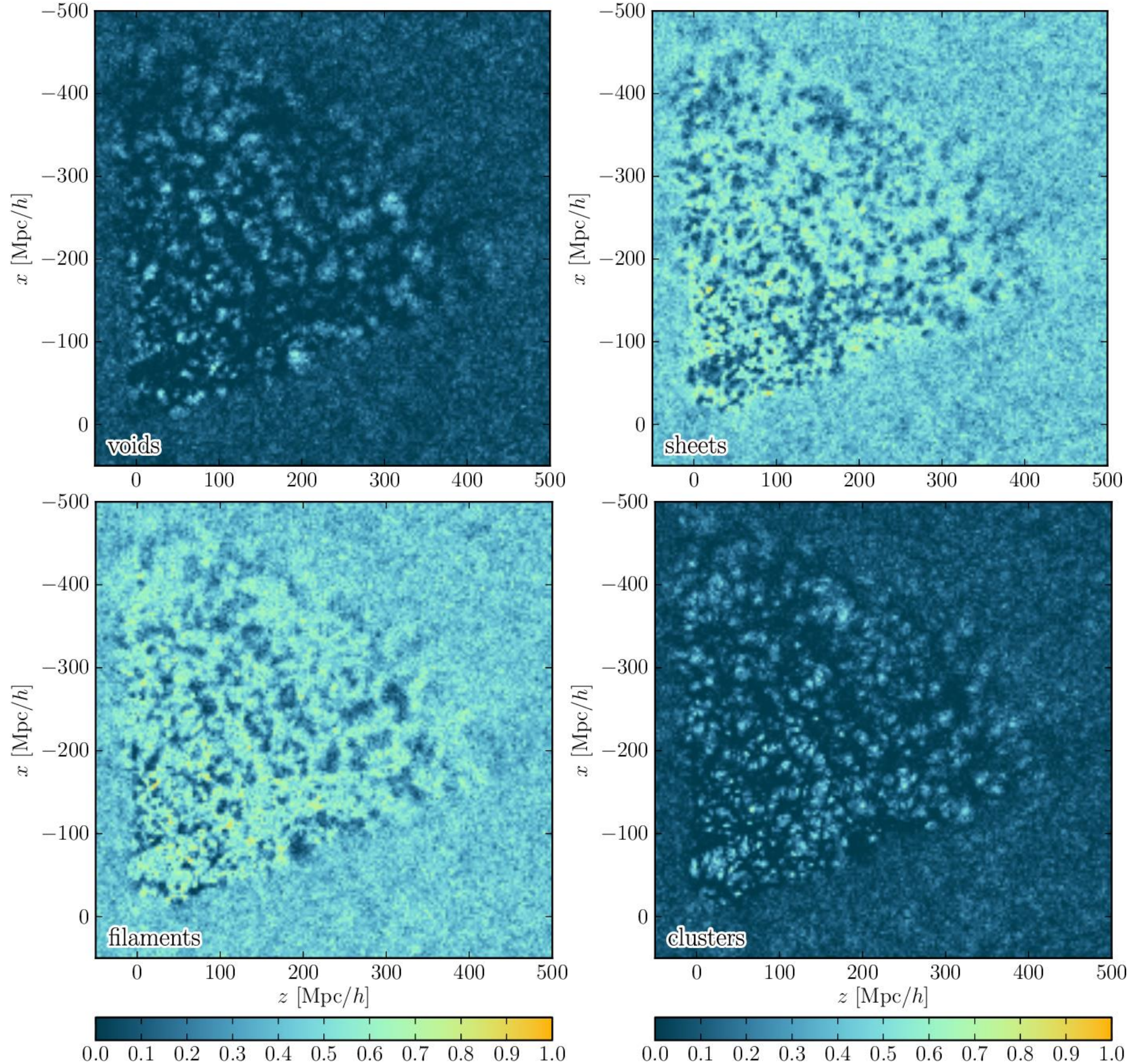
Final conditions



FL, Jasche & Wandelt 2015, arXiv:1502.02690

Dynamic structures inferred by BORG

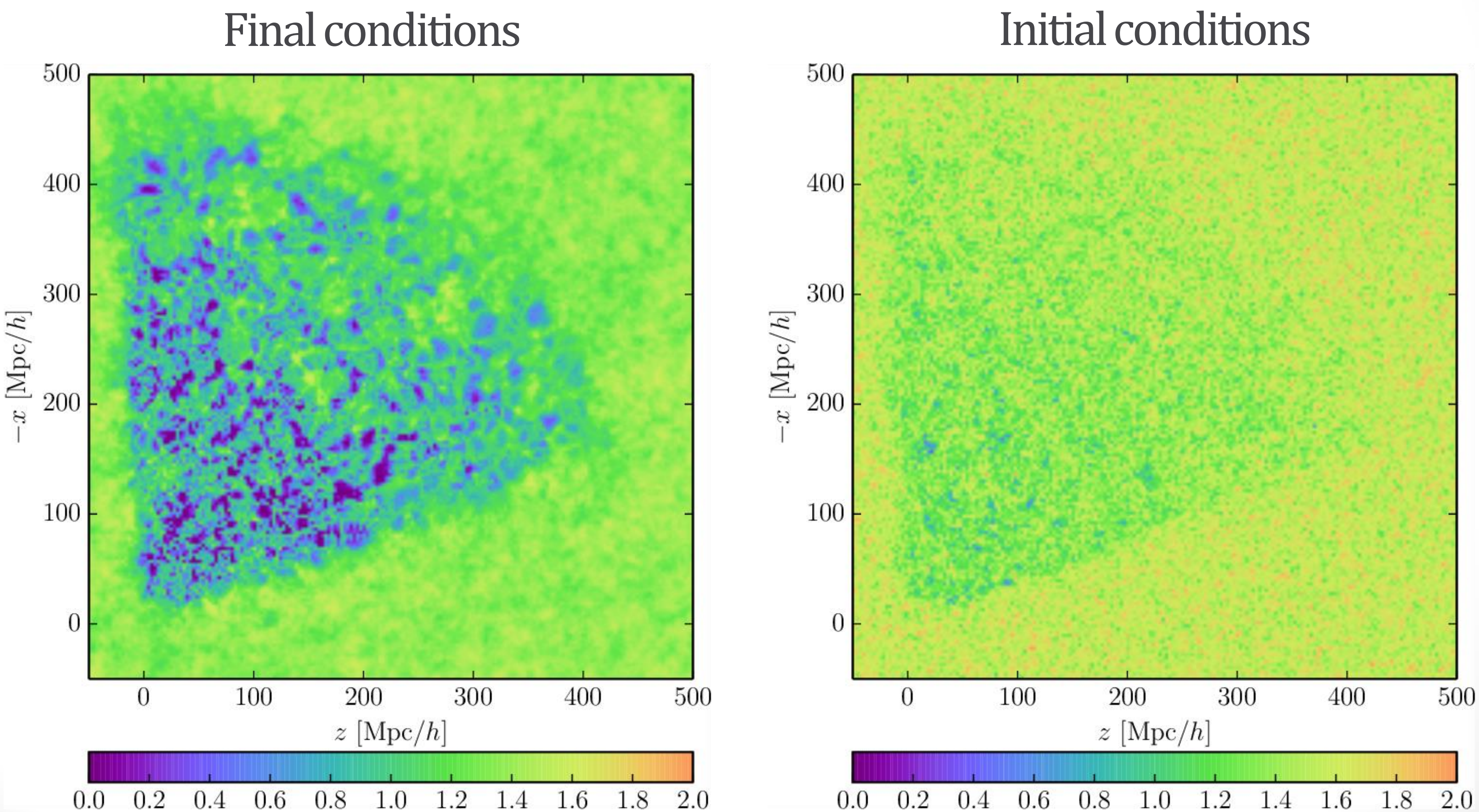
Initial conditions



FL, Jasche & Wandelt 2015, arXiv:1502.02690

Entropy of the structure types pdf

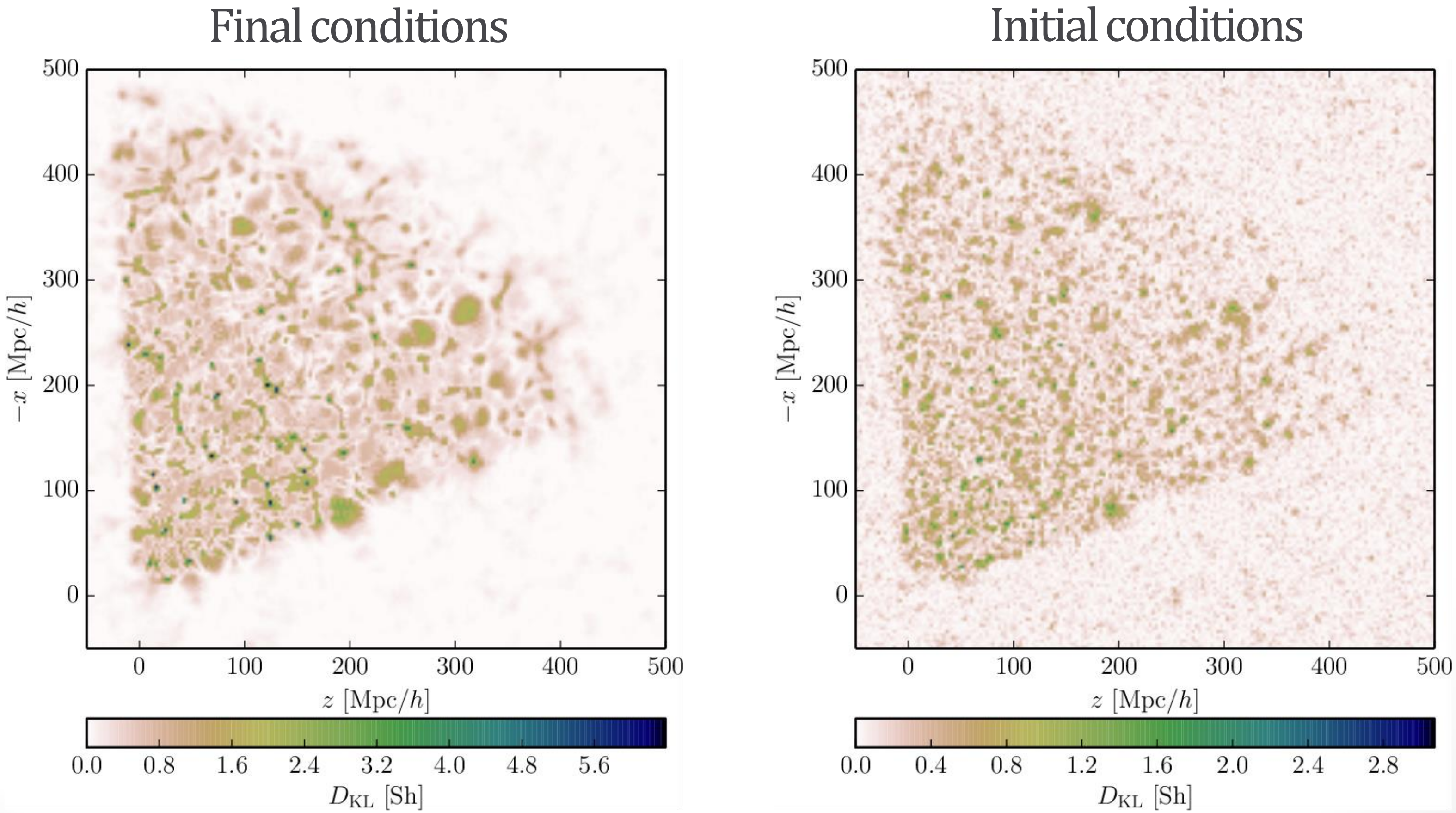
$$H\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)\right] \equiv -\sum_{i=0}^{3} \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)\log_2(\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d))$$ in shannons (Sh)



FL, Jasche & Wandelt 2015, arXiv:1502.02690

Kullback-Leibler divergence posterior/prior

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

FL, Jasche & Wandelt, in prep.

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"

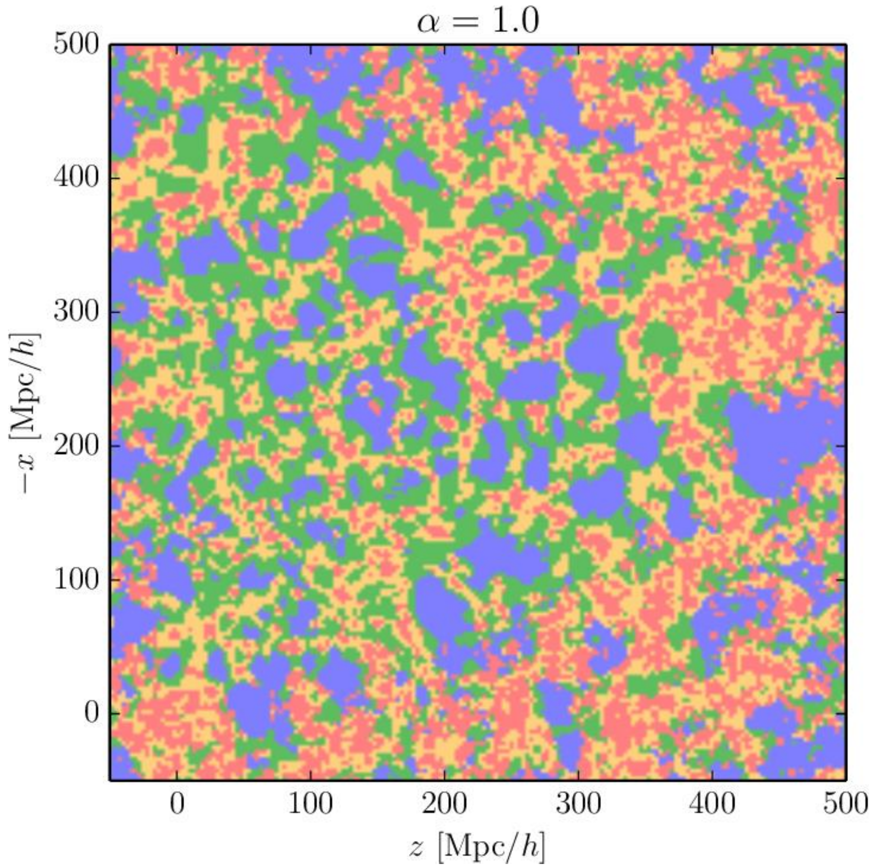
$$U(a_{-1}) = 0$$ "Not playing the game"

- 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

FL, Jasche & Wandelt, in prep.

Playing the game…

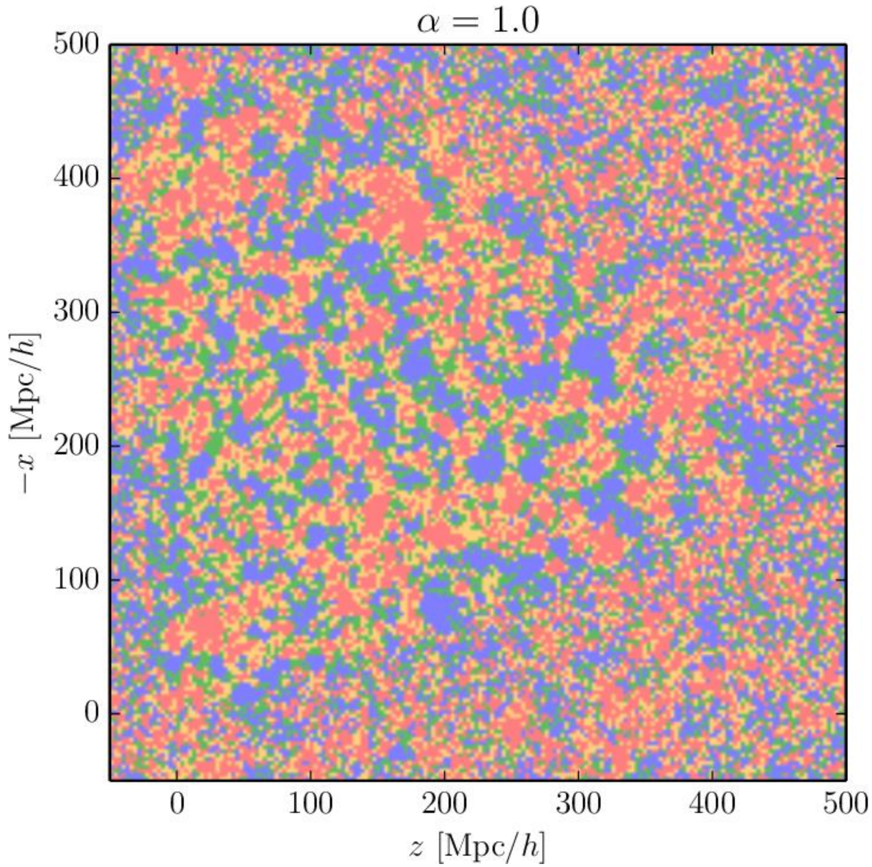
Final conditions



FL, Jasche & Wandelt, in prep.

Playing the game…

Initial conditions



FL, Jasche & Wandelt, in prep.

Summary & Conclusions

- Bayesian large-scale structure inference in 10 millions dimensions is possible!
  - Uncertainty quantification (noise, survey geometry, selection effects and biases)
  - Non-linear and non-Gaussian inference with improving techniques

- Application to data: four-dimensional chronocosmography
  - Simultaneous analysis of the morphology and formation history of the large-scale structure
  - Physical reconstruction of the initial conditions
  - Inference of cosmic voids at the level of the dark matter distribution
  - Characterization of the dynamic cosmic web underlying galaxies

Mapping the Universe: epilogue?