Probabilistic large-scale structure inference, cosmic web analysis and information theory

Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

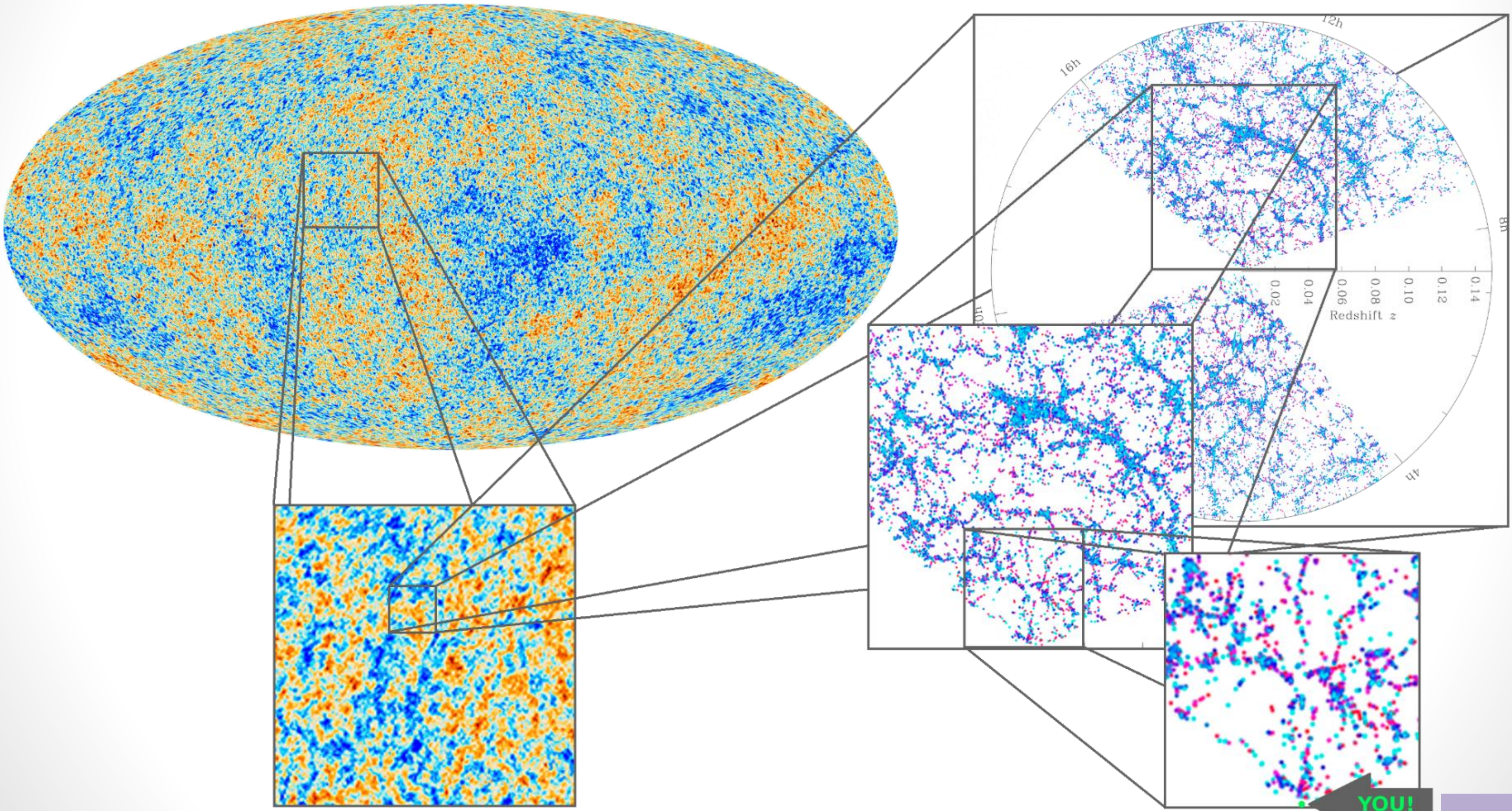
April 25th, 2016

In collaboration with:

Jens Jasche (ExC Universe, Garching), Guilhem Lavaux (IAP),
Will Percival (ICG), Benjamin Wandelt (IAP/U. Illinois)

The big picture: the Universe is highly structured

You are here. Make the best of it…



Planck collaboration (2013-2015)   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?

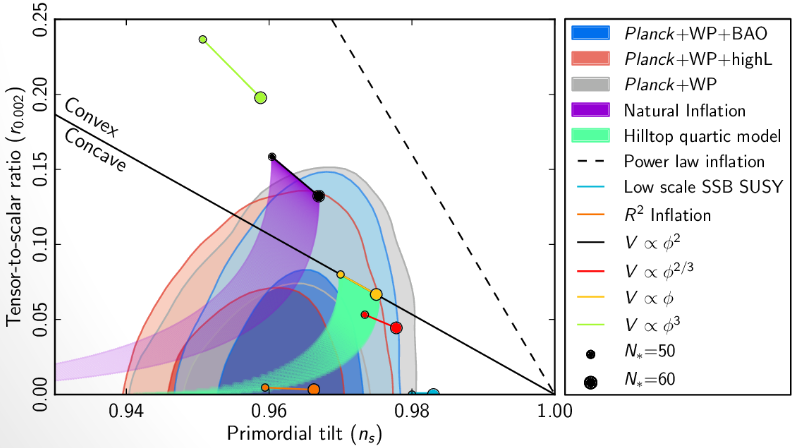
- **How did the large-scale structure take shape?**
  - What is the physics of dark matter and dark energy?

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

Redshift range	Volume (Gpc^3)	k_{max} (Mpc/h)$^{-1}$	N_{modes}
0-1	50	0.15	10^7
1-2	140	0.5	$5x10^8$
2-3	160	1.3	10^{10}

M. Zaldarriaga

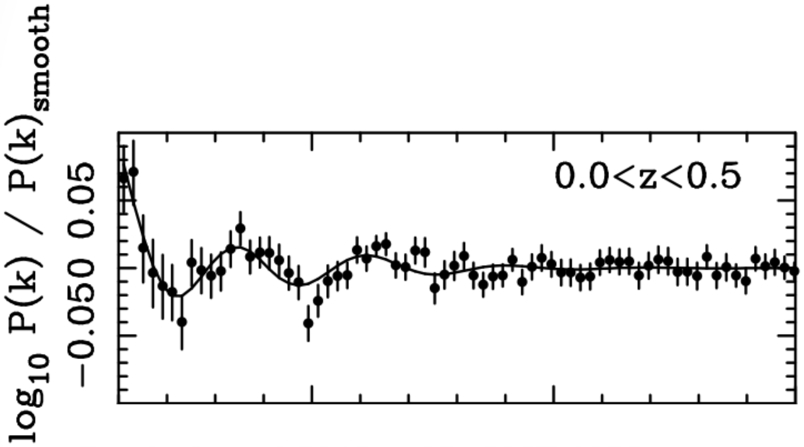
- Precise tests require many modes.
- In 3D galaxy surveys, the number of modes usable scales as $k^3_{\rm max}$.

- 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?"

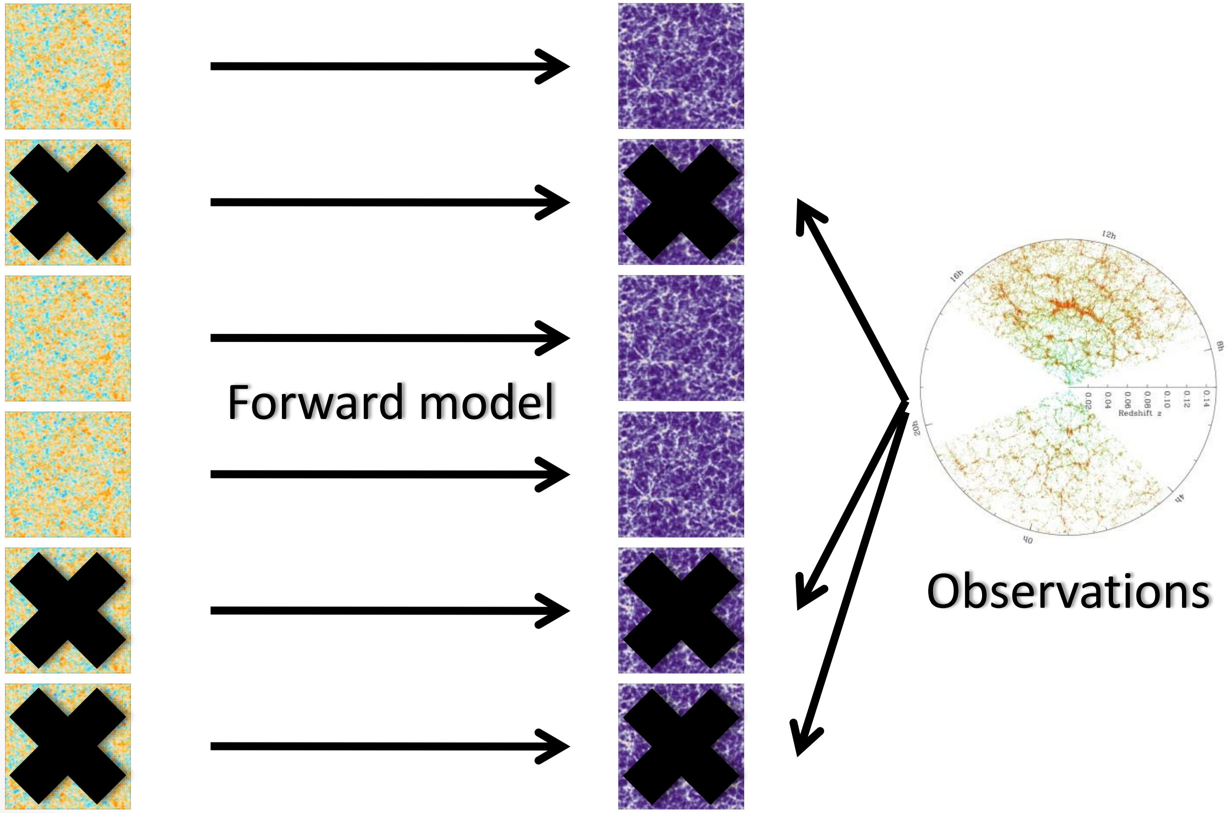
Bayes' theorem: $\mathcal{P}(s|d)\mathcal{P}(d) = \mathcal{P}(d|s)\mathcal{P}(s)$

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

⇨ How to do that?

Bayesian forward modeling: the ideal scenario

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

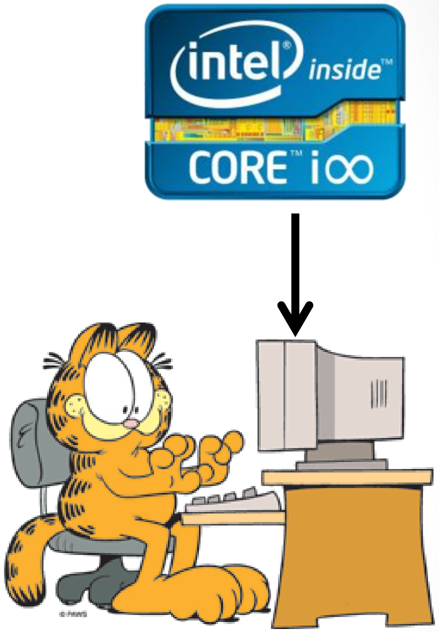


Forward model

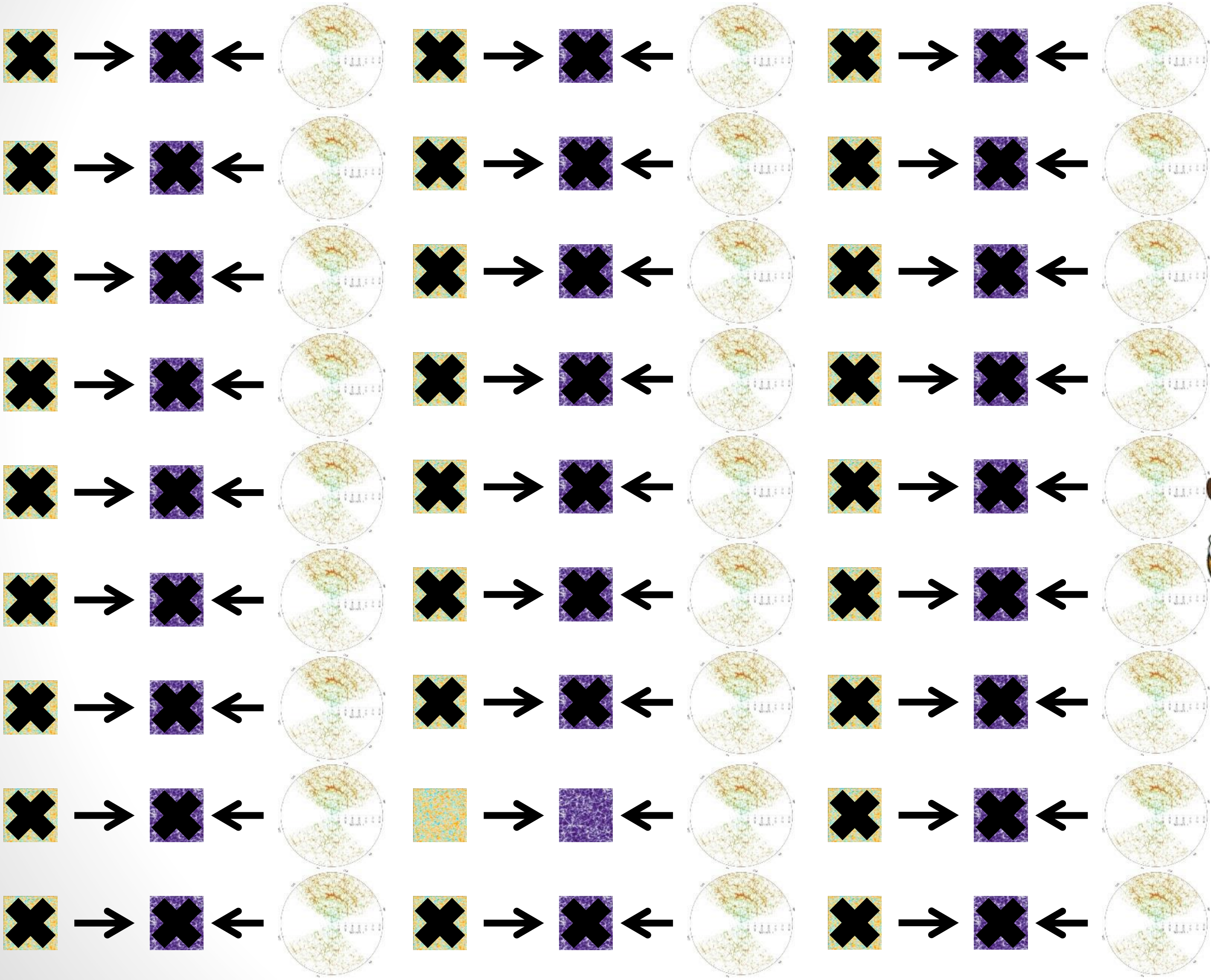
Observations

All possible ICs

All possible FCs



Bayesian forward modeling: the ideal scenario



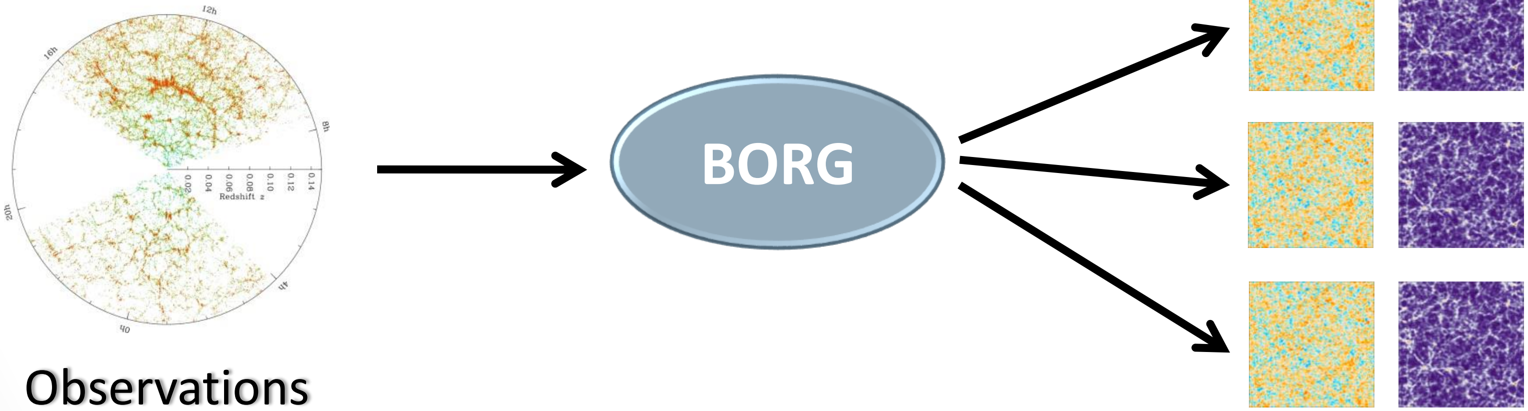
$d \approx 10^7$

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)



BORG

Observations

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

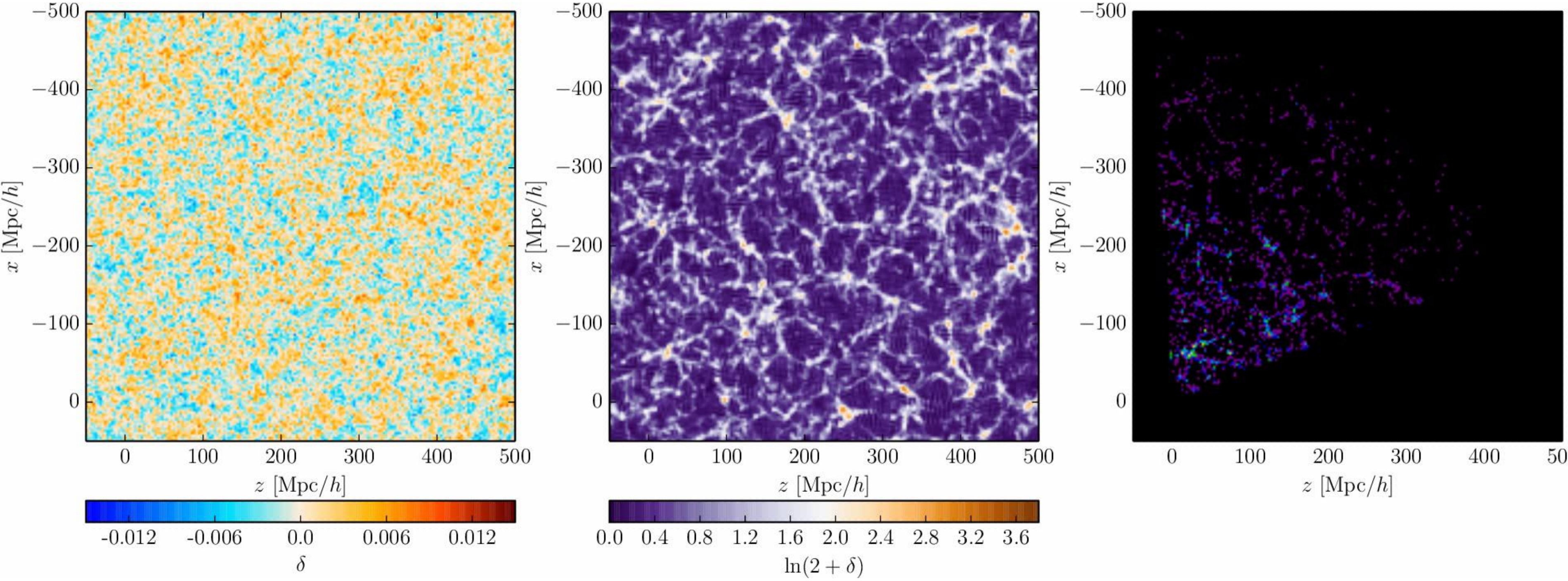
Samples of possible 4D states

Jasche & Wandelt 2013, arXiv:1203.3639

Jasche, FL & Wandelt 2015, arXiv:1409.6308

CHRONO-COSMOGRAPHY

BORG at work: SDSS chrono-cosmography



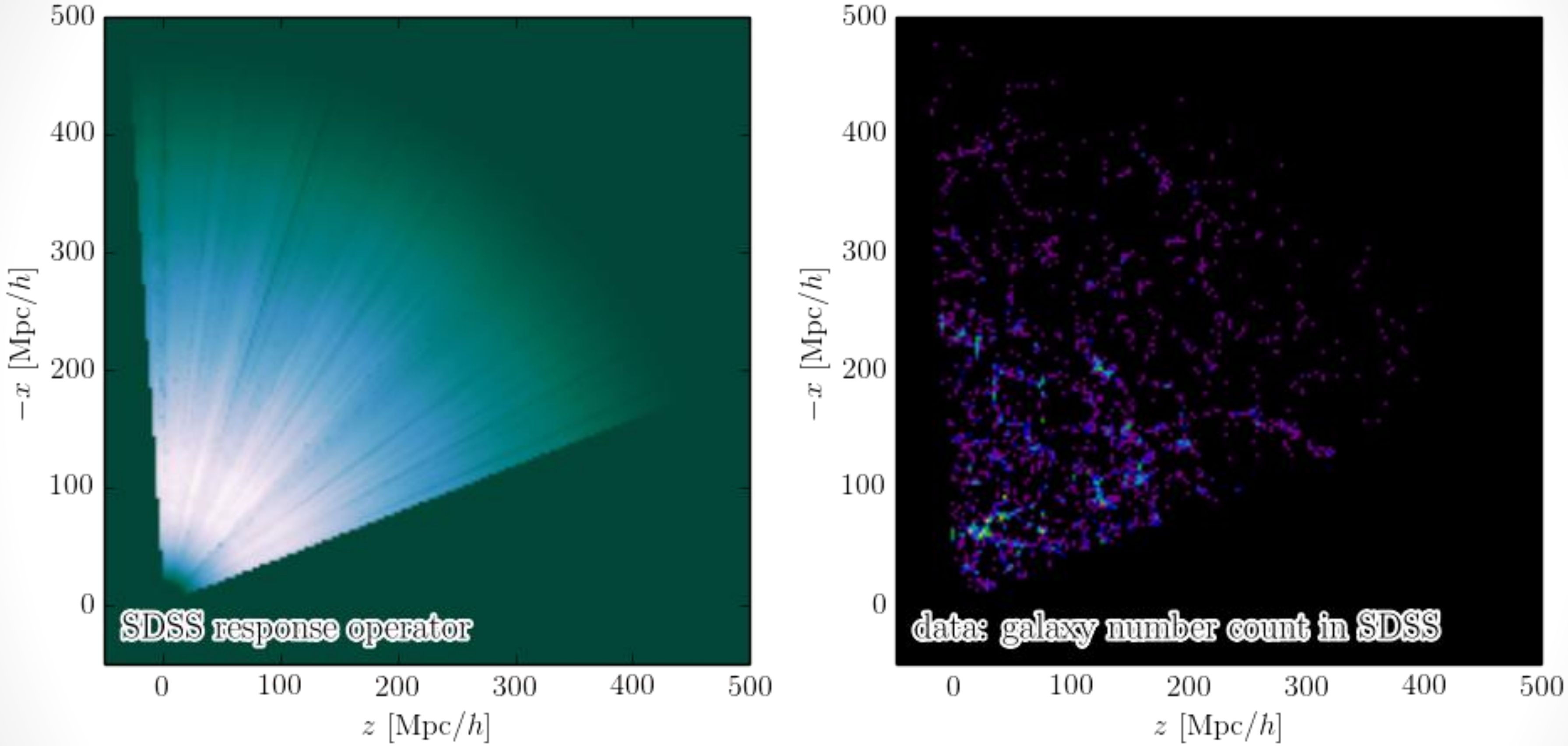
Initial conditions **Final conditions** **Observations**

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

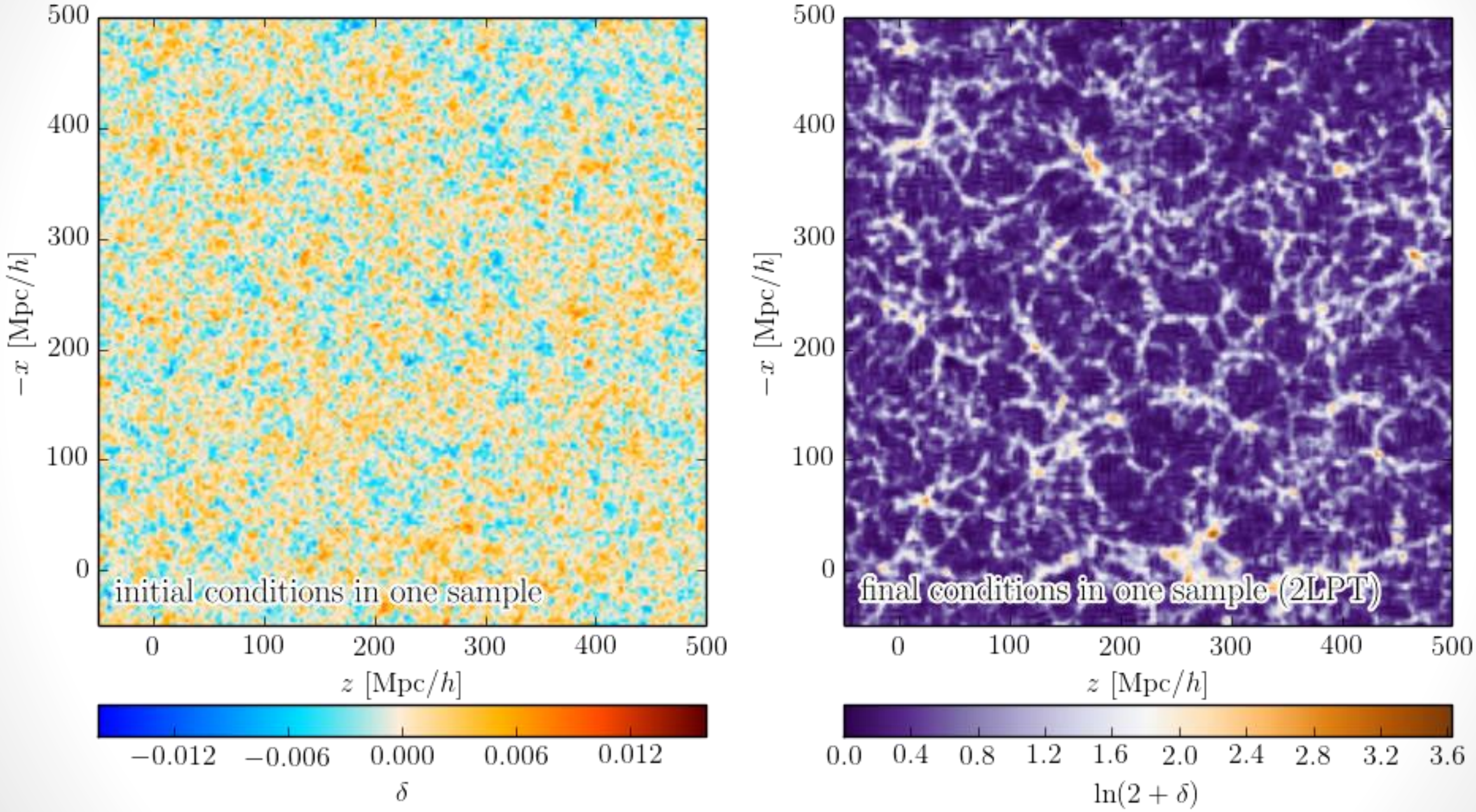
Bayesian chrono-cosmography from SDSS DR7



Data

Jasche, FL & Wandelt 2015, arXiv:1409.6308

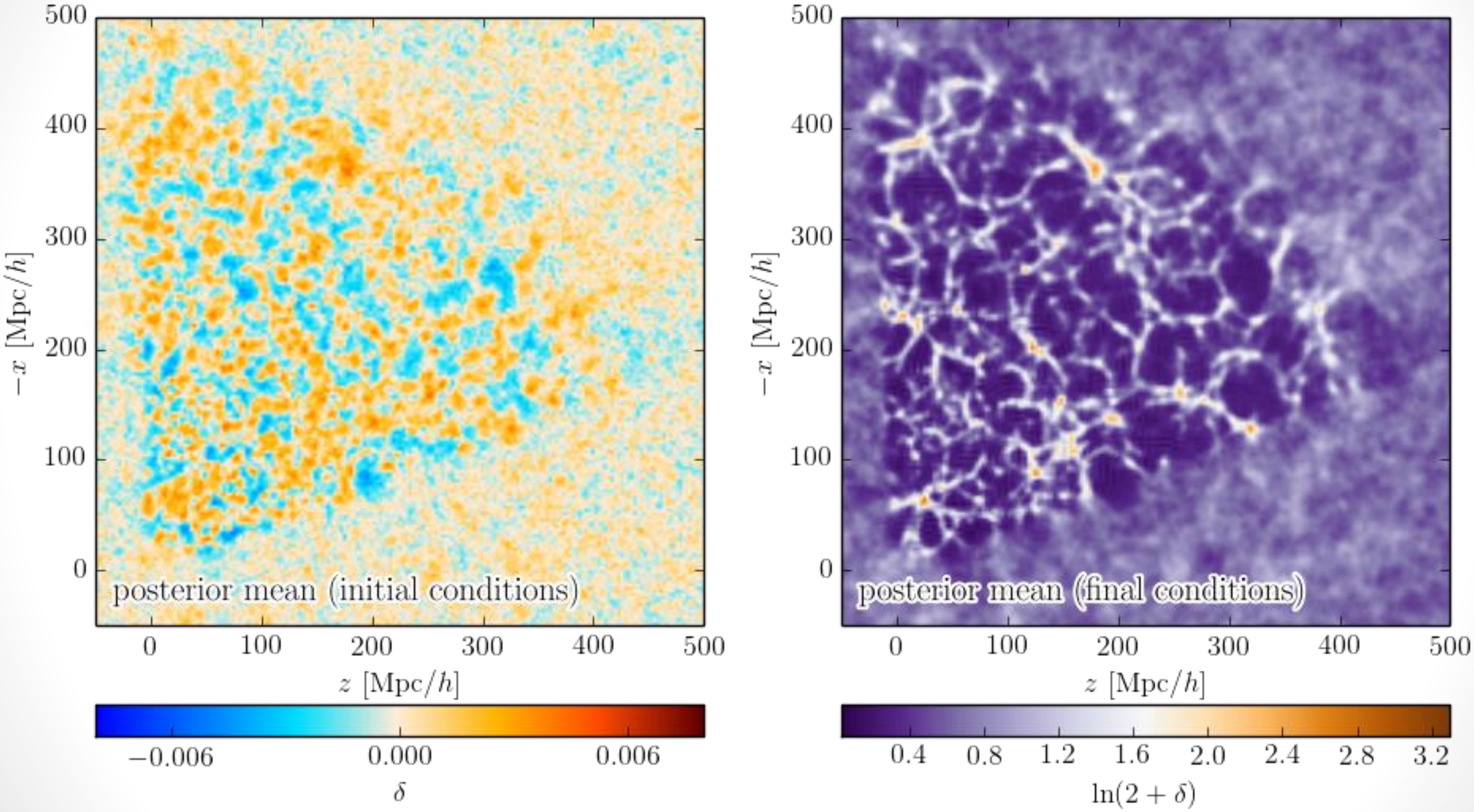
Bayesian chrono-cosmography from SDSS DR7



One sample

Jasche, FL & Wandelt 2015, arXiv:1409.6308

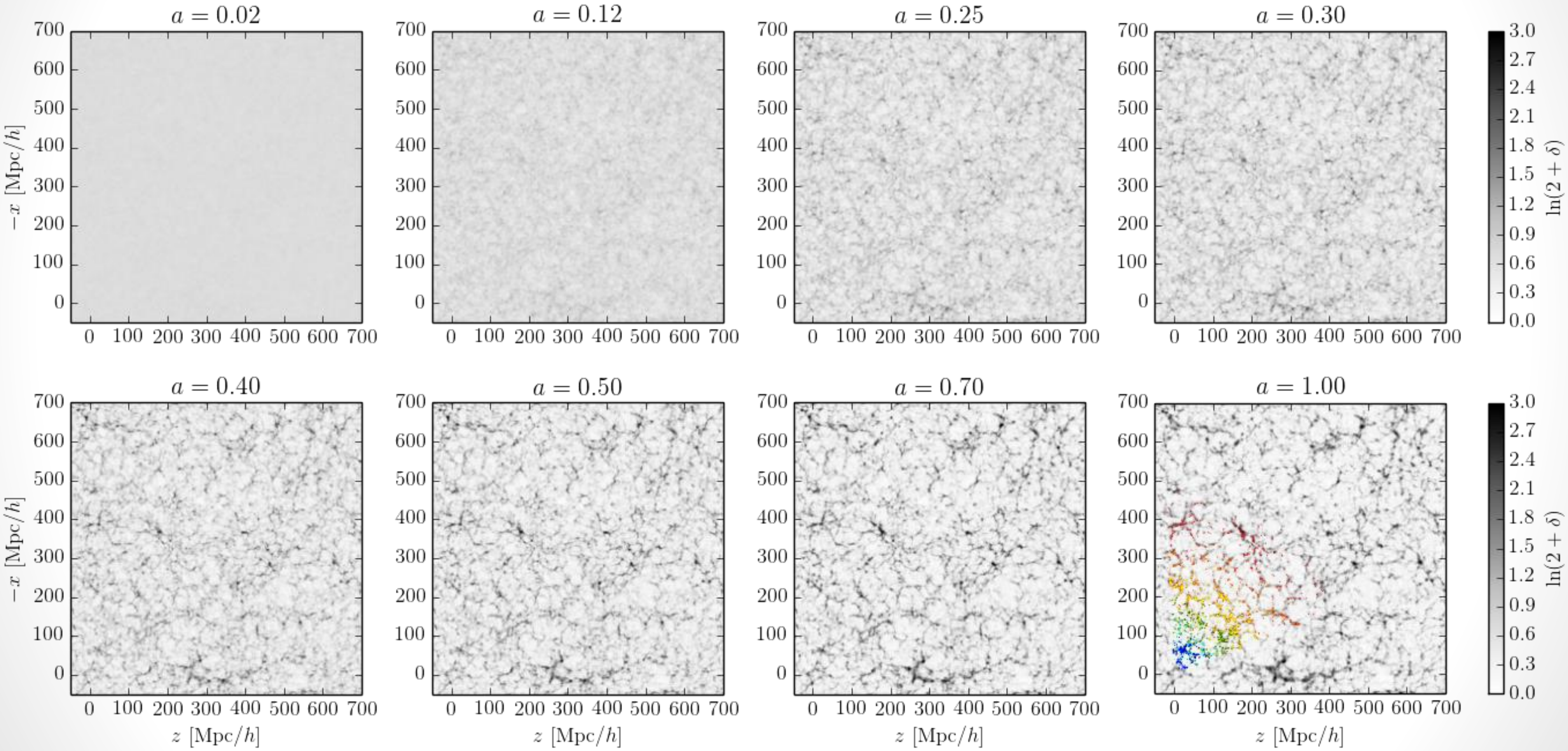
Bayesian chrono-cosmography from SDSS DR7



Posterior mean

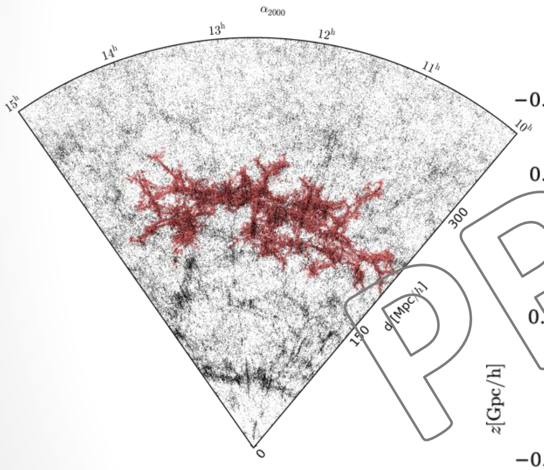
Jasche, FL & Wandelt 2015, arXiv:1409.6308

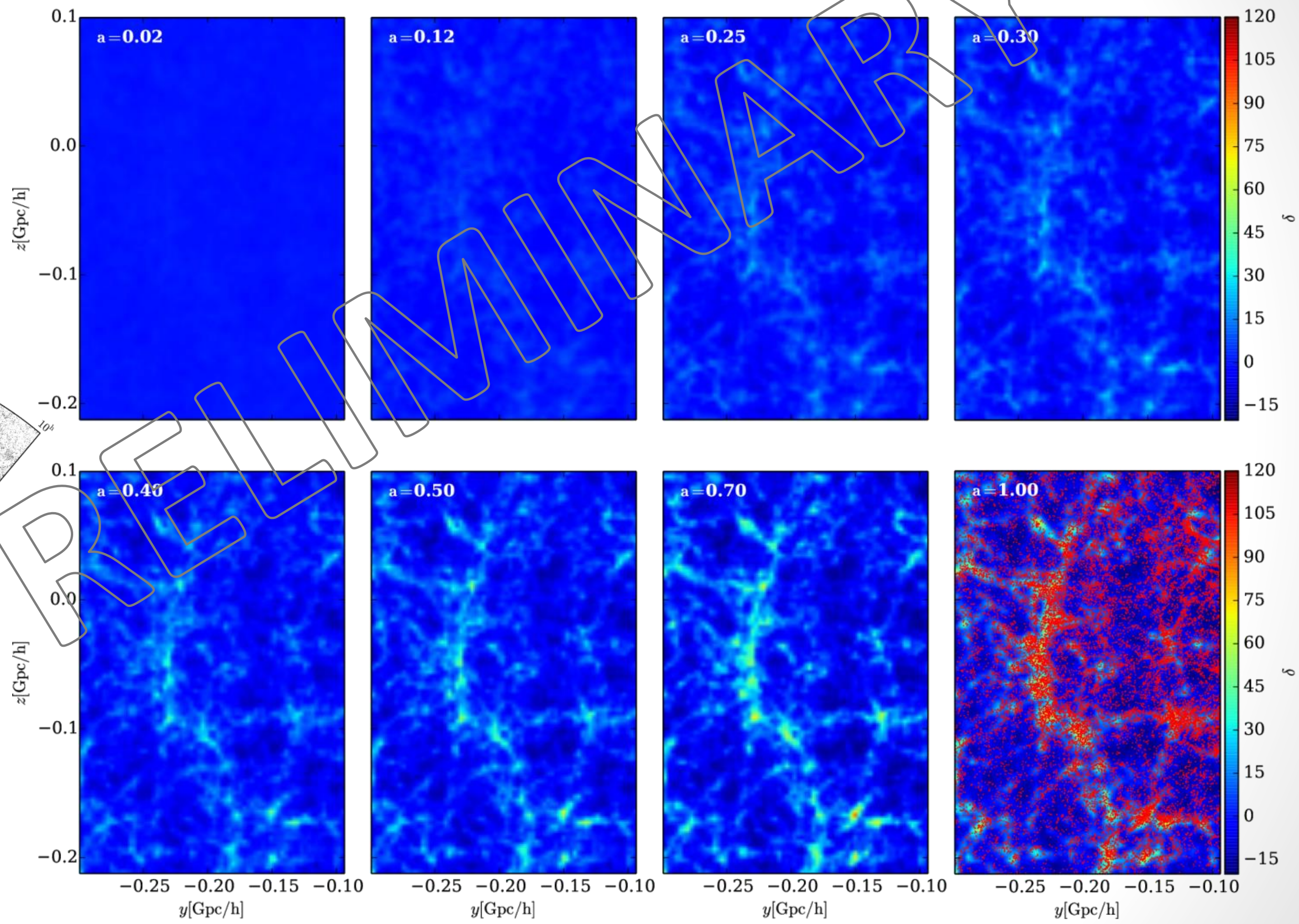
Evolution of cosmic structure



Jasche, FL & Wandelt 2015, arXiv:1409.6308

The formation history of the Sloan Great Wall

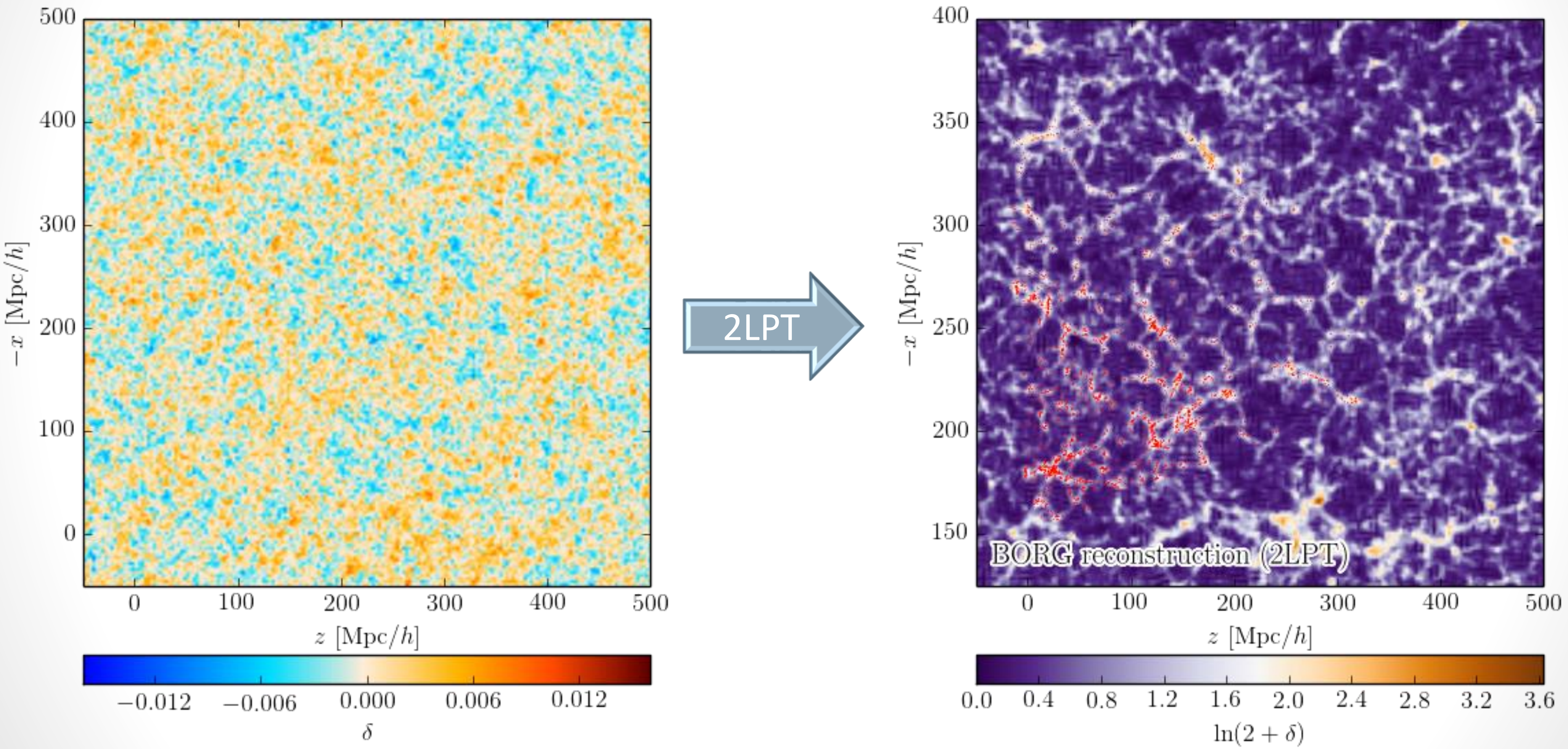




Jasche, Romano-Díaz, FL & Wandelt, in prep.

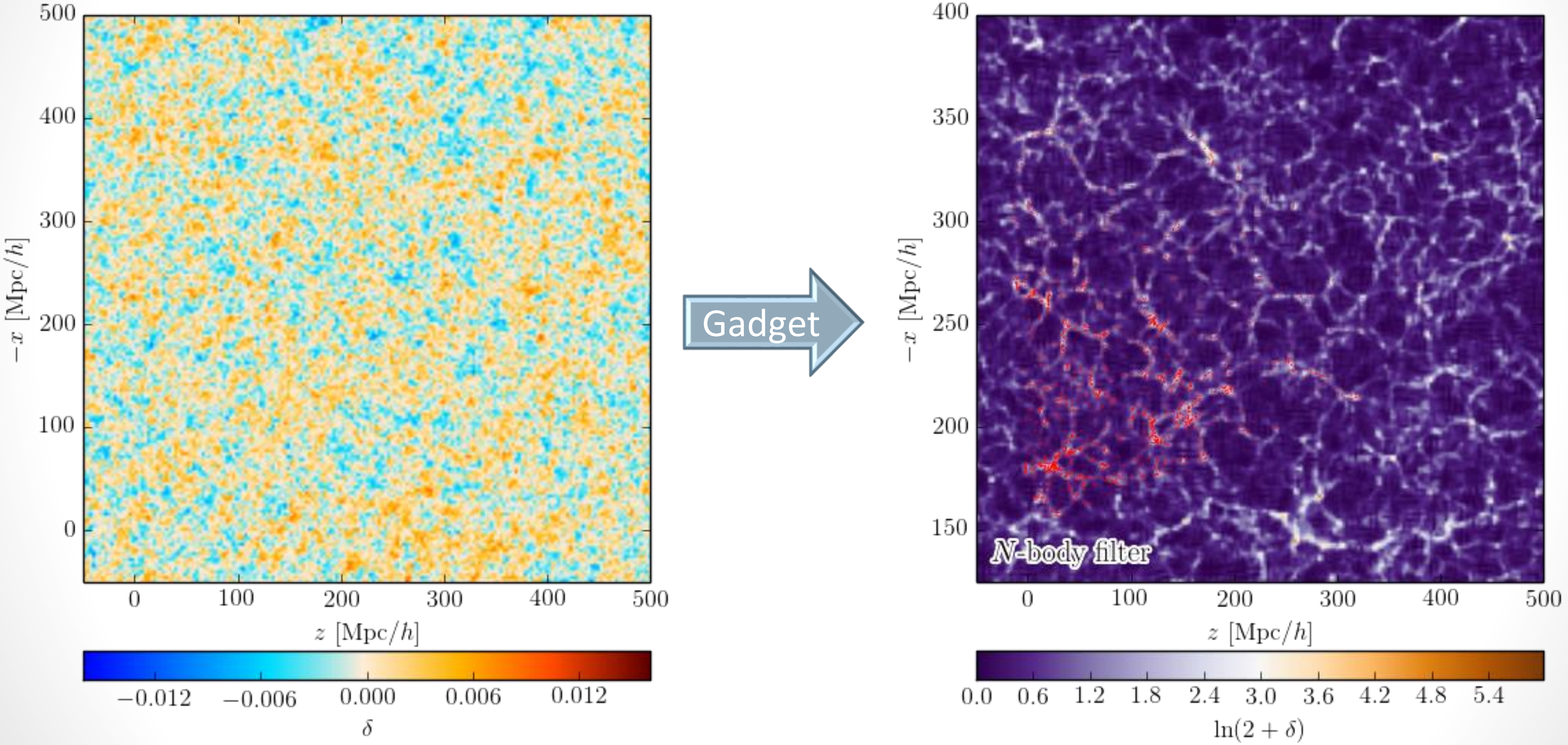
THE NON-LINEAR REGIME OF STRUCTURE FORMATION

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):

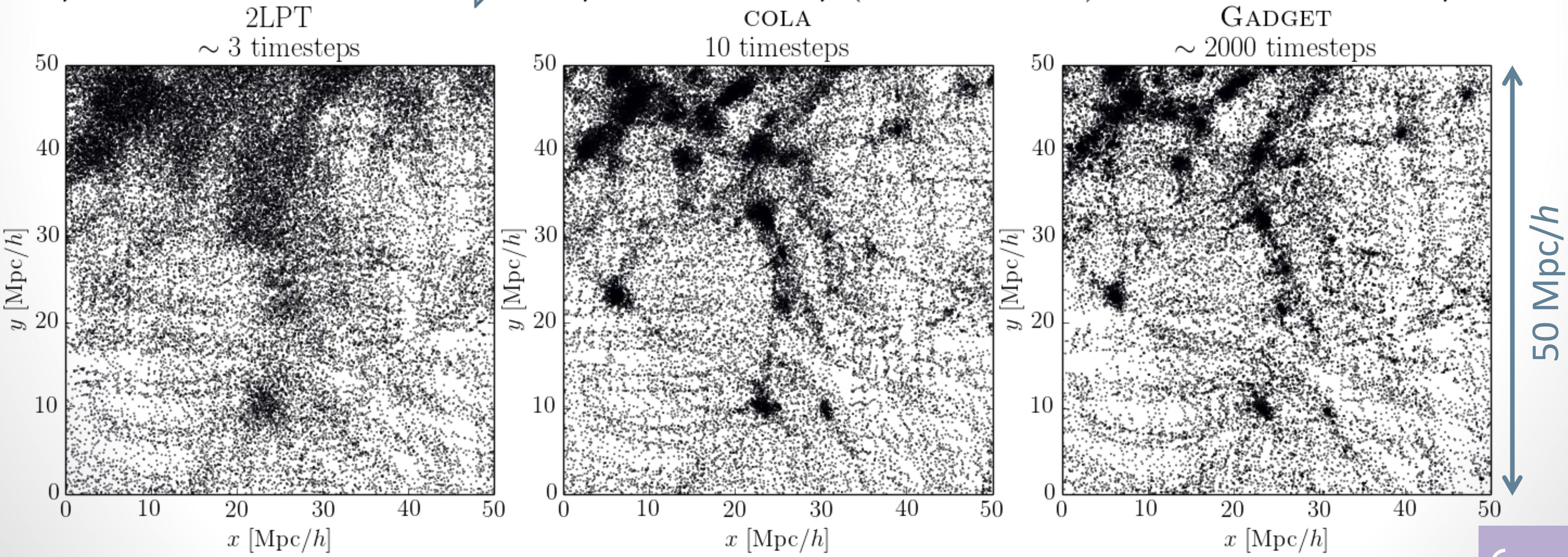
Standard:

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

⇒

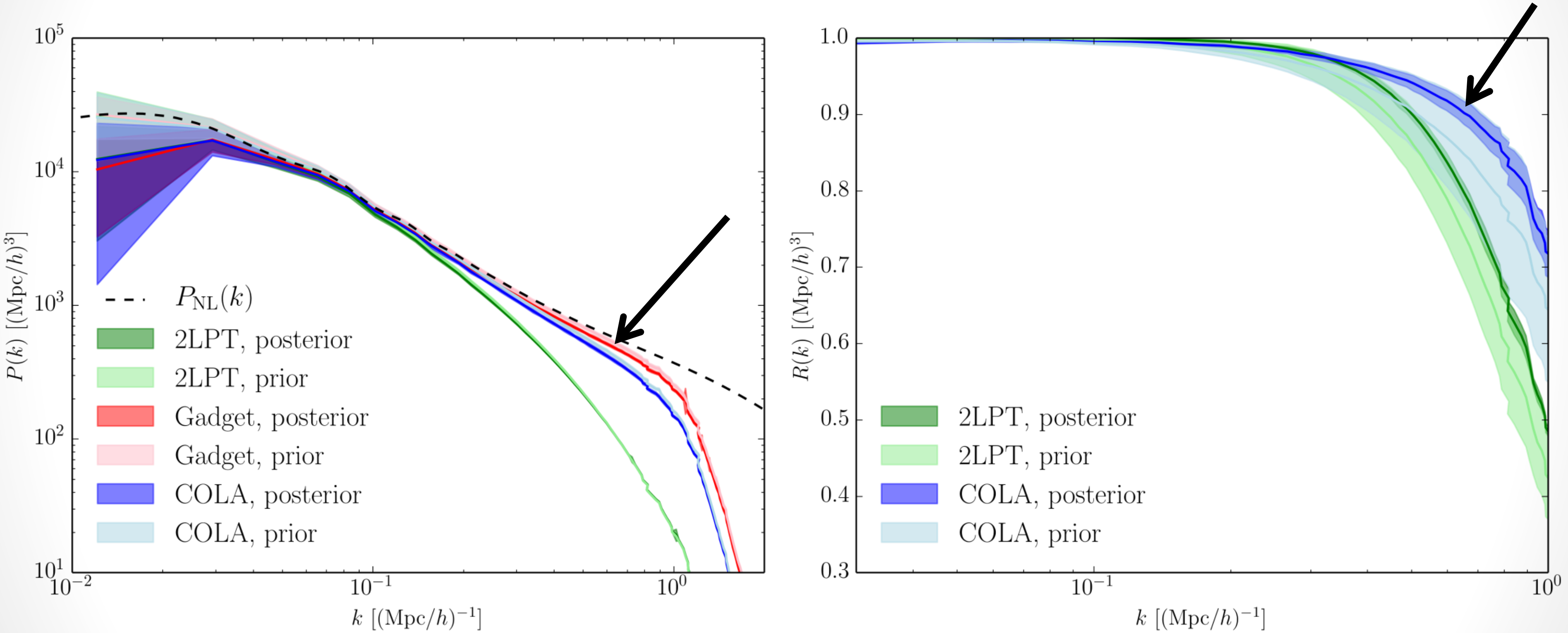
Modified:

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



Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322

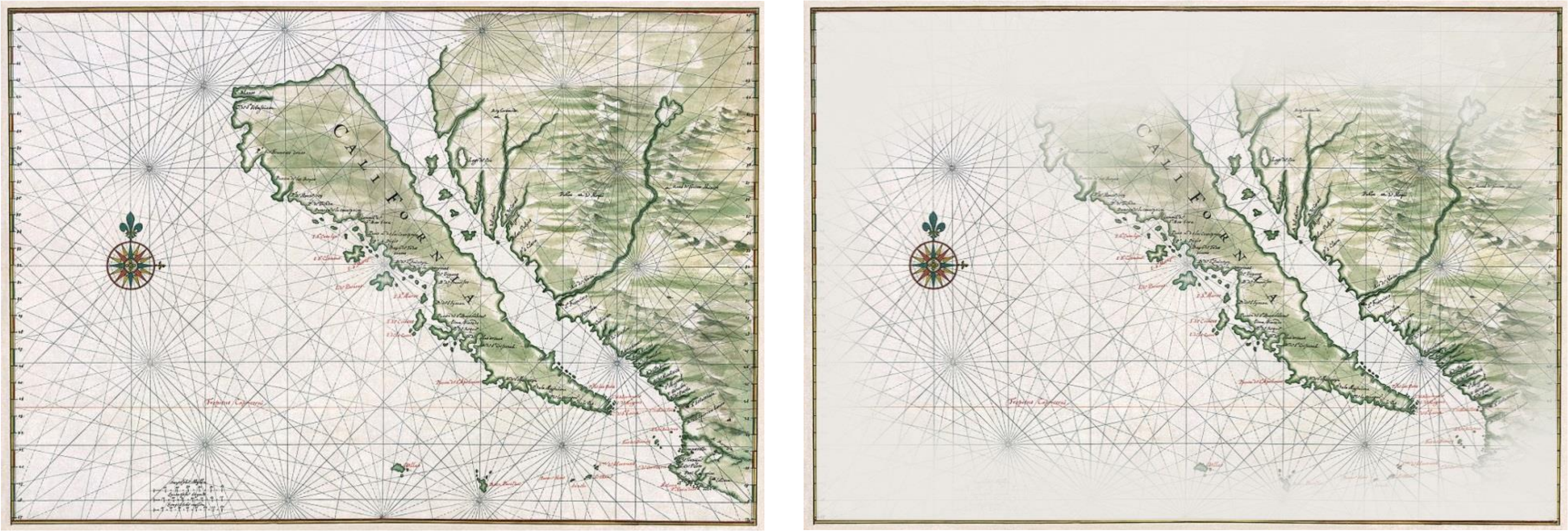
Non-linear filtering improves the fit



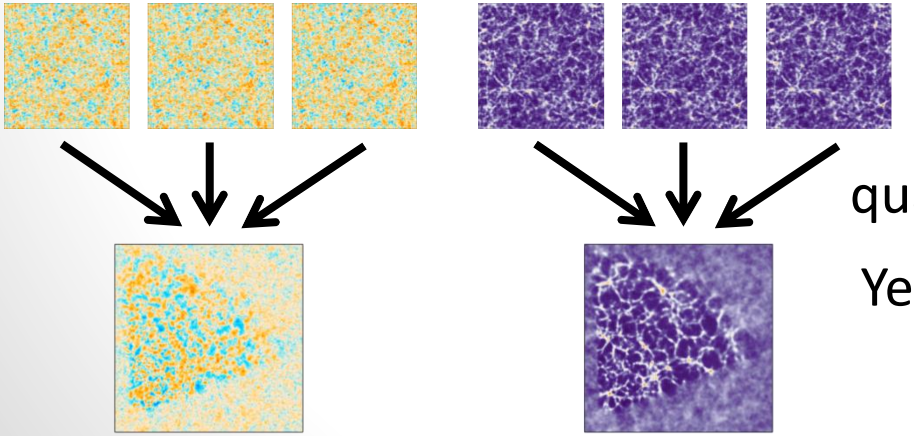
FL, Jasche, Sutter, Hamaus & Wandelt 2014, arXiv:1410.0355

HOW IS THE COSMIC WEB WOVEN?

Uncertainty quantification



Uncertainty quantification is crucial!



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

Yes, and this is what yields a connection with **information theory**!

Cosmic web classification procedures

void, sheet, filament, cluster?

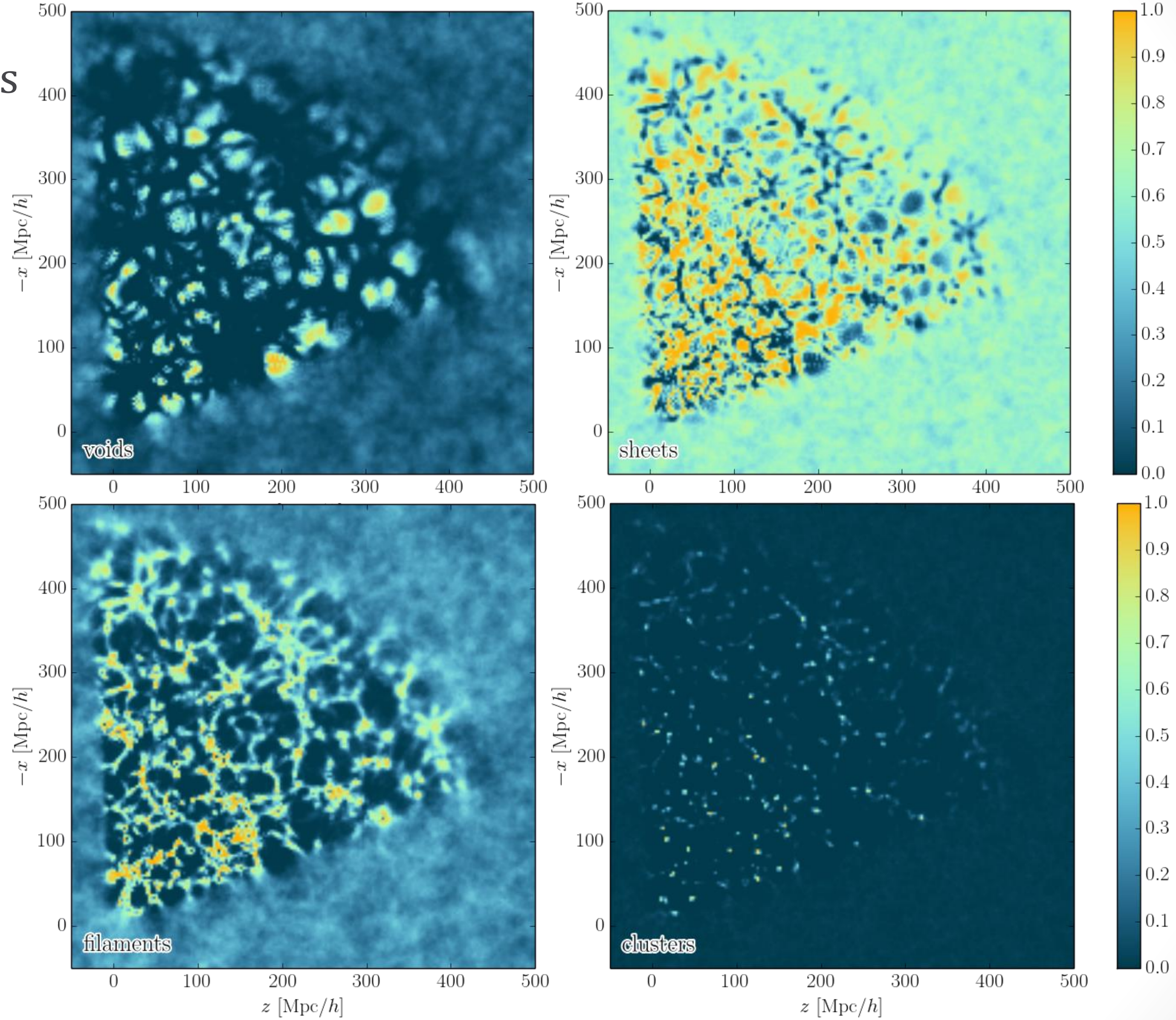
- The **T-web**:

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

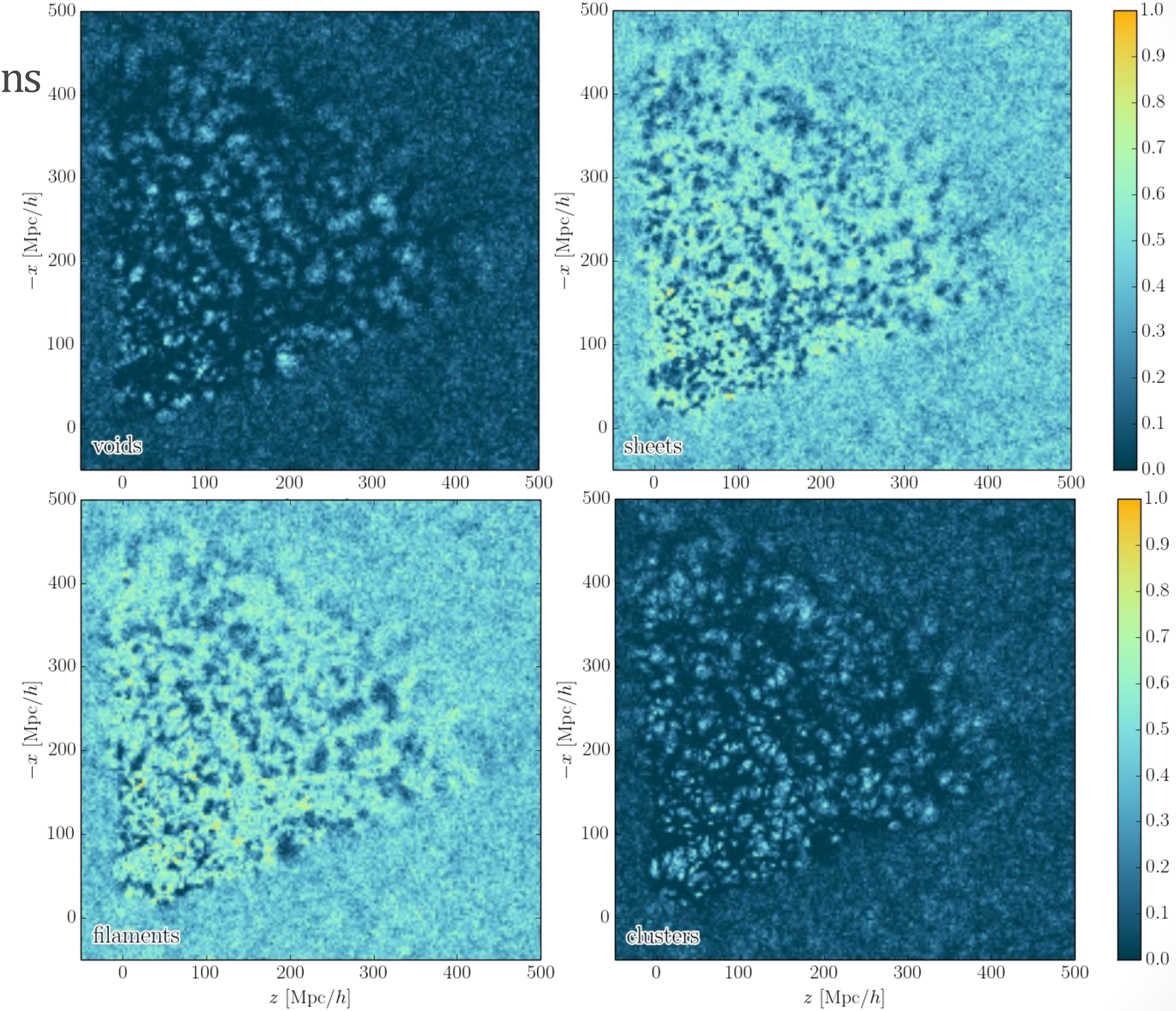
Final conditions



FL, Jasche & Wandelt 2015, arXiv:1502.02690

T-web structures inferred by BORG

Initial conditions



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”:

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

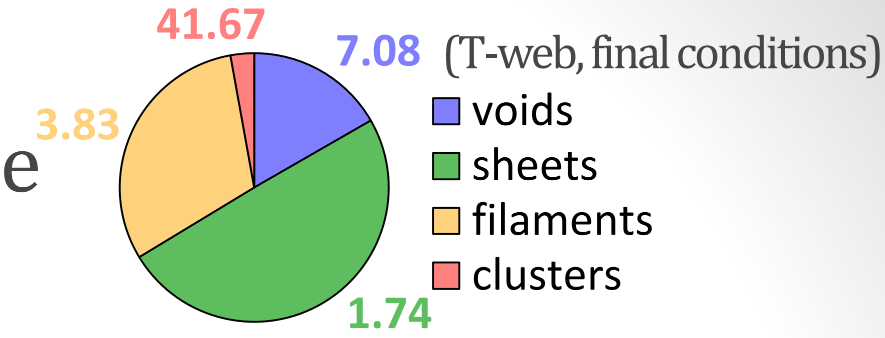
A problem of **Bayesian decision theory**:
one should take the action that maximizes the utility

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

- How to write down the gain functions?

FL, Jasche & Wandelt 2015, arXiv:1503.00730

Gambling with the Universe



- One proposal:

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

"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"

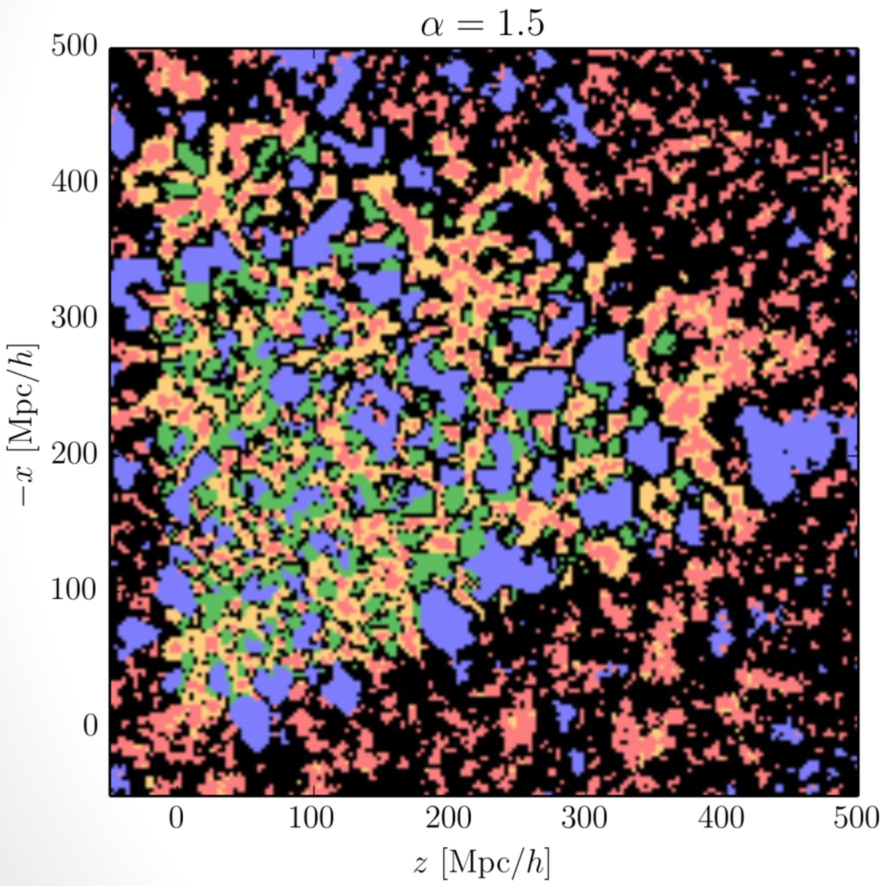
- With $\alpha = 1$, it's a *fair game* ⇨ always play
  ⇨ "**speculative map**" of the LSS
- Values $\alpha > 1$ represent an *aversion for risk*
  ⇨ increasingly "**conservative maps**" of the LSS

FL, Jasche & Wandelt 2015, arXiv:1503.00730

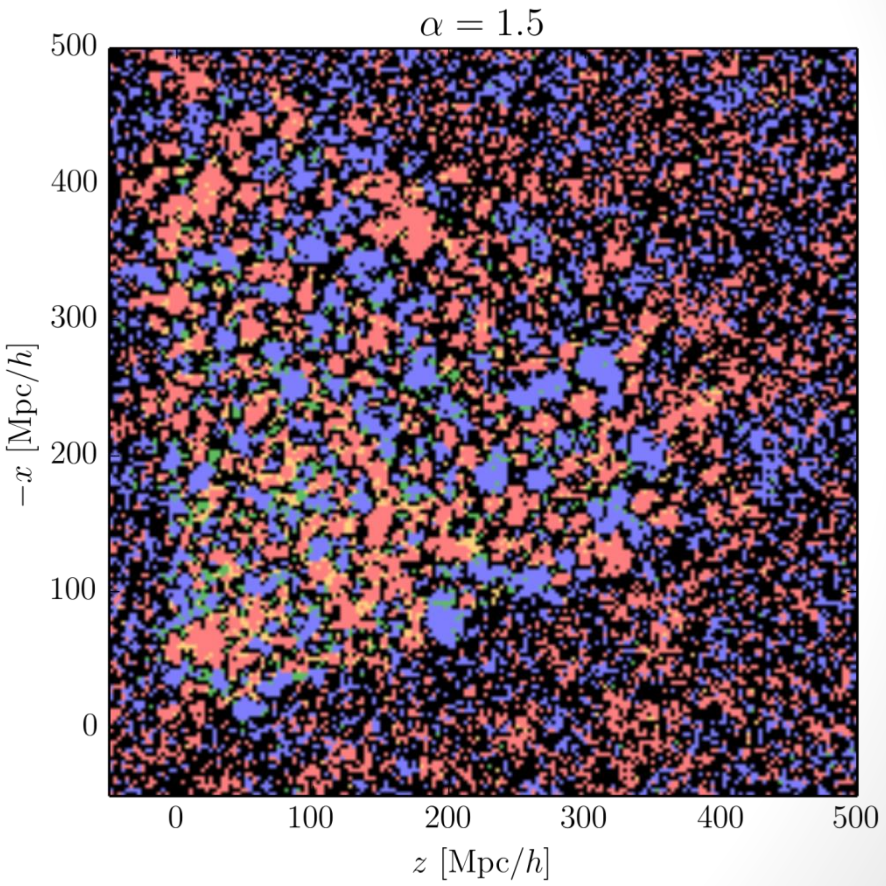
Playing the game…

🟦 voids
🟨 filaments
⬛ undecided
🟩 sheets
🟥 clusters

Final conditions



Initial conditions



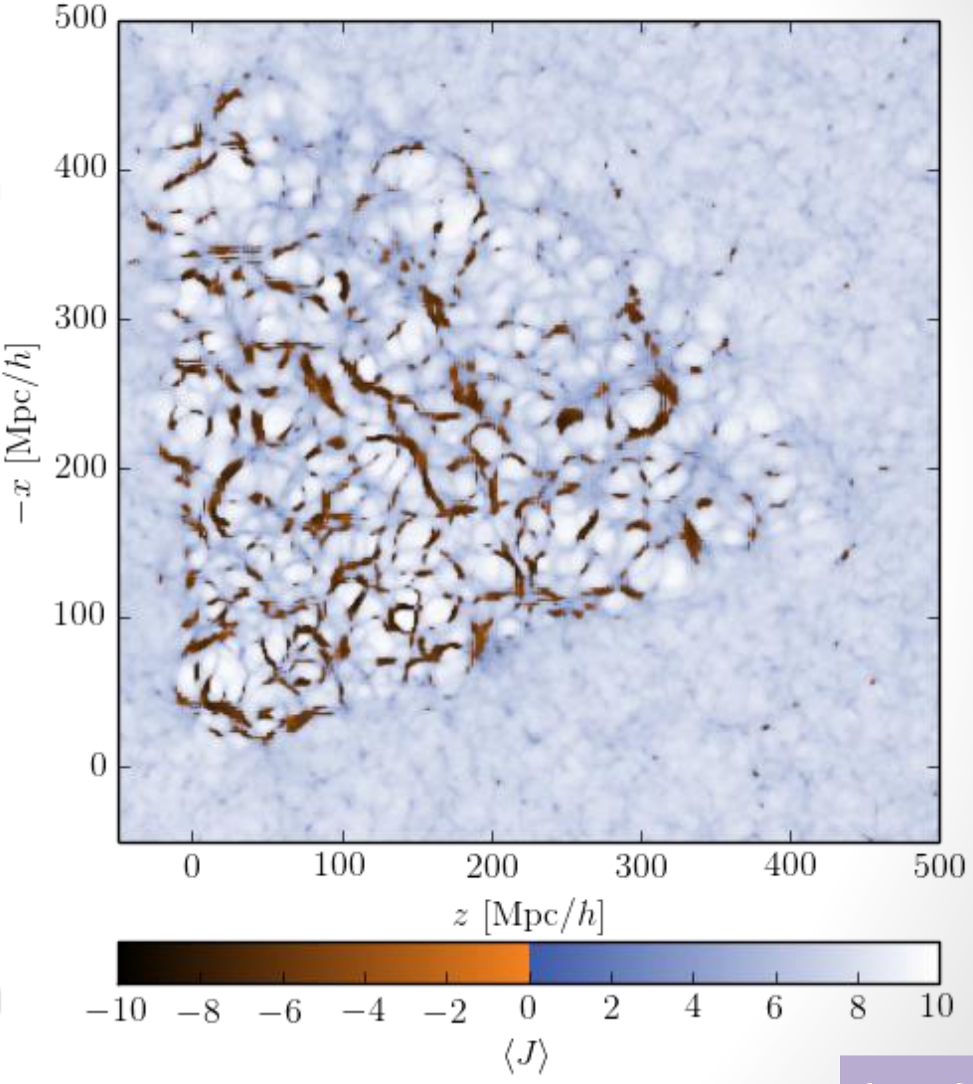
FL, Jasche & Wandelt 2015, arXiv:1503.00730

Inference of the dark matter phase-space sheet

- The dark matter phase-space sheet has been studied so far in simulations

e.g. Neyrinck 2012, arXiv:1202.3364
Abel, Hahn & Kaehler 2012, arXiv:1111.3944
Shandarin, Habib & Heitmann 2012, arXiv:1111.2366

- BORG infers **Lagrangian dynamics** in real data
- This is opening the way to **new confrontations** between data and theory
- Identified structures have a direct **physical interpretation**



FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

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** :

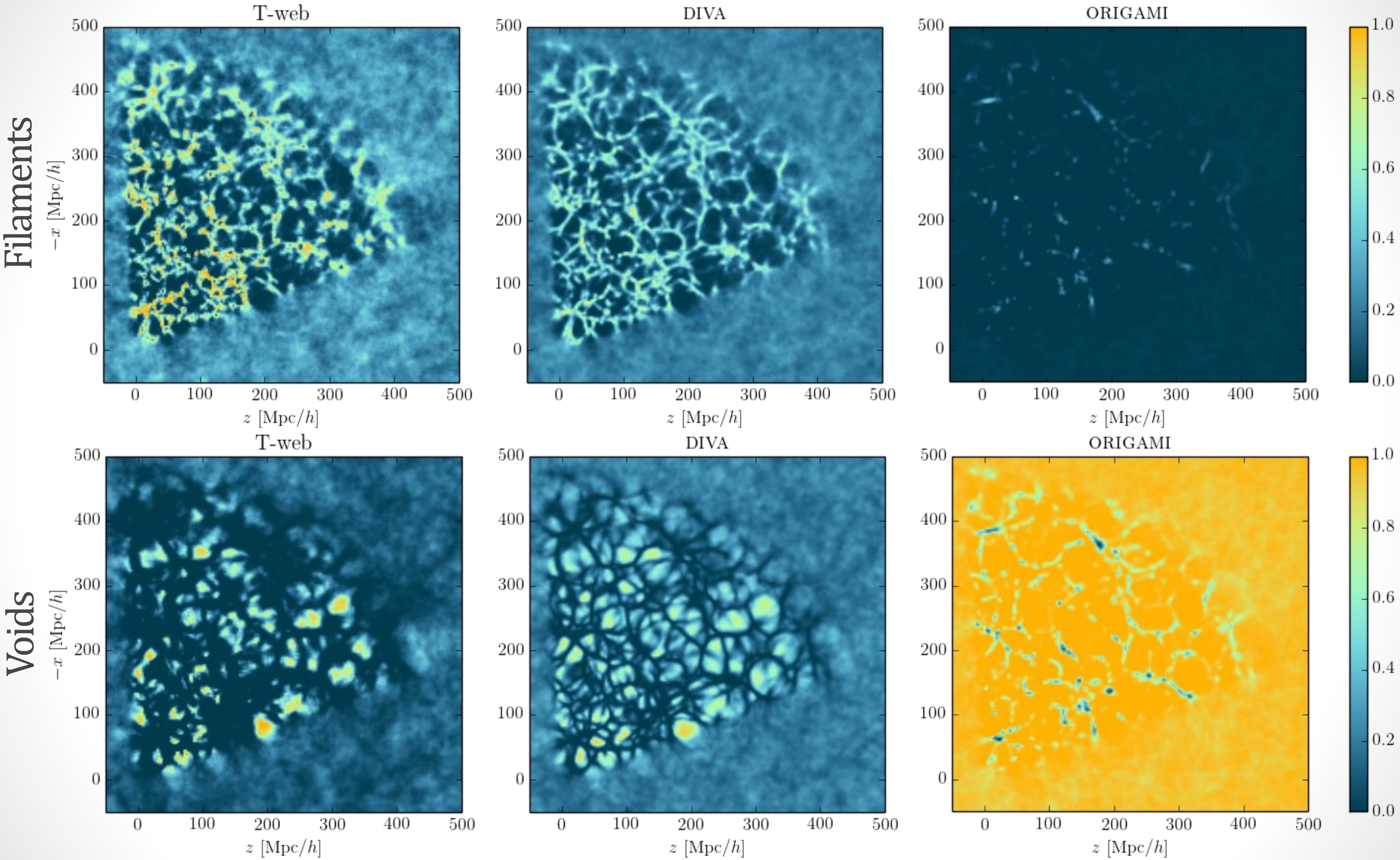
uses the dark matter "phase-space sheet" (number of orthogonal axes along which there is shell-crossing)

Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

Lagrangian classifiers

now usable in real data!

Comparing classifiers



FL, Jasche & Wandelt 2015, arXiv:1502.02690

FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

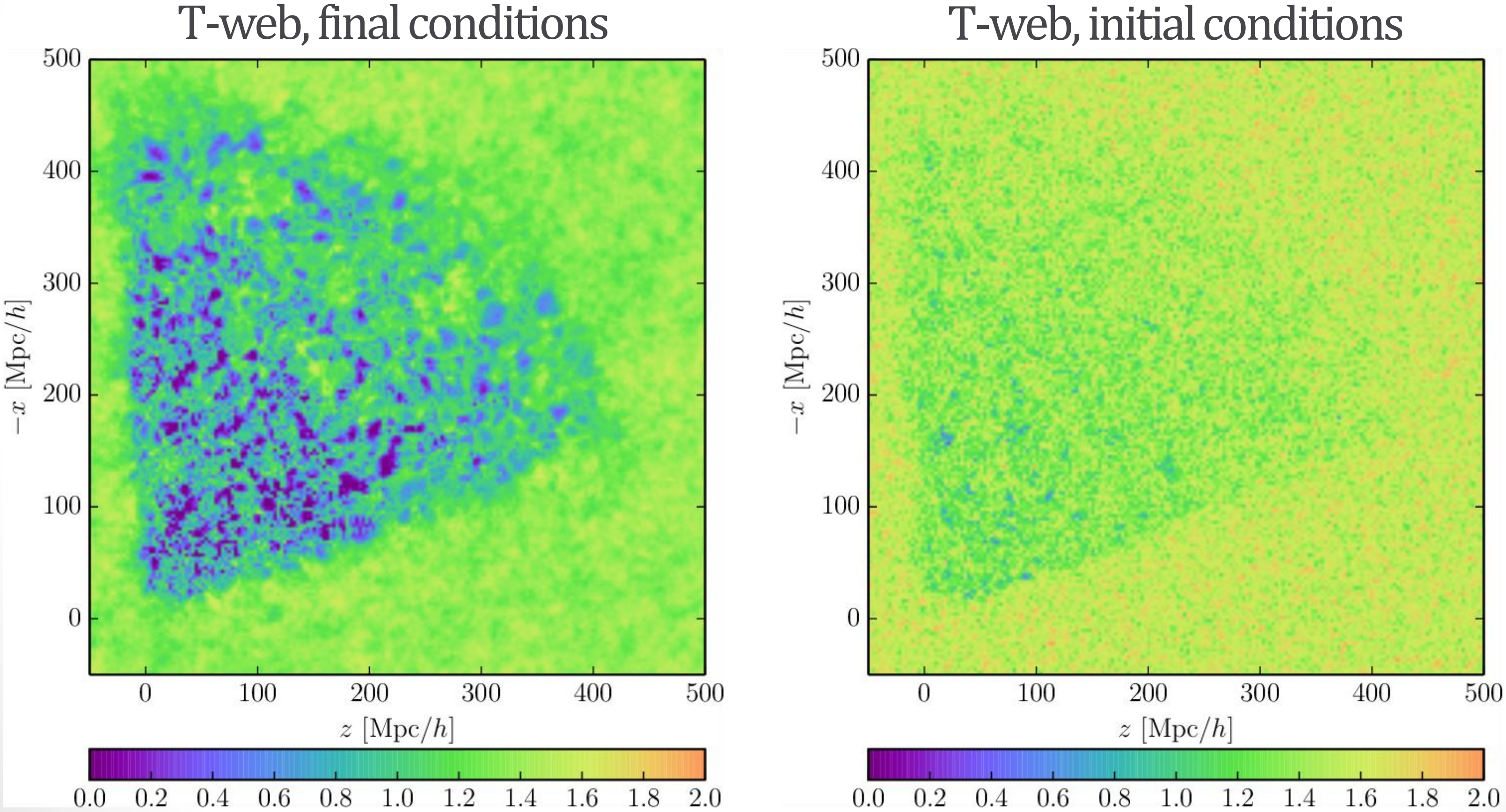
COSMIC WEB ANALYSIS AND INFORMATION THEORY

What is the information content of these maps?

Shannon entropy

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

in shannons (Sh)



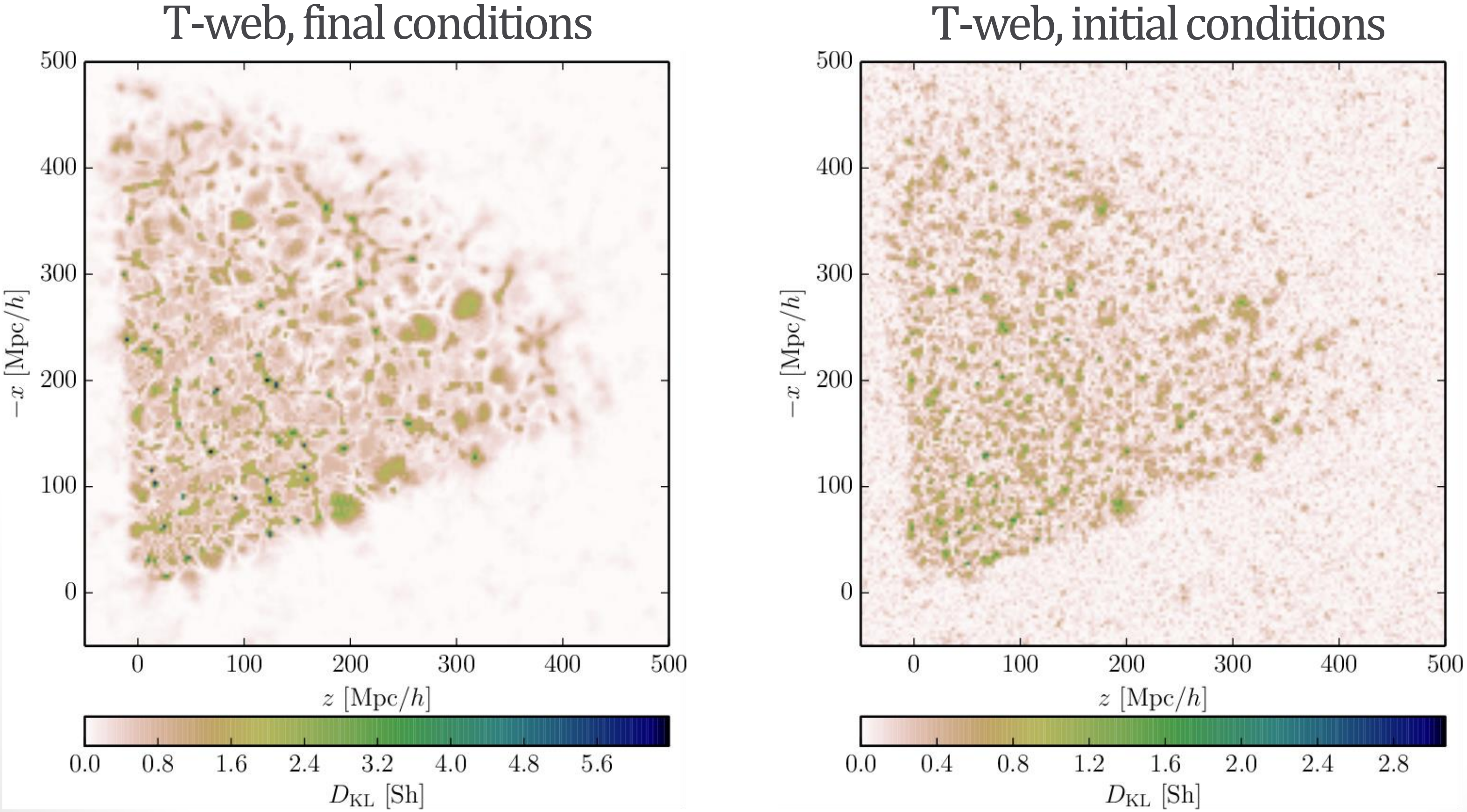
FL, Jasche & Wandelt 2015, arXiv:1502.02690

How much did the data surprise us?

information gain a.k.a. relative entropy or Kullback-Leibler divergence posterior/prior

$$D_{\mathrm{KL}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)||\mathcal{P}(\mathrm{T})\right] = \sum_i \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)\log_2\left(\frac{\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)}{\mathcal{P}(\mathrm{T}_i)}\right)$$

in Sh

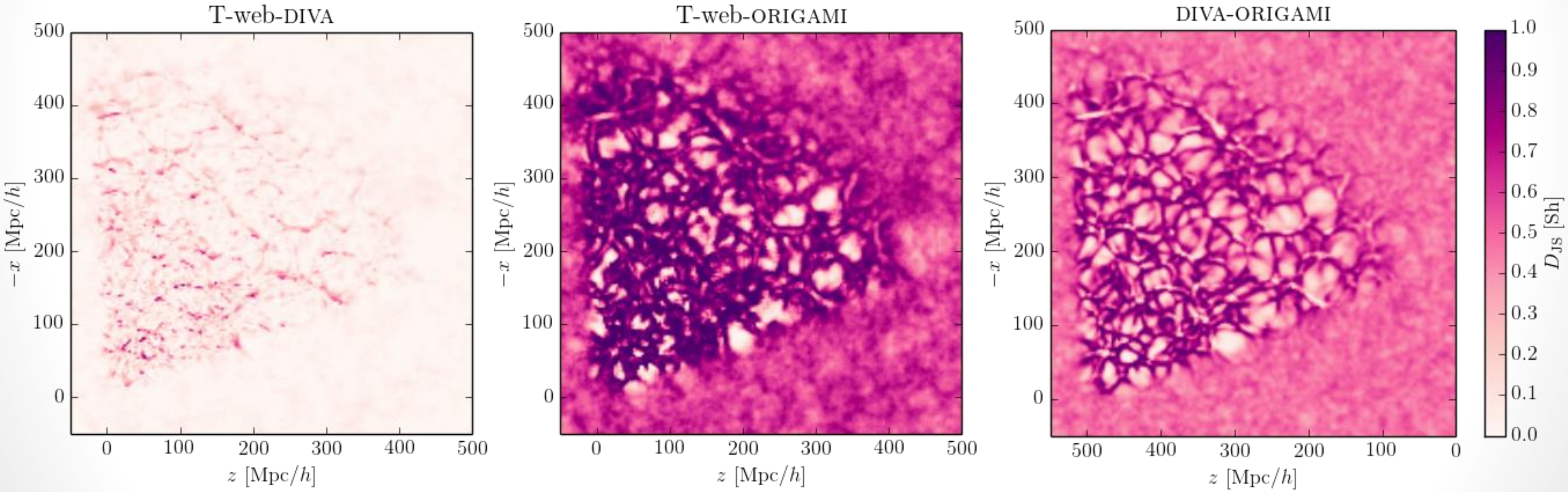


FL, Jasche & Wandelt 2015, arXiv:1502.02690

How similar are different classifications?

Jensen-Shannon divergence

$$D_{\mathrm{JS}}[\mathcal{P}:\mathcal{Q}] \equiv \frac{1}{2} D_{\mathrm{KL}}\left[\mathcal{P}||\frac{\mathcal{P}+\mathcal{Q}}{2}\right] + \frac{1}{2} D_{\mathrm{KL}}\left[\mathcal{Q}||\frac{\mathcal{P}+\mathcal{Q}}{2}\right]$$



(more about the Jensen-Shannon divergence later)

FL, Lavaux, Jasche & Wandelt, in prep.

Which is the best classifier?

- Can we extend the decision problem to the space of classifiers?
- As before, the idea is to maximize a utility function

$$U(\xi) = \langle U(d, \mathrm{T}, \xi) \rangle_{\mathcal{P}(d,\mathrm{T}|\xi)}$$

- An important notion: the **mutual information** between two random variables

$$I\left[X:Y\right] \equiv D_{\mathrm{KL}}[\mathcal{P}(x,y)||\mathcal{P}(x)\mathcal{P}(y)]$$
$$= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \mathcal{P}(x,y) \log_2 \left(\frac{\mathcal{P}(x,y)}{\mathcal{P}(x)\mathcal{P}(y)} \right)$$

- Property: $I\left[X:Y\right] = \langle D_{\mathrm{KL}}[\mathcal{P}(x|y)||\mathcal{P}(x)] \rangle_{\mathcal{P}(Y)}$

Mutual information is the expectation of the Kullback-Leibler divergence of the conditional from the unconditional distribution.

FL, Lavaux, Jasche & Wandelt, in prep.

Bayesian problems

1. Optimal parameter inference

   example: information content of cosmic web maps

2. Model selection

   example: dark energy models

3. Prediction of future observations

   example: galaxy properties

(Some numerical results for classifier utilities in the upcoming paper)

FL, Lavaux, Jasche & Wandelt, in prep.

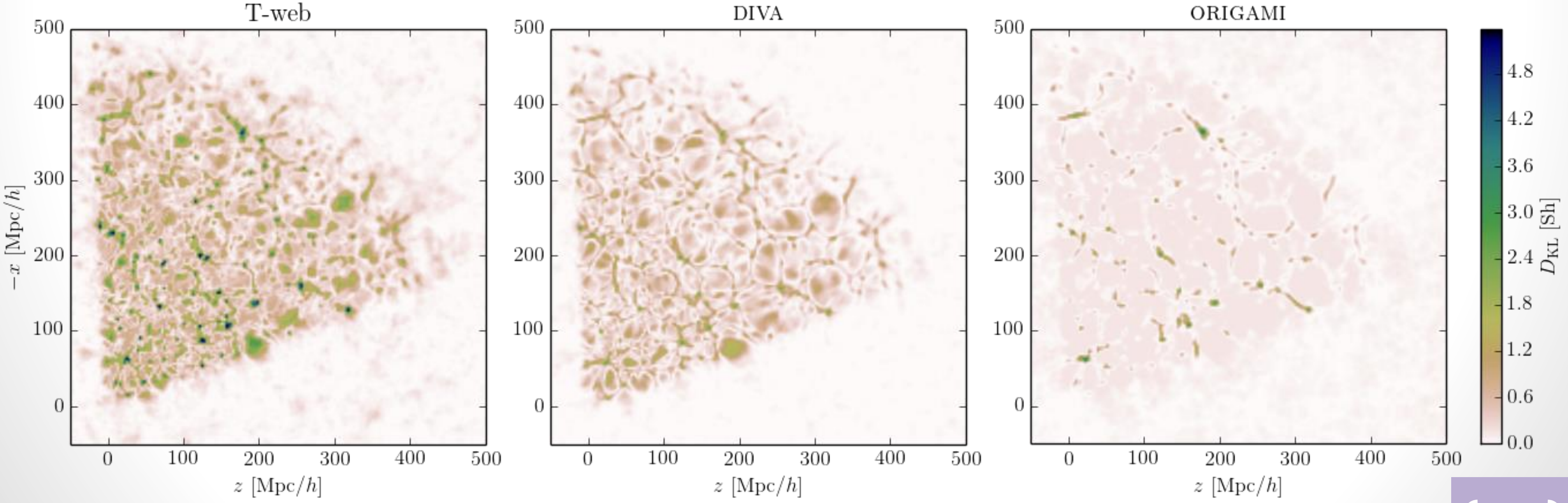
1. Utility for parameter inference:

cosmic web analysis

- In analogy with the formalism of **Bayesian experimental design**: maximize the **expected information gain** for cosmic web maps $U_1(d,\xi)(\vec{x}_k) = D_{\mathrm{KL}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\xi)||\mathcal{P}(\mathrm{T}|\xi)\right]$

$$\boxed{U_1(\xi) = I[\mathrm{T}\!:\!d|\xi]}$$

classification data



FL, Lavaux, Jasche & Wandelt, in prep.

2. Utility for model selection:

dark energy equation of state

- For example, consider three dark energy models with

$$w = -0.9, w = -1, w = -1.1$$

- The **Jensen-Shannon divergence** between posterior predictive distributions can be used as an approximate **predictor for the change in the Bayes factor**

Vanlier *et al.* 2014, BMC Syst Biol 8, 20 (2014)

- In analogy: $U_2(d,\xi)(\vec{x}_k) = D_{\mathrm{JS}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\mathcal{M}_1) : \mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\mathcal{M}_2)|\xi\right]$

$$\boxed{U_2(\xi) = I\left[\underset{\substack{\nearrow \\ \text{model classifier}}}{\mathcal{M}} : \underset{\substack{\nwarrow \\ \text{mixture distribution}}}{\mathcal{R}(d)}|\xi\right]}$$

$$\mathcal{R}(d) \equiv \frac{\mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\mathcal{M}_1) + \mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\mathcal{M}_2)}{2}$$

FL, Lavaux, Jasche & Wandelt, in prep.

3. Utility for prediction of new data:

galaxy colors

- Maximize the **expected information gain** for some new quantity $U_3(d, \mathrm{T}, \xi) = D_{\mathrm{KL}}\left[\mathcal{P}(c|d, \mathrm{T}, \xi)||\mathcal{P}(c|\xi)\right]$

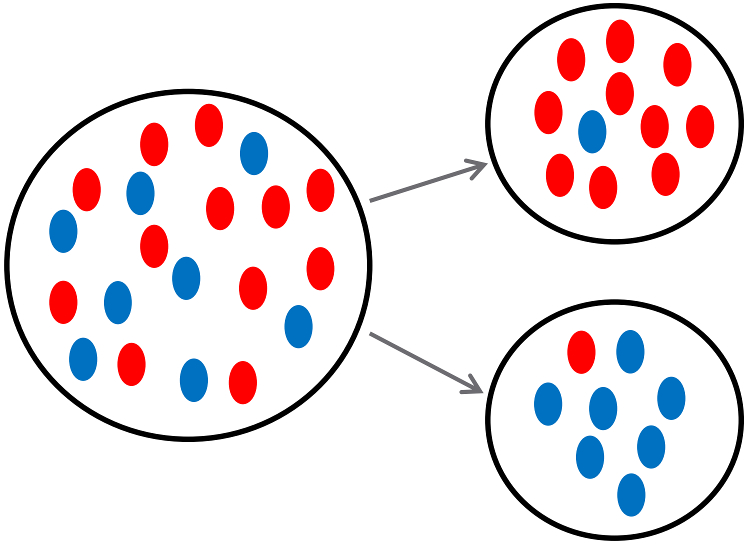
$$\boxed{U_3(\xi) = I[\underset{\substack{\nearrow \\ \text{predicted data}}}{c} : \underset{\substack{\nwarrow \\ \text{classification}}}{\mathrm{T}}|\xi]}$$

FL, Lavaux, Jasche & Wandelt, in prep.

3. Utility for prediction of new data:

galaxy colors

- How to compute the information gain?



child1 entropy:

$$H = -\frac{10}{11}\log_2\left(\frac{10}{11}\right) - \frac{1}{11}\log_2\left(\frac{1}{11}\right) = 0.4395$$

child2 entropy:

$$H = -\frac{8}{9}\log_2\left(\frac{8}{9}\right) - \frac{1}{9}\log_2\left(\frac{1}{9}\right) = 0.5033$$

parent entropy:

$$H = -\frac{8}{20}\log_2\left(\frac{8}{20}\right) - \frac{12}{20}\log_2\left(\frac{12}{20}\right) = 0.9709$$

weighted average entropy of children:

$$\frac{11}{20} \times 0.4395 + \frac{9}{20} \times 0.5033 = 0.4682$$

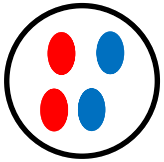
information gain for this split: $0.9709 - 0.4682 = 0.5027\ \mathrm{Sh}$

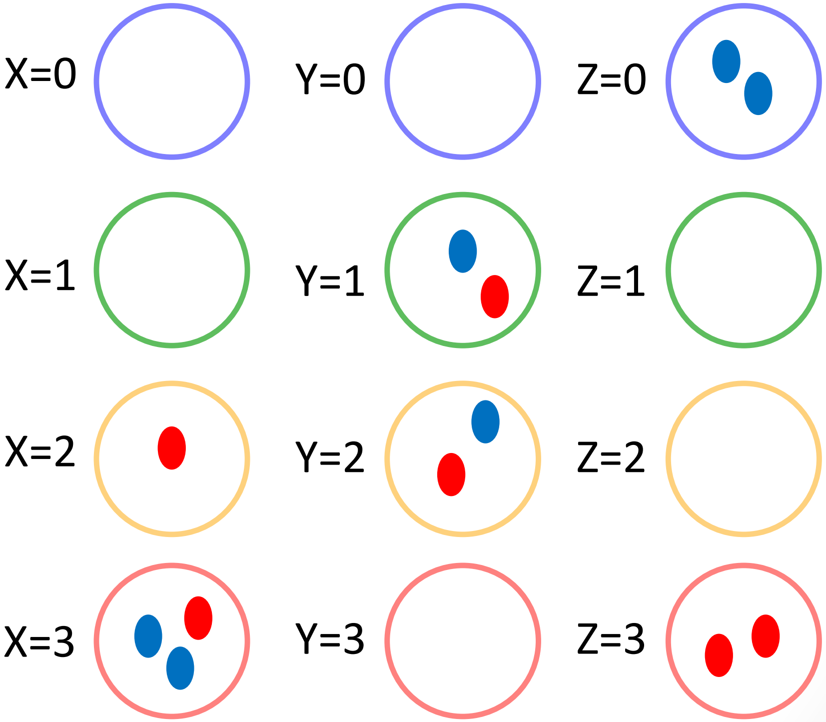
3. Utility for prediction of new data:

galaxy colors

- A **supervised machine learning** problem!
  - 3 **features** = classifications (T-web, DIVA, ORIGAMI) with
  - 4 **possible values** (void, sheet, filament, cluster)
  - 2 **classes** (red, blue)

X	Y	Z	C
3	2	3	I
3	1	3	I
2	2	0	II
3	1	0	II





$\mathrm{IG} = 0.3112$ $\mathrm{IG} = 0.$ $\mathrm{IG} = 1.$

no gain: worst

best!

FL, Lavaux, Jasche & Wandelt, in prep.

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
- Probabilistic analysis of the cosmic web yields a data-supported **connection between cosmology and information theory**