The initial conditions and the large-scale structure of the Universe

Florent Leclercq

Institut d’Astrophysique de Paris

June 12th, 2013

with: Jens Jasche (IAP), Héctor Gil-Marín (U. Portsmouth/U. Barcelona), Svetlin Tassev (U. Princeton), Benjamin Wandelt (IAP/U. Illinois), Matias Zaldarriaga (IAS Princeton)

Why cosmology?

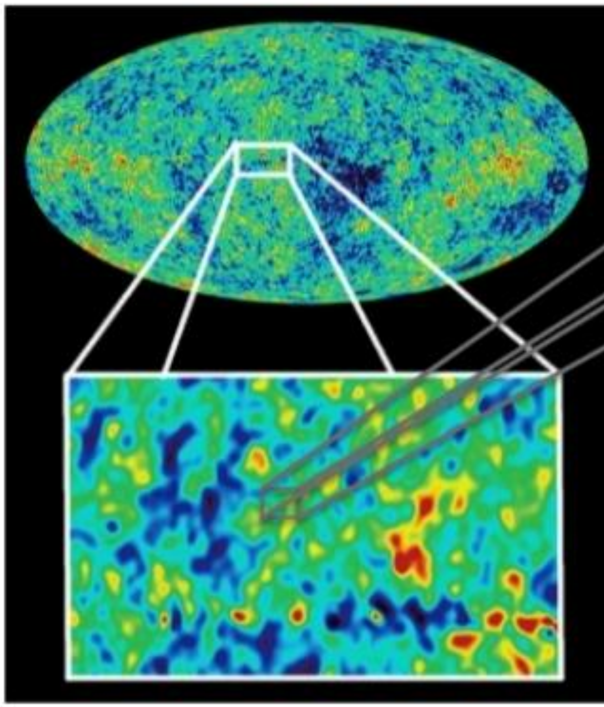
- Cosmology is the science of the Universe as a *physical system*, where the Universe means "*everything that exists in the physical sense*".
- Important ideas:
  - The Universe *in its globality* can be treated as a physical system
  - Science can deal with *times and places we cannot experience* (the observable Universe is a strict subset of the Universe)

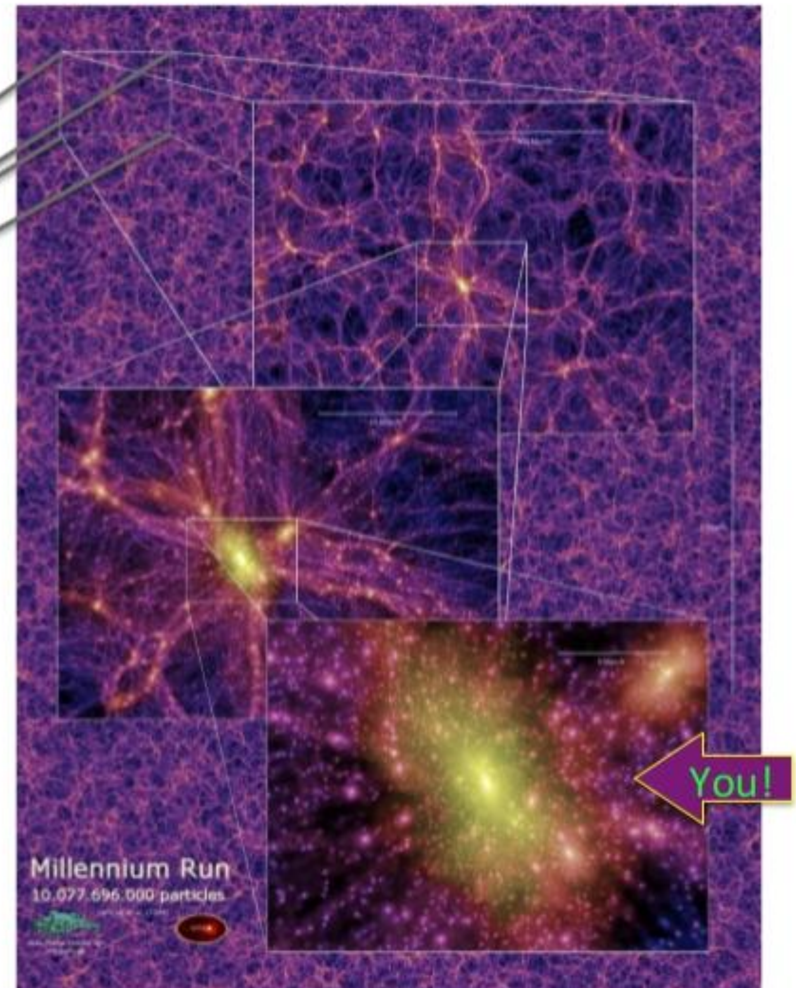
The specificities of cosmology

- *Unicity*. The experience is unique and irreproducible by physical experimentation. The properties of the Universe cannot be determined statistically on a set.
- *Observer*. The observer is part of the physical system described. The measure may have an influence on the system.
- *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.
- *Initial conditions*. The is no exteriority nor anteriority, which makes the initial conditions of the Universe particular with respect to other physical phenomena.

The inhomogeneous Universe

You are here, make the best of it…





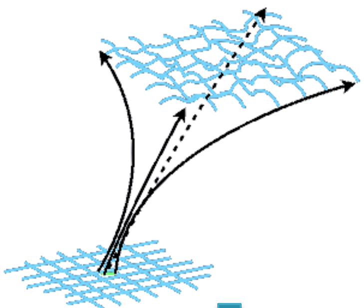
A call to modesty…

"*Hominem te esse*"

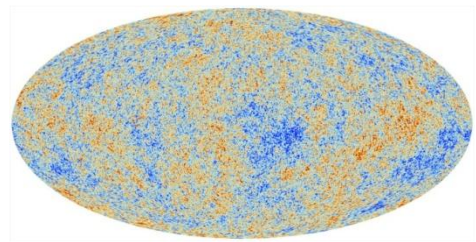
"*Memento mori*"

Inflation as the origin of structure

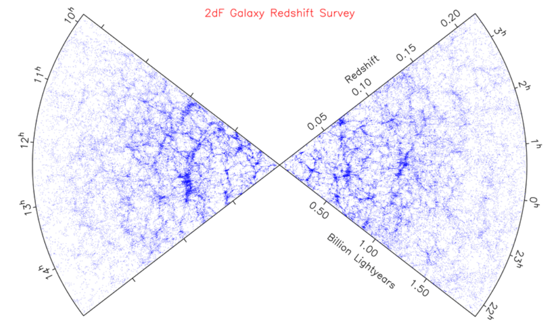
quantum fluctuations



matter and energy perturbations



clusters, halos, galaxies



inflation

gravitational collapse

statistics of these

probe of this era

$\delta\varphi$

inflaton field

Φ

curvature potential

δT CMB

$\boxed{\begin{matrix}\delta\rho \\ \delta n\end{matrix}}$

matter

galaxies

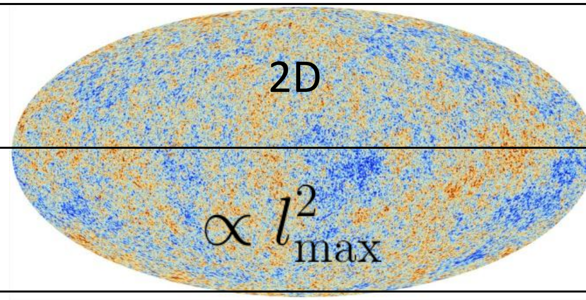
powerful probes of inflation

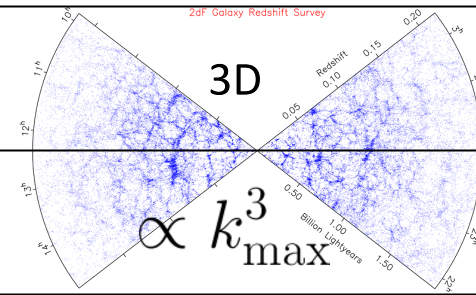
if we can accurately model some relationships

Cosmostatistics of the initial conditions

- *Initial conditions* : ICs for *gravitational evolution*. AFTER the inflationary phase and the Hot Big Bang phenomena (*primordial nucleosynthesis, decoupling and 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

Showdown: CMB versus LSS





	CMB	LSS
Dimension	2D	3D
Number of modes	$\propto l_{\max}^2$	$\propto k_{\max}^3$
Systematics and selection functions	Relatively clean	Relatively messy
Temporal evolution	✗	✓
Color, magnitude, redshift space distortions, bias	✗	✓

Vanilla inflation predicts

(single scalar field, slow-roll regime, canonical kinetic term, Bunch-Davies vacuum, in Einsteinian General Relativity)

- Flat, homogeneous and isotropic Universe
- Nearly scale-invariant scalar perturbations, adiabatic, nearly Gaussian-distributed

$$n_{\mathrm{S}} - 1 = 2\eta - 6\epsilon$$

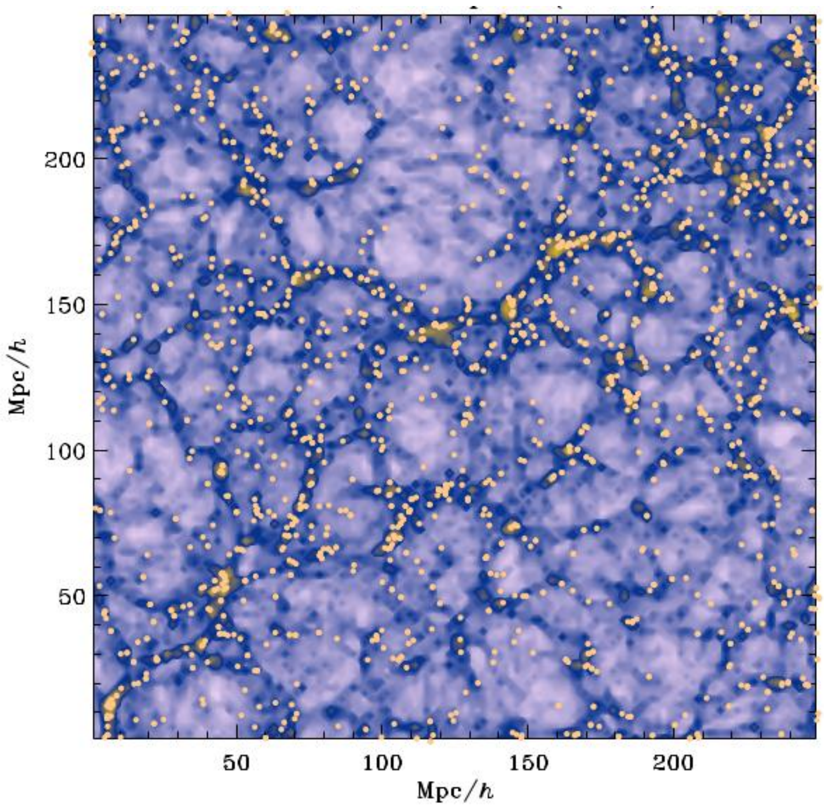
- Negligible amount of tensor perturbations

$$n_{\mathrm{T}} = -2\epsilon$$

Any departures shed direct light on inflation!

- Primordial non-Gaussianity
- Isocurvature modes
- Primordial gravitational waves

Large-scale structure in the Universe



Blue: matter distribution
Orange: dark matter halos / galaxies

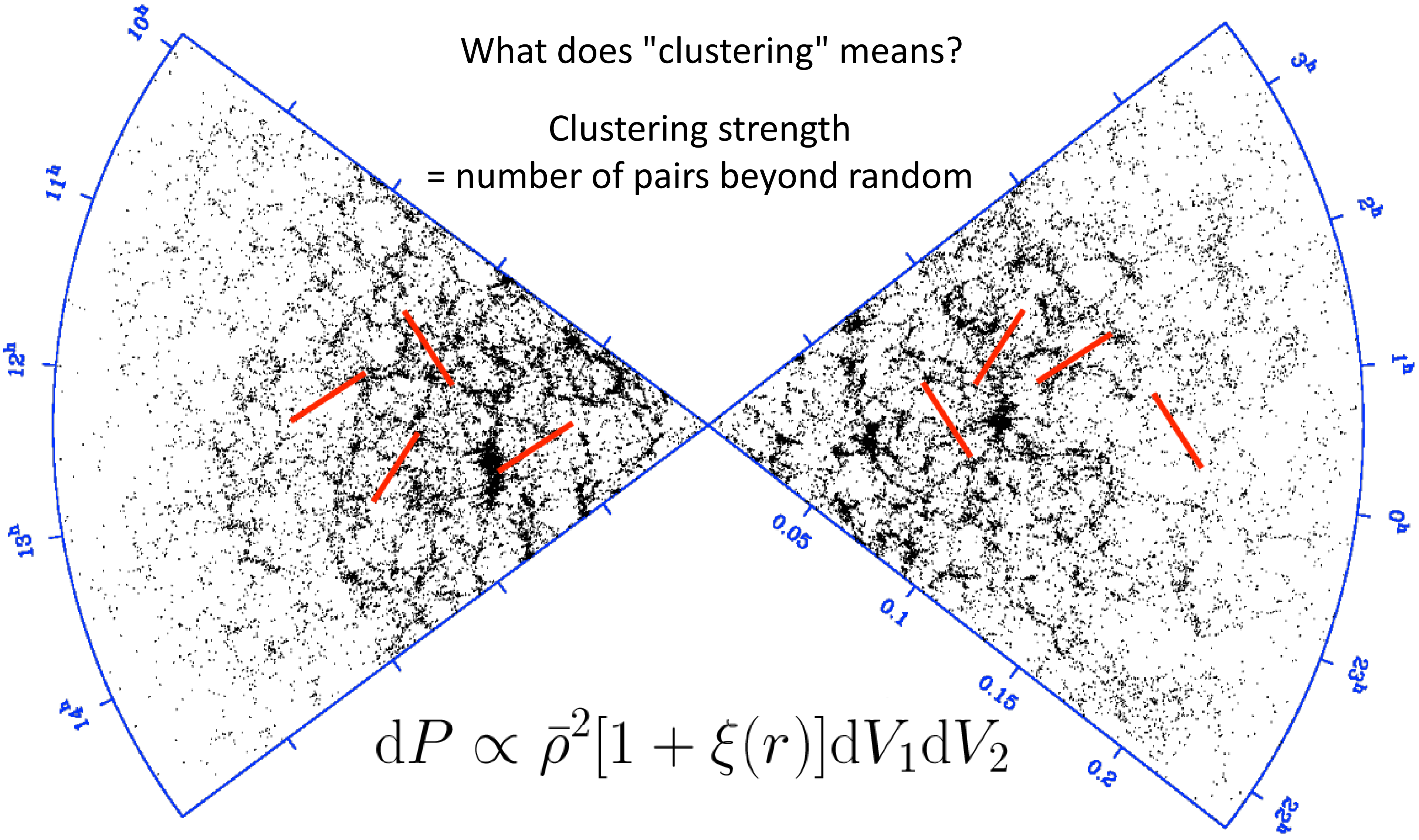
- Halos trace mass distribution (of *dark matter*).
- Halos are NOT randomly distributed: there exists a Large Scale Structure of the Universe
- How do we analyze this structure quantitatively?

Correlation functions and **Fourier analysis**

Two-point information: the amplitude of clustering

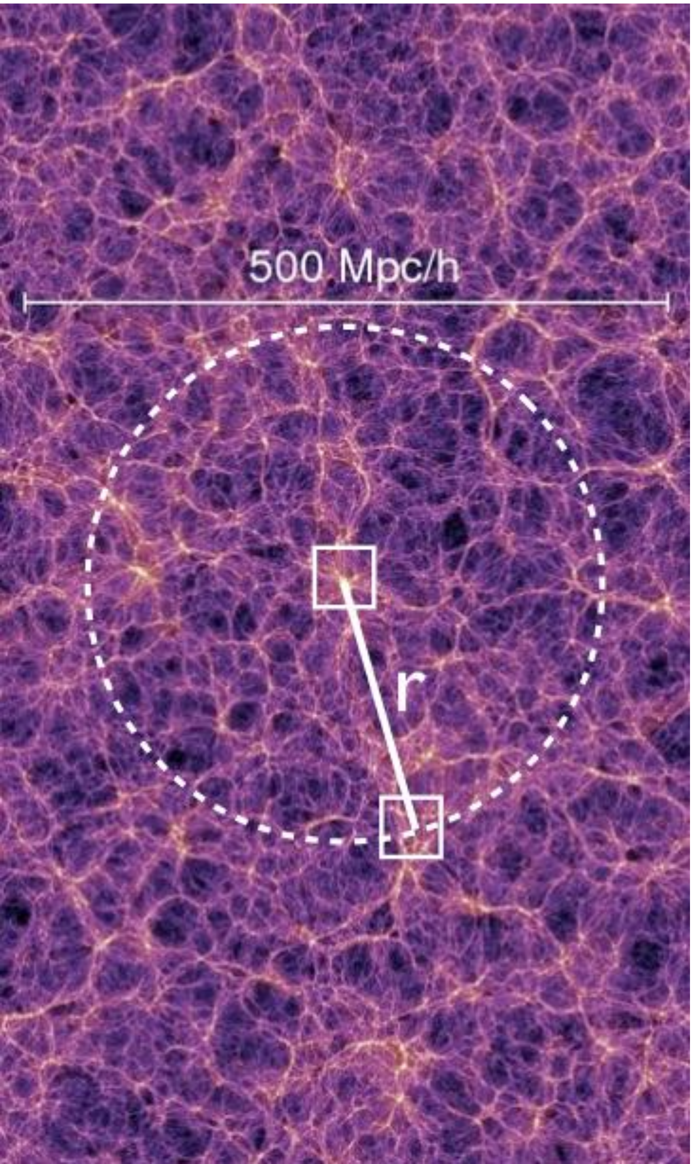
What does "clustering" means?

Clustering strength
= number of pairs beyond random



$$\mathrm{d}P \propto \bar{\rho}^2[1+\xi(r)]\mathrm{d}V_1\mathrm{d}V_2$$

The density contrast field



- The "probability of seeing" structure can be recast in terms of the density contrast:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}$$

- The correlation function is the real-space two-point statistics of the field:

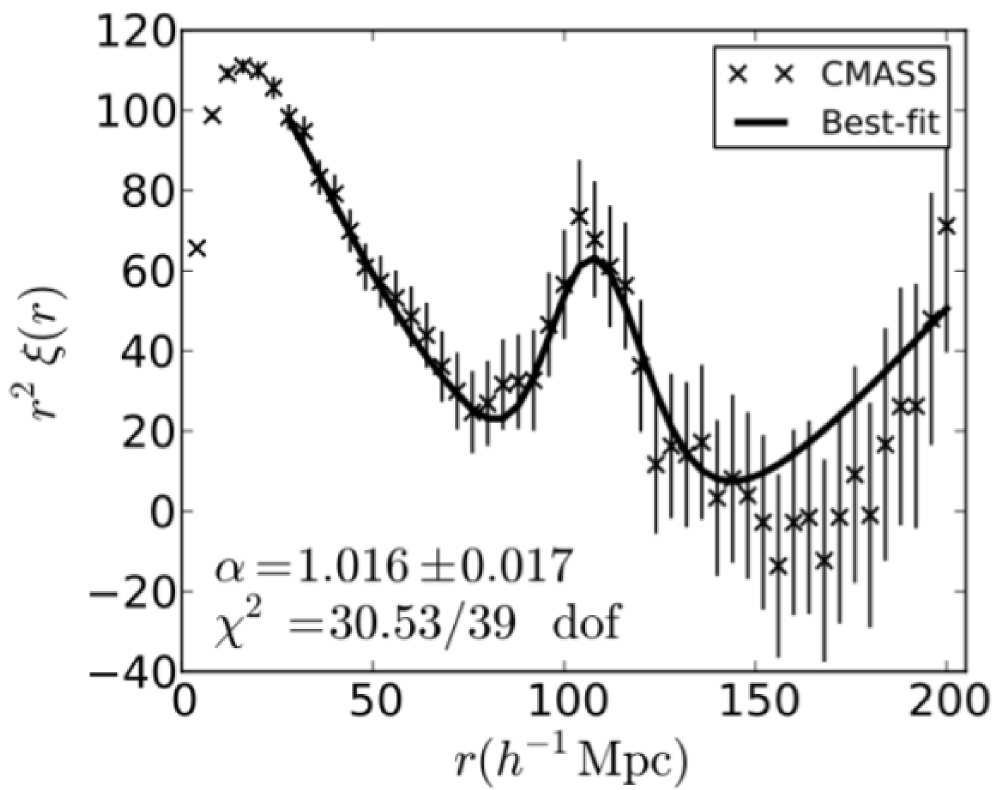
$$\xi(r) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle$$

Real-space correlation function

$$\begin{aligned} \xi(\mathbf{x_1}, \mathbf{x_2}) &= \langle \delta(\mathbf{x_1}) \delta(\mathbf{x_2}) \rangle \\ &= \xi(\mathbf{x_1} - \mathbf{x_2}) \\ &= \xi(|\mathbf{x_1} - \mathbf{x_2}|) \end{aligned}$$

from statistical homogeneity

from statistical isotropy



Power spectrum

- The power spectrum is the Fourier transform of the real-space correlation function

$$P(k) = \langle \delta(\mathbf{k}) \delta(\mathbf{k}) \rangle$$

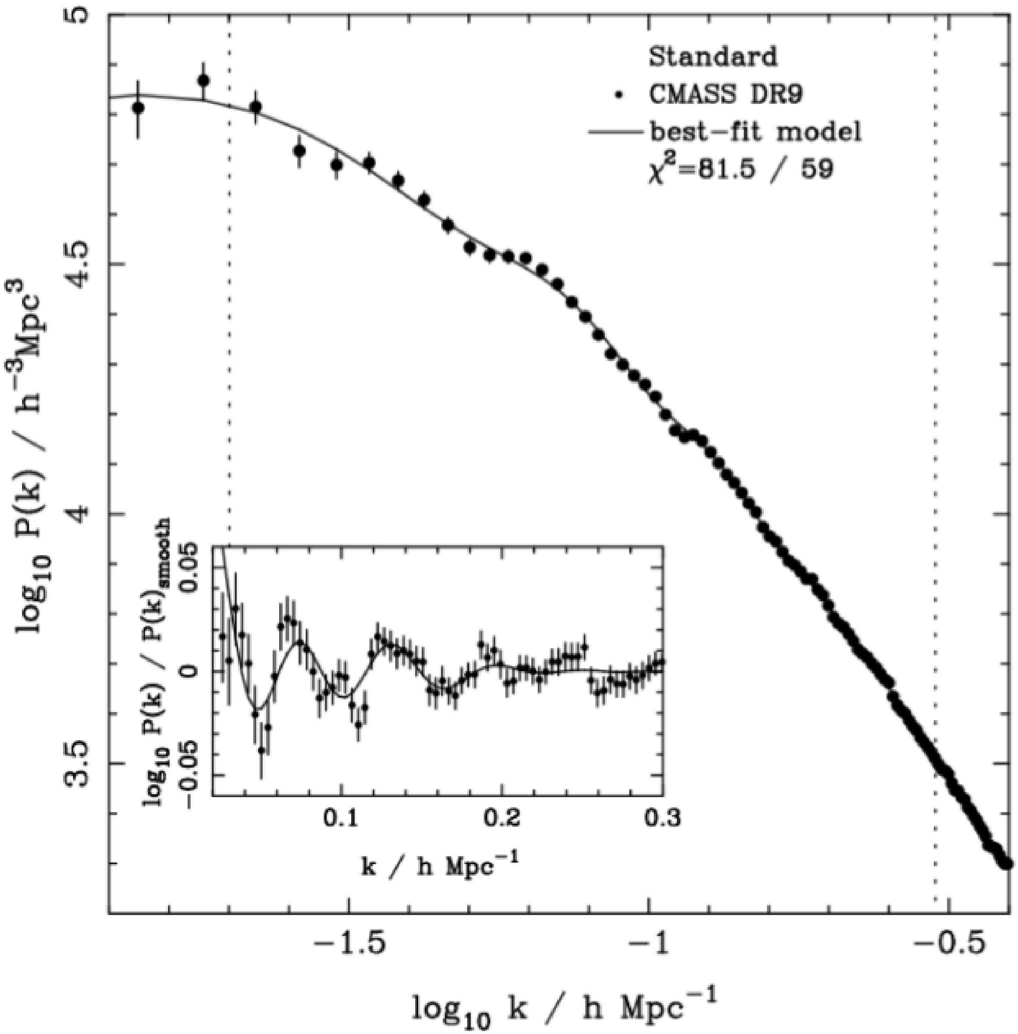
- Gives the clustering strength ("power") at different scales
- By analogy, one can think of "throwing down" Fourier modes rather than "sticks"

Power spectrum

$$\langle\delta(\mathbf{k_1})\delta(\mathbf{k_2})\rangle = (2\pi)^3\delta_D(\mathbf{k_1}-\mathbf{k_2})P(k_1)$$

from statistical isotropy

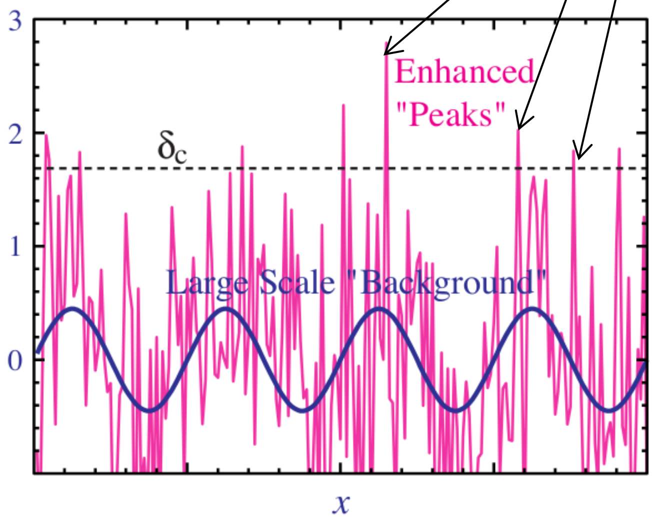
from statistical homogeneity



Caveat I: the bias problem

- Halo/galaxy formation takes place in overdense peaks of the underlying matter distribution

first sites of galaxy formation



Gaussian fluctuations on various scale (described by the power spectrum)

first sites of snowfall

Caveat I: the bias problem

- Galaxy bias = relationship between galaxy and matter over-density fields
  - Local bias

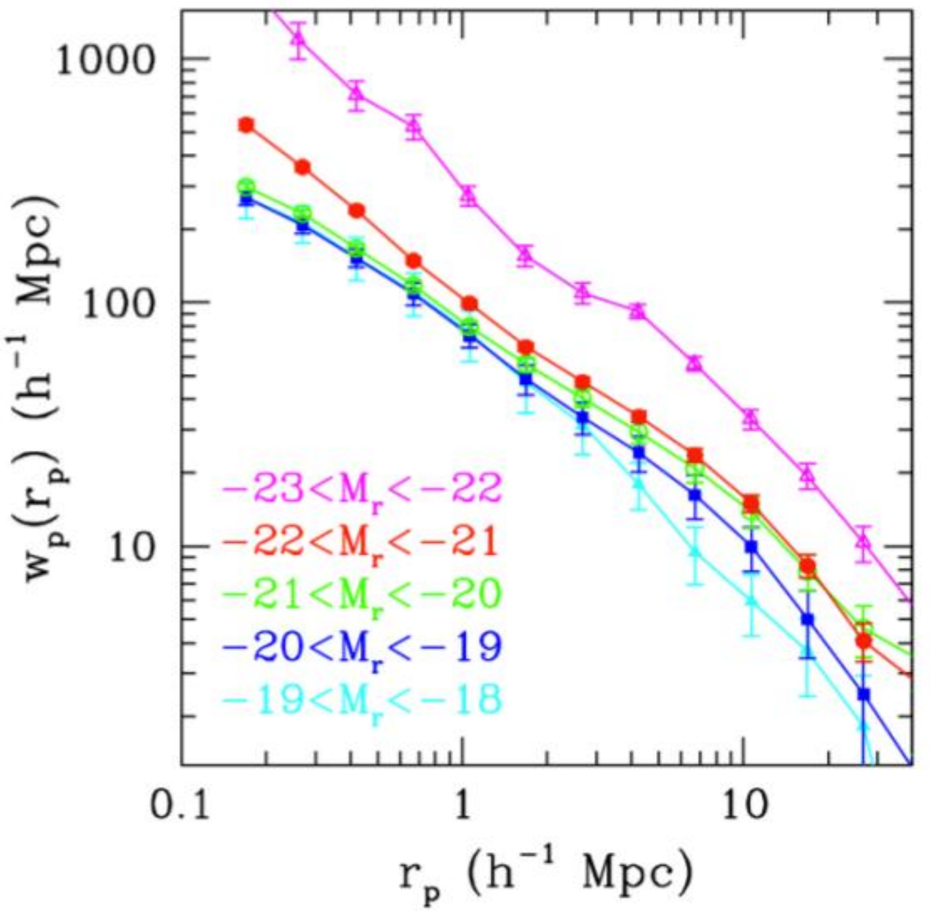
$$\delta_{\rm g}(z,k) = b(z,k)\delta_{\rm m}(z,k)$$

  - Scale-independent local bias

$$\delta_{\rm g}(z,k) = b(z)\delta_{\rm m}(z,k)$$

- Translates to a perfect degeneracy in the power spectrum

$$P_{\rm g}(z,k) = b^2(z)P_{\rm m}(z,k)$$



Zehavi et al. 2010, arXiv:1005.2413

Caveat II: Redshift space distortions

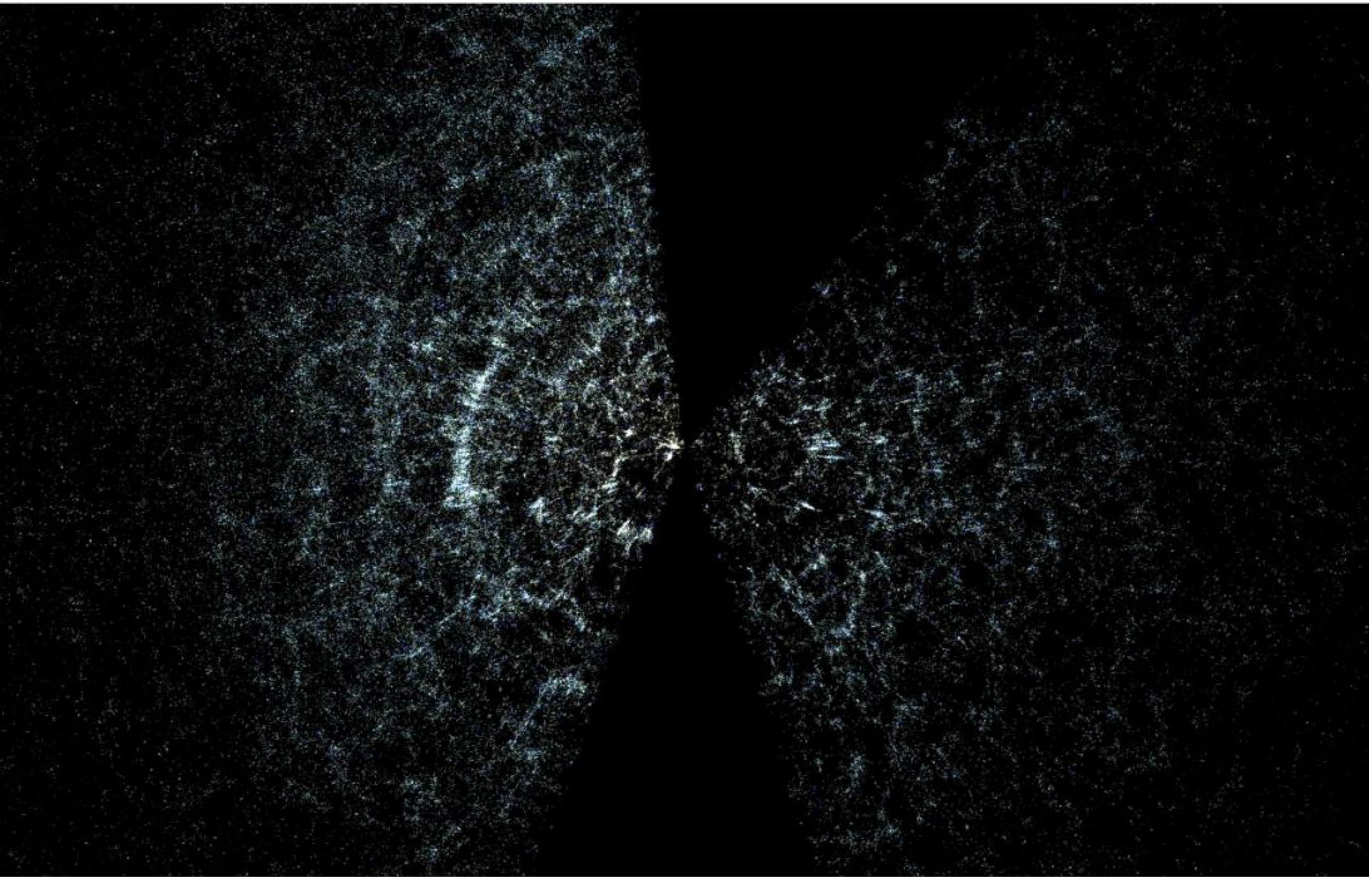


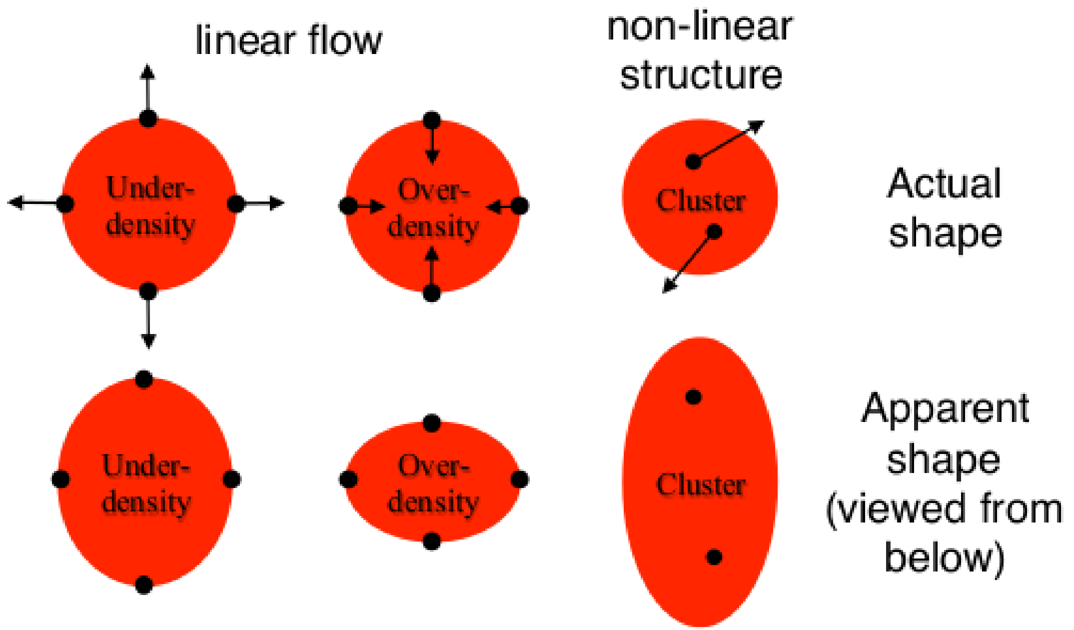
Image of the SDSS, from U. Chicago

Caveat II: Redshift space distortions

- Fingers-of-God: random velocity dispersion in galaxy clusters that deviates a galaxy's velocity from pure Hubble flow ⇨ streches out clusters in redshift space
- Kaiser effect: peculiar velocity of galaxies bound to a central mass that undergoes infall (peculiar velocities are coherent toward the central mass, not random)

Caveat II: Redshift space distortions

- Statistical comparison of the "apparent" structure across and along the line-of-sight
- Linear growth of structure enhances clustering signal, but only along line-of-sight



Caveat II: Redshift space distortions

- Redshift space distortions theory: $\mathbf{s} = \mathbf{r} + v\,\hat{\mathbf{r}}_{\rm los}$

$$\mu = \cos\alpha$$

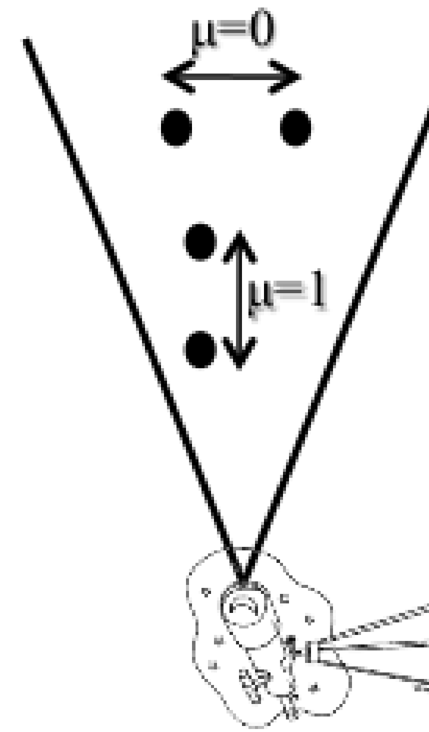
$$\theta = \nabla \cdot \mathbf{v}$$

$$\delta_{\rm g}^{\rm s} = \delta_{\rm g}^{\rm r} - \mu^2\theta = \delta_{\rm g}^{\rm r}(1 + f\mu^2)$$

$$\delta_{\rm g} = b\delta, \quad \theta = -f\delta, \quad f \equiv \frac{{\rm d}\ln D}{{\rm d}\ln a}$$

$$P_{\rm g}^{\rm s}(\mu) = P_{\rm gg} + 2\mu^2 P_{{\rm g}\theta} + \mu^4 P_{\theta\theta}$$

$$\Rightarrow \quad P_{\rm g}^{\rm s}(k,\mu) = [b + f\mu^2]^2\, P_{\rm m}(k)$$



Kaiser 1987, MNRAS, 227, 1

Cosmology with the linear galaxy power spectrum

- Linear growth of structure:

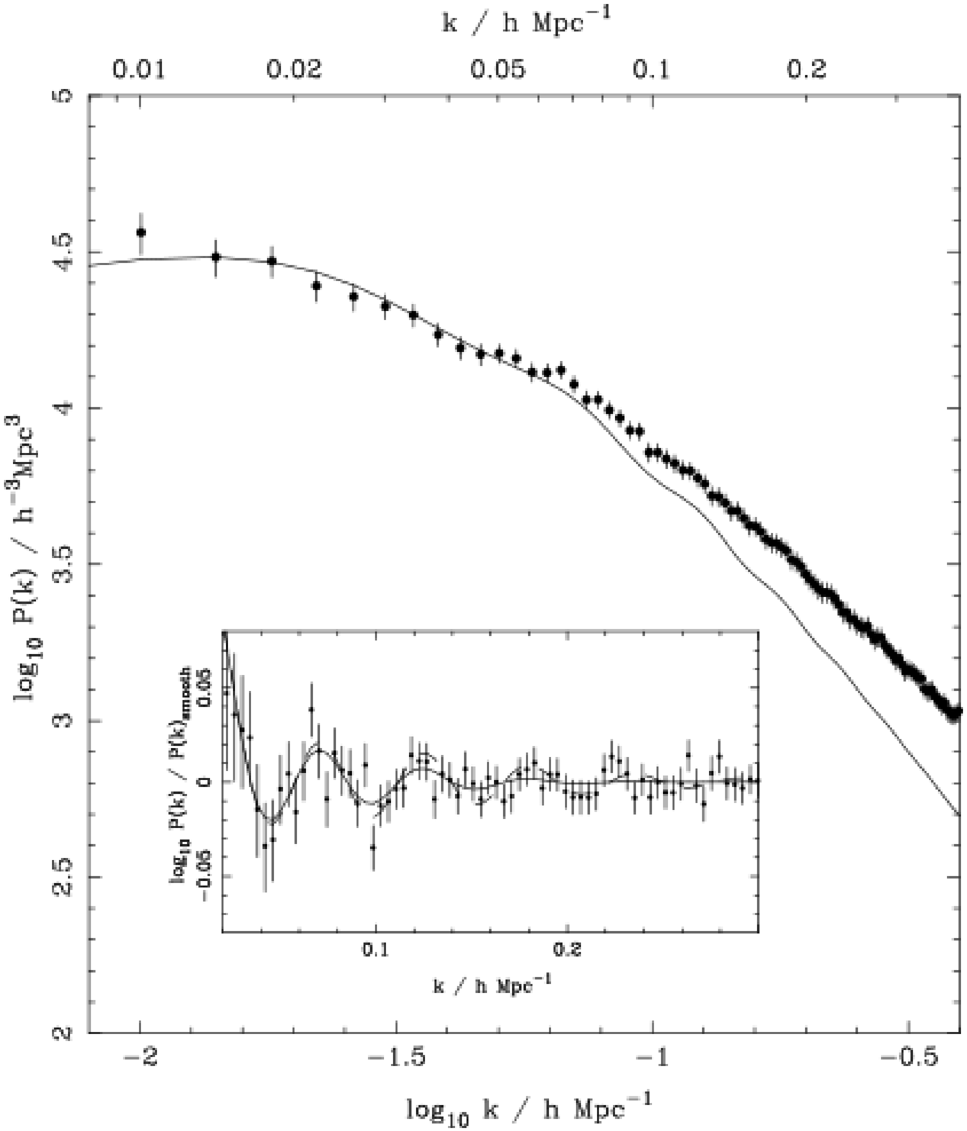
$$\delta(a) = D(a)\delta_i$$
$$P_{\rm m}(k,a) = D(a)^2 P_i(k)$$

- Putting it all together:

$$P_{\rm g}^{\rm s}(k,\mu,a) = D(a)^2[b(a) + f(a)\mu^2]^2\, P_i(k)$$

k = comoving wavenumber
μ = cos(angle to line-of-sight)
a = cosmological scale factor
b = galaxy bias factor
D = linear growth rate
f = dlnD/dlna

Cosmology with the linear galaxy power spectrum



$$P_{\rm g}^{\rm s}(k,\mu,a) = D(a)^2[b(a)+f(a)\mu^2]^2\,P_i(k)$$

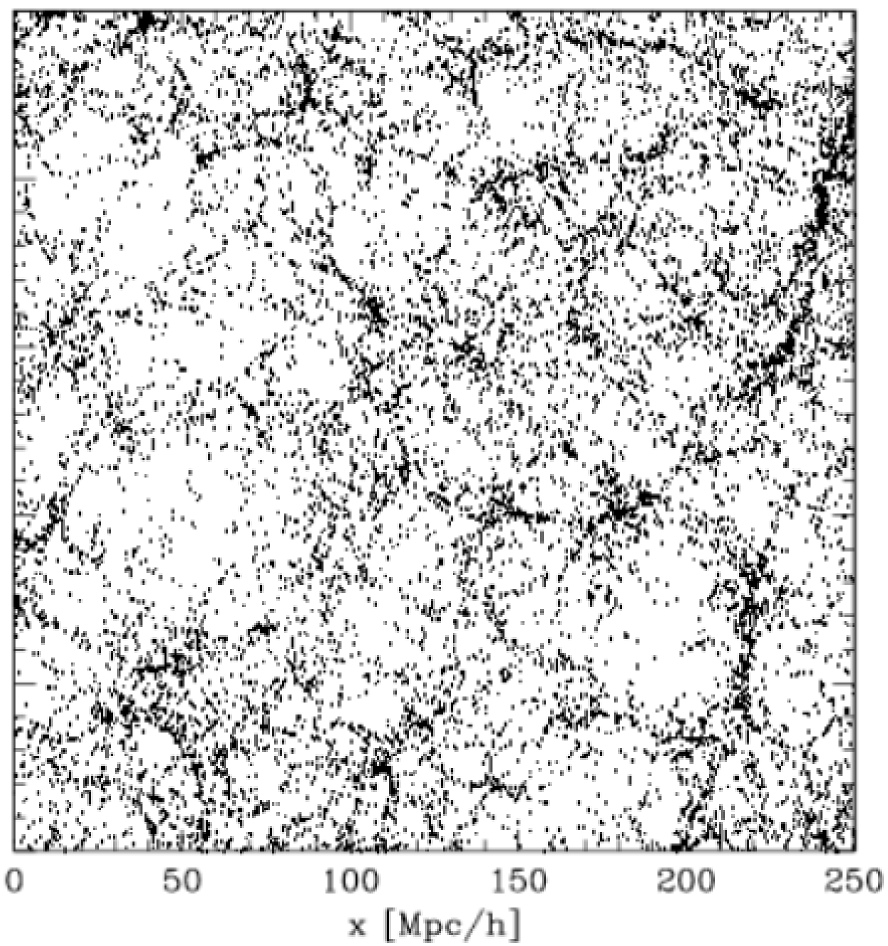
If the LSS was a Gaussian Random Field, we'd be done!

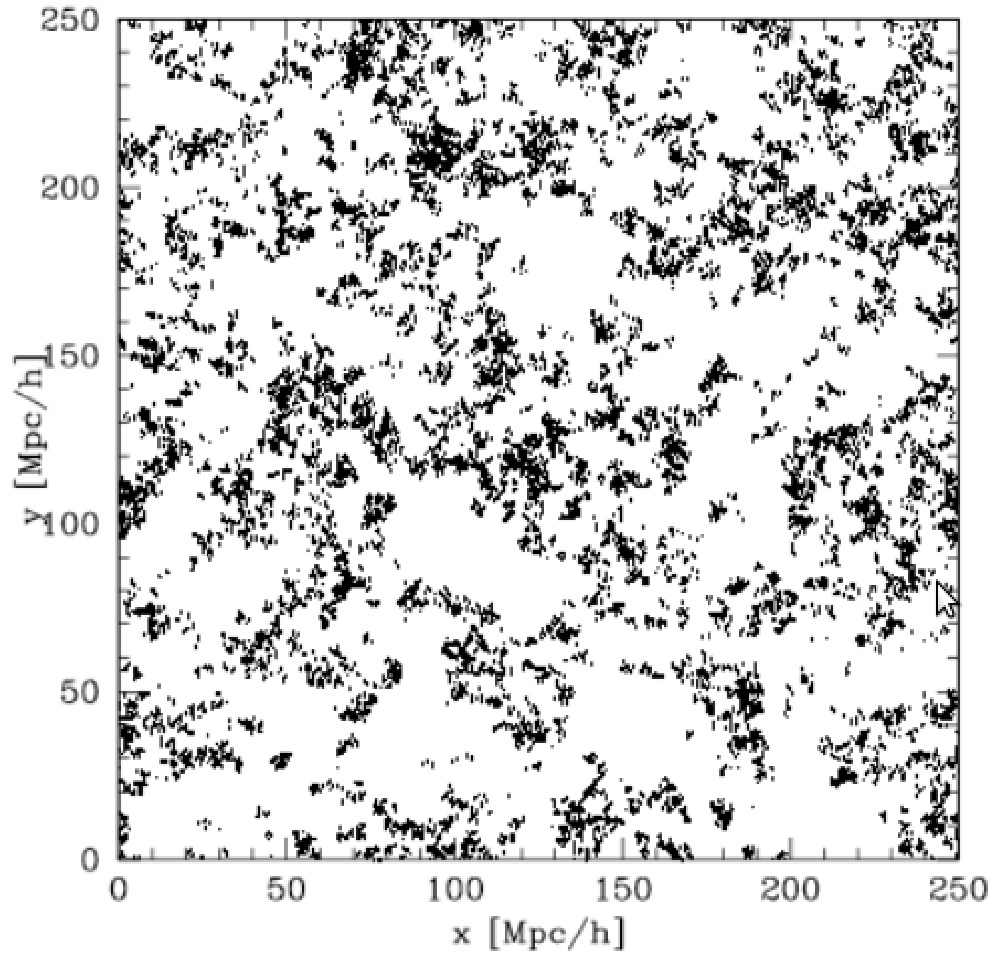
… but it's not:

- Non-linear gravitational evolution
- Primordial non-Gaussianity (?)

Percival et al 2006, arXiv:astro-ph/0608636

Gaussian vs Non-Gaussian information

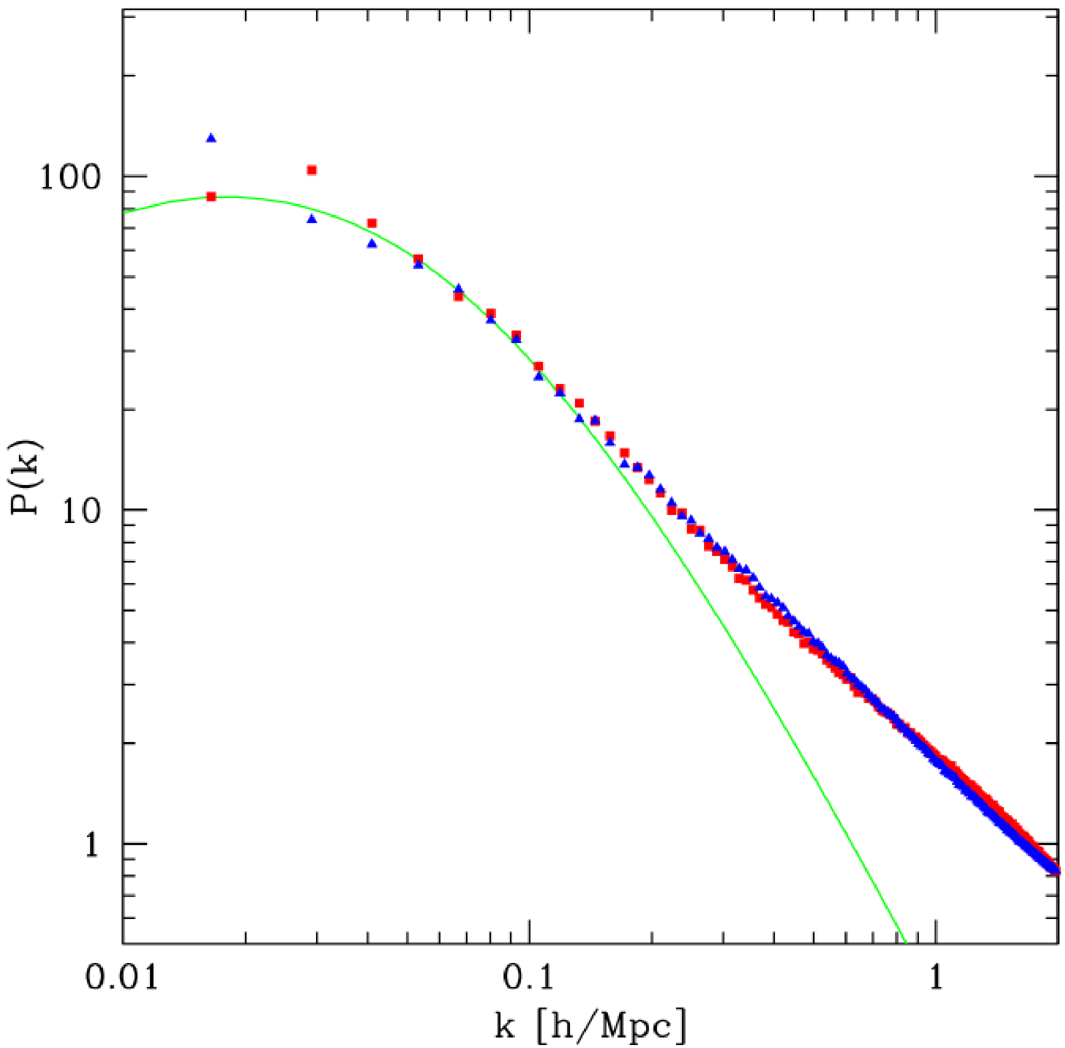




from R. Scoccimarro

Gaussian vs Non-Gaussian information

The two distribution have about the same power spectrum!

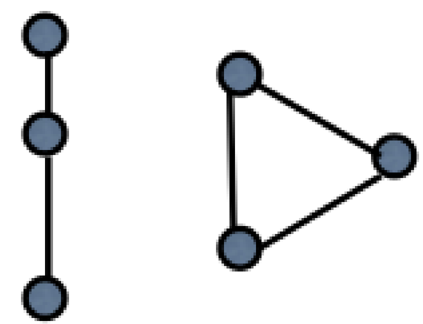


from R. Scoccimarro

The three-point function, the bispectrum

- Three-point statistics are the lowest order measure of the shape of structures (filaments, walls, halos) generated by gravitational instability.
- Indeed, with two points one can only form a single shape: a line. Three-points form a triangle, so we got different triangle shapes.

  e.g. if filamentary structures are predominant, then the bispectrum should be larger for collinear triangles than equilateral.



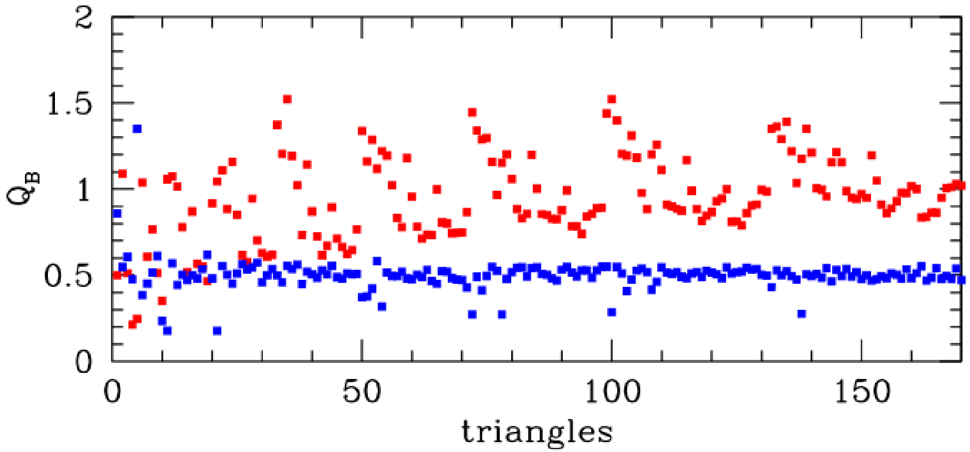
- A limitation of three-point statistics is that three points are always in a plane. In order to better probe the “three-dimensional” shape of structures, one needs to go to the four-point function.

Gaussian vs Non-Gaussian information

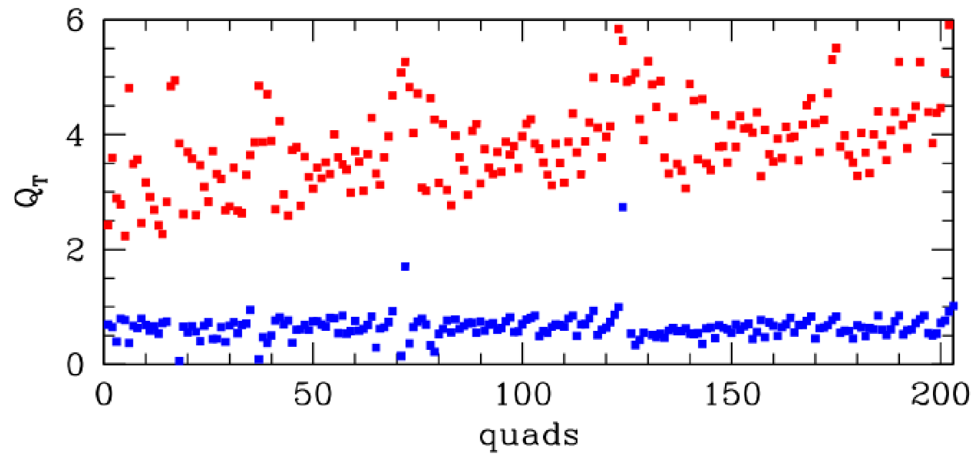
The two distributions can be distinguished easily

by higher-order correlations!

bispectrum



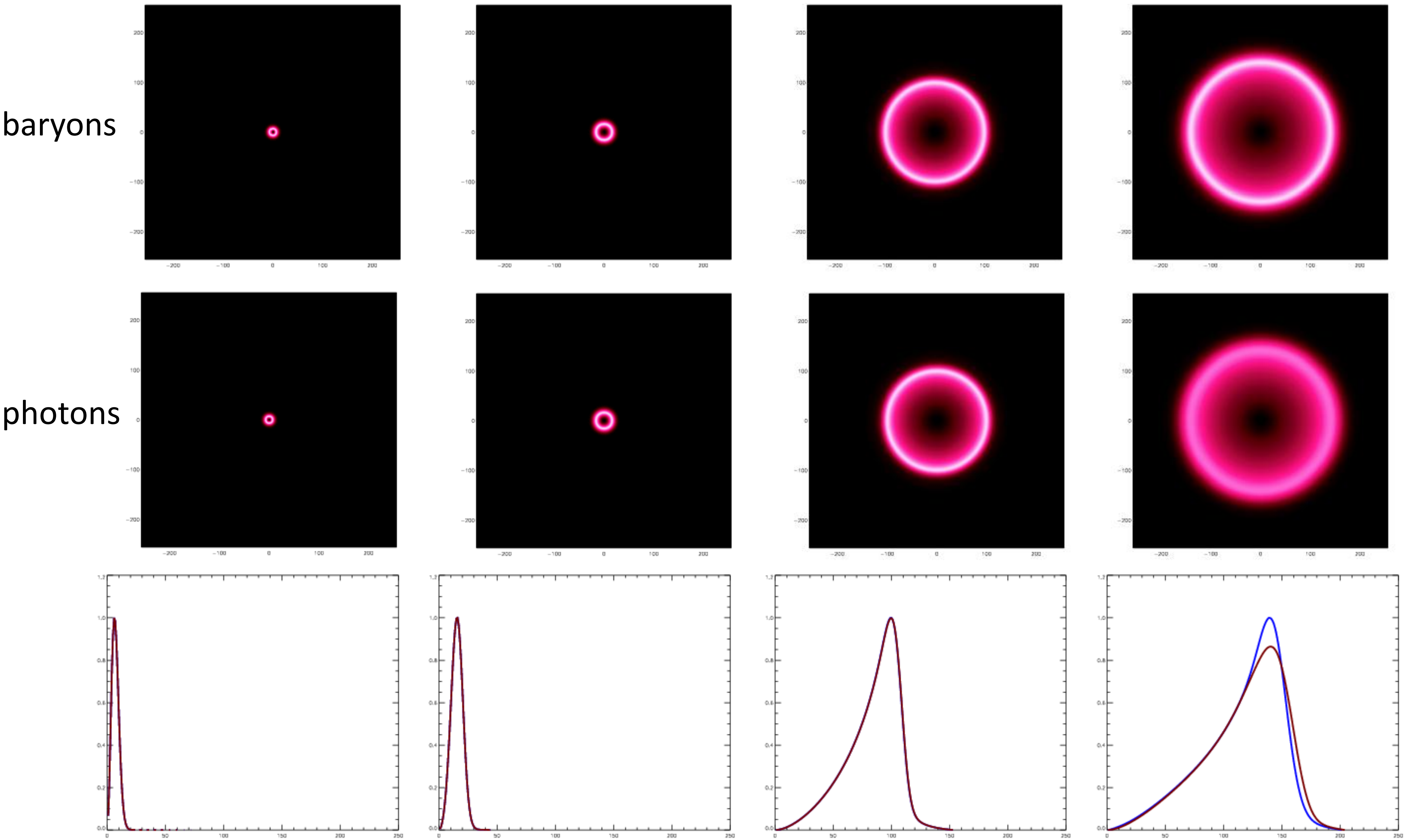
trispectrum



from R. Scoccimarro

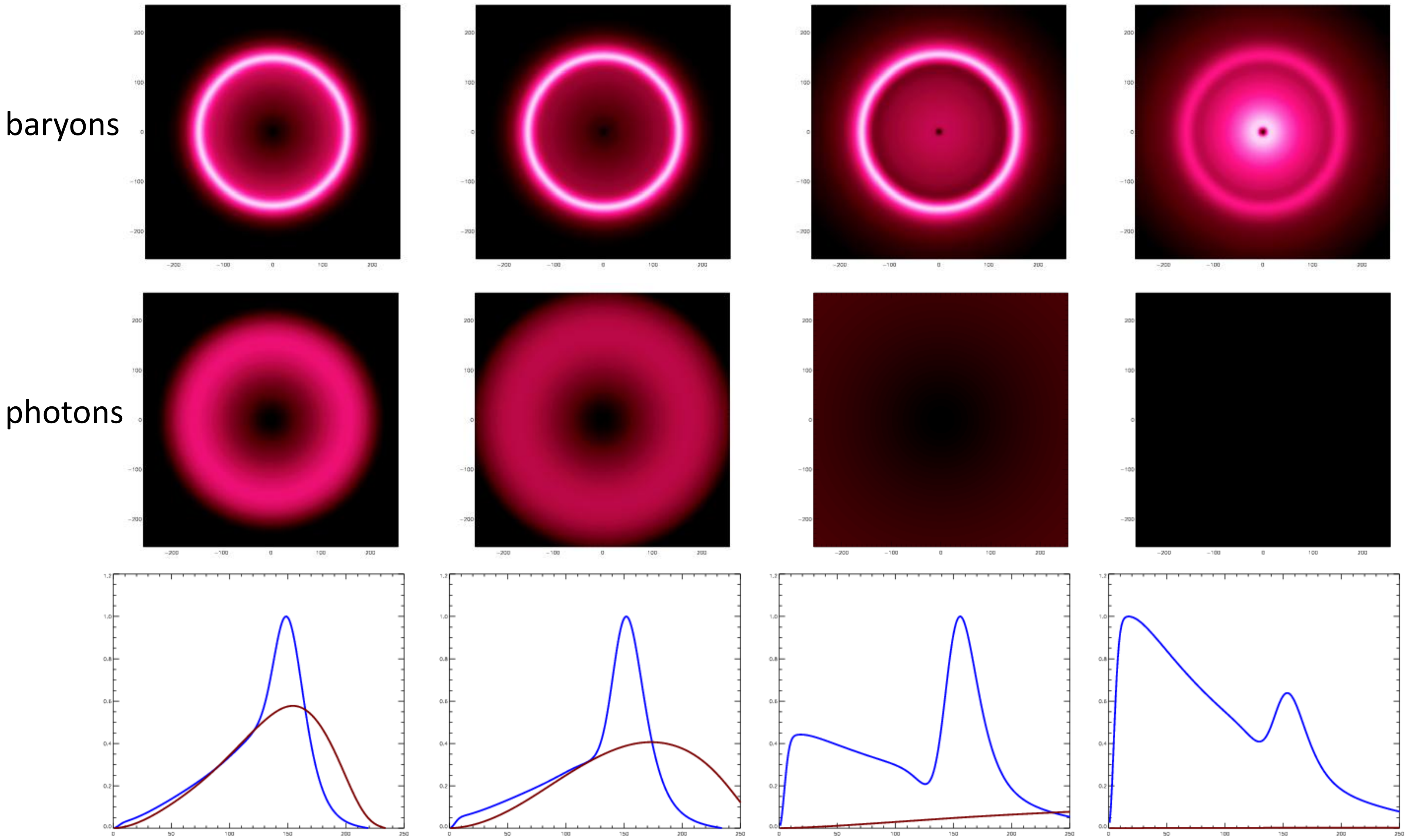
Baryon Acoustic Oscillations (BAO)

images from M. White



Baryon Acoustic Oscillations (BAO)

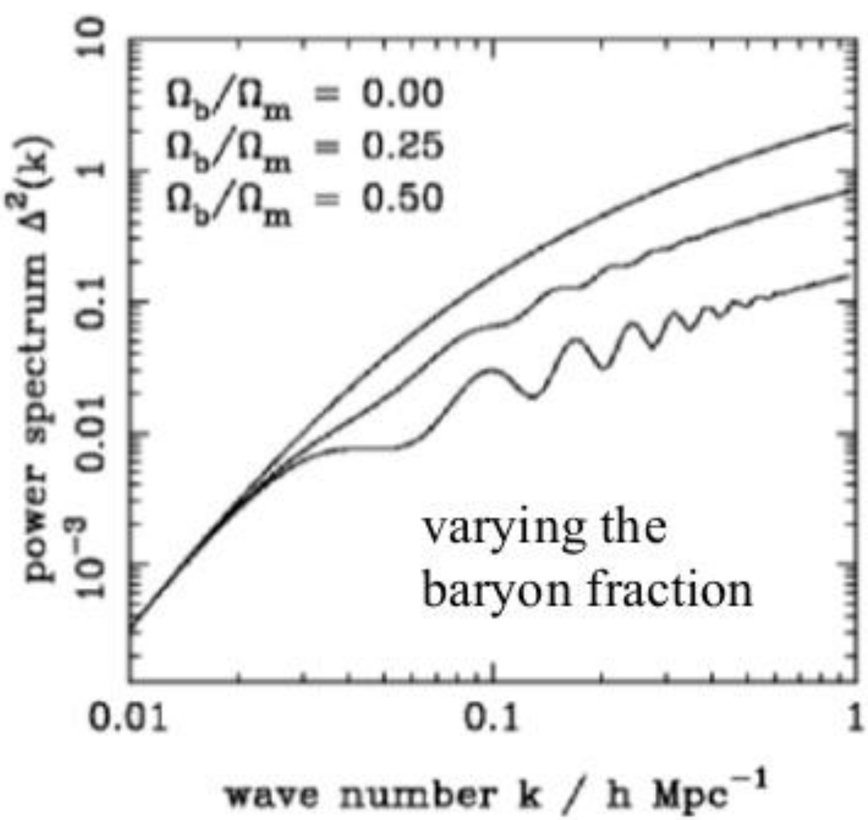
images from M. White



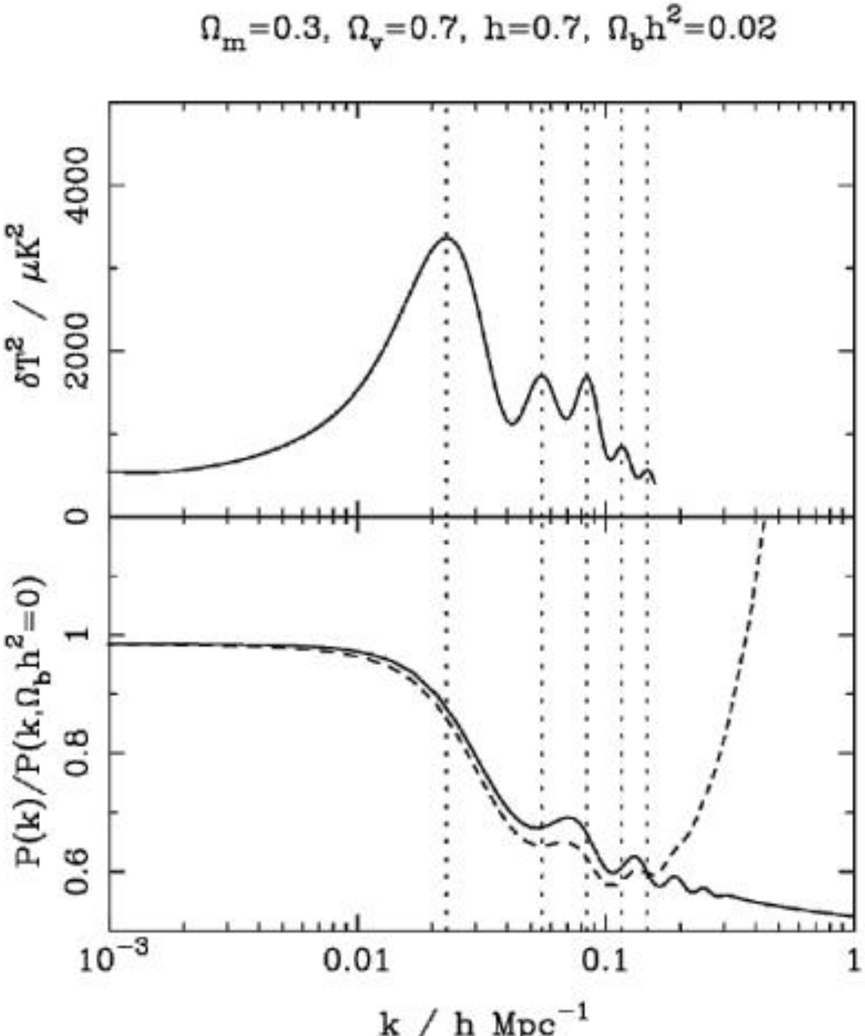
BAO: Wiggles in the power spectrum

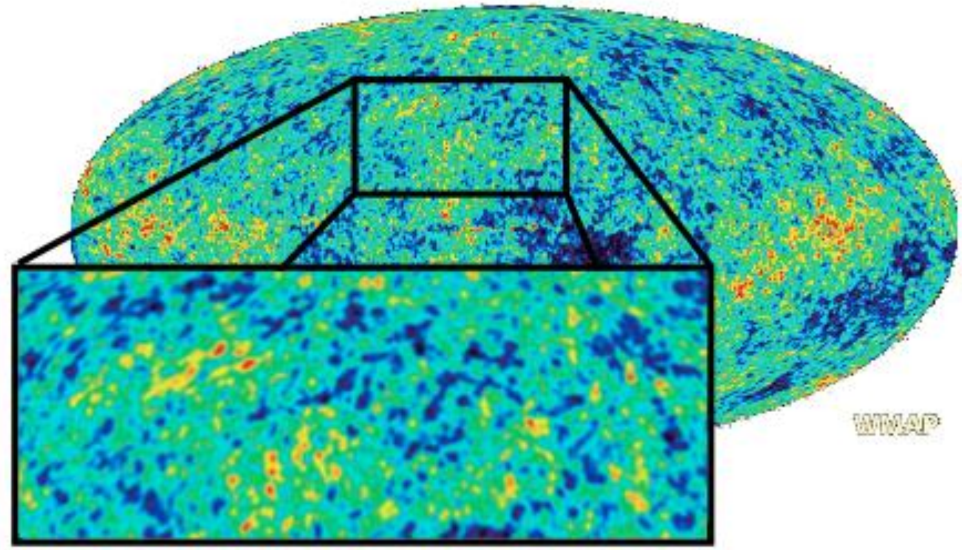
- comoving sound horizon: ~115 Mpc/h
- BAO wavelength: 0.06 h/Mpc

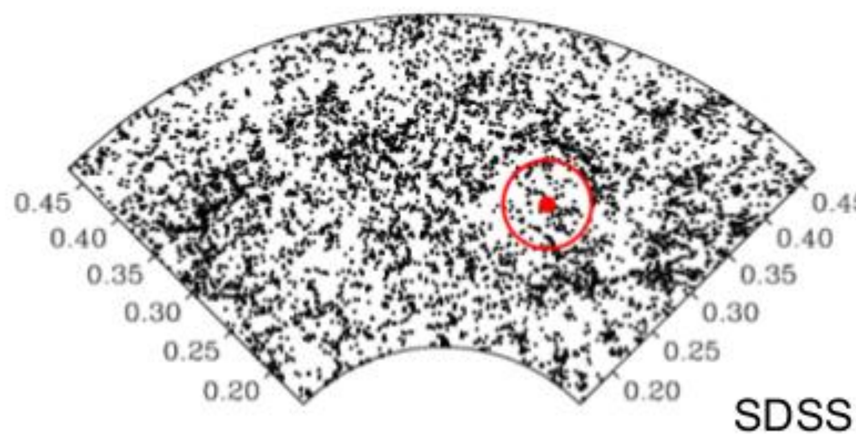
Eisenstein et al 2005, arXiv:astro-ph/0501171



BAO: relationship between BAO and LSS power spectra



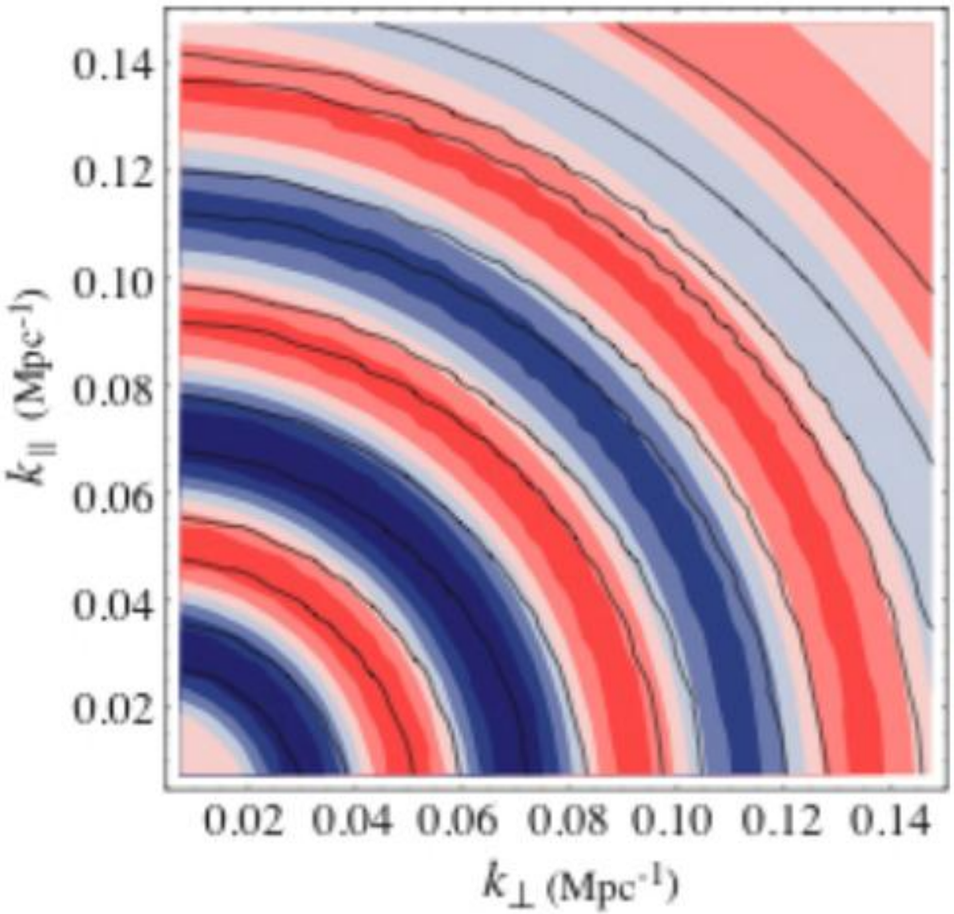




from W. Percival

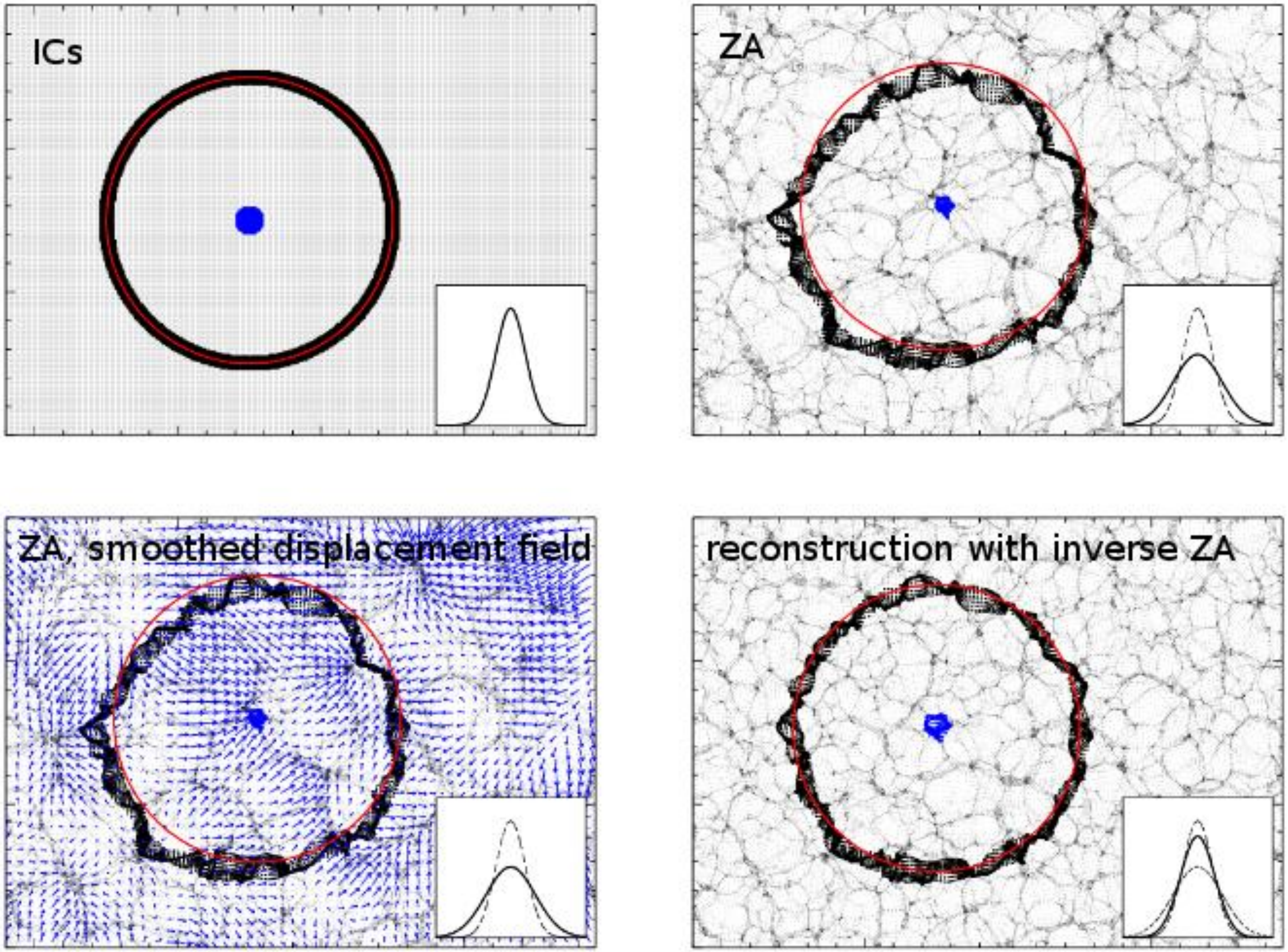
BAO as a standard ruler

- Changes in cosmological model alter differently the BAO scale in radial and angular direction
- This difference is known as the Alcock-Paczynski effect



Hu & Haiman 2003, arXiv:astro-ph/0306053

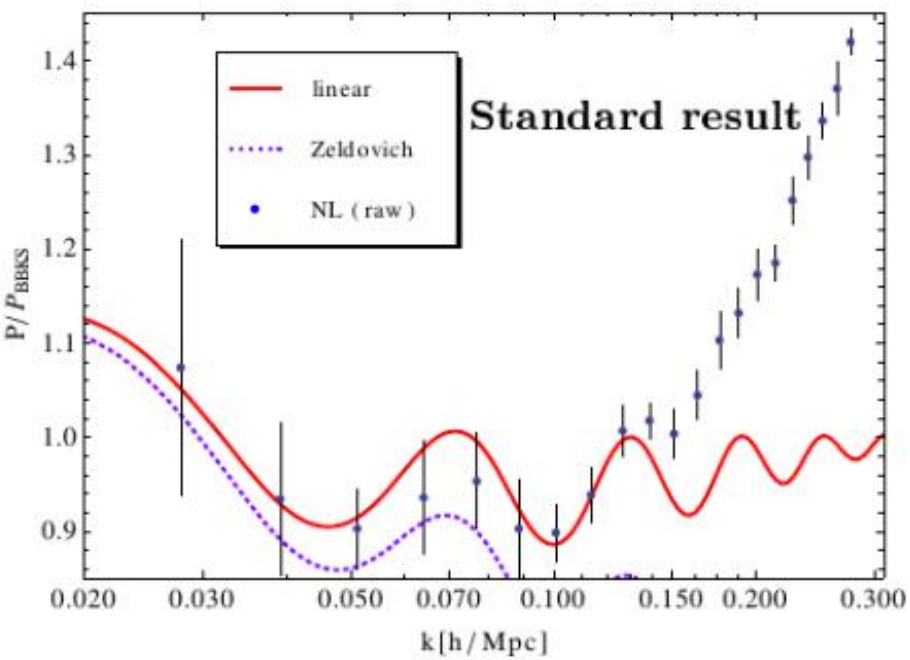
BAO peak reconstruction with Lagrangian perturbation theory

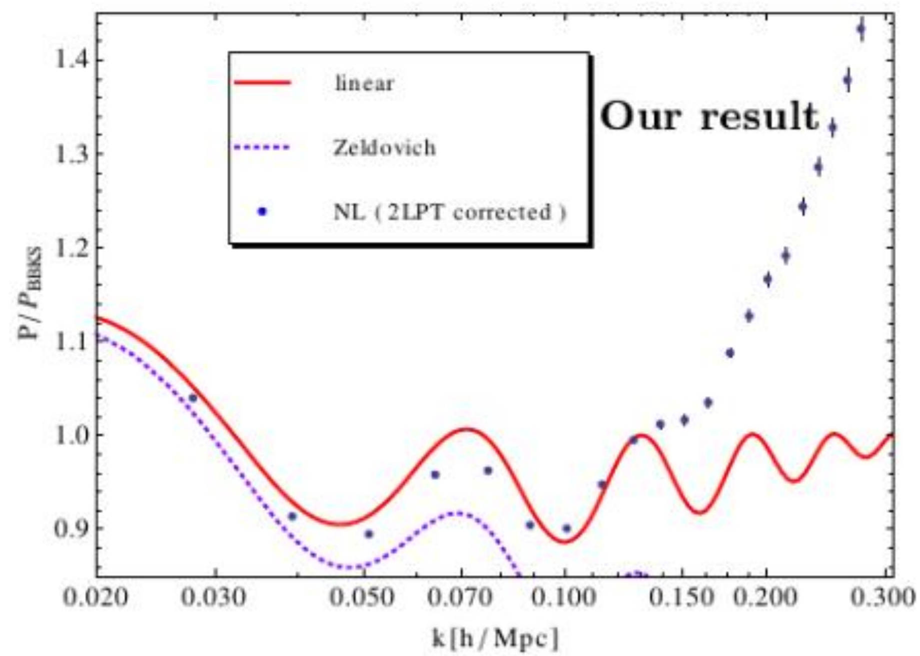


Padmanabhan et al 2012, arXiv:1202.0090

BAO peak reconstruction: Lagrangian space + mode-coupling residual

$$\delta_{NL}(\boldsymbol{k},\eta) \approx \tilde{R}(k,\eta)\delta_Z(\boldsymbol{k},\eta) + \delta_{MC}(\boldsymbol{k},\eta)$$

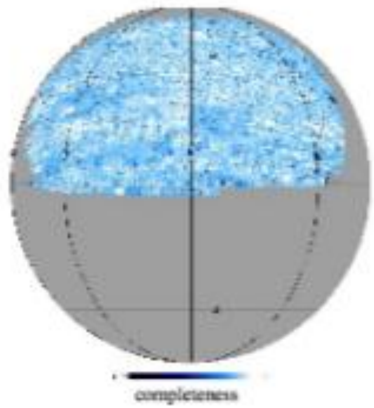




Tassev & Zaldarriaga, arXiv:1203.5785, 1203.6066

Bayesian inference of the ICs

- Why do we need Bayesian inference?
  Inference of sinals = ill-posed problem
  - Noise
  - Incomplete observations: survey geometry, selection effects, biases, cosmic variance
  - Systematic uncertainties



⇨ No unique recovery is possible!

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

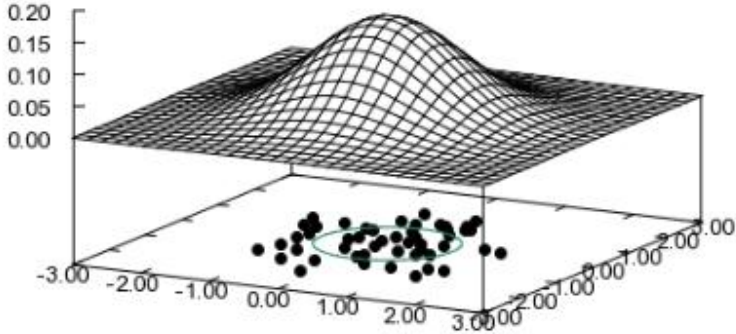
from J. Jasche

Bayesian inference of the ICs

- Problems:
  - High dimensional (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 of the posterior

$$\mathcal{P}(s|d) \rightarrow \mathcal{P}_N(s|d) = \frac{1}{N}\sum_{i=1}^{N}\delta^D(s - s_i)$$



- But how to "get the dots" ?

from J. Jasche

Markov Chain Monte Carlo method

- MCMC method maps the likelihood surface by building a chain of parameter values whose density at any location is proportional to the likelihood at that location p(x)

given a chain at parameter x, and a candidate for the next step x', then x' is accepted with probability

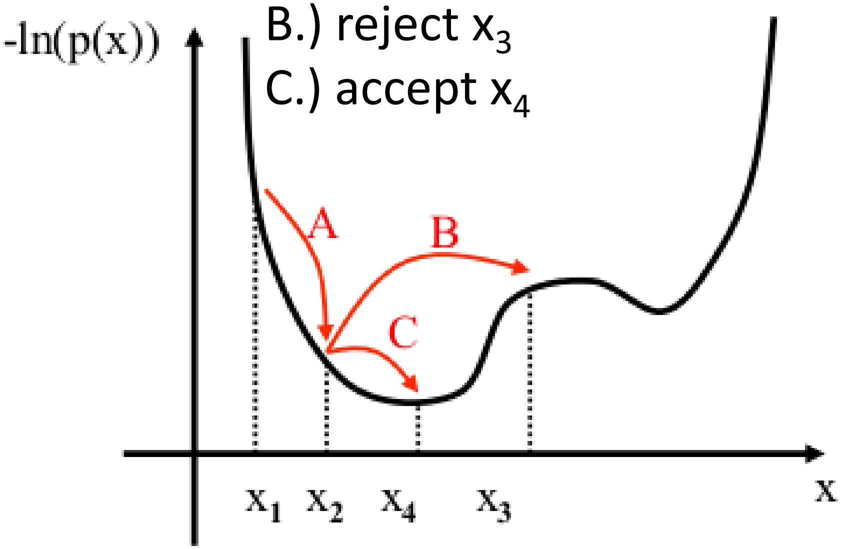
- 1 if p(x') > p(x)
- p(x')/p(x) otherwise

an example chain starting at x_1:

A.) accept x_2

B.) reject x_3

C.) accept x_4



- Hamiltonian sampling: choose your steps such as the acceptance rate is always 1!

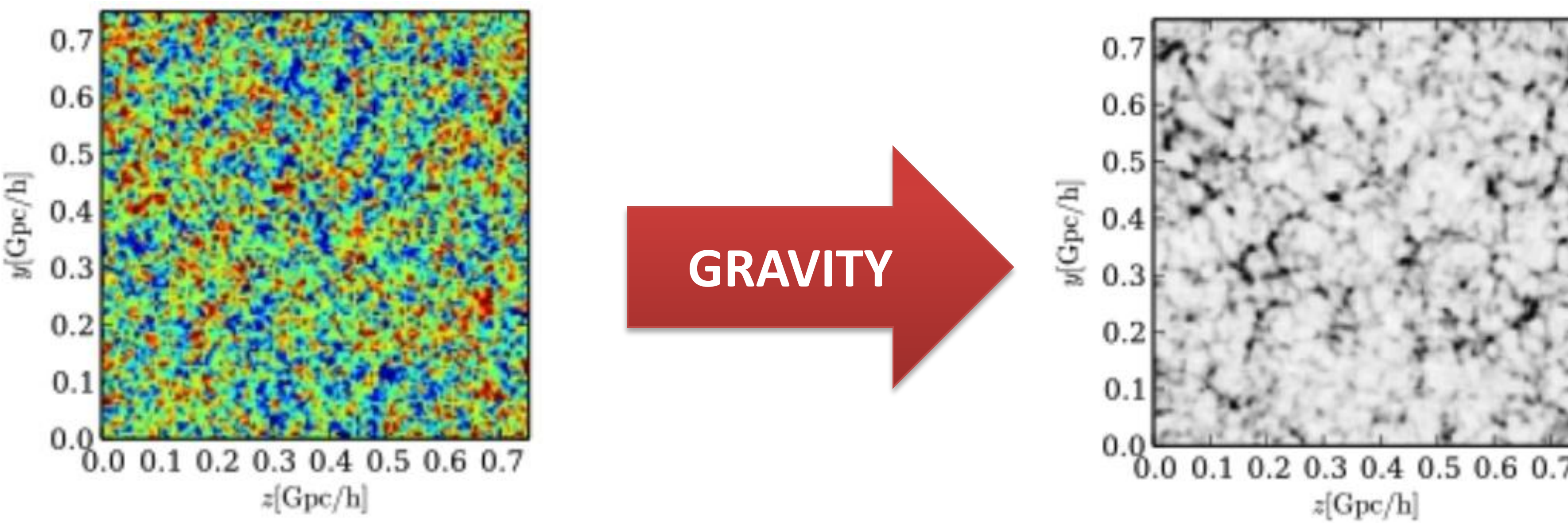
Chain: x_1, x_2, x_2, x_4, …

4D physical inference of the ICs

- Physical motivation:
  - Complex final state
  - Simple initial state
- A "direct only" problem

Initial state

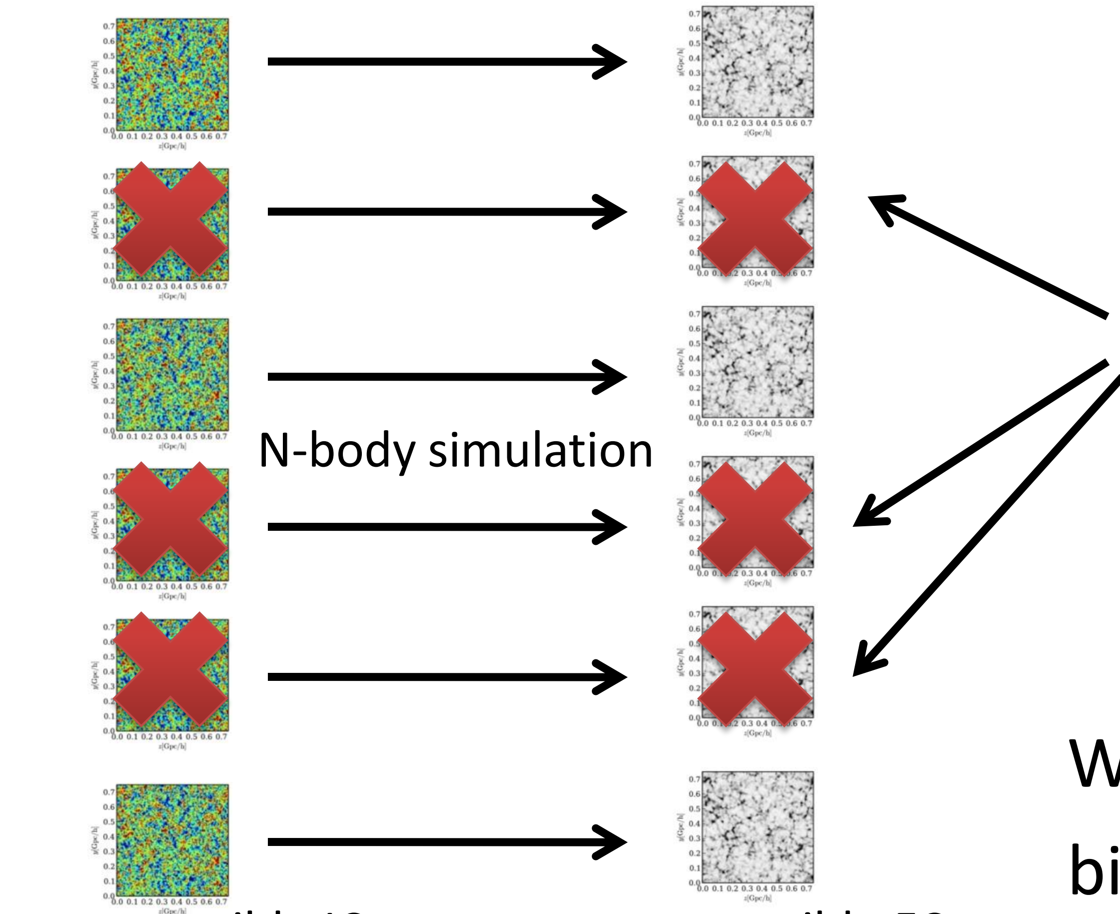
Final state

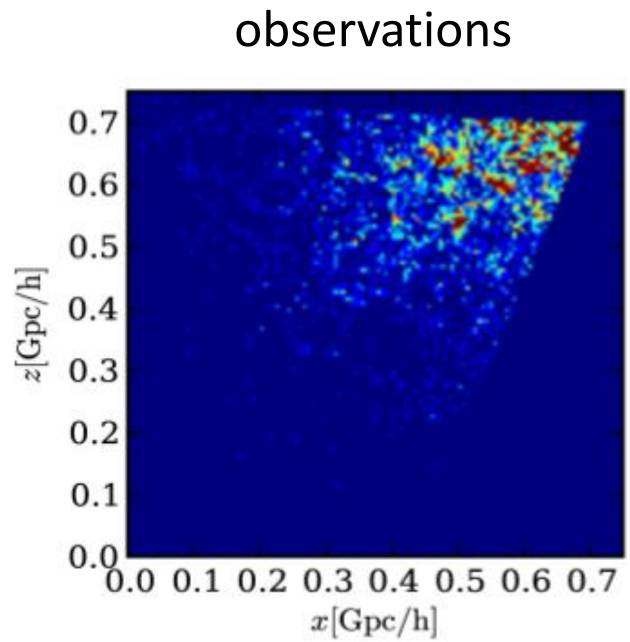


from J. Jasche

4D physical inference of the ICs

- The ideal scenario:



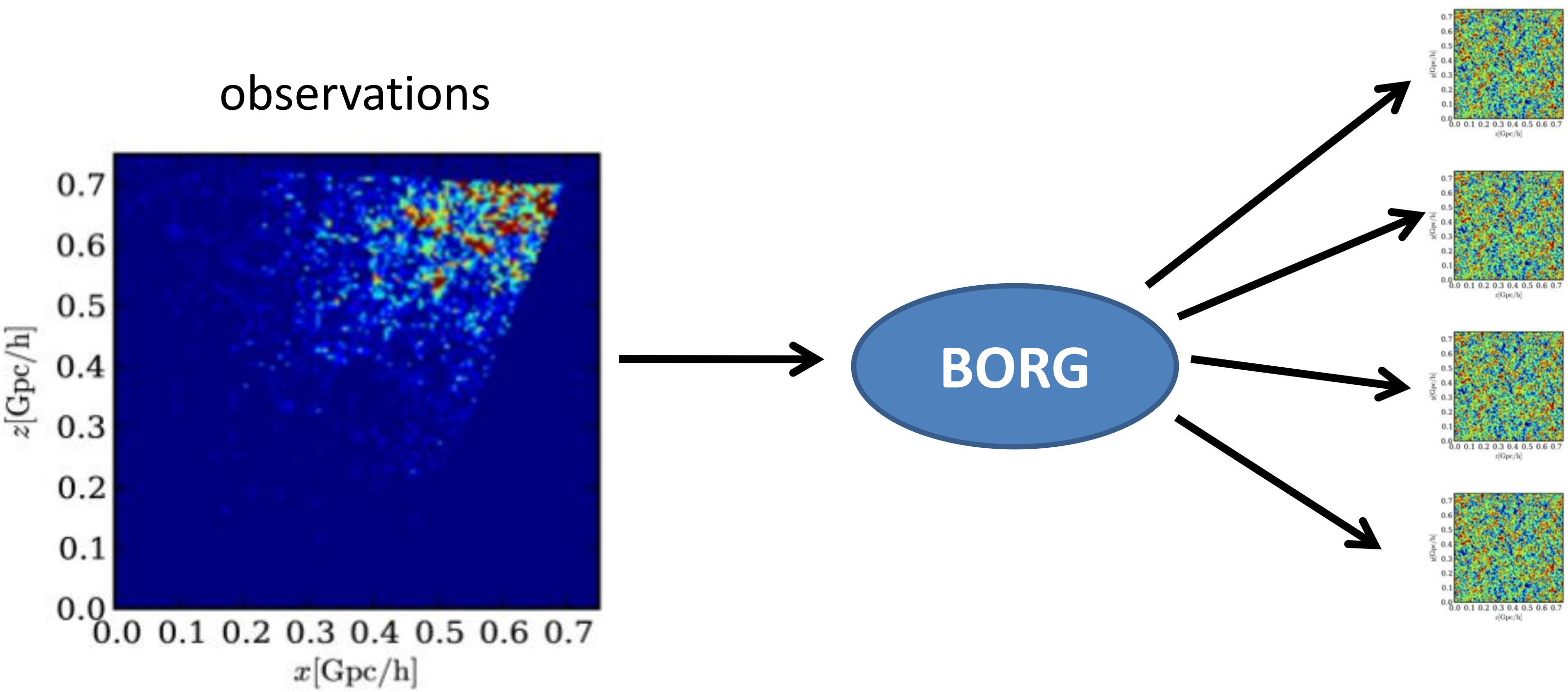


We need a very, very, very big computer!

from J. Jasche

BORG: Bayesian Origin Reconstruction from Galaxies

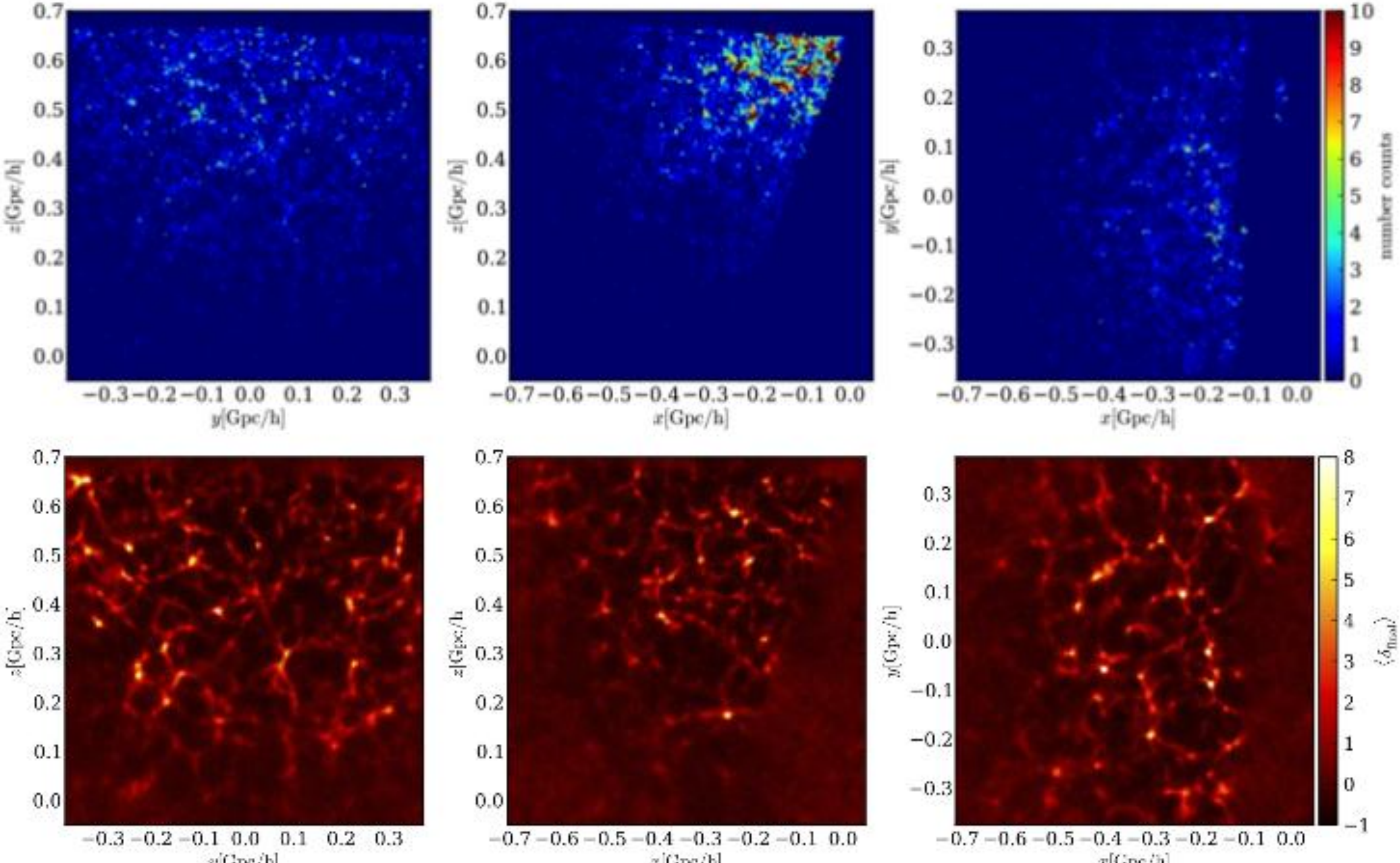
- MCMC with Hamiltonian sampling
- Second-order Lagrangian perturbation theory



Jasche & Wandelt, arXiv:1203.3639 from J. Jasche

BORG at work (movie)

BORG: data vs reconstruction

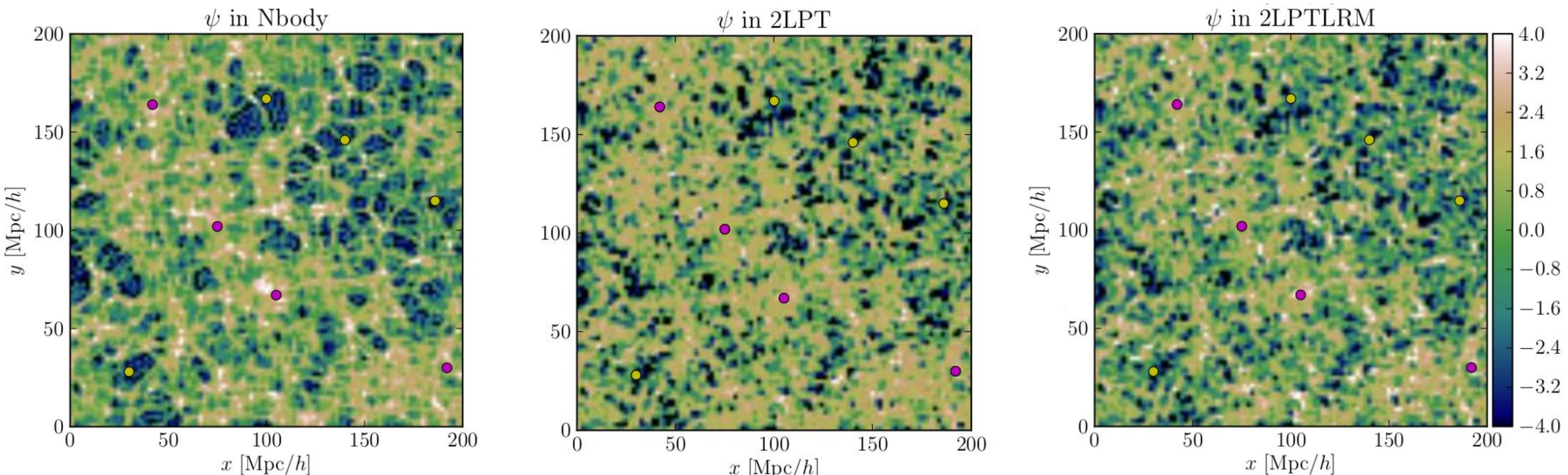


Jasche & Wandelt, arXiv:1203.3639

Beyond 2LPT?

- Recall the number of usable modes goes like k^3
- We need numerically efficient and flexible tools to model cosmic structure formation in the NL regime
- A proposal: remapping of 2LPT in the mildly non-linear regime

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



Cooking-up the mildly non-linear regime

- How to model the mode-coupling terms?
- Recipe:
  - Treat larges scales with LPT…
  - … and small scales (MC terms) with full gravity (N-body simulation
- Do time-stepping for "mode-coupling" residual. Standard N-body codes time-step the full displacement, so a lot of useless time-steps at early times/large scales just to recover the standard (linear/order 2) growth factor.

COLA (COmoving Lagrangian Acceleration)

- Write the displacement vector as:

$$\mathbf{s} = \mathbf{s}_{\mathrm{LPT}} + \mathbf{s}_{\mathrm{MC}}$$

Tassev & Zaldarriaga, arXiv:1203.5785

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

Standard:

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

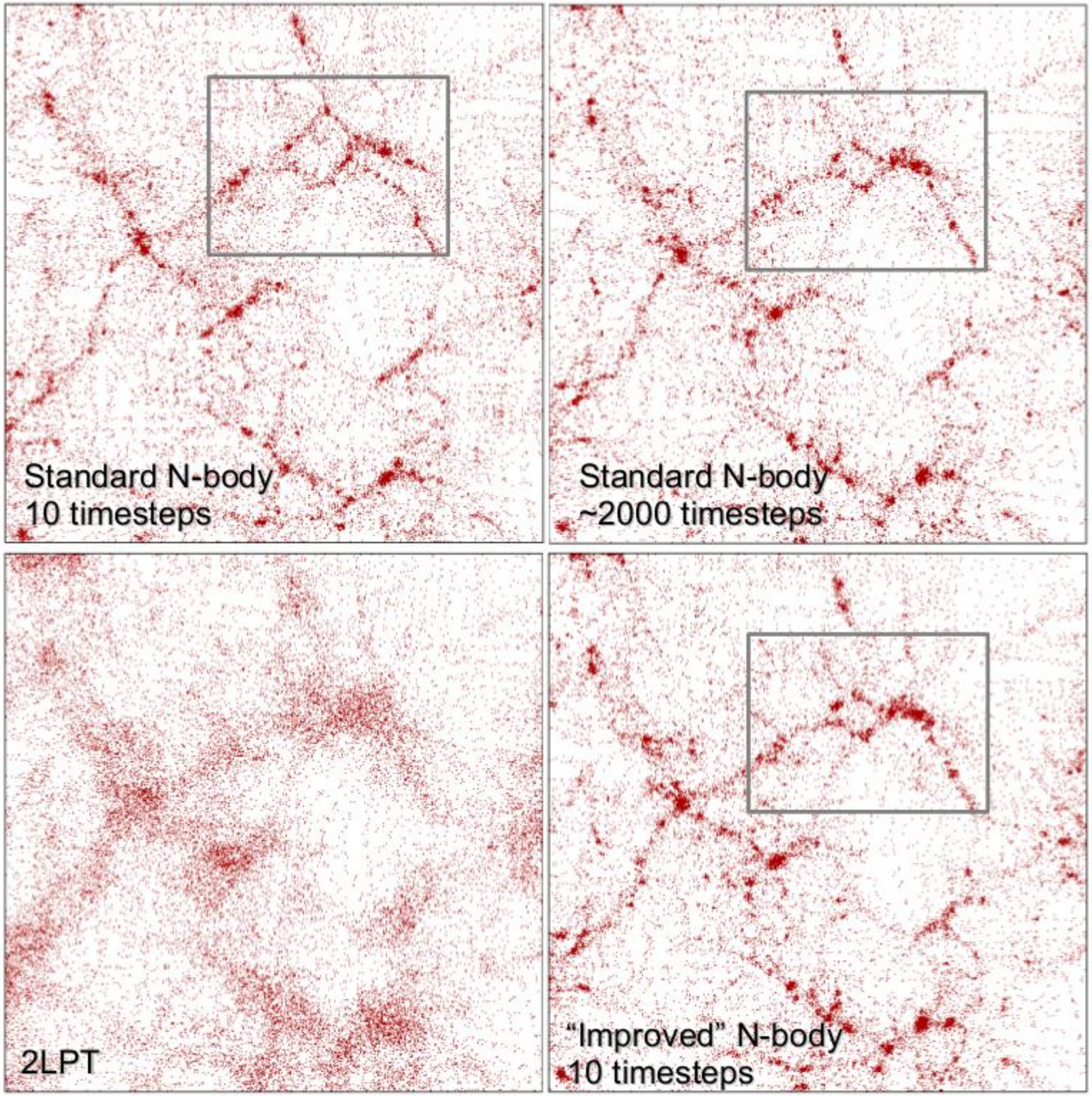
Modified:

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

Tassev, Zaldarriaga, Eisenstein 2013, arXiv:1301.0322

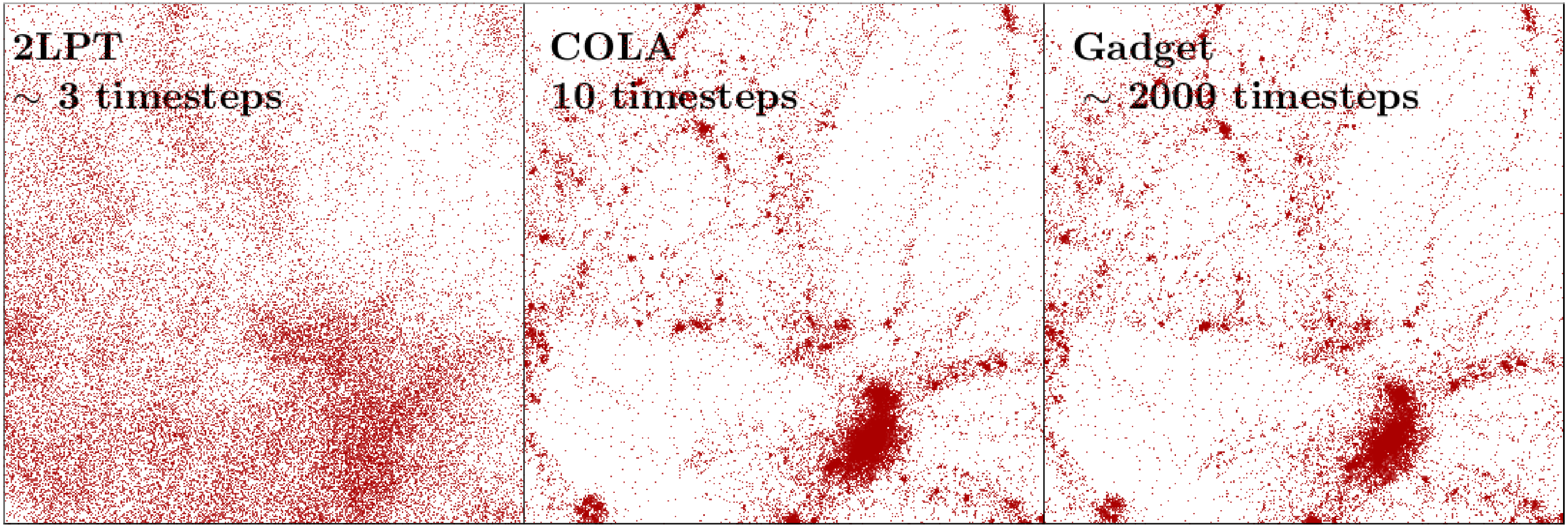
COLA: linear and mildly non-linear scales

Slices through simulations
(L = 100 Mpc/h,
thickness = 16 Mpc/h,
particle mass = $5.7\text{x}10^{11}$ Ms/h)



Tassev, Zaldarriaga, Eisenstein 2013,
arXiv:1301.0322

COLA: small scales



Slices through simulations (L=20 Mpc/h, thickness = 3Mpc/h, particle mass = 4.6×10^9 Ms/h)

Tassev, Zaldarriaga, Eisenstein 2013, arXiv:1301.0322

Concluding thoughts

- Next steps: incorporate into mock catalog generating pipeline and into Bayesian codes

- Non-linear cosmological inference of the initial conditions of the Universe becomes feasible.
  - BAO, clusters, voids
  - Non-Gaussianity
  - Isocurvature perturbations
  - Gravitational waves in LSS...

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