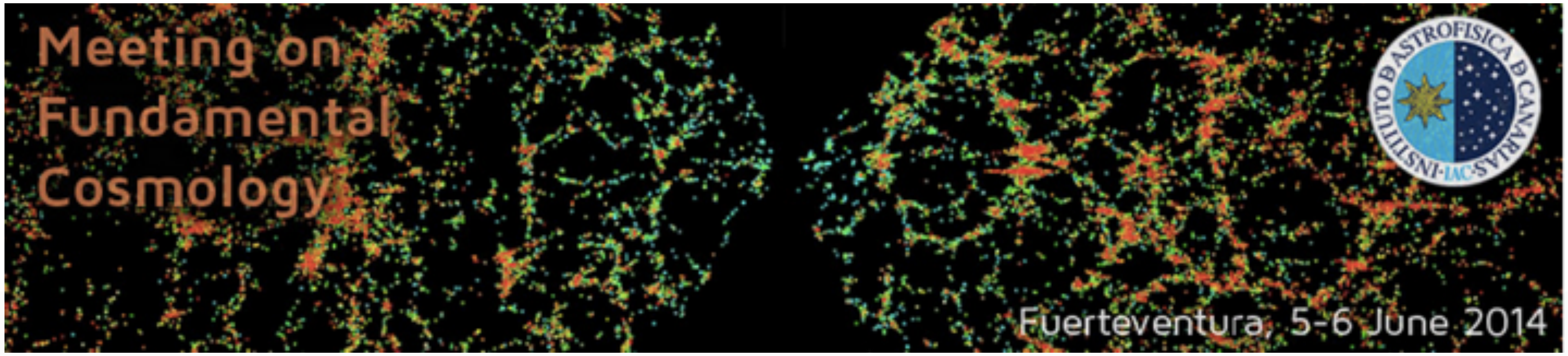


Mock galaxy catalogs with perturbation theory

AIP

Leibniz Institute for Astrophysics in Potsdam

Francisco-Shu Kitaura

Karl-Schwarzschild fellow

member of BOSS, J-PAS, eROSITA, 4MOST
co-lead of reconstruction WG in EUCLID

collaborators on this project

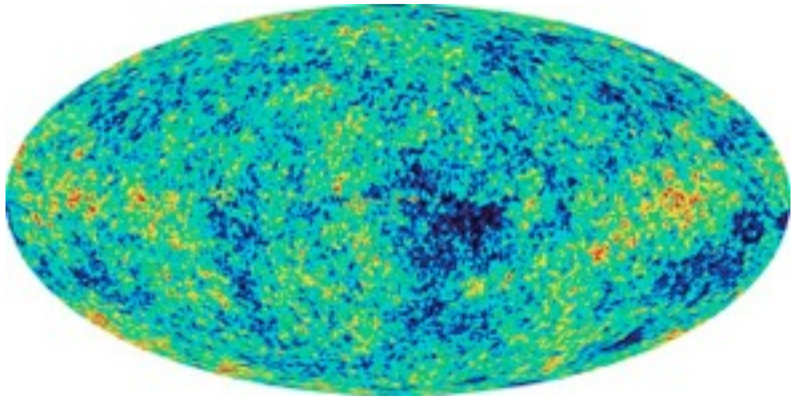
Francisco Prada   Gustavo Yepes   Volker Müller
Claudia Scoccola   Hector Gil-Marin   Chia-Hsun Chuan
Sergio Rodriguez
Steffen Heß   Stefan Gottlöber   Anatoly Klypin

How can we generate a mock galaxy catalog for present large-scale structure surveys?

(BOSS,J-PAS,DESI,EUCLID,4MOST,WEAVE,DES,…)

The ingredients

1) We start from the primordial fluctuations



Planck -8.9 < fNL < 14.3 (2 sigma)

(Suyama et al., 2013)

WMAP ~300 000 years after the Big Bang

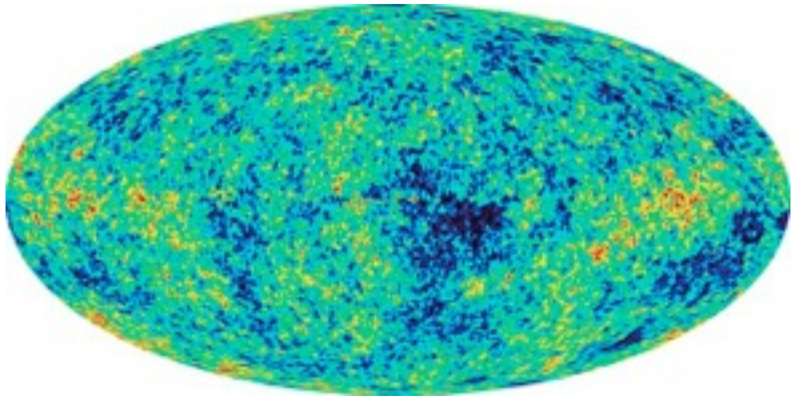
The ingredients

1) We start from the primordial fluctuations

   i) Gaussian

   ii) or non-Gaussian

Scoccimarro, Hui, Manera, Chang Chang 2012



Planck $-8.9 < fNL < 14.3$ (2 sigma)

(Suyama et al., 2013)

WMAP ~300 000 years after the Big Bang

2) We simulate structure formation

Why not simulate everything?

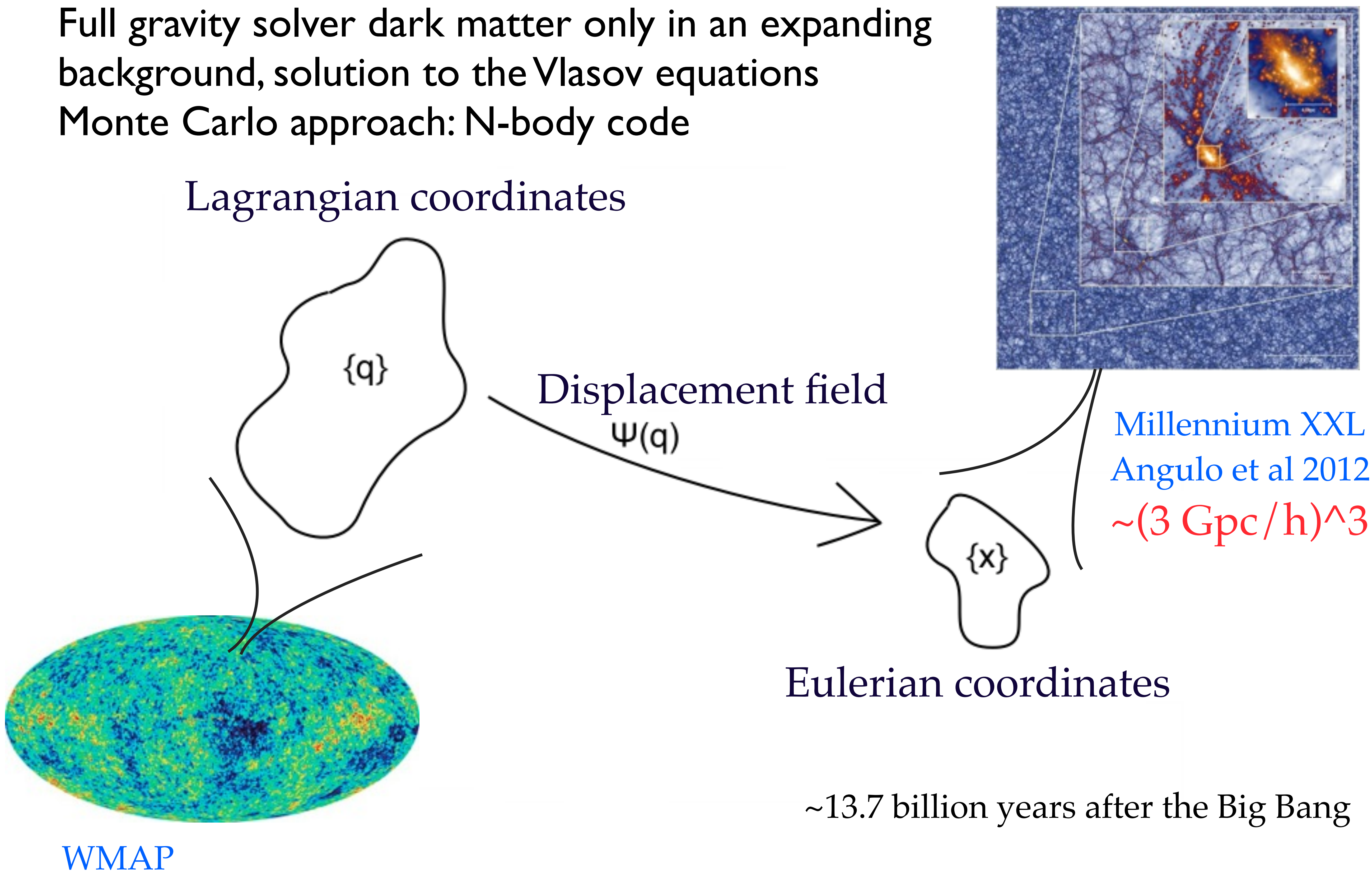
Full gravity and hydrodynamical solver in an expanding background

Vogelsberger et al 2014

about (70 Mpc/h)^3 !! too small volumes for our purposes!!

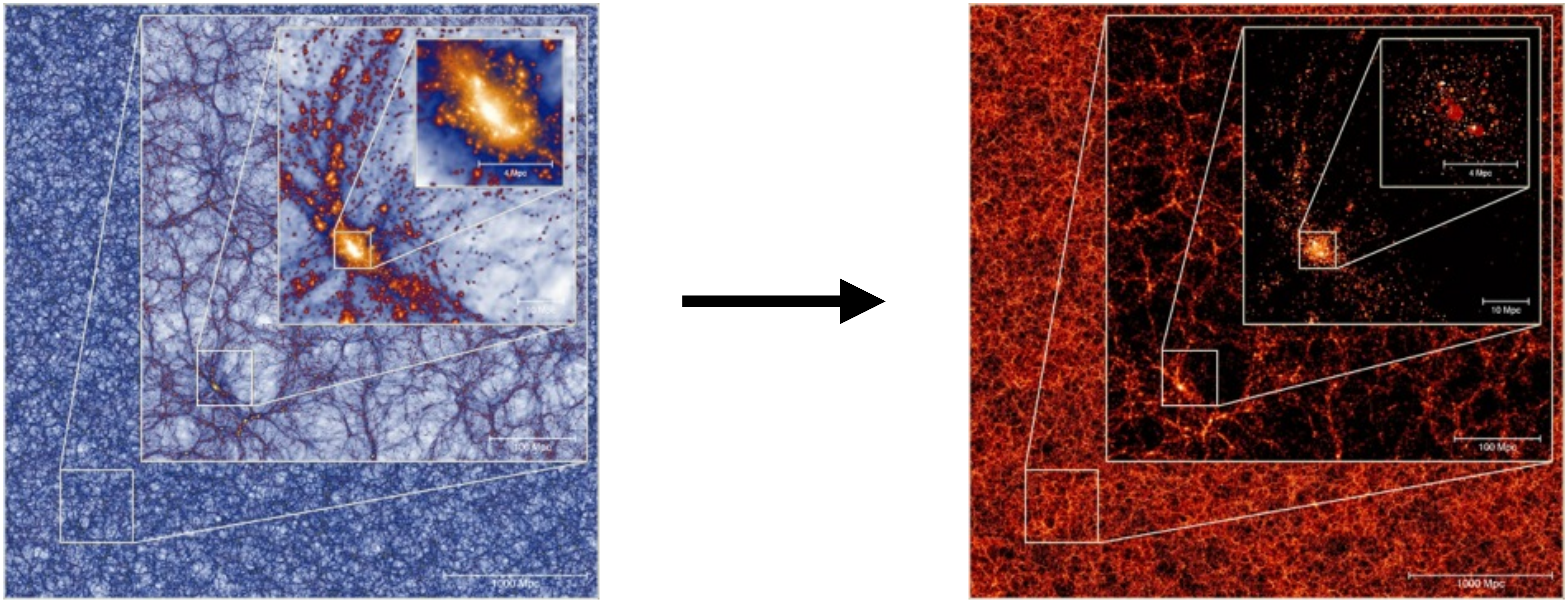
2) We simulate structure formation

Full gravity solver dark matter only in an expanding background, solution to the Vlasov equations
Monte Carlo approach: N-body code



3) We run a halo finder

Knebe et al. 2011



Millennium XXL
Angulo et al 2012

4) We produce a galaxy catalog based on either:

- semi-analytic model

Cole 2000; Haton 2003; Croton 2006; Bower 2006;
Monaco 2007; Benson 2010; deLucia & Blaizot 2007; Baugh 2006

- halo occupation distribution

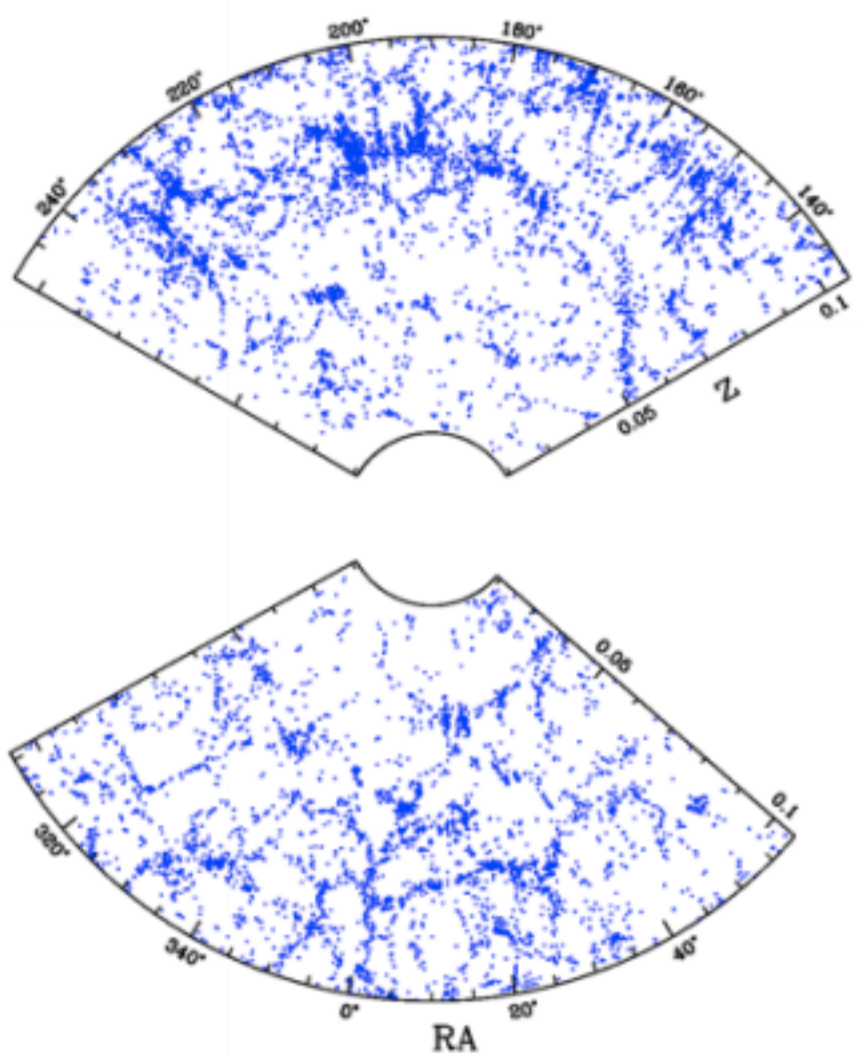
Berlind & Weinberg 2002; Zheng 2005; Zehavi et al 2011

- abundance matching

Kratsov 04, Tasitsiomi 04, Vale&Ostriker04, Conroy 06,
Behroozi10, Trujillo-Gomez 11, Nuza 12

5) We apply observational effects

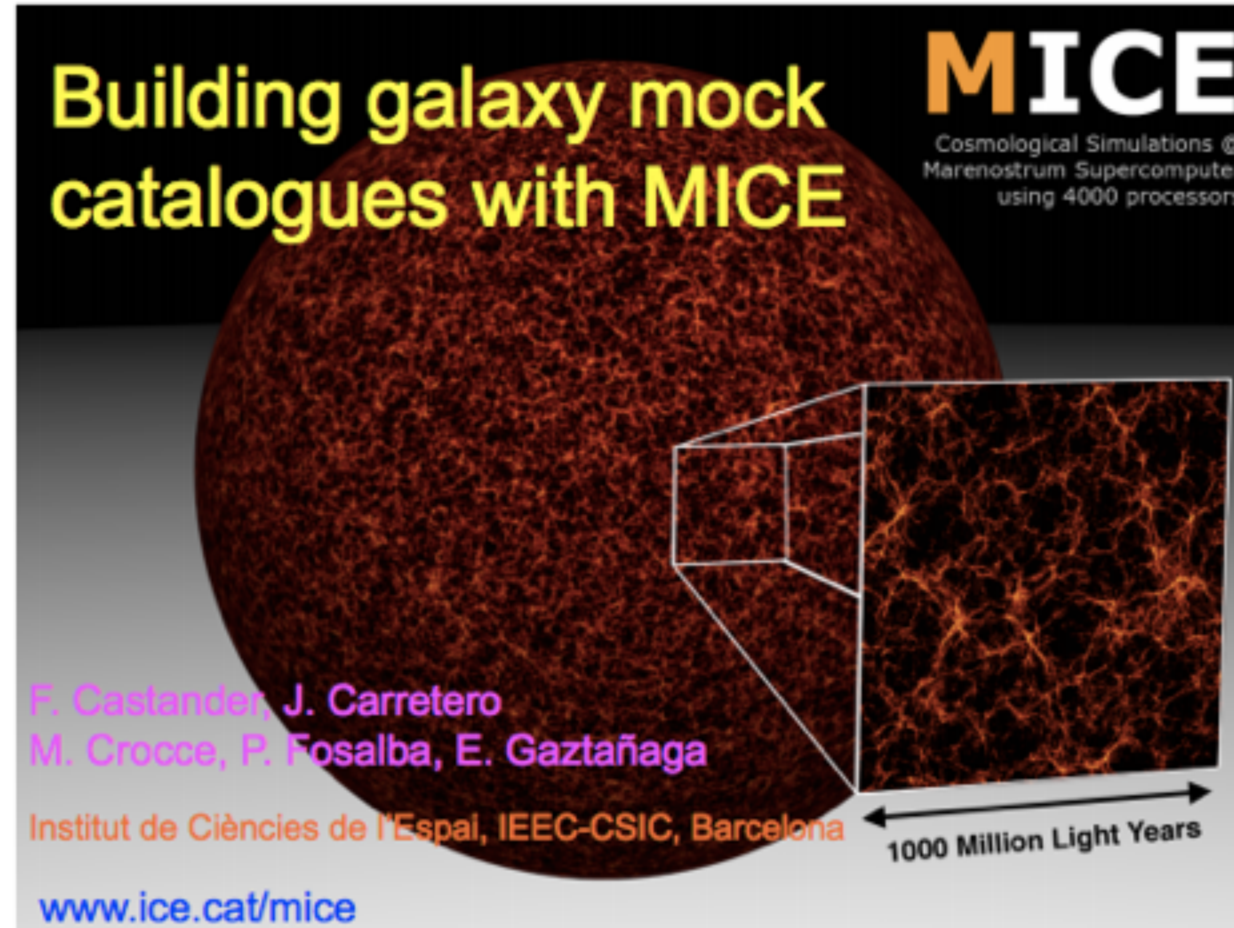
- selection function (magnitude cut)
- survey mask
- redshift-space distortions



Marc Manera SDSS

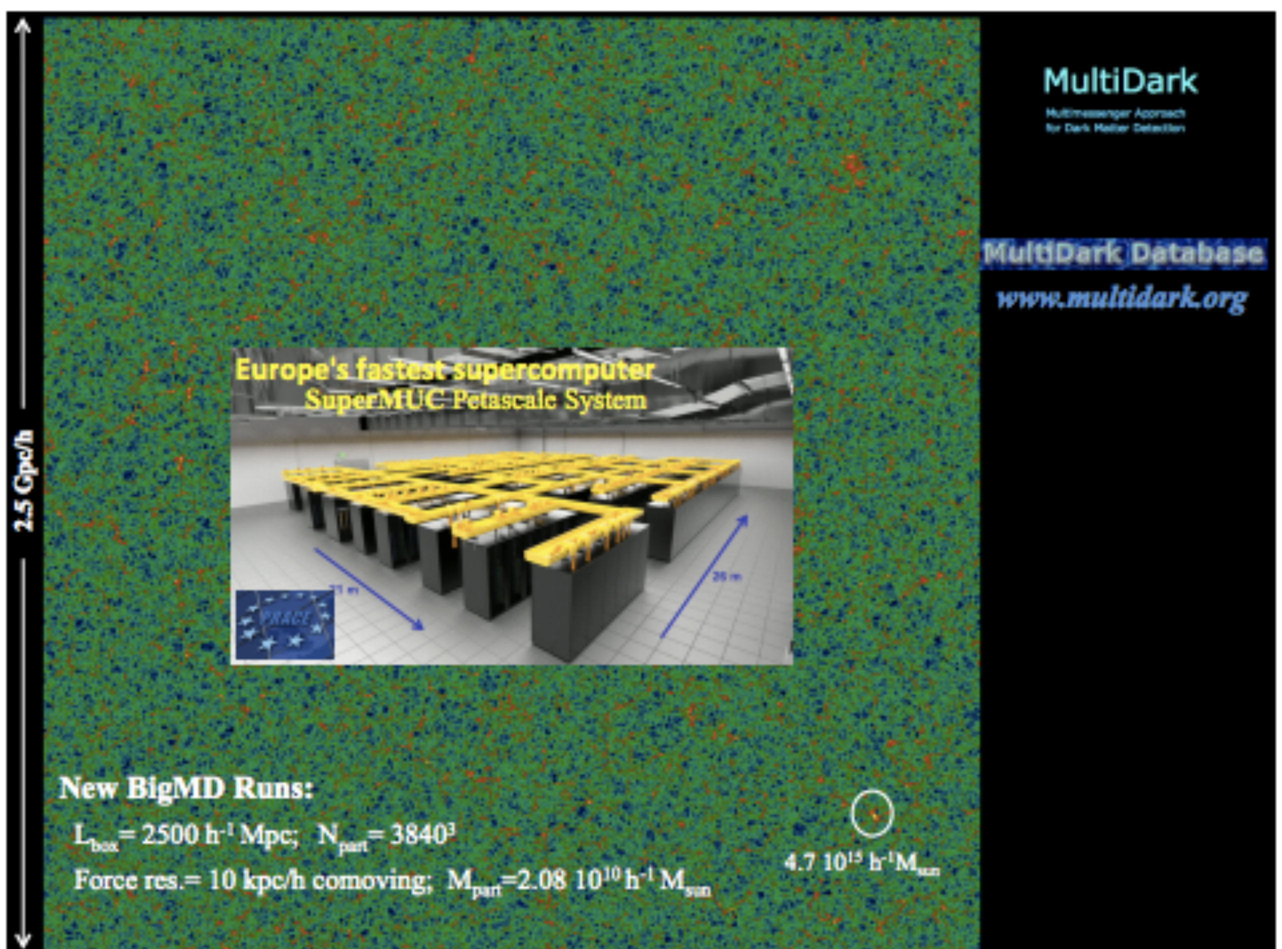
extraordinary effort in Spain!

remember Pablo Fosalba’s talk



or with people from Spain involved:
BigMultiDark / MICE /
MXXL Raul Angulo /
PThalos Marc Manera /
PATCHY FK

BigMultiDark



see Paco Prada’s talk ~(2.5 Gpc/h)^3

MICE ~(3 Gpc/h)^3

END OF MY TALK?

However,
this effort can only be done a couple of times,
but not thousands of times!

we need to scan different cosmologies,
different seeds (cosmic variance)

to estimate error bars/covariance matrices
to measurements based on our unique Universe;

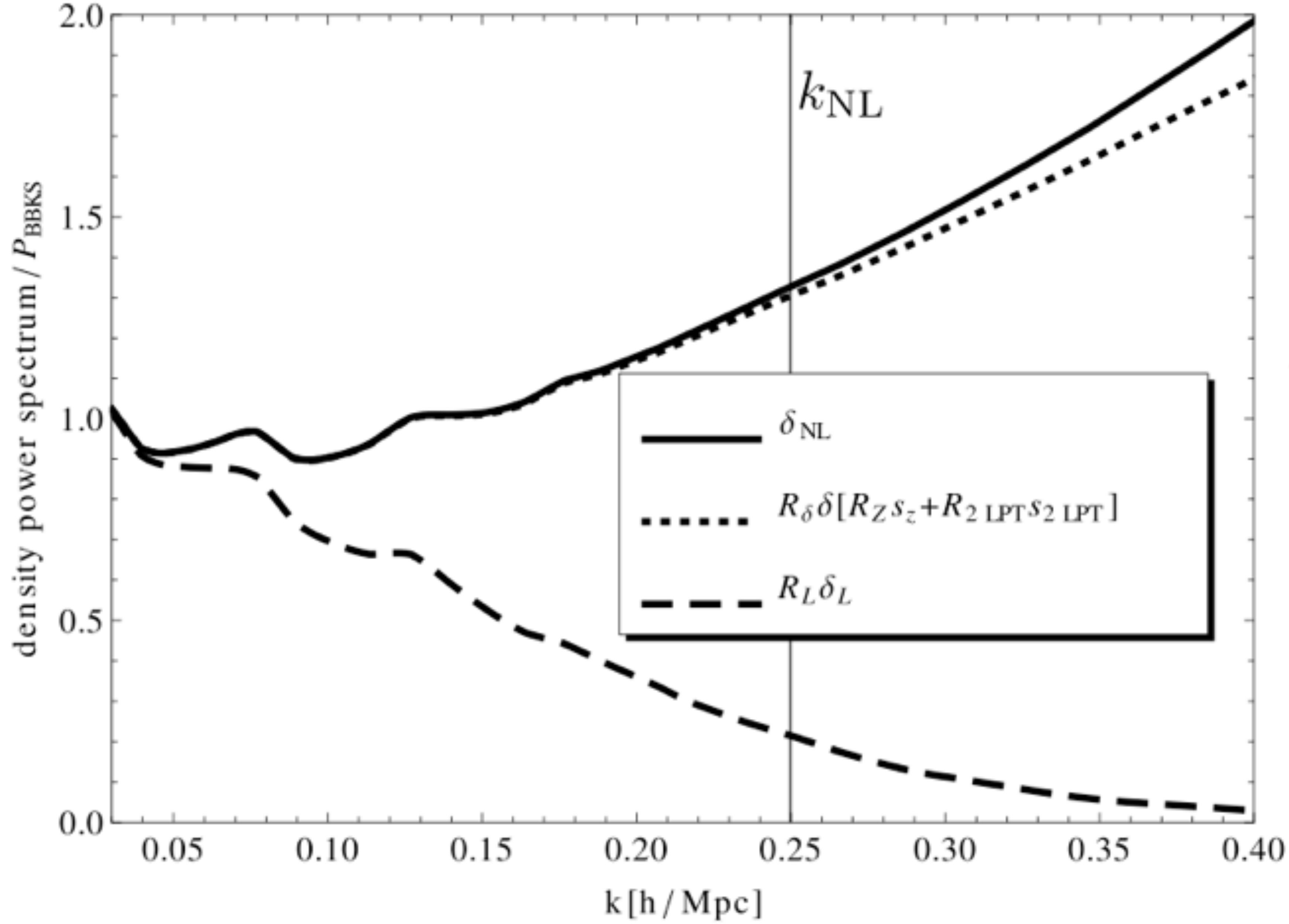
to test analysis tools with different realisations
in a statistical relevant way

What is the bottle-neck of the computations?

What is the bottle-neck of the computations?

1. the gravity solver

Perturbative approaches to model BAOs



$$\delta(\boldsymbol{k}) = R_\delta(k)\delta_1(\boldsymbol{k})$$

$$R_\delta(k,\eta) = \exp(0.58d)$$

Tassev & Zaldarriaga 2012

deviation in the power spectrum due to perturbative approach!

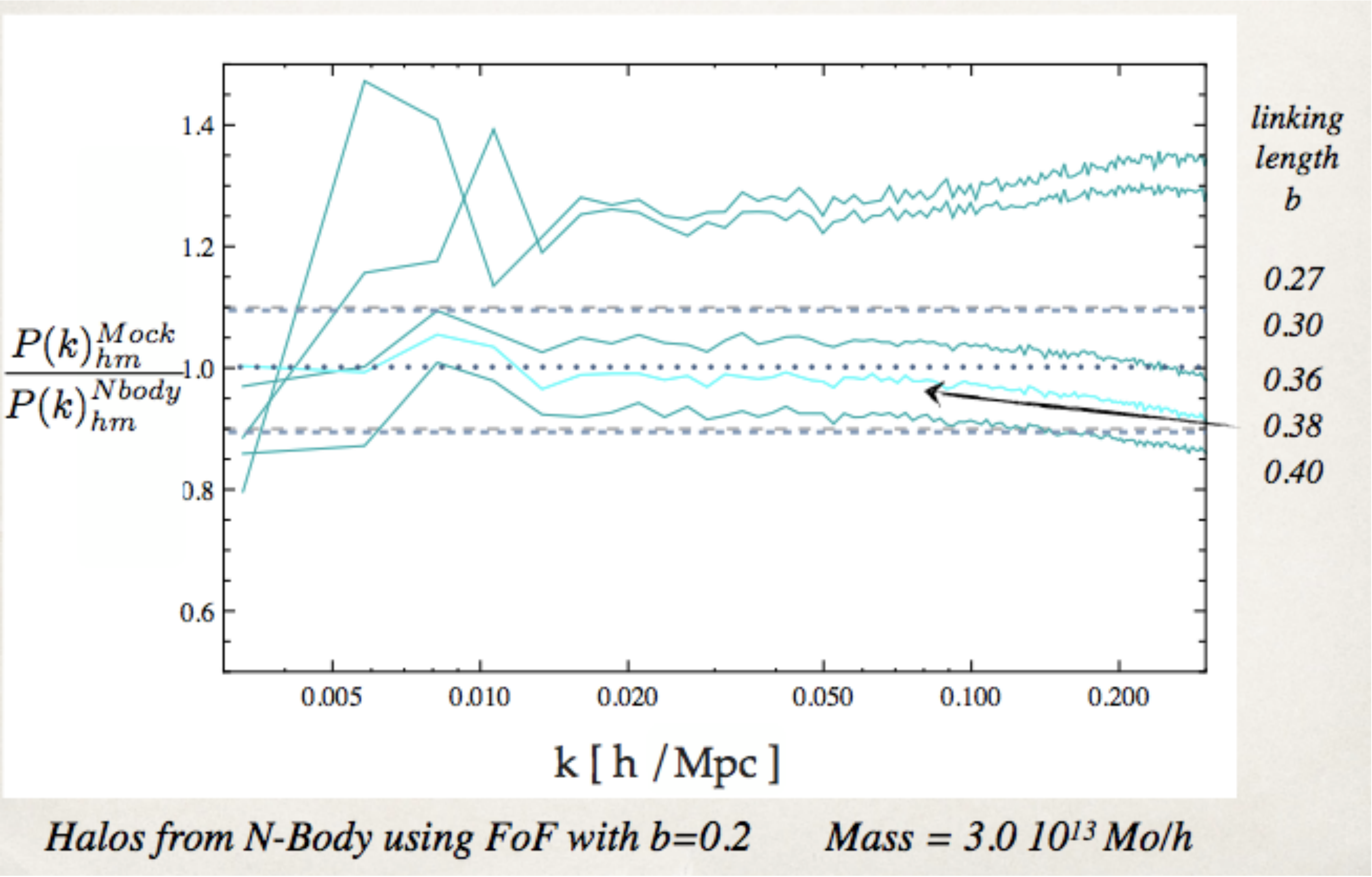
let us use approximate gravity solvers!

pioneering works by

1. Scoccimarro & Sheth 2002 (PThalos)
   2LPT
2. Monaco et al 2002 (Pinocchio)
   Zeldovich (being improved to 2LPT)
   includes merging histories

3. Tassev et al 2013 (COLA)
   make N-body code faster
4. White et al 2014 (QPM)
   use quasi PM solver

PThalos



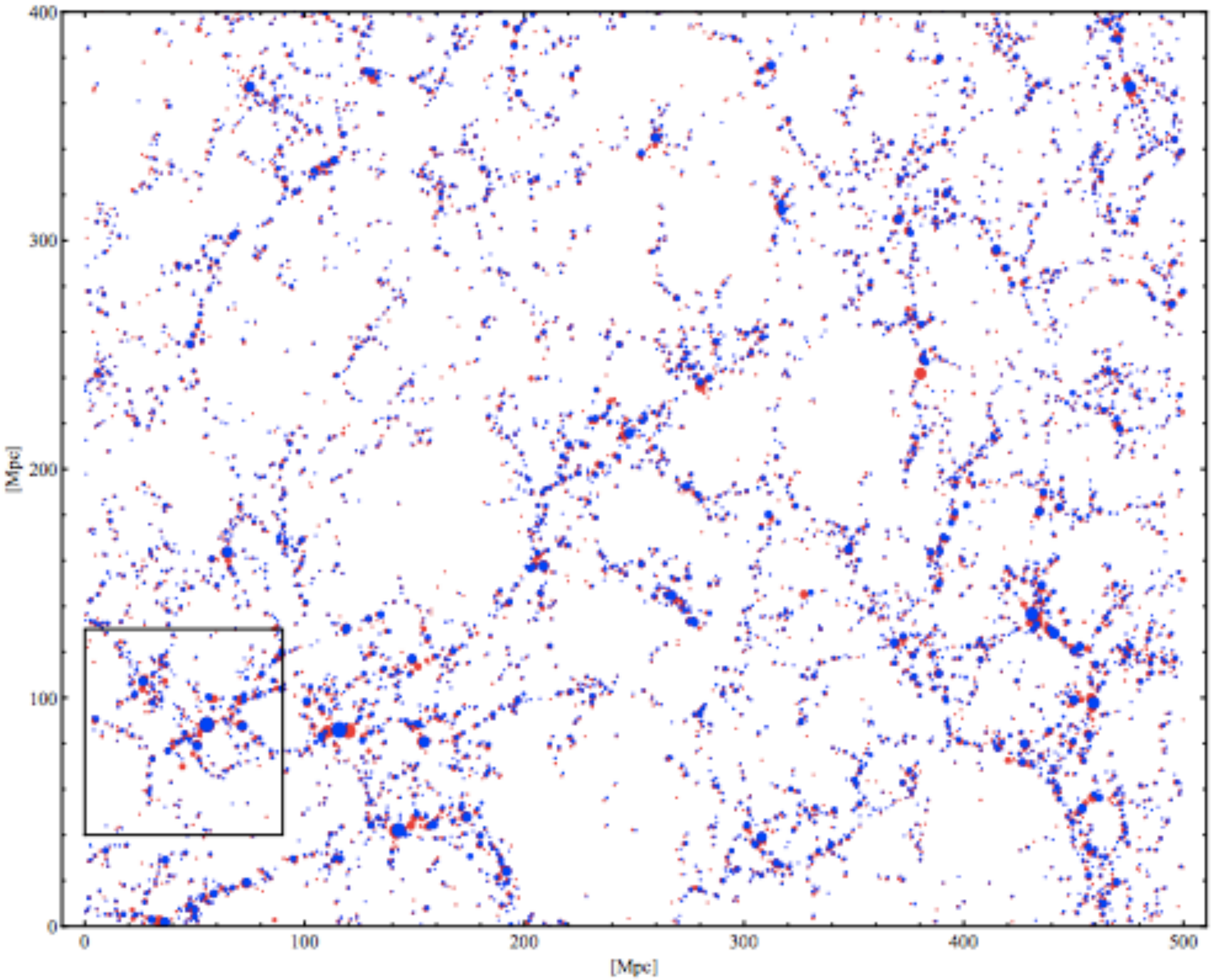
calibration with N-body
1000 mocks done!

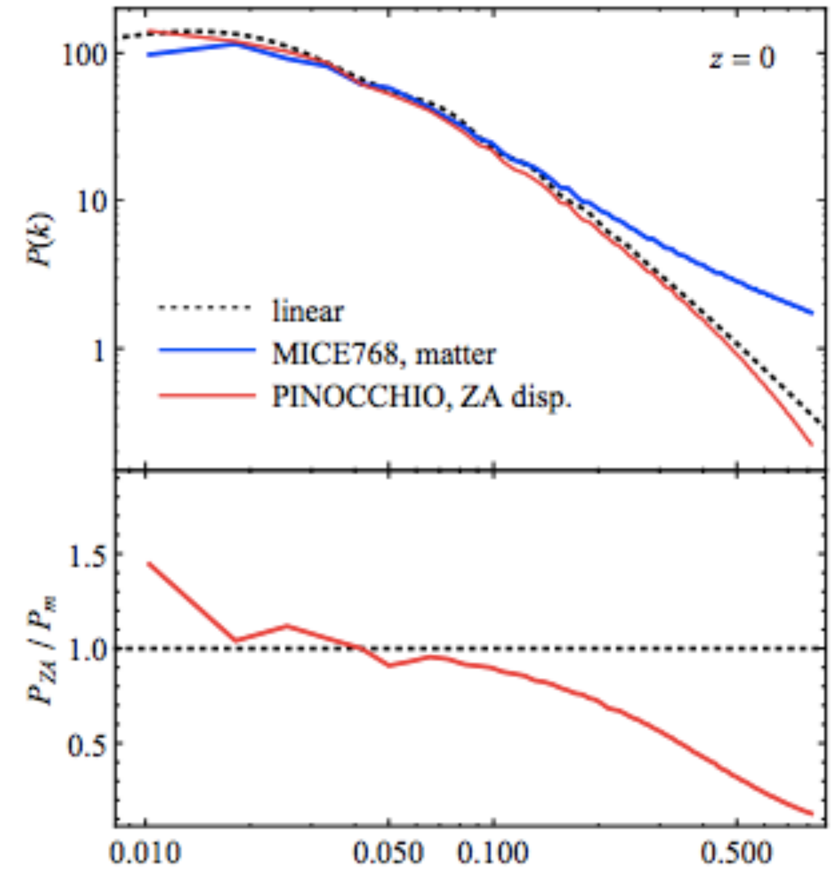
about 10 % deviation in the power-spectrum at k~0.25
needs to resolve halos

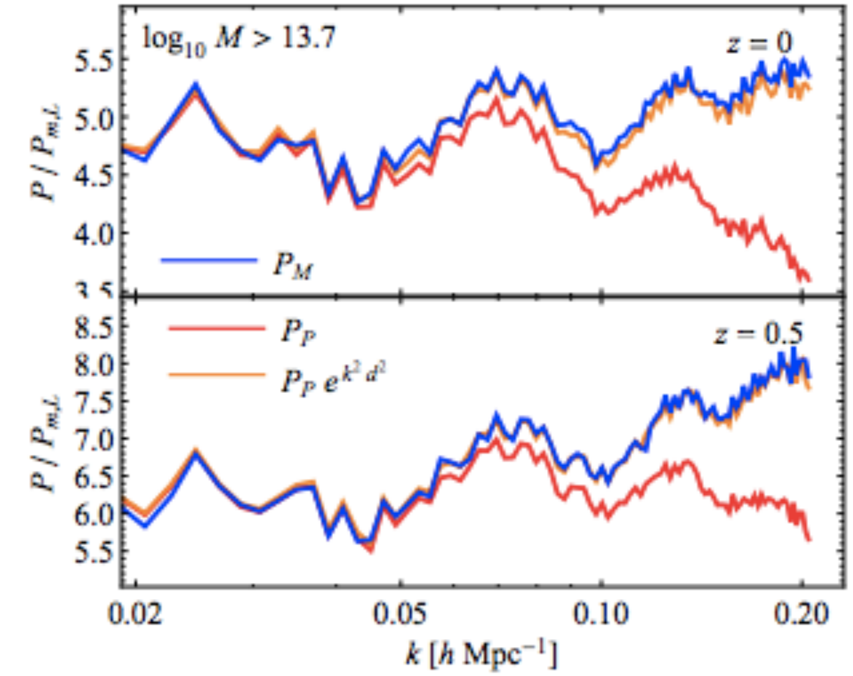
see galaxy mocks from Manera M. & BOSS collaboration 2012 &2014, MNRAS

Pinocchio

Monaco et al 2013







fast + predictive (merger trees)

deviation in the power spectrum
needs to resolve halos

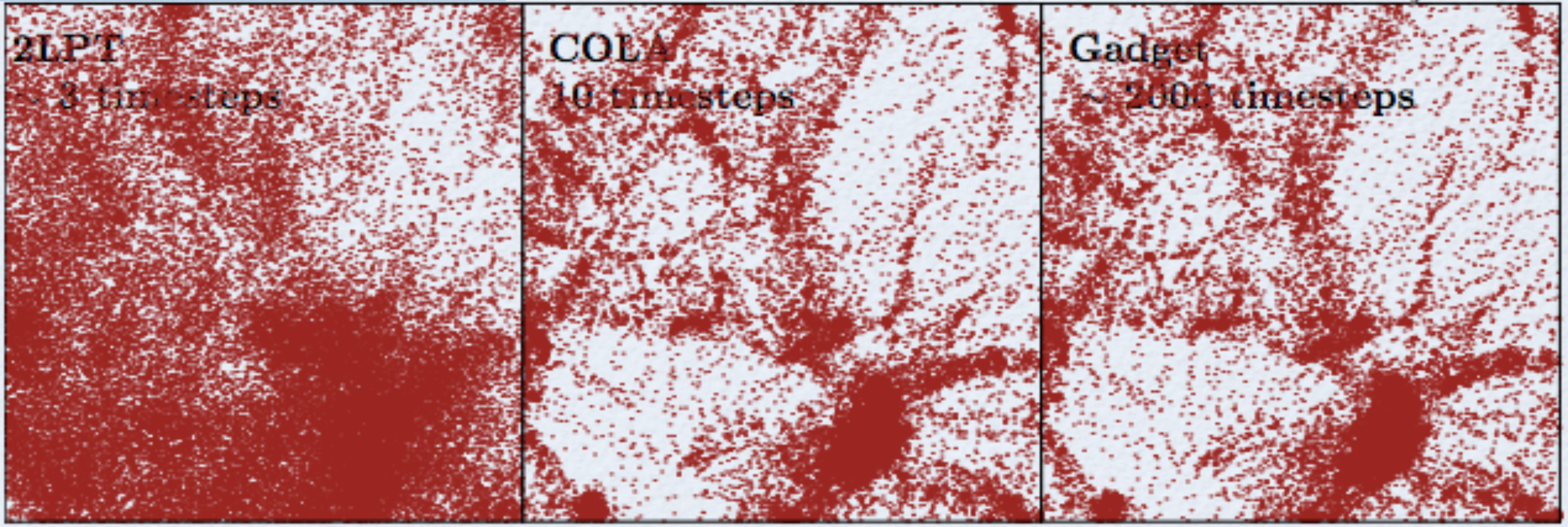
COLA 10-timestep simulations

Koda et al 2014

WiZ-COLA simulation

2LPT + Particle Mesh force correction

Tassev et al. 2013 JCAP **06** 36



2LPT (*left*): very fast but resolves high-mass haloes only ~$10^{13} M_\odot$

Manera et al. 2013 MNRAS **428** 1036 (BOSS)

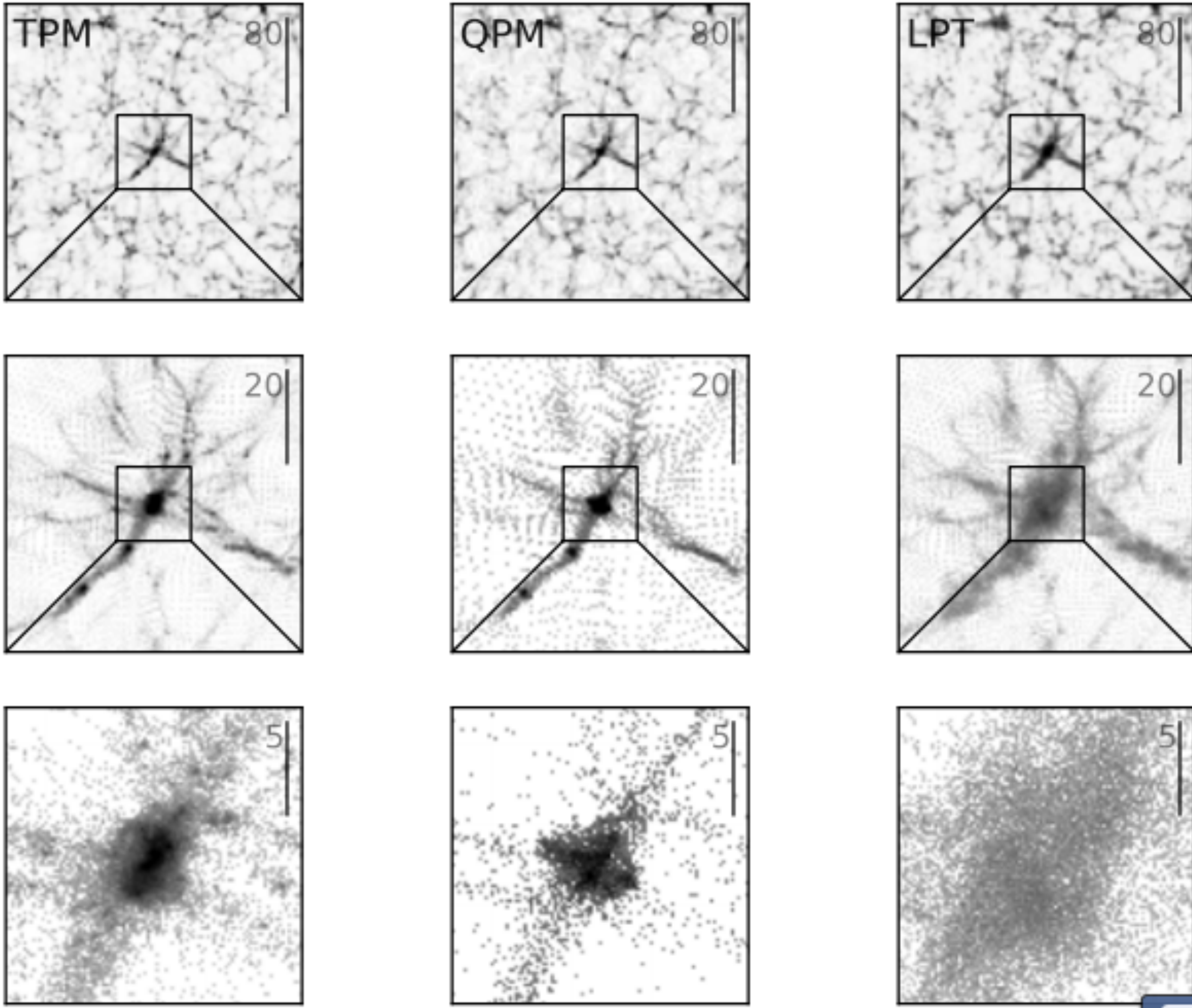
COLA (*middle*): COmoving Lagrangian Acceleration (COLA) method.

- resolves halo mass (~ 5%)
- accurate large-scale structure
- more than 100 times faster than usual simulations
- with 2LPT, larger timestep is allowed at high z → better halo mass

faster than common N-body,
but needs to resolve halos

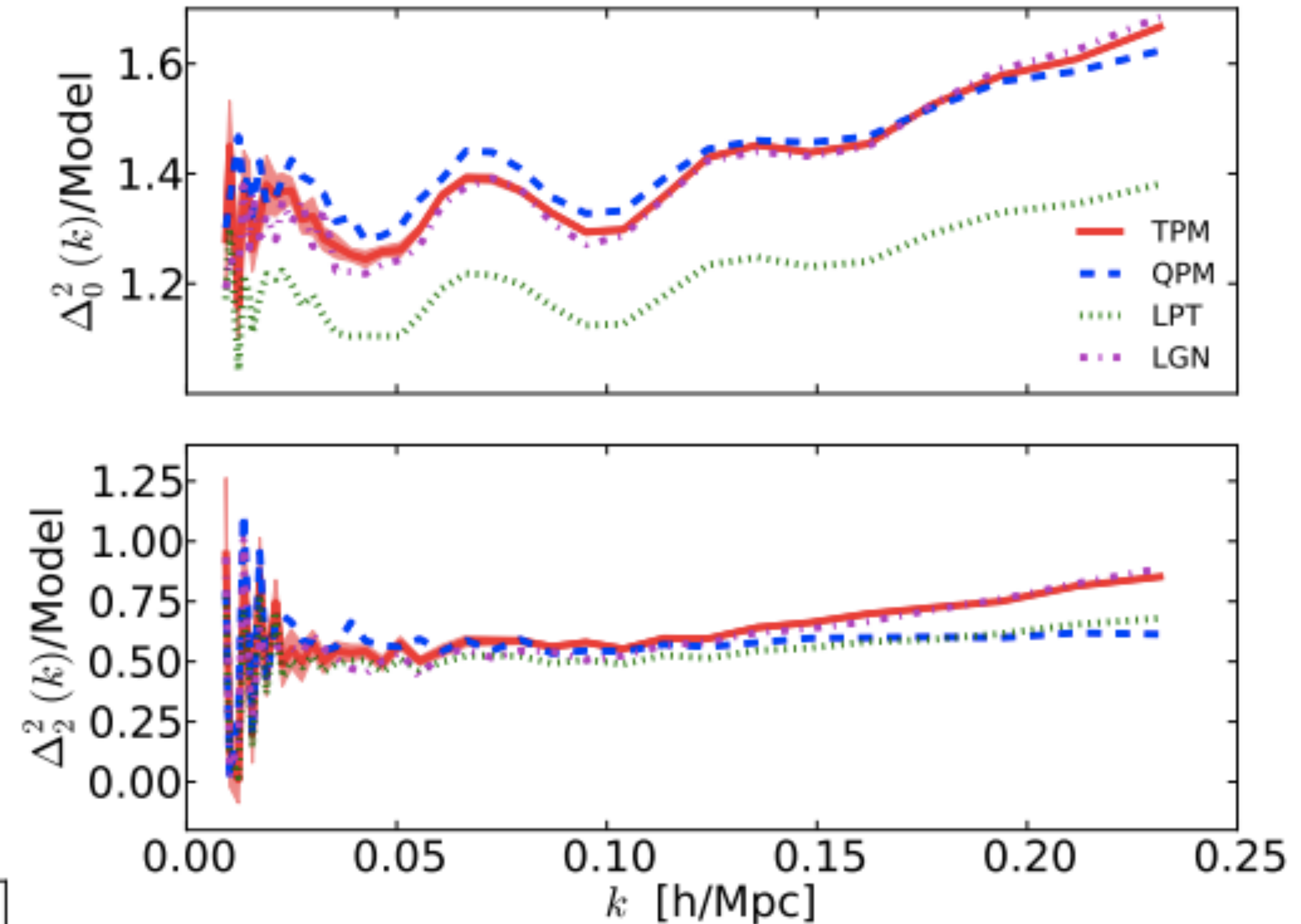
QPM

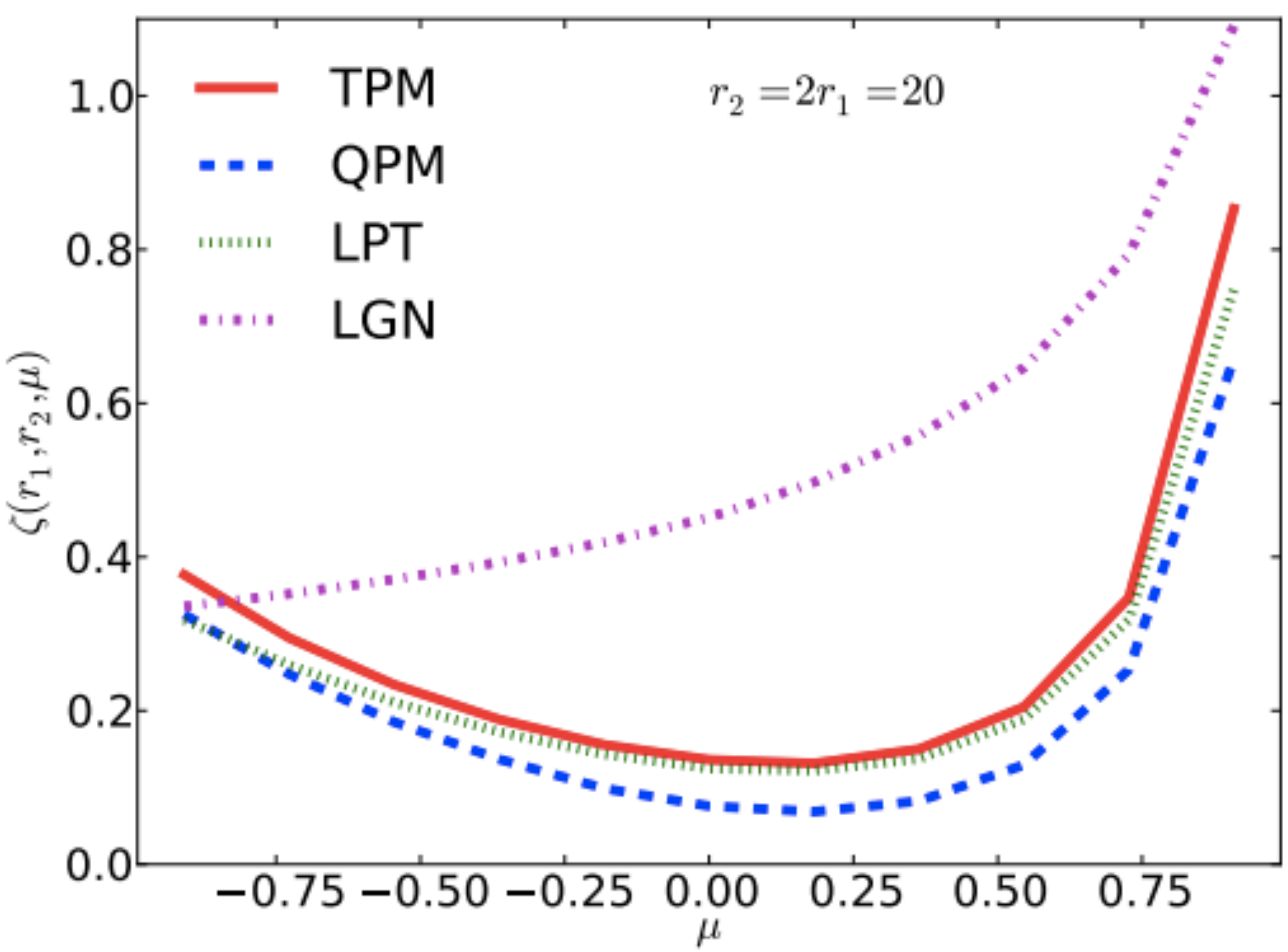
White et al 2013



QPM

White et al 2013





fast, but approximate PM method leads to worse 3pt statistics than LPT

is the BAO damping correctly modelled?

What is the bottle-neck of the computations?

1. the gravity solver
2. resolution (number of particles)

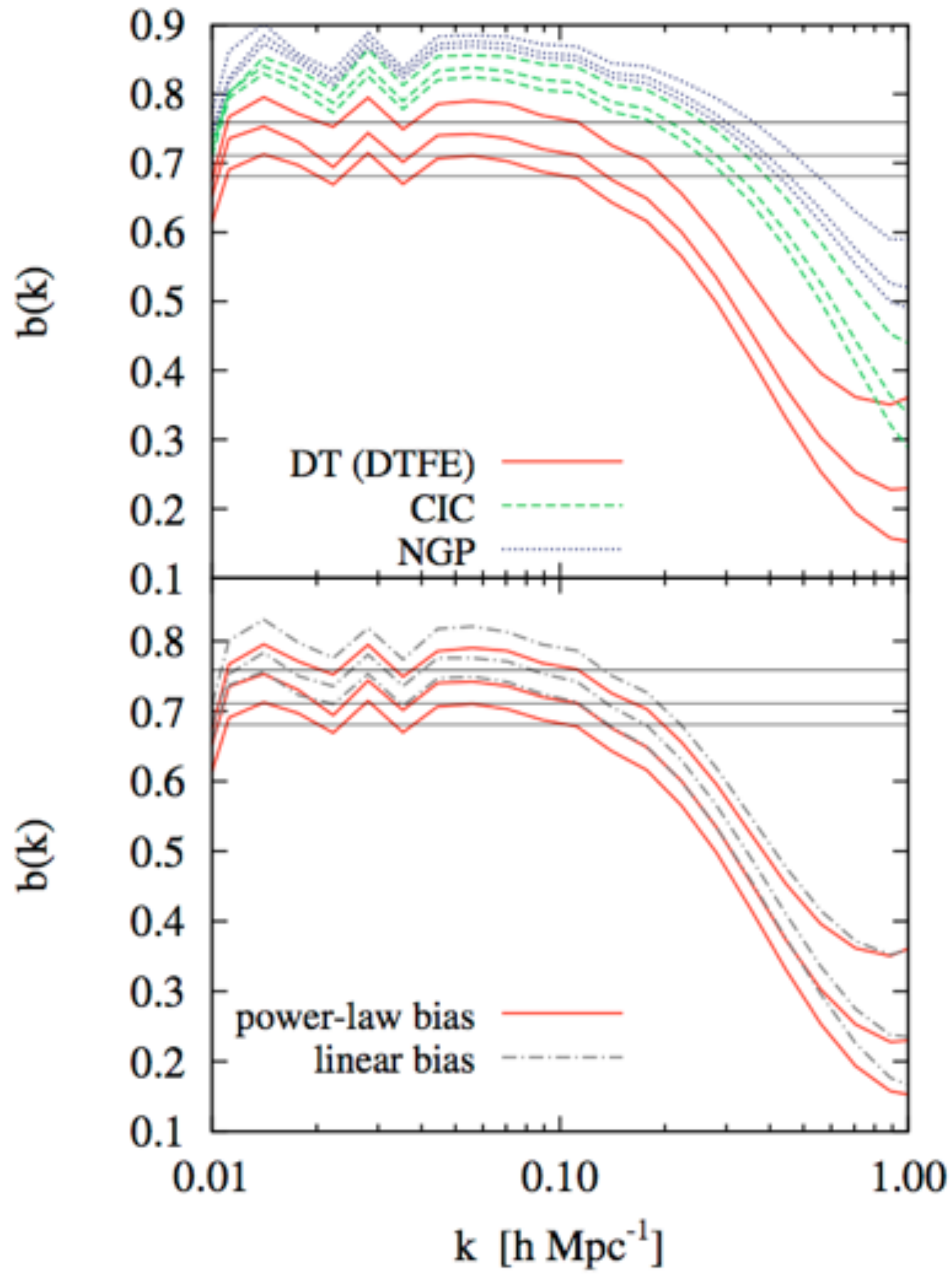
Let us use low resolution simulations and augment the missing halos!

Nonlinear scale-dependent biasing

exponential bias

<N>=f (1+delta)$^{\text{biasE}}$

de la Torre, S. & Peacock, J. 2012
Cen, R. & Ostriker, J. P. 1993



is equivalent to an infinite expansion a la Fry & Gaztañaga 1993 and can be interpreted as a Lagrangian bias in the lognormal approximation

Angulo et al 2014 with Fry & Gaztañaga 1993

Let us ask a different question:

How can I generate a distribution of halos/galaxies in a statistical way?

(of a certain type within a certain mass range) the answer is:

we need all the higher order statistics from a reference sample
+ a way to draw from such a higher order PDF:

$$N_h \curvearrowleft \mathcal{P}(N_h|\xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \dots)$$

How can we do it?
Method 1:

If we know the higher order correlations for the logarithm of the density field, we can perform multidimensional Edgeworth expansions:

$$
\begin{aligned}
&P(\mathbf{\Phi}) = (\det(\mathbf{S}))^{-1/2} G(\boldsymbol{\nu}) \qquad (43)\\
&\times \Big[1 + \frac{1}{3!} \sum_{i'j'k'} \langle \Phi_{i'} \Phi_{j'} \Phi_{k'} \rangle_c \sum_{ijk} S_{ii'}^{-1/2} S_{jj'}^{-1/2} S_{kk'}^{-1/2} h_{ijk}(\boldsymbol{\nu})\\
&+ \frac{1}{4!} \sum_{i'j'k'l'} \langle \Phi_{i'} \Phi_{j'} \Phi_{k'} \Phi_{l'} \rangle_c \sum_{ijkl} S_{ii'}^{-1/2} S_{jj'}^{-1/2} S_{kk'}^{-1/2} S_{ll'}^{-1/2} h_{ijkl}(\boldsymbol{\nu})\\
&+ \frac{1}{6!} \sum_{i'j'k'l'm'n'}\\
&\times \Big[\frac{1}{3!3!2} \sum_{j_1 \ldots j_6 \in [1,\ldots,6]} \tilde{\epsilon}_{j_1 \ldots j_6} \langle \Phi_{i'_{j_1}} \Phi_{i'_{j_2}} \Phi_{i'_{j_3}} \rangle_c \langle \Phi_{i'_{j_4}} \Phi_{i'_{j_5}} \Phi_{i'_{j_6}} \rangle_c \Big]_{10}\\
&\times \sum_{ijklmn} S_{ii'}^{-1/2} S_{jj'}^{-1/2} S_{kk'}^{-1/2} S_{ll'}^{-1/2} S_{mm'}^{-1/2} S_{nn'}^{-1/2} h_{ijklmn}(\boldsymbol{\nu}) + \ldots \Big],
\end{aligned}
$$

Kitaura 2012

Colombi 1994

complicated + expensive!

How can we do it?
Method 1:

If we know the higher order correlations for the logarithm of the density field, we can perform multidimensional Edgeworth expansions:

$$P(\boldsymbol{\nu}) = G(\boldsymbol{\nu})\left[1 + \mathcal{S}(\boldsymbol{\nu}) + \mathcal{K}(\boldsymbol{\nu}) + \ldots\right]$$

Kitaura 2012

$$\boldsymbol{\Phi} \equiv \ln \boldsymbol{\rho} - \langle \ln \boldsymbol{\rho} \rangle \qquad \nu_i \equiv \sum_j S_{ij}^{-1/2} \Phi_j$$

$$\mathcal{S}(\boldsymbol{\nu}) \equiv \frac{1}{3!}\sum_{ijk} \kappa_{ijk} h_{ijk}(\boldsymbol{\nu}) = \frac{1}{3!}\sum_{i'j'k'} \xi_{i'j'k'} \tilde{h}_{i'j'k'}(\boldsymbol{\nu}) \qquad (60)$$

$$= \frac{Q_3}{3!}\sum_{i'j'k'} \left[S_{i'j'}S_{i'k'} + S_{i'j'}S_{j'k'} + S_{i'k'}S_{j'k'}\right] \tilde{h}_{i'j'k'}(\boldsymbol{\nu})$$

$$= Q_3 \left[\frac{1}{2}\sum_i \Phi_i^2 \eta_i - \frac{1}{2}\sum_i S_{ii}\eta_i - \sum_i \Phi_i\right],$$

$$\eta_i \equiv \sum_j S_{ij}^{-1} \Phi_j$$

Let us imagine we would know the halo/galaxy density field, i.e. the expected number of halos/galaxies per finite volume (cell).

Stochastic biasing

$$N_h \curvearrowleft \mathcal{P}(N_h|\rho_h)$$

caution! we still need to know the deviation from Poissonity!

over-dispersion modelled by the NB PDF:

$$P(N_i \mid \lambda_i, \beta) = \frac{\lambda_i^{N_i}}{N_i!} \frac{\Gamma(\beta + N_i)}{\Gamma(\beta)(\beta + \lambda)^{N_i}} \frac{1}{(1 + \lambda/\beta)^{\beta}}$$

Kitaura, Yepes & Prada 2014, MNRAS

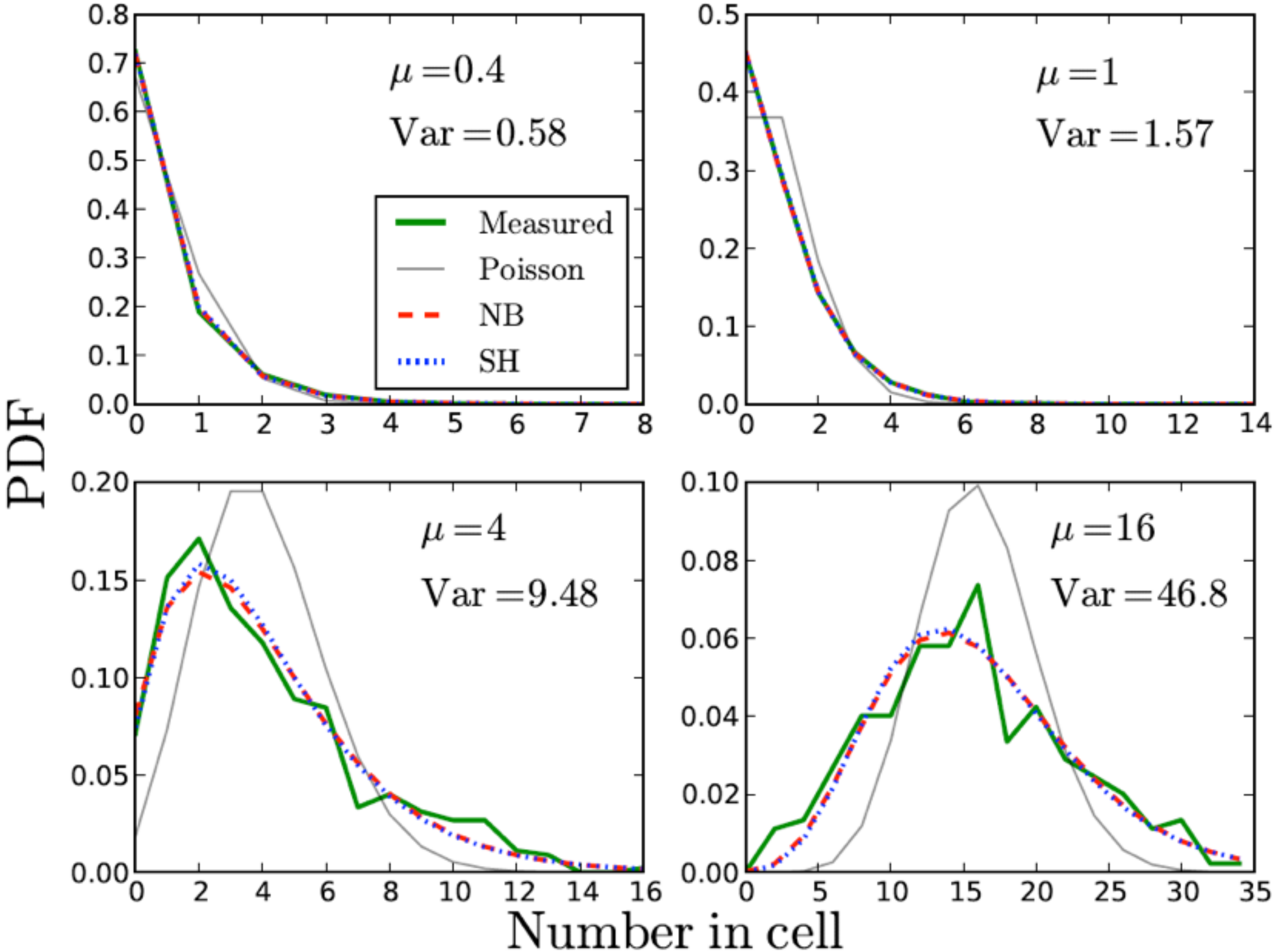
non-Poissonian PDFs:
W.C. Saslaw, A.J.S. Hamilton, 1984, ApJ, 276, 13
Sheth R. K., 1995, MNRAS, 274, 213

stochastic bias:
Dekel A., Lahav O., 1999, ApJ, 520, 24
Sheth R. K., Lemson G., 1999, MNRAS, 304, 767

and many more see references in e.g. Kitaura et al 2013

Somerville et al 2001, MNRAS, 320, 289
Casas-Miranda et al 2002, MNRAS, 333, 730

Neyrinck M et al 2013; Aragon-Calvo M. 2013



Deterministic biasing

we need to know the (deterministic) biasing,
but this implies knowing all the higher order correlation functions!

$$B(\rho_h|\rho_{\mathrm{M}}) = B(\rho_h|\rho_{\mathrm{M}}, \xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \ldots)$$

$$N_h \curvearrowleft \mathcal{P}(N_h|B(\rho_h|\rho_{\mathrm{M}}, \xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \ldots))$$

Deterministic biasing parametrization

Fry & Gaztañaga 1993

$$\rho_h = f_h^a \sum_i a_i \delta_{\mathrm{M}}^i$$

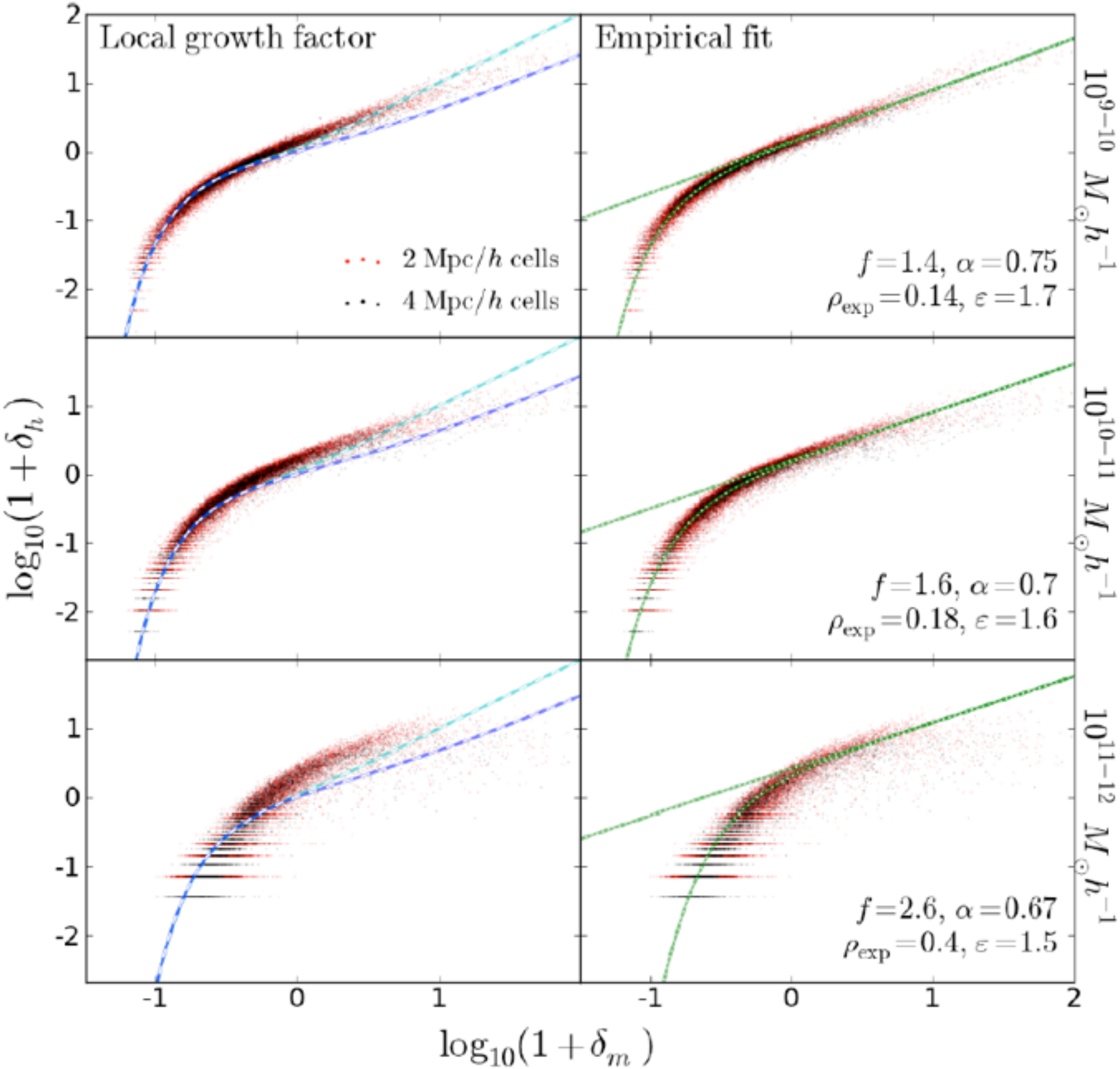
Cen & Ostriker 1993

$$\rho_h = f_h^b \exp\left[\sum_i b_i \log\left(1+\delta_{\mathrm{M}}\right)^i\right]$$

Kitaura, Yepes & Prada 2014; Neyrinck et al 2014

$$\rho_h = f_h\, \theta(\rho_{\mathrm{M}} - \rho_{\mathrm{th}})\, \rho_{\mathrm{M}}^{\alpha} \exp\left[-\left(\frac{\rho_{\mathrm{M}}}{\rho_\epsilon}\right)^{\epsilon}\right]$$

Neyrinck M et al 2013; Aragon-Calvo M. 2013



How can we do it?
Method II:

given the dark matter field
(from low resolution N-body or approximate solver)
parametrize the bias
and constrain the bias parameters, in such a way that the higher order correlation functions are matched

expensive!

How can we do it?
Method III:

Kitaura et al in prep

$$\bar{N}_h = \langle \rho_h \rangle \leftarrow \xi_1^h$$

$$P_h(k) \leftarrow \xi_2^h$$

$$\mathcal{P}^1(\rho) = \int_{-\sqrt{-1}\infty}^{\sqrt{-1}\infty} \frac{\mathrm{d}t}{2\pi\sqrt{-1}} \exp\left(t\rho + \mathcal{C}(t)\right)$$

$$\mathcal{P}_h^1\left(B\left(\rho_h|\rho_{\mathrm{M}}\right)\right) \leftarrow \{\xi_1^h, \xi_2^h, \xi_3^h, \xi_4^h, \ldots\}$$

$$N_h \curvearrowleft \mathcal{P}(N_h|B(\rho_h|\rho_{\mathrm{M}}, \bar{N}_h, P_h(k), \mathcal{P}_h^1))$$

our approach:

Let us use low resolution + approximate gravity solvers and augment all the missing halos!

- The approximate gravity solver accurately models the higher order statistics of the dark matter density field.

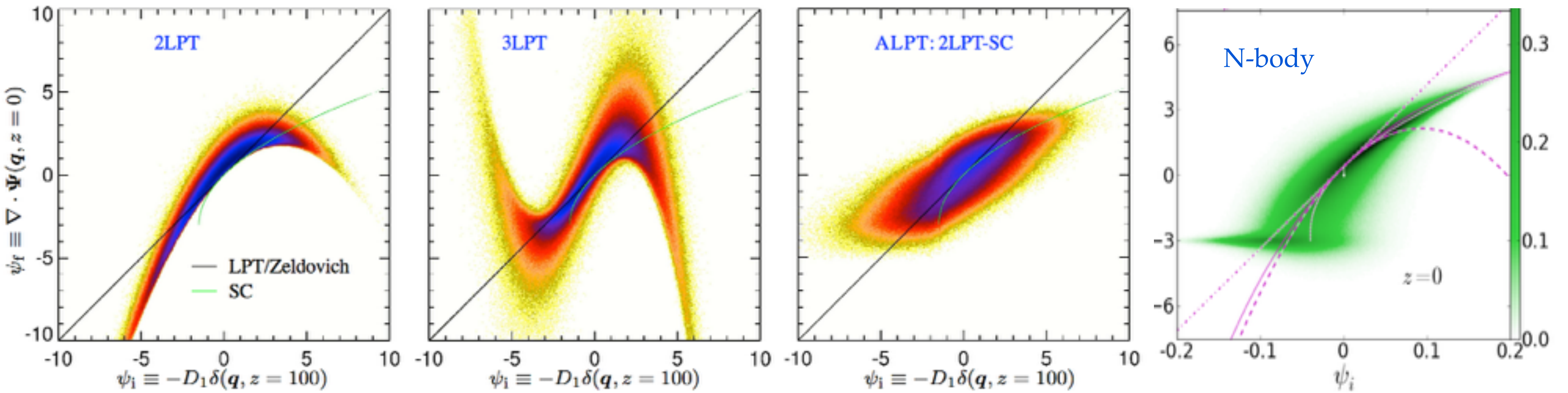
(low N-body resolution or perturbation theory based method)

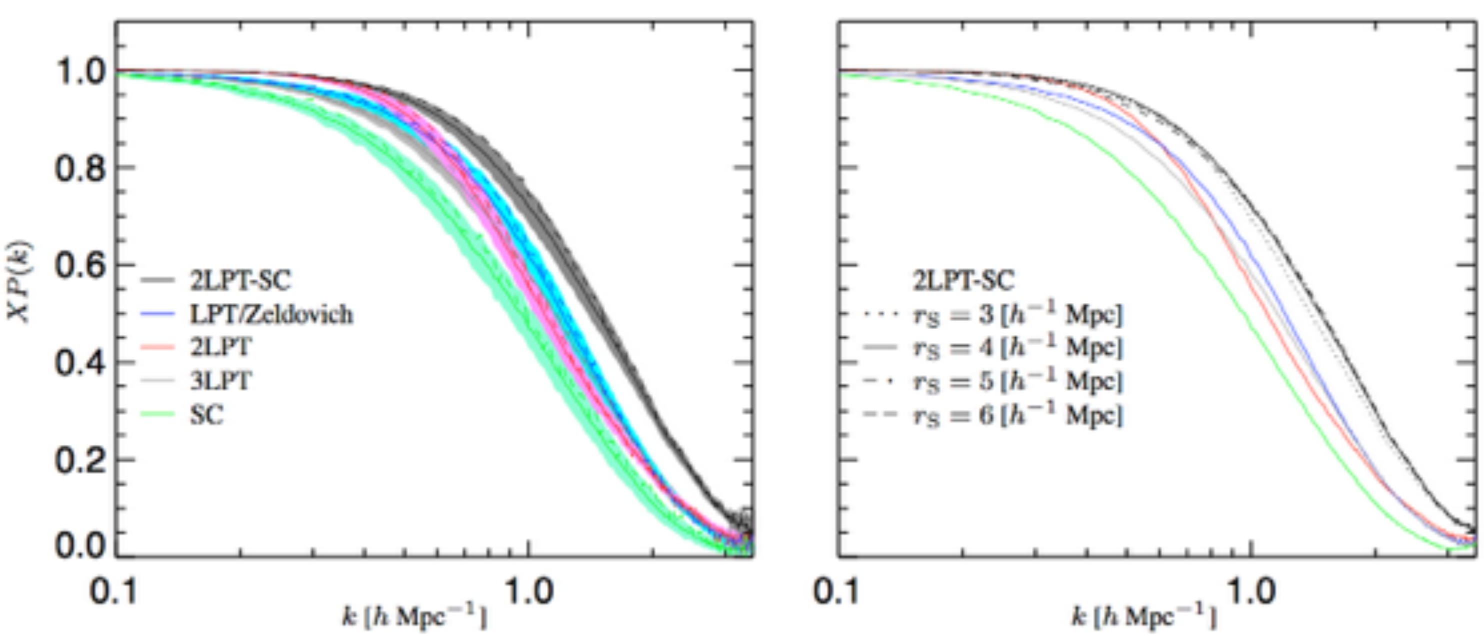
- The biasing model accurately connects the dark matter phase- space distribution with the halo distribution.

Simple efficient accurate one-step gravity solver…

ALPT: Augmented Lagrangian Perturbation Theory

Kitaura F. S. & Heß S. 2013, MNRAS, 435, L78

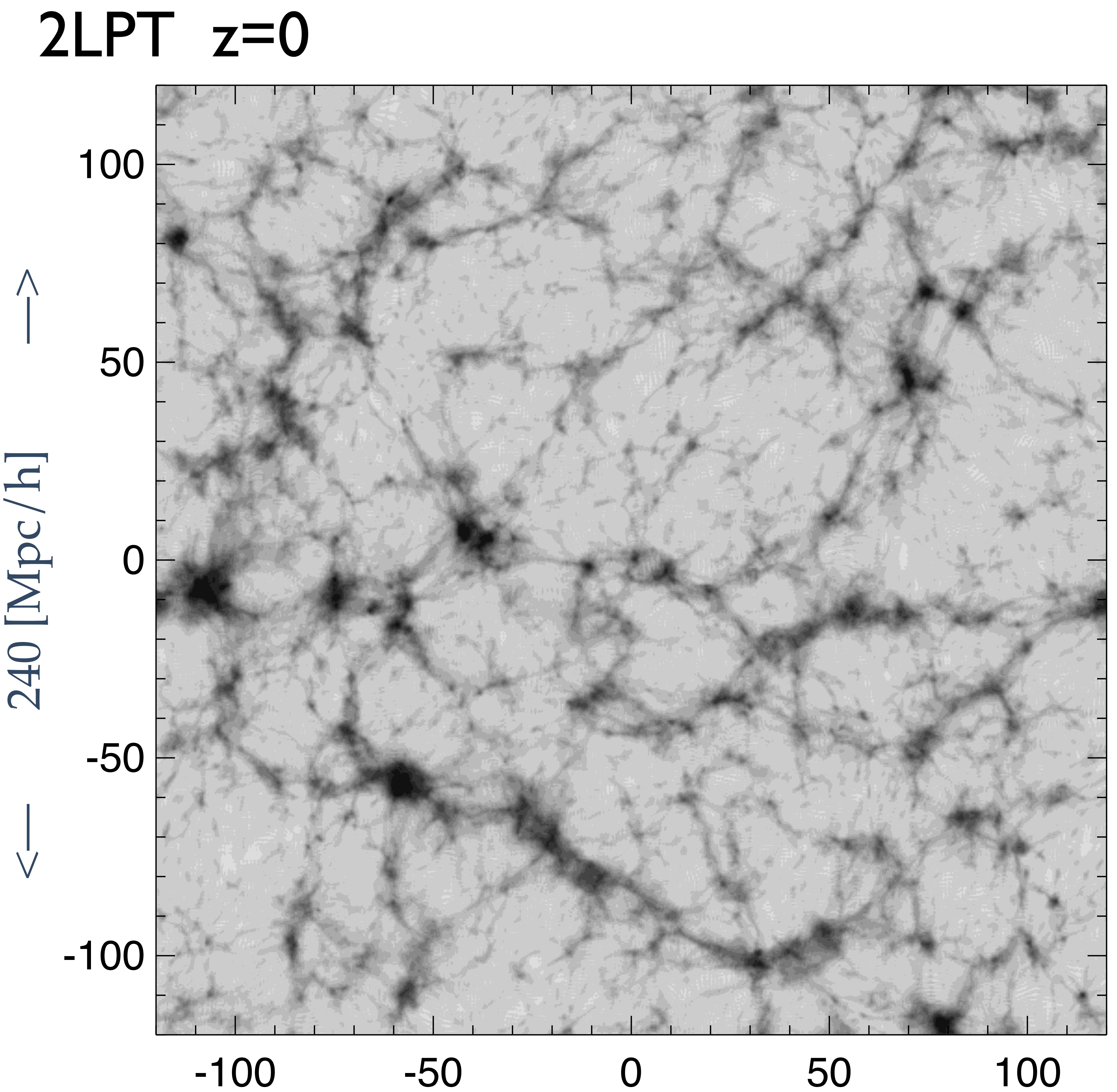




Top: relation between displacement field divergence vs ICs density field
Left: normalised cross-power spectra (propagator) at z=0 for different approximations vs N-body

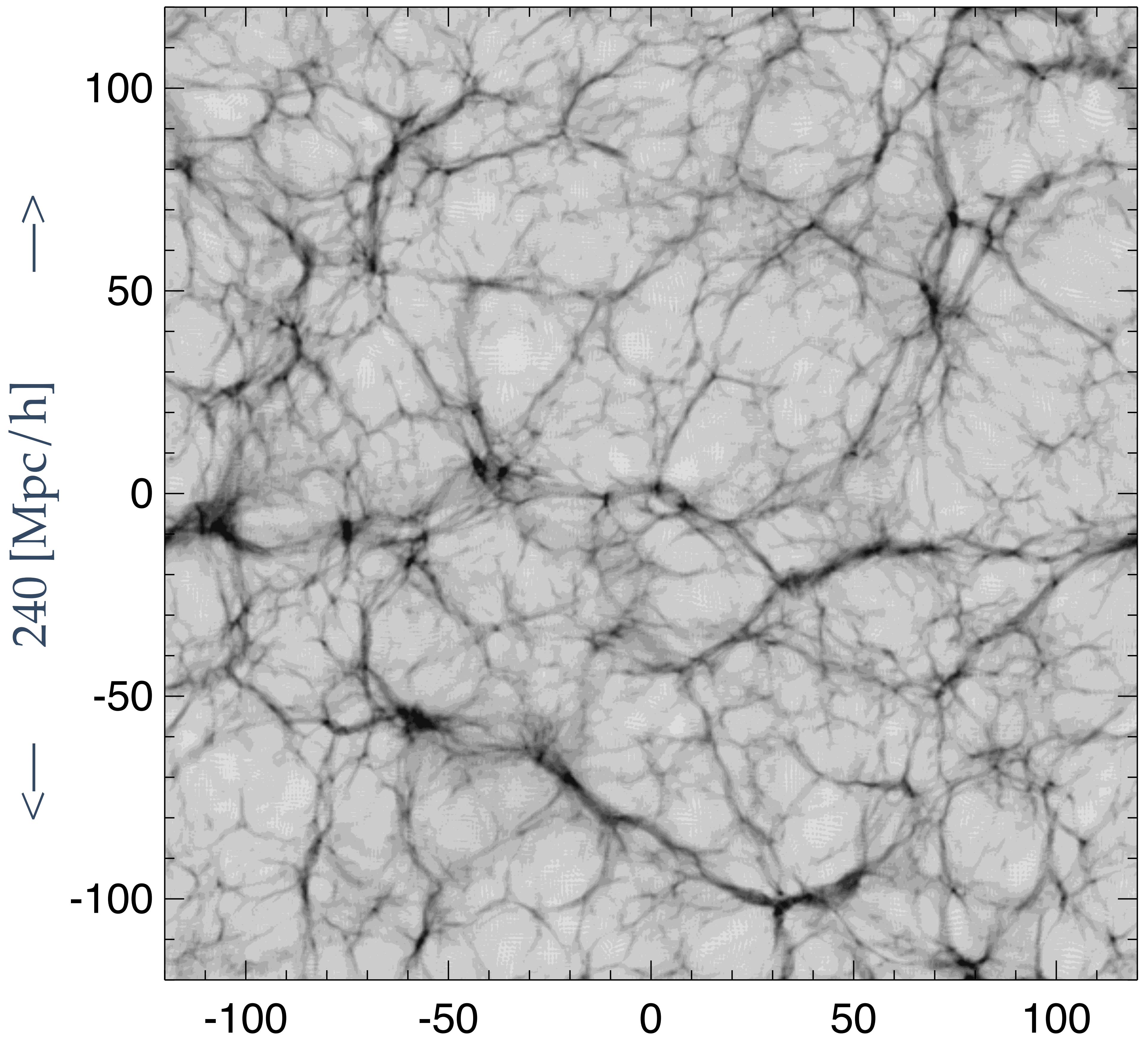
for N-body relation see Neyrinck M. C., 2013, MNRAS, 428, 141
for the spherical collapse relation see also Bernardeau F. 1994; Mohayee R. 2006

see also Tassev S. & Zaldarriaga M., 2012, JCAP, 4, 13 for other LPT improvements with transfer functions (less correlated with the N-body solution than ALPT)



2LPT z=0
240 [Mpc/h]
100
50
0
-50
-100
-100
-50
0
50
100

ALPT z=0



Calibration with N-body simulations for CMASS LRG type galaxies

we are performing mocks for BOSS, 4MOST, JPAS, EUCLID,…

Reference N-body simulation (Gadget): BIGMULTIDARK

Volume: (2500 Mpc/h)^3

Hess et al in prep

Prada et al in prep

Number of particles: 3840^3 (2M cpu hs)

halos selected with bdm (density peaks) according to vmax>~350 km/s (LRGs)

-> I consider 8 sub-volumes of (1250 Mpc/h)^3

Simulations with PATCHY: PerturbAtion Theory Catalog generator of Halo/galaxY distributions

Kitaura, Yepes & Prada 2014, MNRAS

Volume: (1250 Mpc/h)^3

Grid number of cells: 512^3

53 times lower resolution required!

Resolution of the grid: (2.4 Mpc/h)^3

same cosmology (Planck-like)

same redshift: z=0.577

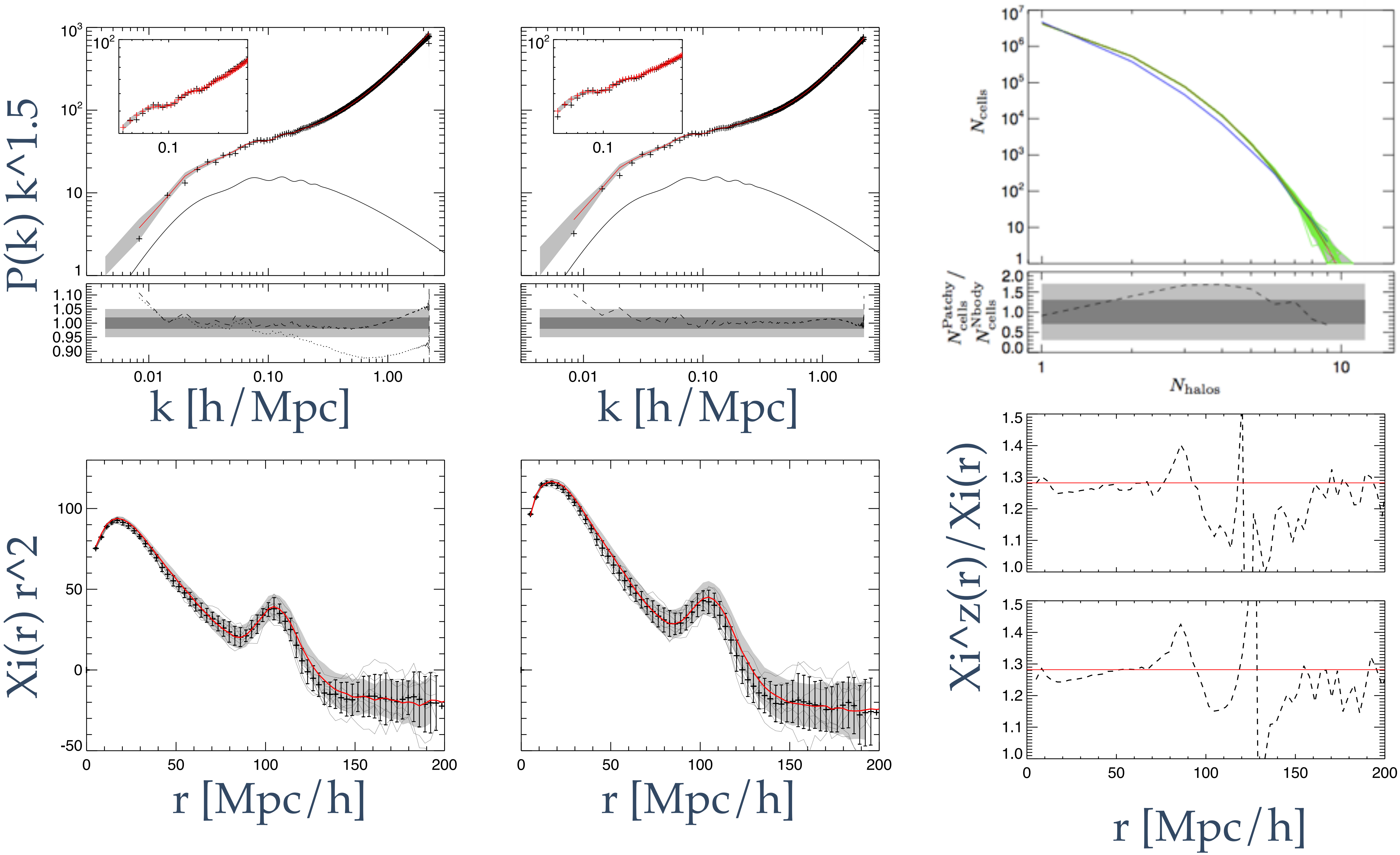
on my laptop (quad core i7, 4 cpus+4 virtual cpus, 8 Gb RAM)

about 15 mins

Calibration with N-body simulations

Kitaura, Yepes & Prada 2014, MNRAS

real-space redshift-space

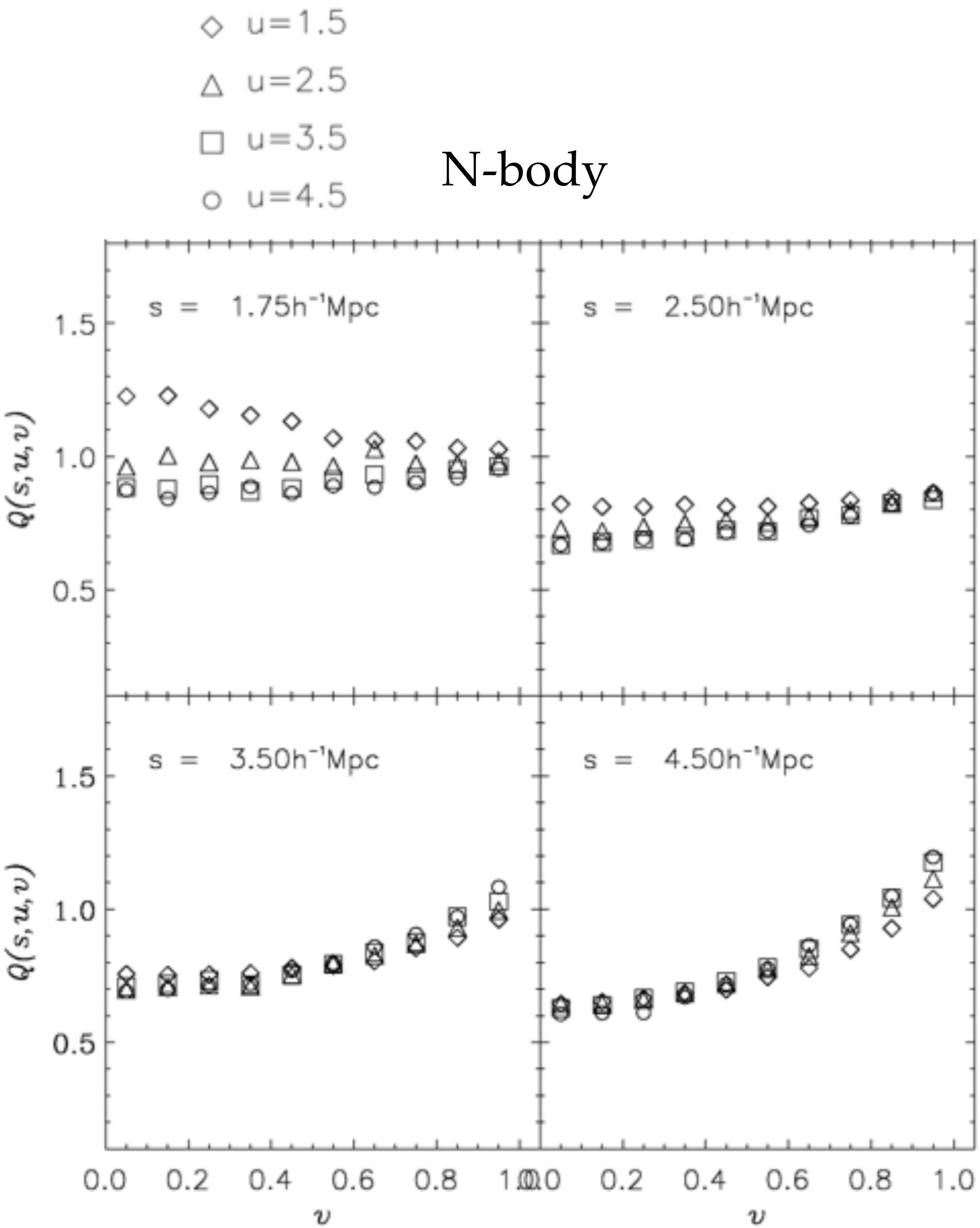


Three-point function

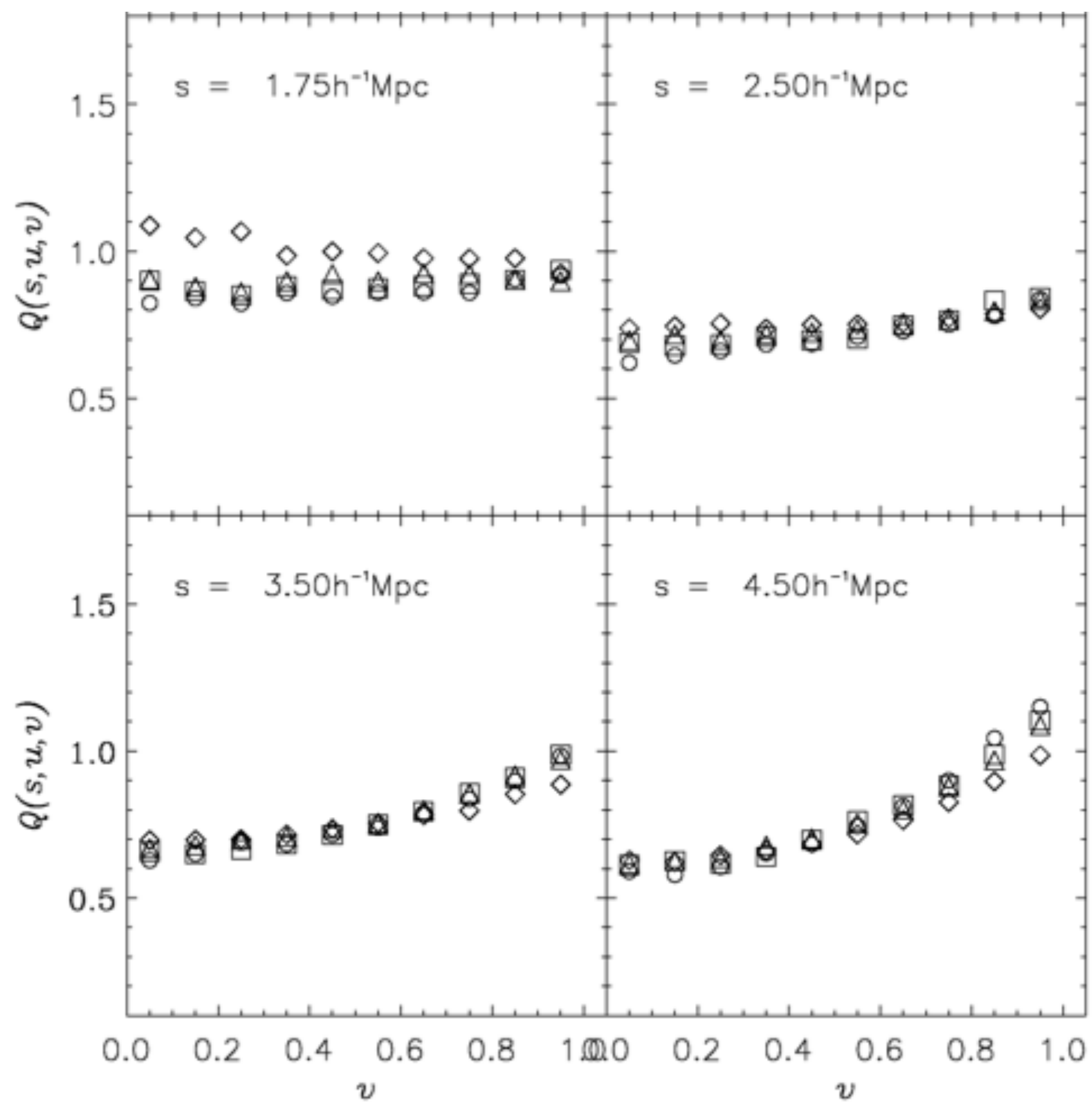
calculations by Volker Müller

Triangle sides: s, su, and s(u+v)

The 3-point function is described by the hierarchical ansatz
Q(s,u,v) = (s,u,v)/((r12) (r23)+c.c.),
i.e. the increase from v=0 to 1 means transition to linear structures.

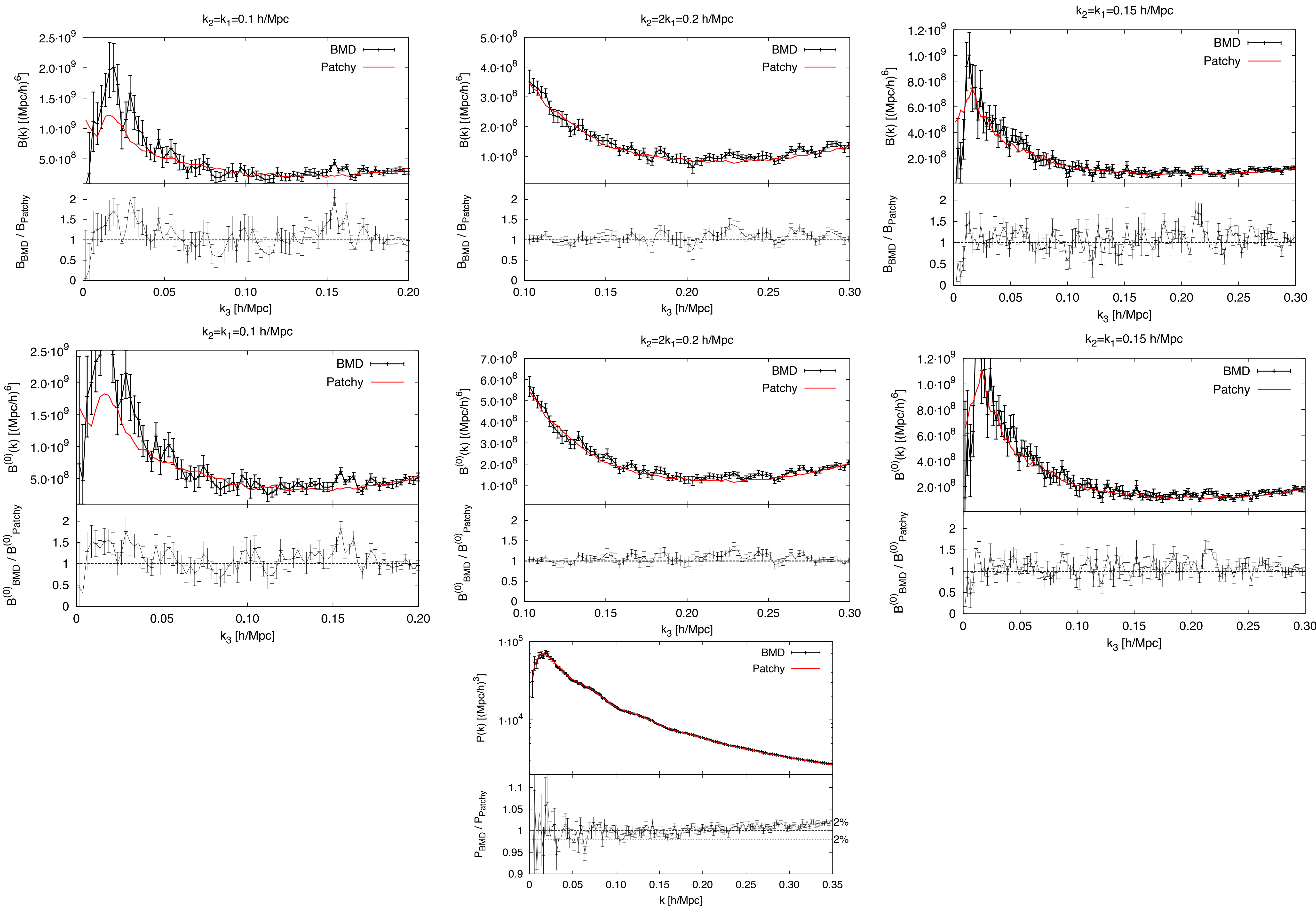


ALPT in PATCHY



calculations by Hector Gil-Marin

Bispectrum



What parameters are found for LRGs?

$$\rho_h = f_h\, \theta(\rho_{\rm M} - \rho_{\rm th})\, \rho_{\rm M}^{\alpha} \exp\left[-\left(\frac{\rho_{\rm M}}{\rho_\epsilon}\right)^{\epsilon}\right]$$

Why do LRGs have a constant linear bias of about 2?

mainly because they reside in the high density peaks!
rho_th and not alpha!

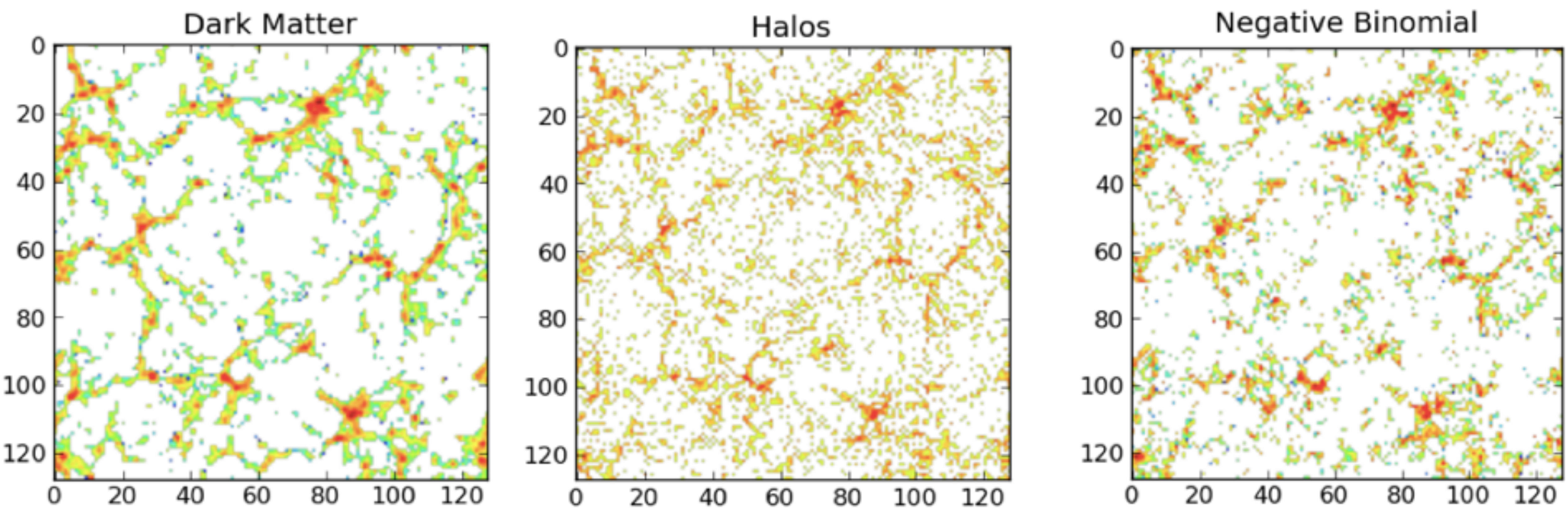
Can we apply such a stochastic nonlinear, scale-dependent biasing model for density reconstructions?

Bayesian density reconstruction from biased galaxy data (e.g. eLGs)

$$\rho_h = f_h\,\theta(\rho_{\rm M} - \rho_{\rm th})\,\rho_{\rm M}^{\alpha}\,\exp\left[-\left(\frac{\rho_{\rm M}}{\rho_{\epsilon}}\right)^{\epsilon}\right]$$

Ata, Kitaura & Müller in prep

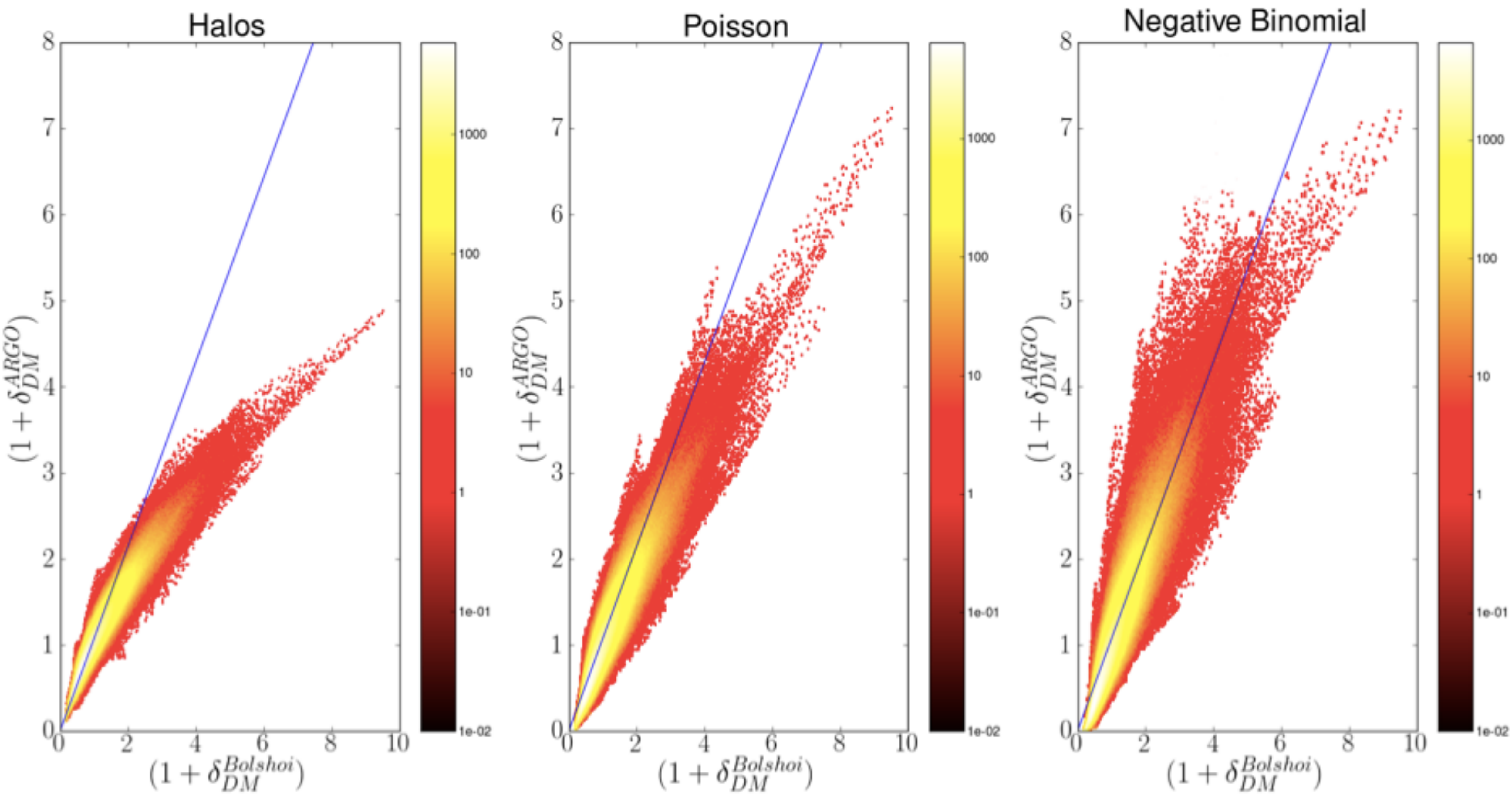
$$P(N_i \mid \lambda_i, \beta) = \frac{\lambda_i^{N_i}}{N_i!}\,\frac{\Gamma(\beta + N_i)}{\Gamma(\beta)(\beta+\lambda)^{N_i}}\,\frac{1}{(1+\lambda/\beta)^{\beta}}$$



Bayesian approach with the ARGO-code

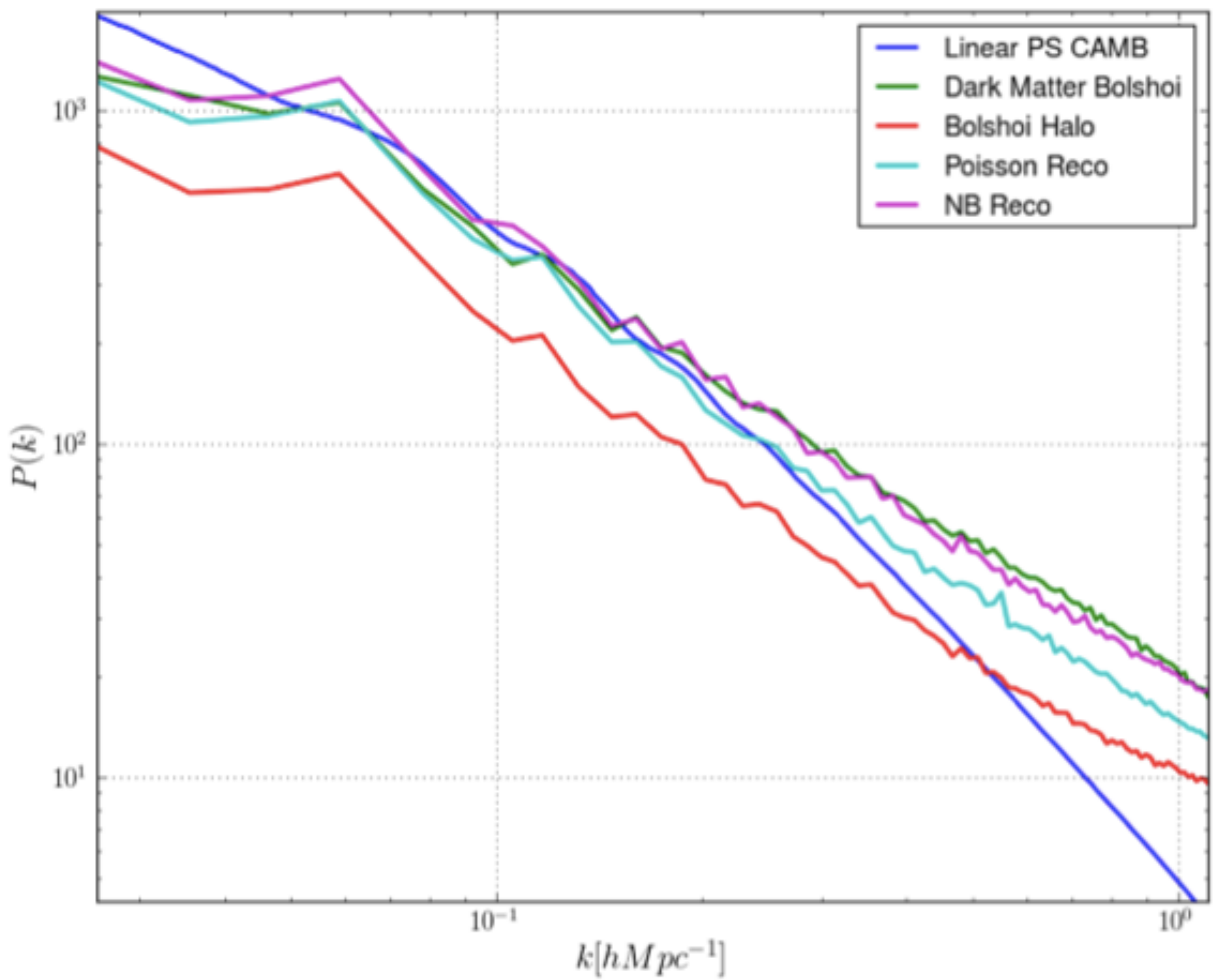
Kitaura & Enßlin 2008; Kitaura et al 2010; Kitaura et al 2012

Bayesian density reconstruction from biased galaxy data (e.g. eLGs)



Ata, Kitaura & Müller in prep

Bayesian density reconstruction from biased galaxy data (e.g. eLGs)



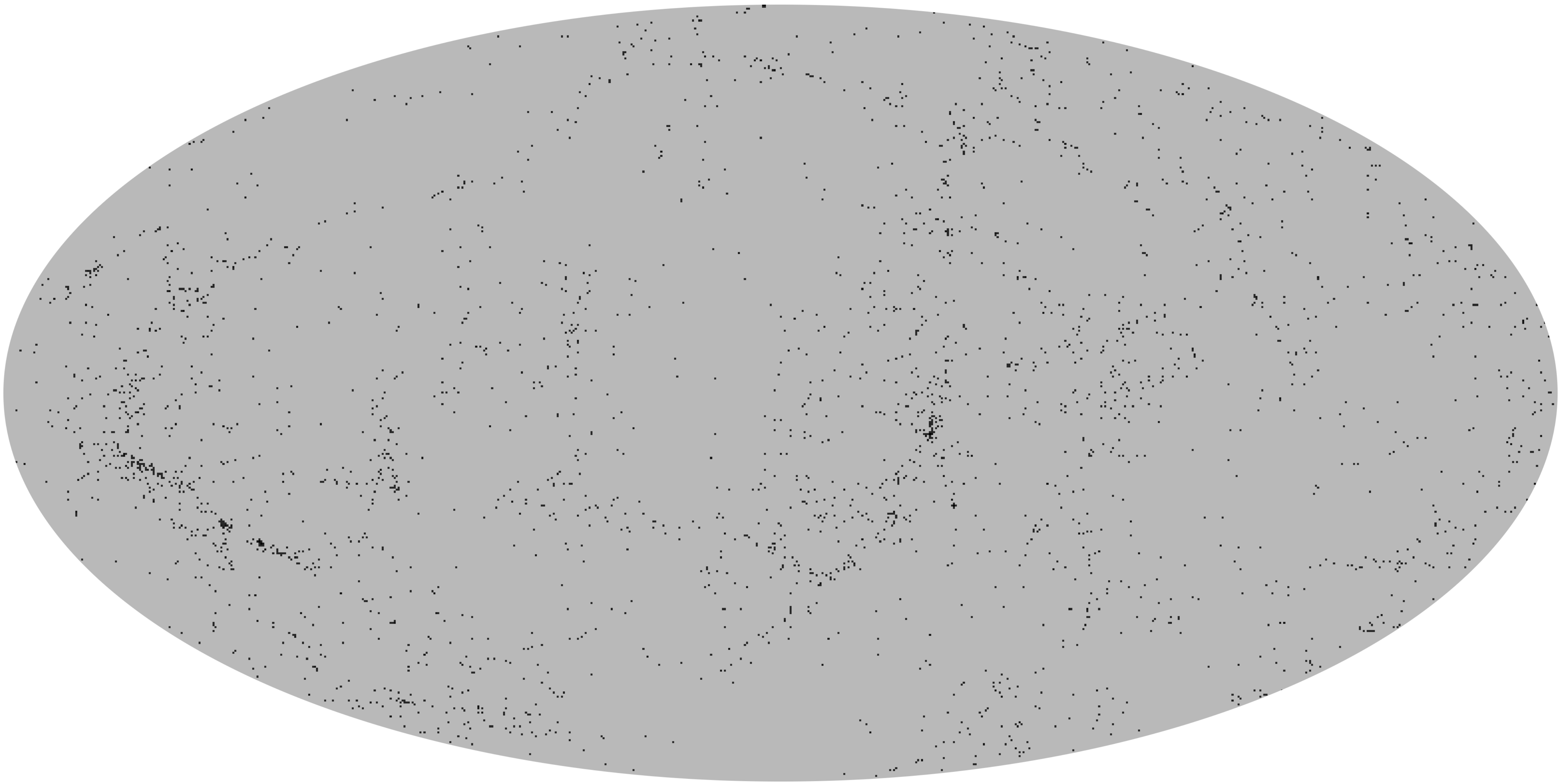
Ata, Kitaura & Müller in prep

Conclusions

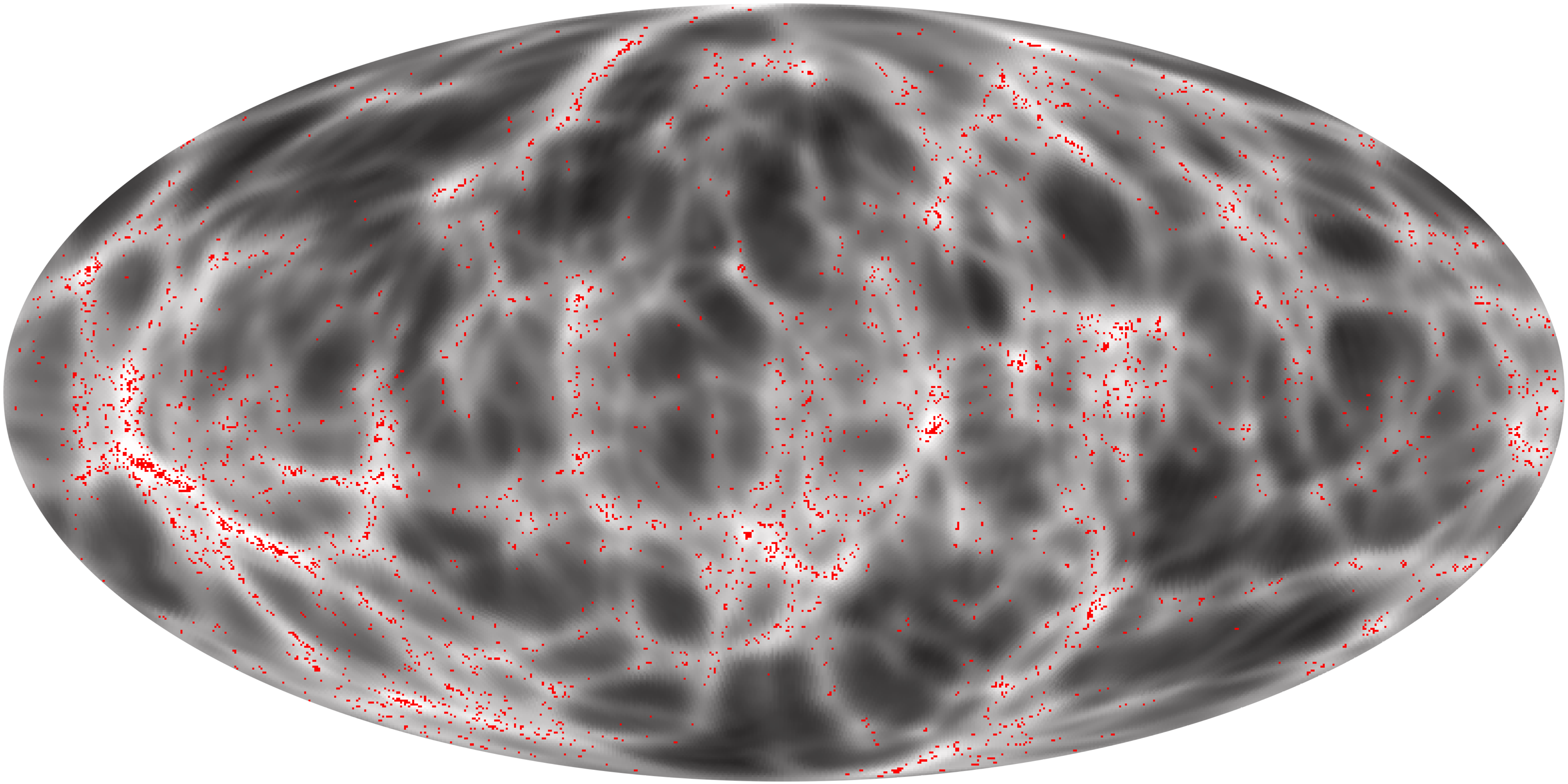
- Mock galaxy catalogs are necessary and important to test our theoretical models, estimate systematic errors/covariances, cosmic variance, etc
- Detailed state-of-the-art simulations are required (full gravity solvers, hydrodynamical simulations, feedback, etc.) to provide reference samples.
- We need thousands of mocks. ⟶ This cannot be done with full solvers!
- Approximate methods are required!
- There is a variety of methods: PThalos, PINOCCHIO, PATCHY, QPM,COLA. This is good!

Regarding PATCHY:

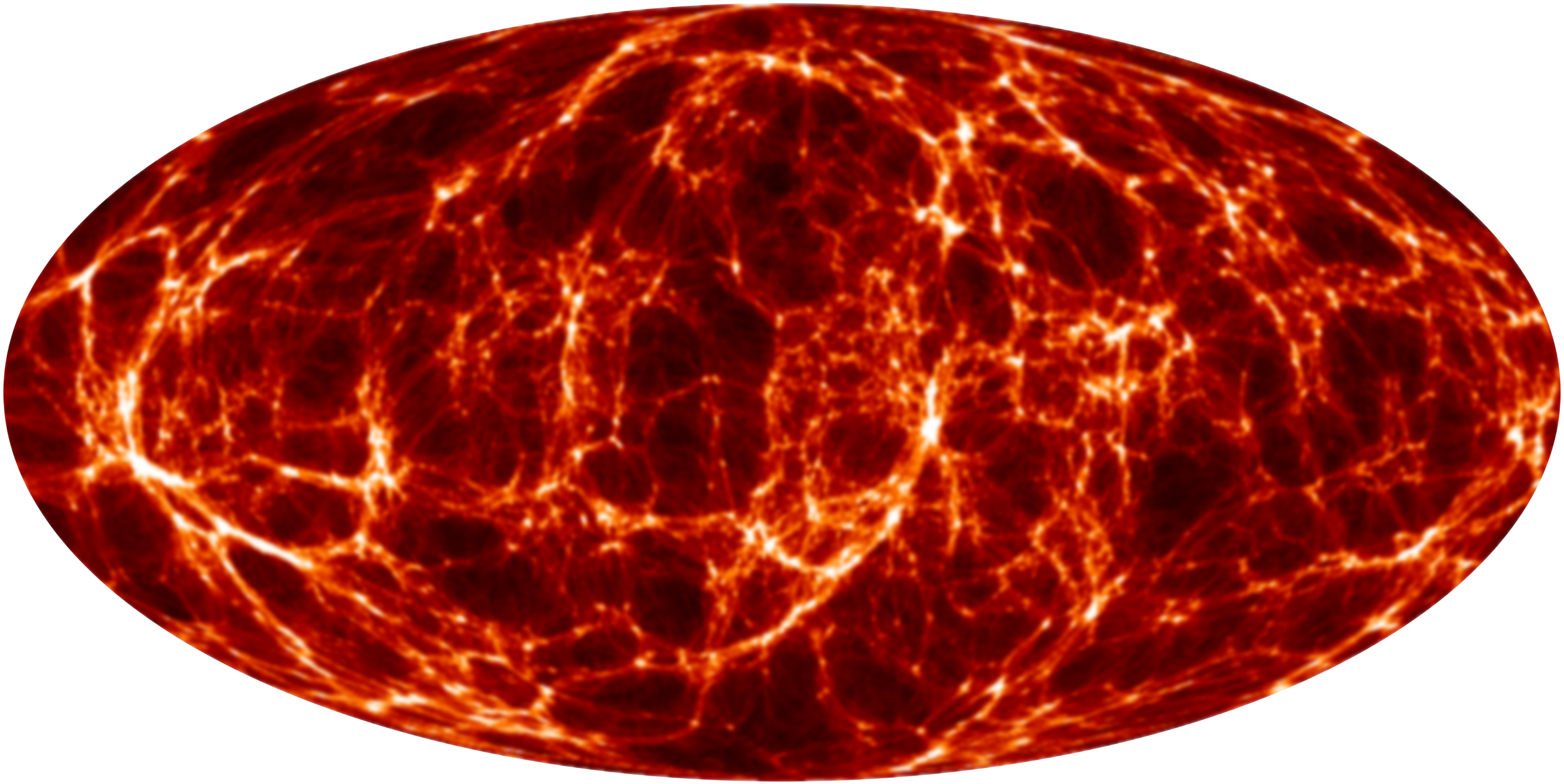
- We can efficiently model structure formation with Augmented Lagrangian Perturbation Theory ⟶ saves computational time
- We can use an effective statistical biasing description (nonlinear, stochastic, scale-dependent) ⟶ saves memory requirements (factor about 50)
- This statistical description can be used for Bayesian inference!
- We have done more than 1000 mocks for BOSS (Scoccola et al in prep)
- PATCHY caveats: PDF still not perfectly fit; needs to be calibrated for each halo population from N-body simulations; not predictive!



real observations!



mean over an ensemble of ALPT reconstruction!



high resolution constrained N-body simulation!

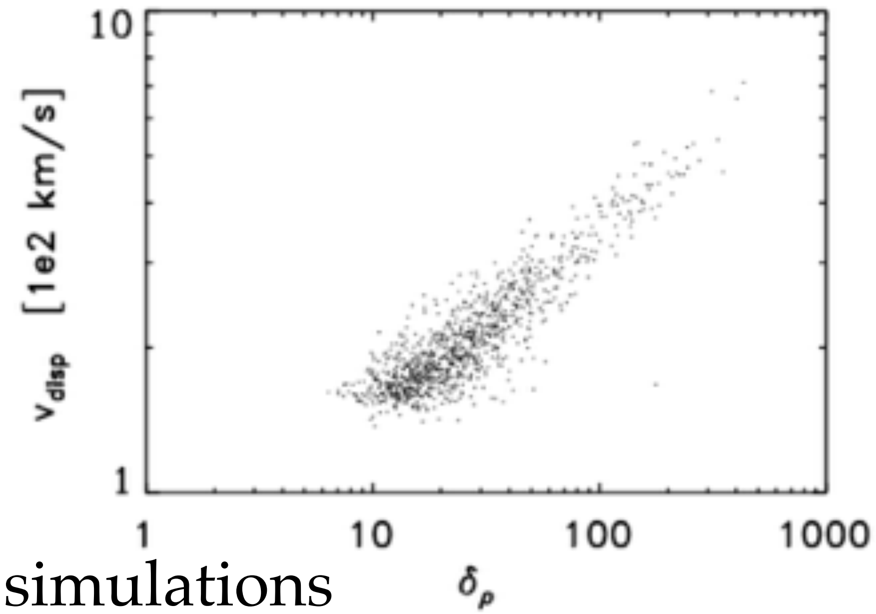
Modelling redshift space distortions

Kitaura F. S. PhD 2007 from MR sim (Hess & Kitaura in prep)

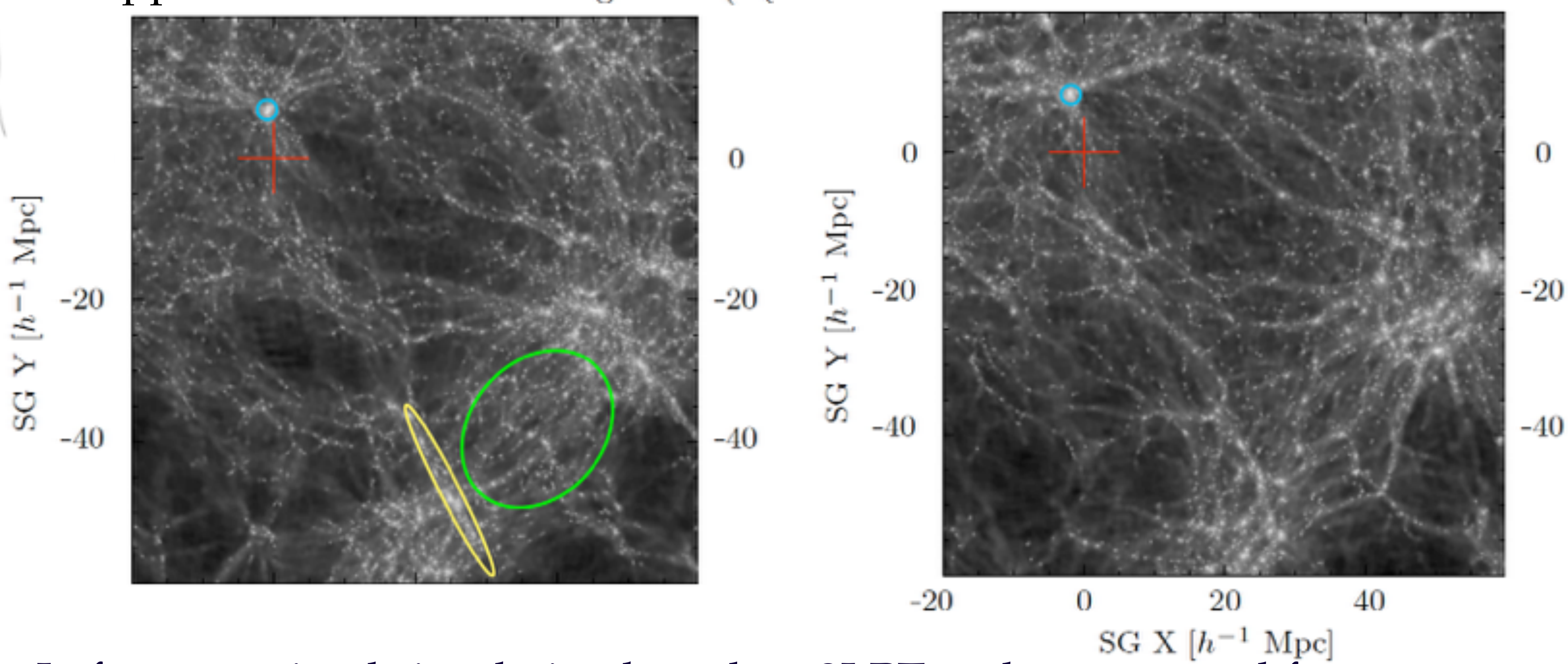
Kitaura, Yepes & Prada 2014, MNRAS

$$\boldsymbol{v} = \boldsymbol{v}^{\rm coh} + \boldsymbol{v}^{\sigma}$$

$$\boldsymbol{v}_r^{\sigma} \equiv (\boldsymbol{v}^{\sigma} \cdot \hat{\boldsymbol{r}})\hat{\boldsymbol{r}}/(Ha) = \mathcal{G}\left(g \times \left(1 + b^{\rm ALPT}\delta^{\rm ALPT}(\boldsymbol{x})\right)^{\gamma}\right)\hat{\boldsymbol{r}}$$



Application of KIGEN to 2MRS: constrained simulations



Left: constrained simulation based on 2LPT and compressed fogs

Right: constrained simulation based on ALPT modelling fogs

Hess Steffen, Kitaura F. S. & Gottlöber S., 2013, MNRAS, 435, 2065