Sufficient Statistics

István Szapudi

Institute for Astronomy
University of Hawaii

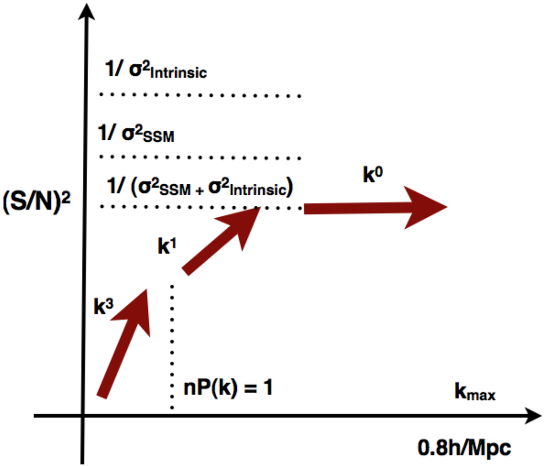
April 11, 2019

Outline

Cosmological Information in LSS Surveys

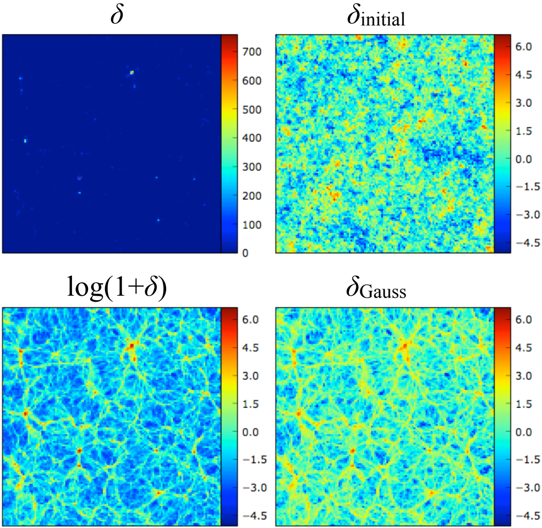
- The standard power spectrum extract a small fraction of the available information from LSS surveys, (unlike CMB)
- Non-Gaussianity, linear modes are used $k \simeq 0.2$
- Good news: many ($\propto k^3$) high k modes are available
- Bad news: plateau in the power spectrum due to SS and IS mode coupling
- Standard idea: use higher order statistics (weakly non-linear)
- Worst news: information content of higher order statistics vanishing in the non-linear regime

Summary of plateaux



Logarithmic mapping

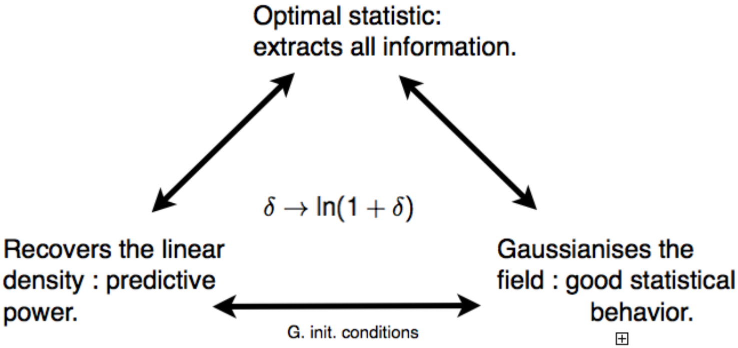
Neyrinck, Szapudi, & Szalay 2009



Sufficient Statistics: All Information on a Parameter

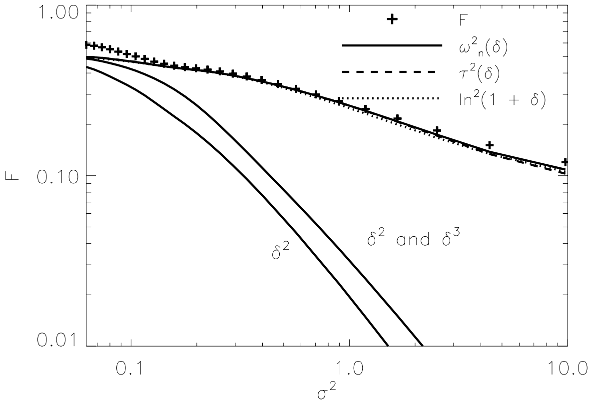
Carron & Szapudi (2013 MNRAS 434, 2961; 2014, MNRAS 439, L11)

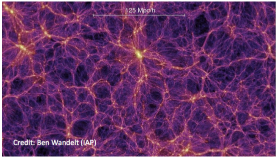
$$\partial_\alpha \ln p(\delta) \simeq \tau^2(\delta) \simeq \left(\frac{(1+\delta)^{(n+1)/3} - 1}{(n+1)/3} \right)^2 \simeq \ln(1+\delta)^2$$



Diagonal covariances matrix.

Info. in the Millenium simulation density field





Analogus to lognormal fields.

Discreteness effects

So we have now a good understanding at the level of the δ field. How to apply these transformation to the observed discrete galaxy field ?

- Now two layers of non-Gaussian statistics : underlying matter field, and discrete sampling effects.

$$P(N|\boldsymbol{\theta}) = \int_0^\infty d\rho \; \underbrace{p(\rho|\boldsymbol{\theta}) P(N|\rho)}_{\propto P(\rho|N)} .$$

- Saddle-point (Laplace) approximation (ρ^* maximum of posterior for ρ) :

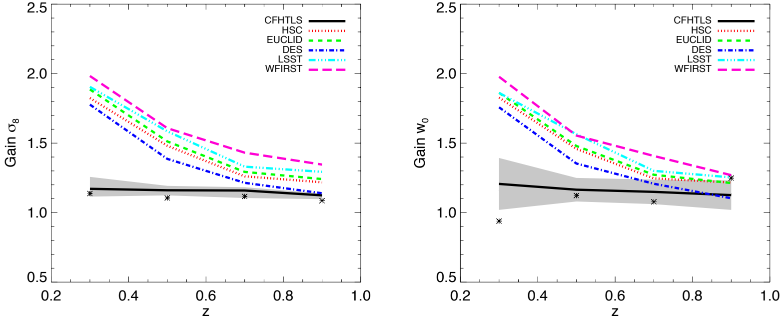
$$\partial_\alpha \ln P(N|\boldsymbol{\theta}) = \underbrace{\partial_\alpha \ln p(\rho^*|\boldsymbol{\theta})}_{\text{original suff. observable}} - \underbrace{\frac{1}{2}\partial_\alpha \ln'' P(\rho^*|N, \boldsymbol{\theta})}_{\text{sensitivity of curv. of posterior}} .$$

- For Poisson sampling it all boils down to extract the mean and variance of $A^* = \ln \rho^*$

Carron & Szapudi (2014), MNRAS 439, L11

Forecasting dark energy EoS:

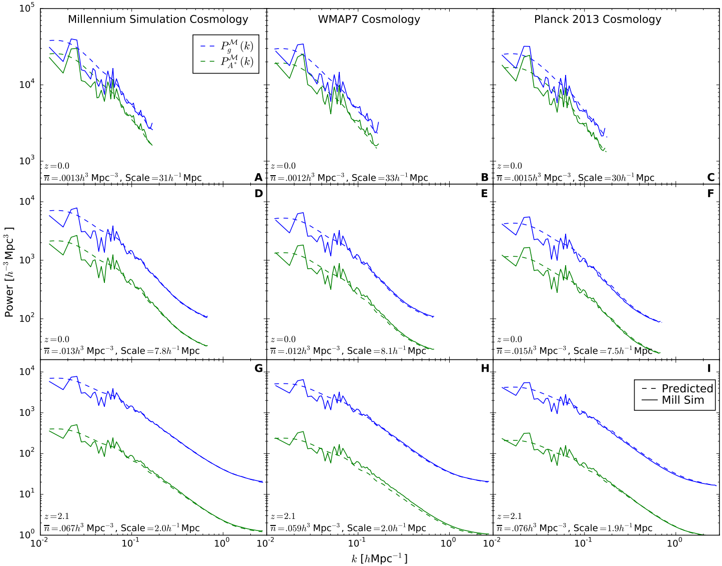
Lognormal is a great approximation in 2D



Survey	Max. Gain
Euclid	1.86
WFIRST-AFTA	1.98
HSC	1.83
LSST	1.86
DES	1.76

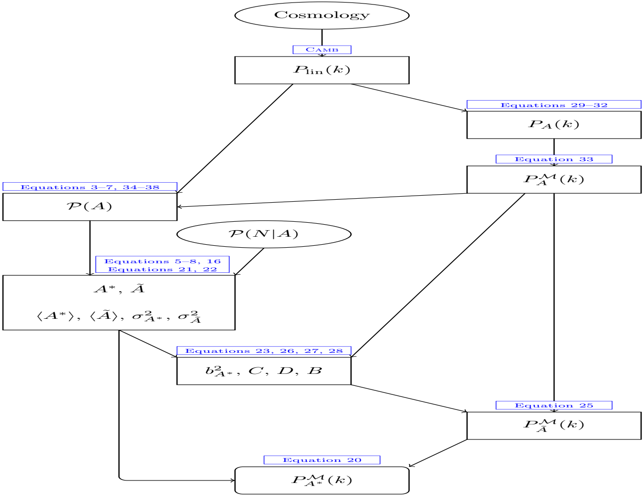
Prediction of the A^* power spectrum

Repp & Szapudi 2018b, 2019



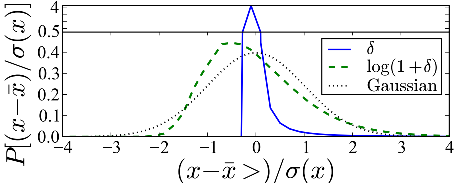
Prediction of the A^* power spectrum

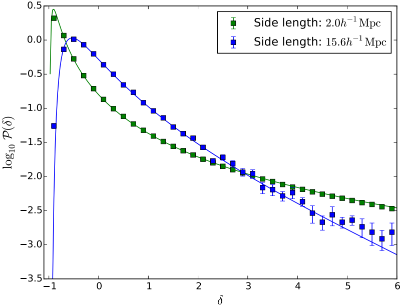
Repp & Szapudi 2018b, 2019



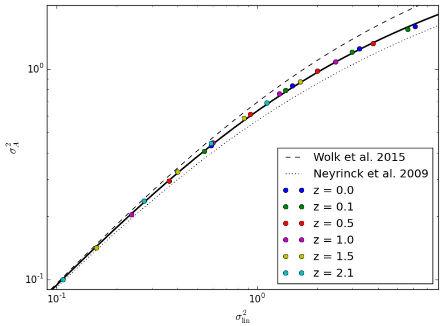
Non-lognormality in 3D

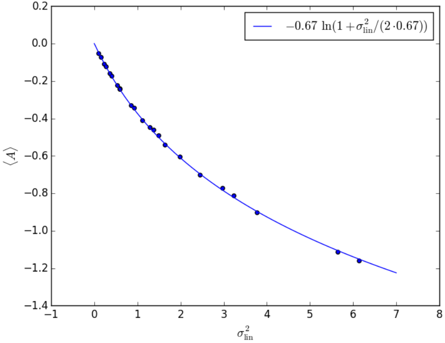
GEV distribution: Repp & Szapudi 2018a





Fitting Formulae





Numerical model for P_A

Repp & Szapudi 2017, motivated by Szapudi and Kaiser 2003

- The shape of the $A = \log(1+\delta)$ power spectrum is approximately the same as the linear δ

$$P_A(k) = N\, C(k)\, \frac{\sigma_A^2}{\sigma_{\rm lin}^2} P_{\rm lin}(k),$$

  where

$$C(k) = \begin{cases} 1 & \text{if } k < 0.15 h\,\text{Mpc}^{-1} \\ (k/0.15)^\alpha & \text{if } k \geq 0.15 h\,\text{Mpc}^{-1} \end{cases},$$

  and

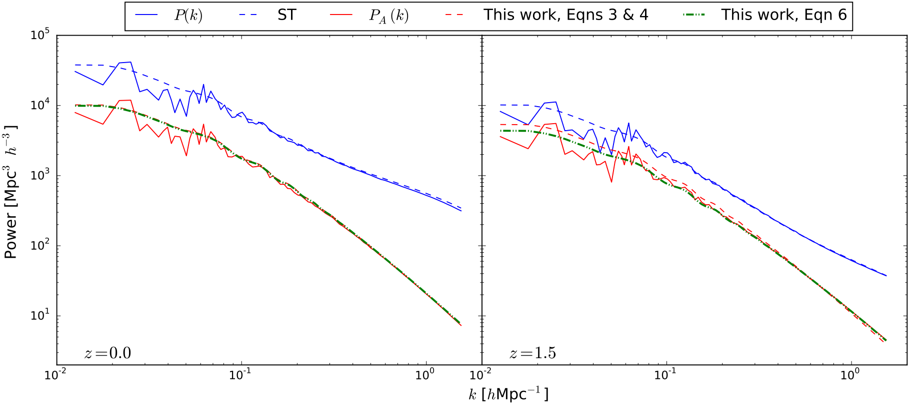
$$N = \frac{\int dk\, k^2 P_{\rm lin}(k)}{\int dk\, k^2 C(k) P_{\rm lin}(k)}, \quad \alpha(z) \simeq [0.02 - 0.14]$$

- where $\sigma_{\rm lin}^2$ is calculated with CAMB and

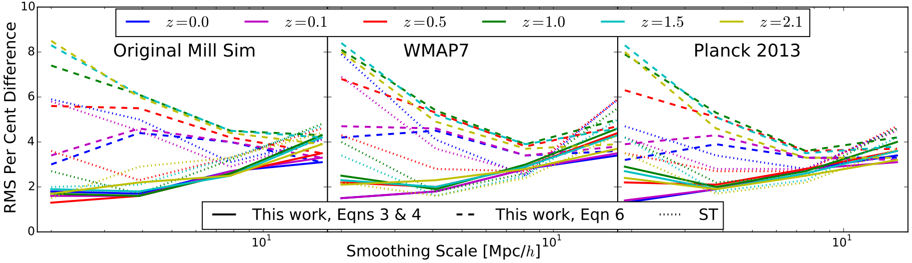
$$\sigma_A^2 = 0.73 \ln\left(1 + \frac{\sigma_{\rm lin}^2}{0.73}\right).$$

- checked for several cosmologies and redshifts between $0 \leq z \leq 2.1$

Precision Prediction for the Log Power Spectrum

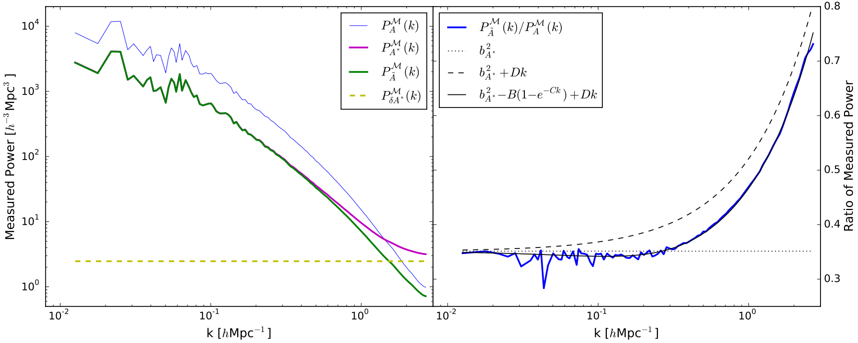


Prediction errors



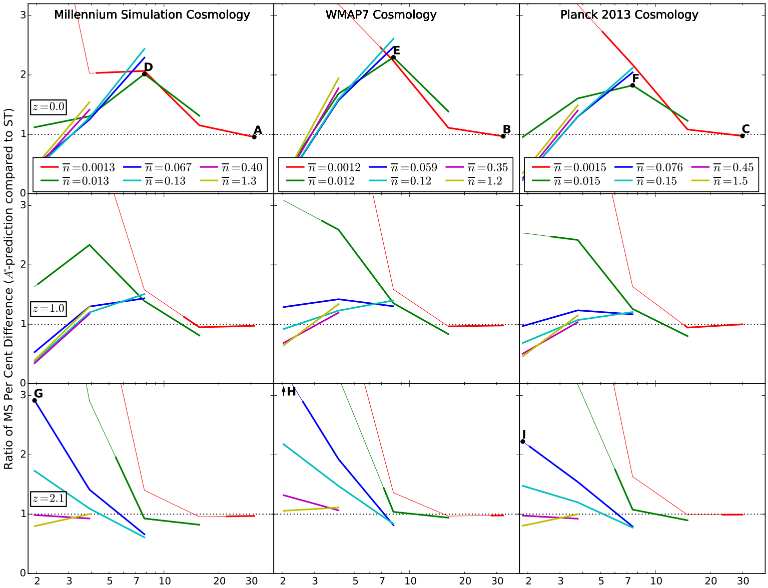
From $\tilde{A}$ to A^*

Repp & Szapudi 2019



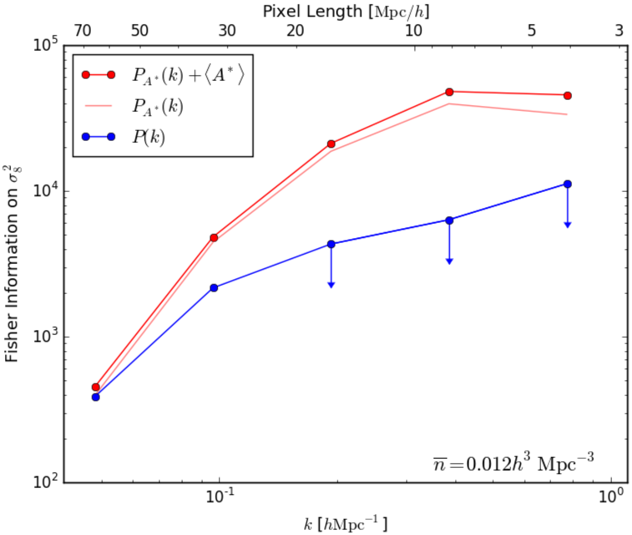
Accuracy

Repp & Szapudi 2019



Forcast for A^* in 3D (preliminary)

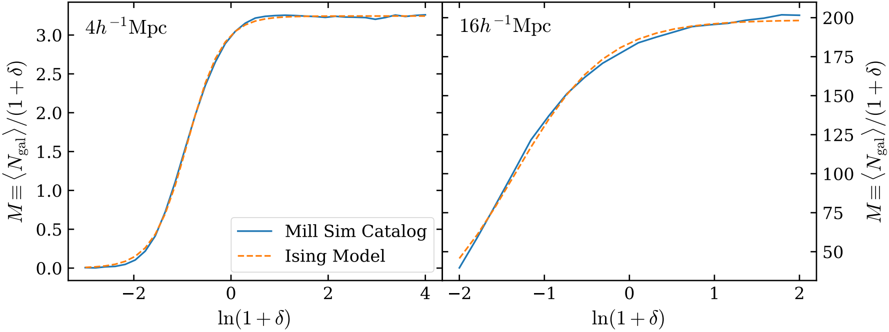
Realistic density but real space no bias



Bias in Simulations

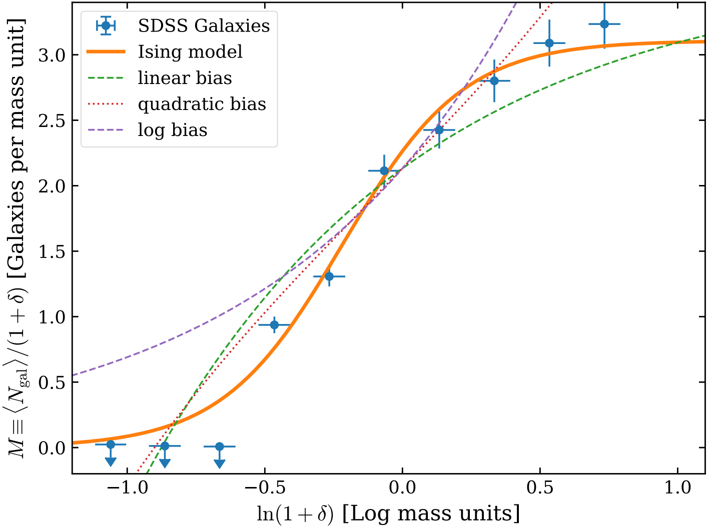
Repp & Szapudi 2019

$$M \equiv \langle N_{\rm gal} \rangle_A \cdot (1+\delta)^{-1} = \frac{b\overline{N}}{1 + \exp\left(\frac{A_t - A}{T}\right)}$$

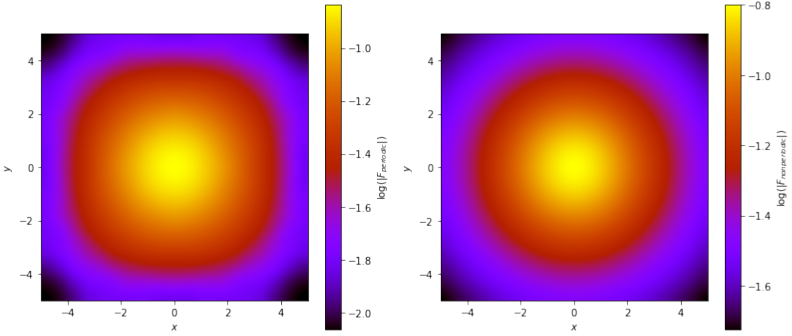


Bias in SDSS

Repp & Szapudi 2019



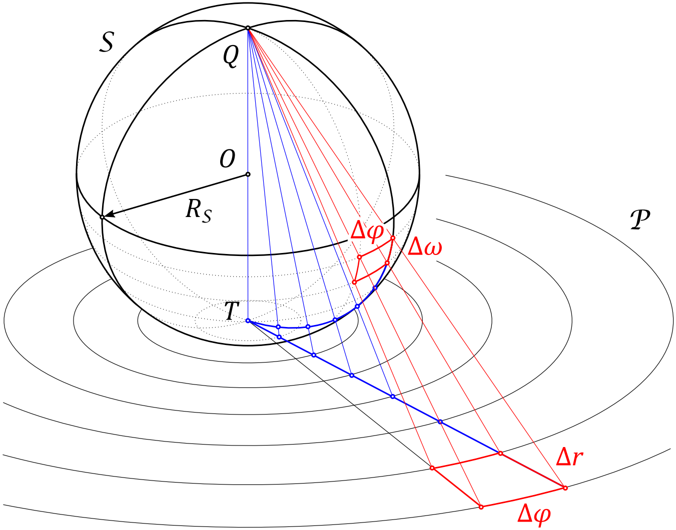
Problems with cosmological simulations



- Forces are wrong. (High Precision Cosmology?)
- Spherical symmetry broken
- T^3 Topology contrary to observations
- Number of modes $\simeq k^3$, too many high k
- Does not match observational geometry

Idea: compactify the infinite volume

Illustrated in 2D

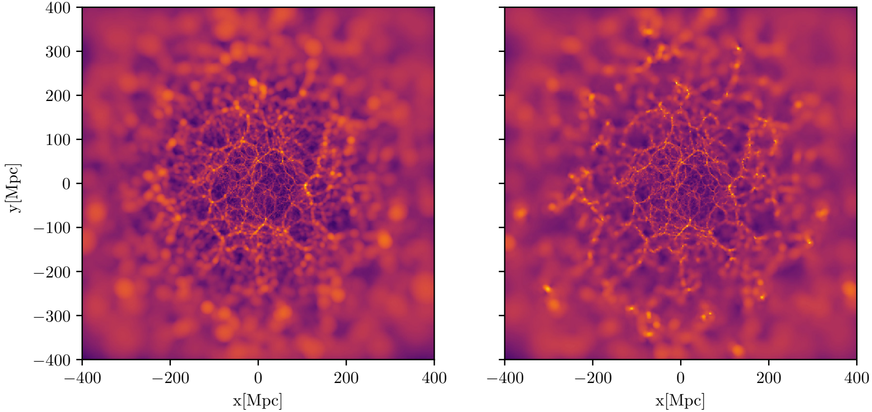


Millennium resimulations

Racz etal 2019

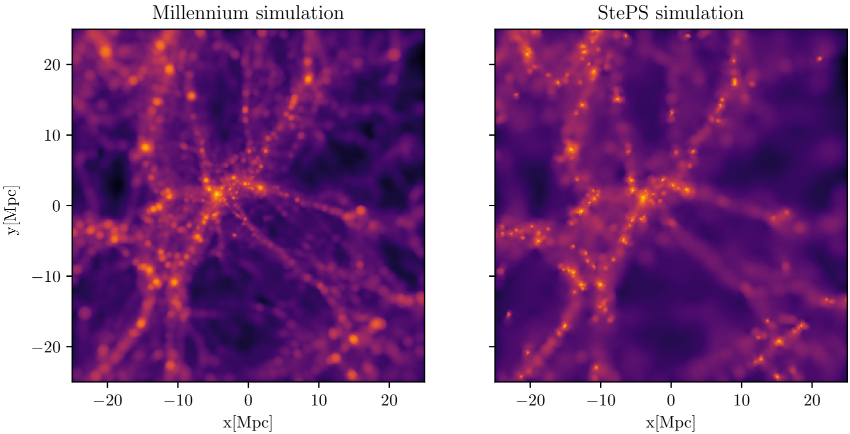
- Millennium Cosmology using the original scripts with StePS
- H_0 = 73km/s/Mpc, Ω_m = 0.25, Ω_Λ = 0.75, σ_8 = 0.9
- 27 volumes to simulate periodic B.C.
- $z_{in} = 127$
- Original: 512 CPU $\times$ 683 hours 0.321 Gpc^3
- StEPS: 12 GPU $\times$ 106 hours 1.35 Gpc^3
- $N \log N$ vs N^2
- multiresolution

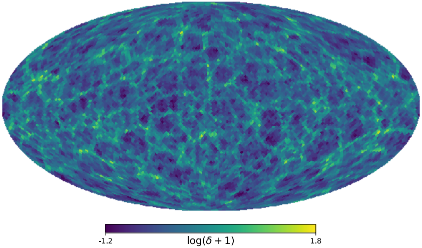
Results:Millennium

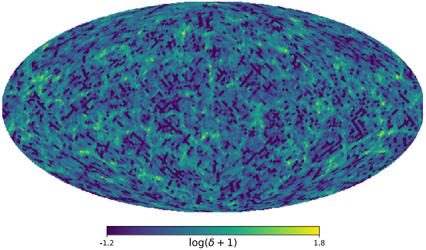


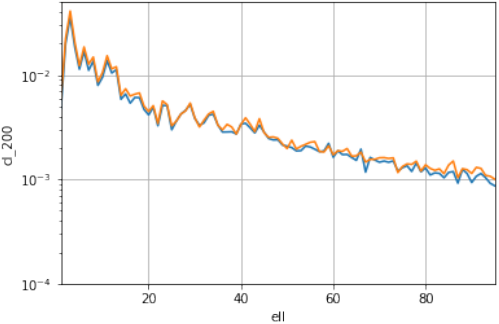
Results

Racz etal 2019, Millennium resimulation




-1.2
log(δ + 1)
1.8


-1.2
log(δ + 1)
1.8


d_200
10⁻²
10⁻³
10⁻⁴
20
40
60
80
ell

Summary

- Sufficient statistics are well approximated with the log transform
- They are well understood in Fisher information theory, and can be predicted and estimated from data
- The log and A^* power spectra are now predicted to percent level accuracy in real space
- The information gain for Euclid like surveys $\gtrsim 2\times$.
- Accurate bias fit inspired by the Ising model
- Compactified cosmological simulation have unprecedented dynamic range for given resources
- Status: tested on 48 GPUs and recovered Millennium as a test
- Fast prediction for parameter estimation
- Super-survey modes/covariance matrices/high dynamic range (e.g., local group, galaxy simulations in cosmological background)
- Next: redshift distortions

RSD (preiminary)