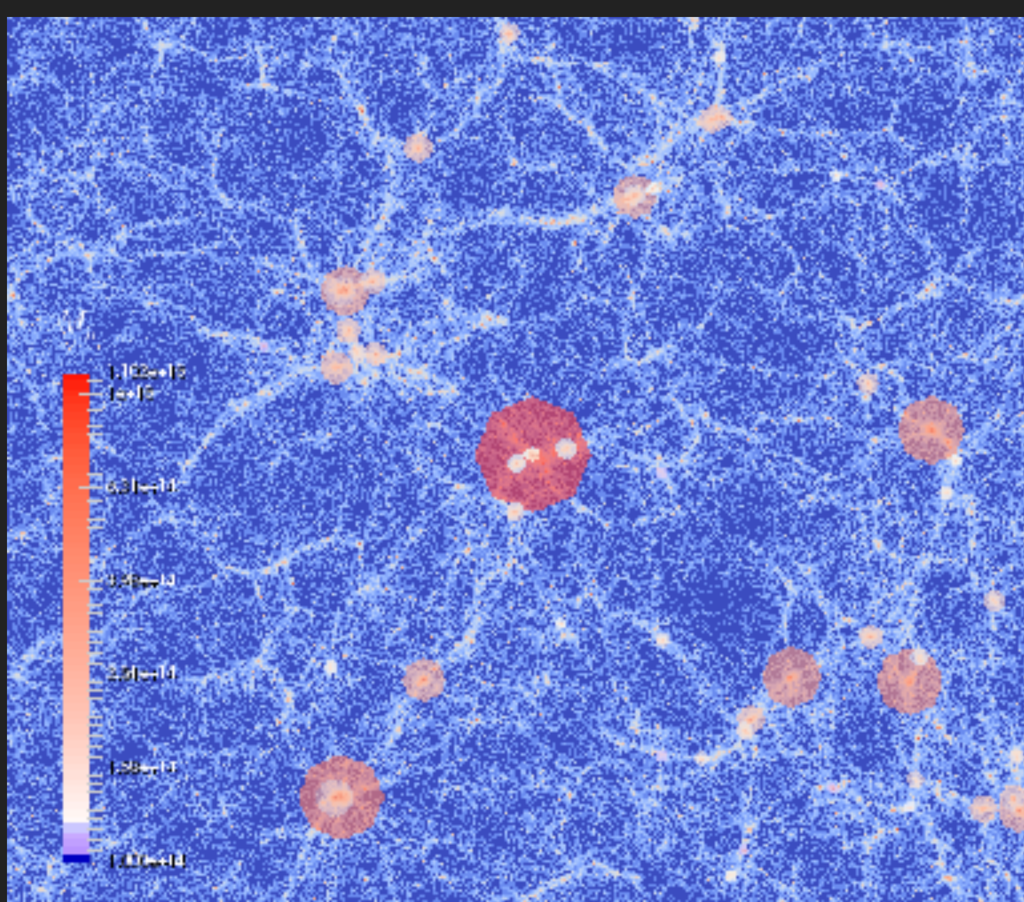
DARK EMULATOR

COSMIC LARGE-SCALE STRUCTURES AND PARAMETER ESTIMATION

TAKAHIRO NISHIMICHI (KAVLI IPMU, JST CREST)



- **Kavli IPMU**

Takahiro Nishimichi*
Masahiro Takada
Naoki Yoshida
Taira Oogi*

- **U. Tokyo**

Ken Osato*
Masamune Oguri

- **NAOJ**

Masato Shirasaki*
Takashi Hamana

- **Hirosaki U.**

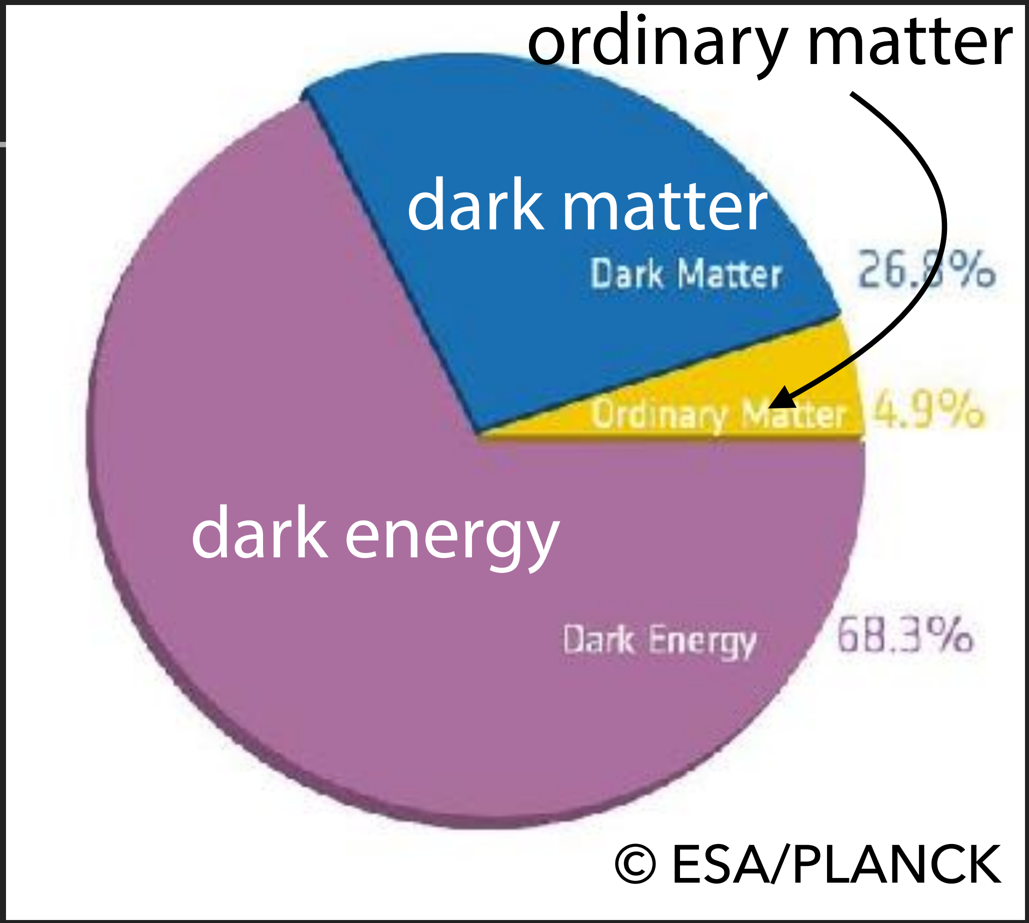
Ryuichi Takahashi*

2016/12/22; data assimilation seminar @ RIKEN

BACKGROUND

COSMOLOGY: CURRENT STATUS

- All sorts of observations can be explained by the standard model with only 6 free parameters
  - and they are very precisely determined
- *Dark components* play the major role in the current Universe



Parameter	68 % limits
$\Omega_b h^2$	0.02225 ± 0.00016
$\Omega_c h^2$	0.1198 ± 0.0015
$100\theta_{MC}$	1.04077 ± 0.00032
τ	0.079 ± 0.017
$\ln(10^{10} A_s)$	3.094 ± 0.034
n_s	0.9645 ± 0.0049
H_0	67.27 ± 0.66
Ω_Λ	0.6844 ± 0.0091
Age/Gyr	13.813 ± 0.026

Planck 2015 cosmological params.

baryon density

dark matter density

cosmic geometry (dark energy encoded here)

optical depth

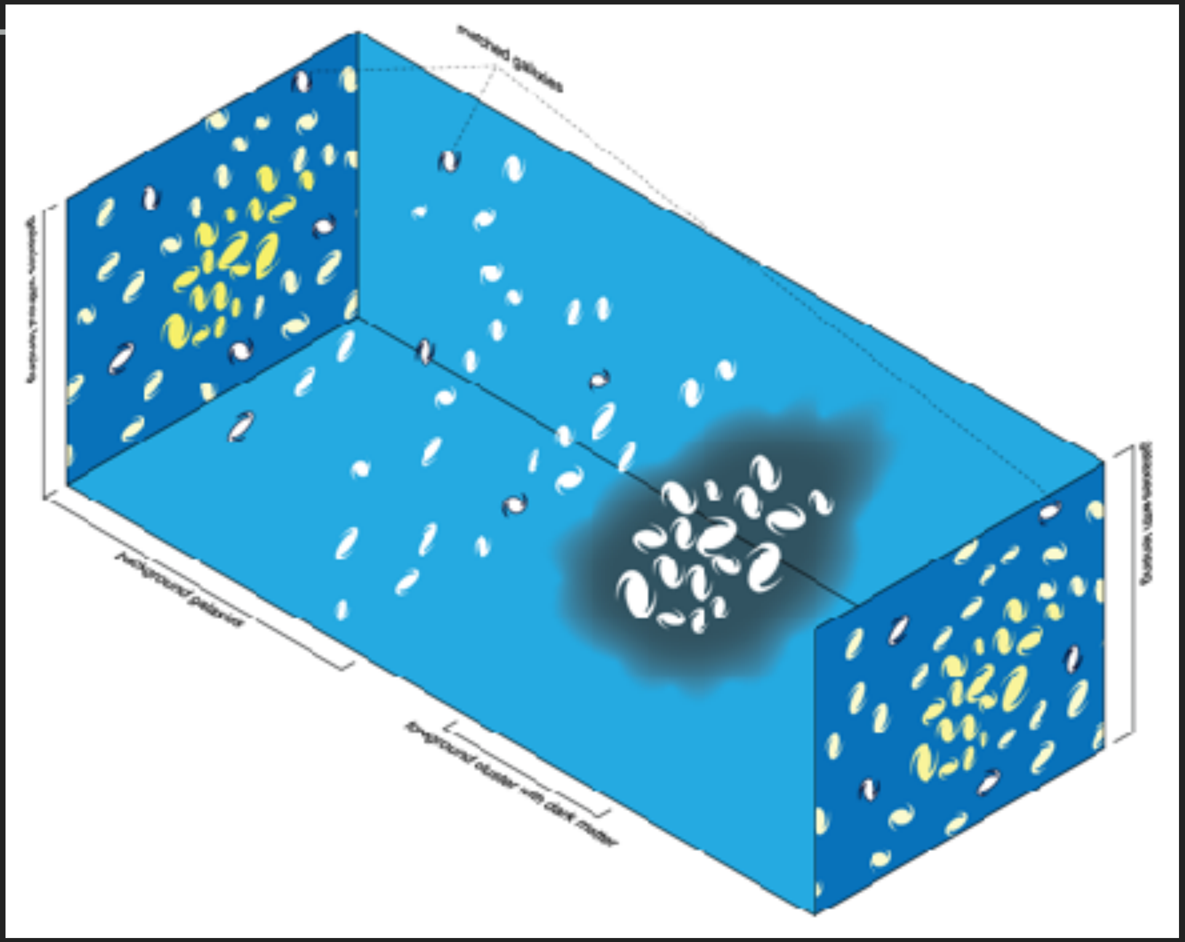
} 2 params for cosmic fluctuations

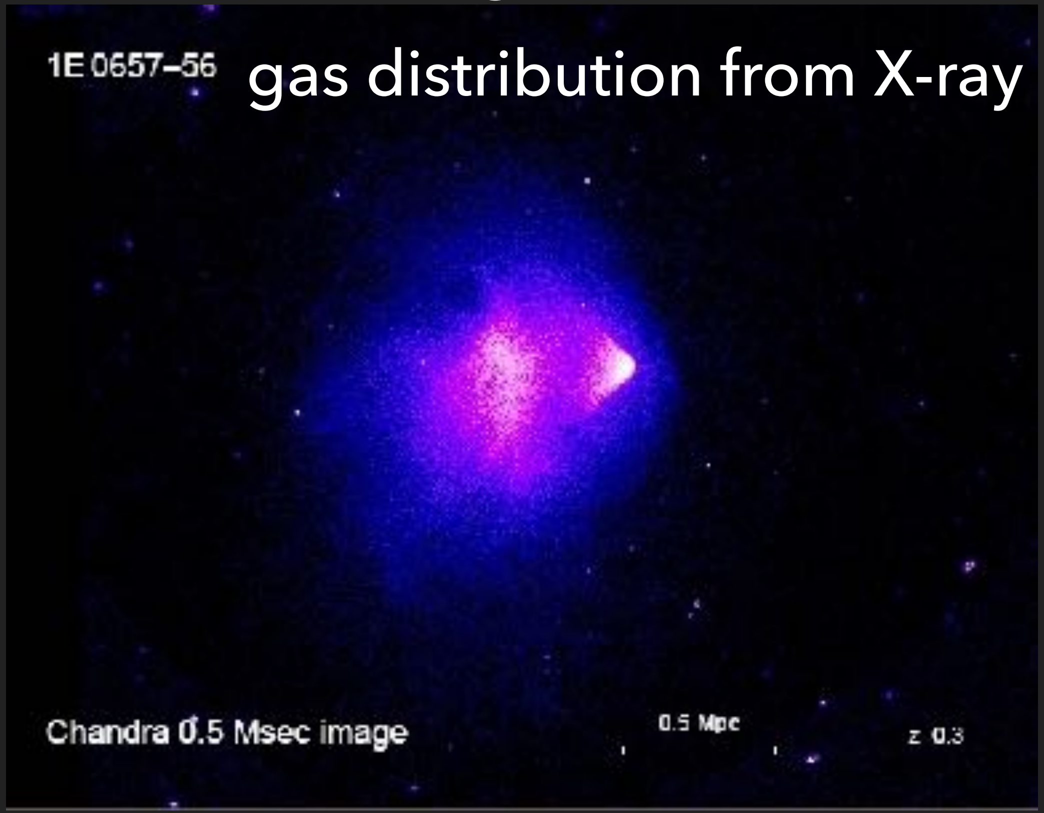
WHAT IS DARK ENERGY?

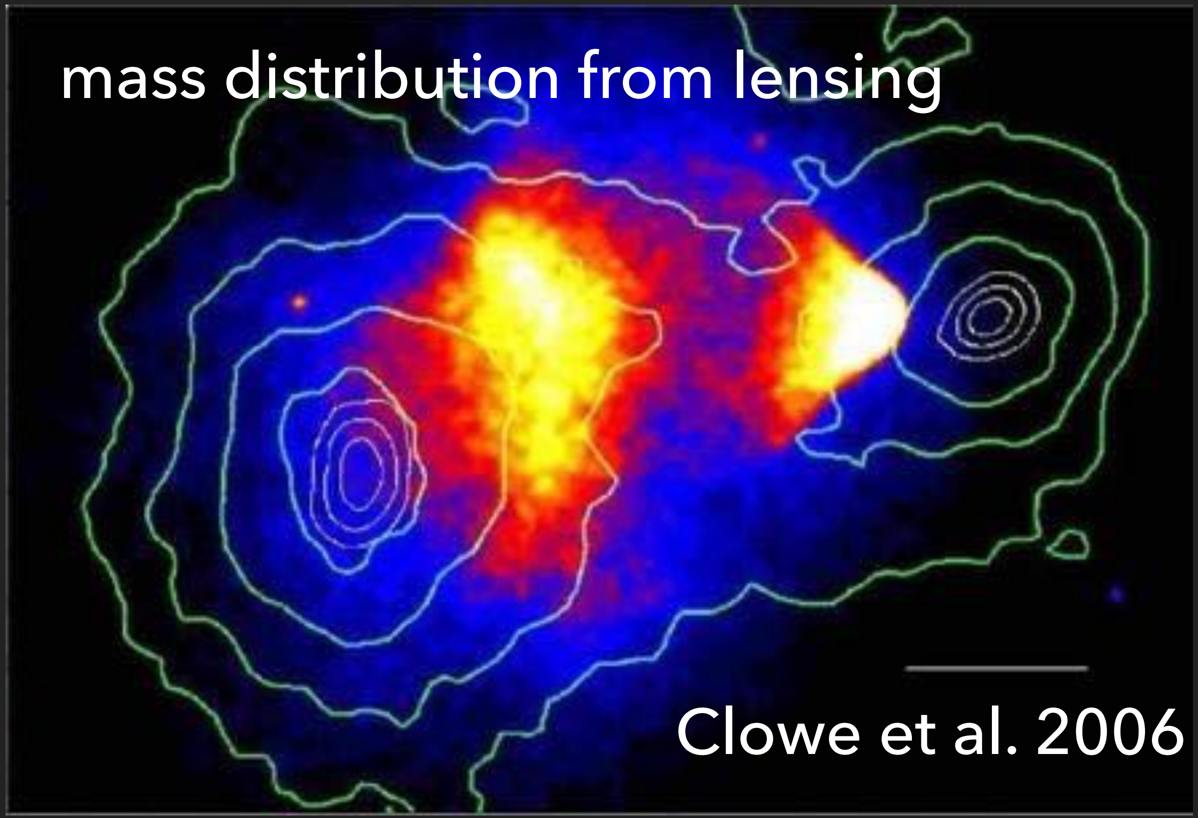
WHAT IS DARK MATTER?

DARK MATTER

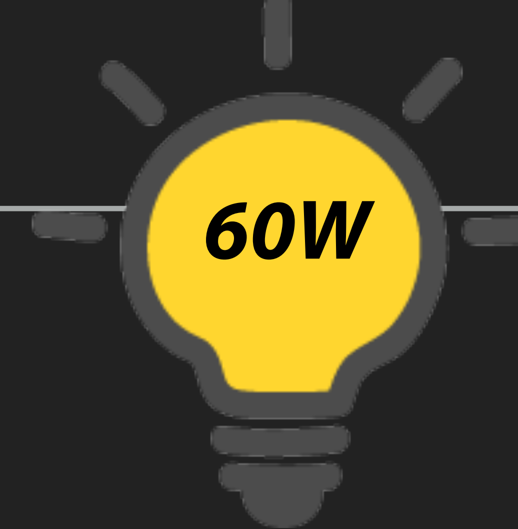
- No light emission but has mass (i.e., gravitational interaction)
- indirectly observable through *gravitational lensing*
  - light path bent by gravity
- ***bullet cluster***
  - *Chandra X-ray data vs lensing data*
  - *small cluster coming from the left and past to the right*

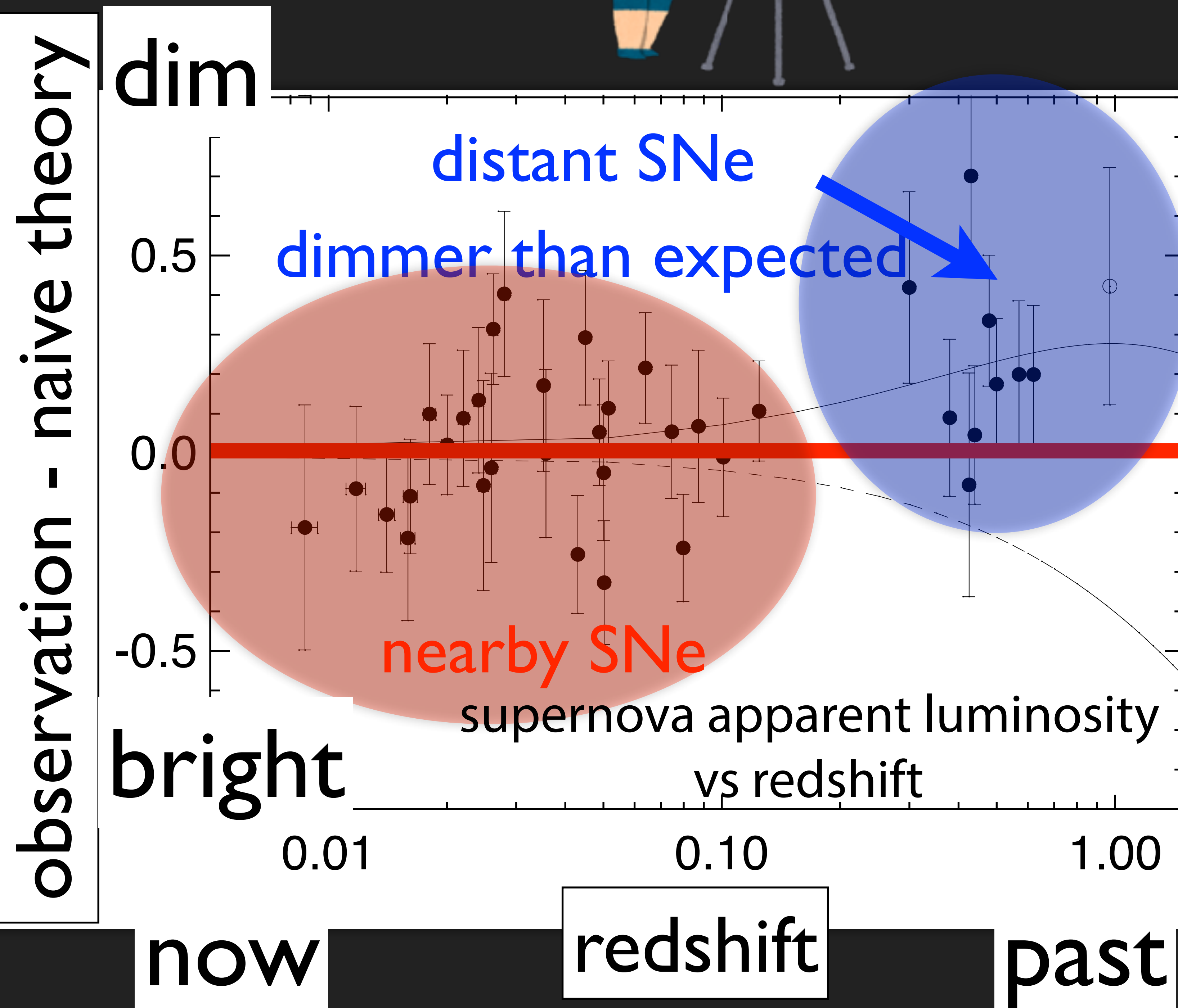






DARK ENERGY





Riess (1969-) Schmidt (1967-)

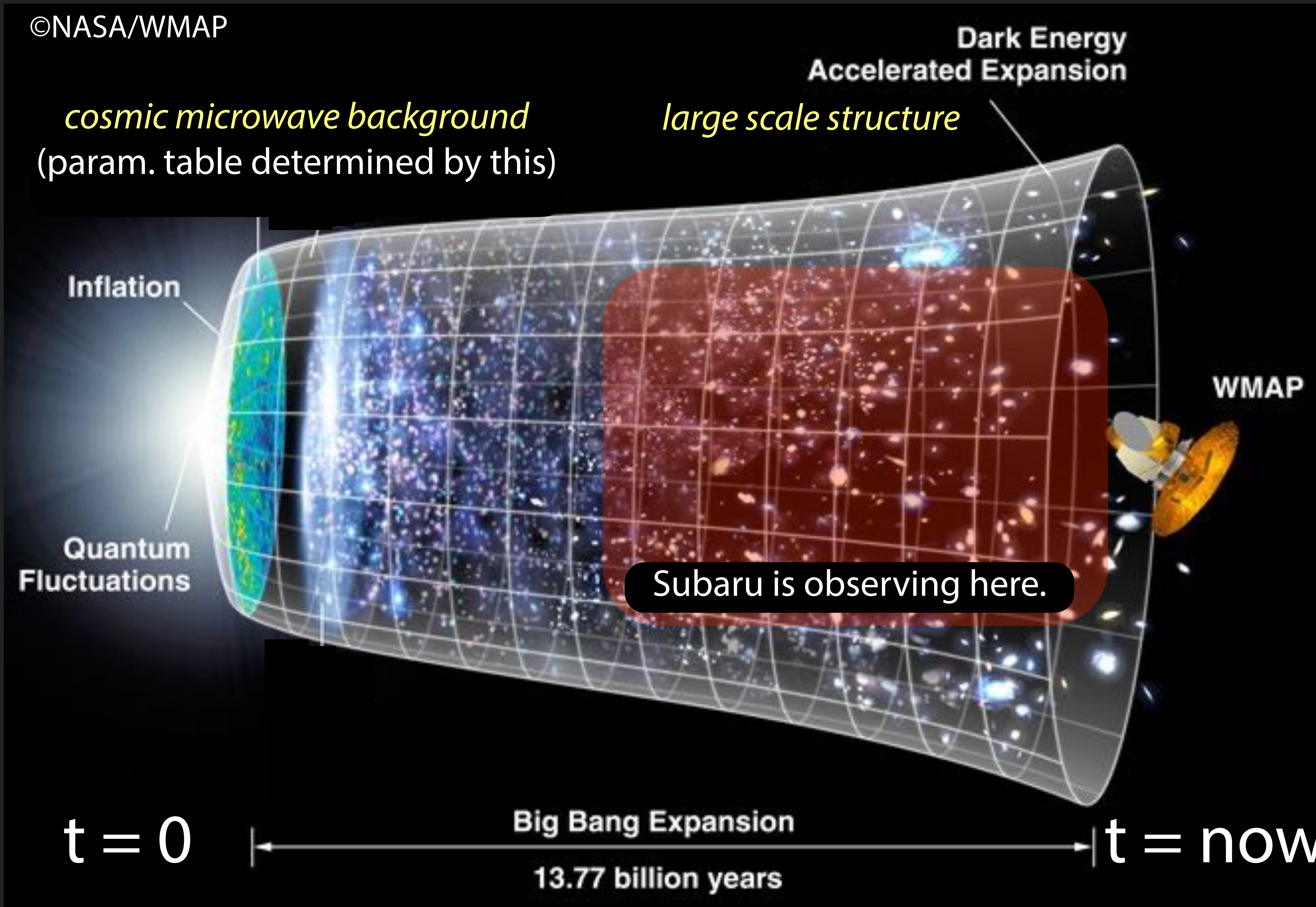
High-Z Supernova Search Team

Nobel Physics Prize in 2011, shared with Perlmutter (Supernova Cosmology Project)

***Something* is accelerating the expansion!**

PROBLEM SETTINGS

OBSERVABLE UNIVERSE

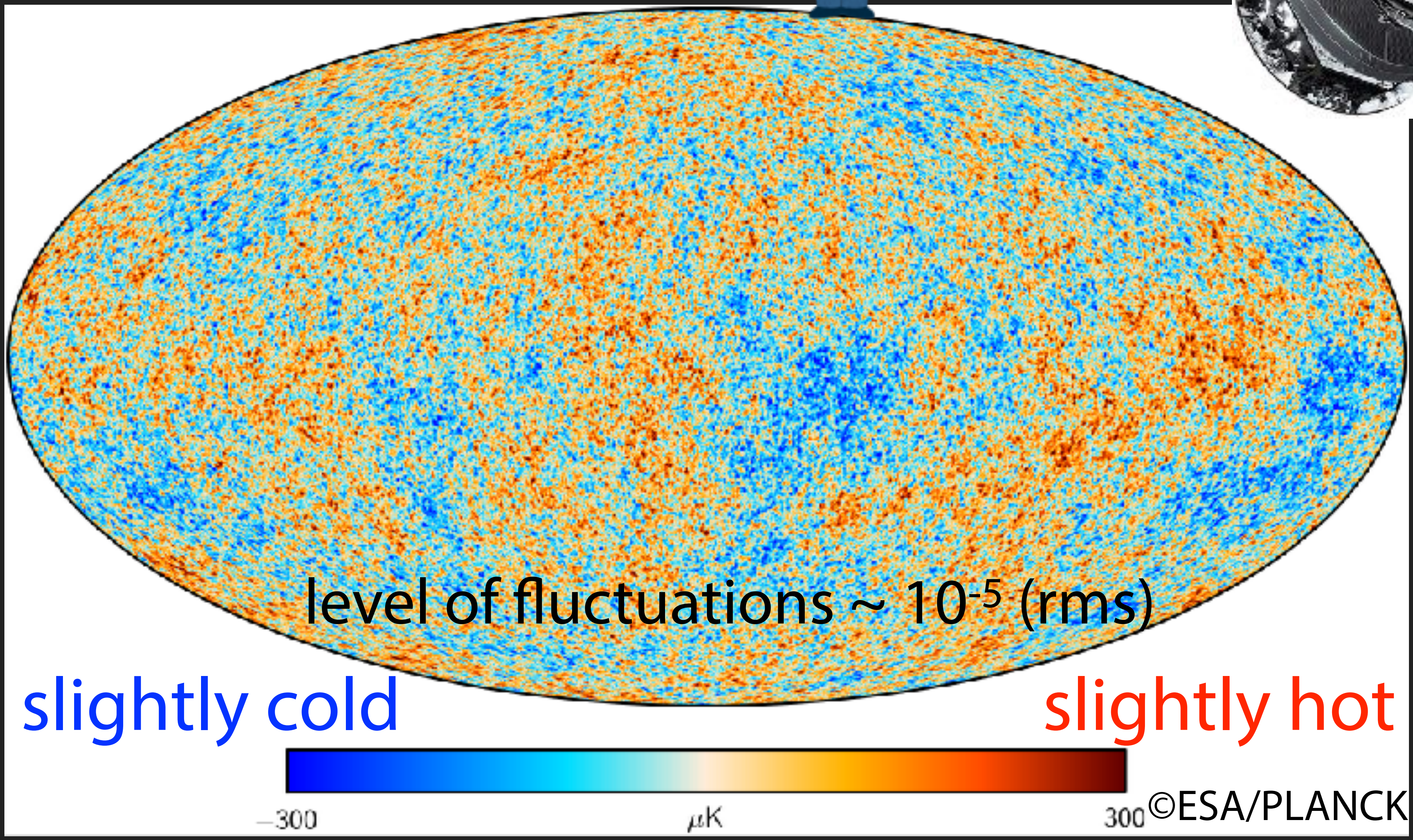


3K COSMIC MICROWAVE BACKGROUND ANISOTROPY

"*initial condition*" of our Universe @ t ~ 380 kyrs

c.f., t_{now} ~ 14 Gyrs

7 hours old ⟷ 30 years old



COSMIC RANDOM FIELDS

- Assume *cosmological principle*
  - no special place
  - *homogeneity* + *isotropy* (in a statistical sense)
- consider the "fluctuations":

$$\delta_A(\mathbf{x},t) = [A(\mathbf{x},t) - A(t)]/A(t)$$

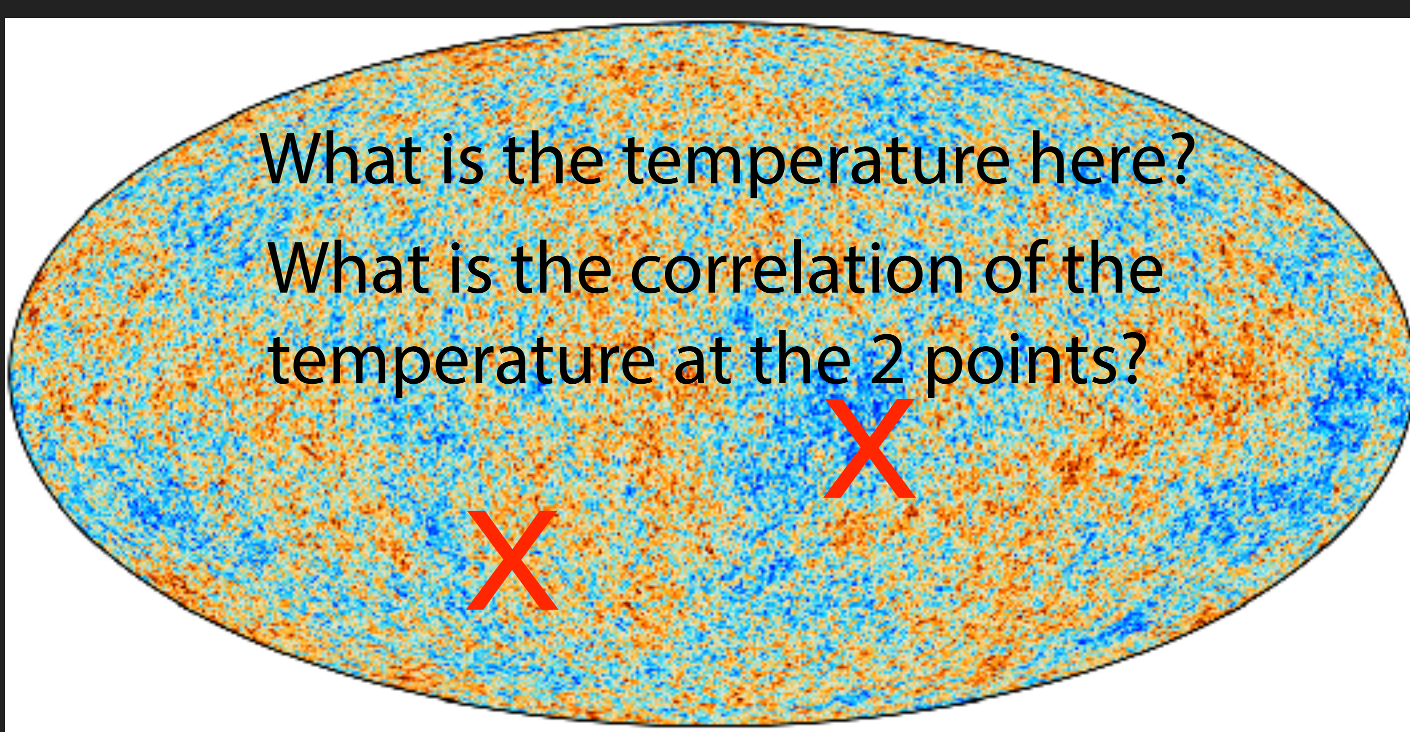
- $P[\delta_A(x_1)] = P[\delta_A(x_2)] = \ldots = P[\delta_A(x_N)]$
- $P[\delta_A(x_1), \delta_A(x_2)] \neq P[\delta_A(x_1)]\, P[\delta_A(x_2)]$

position time

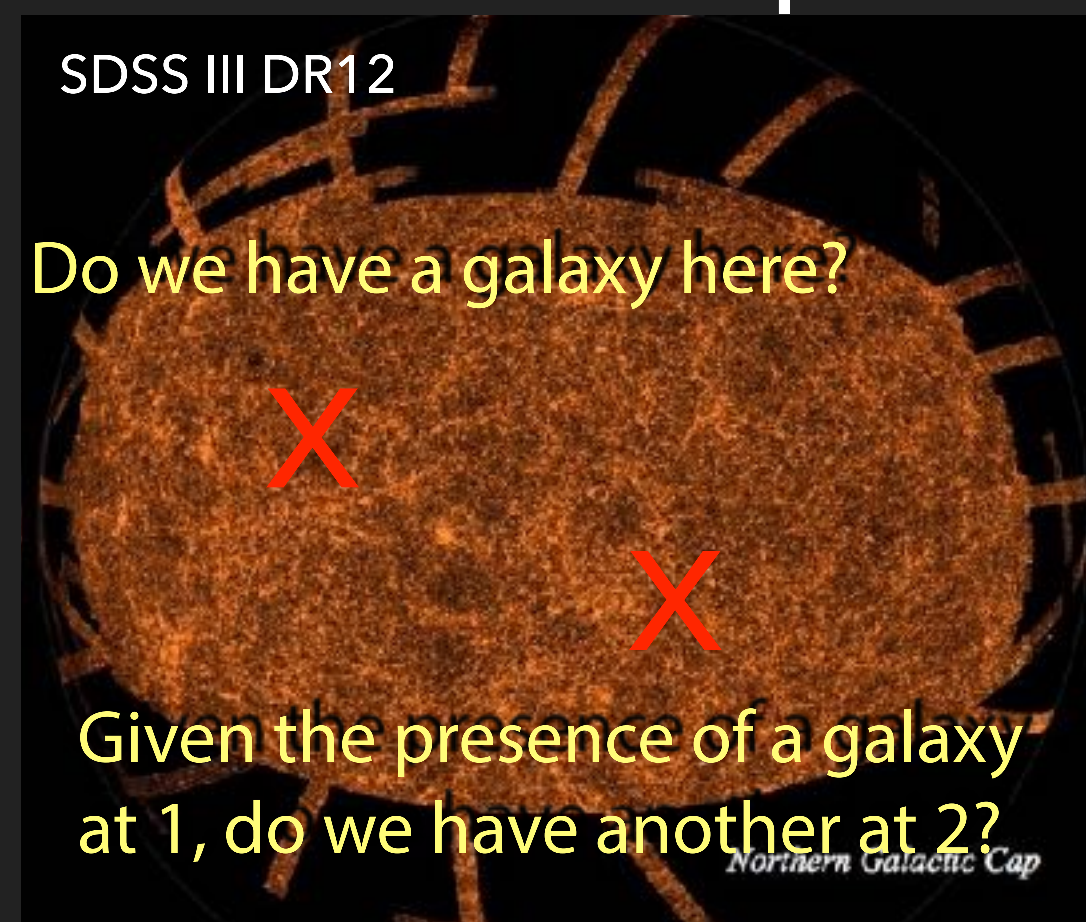
$$\langle A(\mathbf{x},t)\rangle = A(t)$$

<…>: ensemble average

equivalent to spatial average



correlation between positions !



GAUSSIAN RANDOM FIELD AS THE INITIAL CONDITION

- Gaussianity
  - consider N-point correlators:

$$\langle \delta_A(\mathbf{x}_1)\delta_A(\mathbf{x}_2)\ldots\delta_A(\mathbf{x}_N)\rangle_c = 0, \qquad \text{for} \quad \mathrm{N} \geq 3$$

  - The *2-point correlation function* determines everything:

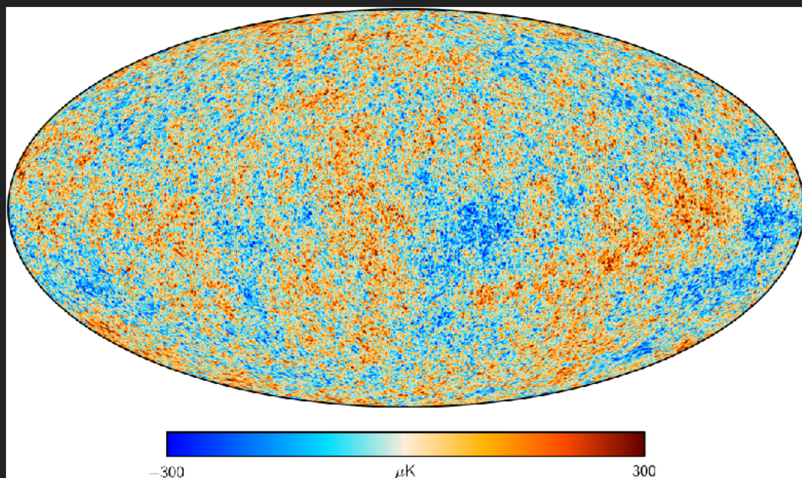
$$\langle \delta_A(\mathbf{x}_1)\delta_A(\mathbf{x}_2)\rangle = \xi_A(\mathbf{x}_1, \mathbf{x}_2) = \xi_A(|\mathbf{x}_1 - \mathbf{x}_2|)$$

statistical isotropy

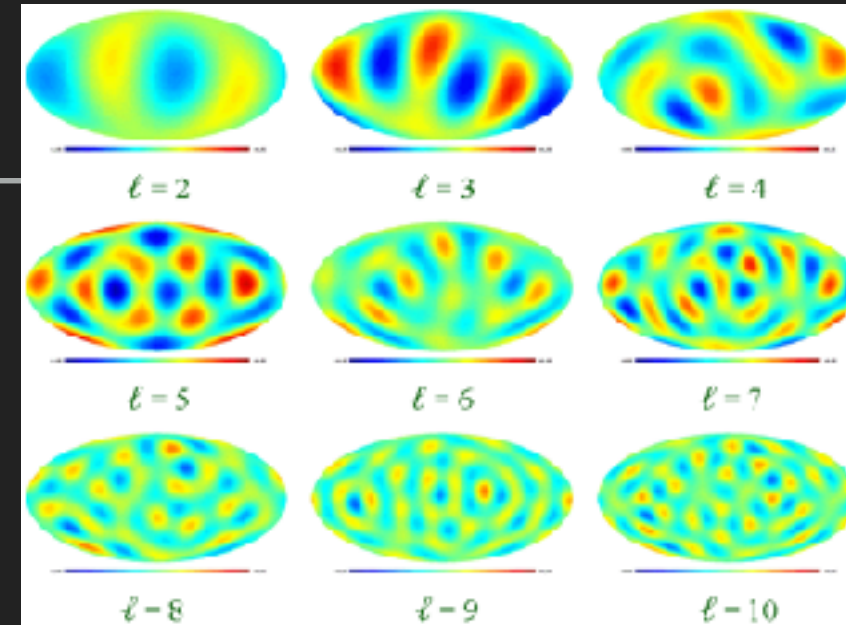
  - or equivalently, the *power spectrum* in Fourier space or in harmonic space
  - In terms of joint probability density functional, this gives

$$P[\delta_A] \propto \exp\left[-\frac{1}{2}\int \mathrm{d}^3x \int \mathrm{d}^3x' \delta_A(\mathbf{x})\xi_A^{-1}(\mathbf{x}, \mathbf{x}')\delta_A(\mathbf{x}')\right]$$

STATISTICS OF CMB

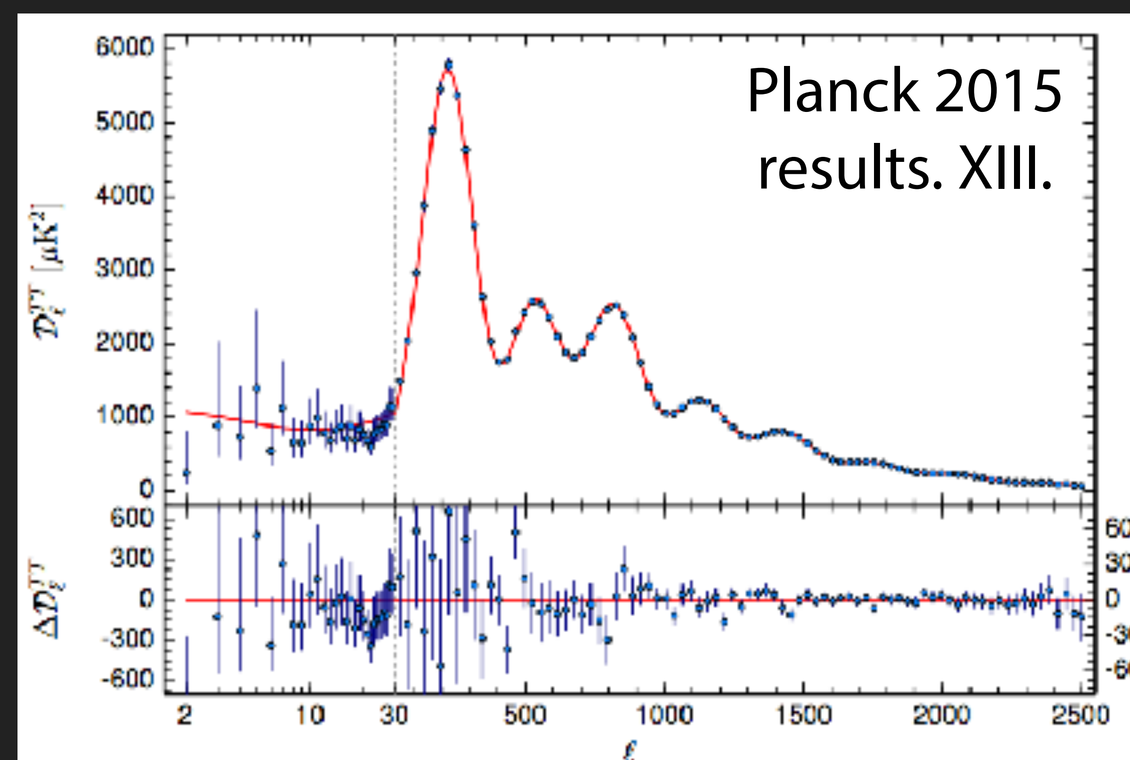


2pt statistics

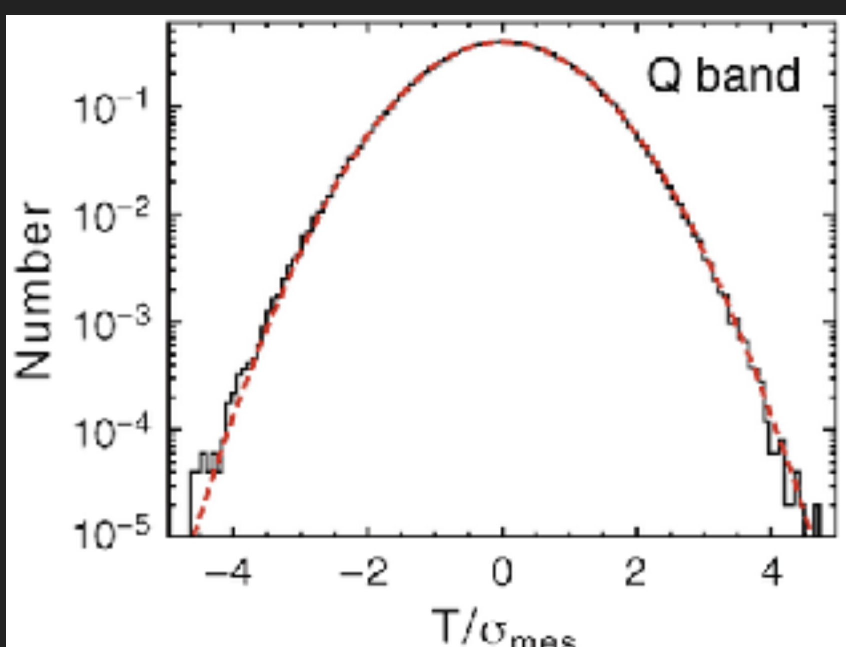


spherical harmonics expansion

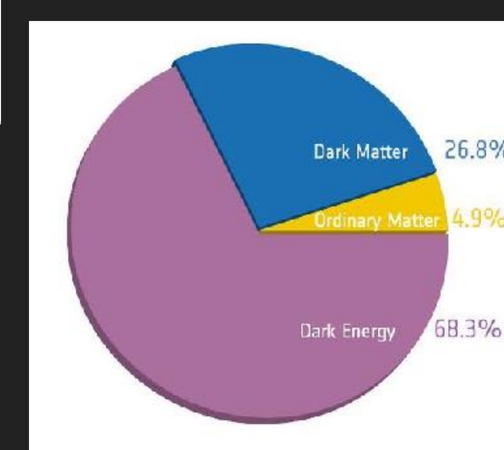
"power spectrum"



1pt probability distribution

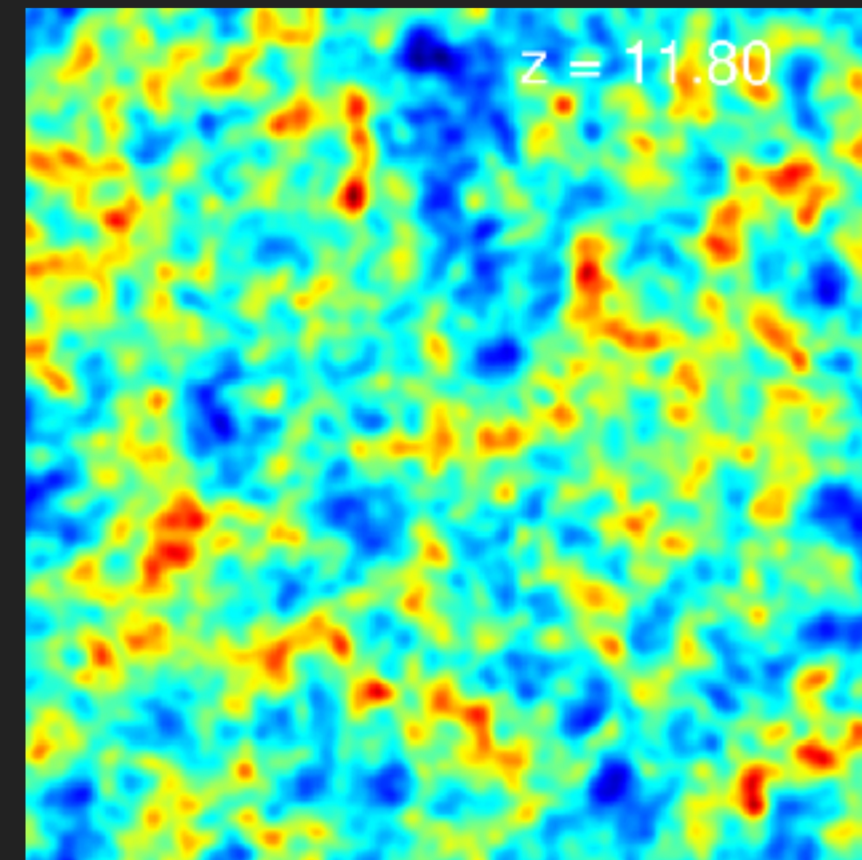


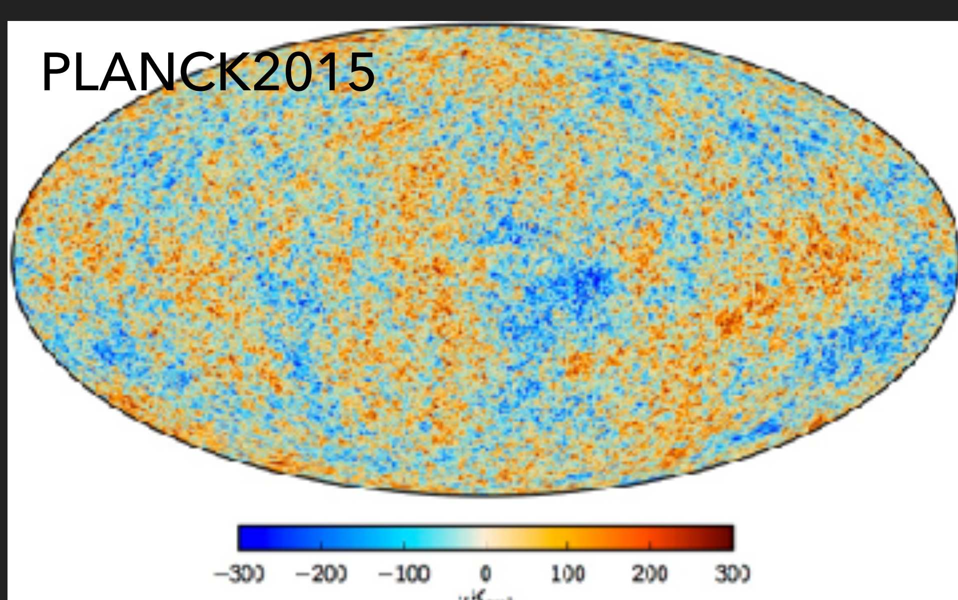
Gaussian Random Field to a very good accuracy (non-Gaussianity<0.1%)



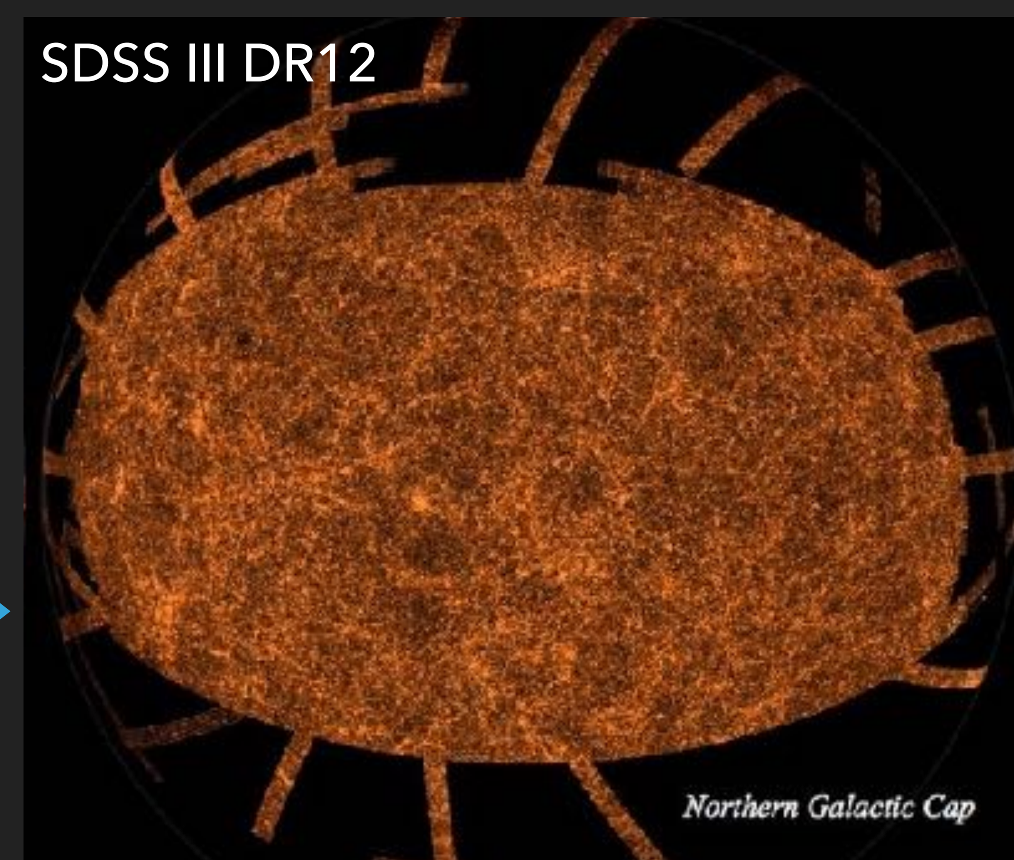
EVOLUTION OF COSMIC FLUCTUATIONS

- Gravitational instability is the driver that forms rich cosmic structures
  - tiny primordial fluctuations seen in the CMB are the seed for the nearby structures
  - Amplified by gravitational instability over the cosmic time
  - Any astronomical structures (stars, galaxies, groups, clusters) originated from this!



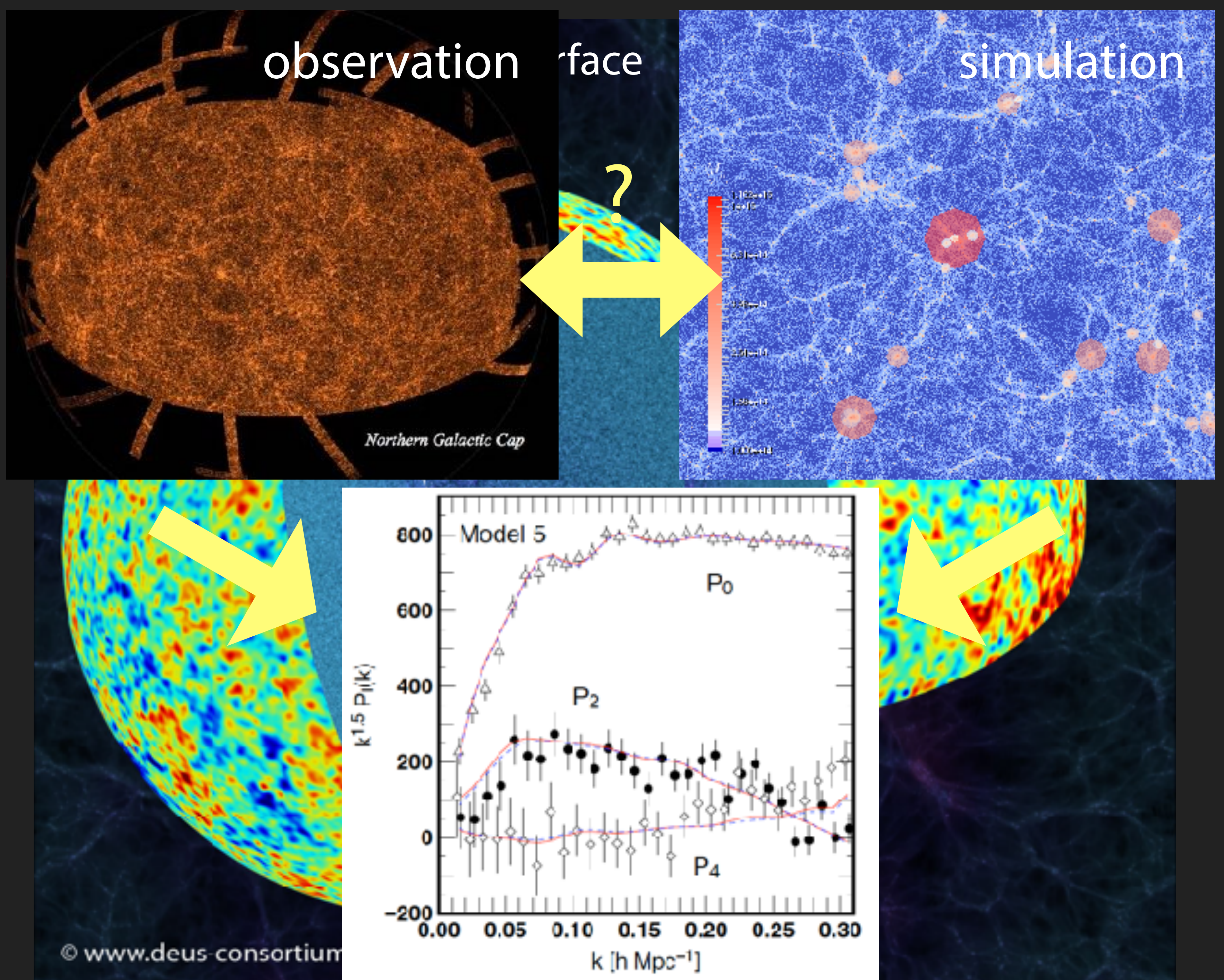


Gravity



EVOLUTION OF COSMIC FLUCTUATIONS (CONTD)

- We do not have the information on the initial condition sufficient for full determination of the *particular random realization* in which we live!



CMB observation

↓

parameter determination with uncertainties

↓

pick one parameter set within the CMB constraint $\chi^2_{\rm red}$

↓

generate a random realization of the initial fluctuations for the given parameters

↓

solve the dynamics

↓

measure statistics

↓

compare with observations

LATE-TIME OBSERVABLES

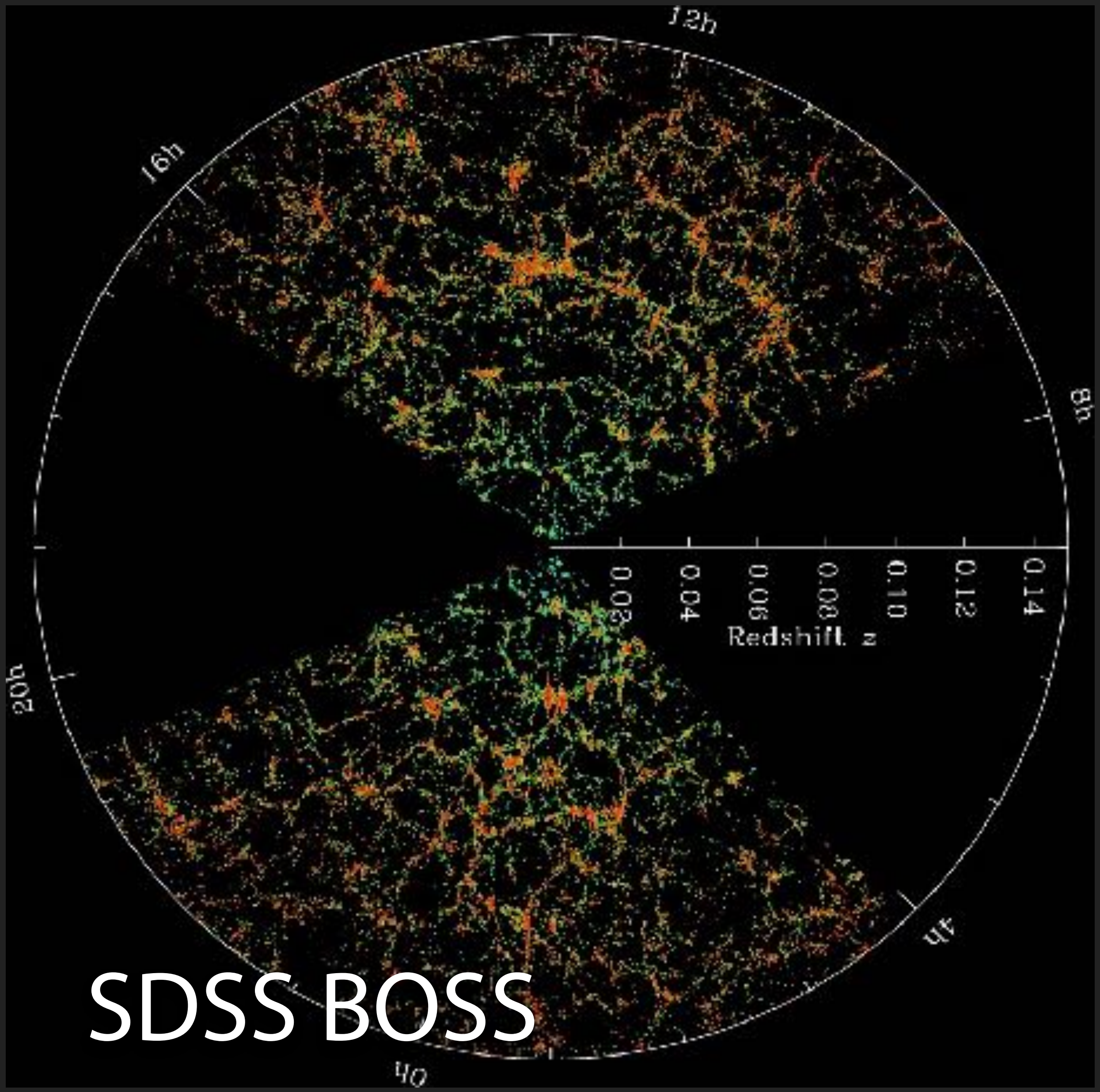
TARGET OBSERVABLES

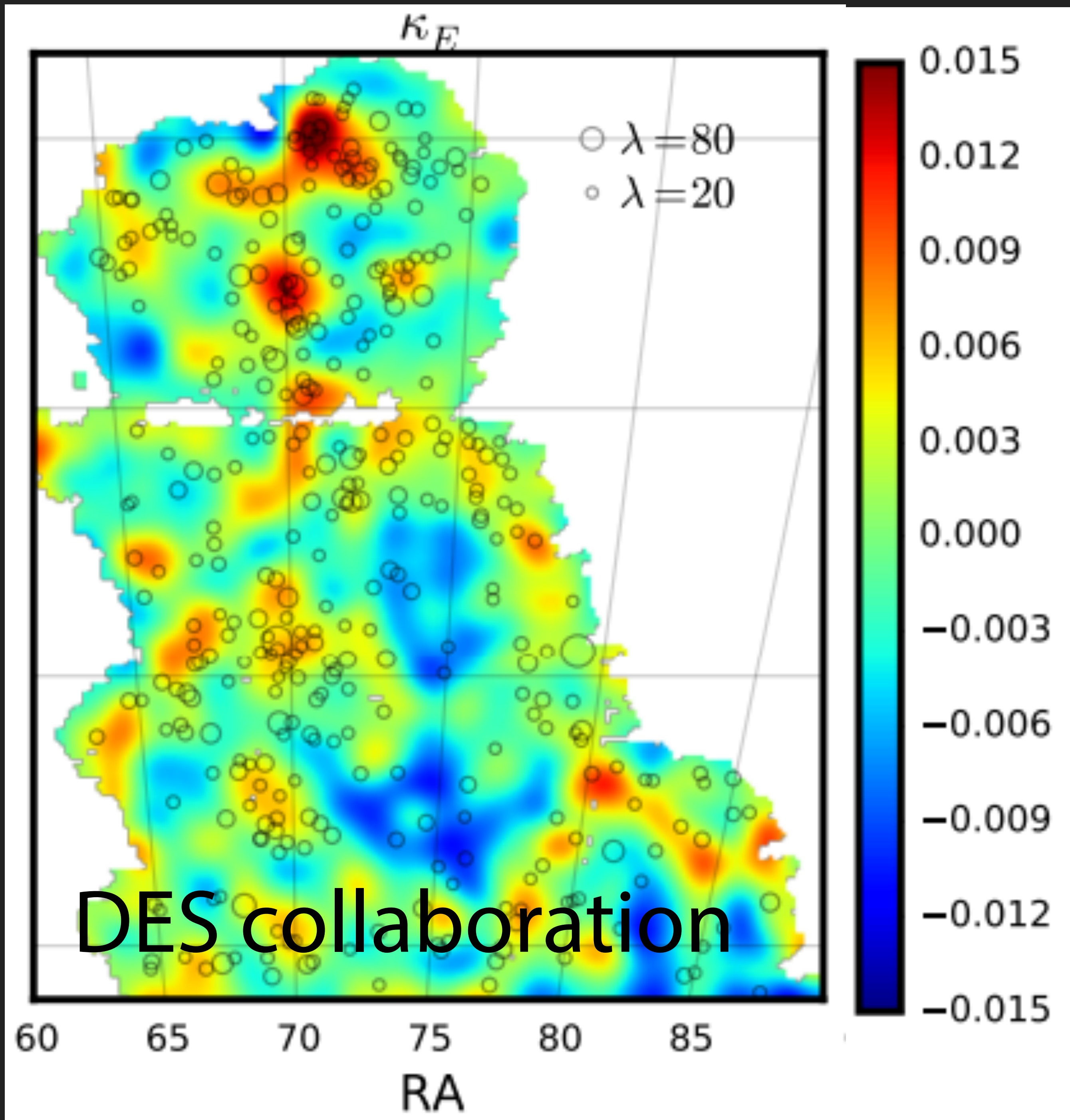
Galaxy clustering

in point process in 3D space

weak gravitational lensing

projected mass distribution on 2D plane

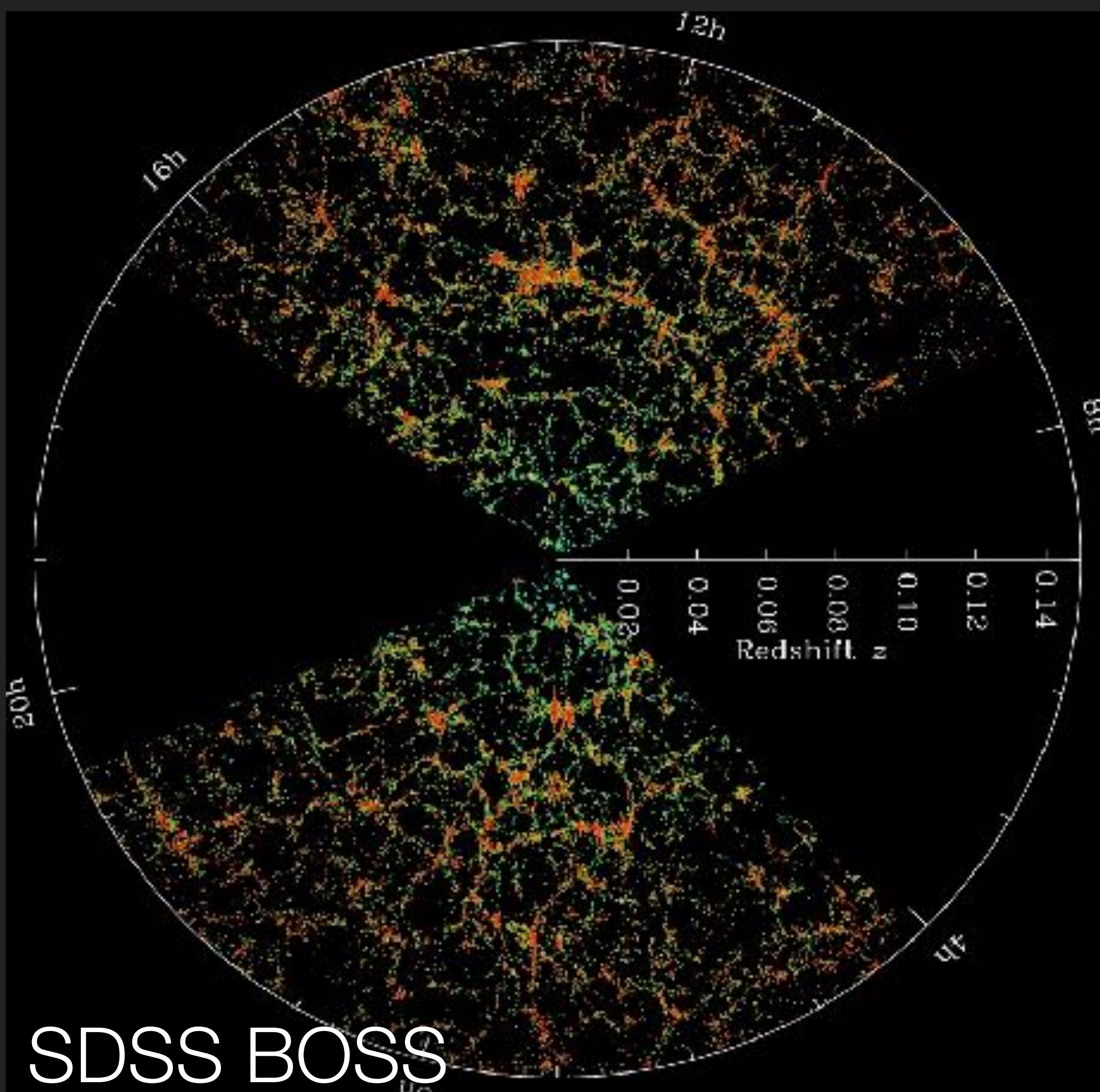




TARGET OBSERVABLES

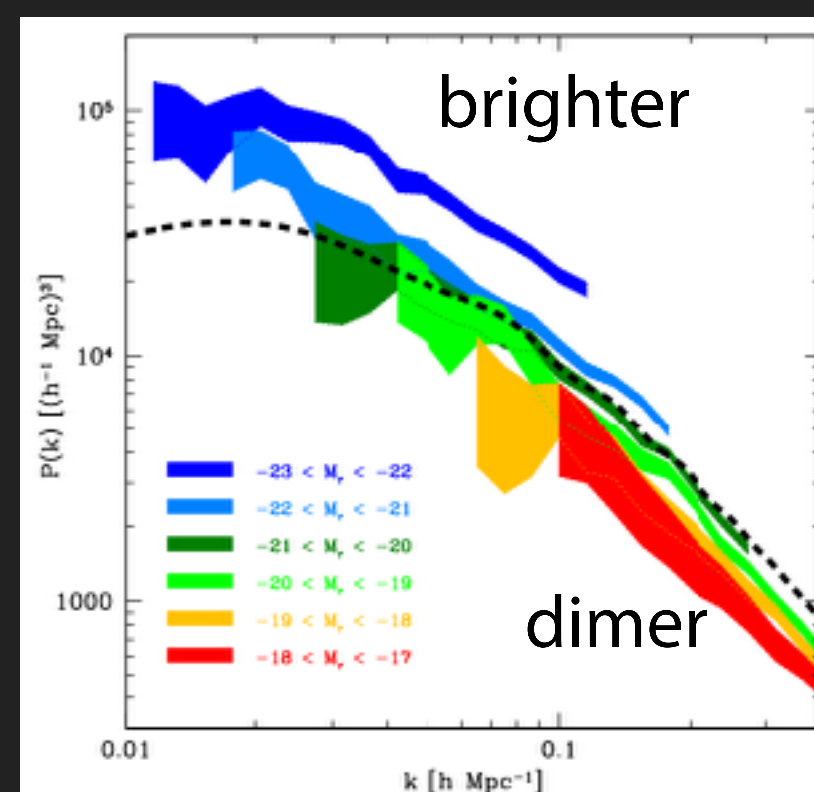
Galaxy clustering

in point process in 3D space



- Basic quantity = Positions
- Additional quantities
  - characteristic of galaxies
  - luminosity, color, morphology …

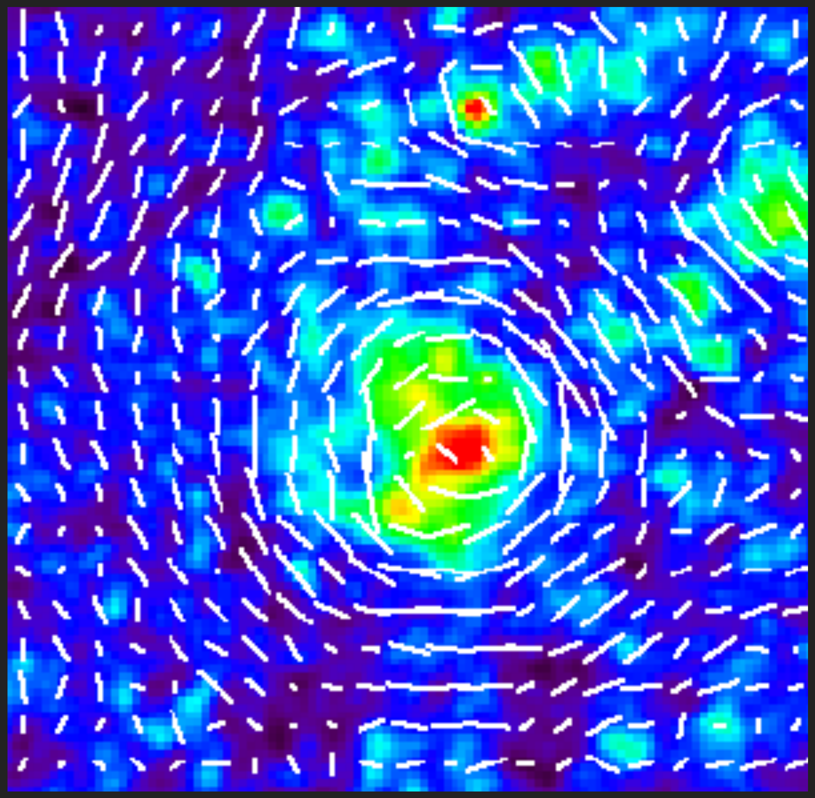
We do not know the relation between the mass density field and galaxy density field! (galaxy bias uncertainty)



galaxy power spectra

Tegmark+'04

TARGET OBSERVABLES

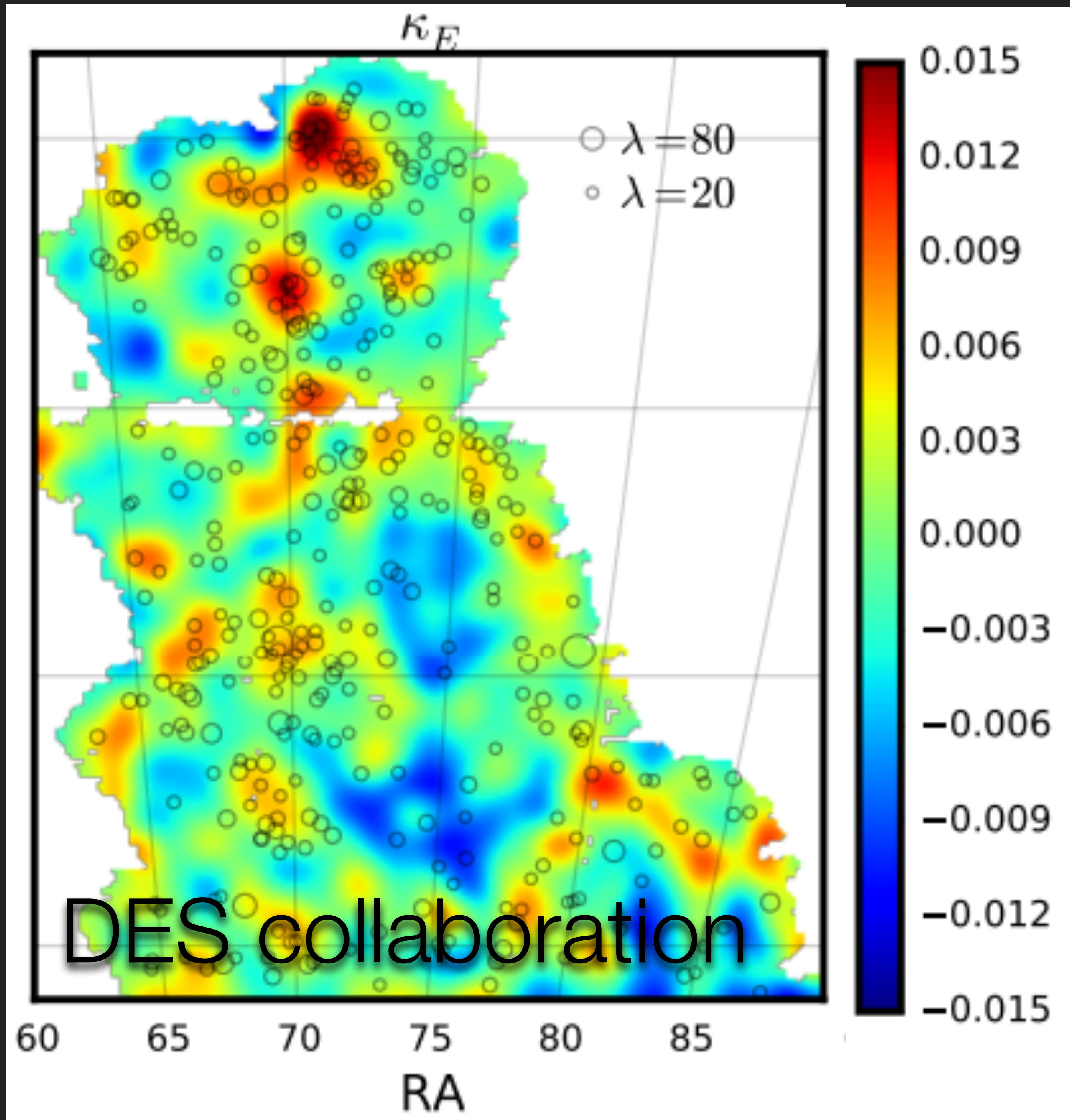


simulation by T. Hamana @ NAOJ

ticks = expected orientation of galaxies

weal gravitational lensing

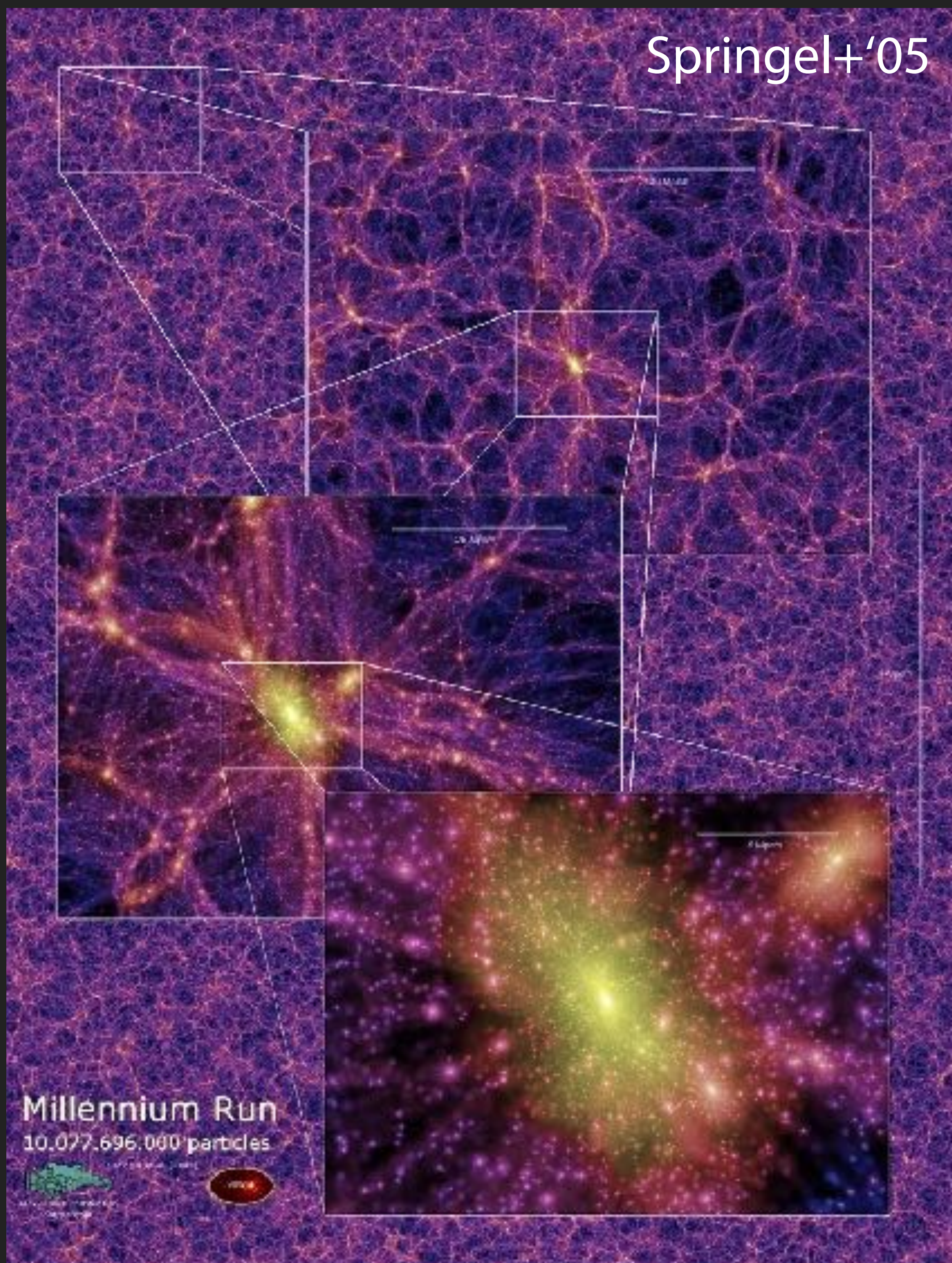
projected mass distribution on 2D plane



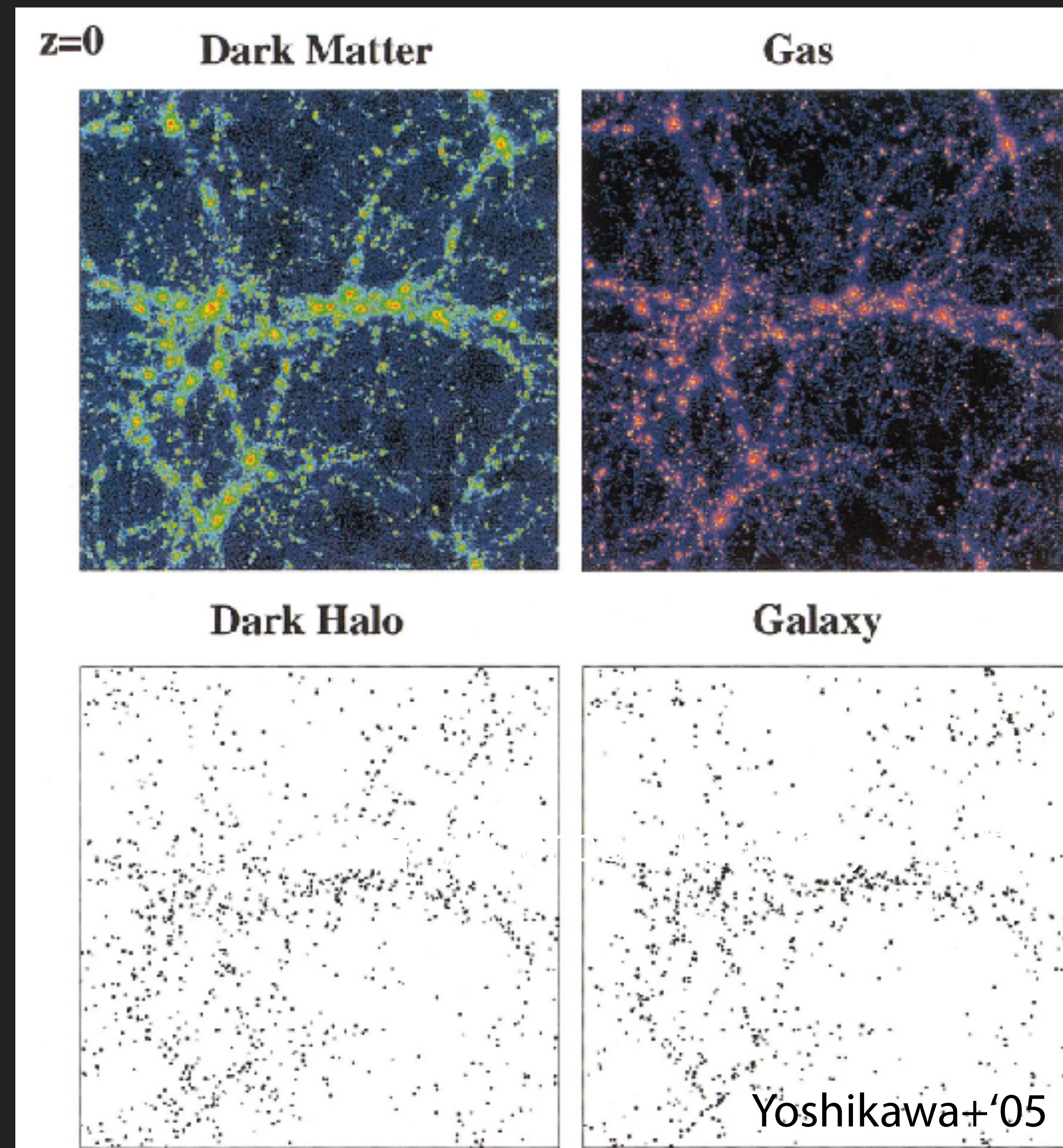
- direct observable
  - shape of background galaxies
- accessible information
  - foreground projected mass map
  - including *dark matter*

WHAT WE CAN AND CANNOT PREDICT

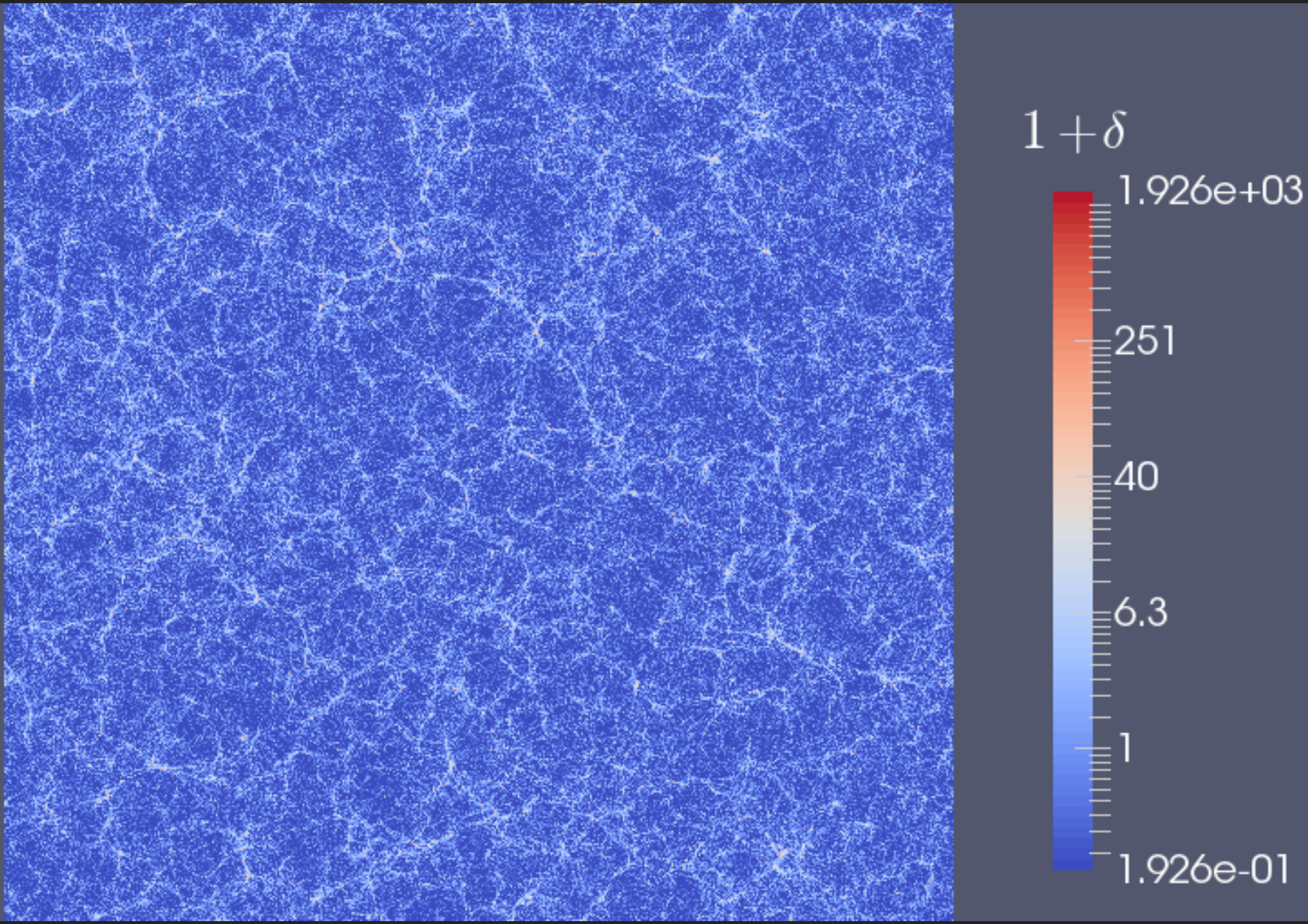
gravity only simulation



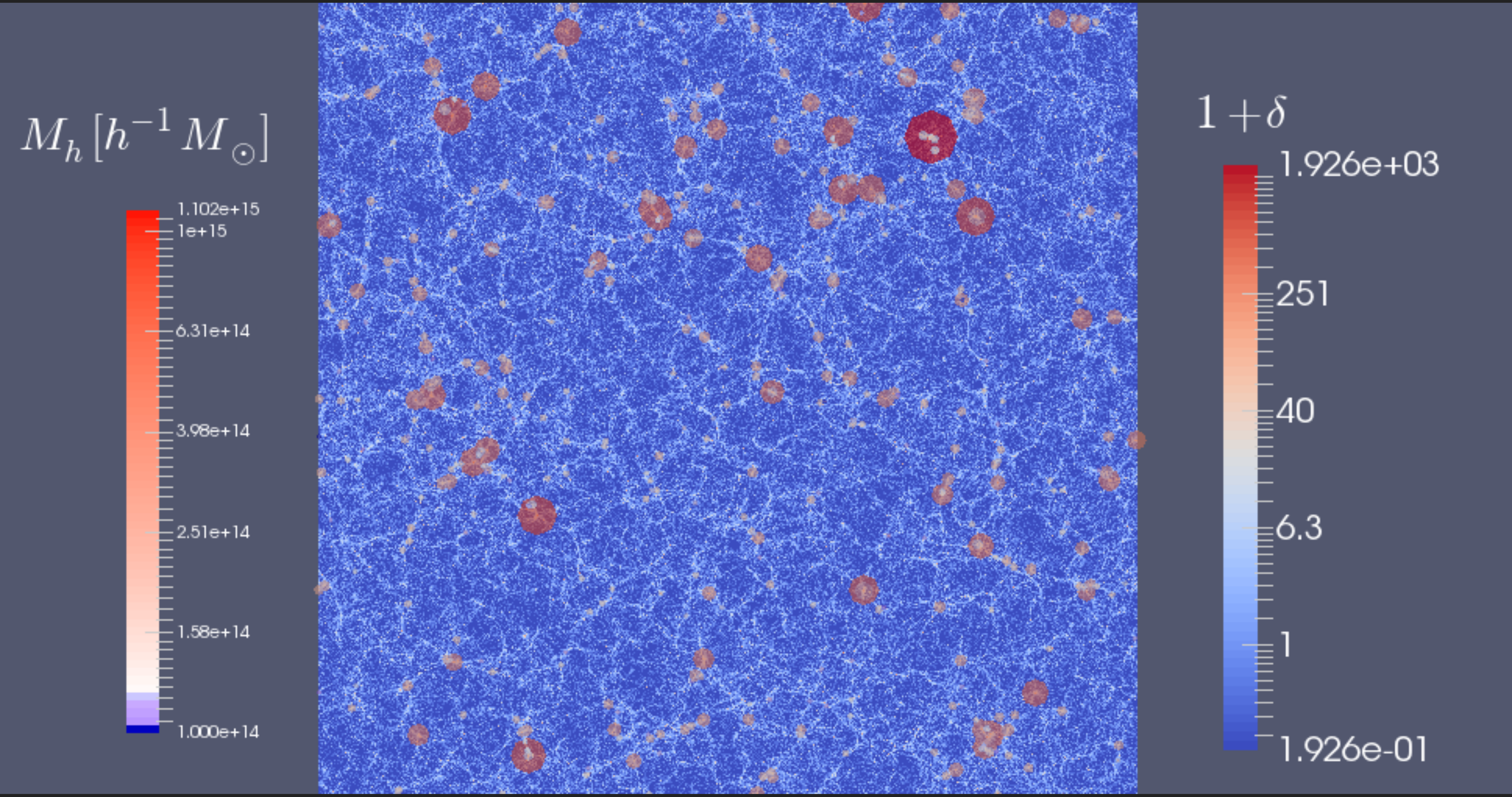
hydrodynamical simulations with various processes



2D SLICE OF COSMIC STRUCTURE



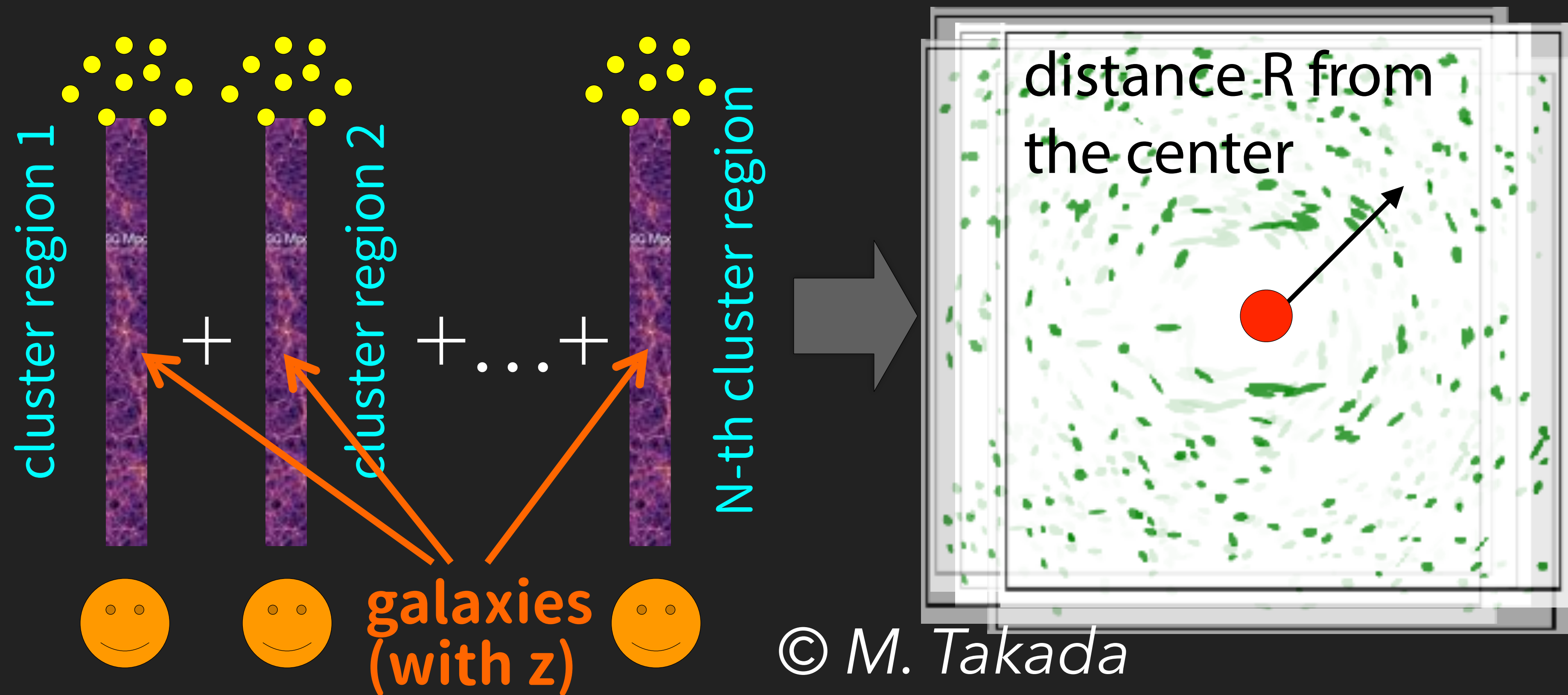
2D SLICE OF COSMIC STRUCTURE



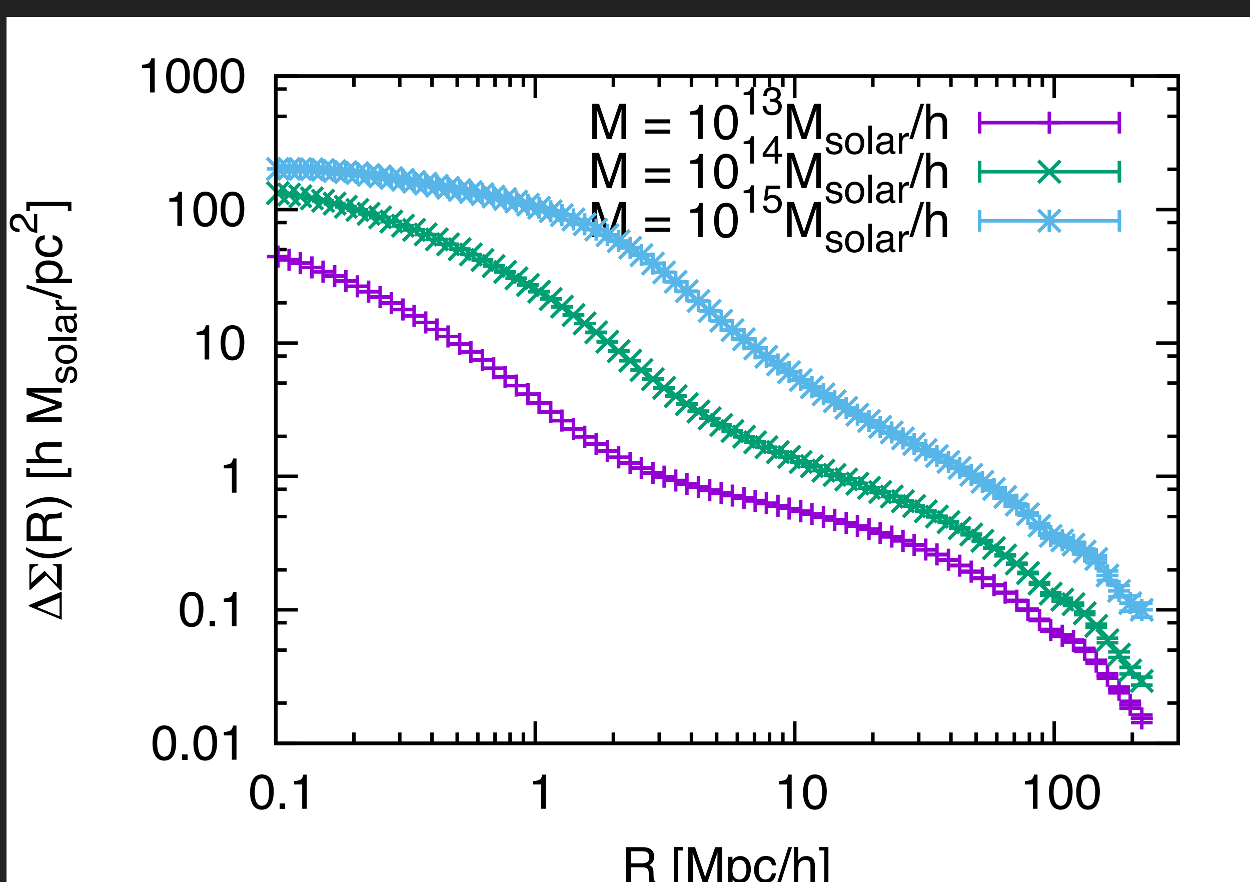
Need an accurate theoretical template as a function of the mass of halos

GALAXY-GALAXY LENSING

e.g., Oguri & Takada '11



stack the images of background galaxies centered at the foreground cluster positions



The cross-correlation signal of the two observables

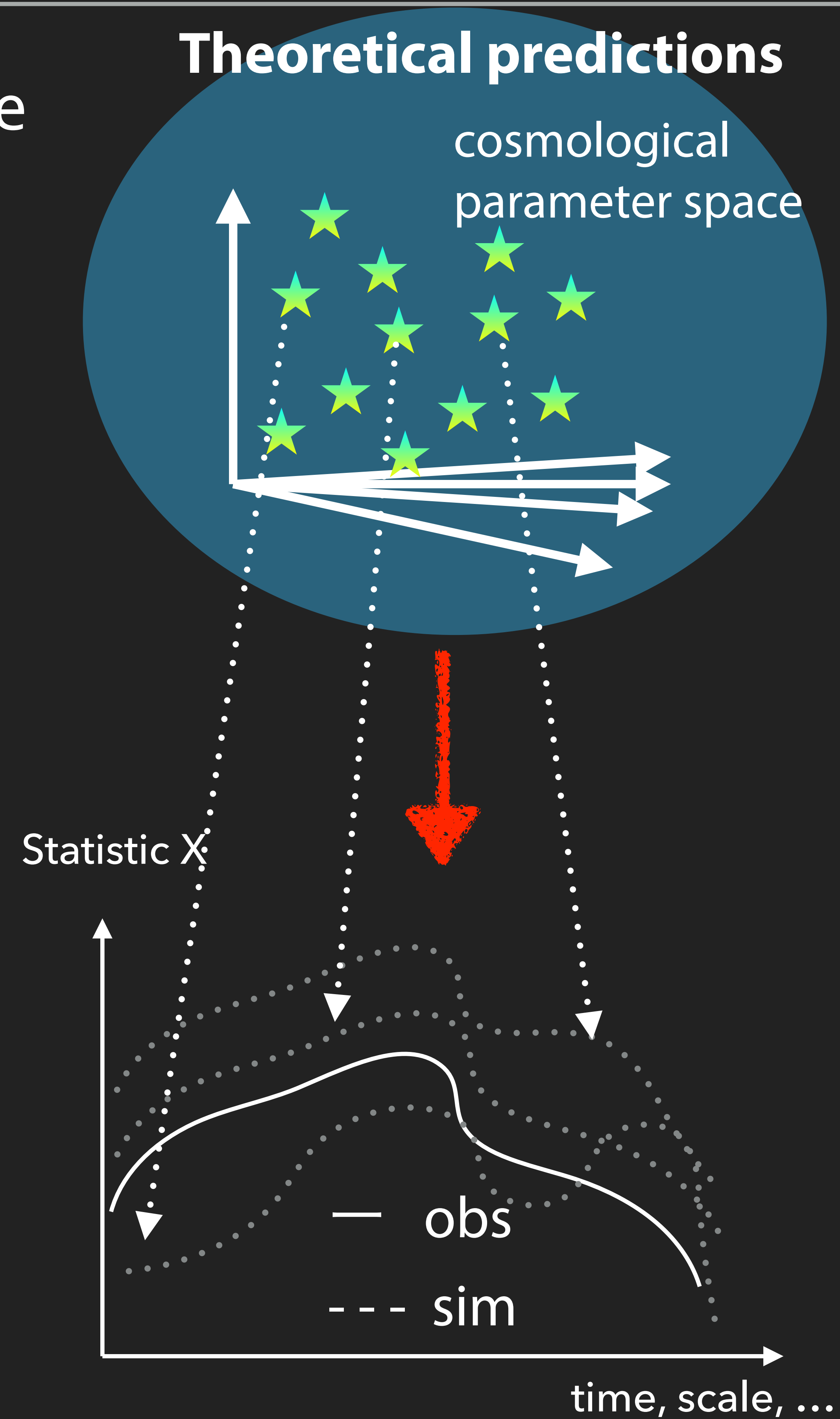
$$\Sigma(R) = \overline{\rho} \int \left[1 + \xi_{\rm gm}\left(\sqrt{R^2 + \Pi^2}\right)\right] d\Pi$$

Break the degeneracy btwn galaxy bias and cosmology!

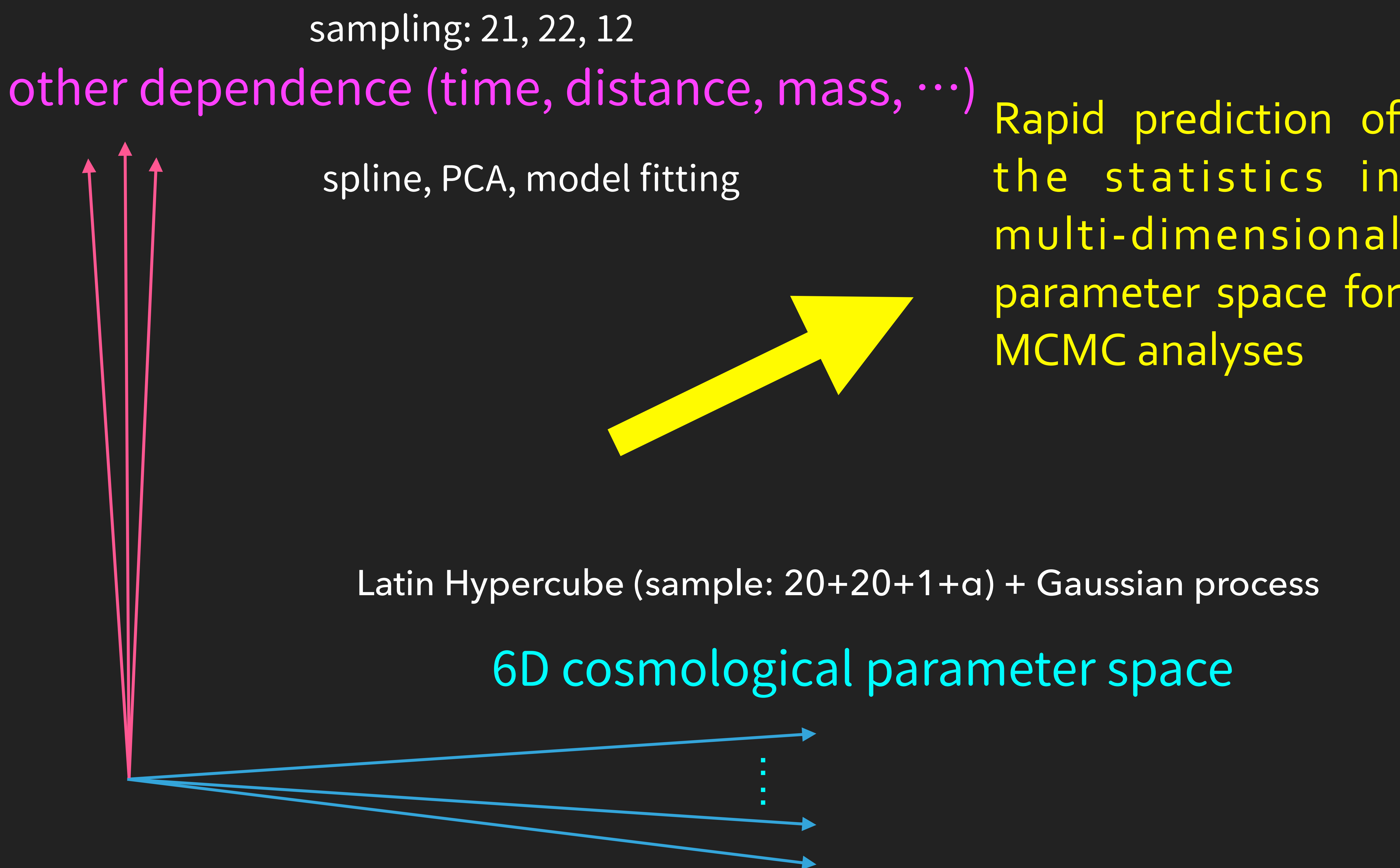
DARK EMULATORS

OBJECTIVE

- Numerical cosmology with large-scale structure data
  - direct confrontation of simulations to observations for model/parameter estimation
- Accurate determination of basic statistical quantities with N-body simulations
  - cosmology dependence
  - machine-learning based approach
- handy numerical codes for statistical analyses
  - rapid emulator written in python

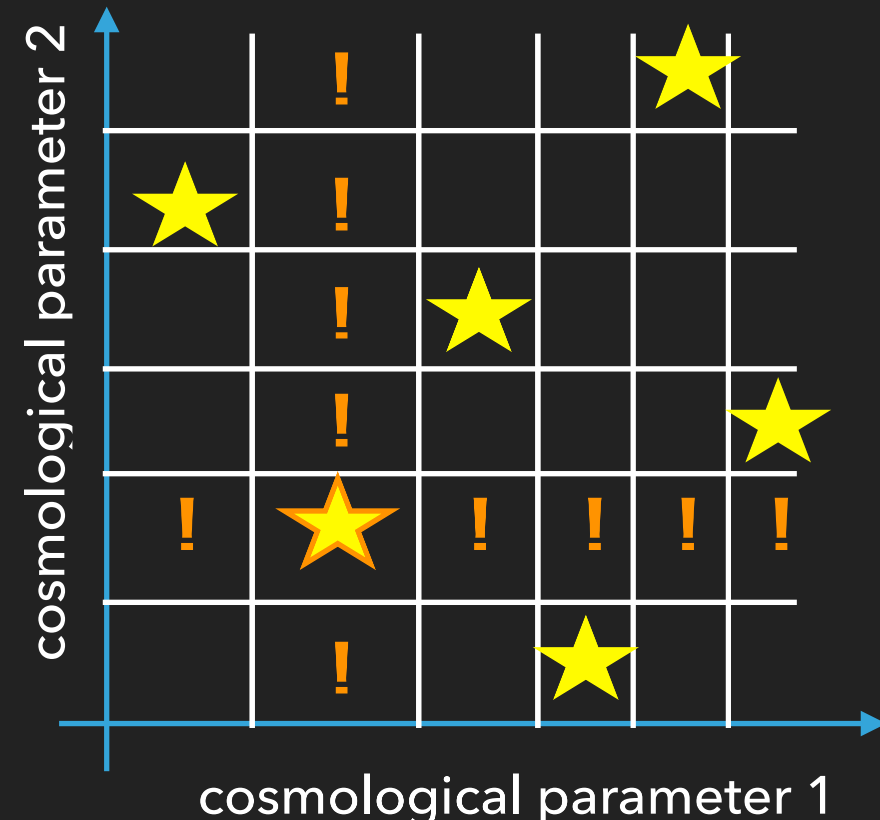


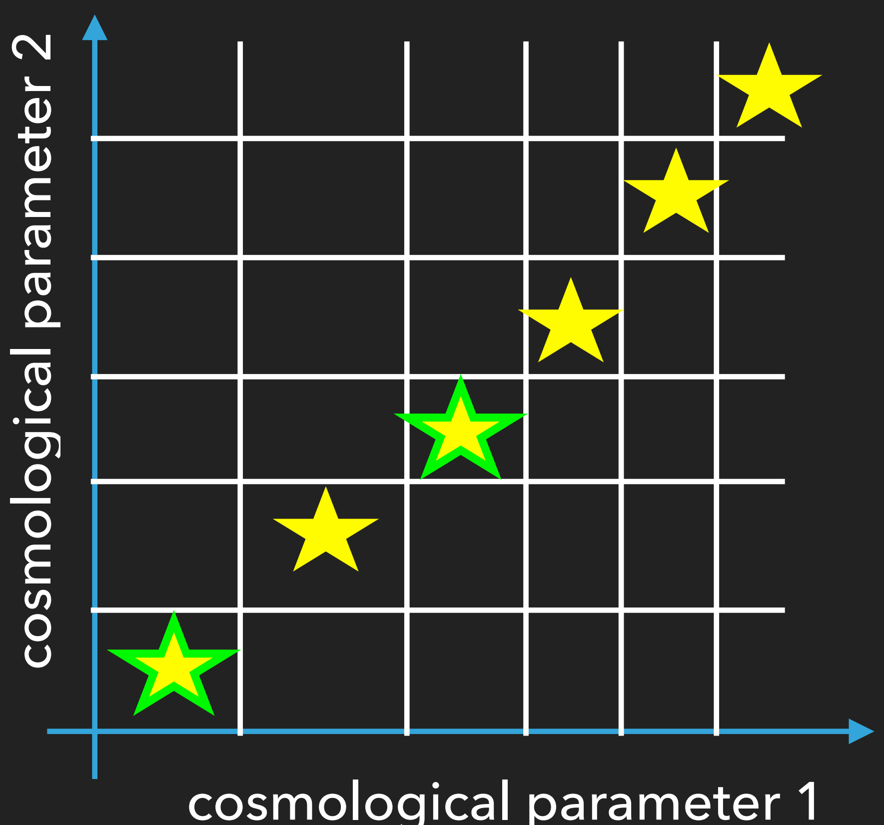
THE SIMULATION DESIGN IN HIGH DIMENSION SPACE

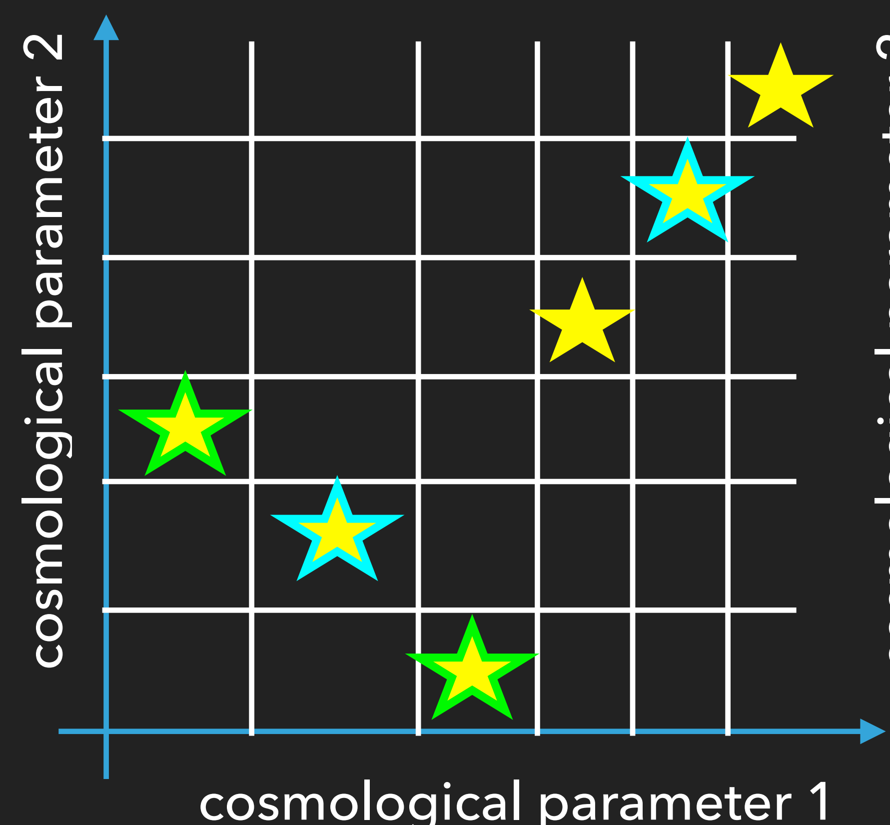


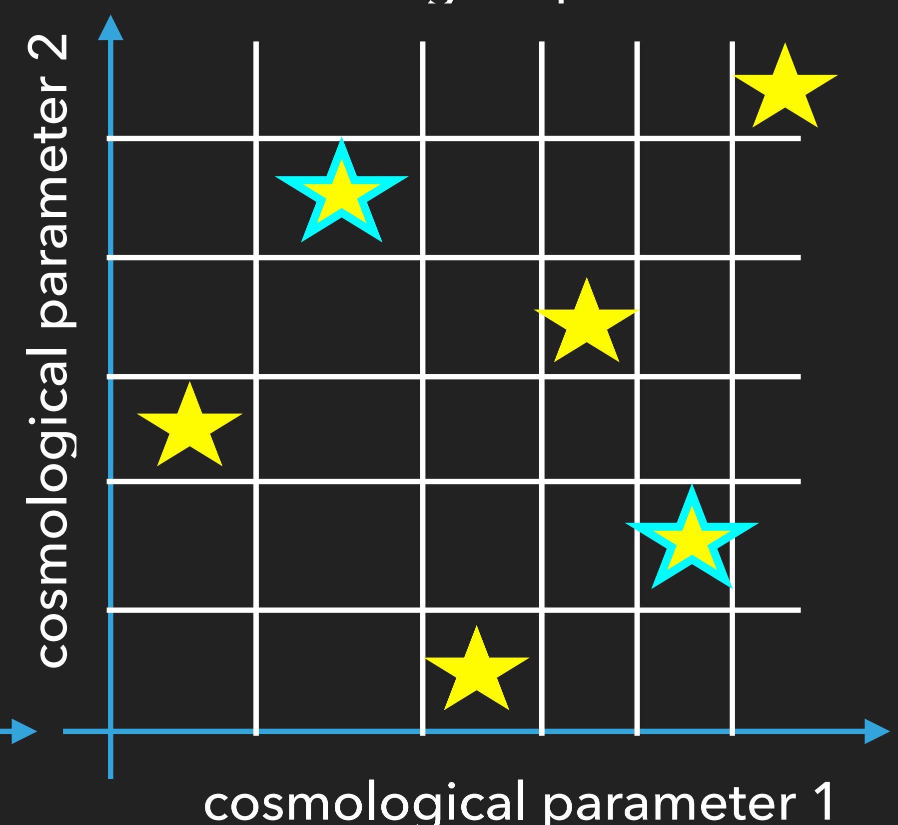
EFFICIENT SAMPLING SCHEME IN MULTI-DIMENSIONAL SPACE

- Latin hypercube designs
  - each sample point is the only one in both the row and the column
  - Such a design is not unique (ex. diagonal design)
- Additional constraint: *maxi-min distance design*
  - maximize the sum of the distances of nearest neighbors



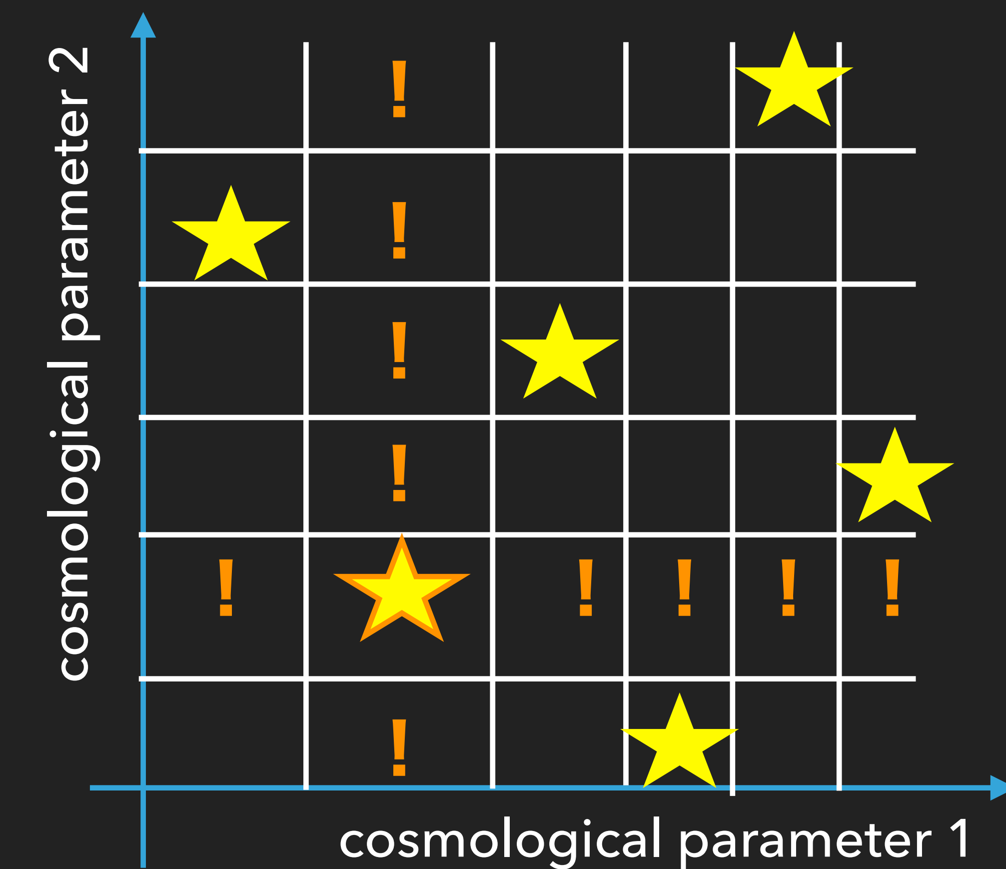




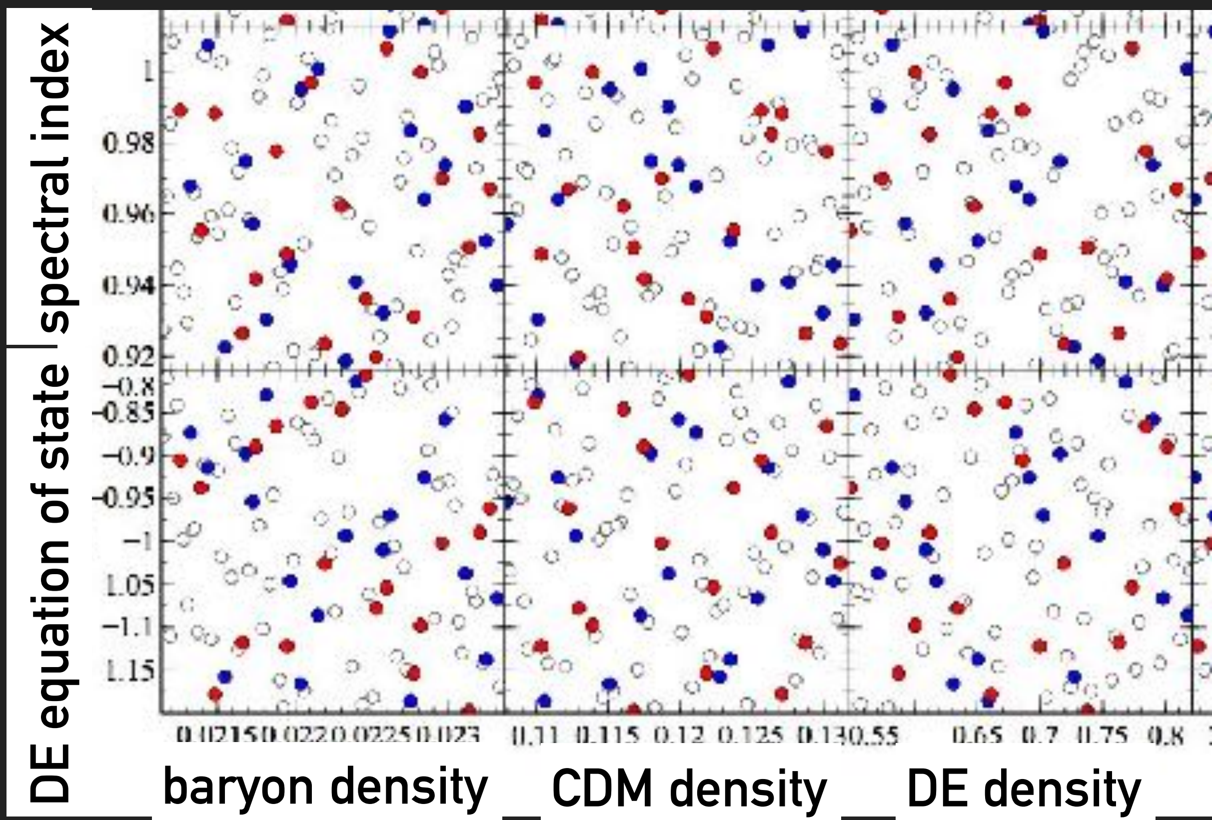


EFFICIENT SAMPLING SCHEME (CONTD)

- "Sliced" LHDs
  - a hierarchical design proposed by Ba, Brenneman&Myers'15
  - 100 sampling points in total in a LHD
  - Each of the 20 points are LHDs (e.g., red/ blue points)
  - **Multiple purpose for different slices**
    - for instance
    - 20 training set
    - 20 validation set



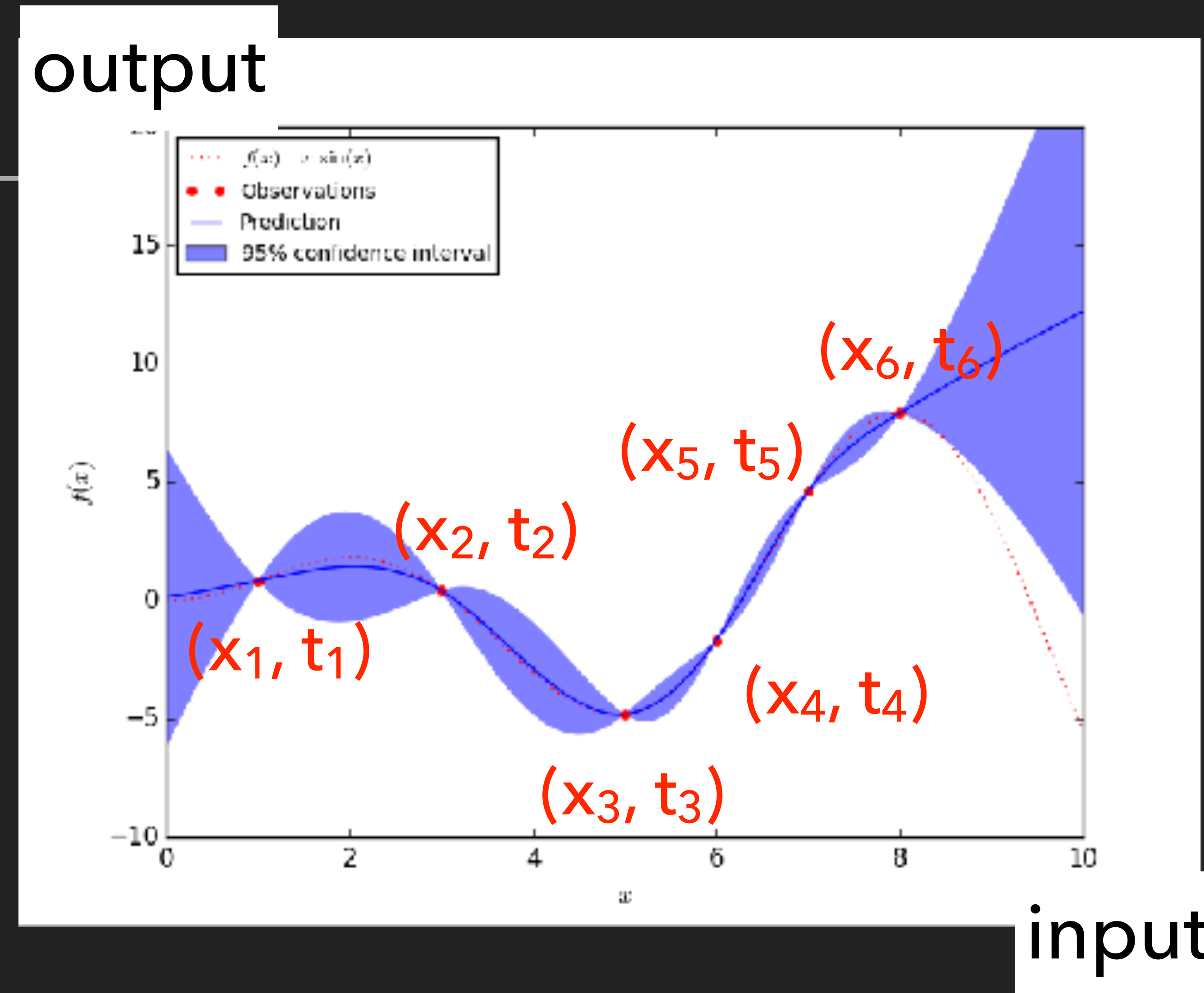
$\omega_b = \Omega_b h^2$: ±5%	$\ln(10^{10} A_s)$: ±20%
$\omega_c = \Omega_c h^2$: ±10%	n_s: ±5%
Ω_Λ: ±20%	w: ±20%



GAUSSIAN PROCESS

- A machine-learning technique that interpolates in functional space
  - non-parametric bayesian inference
  - good scaling in multi-D space
- Learn unknown "complexity" of the function from the data themselves
  - Characterization by the covariance function with a small number of hyper parameters θ
  - Estimate θ from data (x_i, t_i)

ex. $$C(\mathbf{x},\mathbf{x}';\boldsymbol{\theta}) = \theta_1 \exp\left[-\frac{1}{2}\sum_{i=1}^{I}\frac{(x_i - x_i')^2}{r_i^2}\right] + \theta_2.$$



- Given new x_{N+1} predict t_{N+1}
  - Use (x_i, t_i) and θ and solve another bayesian inference problem

$$P(t_{N+1}\,|\,\mathbf{t}_N) \propto \exp\left[-\frac{1}{2}\begin{bmatrix}\mathbf{t}_N & t_{N+1}\end{bmatrix}\mathbf{C}_{N+1}^{-1}\begin{bmatrix}\mathbf{t}_N \\ t_{N+1}\end{bmatrix}\right].$$

answer:

$$\hat{t}_{N+1} = \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{t}_N$$
$$\sigma^2_{\hat{t}_{N+1}} = \kappa - \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{k}.$$

SIMULATION SPEC

- N of particles: 2048^3
- box size: 1, 2 and 4 h^{-1}Gpc

  resolve a 10^{12} $h^{-1}M_{solar}$ halo with ~100 particles in the high resolution runs

- 2nd-order Lagrangian PT initial condition @ $z_{in}=59$

  (vary slightly for different cosmologies to keep the RMS displacement about 25% of the inter-particle separation)

- Tree-PM force by L-Gadget2 (w/ 4096^3 PM mesh)
- 21 outputs in $0 \leqq z \leqq 1.5$ (equispaced in linear growth factor)

- Data compression (256GB -> 48GB par snapshot)
  - positions -> displacement (16 bits par dimension; accuracy ~1h^{-1}kpc)
  - velocity: discard after halo identification
  - ID: rearrange the order of particles by ID and then discard
  - consuming ~200TB (~observational data)

SIMULATION PIPELINE

~2TB, 2+2days / 1 run

on analysis servers

prediction

density estimate

final catalog

data analysis

mass/profile determination

time evolution

Gadget2

rockstar

IC generator

subfind

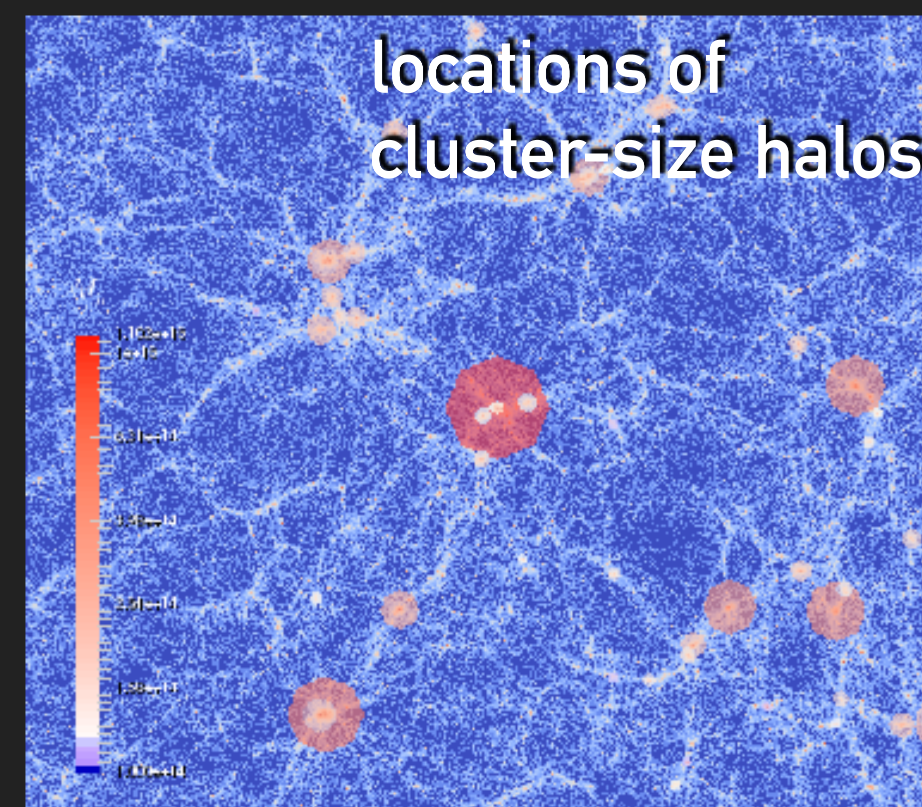
create "universes"

FOF (on the fly)

object identification

on supercomputer

(XC30 @NAOJ; using 648 CPU cores)



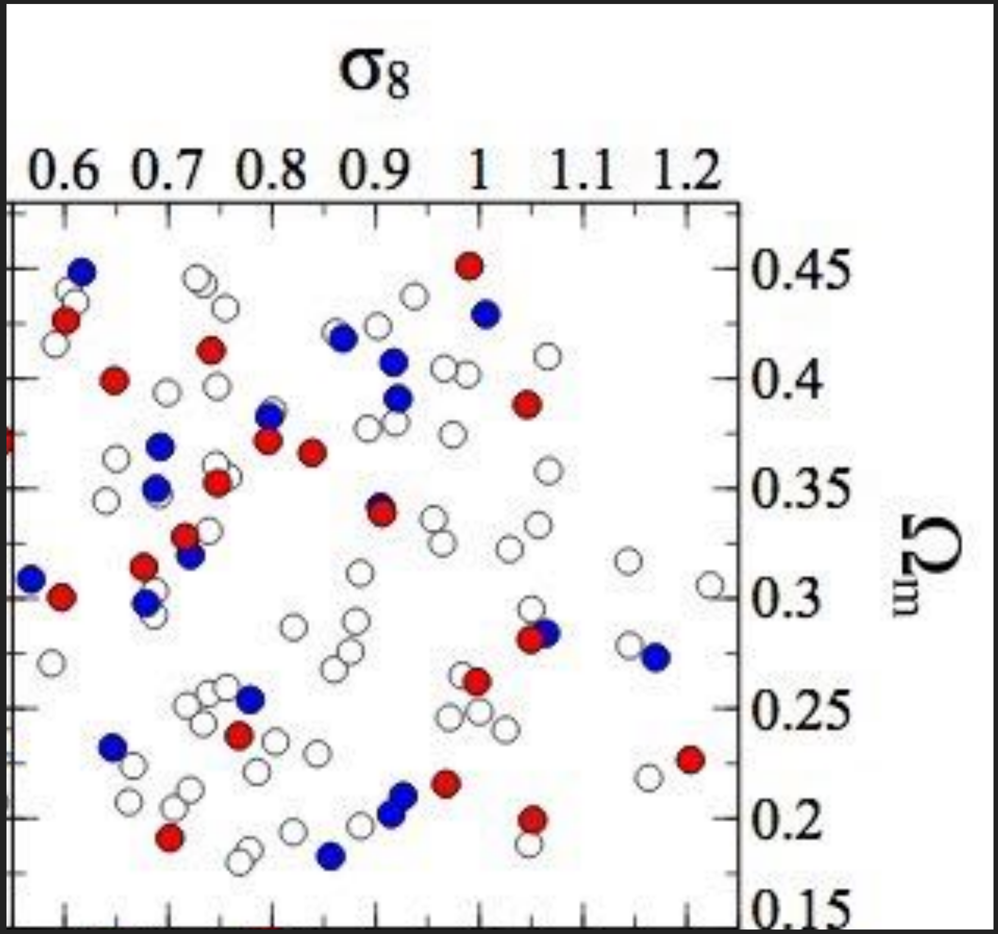
- identify and store more than ***10 billion*** halos in total

SIMULATION STATUS

122 sims are available in total

fiducial model

- PLANCK15 flat ΛCDM
- 24+1 realizations done
- test of statistical error
- tests/development of analysis codes



varied cosmology

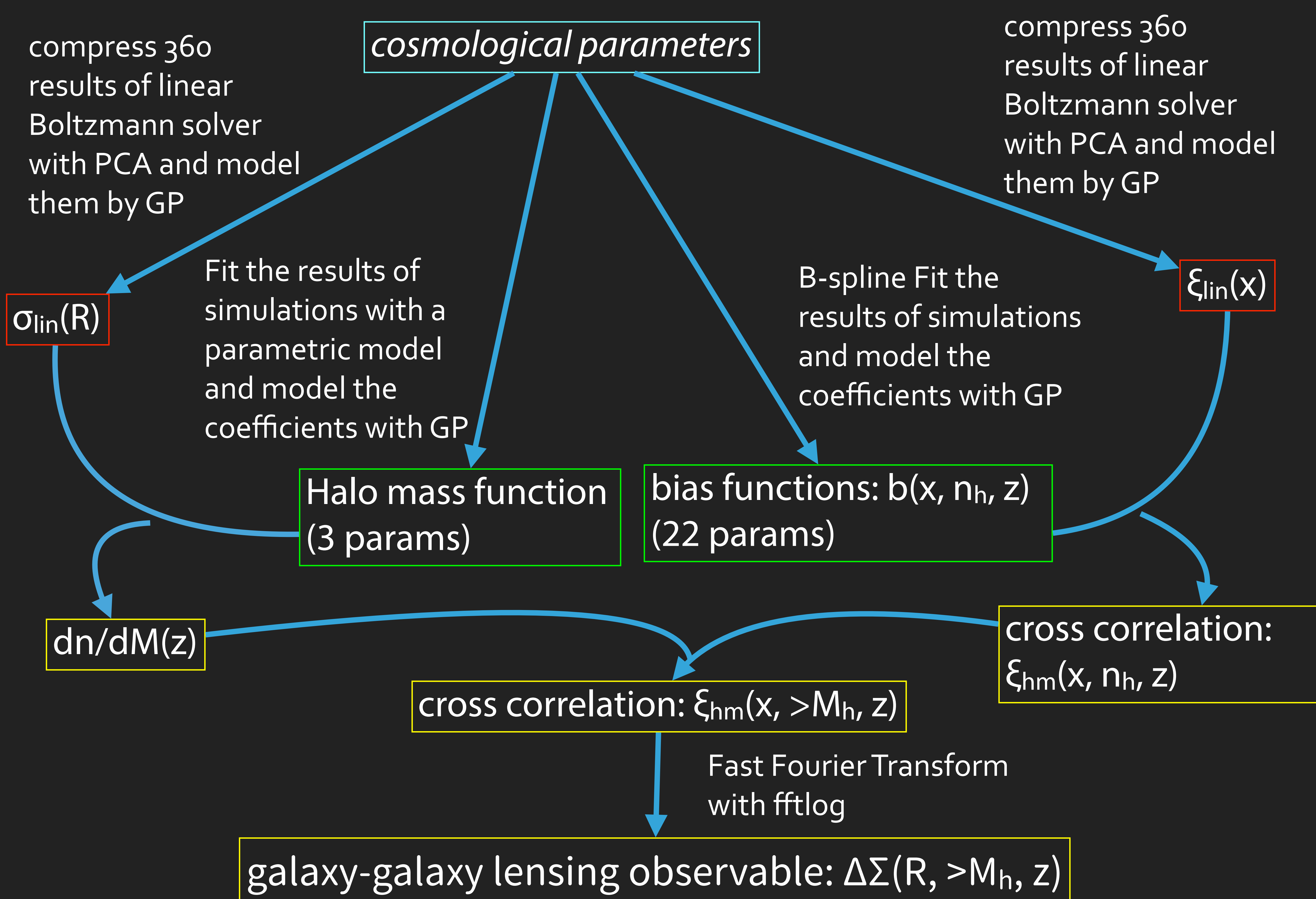
high resolution runs

- g-g lensing
- 2 initial conditions
  - fixed random phase (20 done red)
  - varied random phase (40 done red+blue)

low resolution runs

- calibration of largest scales
- 37 realizations done
- in progress (60 white points)

EMULATOR DESIGN



jupyter dsigma_demo Last Checkpoint: 3 minutes ago (autosaved)

File Edit View Insert Cell Kernel Help Python 2

Read out pre-computed PCA basis function + GP

In [3]:
```
basis = [[joblib.load('basis/S%03d_dens%02d_basis.pkl' % (snap,ndens)) for ndens in range(0,10)] for snap in range(0,21)]
dsig_gp = [[joblib.load('gp/S%03d_dens%02d_gp.pkl' % (snap,ndens)) for ndens in range(0,10)] for snap in range(0,21)]
```

Prepare a table for 3D spline

In [96]:
```
def prep_dsigma_template(params):
    global d_gp
    d_gp = np.zeros((21,10,81))
    for snap in range(0,21):
        for ndens in range(0,10):
            d_gp_pca=dsig_gp[snap][ndens].predict(params)
            d_gp[snap,ndens,] = basis[snap][ndens].inverse_transform(d_gp_pca)[0,]
```

Inputs:

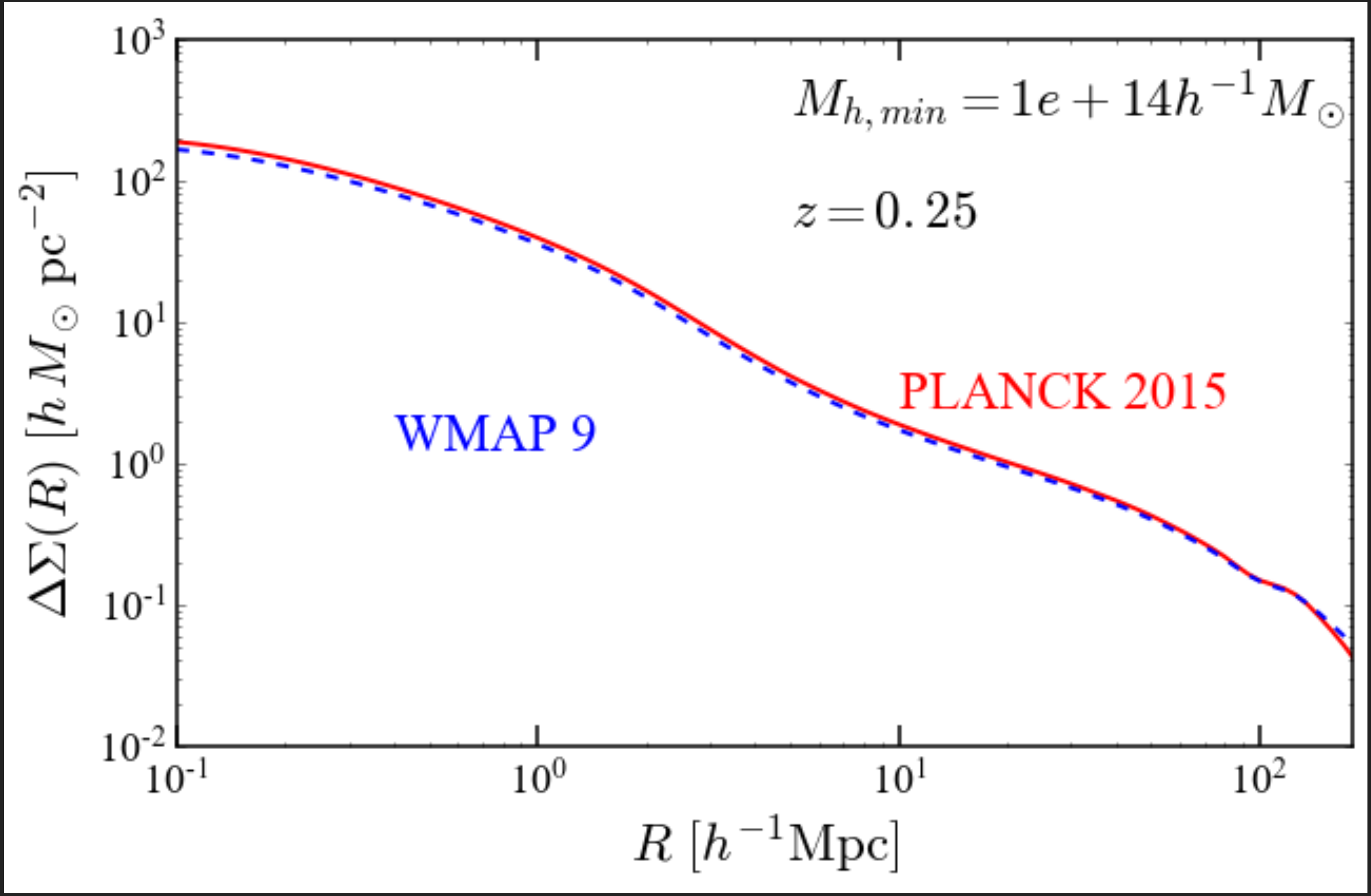
- scale factor (time)
- number density (or halo mass)
- projected distance

In [59]:
```
z=0.25
ascale = 1./(1+z)
Rplot = np.logspace(-1,2.3,100) # in h^{-1}Mpc
Mmin = 1e14 # in h^{-1}M_\odot
```

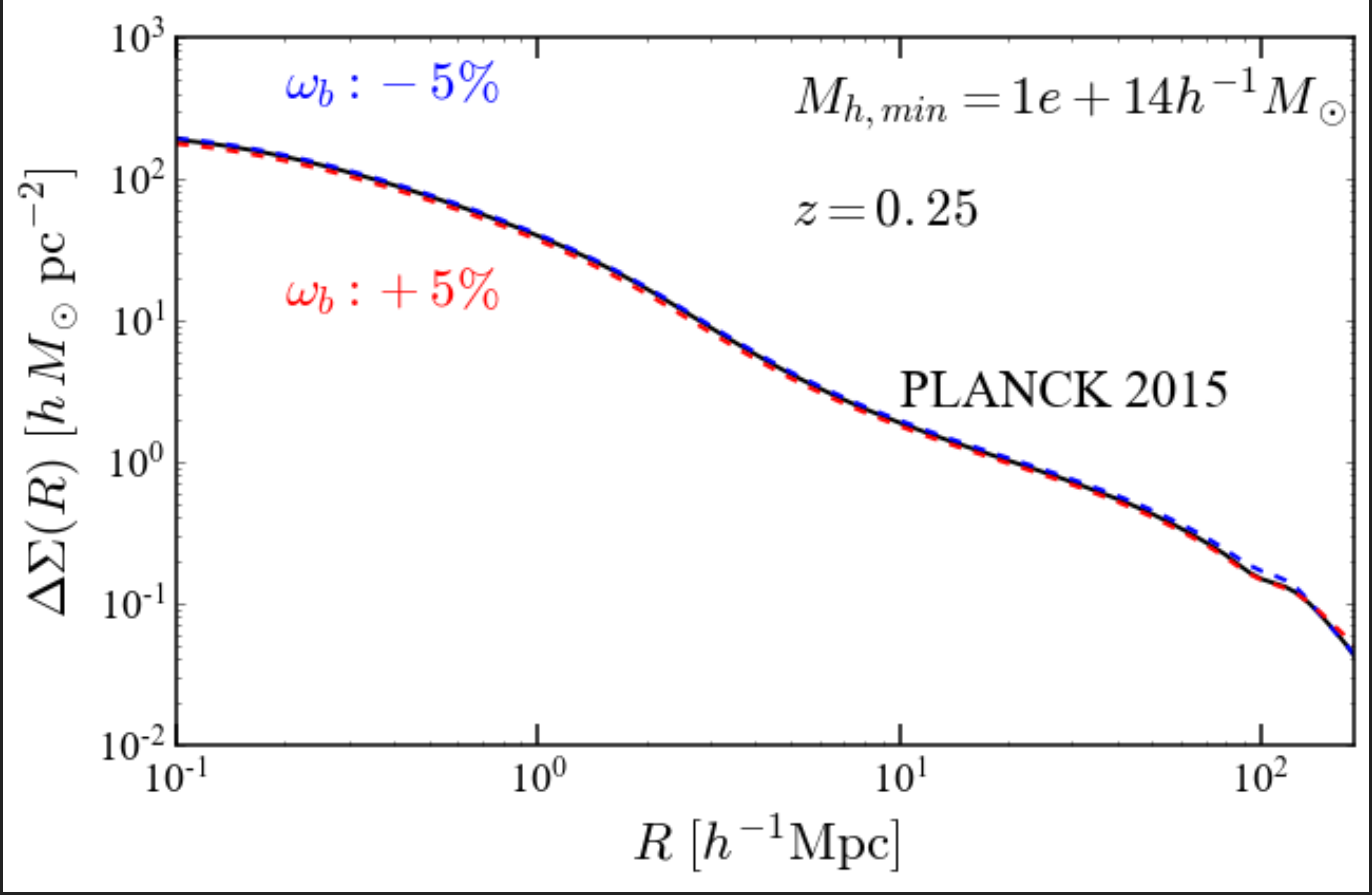
```
# PLANCK 2015
cparam = np.array([[0.02225,0.1198,0.6844,3.094,0.9645,-1]])
set_cosmo(cparam)
set_redshift(z)
lognh = mh_to_logdens(Mmin)
plt.loglog(Rplot,get_dsigma(ascale, lognh, Rplot),lw=2,color='red')
```

- set_cosmo(cparam): give your cosmological params ~5s
- set_redshift(z): and redshifts ~600ms; HMF GP called inside
- lognh = mh_to_logdens(Mmin): convert M_min to n_h ~50μs
- get_dsigma(ascale, lognh, Rplot): Evaluate !! ~1ms

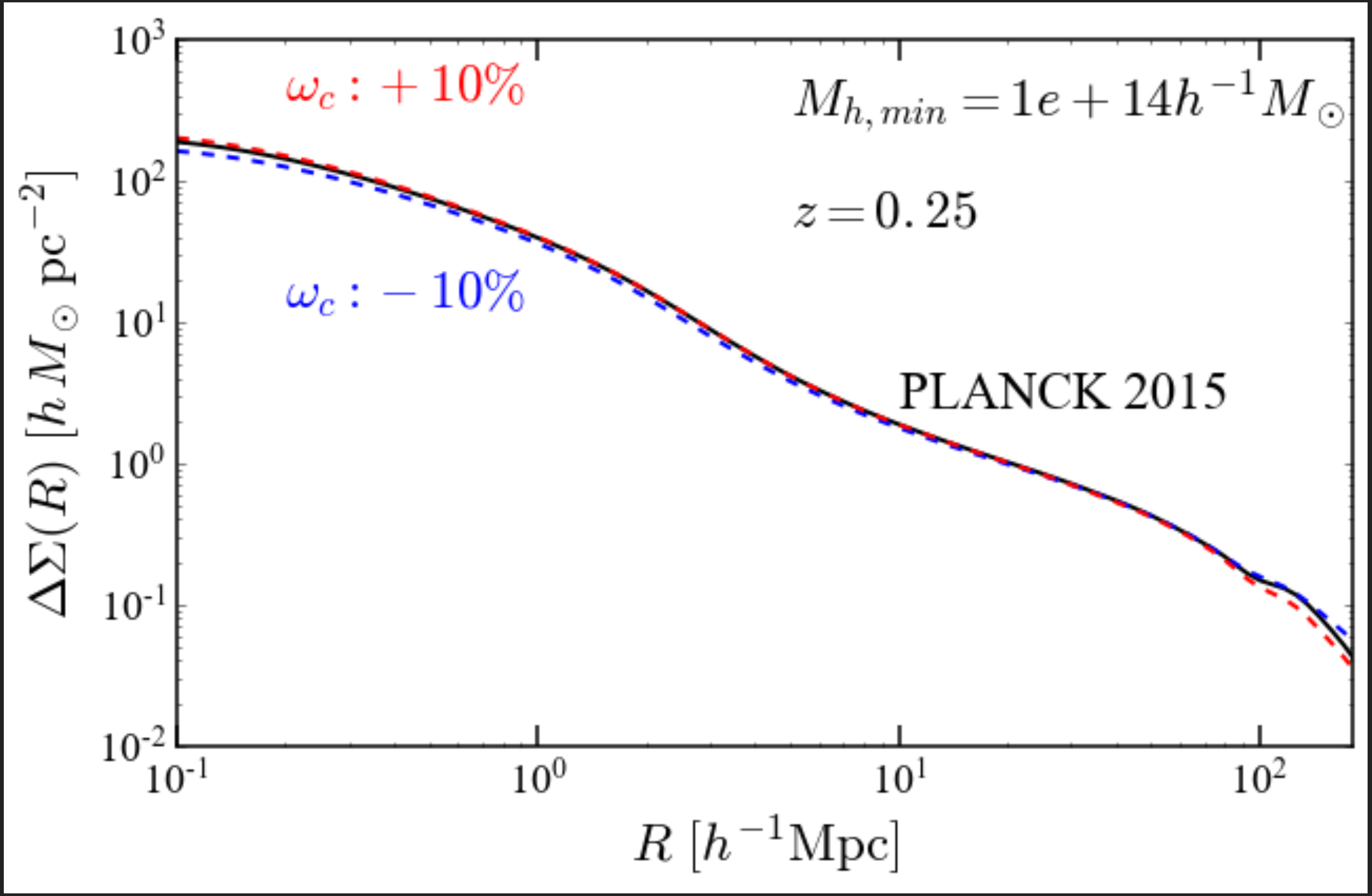
OUR ΔΣ EMULATOR DEMONSTRATIONS



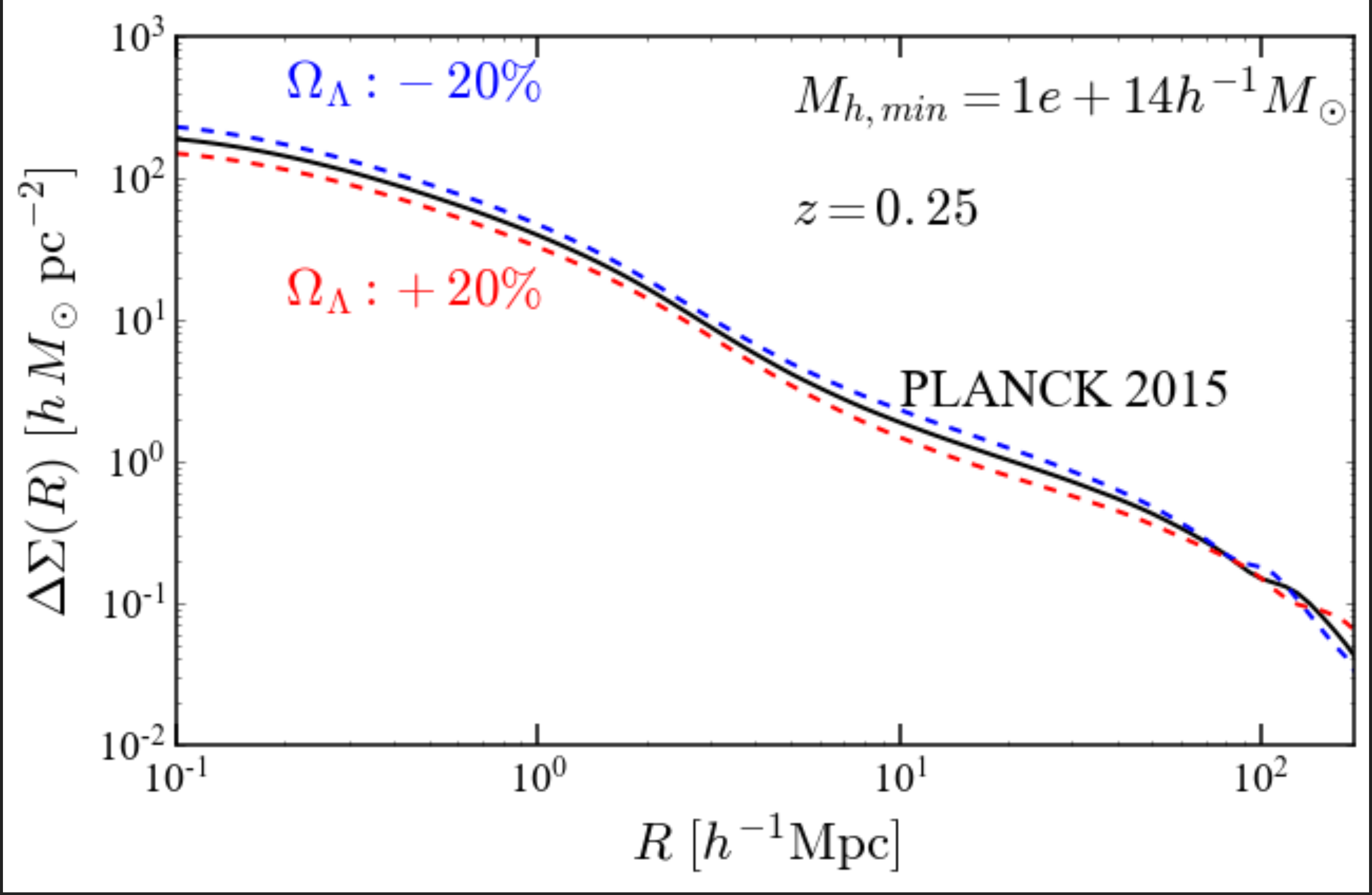
OUR ΔΣ EMULATOR DEMONSTRATIONS



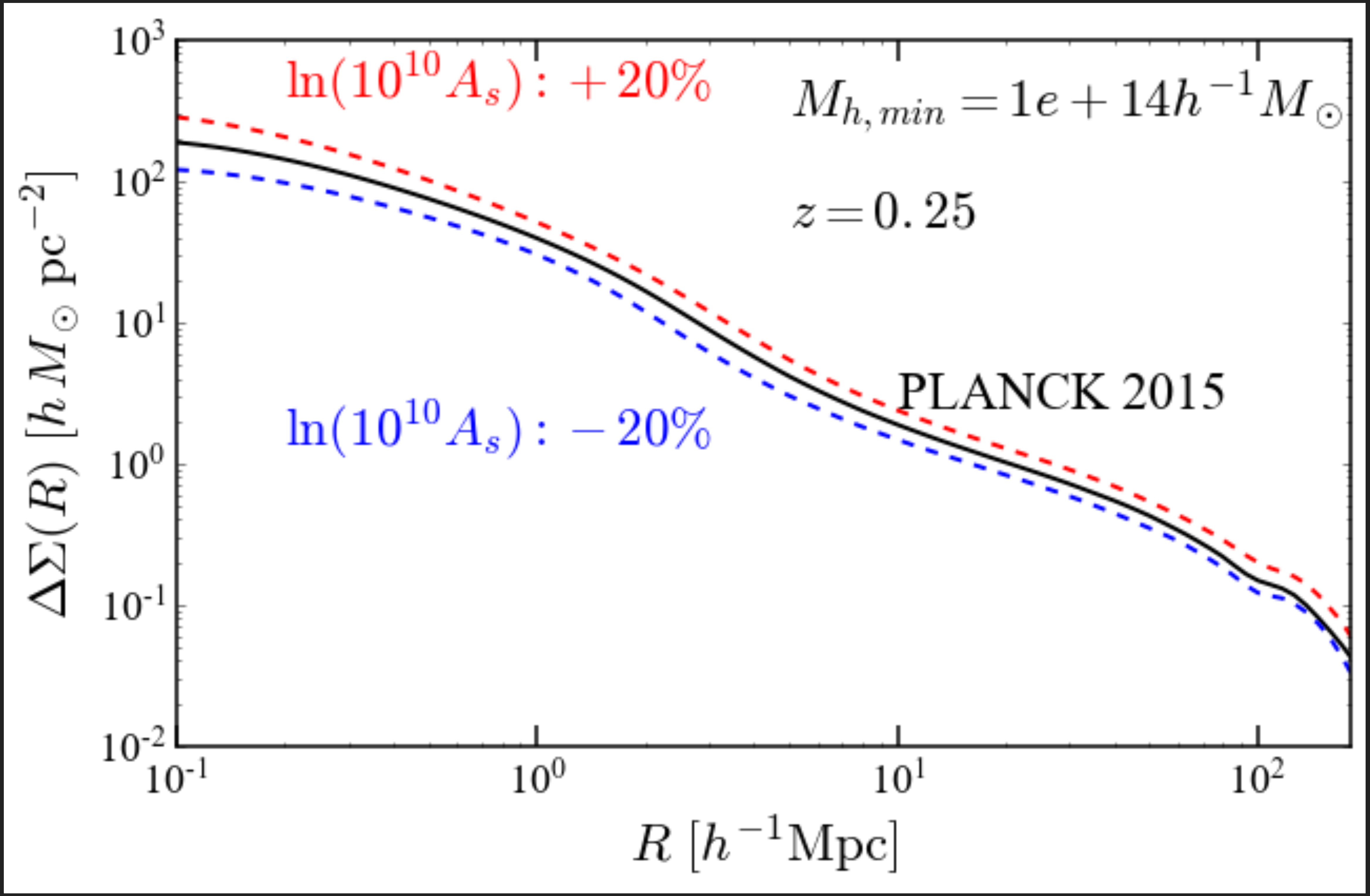
OUR ΔΣ EMULATOR DEMONSTRATIONS



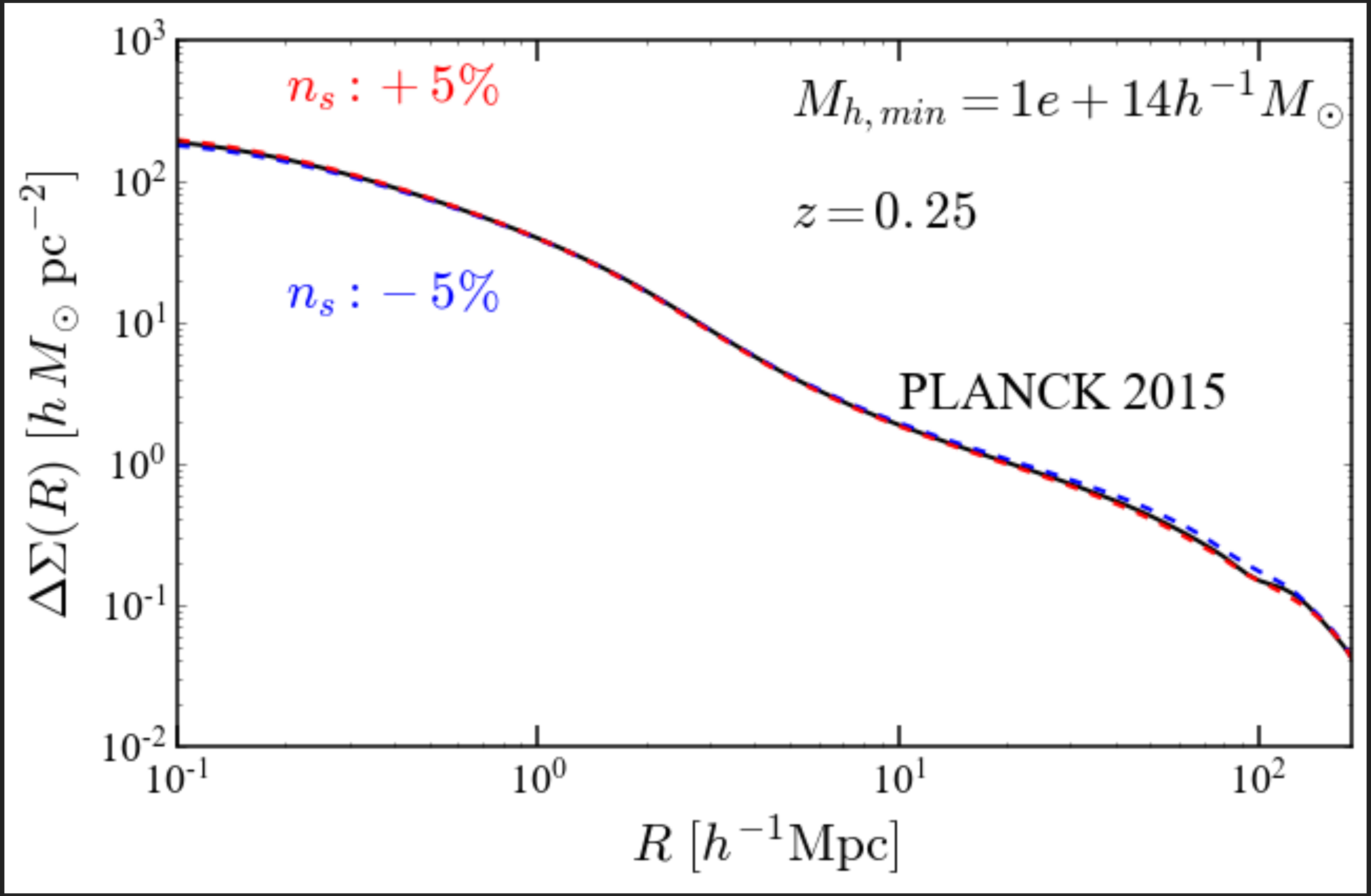
OUR ΔΣ EMULATOR DEMONSTRATIONS



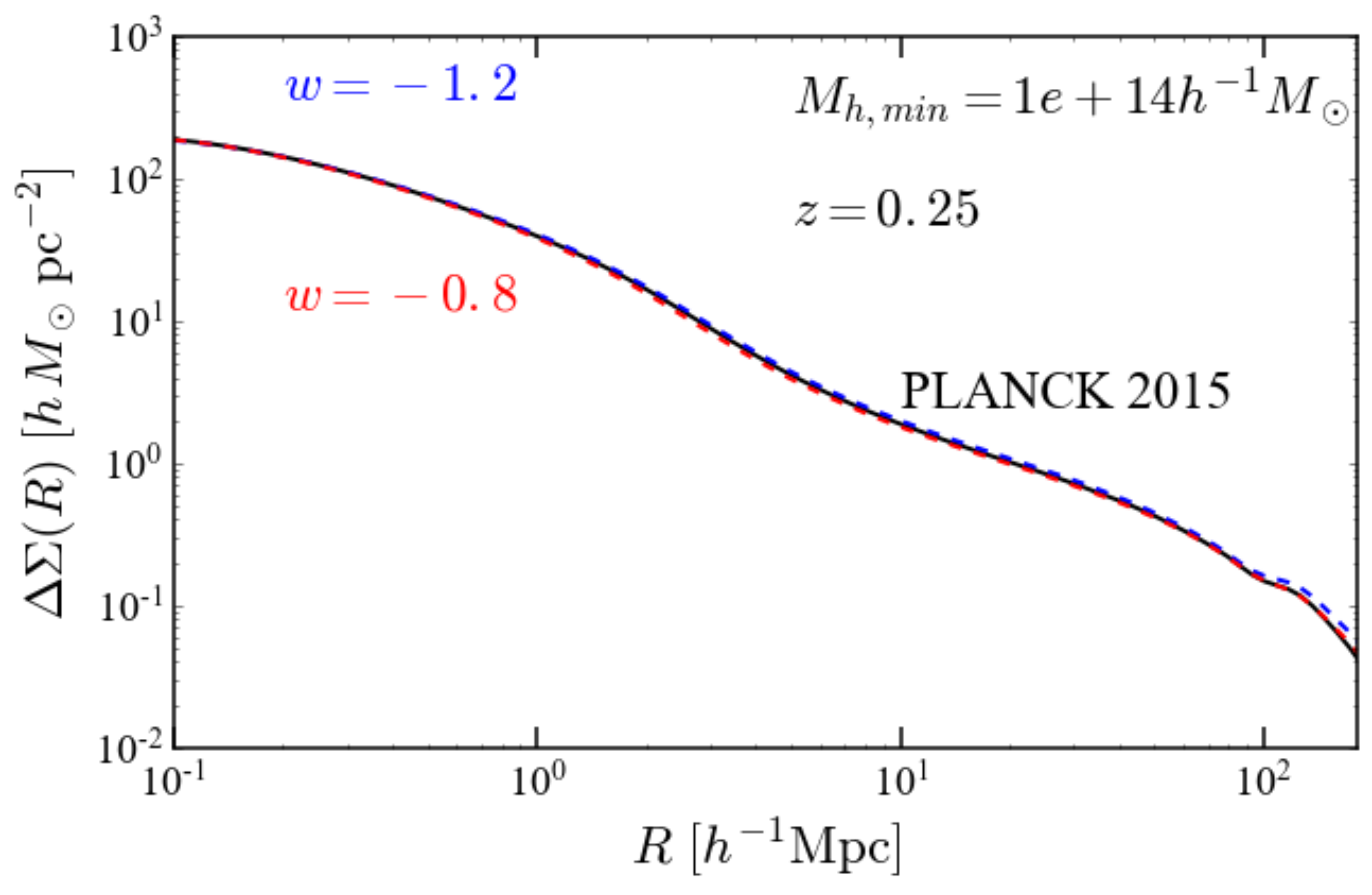
OUR ΔΣ EMULATOR DEMONSTRATIONS



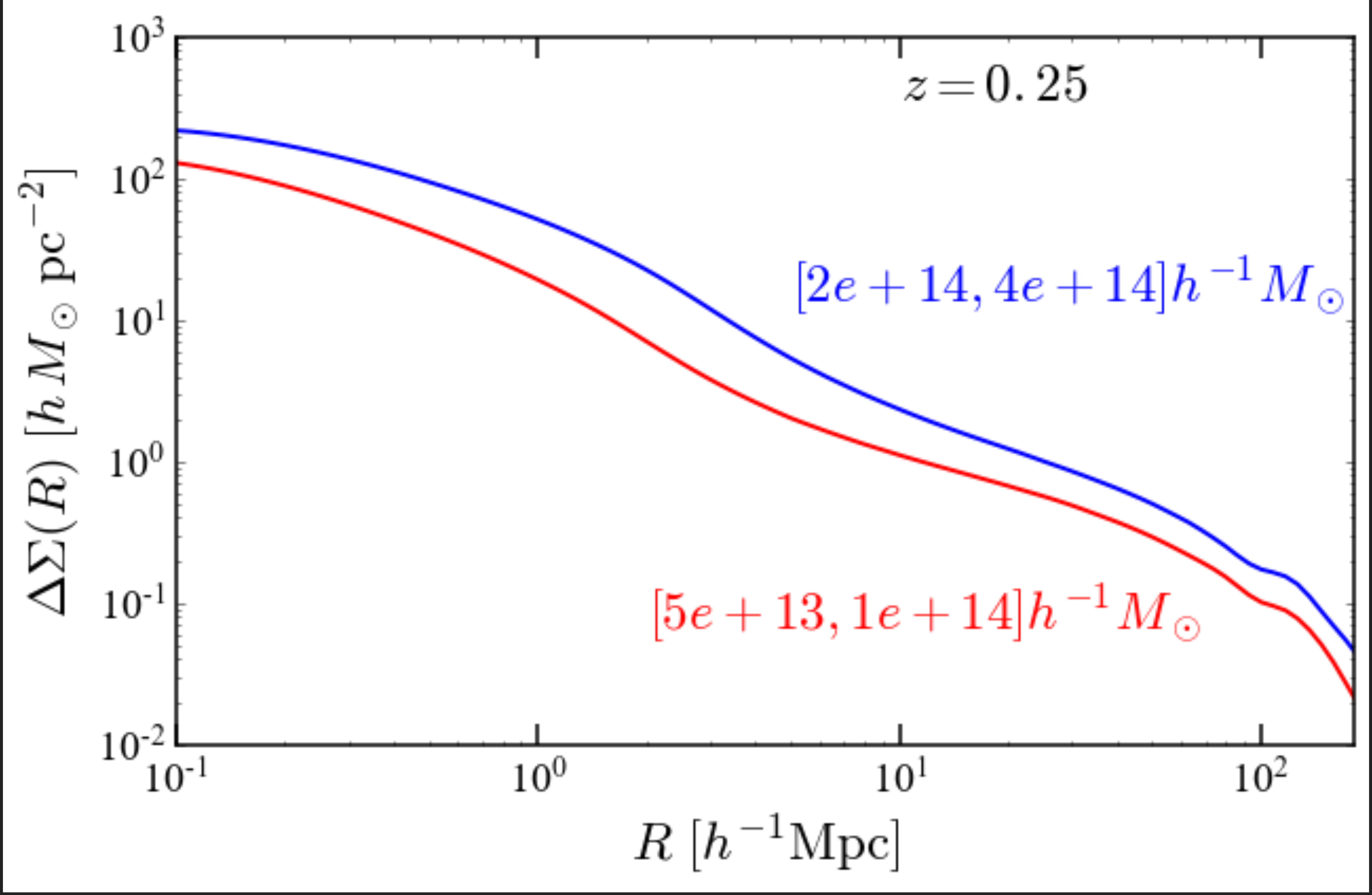
OUR ΔΣ EMULATOR DEMONSTRATIONS



OUR ΔΣ EMULATOR DEMONSTRATIONS



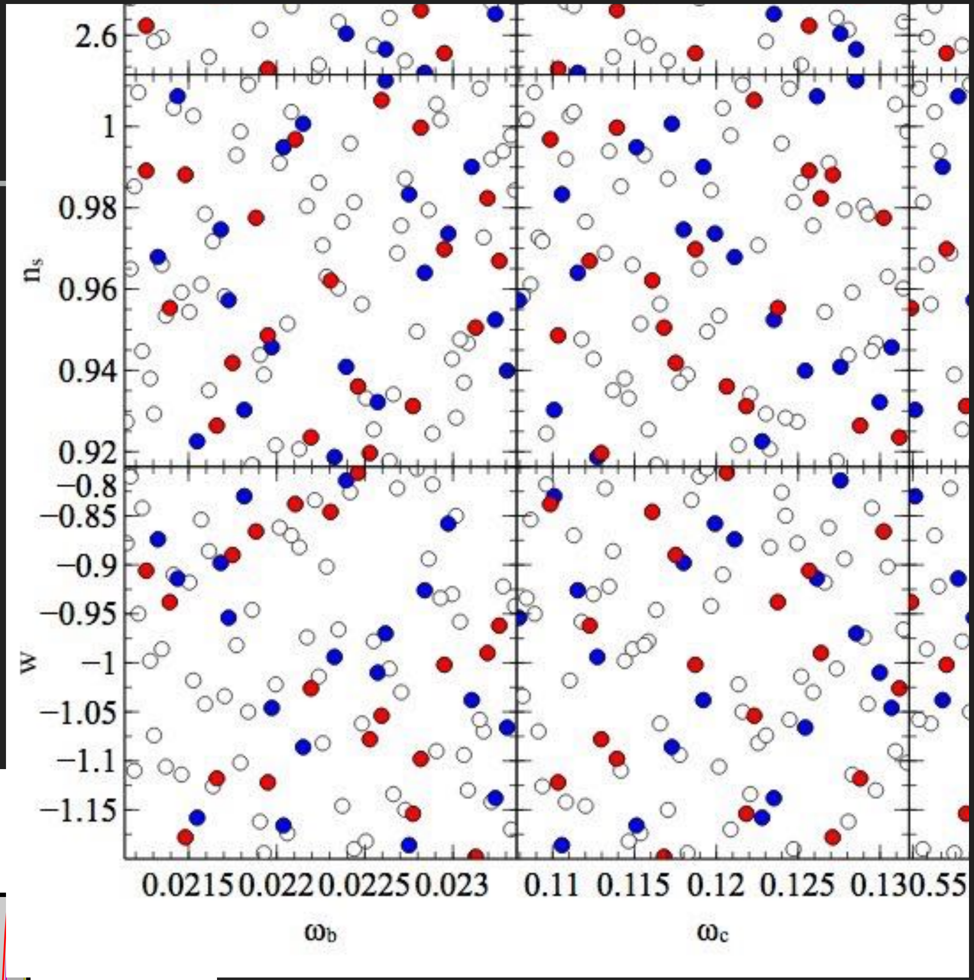
OUR ΔΣ EMULATOR DEMONSTRATIONS



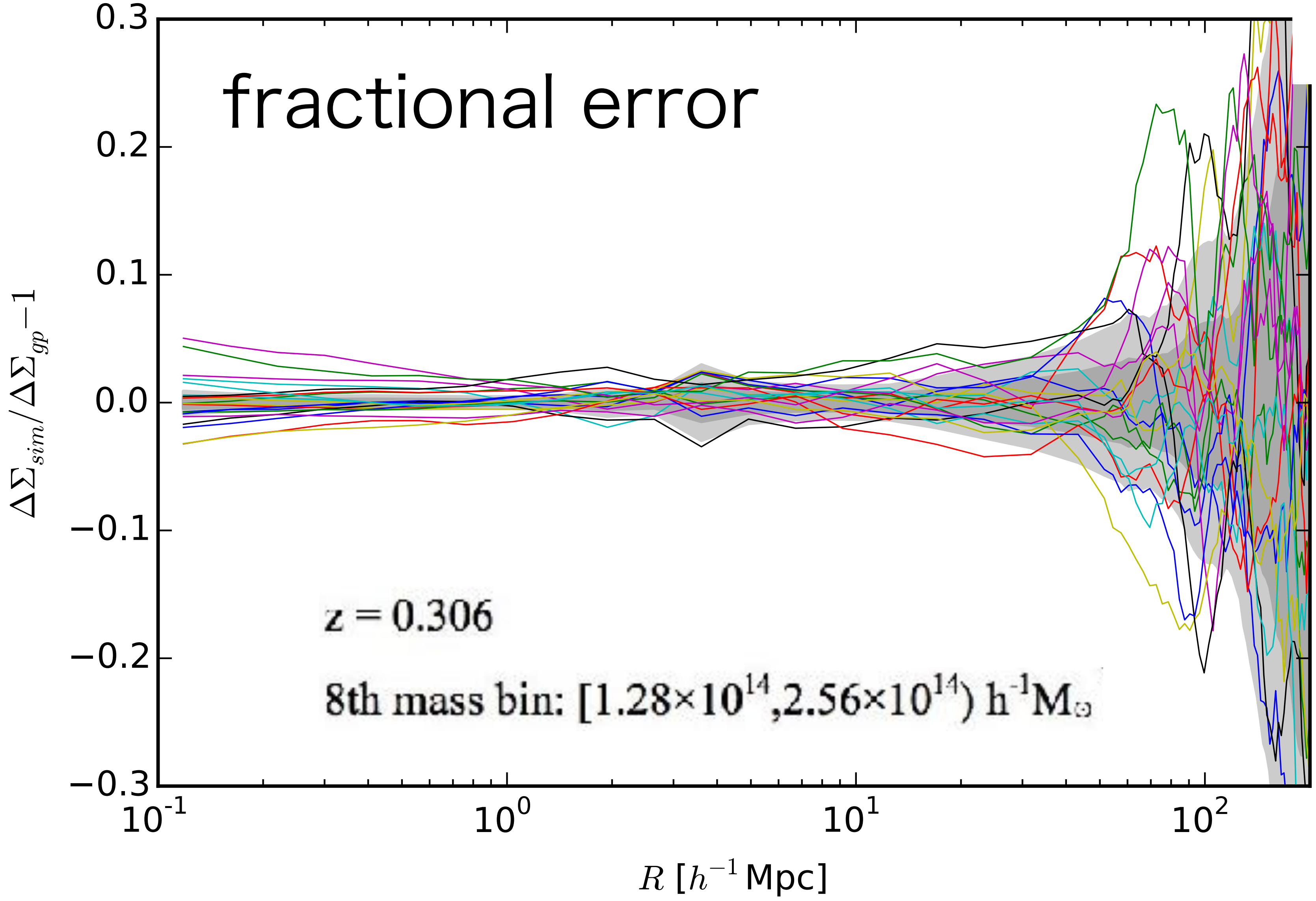
エミュレータ

GAUSSIAN PROCESS ACCURACY

training with 20 models (red)
validation with 20 other models (blue)



random number varied



SUMMARY

- Modeling the halo mass function and galaxy-galaxy lensing signal
  - Latin hypercube design + fitting/GP/spline
  - handy emulator in python almost ready
  - accuracy test undergoing, roughly 2-3% accuracy
- To come
  - Apply to real data and extract cosmological information
  - RSD emulator to combine g-g lensing and 3D clustering
  - further extension under discussion
    - e.g., non-flat, w0-wa cosmologies

MCMC cosmology with N-body emulators

R. Murata

Average DM distribution around massive clusters