Prospects for CMB lensing-galaxy clustering cross-correlations & initial condition reconstruction

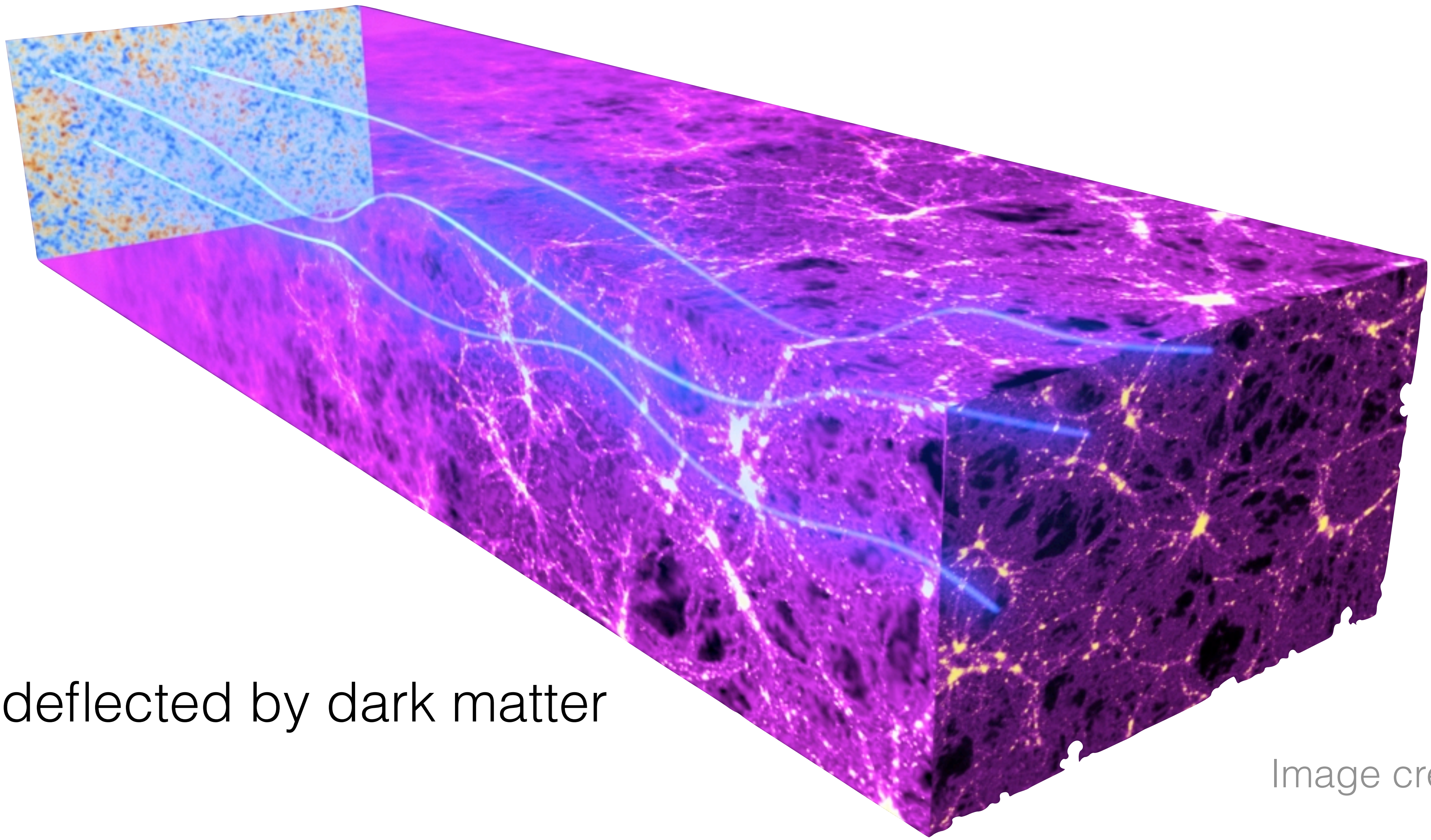
Marcel Schmittfull
Institute for Advanced Study

With Tobias Baldauf, Uros Seljak, and Matias Zaldarriaga

LIneA webinar, August 30th 2018

CMB lensing

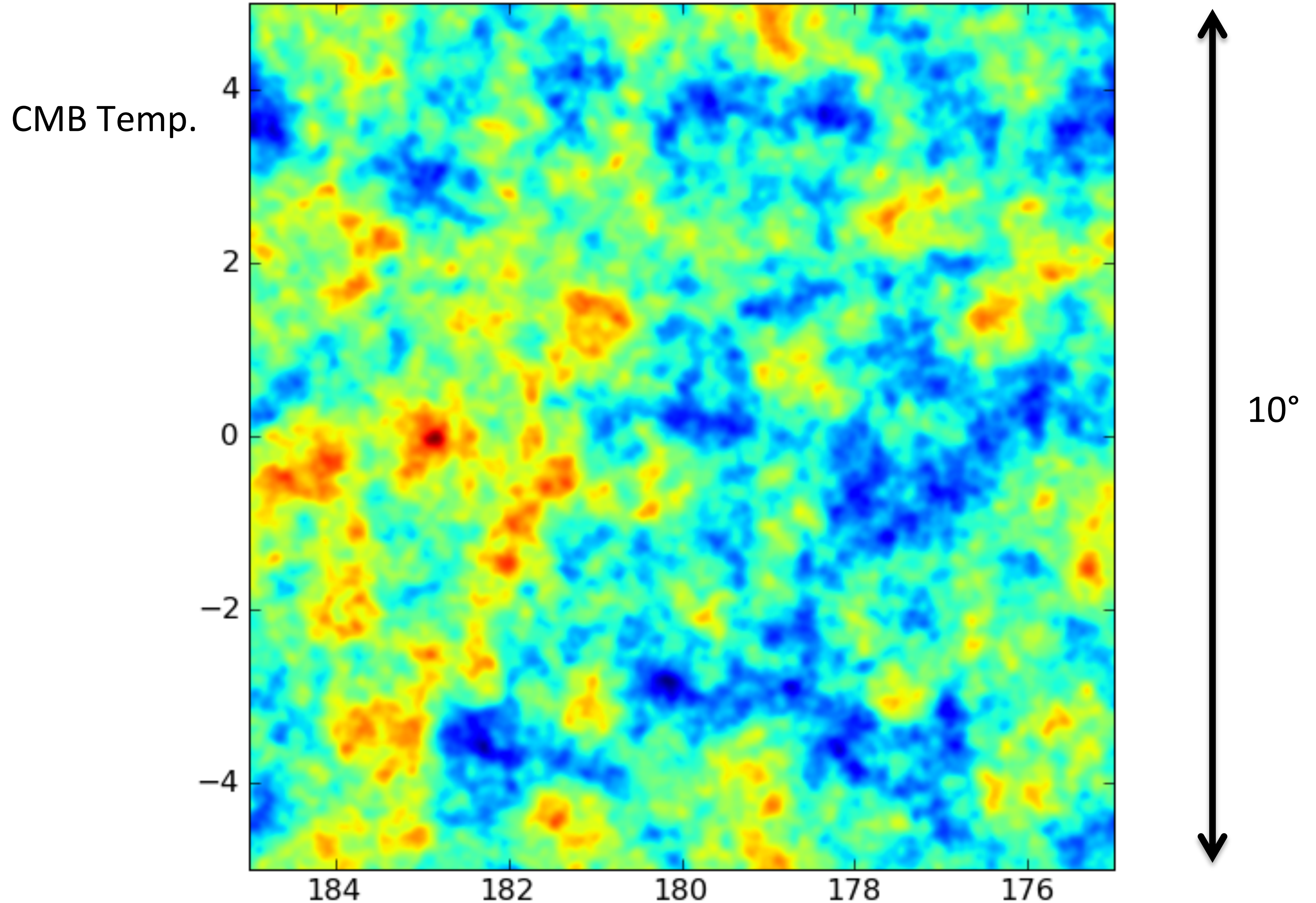
Oldest light we can observe: CMB



deflected by dark matter

Image credit: ESA

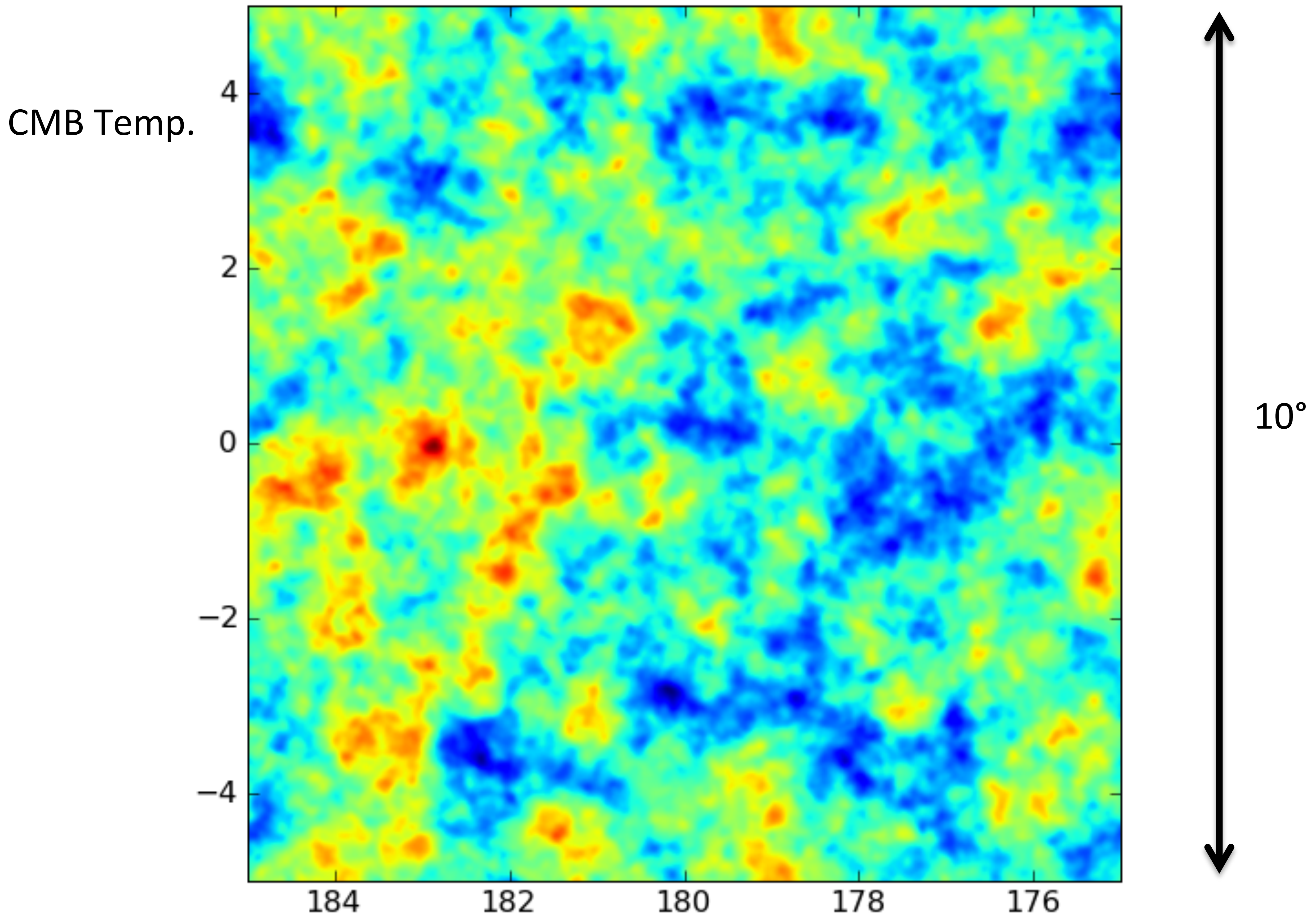
Unlensed CMB



$T(\hat{\mathbf{n}})_{\mathrm{unlensed}}$

Slide credit: Blake Sherwin

Lensed CMB



$$T(\hat{\mathbf{n}})_{\rm lensed} = T(\hat{\mathbf{n}} + \mathbf{d}(\hat{\mathbf{n}}))_{\rm unlensed}$$

Slide credit: Blake Sherwin

Many unlensed patches

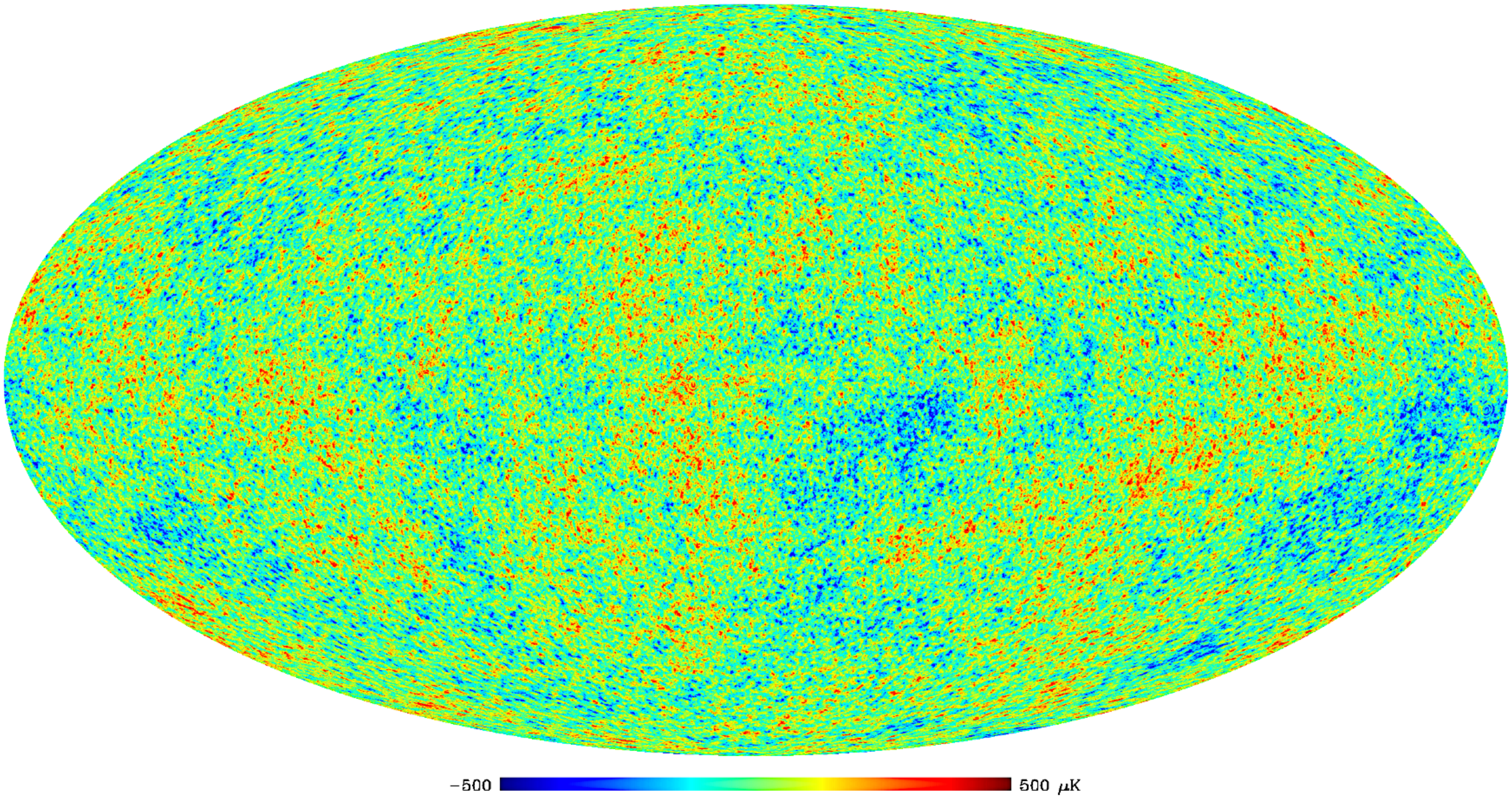
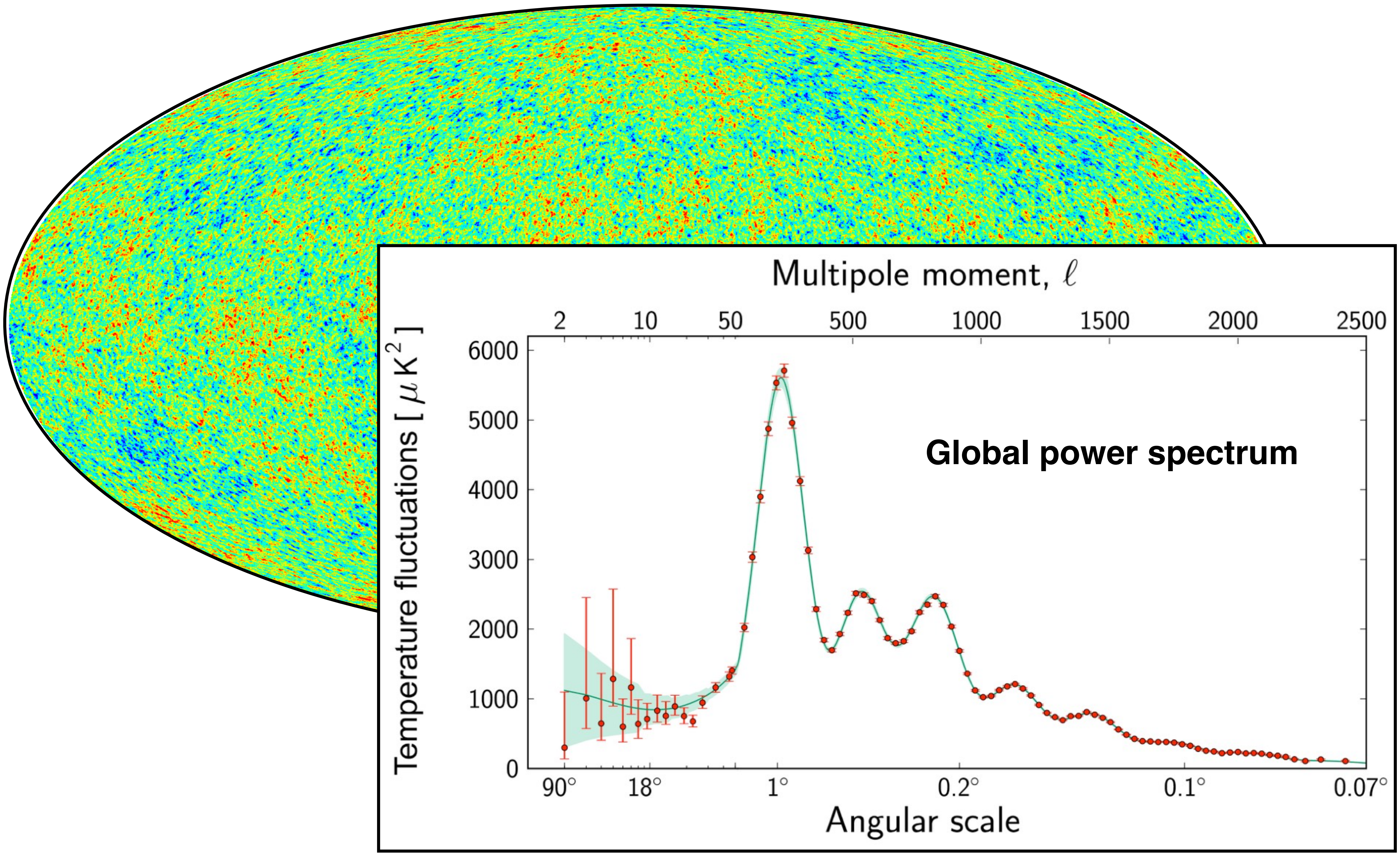
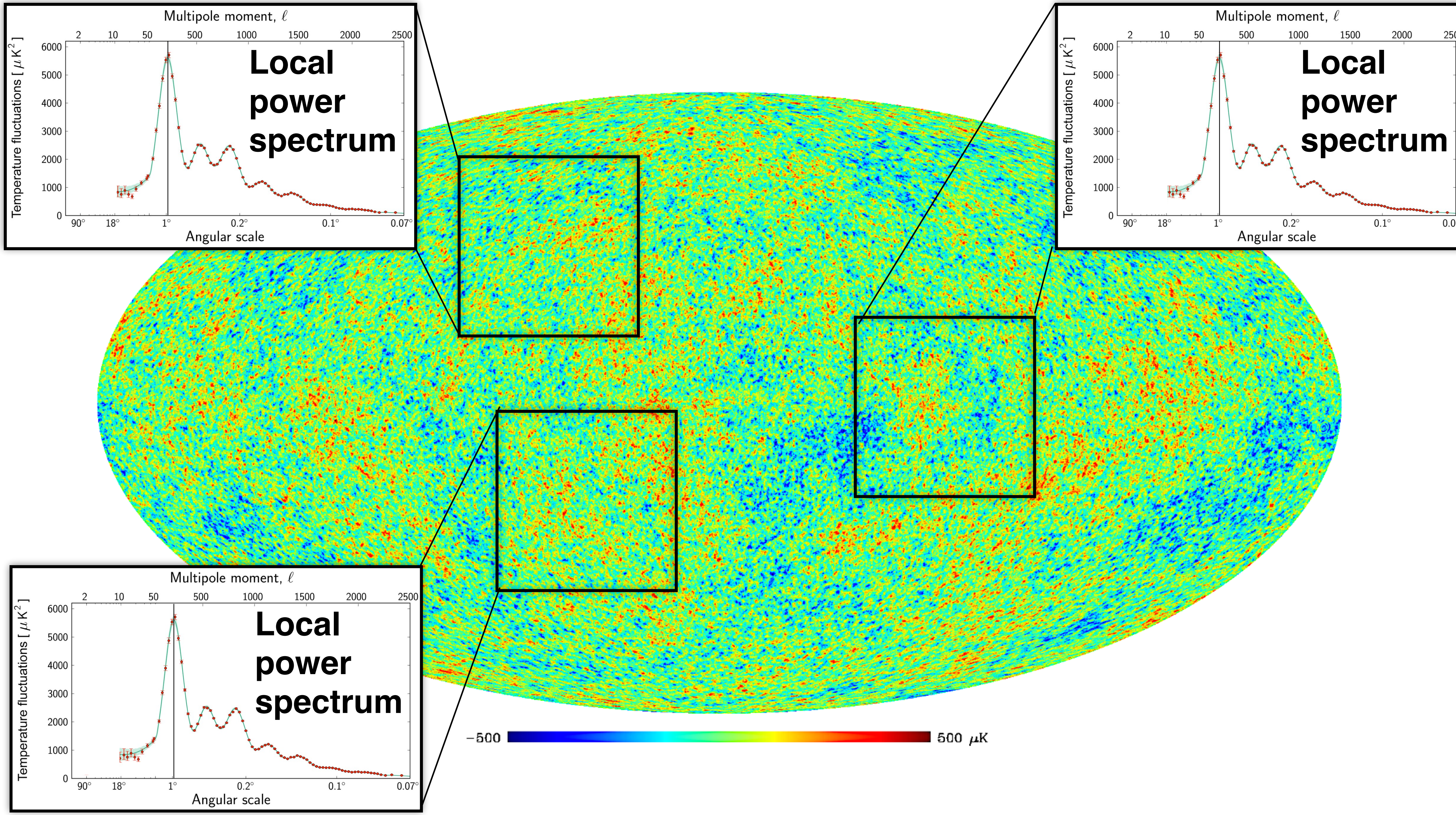


Image credit: cosmostat.org

Many unlensed patches

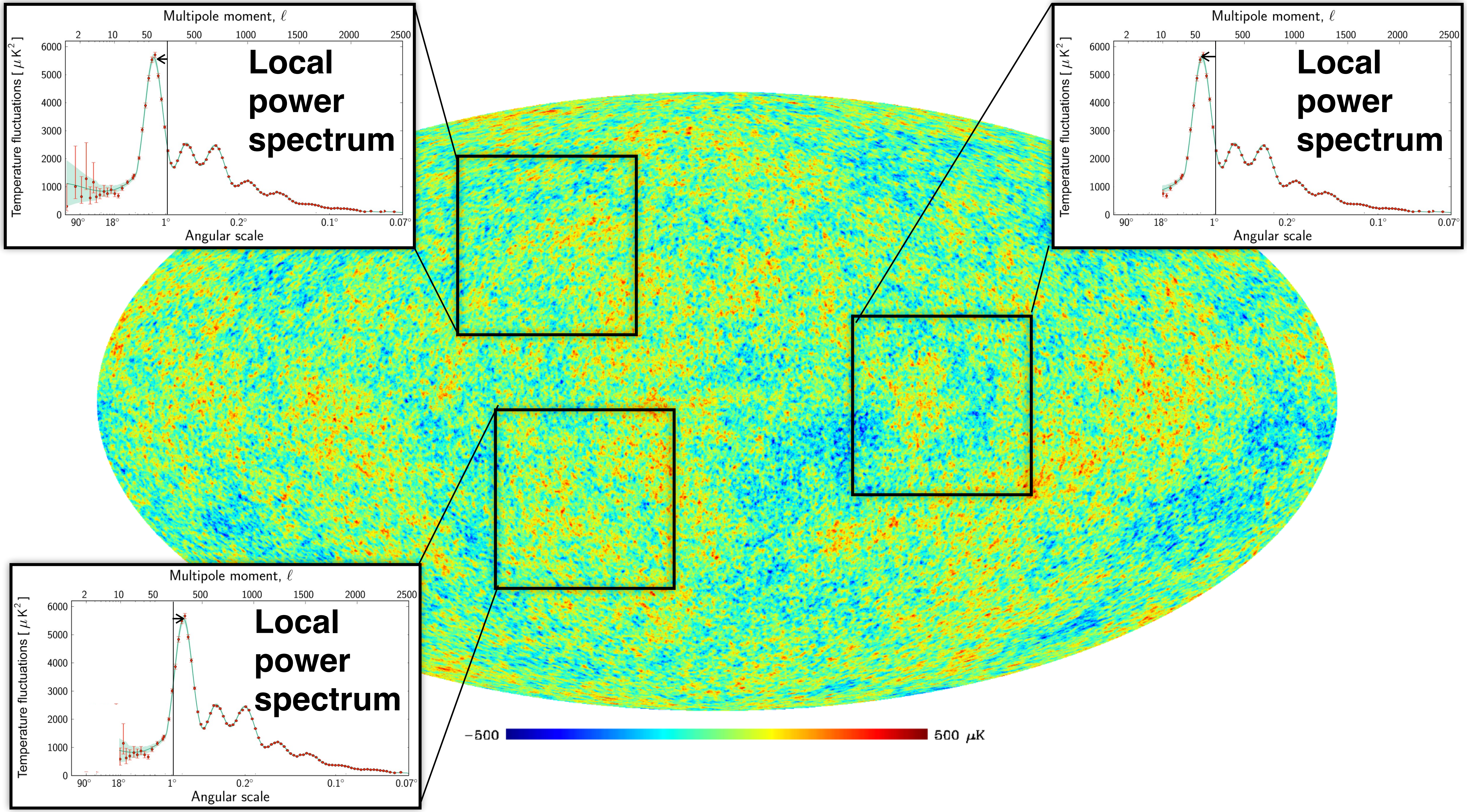


Many unlensed patches



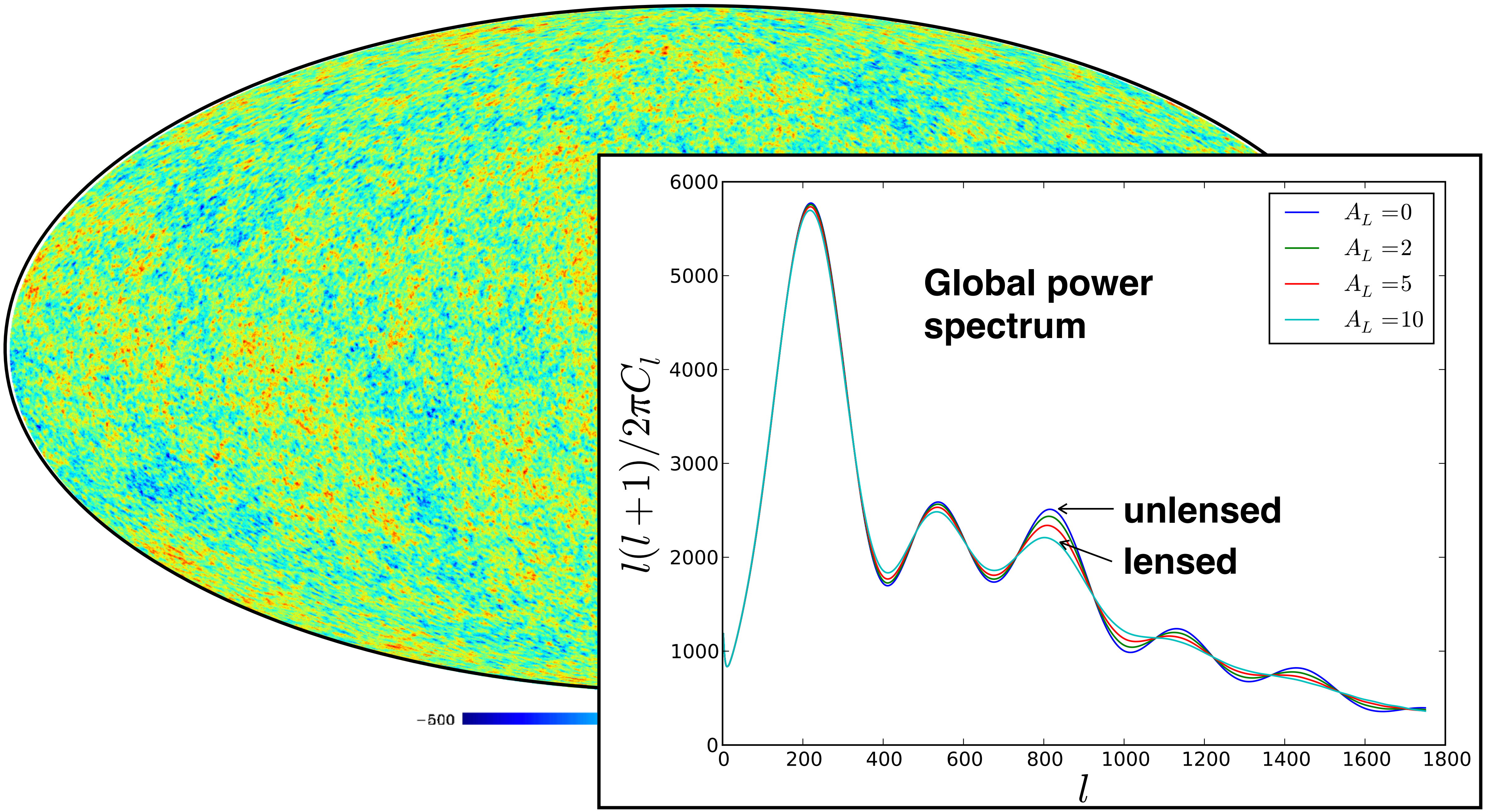
Local power spectrum is the same in each patch

Many lensed patches



Local power is magnified or de-magnified

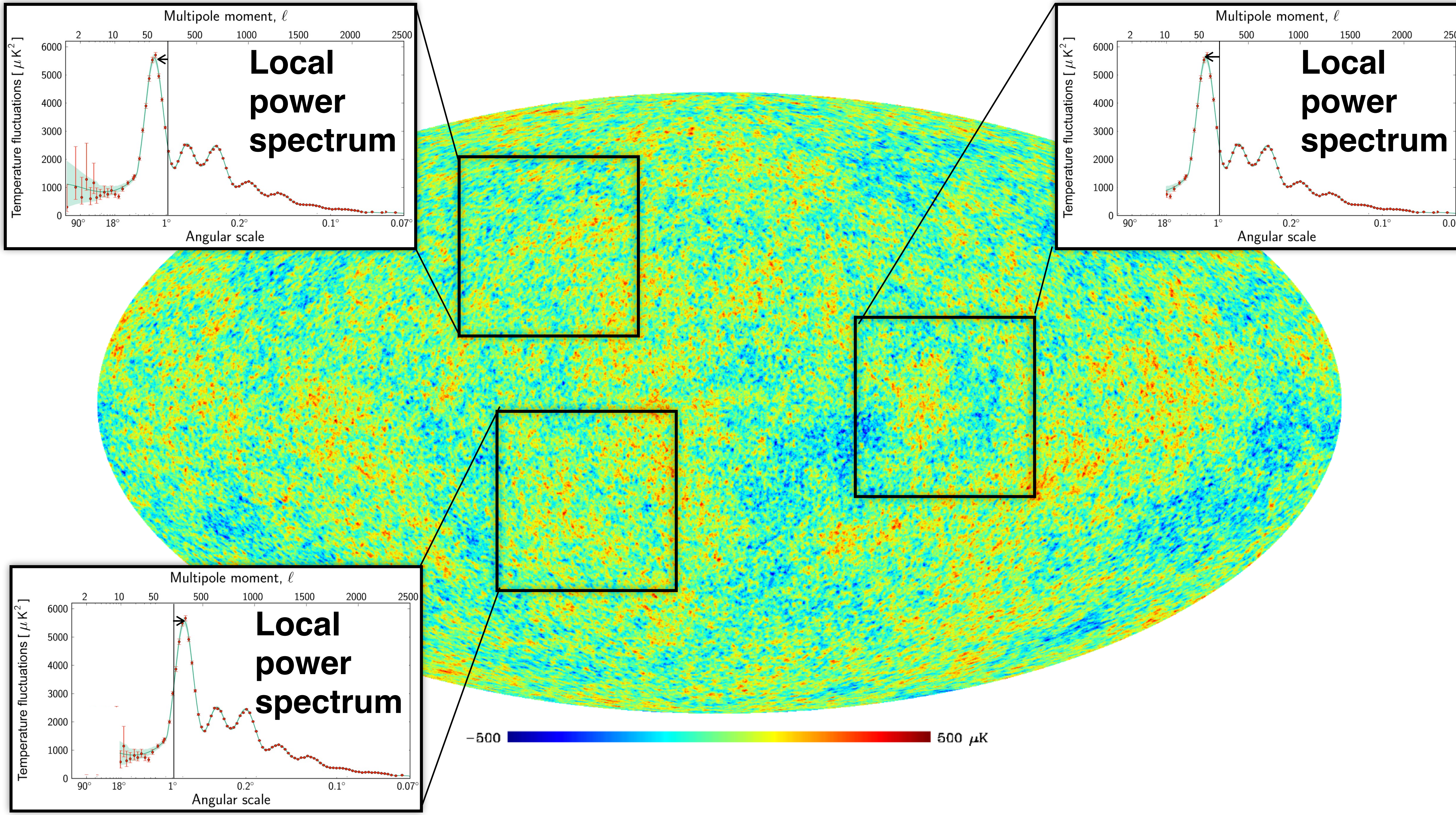
Many lensed patches



Acoustic peaks of global power are smeared out

Image credit: Smidt+ (2010)

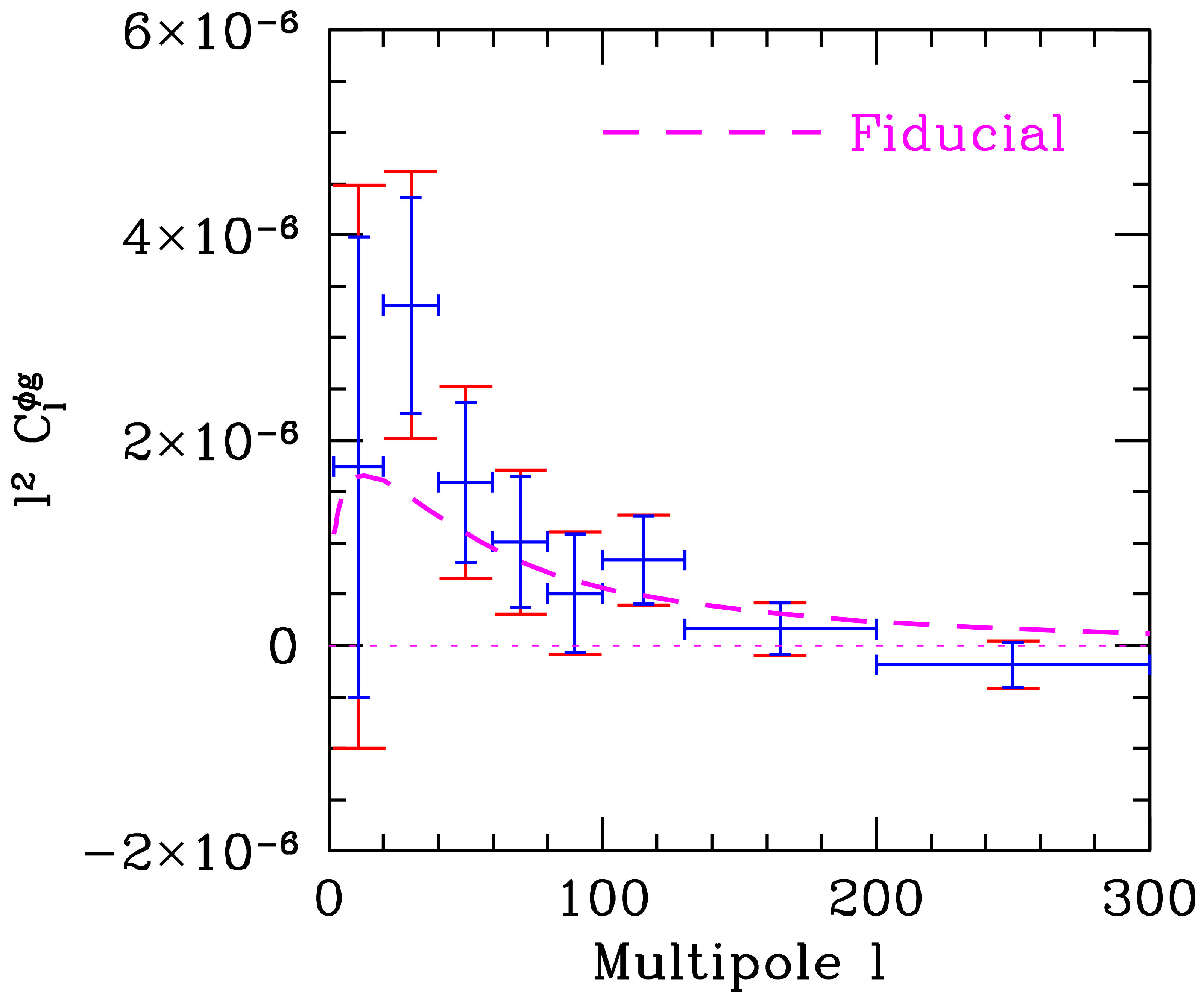
Many lensed patches

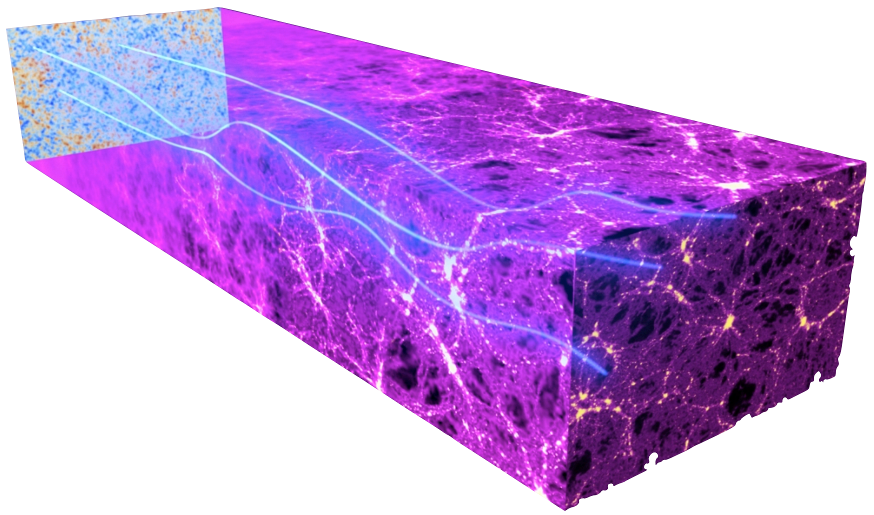


Rather than averaging the modulation, we can measure it as a signal -> CMB lensing map

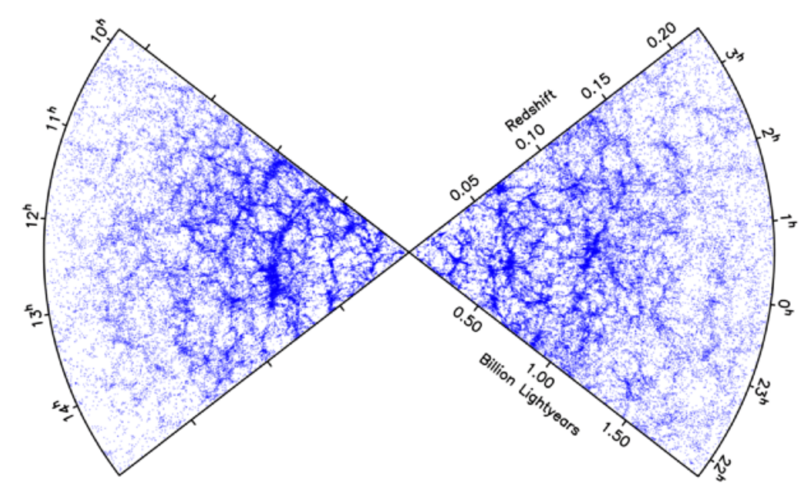
First detection: X-correl with galaxies

WMAP CMB lensing X NVSS galaxies





X



Smith, Zahn & Doré (2007); Hirata, Ho et al. (2008)

Now these are ~20-sigma signals

Planck CMB lensing X {NVSS, MaxBCG, SDSS, WISE}

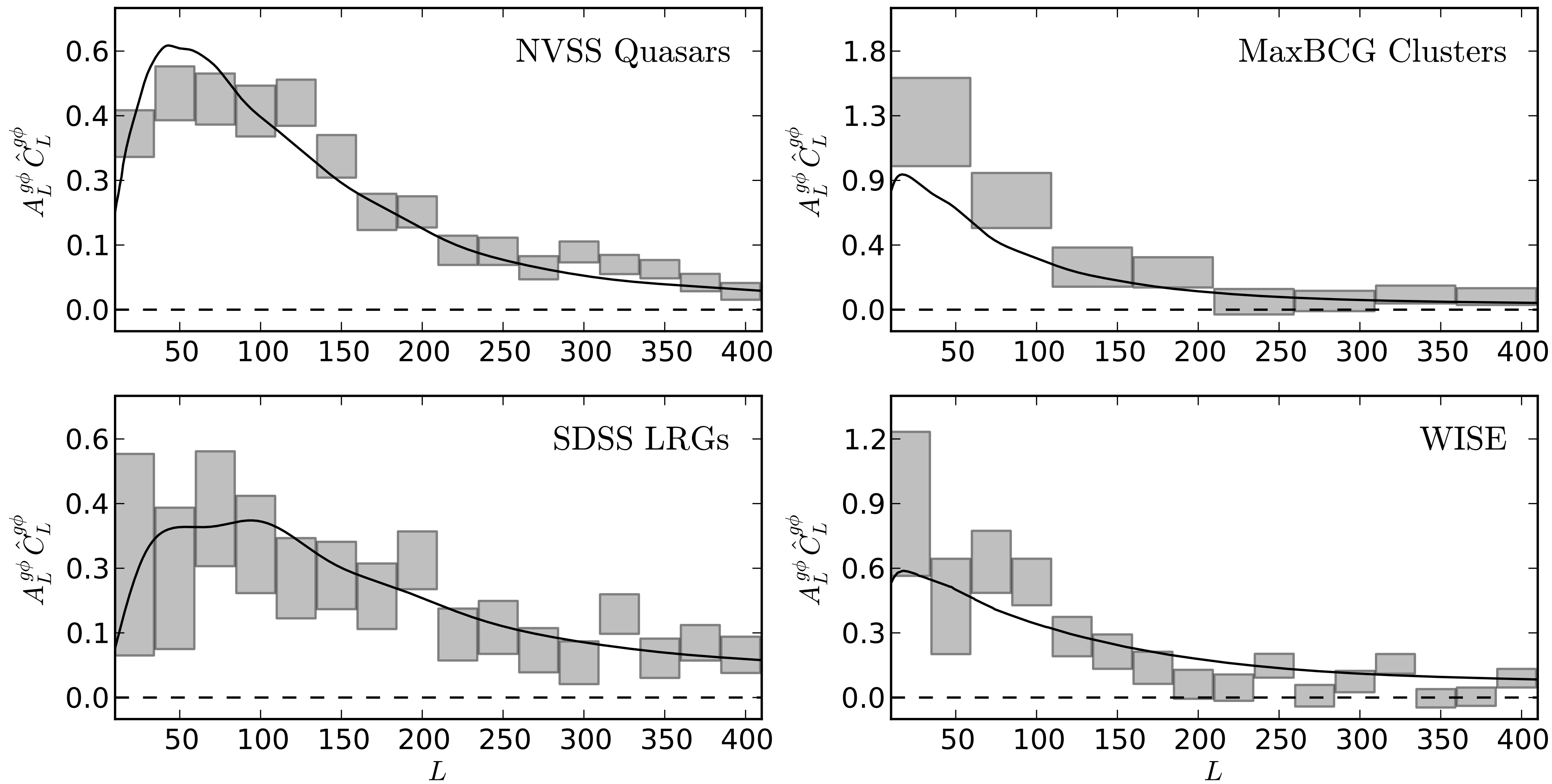


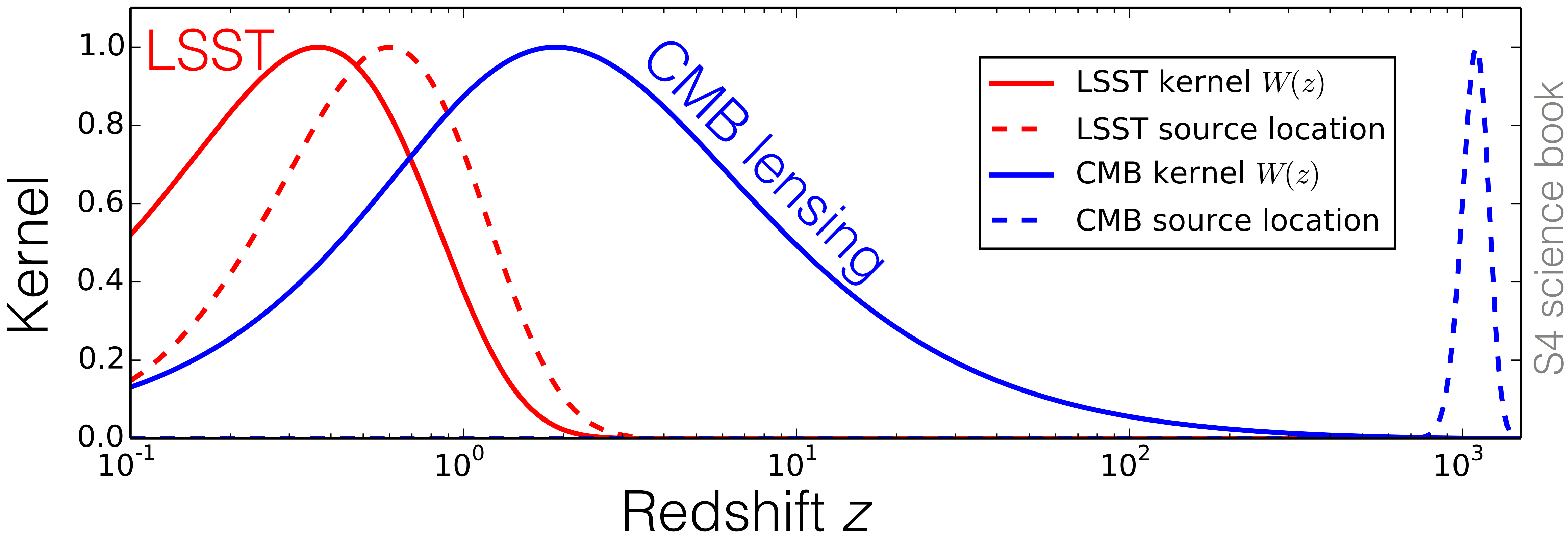
Fig. 17. Cross-spectra of the *Planck* MV lensing potential with several galaxy catalogs, scaled by the signal-to-noise weighting factor $A_L^{g\phi}$ defined in Eq. (52). Cross-correlations are detected at approximately 20σ significance for the NVSS quasar catalog, 10σ for SDSS LRGs, and 7σ for both MaxBCG and WISE.

Planck lensing paper (2013). More recently: {Planck15, ACT, SPT} X {BOSS, CFHTLenS, DES}; lots of great work by Alex van Engelen, Blake Sherwin, and others

Future

CMB lensing maps will soon be signal-dominated
(e.g. Simons Observatory & CMB-S4)

Galaxy surveys collect more galaxies at high redshift
(e.g. DESI, Euclid & LSST)



=> Expect large cross-correlation signal

What can we learn?

Matter amplitude $\sigma_8(z)$

Expansion history / dark energy

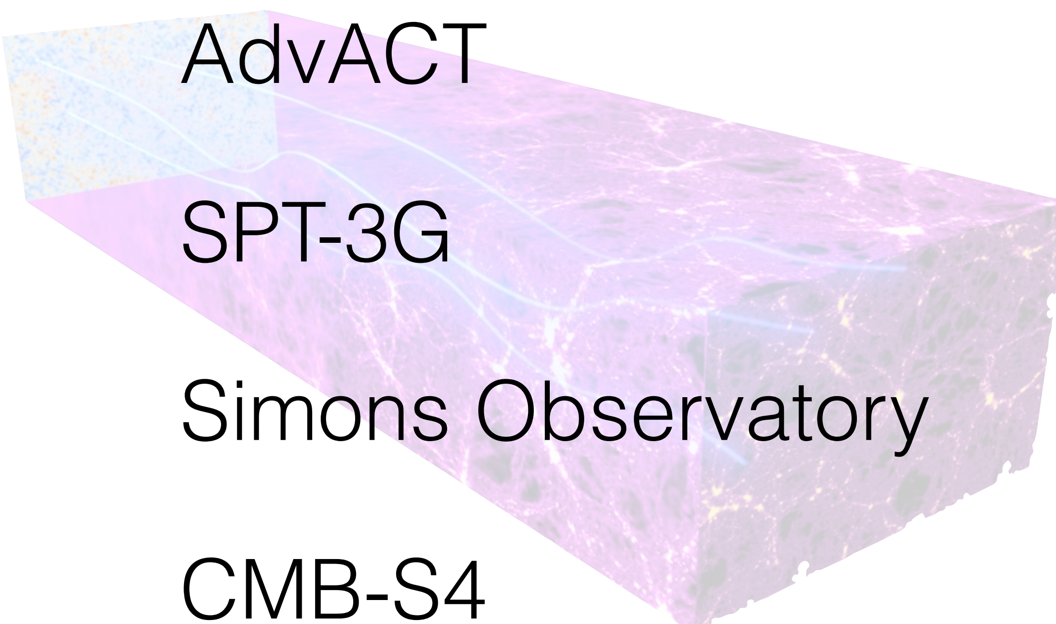
Sum of neutrino masses

Primordial non-Gaussianity / inflation

Galaxy bias and galaxy formation

More?

The future is bright



CMB lensing

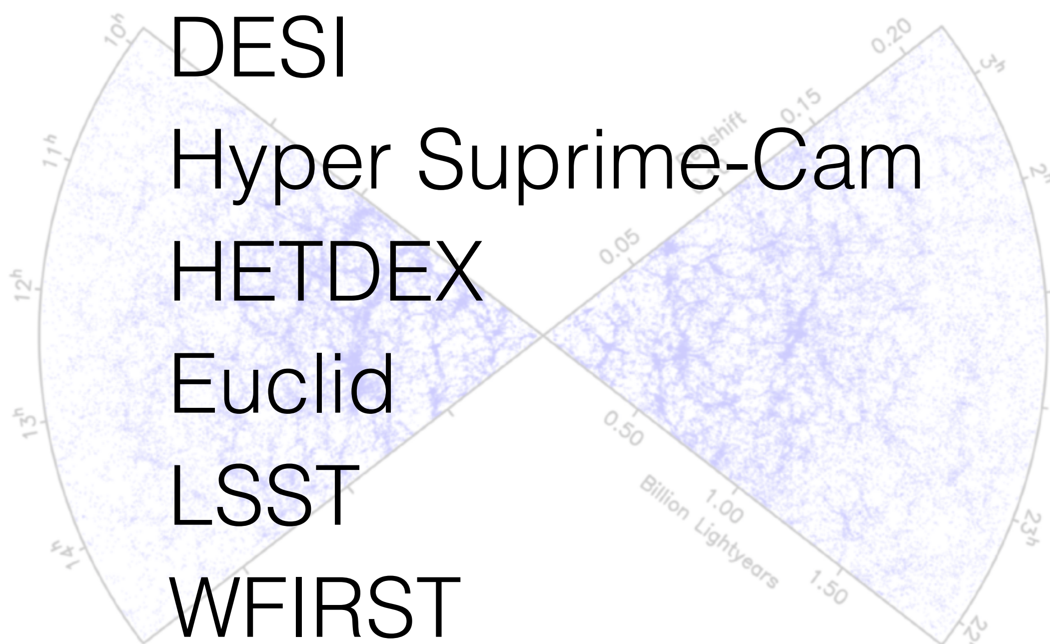
AdvACT

SPT-3G

Simons Observatory

CMB-S4

X



Galaxies

eBOSS

DESI

Hyper Suprime-Cam

HETDEX

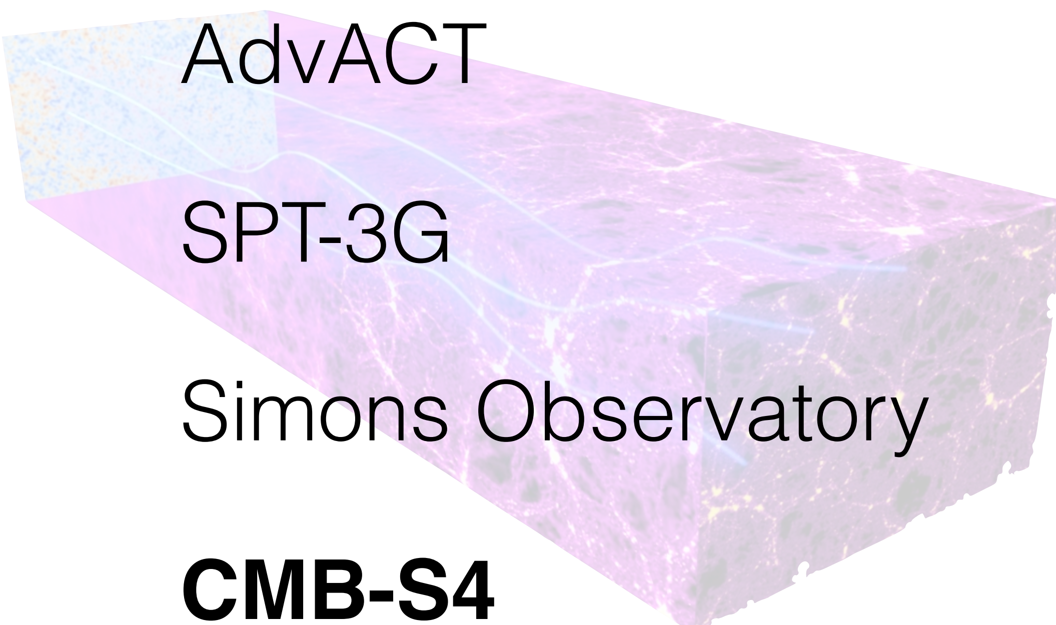
Euclid

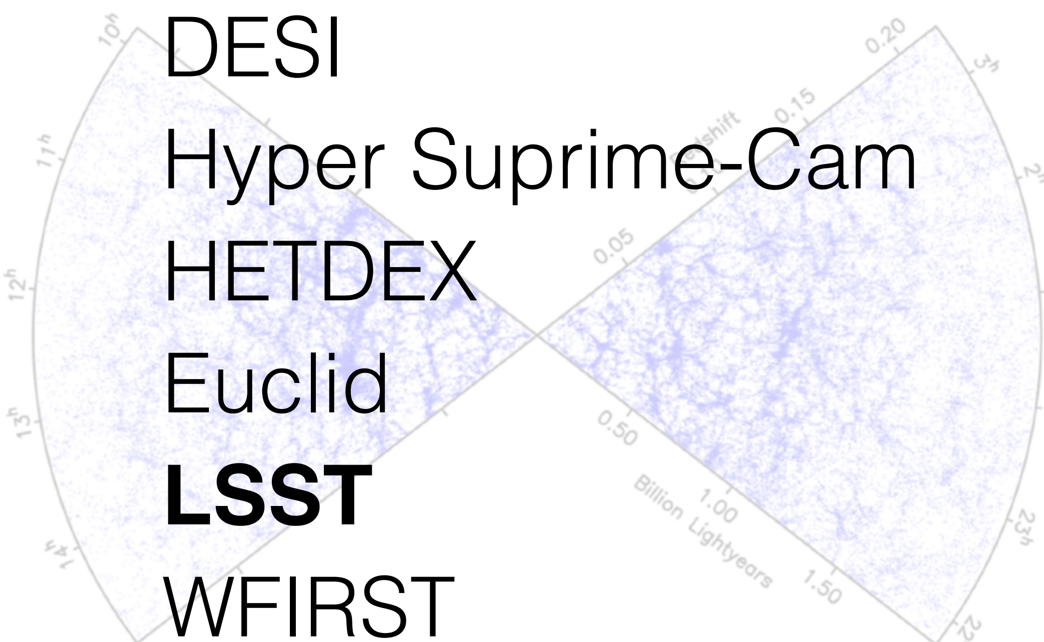
LSST

WFIRST

SPHEREx?

The future is bright





<u>CMB lensing</u>

AdvACT

SPT-3G

Simons Observatory

CMB-S4

X

<u>Galaxies</u>

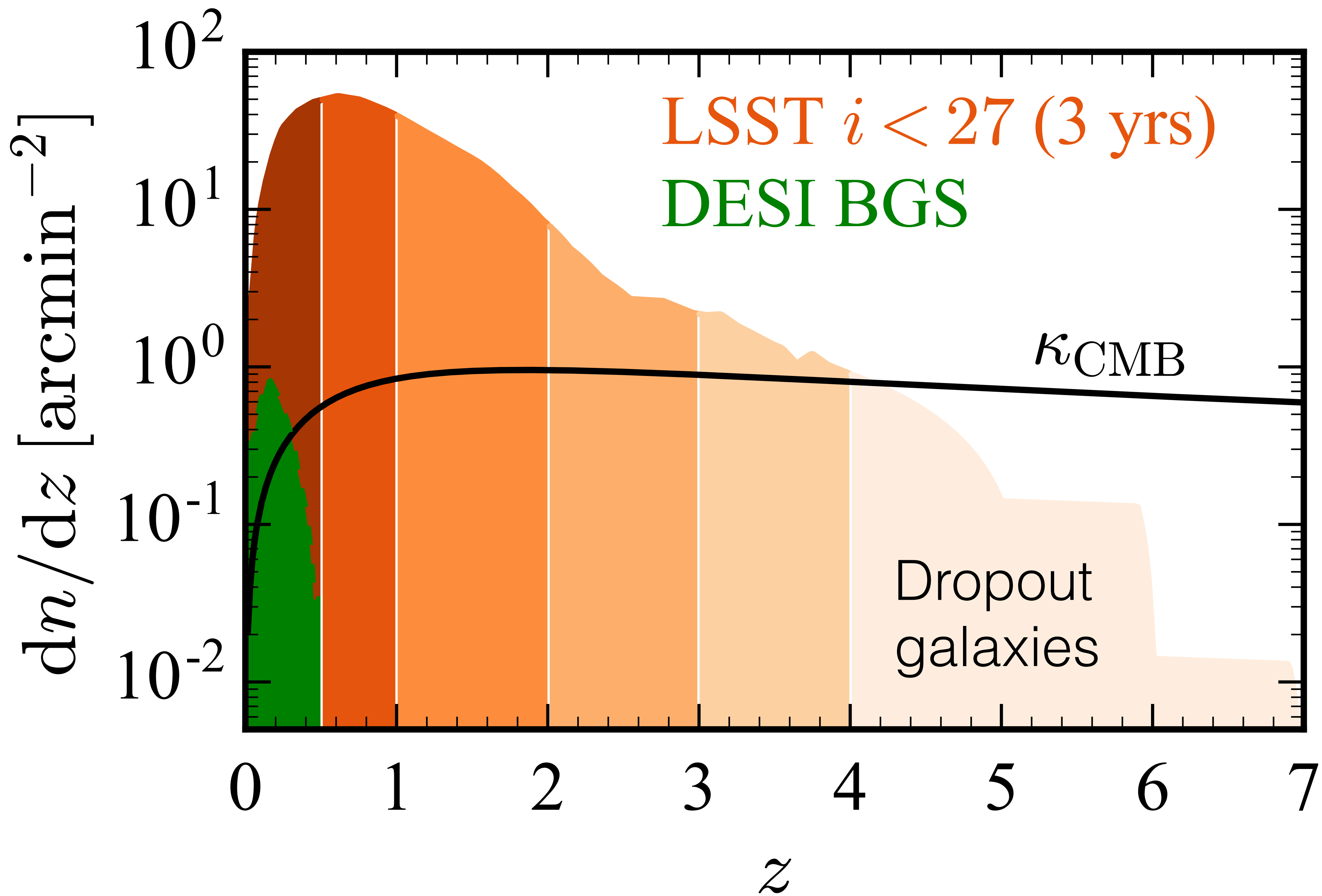
eBOSS
DESI
Hyper Suprime-Cam
HETDEX
Euclid
LSST
WFIRST
SPHEREx

A forecast for CMB-S4 X LSST

Driving question:

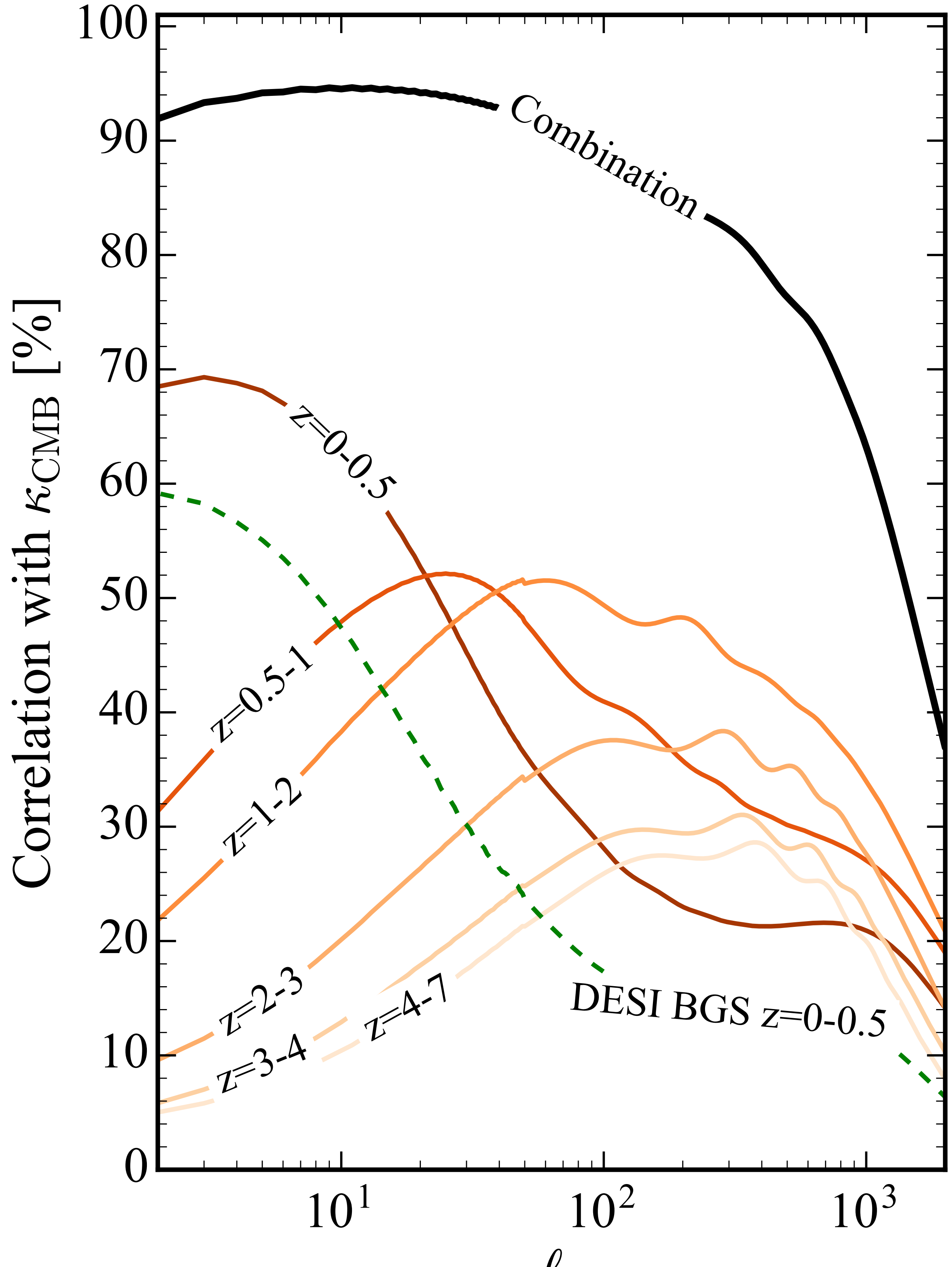
If our models all work and we can mitigate all systematics, and if CMB-S4 and LSST will deliver, what can we hope for?

LSST number density



Following Gorecki et al. (2014), adding z>4 dropouts; 66 arcmin^{-2}; MS & Seljak 1710.09465

Correlation of CMB lensing and galaxies

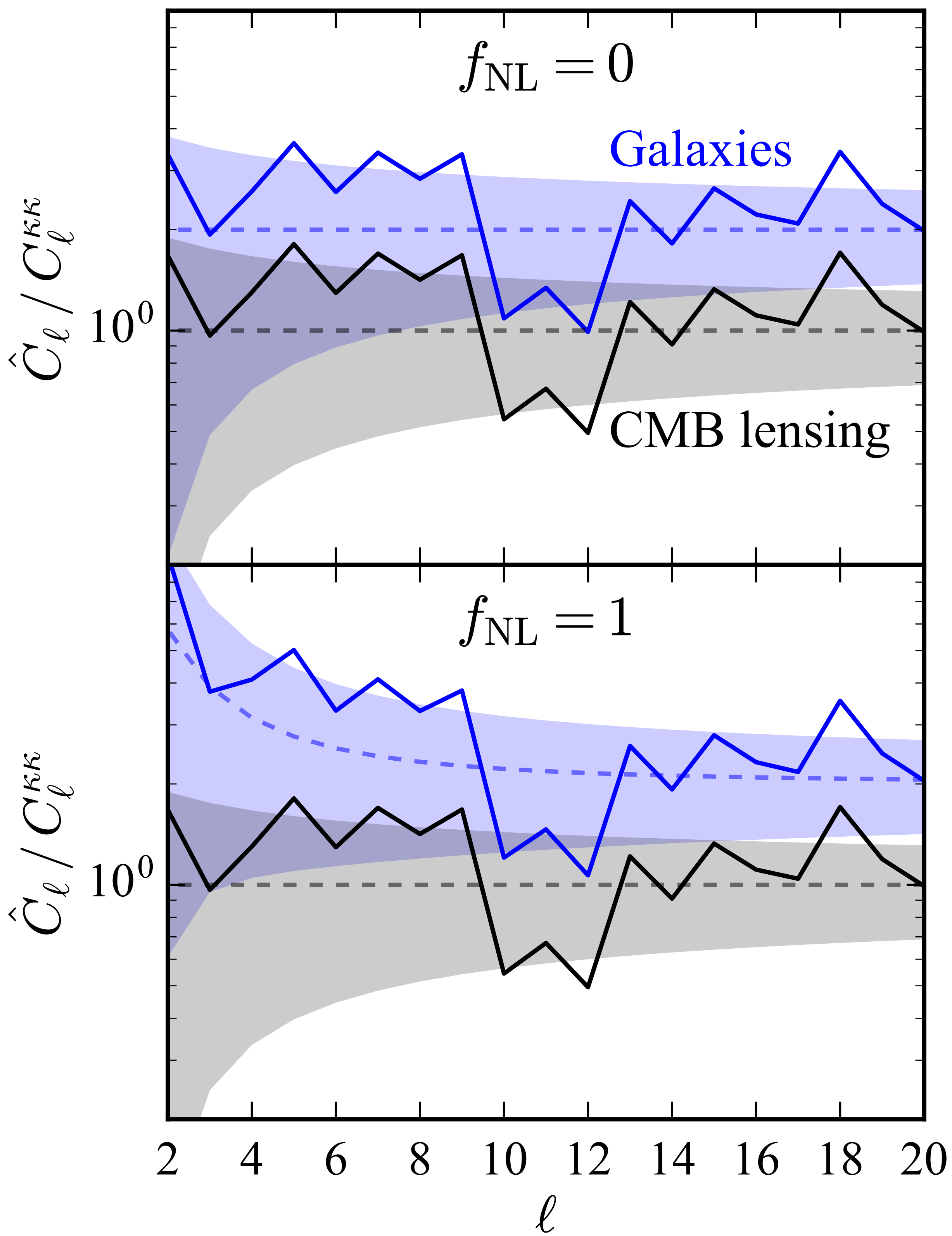


CMB lensing and galaxy maps are up to 95% correlated

$$r_\ell = \frac{C_\ell^{\kappa g}}{\sqrt{\hat{C}_\ell^{\kappa\kappa} \hat{C}_\ell^{gg}}}$$

MS & Seljak 1710.09465

May cancel cosmic variance



Dalal+ (2008), Seljak (2009), McDonald & Seljak (2009), MS & Seljak 1710.09465

May cancel cosmic variance

Ratio galaxies / CMB lensing has no cosmic variance if the two are perfectly correlated (*r*=1)

SNR per mode is

$$[\mathrm{SNR}(f_{\mathrm{NL}})]^2 \simeq \frac{2-r_\ell^2}{1-r_\ell^2} \left(\frac{\partial \ln C_\ell^{\kappa g}}{\partial f_{\mathrm{NL}}}\right)^2$$

$$\propto \frac{1}{1-r_\ell^2} \quad \text{for}\, r \to 1$$

Seljak (2009), McDonald & Seljak (2009), MS & Seljak 1710.09465

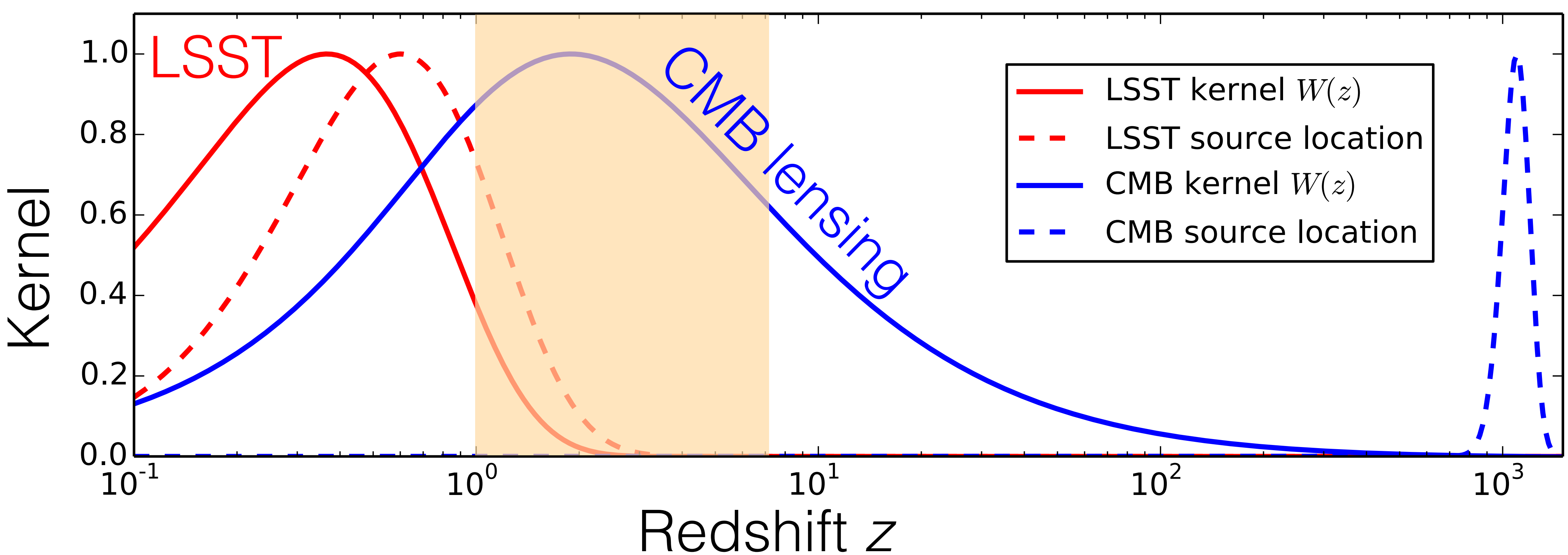
Dropout galaxies at z=4-7

High-z galaxies improve cross-correlation with CMB lensing

Imaging surveys like LSST can use dropout technique to include Lyman break galaxies (LBGs) at z=4-7

HSC/Goldrush found 0.5 million z=4-7 LBGs on 100 deg^2, so LSST could see ~100 million on 18,000 deg^2

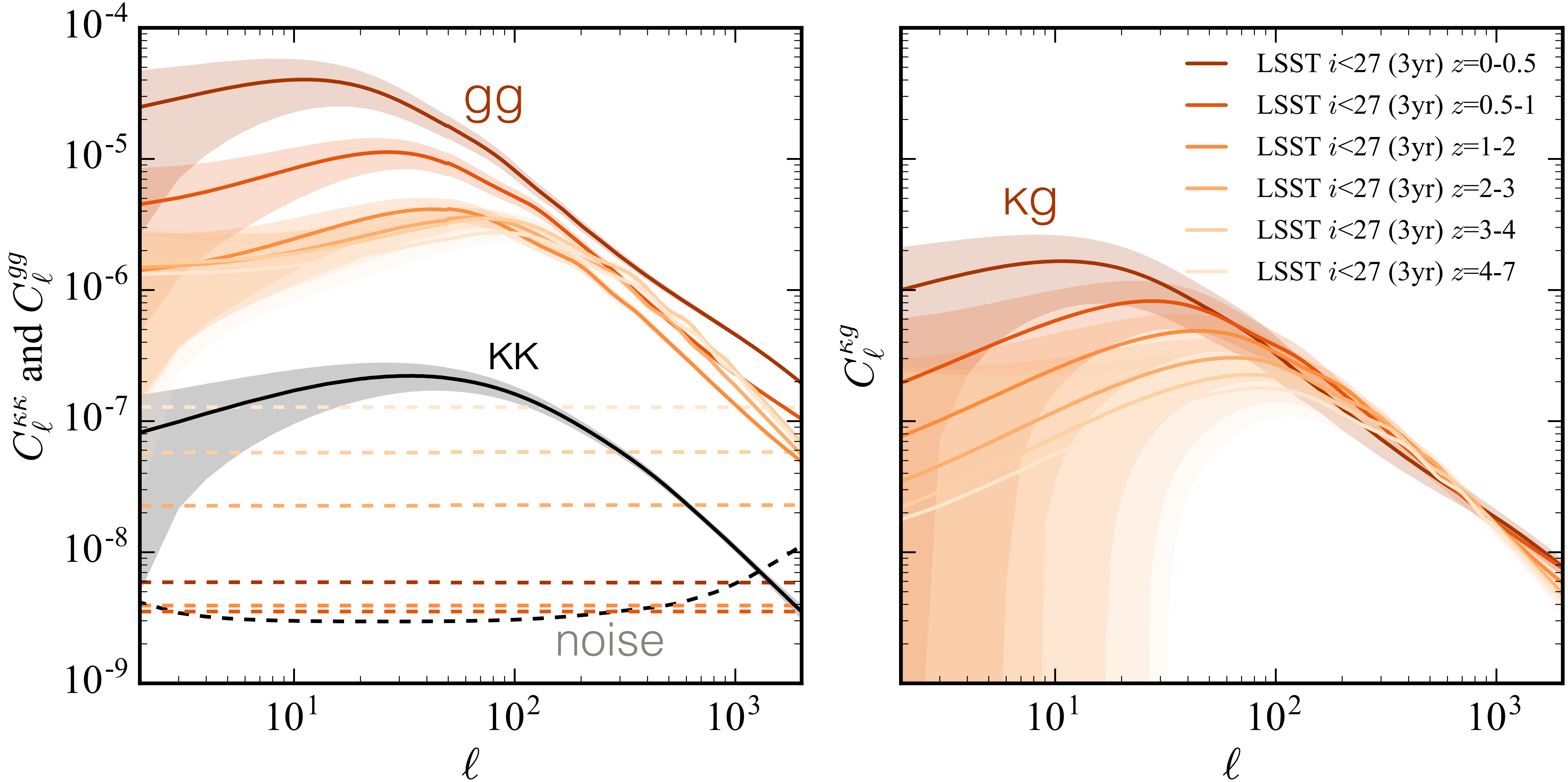
Ono+ (2018)
1704.06004



S4 science book

Power spectra: CMB-S4 & LSST

MS & Seljak 1710.09465



Low ell: $C_\ell^{XX'} = \frac{2}{\pi}\int_0^\infty \frac{\mathrm{d}k}{k}\Delta_{X,\ell}(k)\Delta_{X',\ell}(k)k^3 P_{\delta_X\delta_{X'}}(k, z=0),$

$\Delta_{X,\ell}(k) \equiv \int_0^\infty \mathrm{d}\chi\, \bar{W}_X(\chi) j_\ell(k\chi).$

High ell: $C_\ell^{XX'} = \int_z P_{\delta_X\delta_{X'}}(k=\ell/\chi(z), z) \times W_X(z)b_X(z)W_{X'}(z)b_{X'}(z)$

SNR of auto-power spectra

SNR of C^{XX}	$\ell_{\rm max}$ 500	1000	2000
$\kappa_{\rm CMB}$	233	406	539
BOSS LRG z=0-0.9	140	187	230
SDSS $r < 22$ z=0-0.5	247	487	936
SDSS $r < 22$ z=0.5-0.8	247	487	936
DESI BGS z=0-0.5	230	417	665
DESI ELG z=0.6-0.8	158	210	256
DESI ELG z=0.8-1.7	150	194	225
DESI LRG z=0.6-1.2	184	267	349
DESI QSO z=0.6-1.9	44.8	48.8	50.8
LSST $i < 27$ (3yr) z=0-0.5	250	496	982
LSST $i < 27$ (3yr) z=0.5-1	250	496	979
LSST $i < 27$ (3yr) z=1-2	249	492	956
LSST $i < 27$ (3yr) z=2-3	245	469	830
LSST $i < 27$ (3yr) z=3-4	239	444	724
LSST $i < 27$ (3yr) z=4-7	224	387	555

MS & Seljak 1710.09465

SNR of kg cross-power spectra

	$\ell_{\rm max}$		
SNR of $C^{\kappa_{\rm CMB}X}$	500	1000	2000
BOSS LRG z=0-0.9	77.3	117	159
SDSS $r < 22$ z=0-0.5	88.3	167	284
SDSS $r < 22$ z=0.5-0.8	88.3	167	284
DESI BGS z=0-0.5	50.1	93.5	144
DESI ELG z=0.6-0.8	50.7	73.5	97
DESI ELG z=0.8-1.7	103	148	185
DESI LRG z=0.6-1.2	86.7	133	182
DESI QSO z=0.6-1.9	74.9	94.5	108
LSST $i < 27$ (3yr) z=0-0.5	78.1	150	258
LSST $i < 27$ (3yr) z=0.5-1	112	202	338
LSST $i < 27$ (3yr) z=1-2	144	259	406
LSST $i < 27$ (3yr) z=2-3	121	219	324
LSST $i < 27$ (3yr) z=3-4	101	182	261
LSST $i < 27$ (3yr) z=4-7	94	167	229

MS & Seljak 1710.09465

Fisher analysis setup

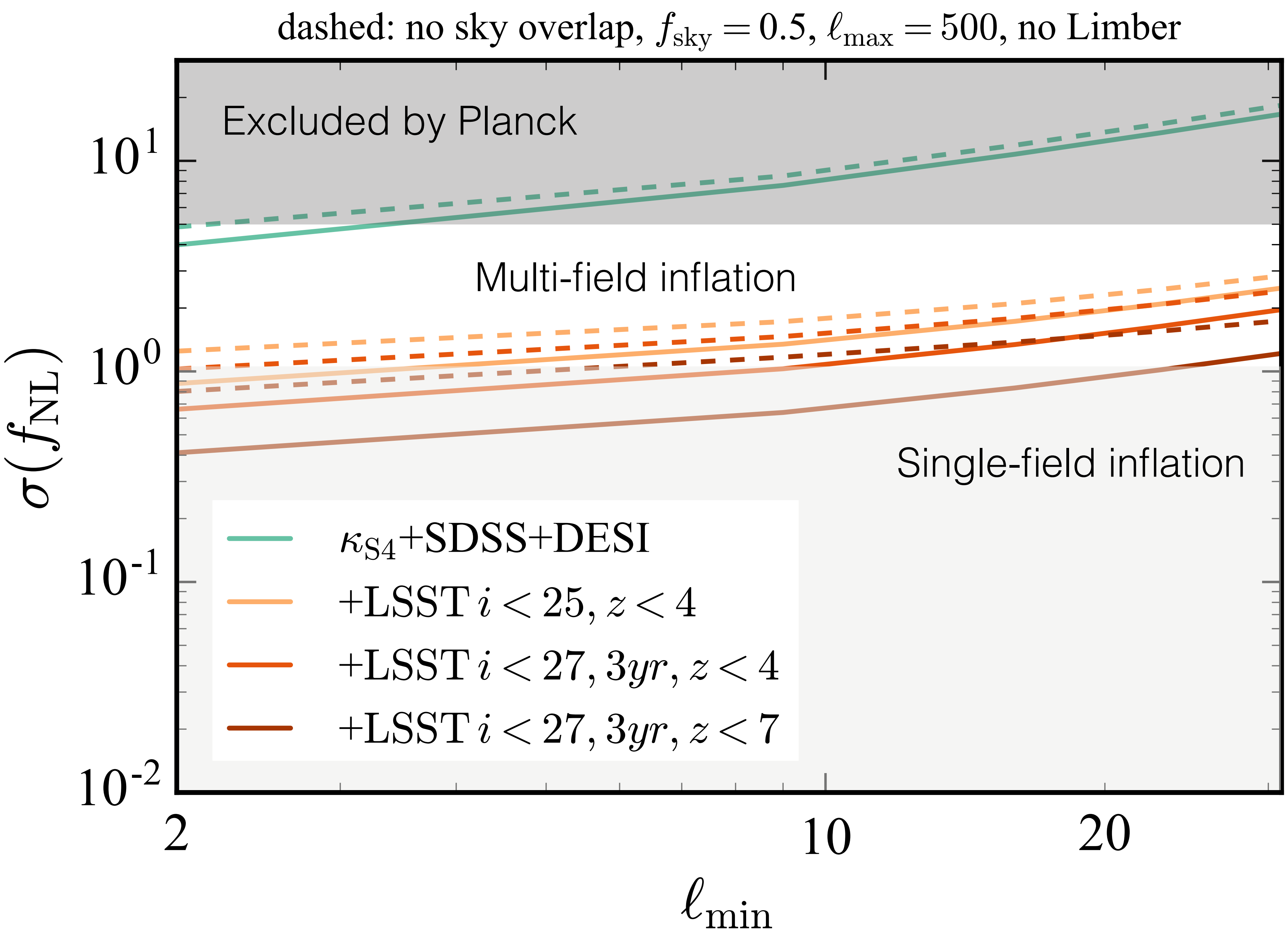
Include all κκ, κg, gg power spectra $\mathbf{d}_\ell = (C_\ell^{11}, C_\ell^{12}, \ldots, C_\ell^{NN})$

For Gaussian covariance, different ell are uncorrelated, so Fisher matrix is

$$F_{ab} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \frac{\partial \mathbf{d}_\ell}{\partial \theta_a} [\mathrm{cov}(\mathbf{d}_\ell, \mathbf{d}_\ell)]^{-1} \frac{\partial \mathbf{d}_\ell}{\partial \theta_b}.$$

N x *N* matrix at every ell

Prospects for local f_{NL}



Dalal+ (2008), Jeong, Komatsu & Jain (2009), Ginnantonio & Percival (2014), MS & Seljak 1710.09465

Prospects for local f_{NL}

S4 + LSST is sensitive to f_{NL}=0.4 (L_{min}=2) - f_{NL}=1 (L_{min}=20)

Without CMB lensing, degrade by factor 10-20

Without sky-overlap, degrade by factor 1.5-2 (SV cancellation)

Without low-L C^{gg}, degrade by factor 2-3

Without z>4 dropout galaxies, degrade by factor 2

MS & Seljak 1710.09465

Challenges for local f_{NL}

Need to measure CMB lensing and galaxy clustering on large scales (L<20)

Star contamination affects low-L gg, potentially mimicking f_{NL}

- Not relevant when just getting upper bound on f_{NL}
- Know direction of our galaxy so could project out modes as in Leistedt et al. (2014)
- Even without low-L gg, f_{NL}=1 is possible

Catastrophic redshift errors

- Hope to calibrate using spec-z surveys
- If global dn/dz known, data can determine outlier fraction so that catastrophic errors don't degrade f_{NL}

-> ask me later

MS & Seljak 1710.09465

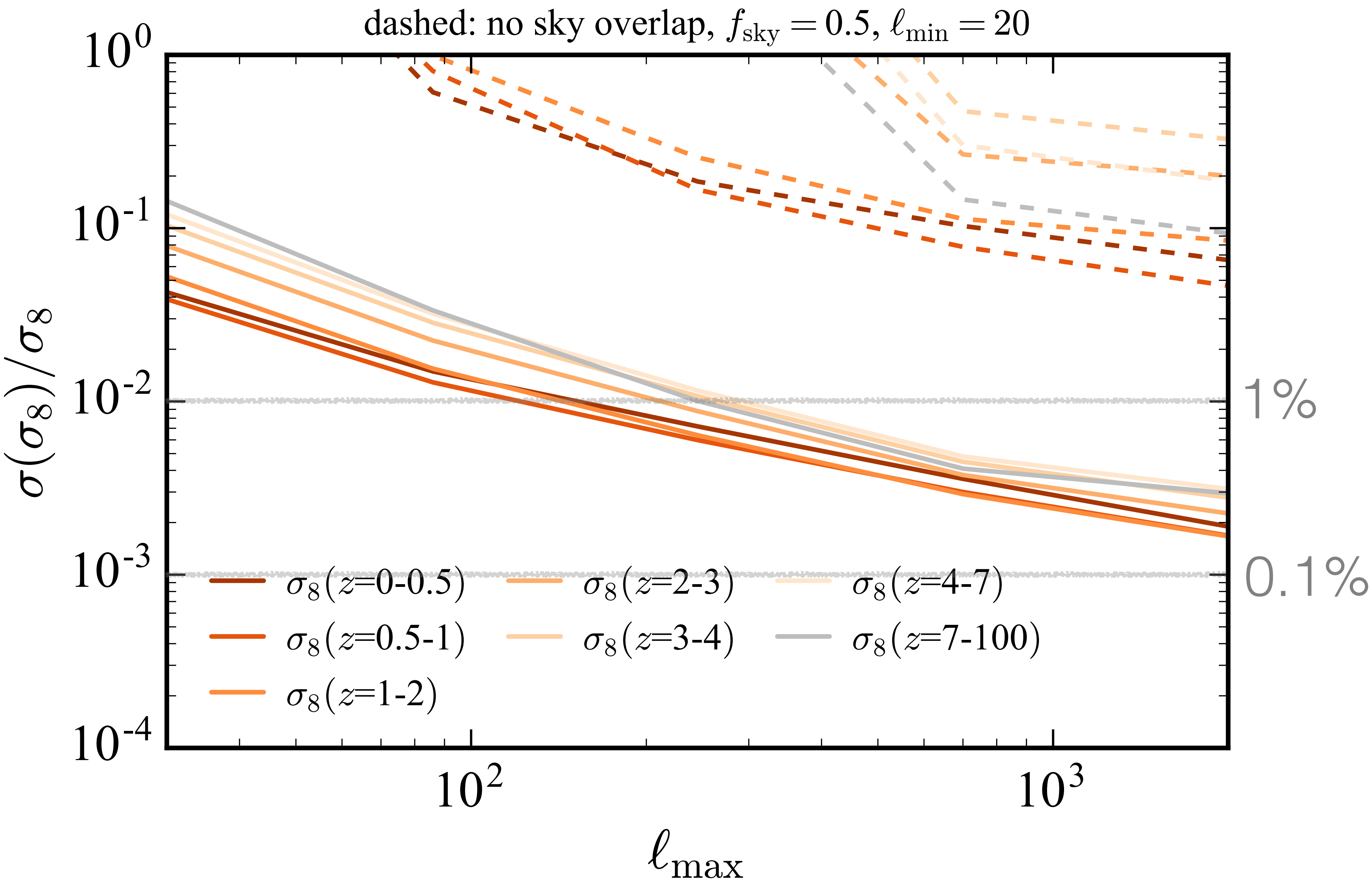
Prospects for matter amplitude $\sigma_8(z)$

$$C^{\kappa\kappa} \propto \sigma_8^2$$

$$C^{\kappa g} \propto b_1 \sigma_8^2$$

$$C^{gg} \propto b_1^2 \sigma_8^2$$

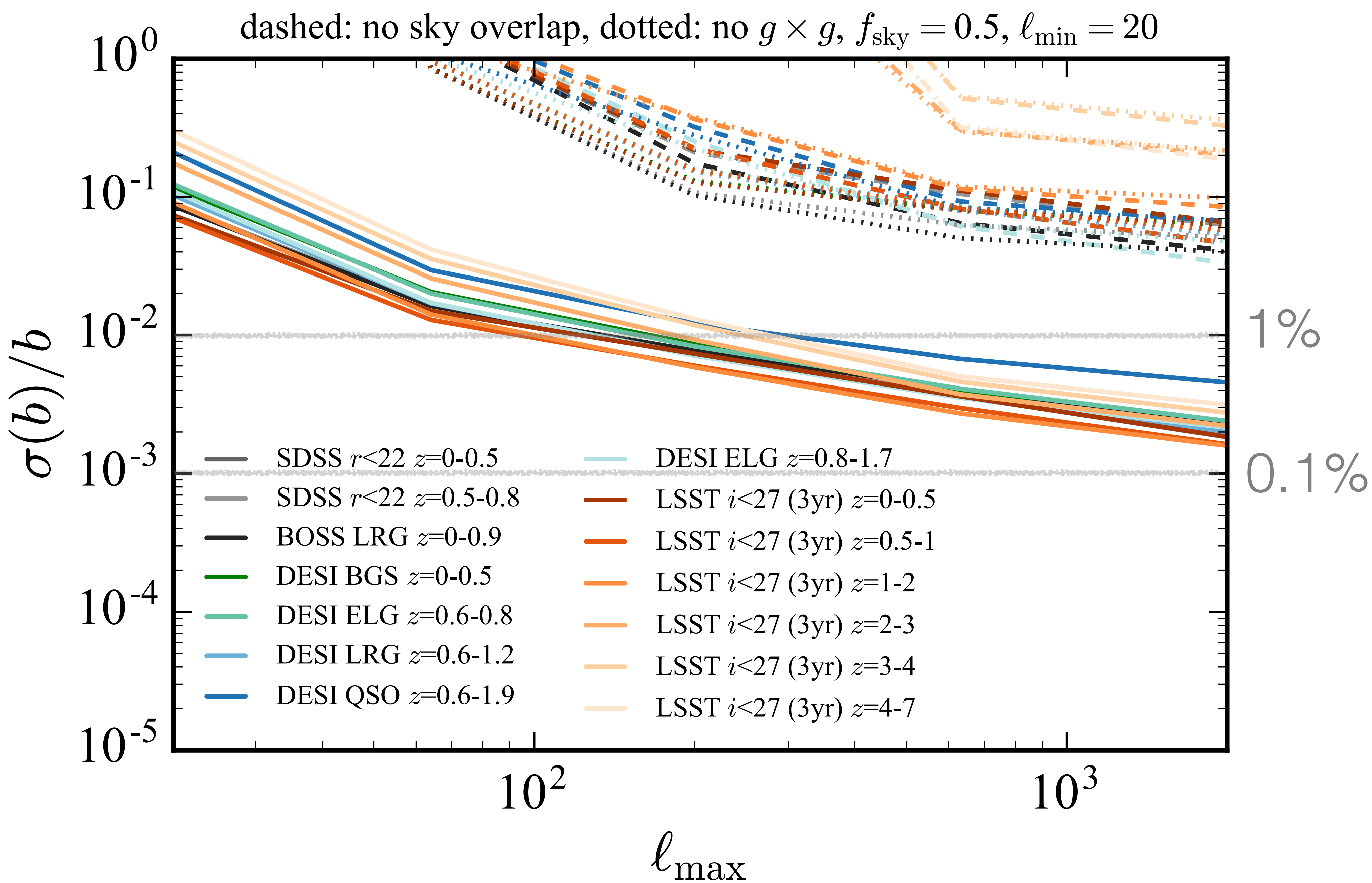
Prospects for $\sigma_8(z)$



Marginalize over one linear bias parameter per redshift bin; fixed cosmology; halofit $P_{mm}(k,z)$; f_{sky}=0.5 for CMB-S4 & LSST

MS & Seljak 1710.09465, see Modi+ (2017) for impact of nonlinear bias

Also get halo bias



Marginalize over one sigma8(z) binned in broad redshift bins

fixed cosmology; halofit $P_{\rm mm}(k,z)$; $f_{\rm sky}$=0.5 for CMB-S4 & LSST

MS & Seljak 1710.09465

Challenges for $\sigma_8(z)$

- Nonlinear halo bias b_2, b_{s2} — Modi, White & Vlah (2017)

  -> Hope for priors from theory, sims, and 3PCF/bispectrum

- Modeling all power spectra to high L_{max}

Conclusions: Part I

CMB-S4 lensing X LSST clustering very promising for measuring primordial non-Gaussianity and growth of structure

Get only slightly worse constraints for Simons Observatory
What about DES instead of LSST?

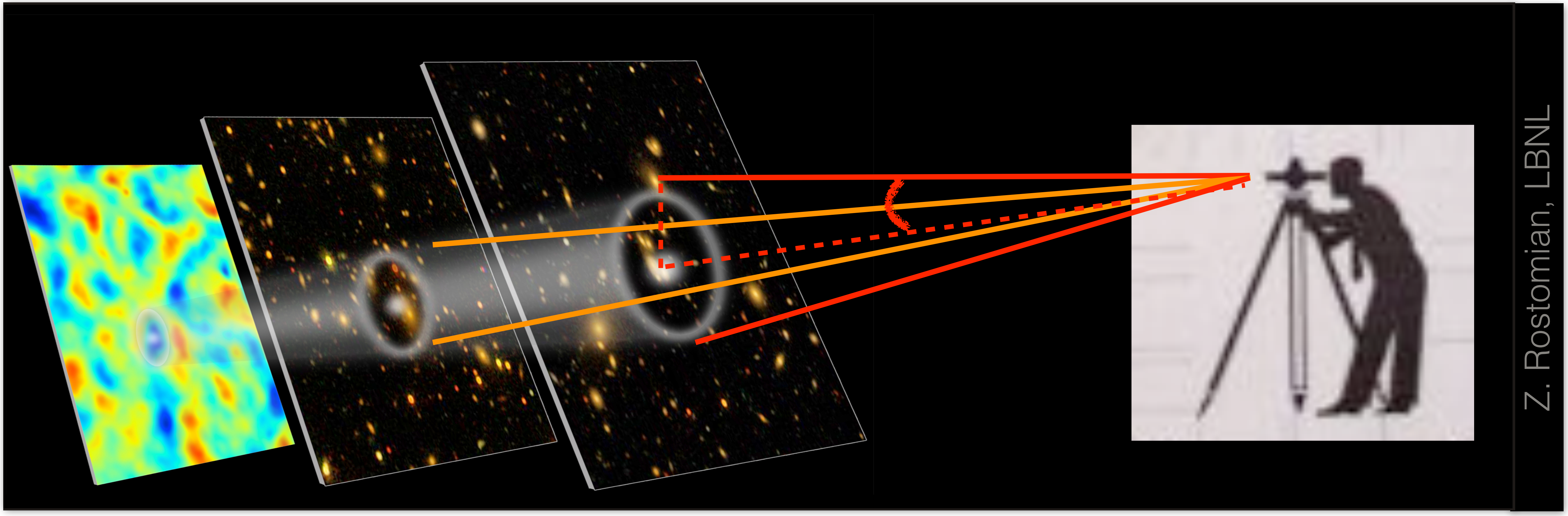
Joint analysis is crucial (factor 10 improvement)

For f_{NL}, need rather low L_{min} and large f_{sky}

Growth measurement is limited by modeling small, nonlinear scales
-> Part II of the talk

Part II
Initial condition reconstruction

Acoustic scale is also imprinted in galaxies: BAO

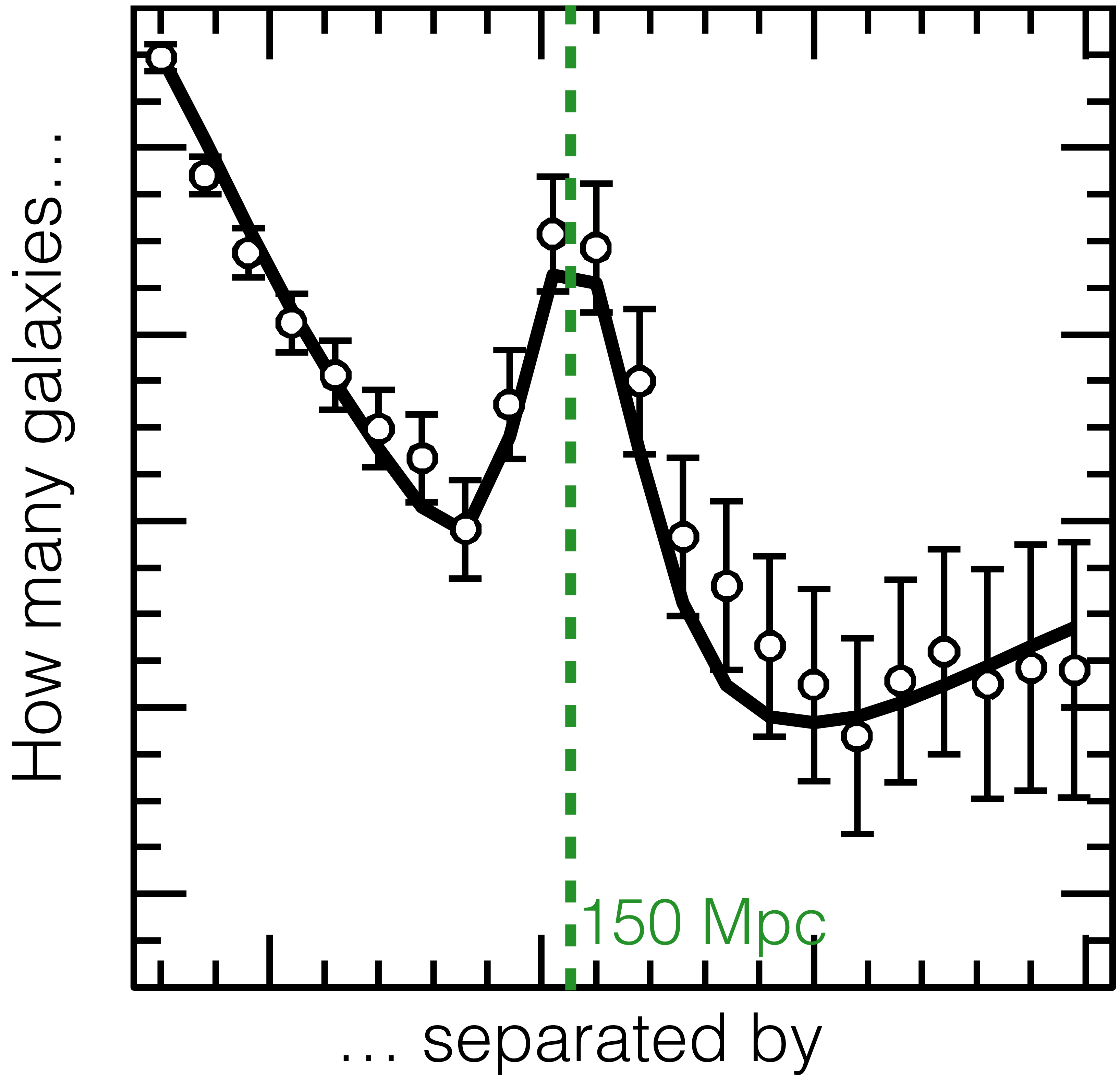


Galaxies more likely separated by 150 rather than 140 or 160 Mpc

$$\text{Distance} \sim \frac{150\,\text{Mpc}}{\text{angle}} \sim \int_0^z \frac{dz'}{H(z')}$$

This measures Hubble parameter (=expansion rate)

Preferred clustering at separation of 150Mpc



Anderson+ (2013) / Sloan Digital Sky Survey

BAO scale is set in the early (linear) Universe

Hot sea of baryons & photons

Driven by photon pressure

Electrons cool,

form hydrogen,

decouple from photons,

& remain in place

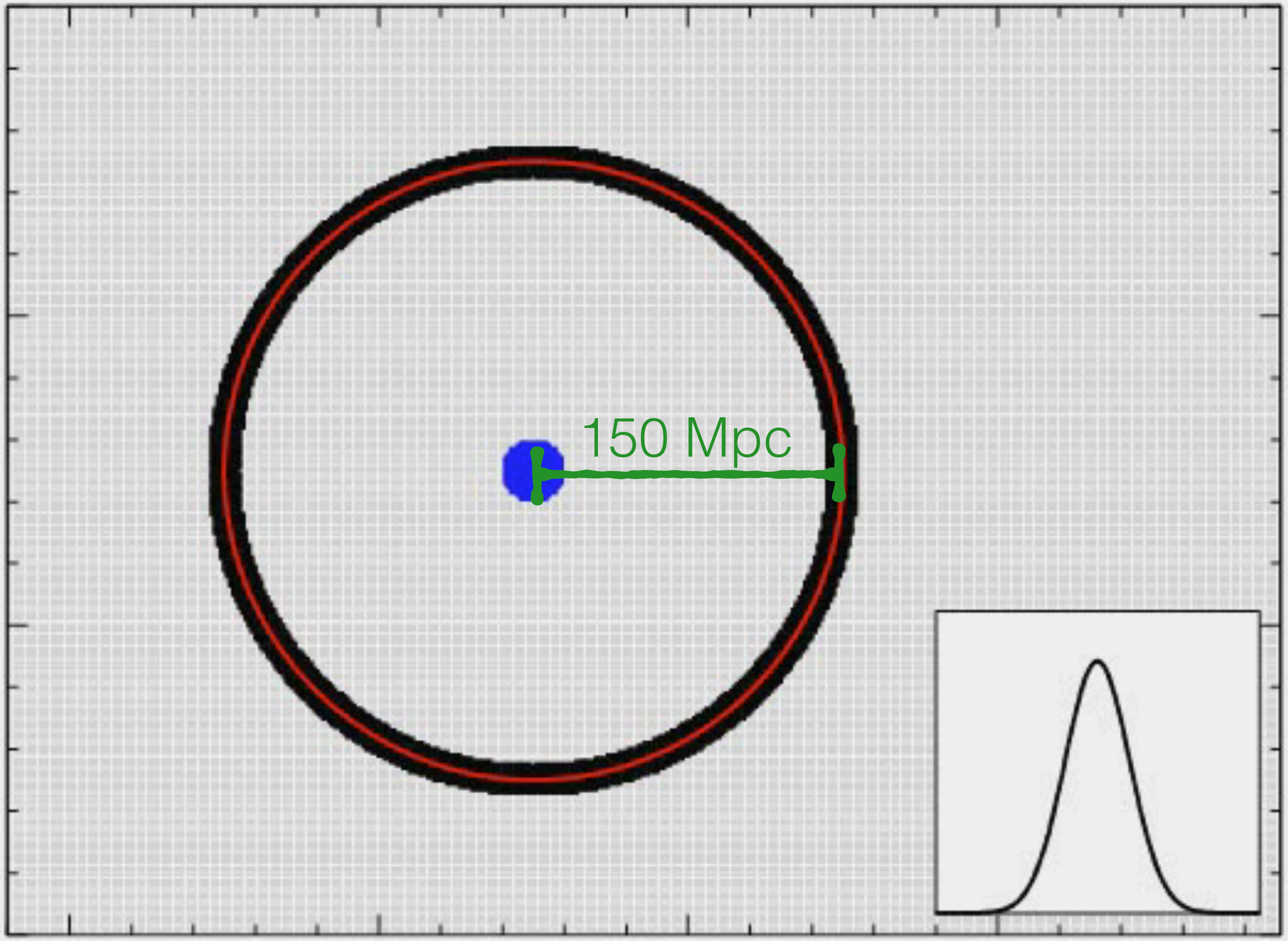
Big bang
[-13.8bn yrs]

Decoupling
[-13.7996bn yrs]

Today

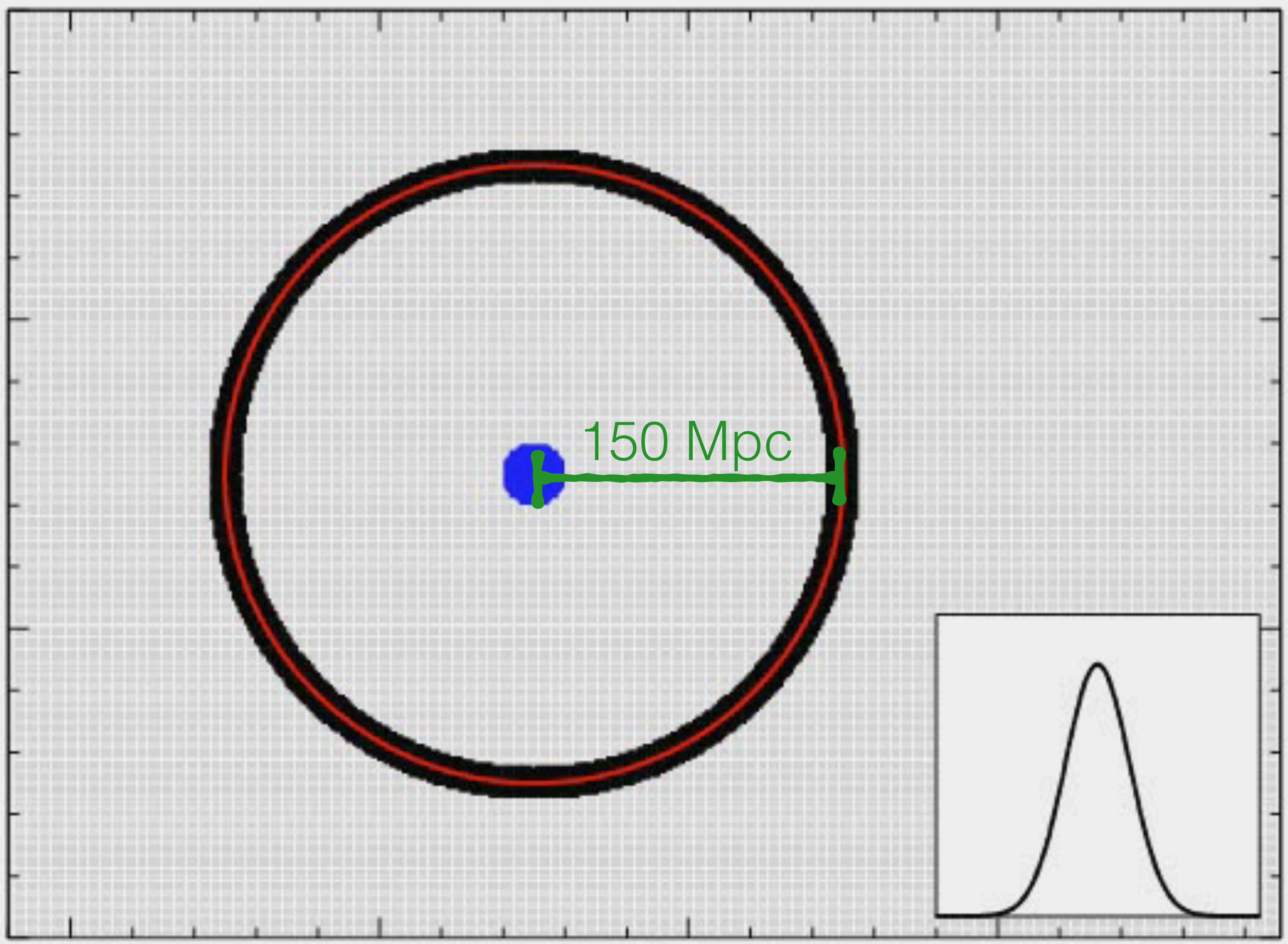
Sound wave travels 150 Mpc:
Baryon-acoustic-oscillation
(BAO) scale

At early times, acoustic scale is the same everywhere



Padmanabhan++ (2012)

Displacements on ~10-150 Mpc modulate this

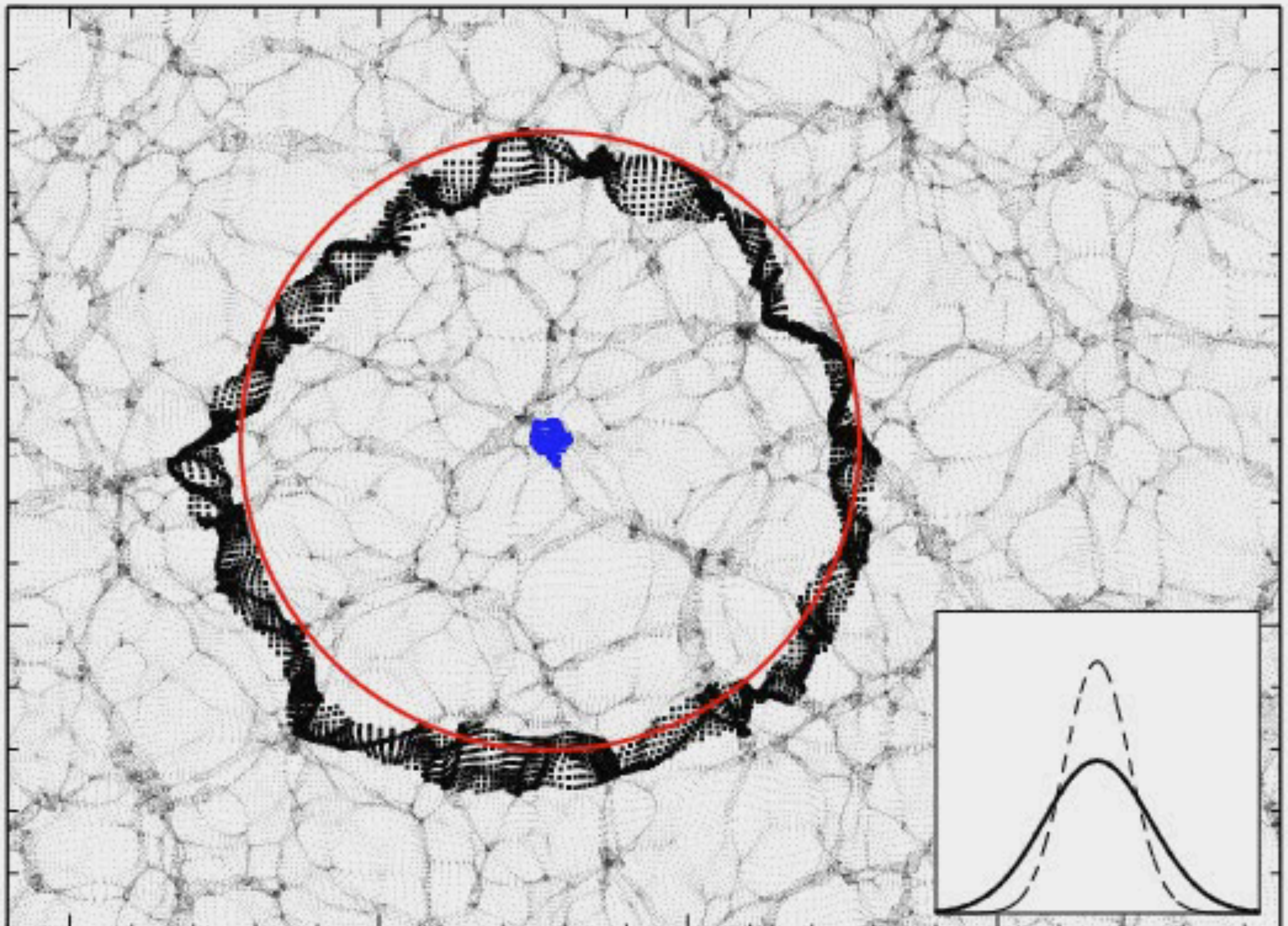


Padmanabhan++ (2012)

Reduce nonlinear dynamics with reconstruction

Estimate potentials and move galaxies back

Eisenstein++ (2007)

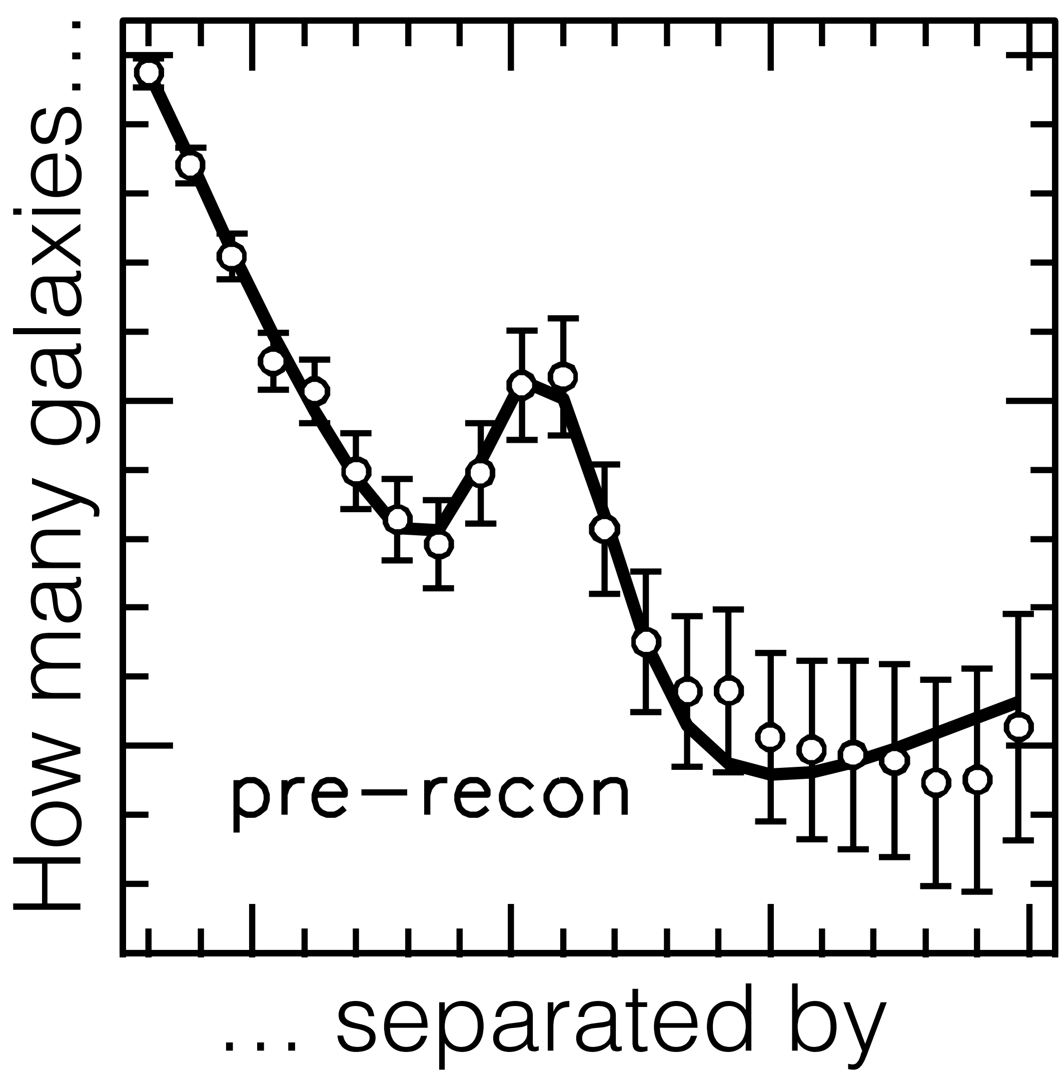


Padmanabhan++ (2012)

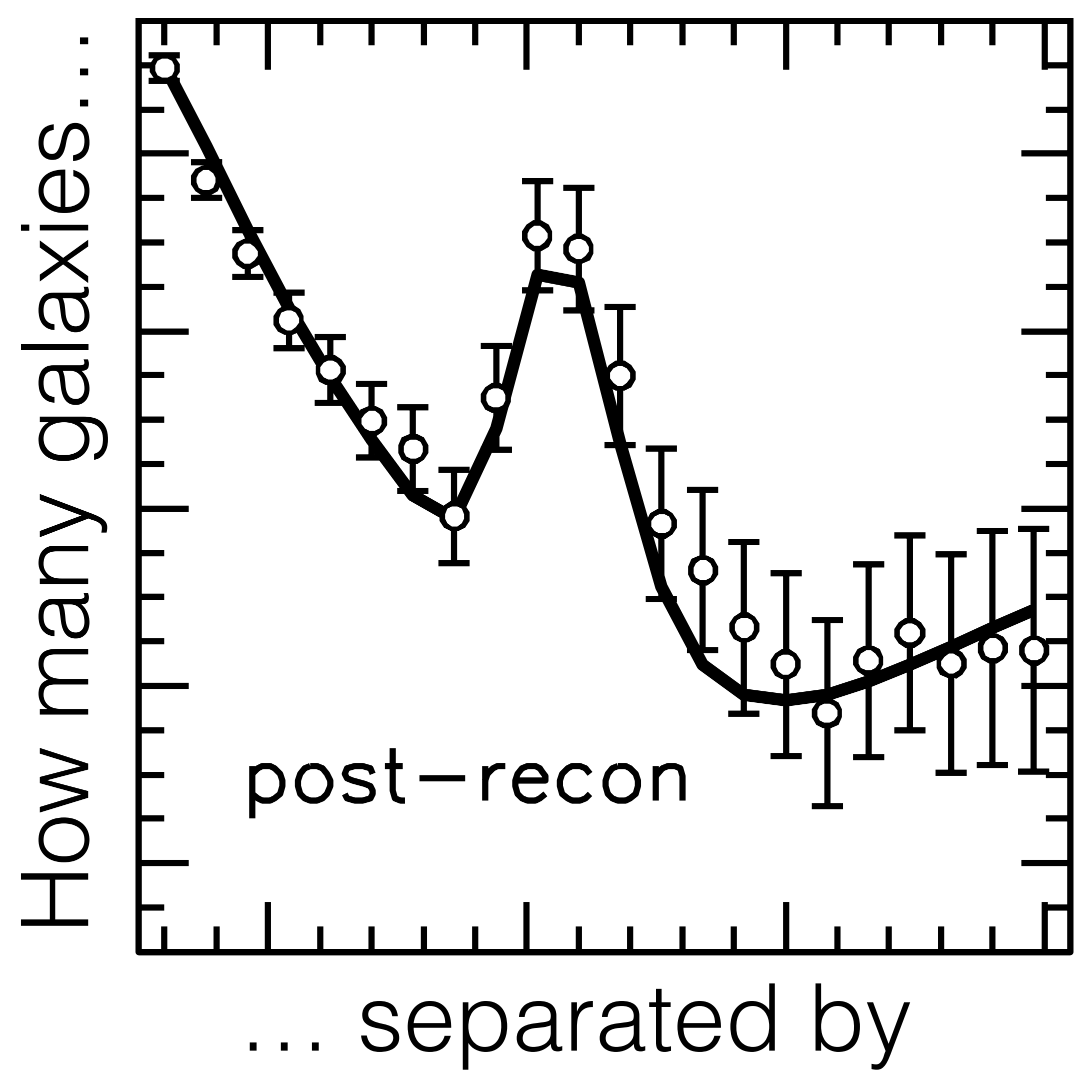
Demonstration of reconstruction on real data

For BOSS DR11 data, signal-to-noise of the distance scale improved by 50%, achieving sub-percent level precision

Observed

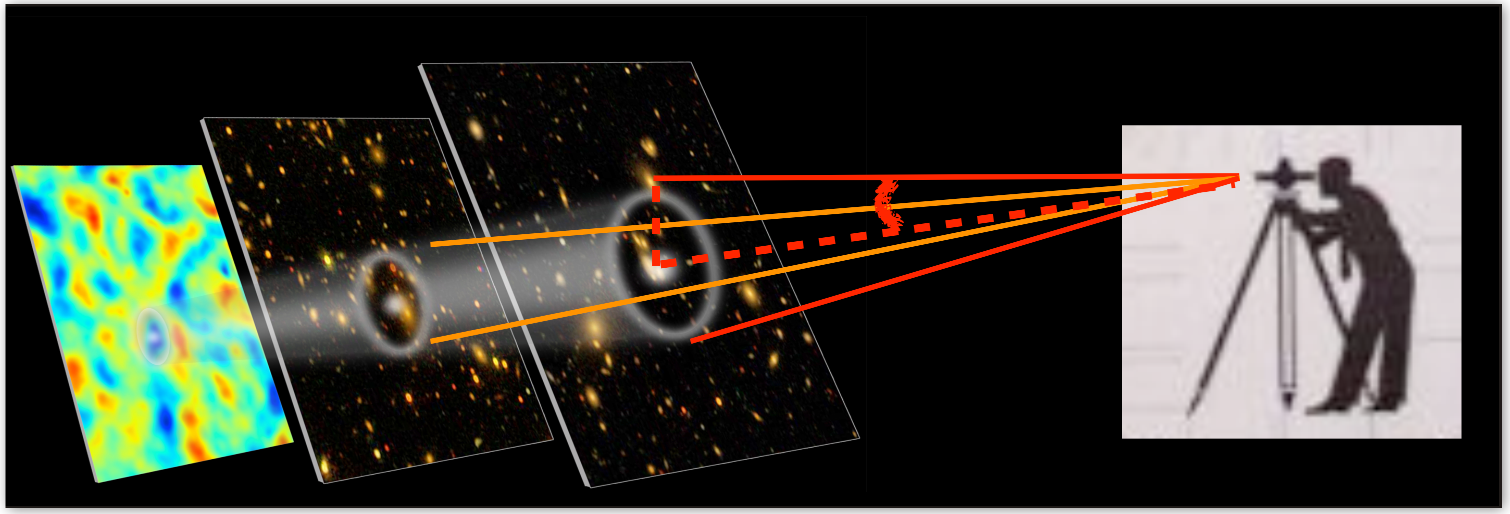


Reconstructed



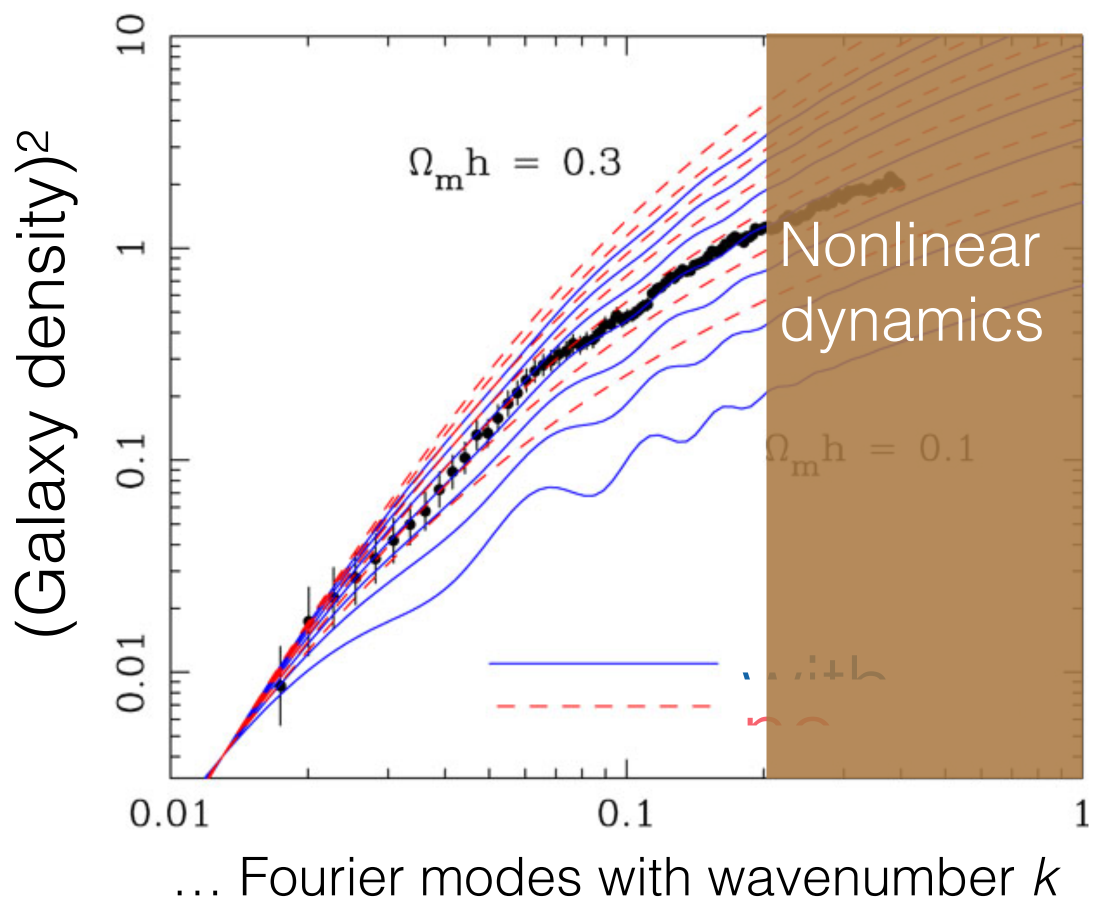
Eisenstein++ (2007), Padmanabhan++ (2012), Anderson++ (2013)

Limiting factor: Structure formation is nonlinear



BAO distance

Nonlinear dynamics smears out primordial BAO scale



Broadband power spectrum

Nonlinear dynamics affects nearby galaxies, so their data is thrown away

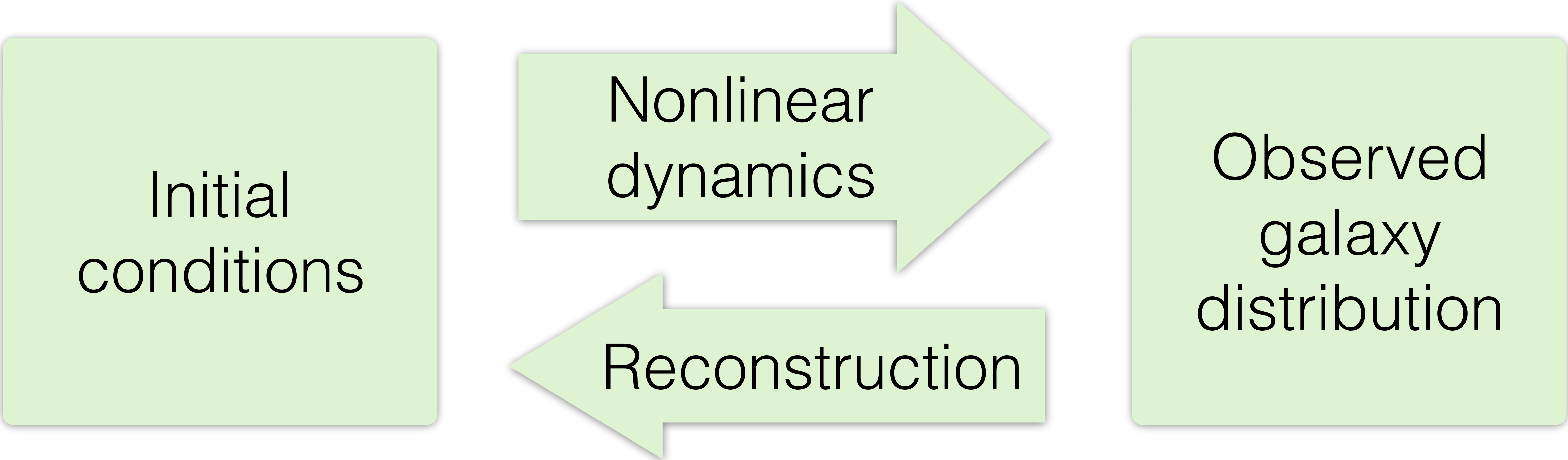
Nonlinear dynamics: What can we do?

(1) Better analytical models

(2) Simulate it all and infer cosmology

(3) Transform data to reduce nonlinear dynamics

(4) Exploit non-Gaussian tails of galaxy distribution



Paradigm 1: Lagrangian reconstruction

Estimate velocities, move galaxies back

*Peebles, PIZA/MAK (e.g. Mohayaee), Eisenstein, Padmanabhan, Tassev, Zaldarriaga, Zhu, X. Wang, UL Pen+, B. Li+, Baldauf, **MS**, …*

Paradigm 2: Forward model & sample

Sample ICs, evolve forward, compare vs observations, iterate

__Jasche__, __Lavaux__, __Leclercq__, Wandelt, Kitaura, HY Wang,…

Paradigm 3: Forward model & optimize

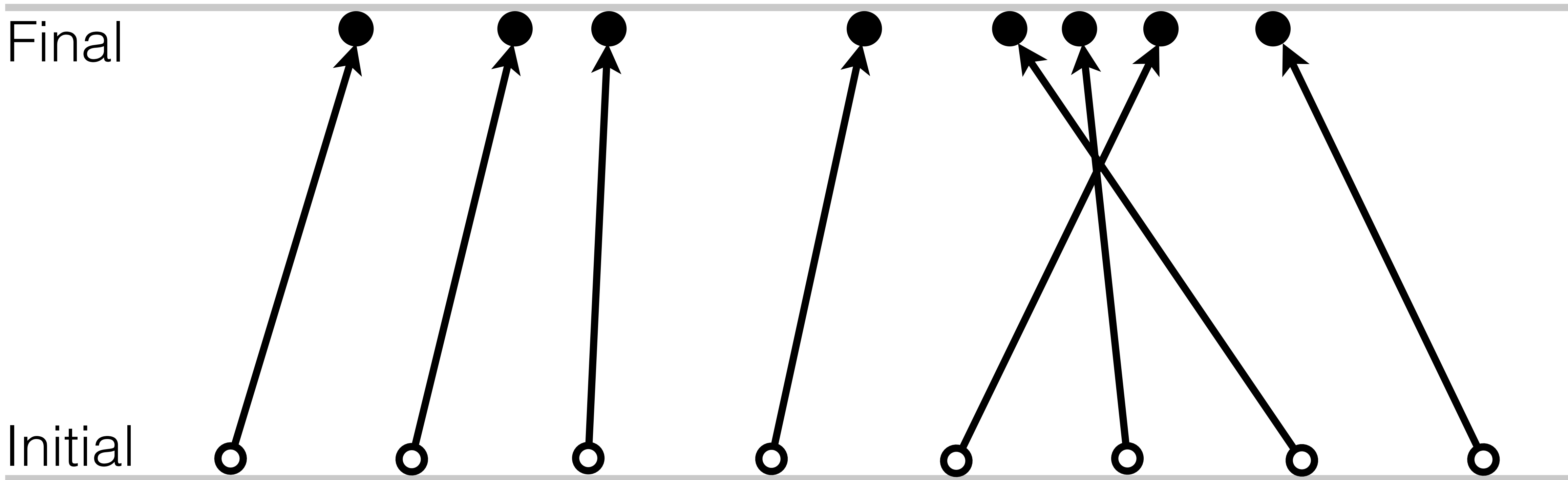
Maximum-likelihood solution by solving optimization problem

Seljak, Aslanyan, Feng, Modi

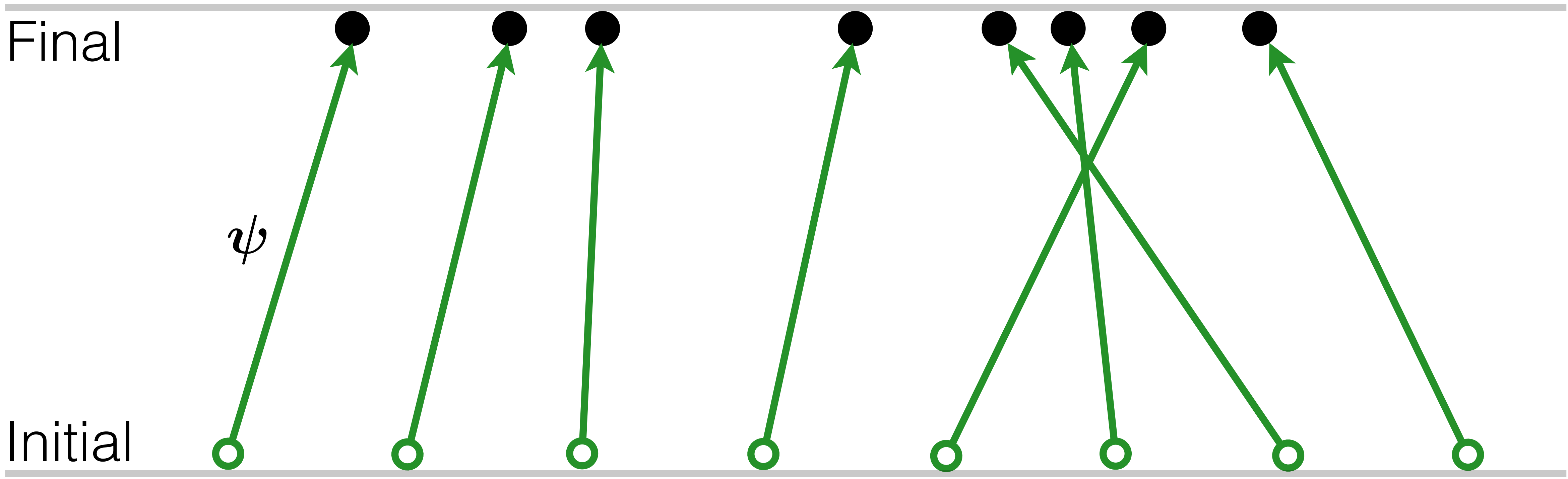
Paradigm 4: ML to go directly to parameters

__Shirley Ho__ +

1-D example

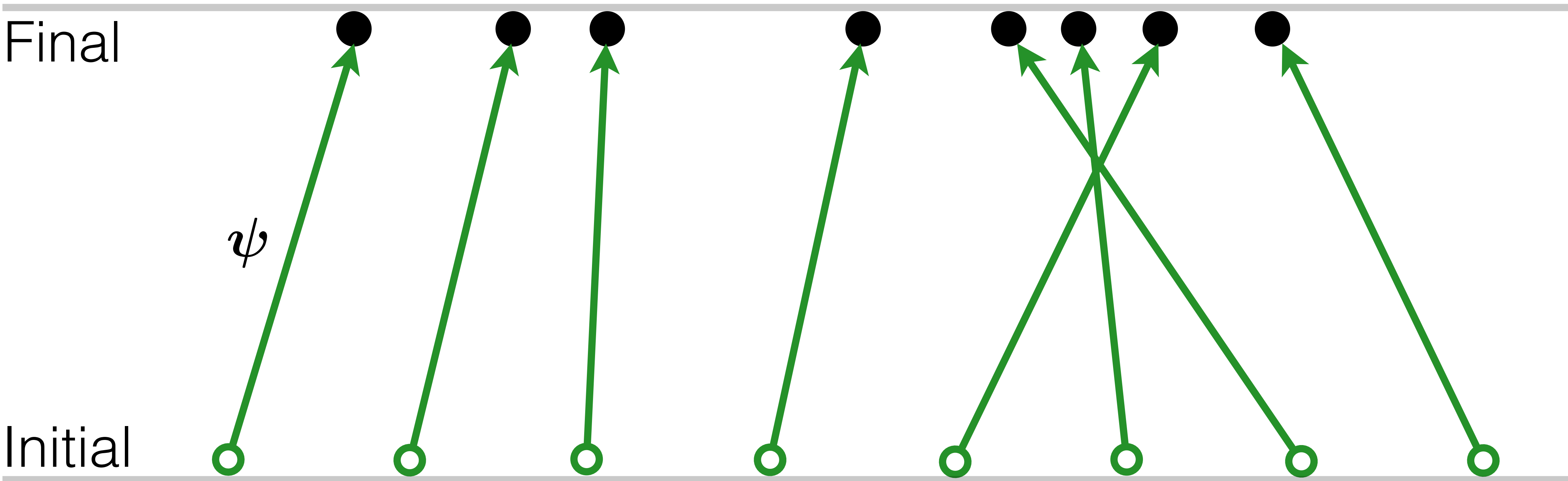


(1) Displacement field is a nonlinear functional of the linear initial density



$$\psi(\mathbf{k}) = \frac{\mathbf{k}}{k^2}\delta_0(\mathbf{k})$$
$$+ \int_{\mathbf{k}_1} \mathbf{L}^{(2)}(\mathbf{k}_1, \mathbf{k}-\mathbf{k}_1)\delta_0(\mathbf{k}_1)\delta_0(\mathbf{k}-\mathbf{k}_1)$$
$$+ \cdots$$

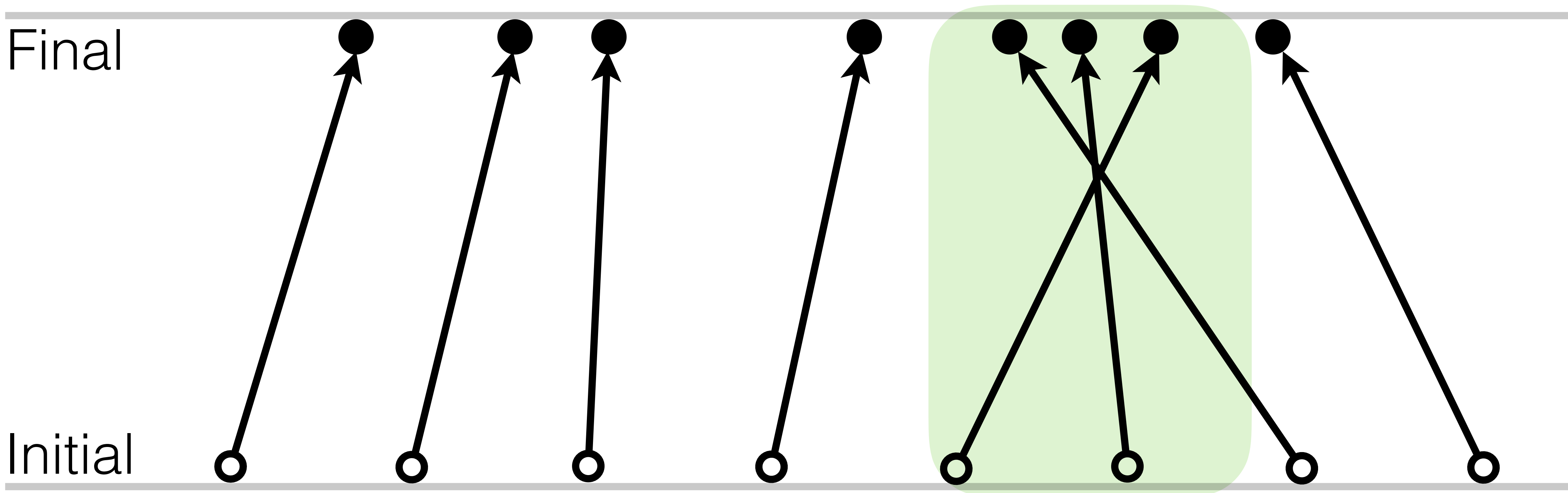
(1) Displacement field is a nonlinear functional of the linear initial density



- Nonlinear terms are small, so displacement is quite linear
- Perturbative modeling works well

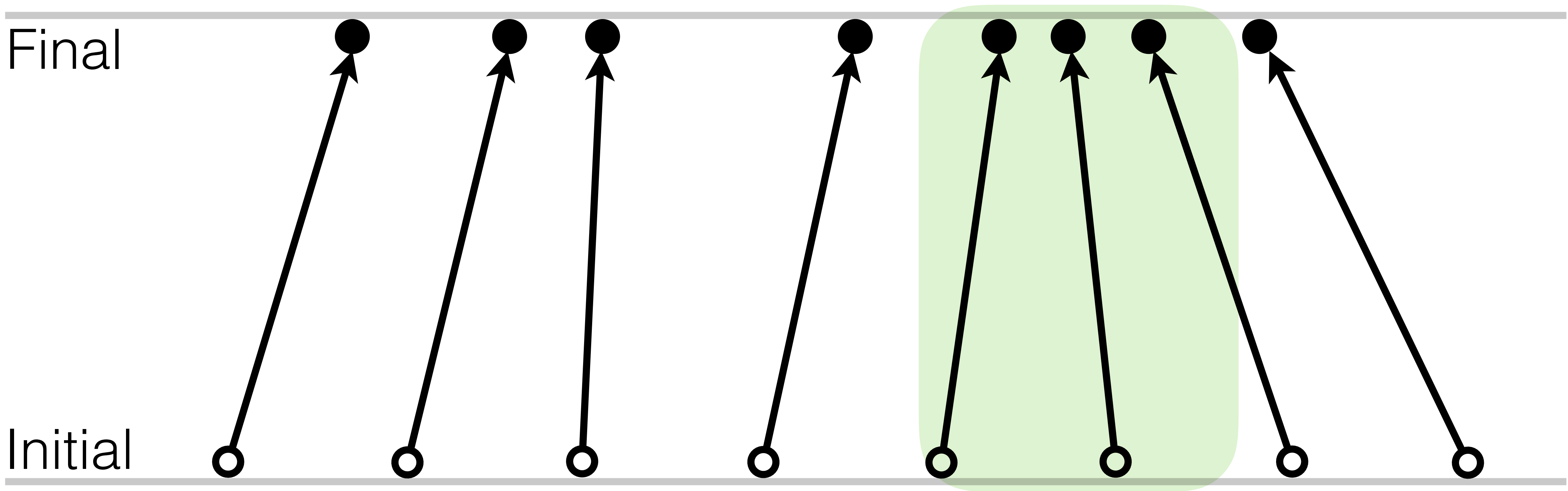
e.g. Baldauf+ (2016)

(2) Shell crossing: Trajectories cross each other



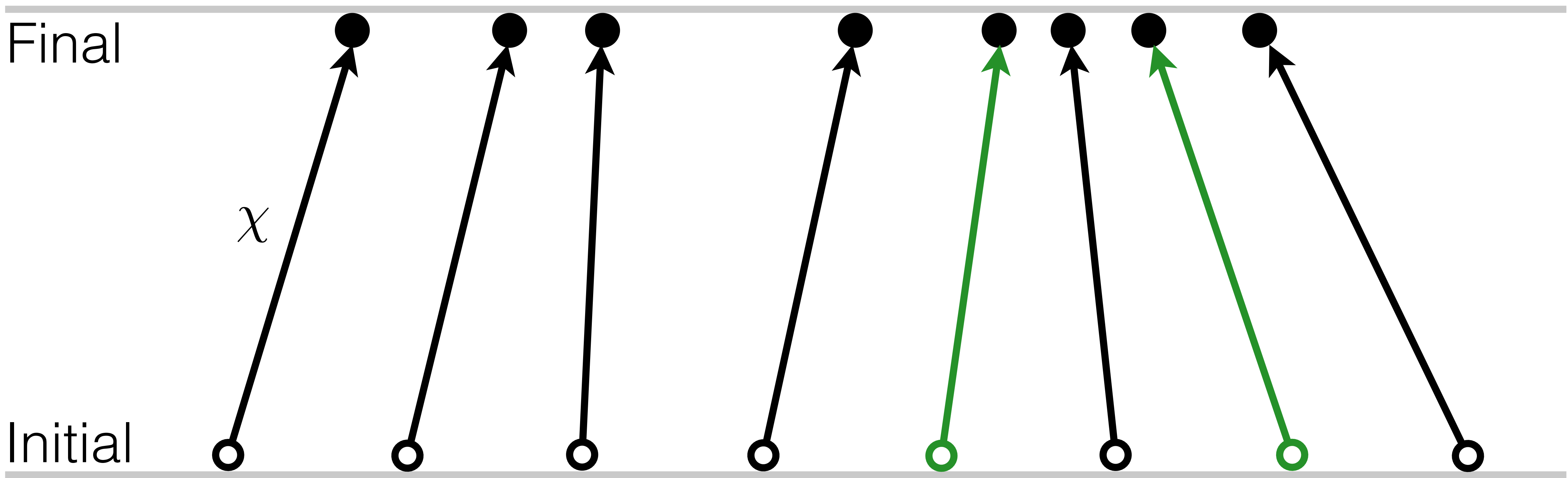
- Strongly nonlinear & difficult to model
- Seems like we cannot tell initial from final position (How many crossings happened?)
- Expect to loose memory of initial conditions

(2) Shell crossing: Trajectories cross each other



- Strongly nonlinear & difficult to model
- Seems like we cannot tell initial from final position (How many crossings happened?)
- Expect to loose memory of initial conditions

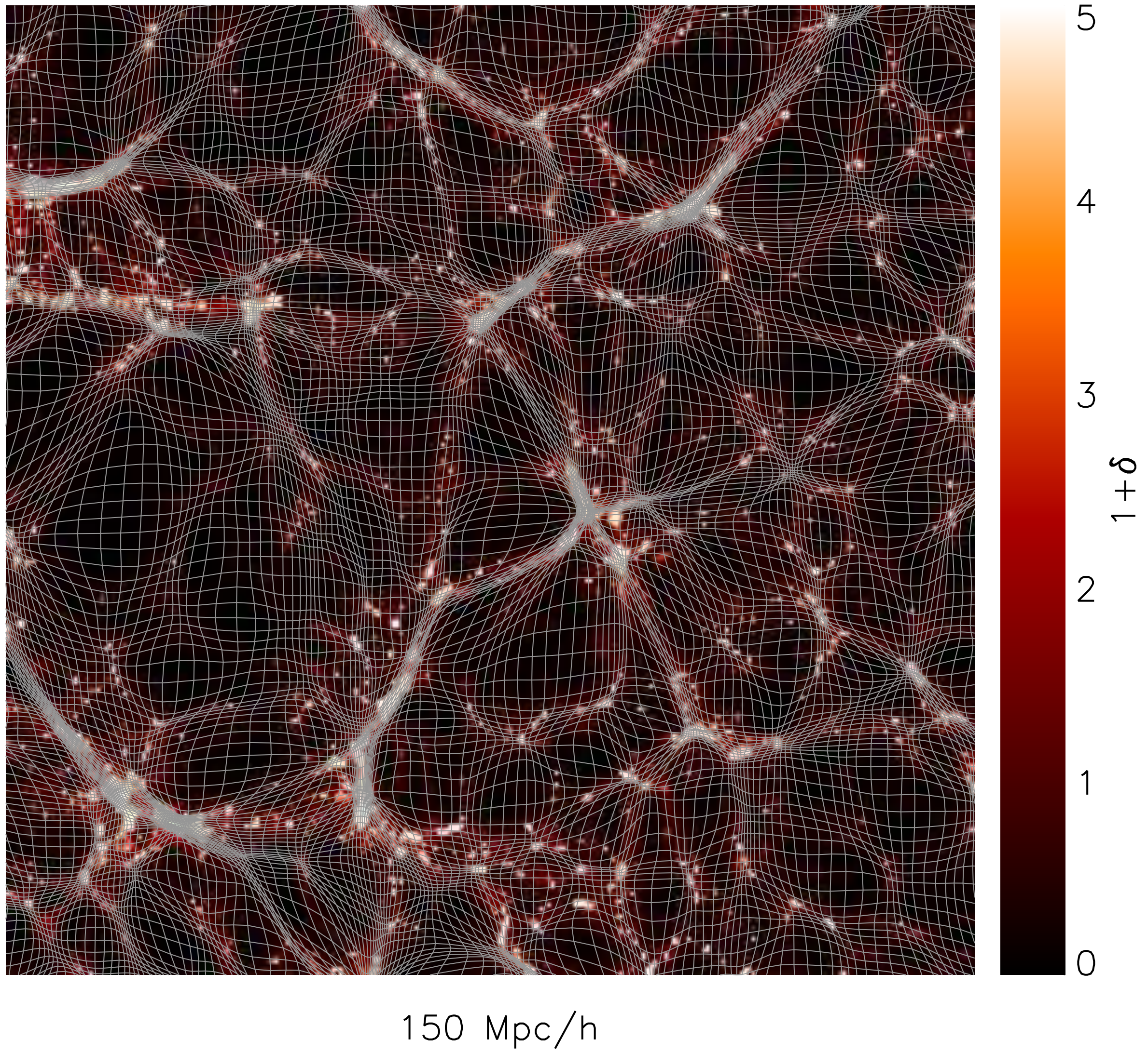
Reconstruction without shell crossing



- Estimate displacement as if there was no shell crossing
- This displacement is pretty linear, so can estimate linear density as

$$\delta_{\rm lin} = \nabla \cdot \chi$$

Algorithm 1: Isobaric/nonlinear reconstruction

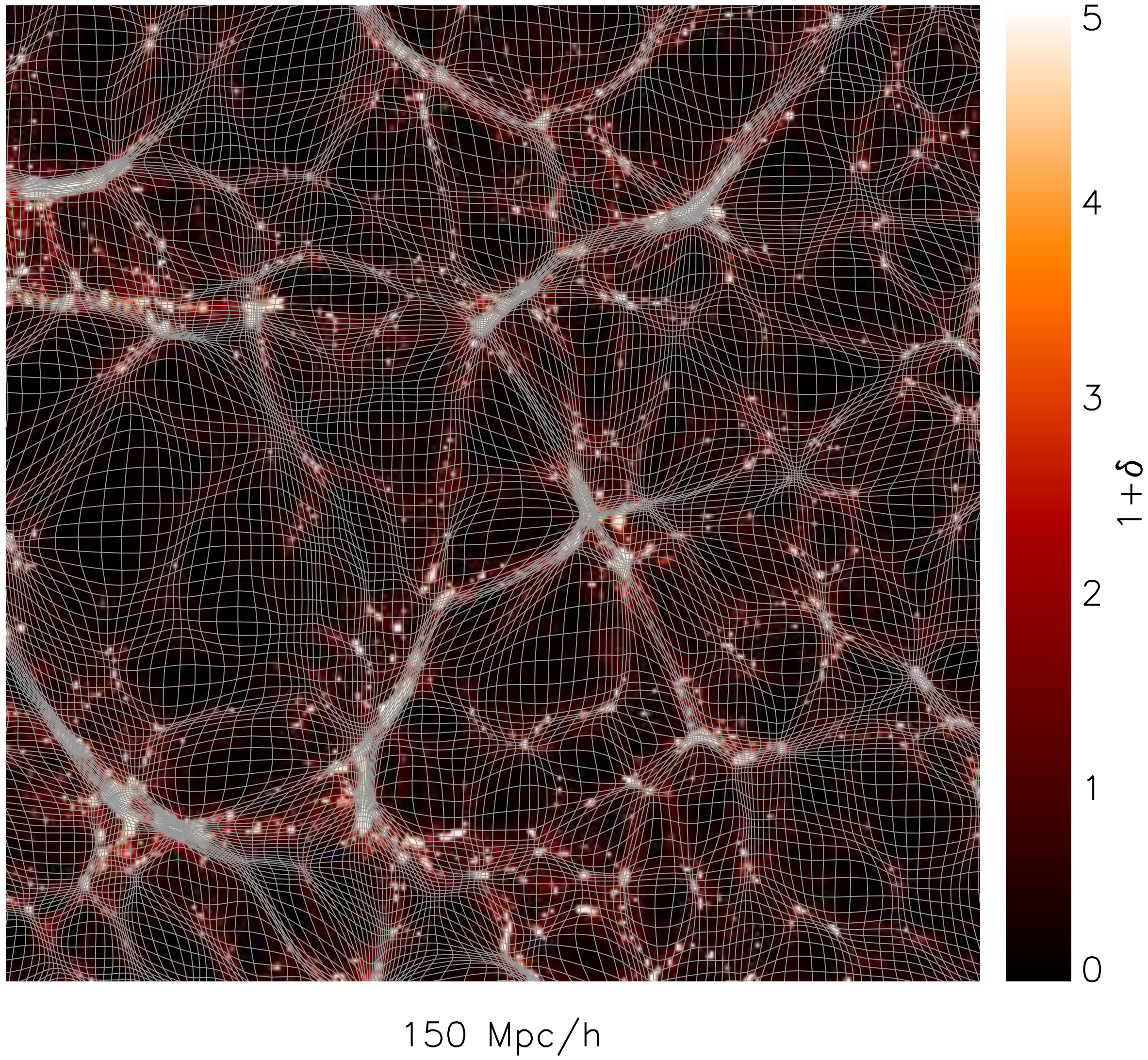


Each volume element has same mass

Get χ by continuously distorting mesh until δ=0 using a moving mesh code

H.M. Zhu, Y. Yu, U.L. Pen, X. Chen & H.R. Yu (2017)
Several more papers with X. Wang, Q. Pan & D. Inman (2017); also PIZA/MAK reconstruction

Algorithm 2: Iterative reconstruction

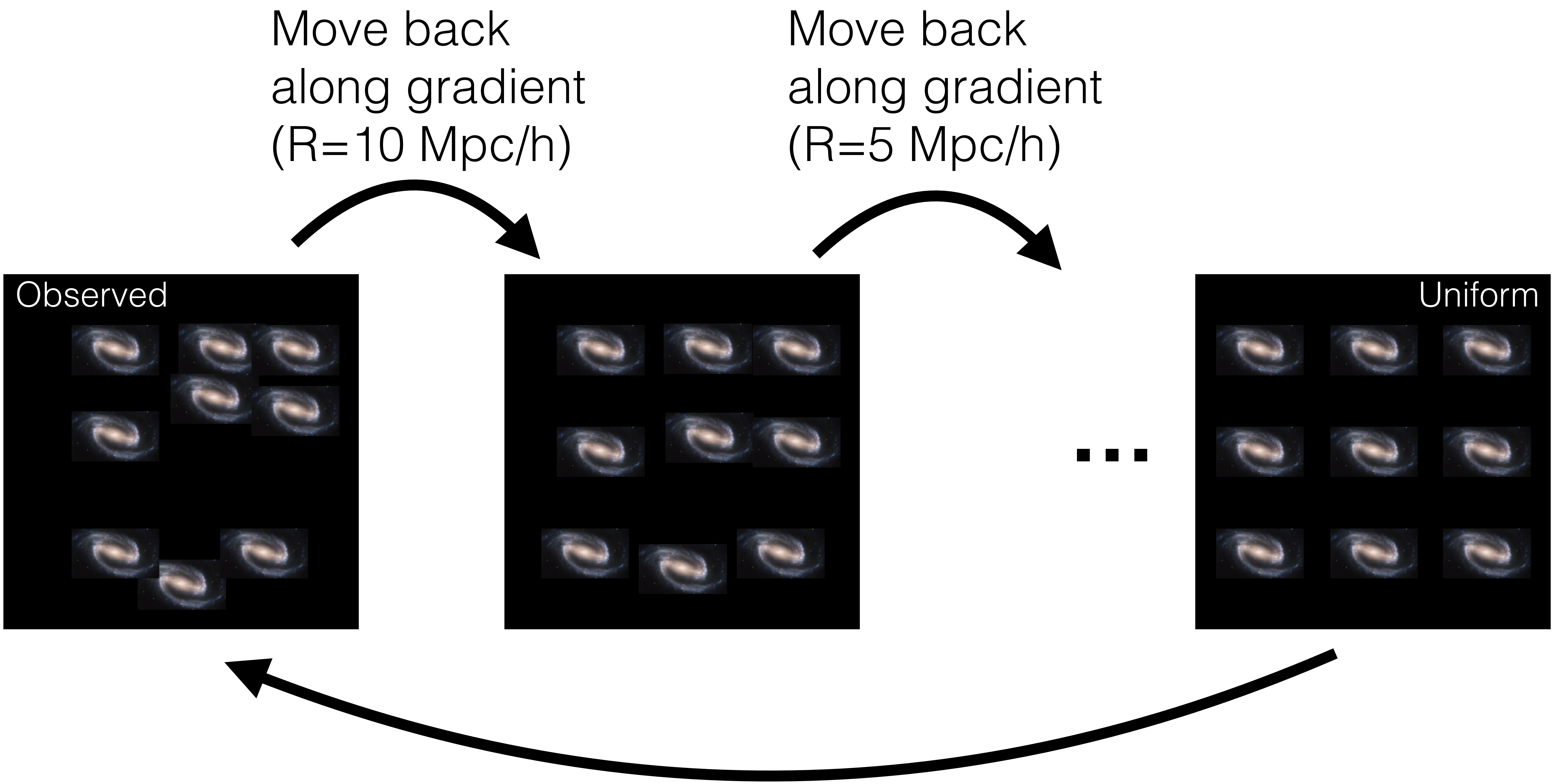


Same idea, but get displacement by iteratively applying Zeldovich displacements

Start with large smoothing scale to achieve coherence on large scales; then decrease smoothing scale iteratively

MS, Baldauf & Zaldarriaga (2017)

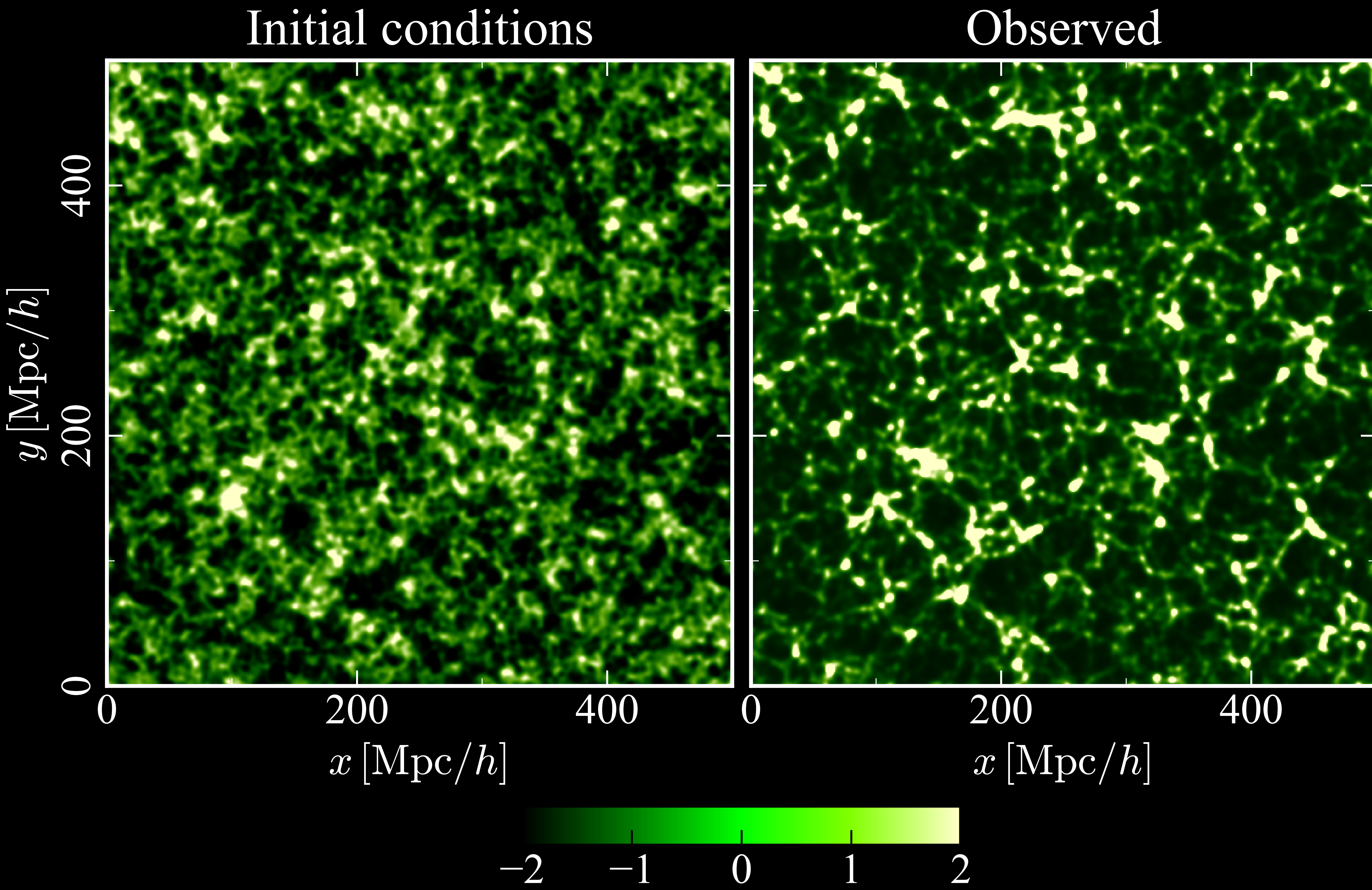
Our reconstruction algorithm



Measure total displacement $\chi(\mathbf{q})$

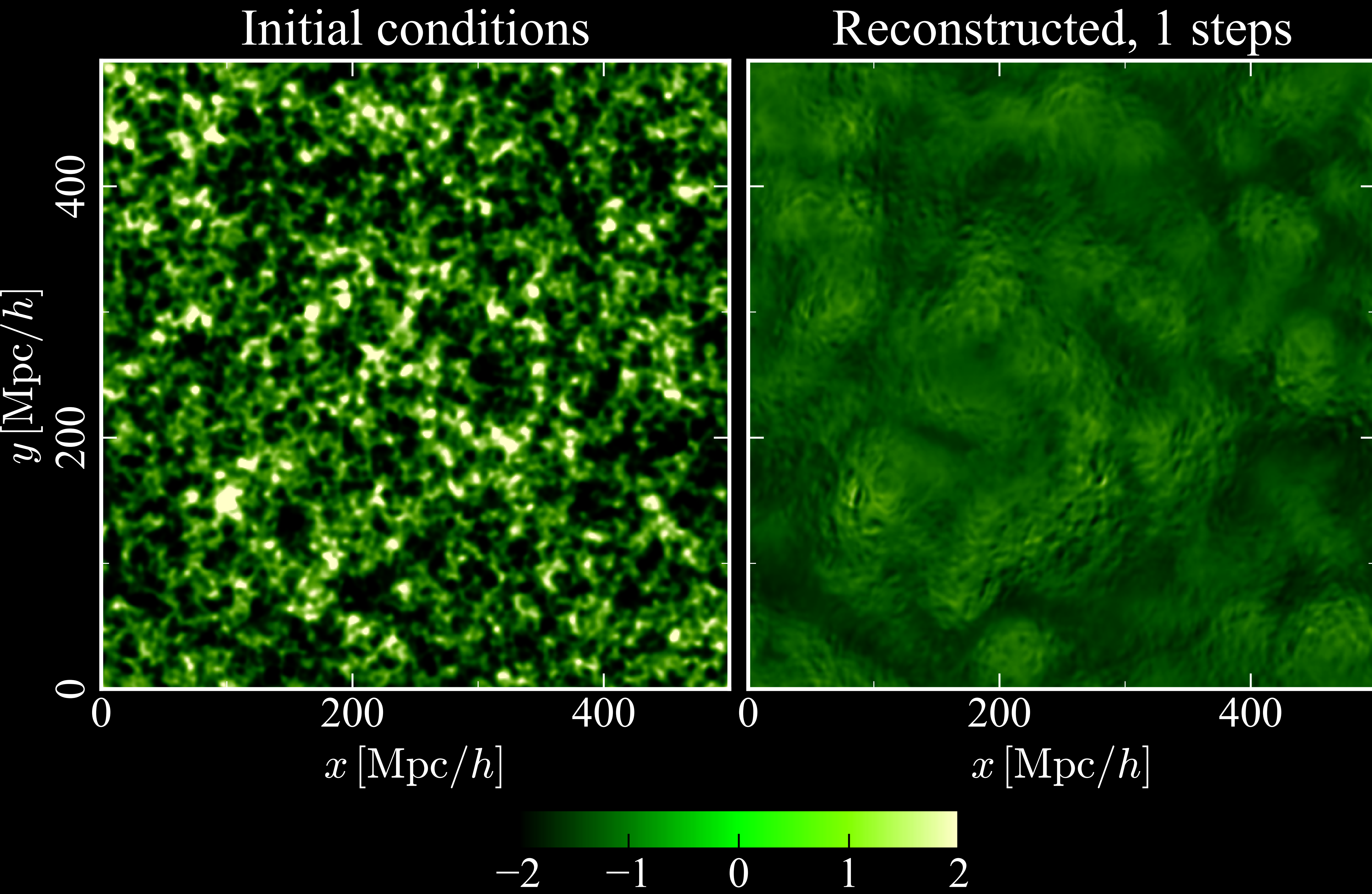
Estimate linear density as $\hat{\delta}_{\rm lin} = \nabla \cdot \chi(\mathbf{q})$

MS, Baldauf & Zaldarriaga (2017)



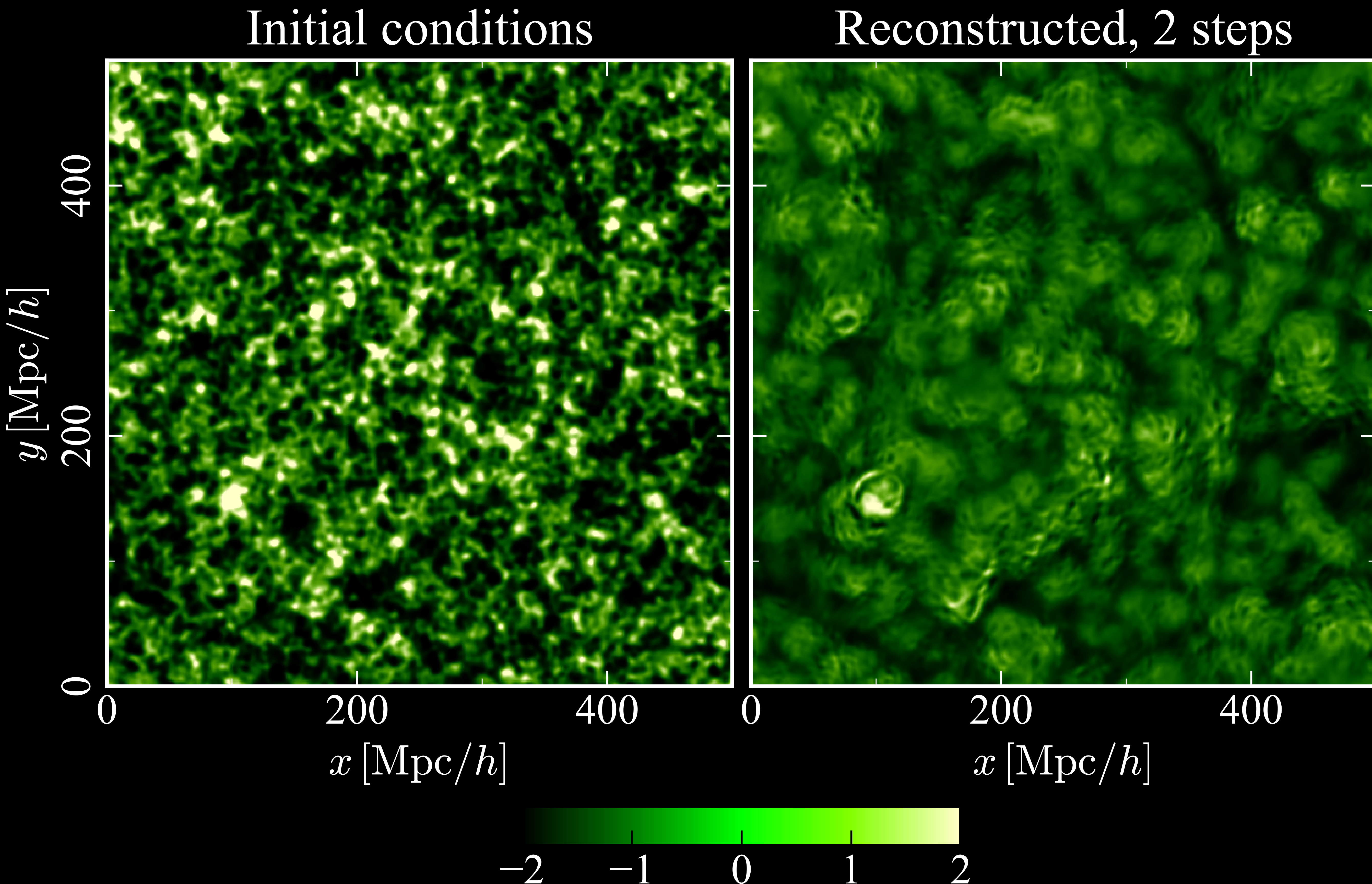
MS, Baldauf & Zaldarriaga (2017)

2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



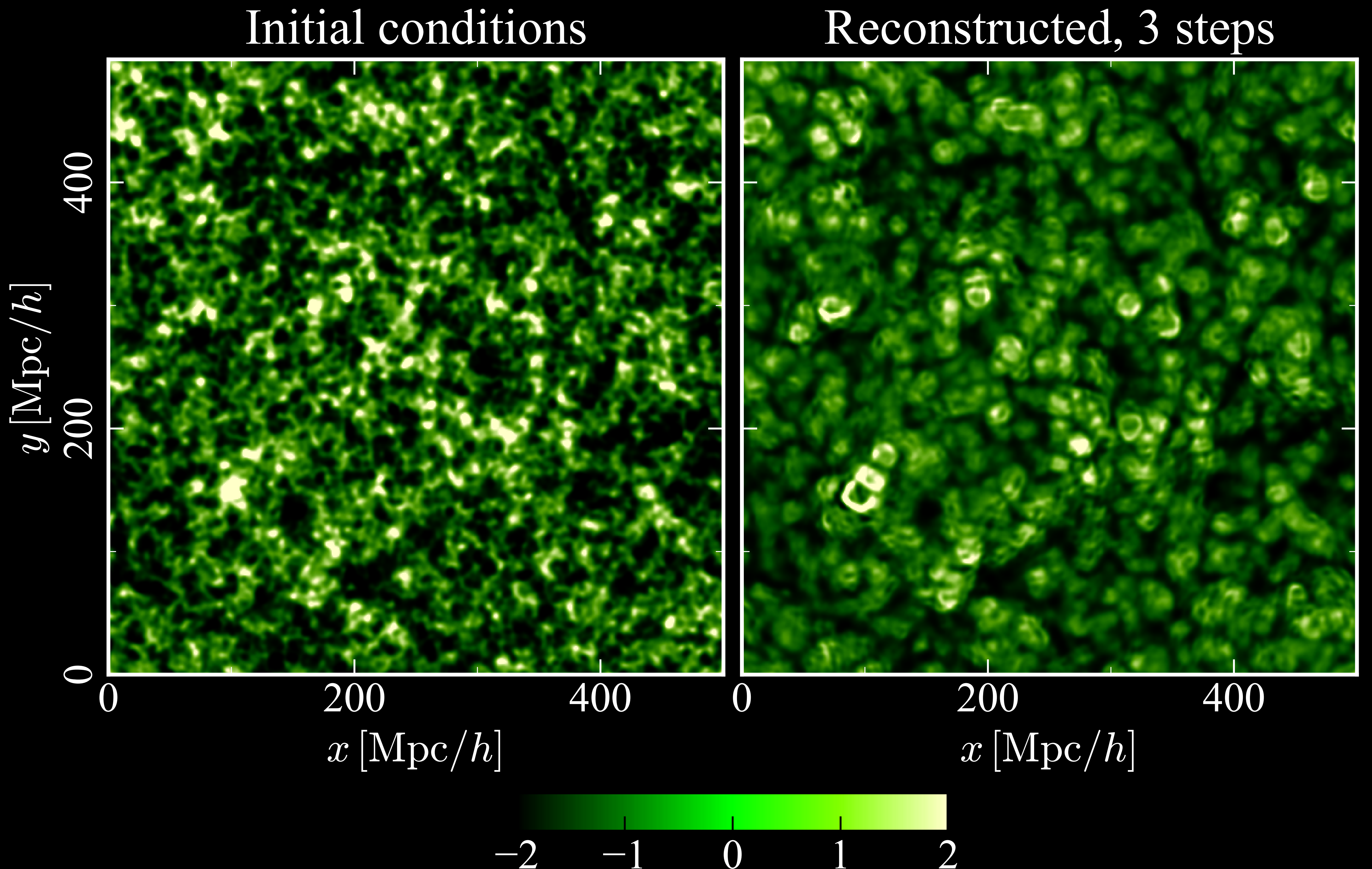
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



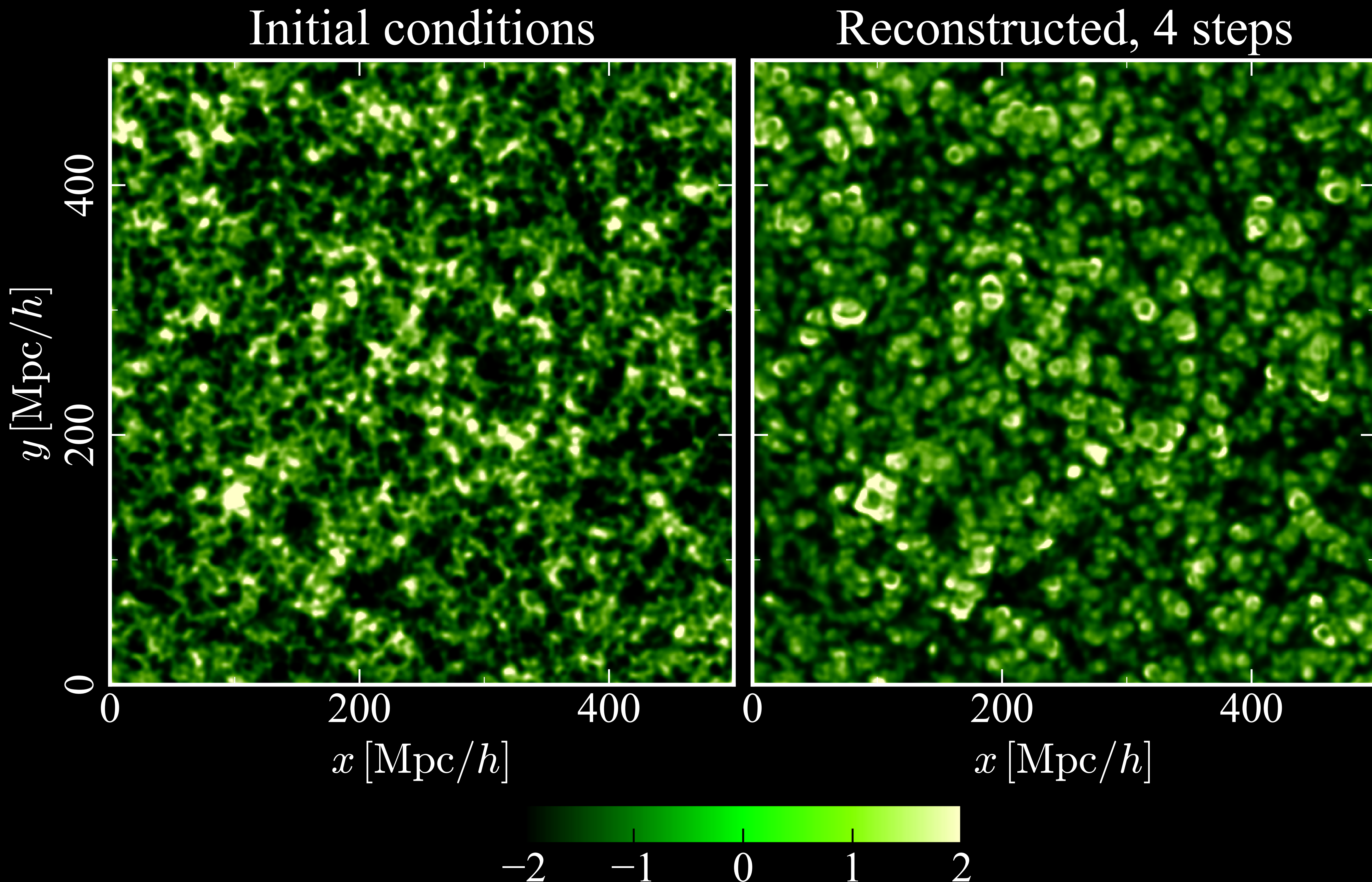
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



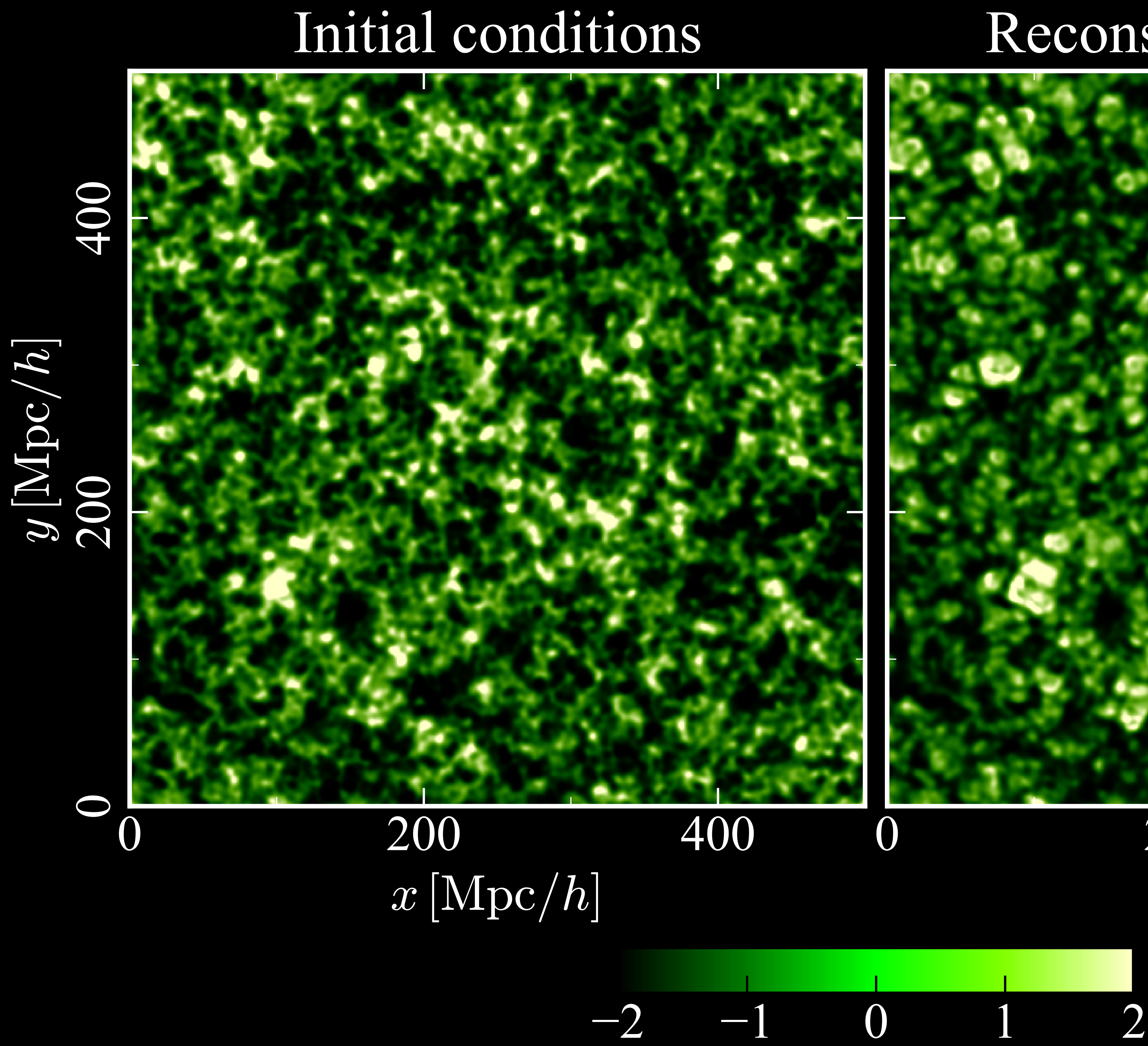
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



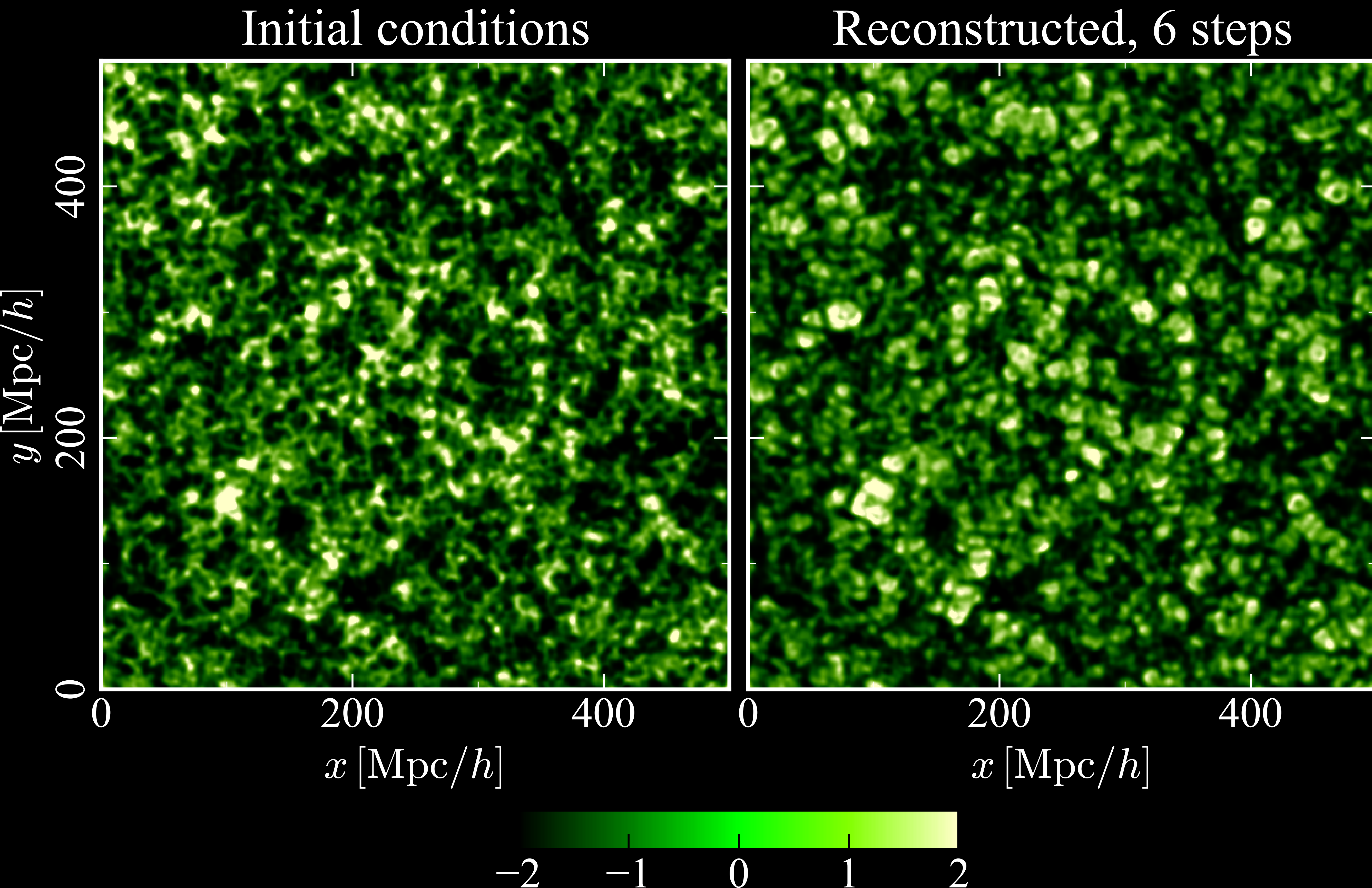
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



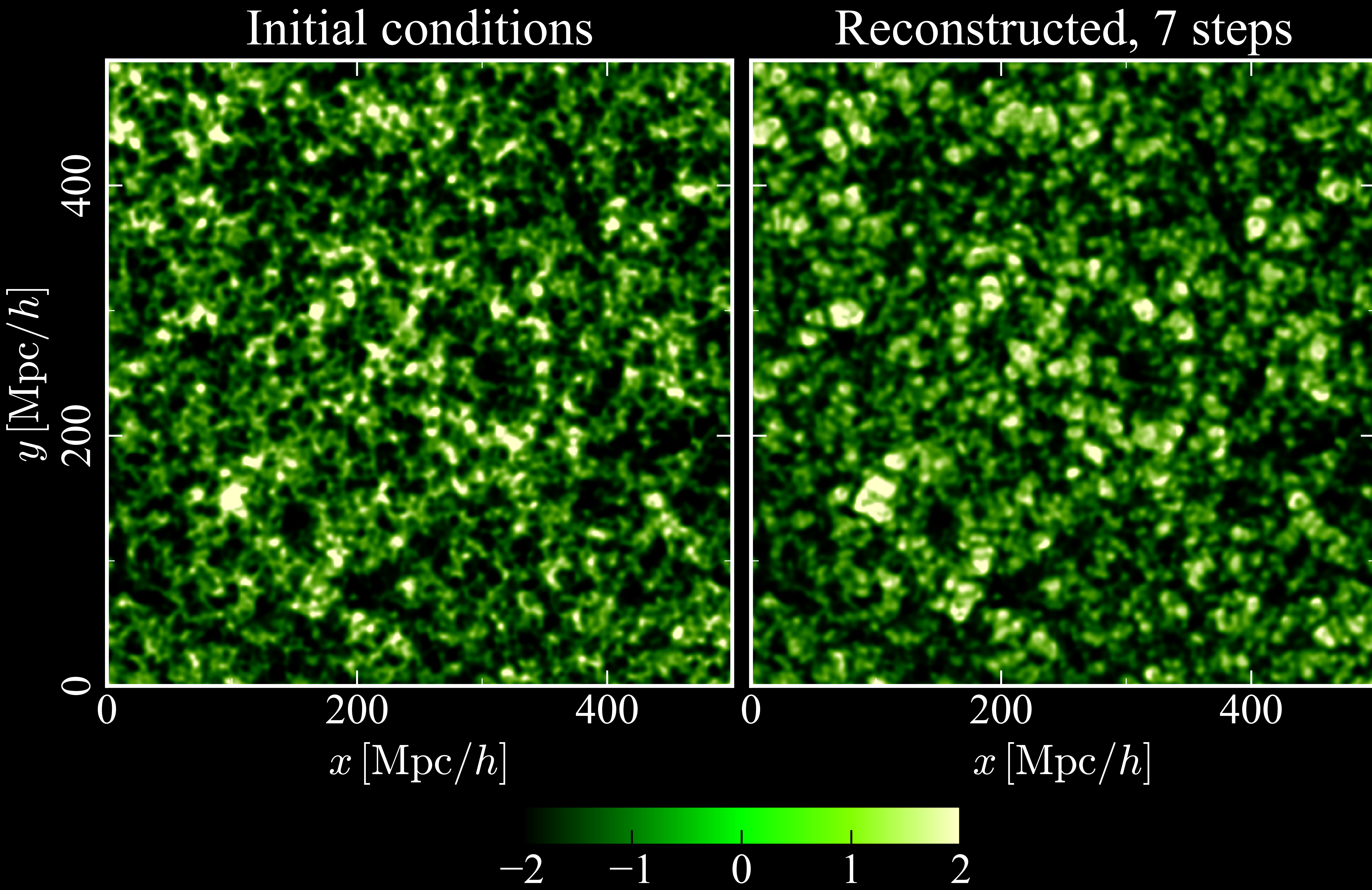
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



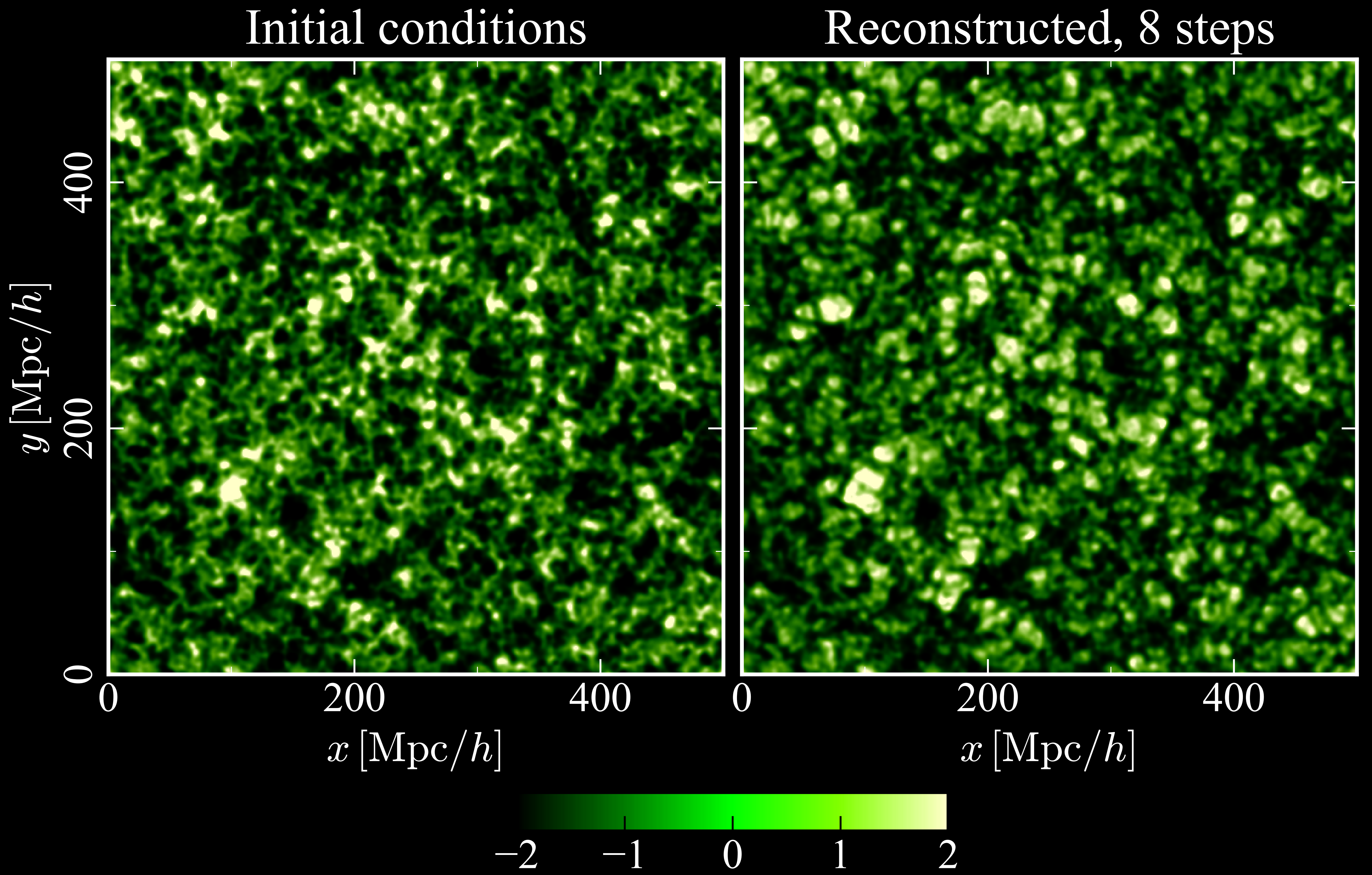
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box



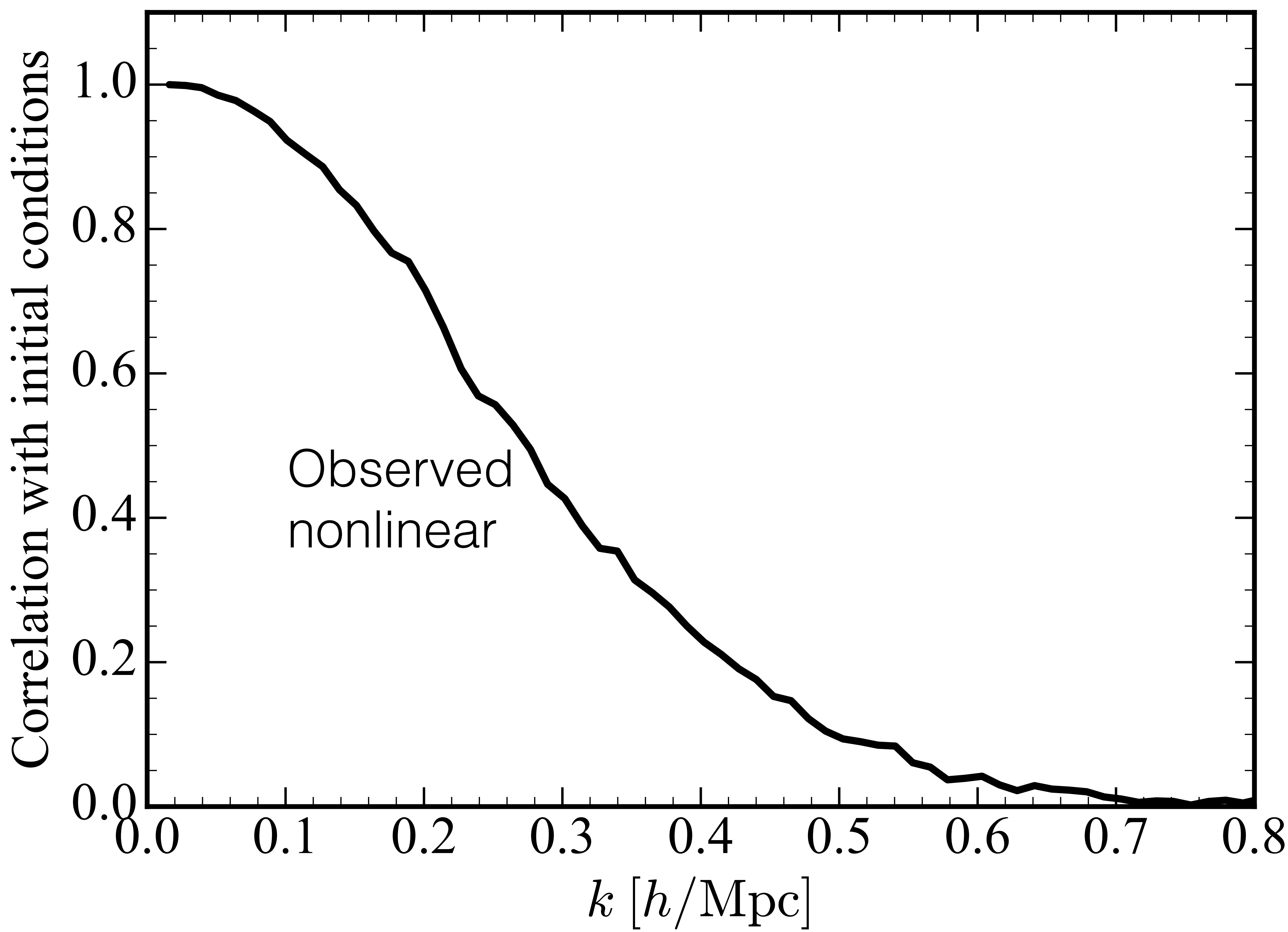
MS, Baldauf & Zaldarriaga (2017)

1st order reconstruction
2D slices of 3D density smoothed with R=2 Mpc/h Gaussian
1% subsample of 2048^3 DM particles in 500 Mpc/h per-side box

Correlation coefficient with initial conditions

Perfect correl.

No correl.

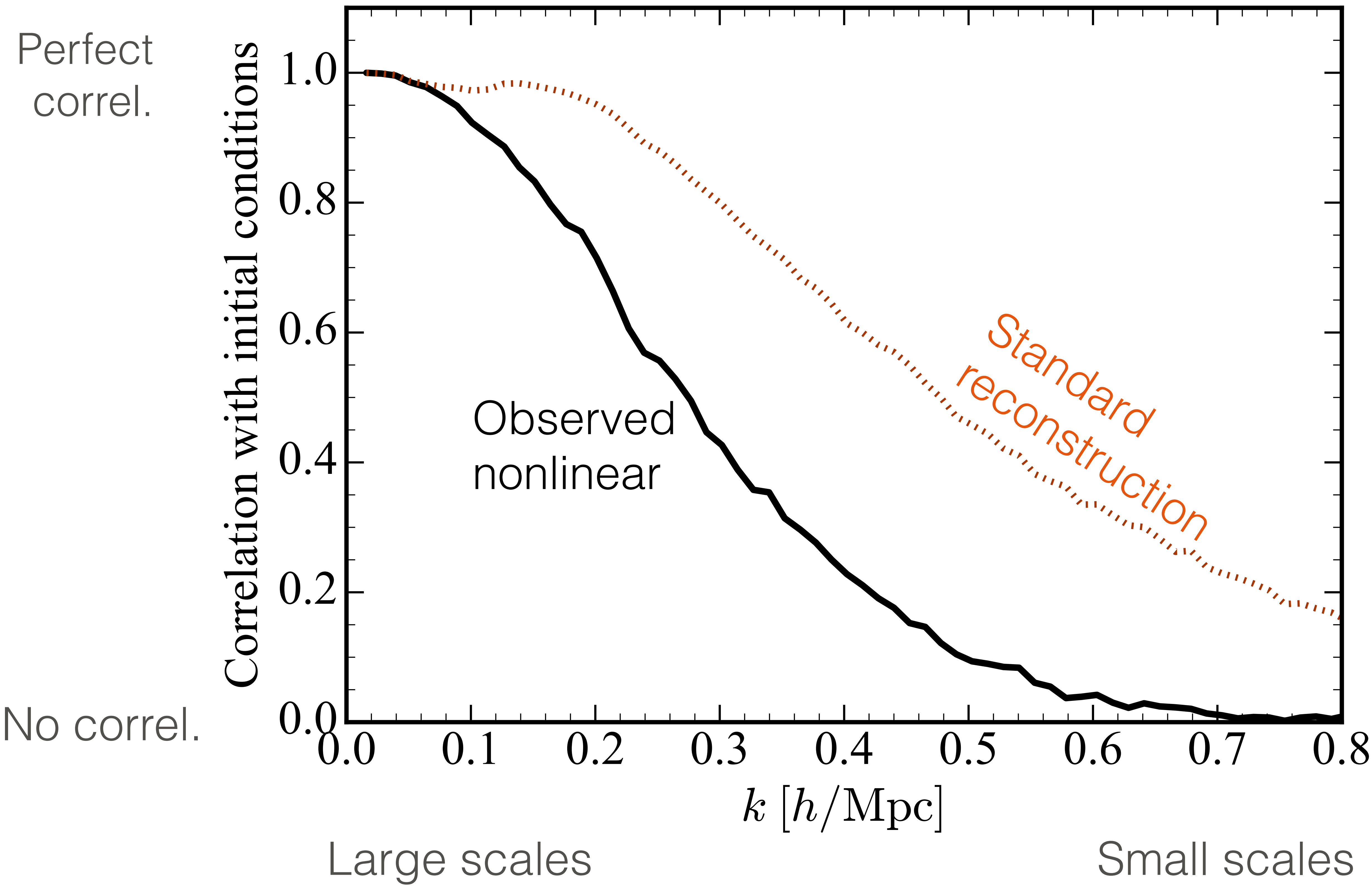


Large scales

Small scales

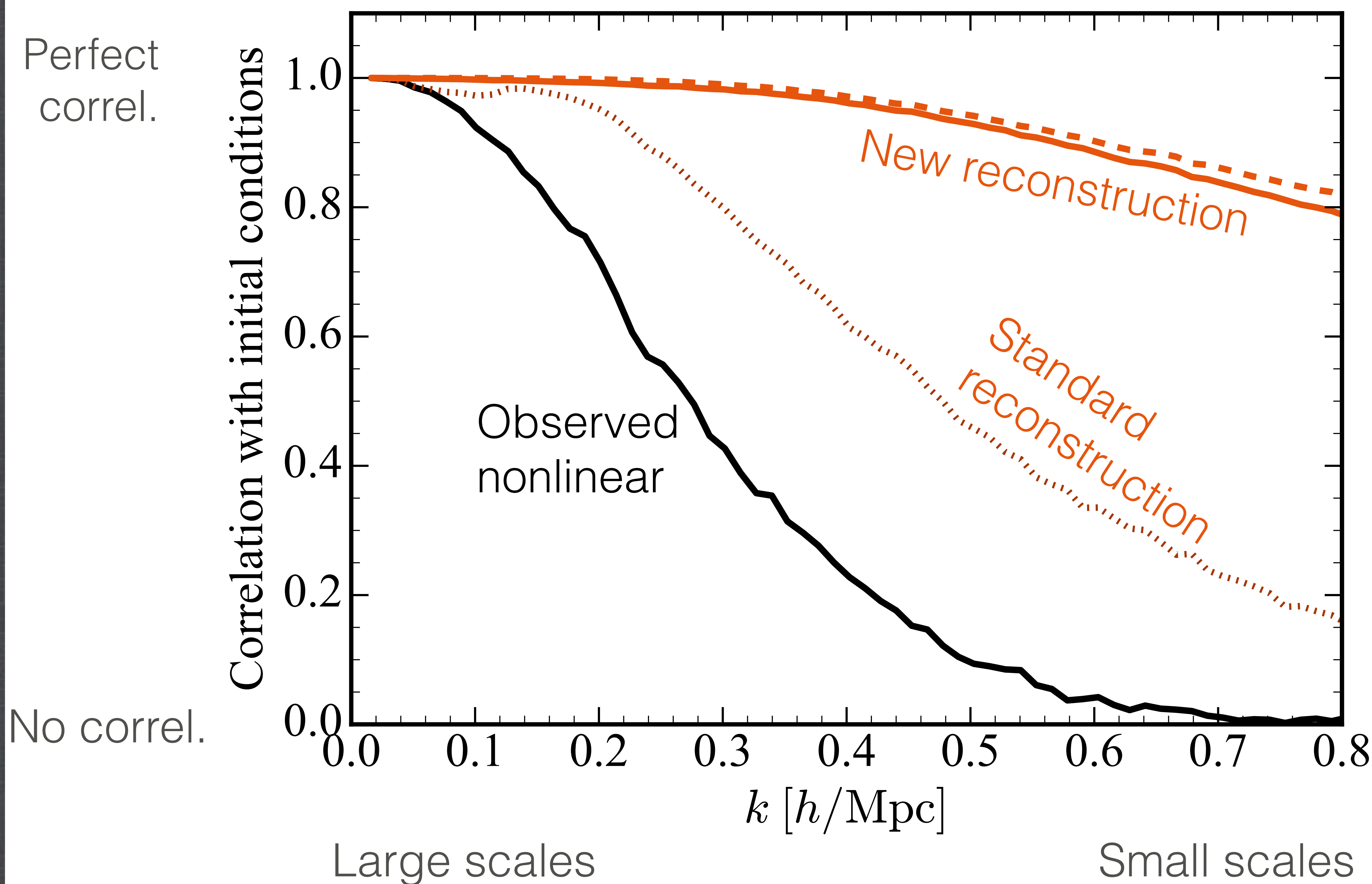
MS, Baldauf, Zaldarriaga (2017); noise-free 4096^3 DM simulations at z=0.6; also see Zhu+ (2017)

Correlation coefficient with initial conditions



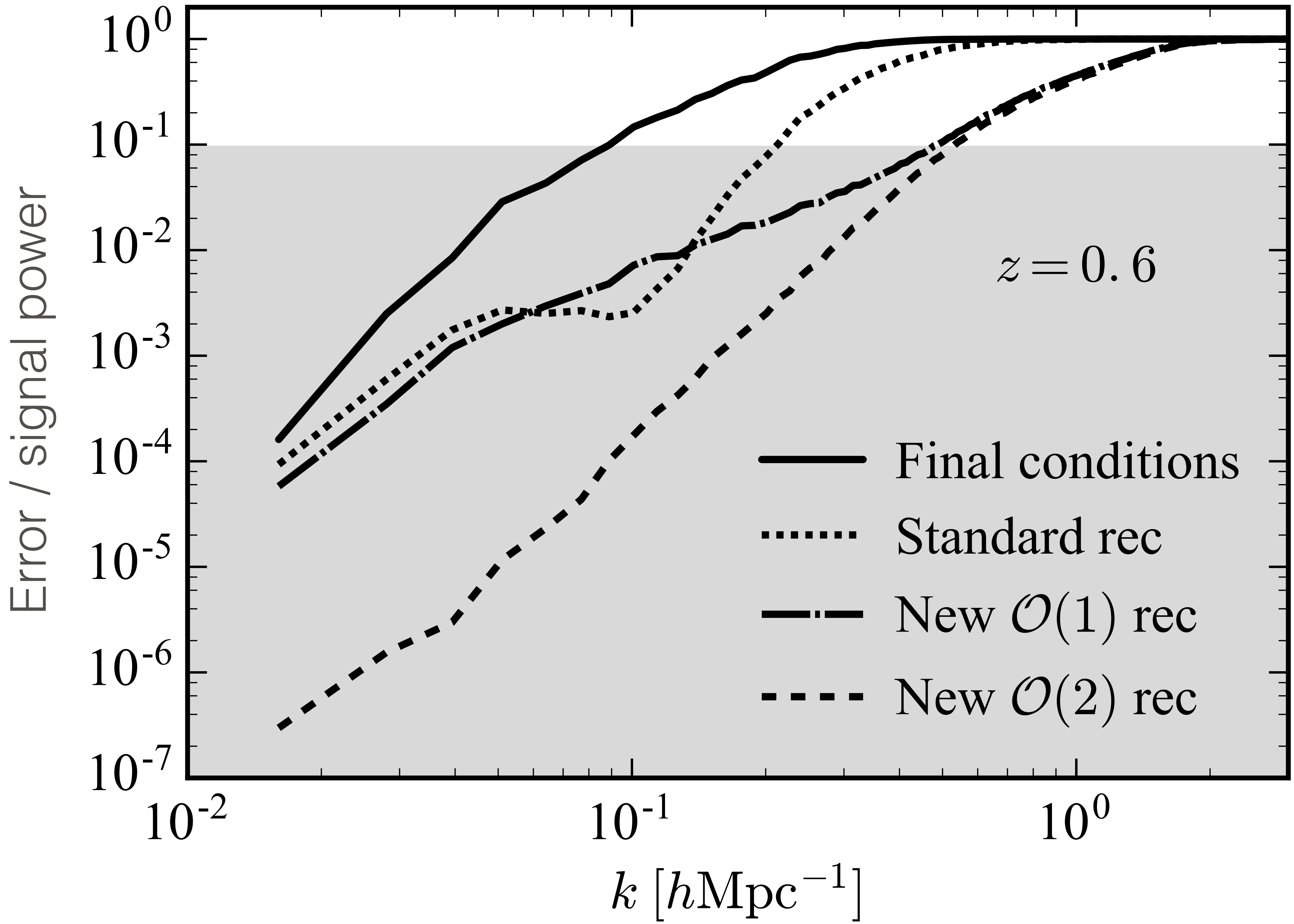
MS, Baldauf, Zaldarriaga (2017); noise-free 4096^3 DM simulations at z=0.6; also see Zhu+ (2017)

Correlation coefficient with initial conditions



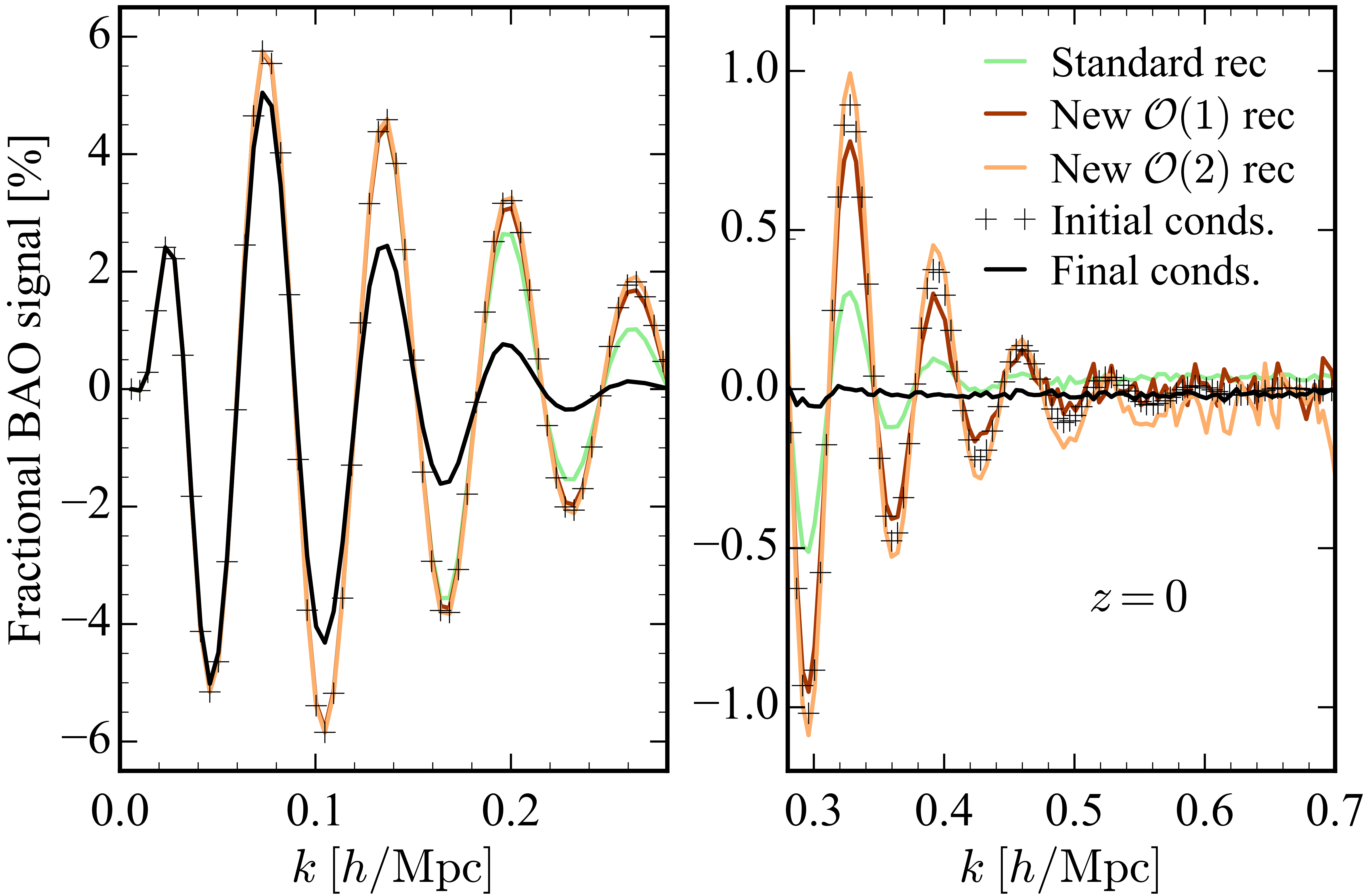
MS, Baldauf, Zaldarriaga (2017); noise-free 4096^3 DM simulations at z=0.6; also see Zhu+ (2017)

Size of fractional mistake (relative to linear)



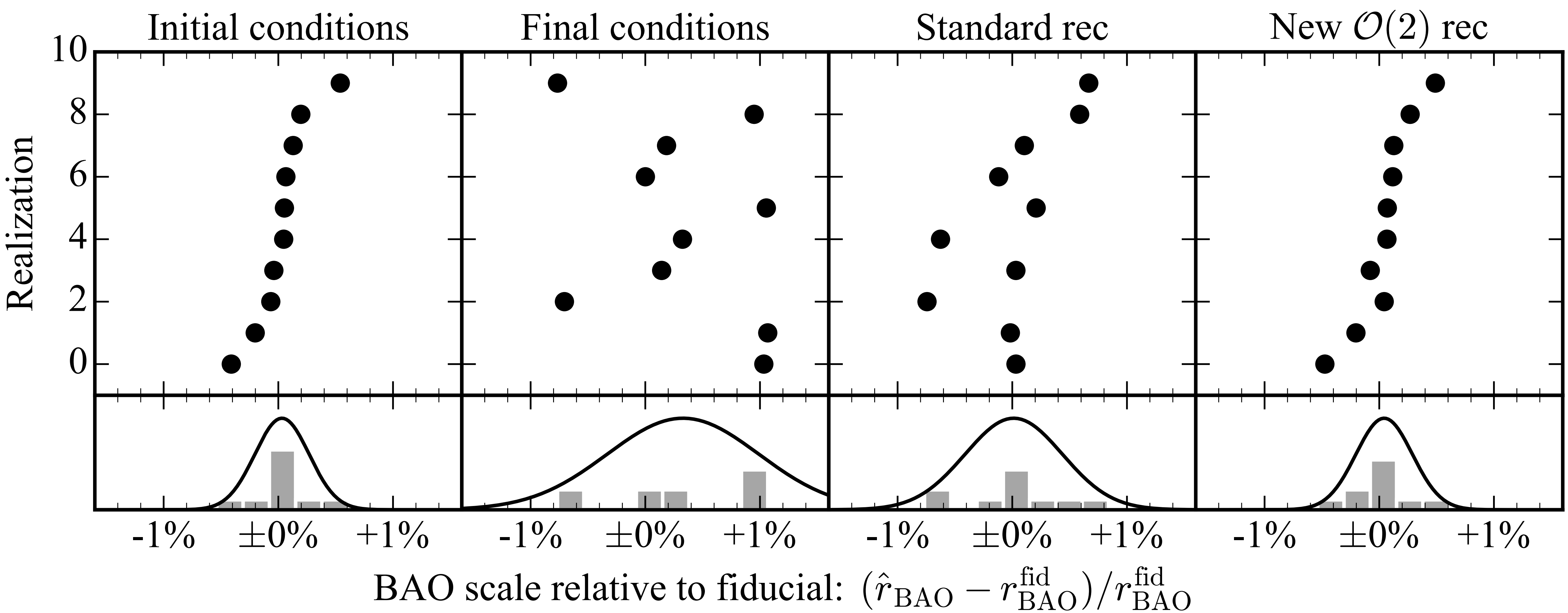
MS, Baldauf, Zaldarriaga (2017)

BAO signal



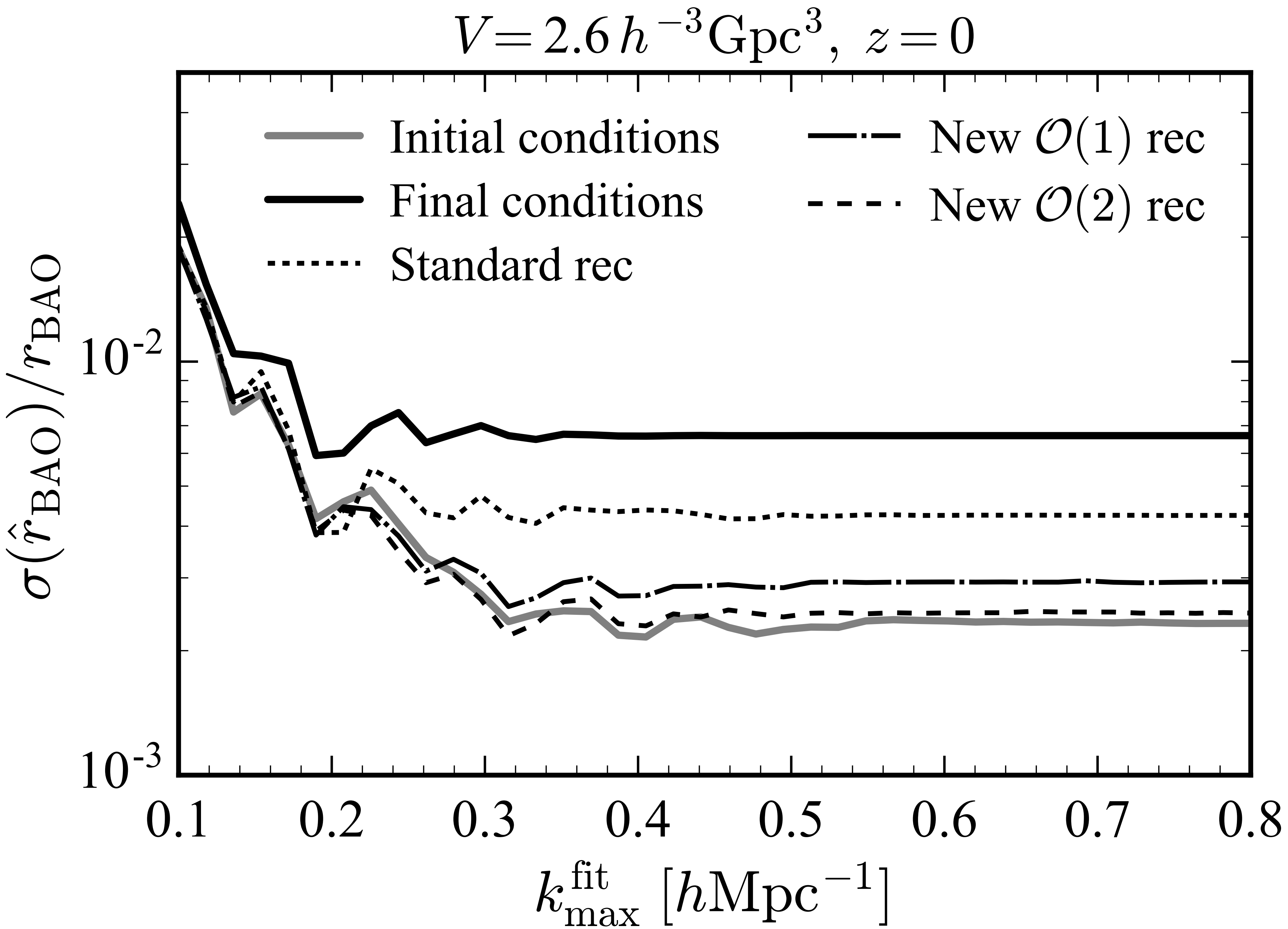
MS, Baldauf, Zaldarriaga (2017); also see Wang, Yu, Zhu, Yu, Pan & Pen (2017)

Best-fit BAO scale in 10 simulations



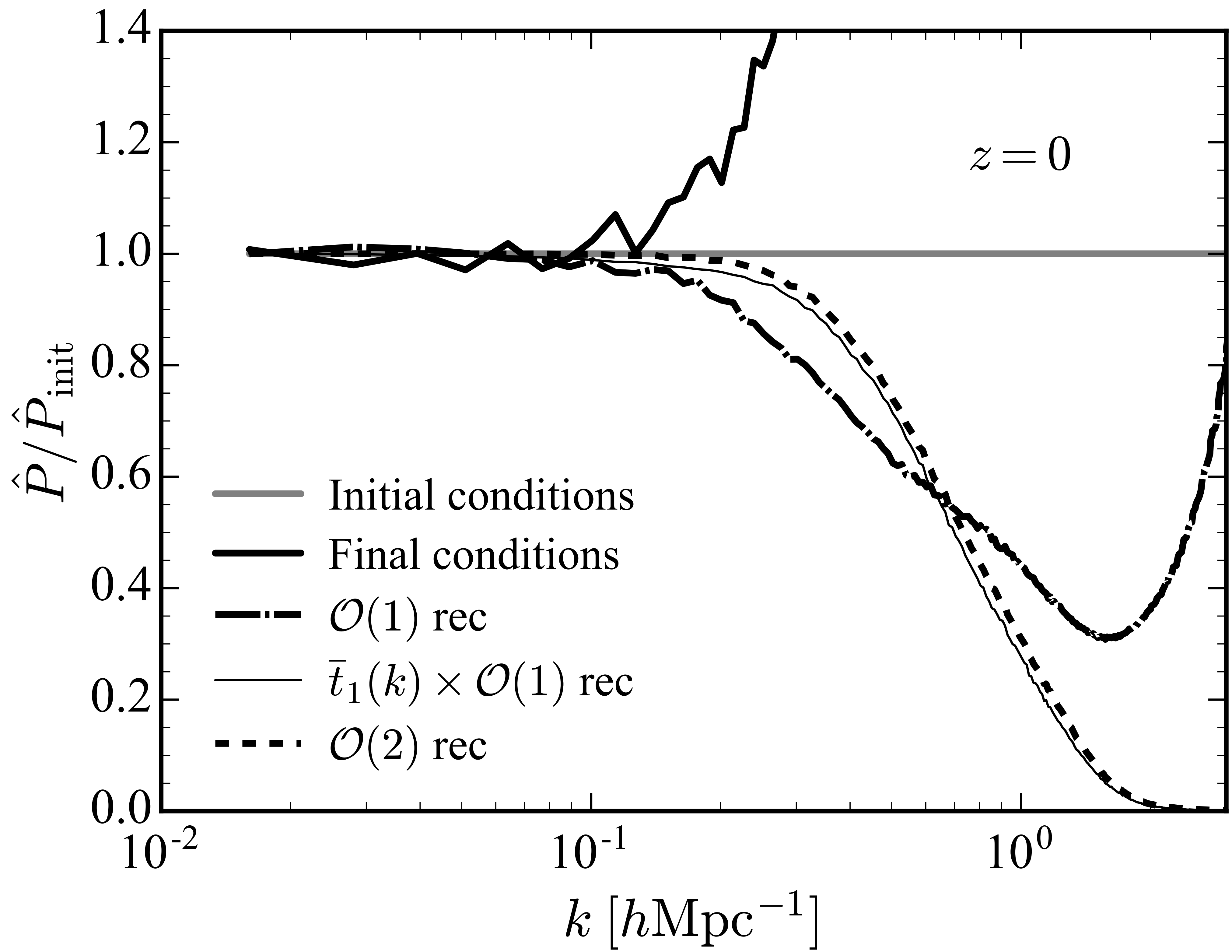
MS, Baldauf, Zaldarriaga (2017); also see Wang, Yu, Zhu, Yu, Pan & Pen (2017)

Fractional error bar of BAO scale



MS, Baldauf, Zaldarriaga (2017); also see Wang, Yu, Zhu, Yu, Pan & Pen (2017)

Broadband power spectrum



MS, Baldauf, Zaldarriaga (2017); also see Pan, Pen, Inman & Yu (2017)

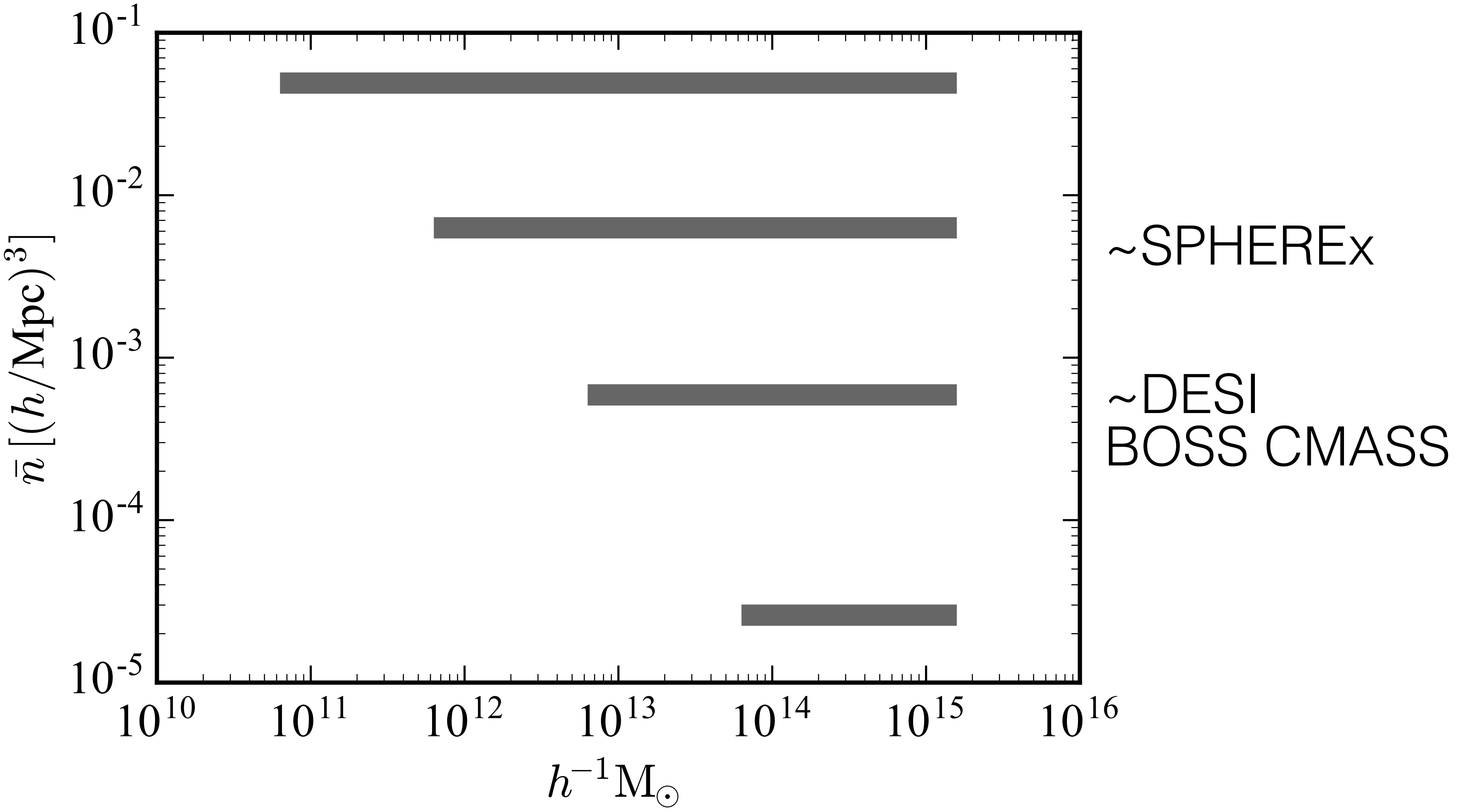
Challenges

Add realism:

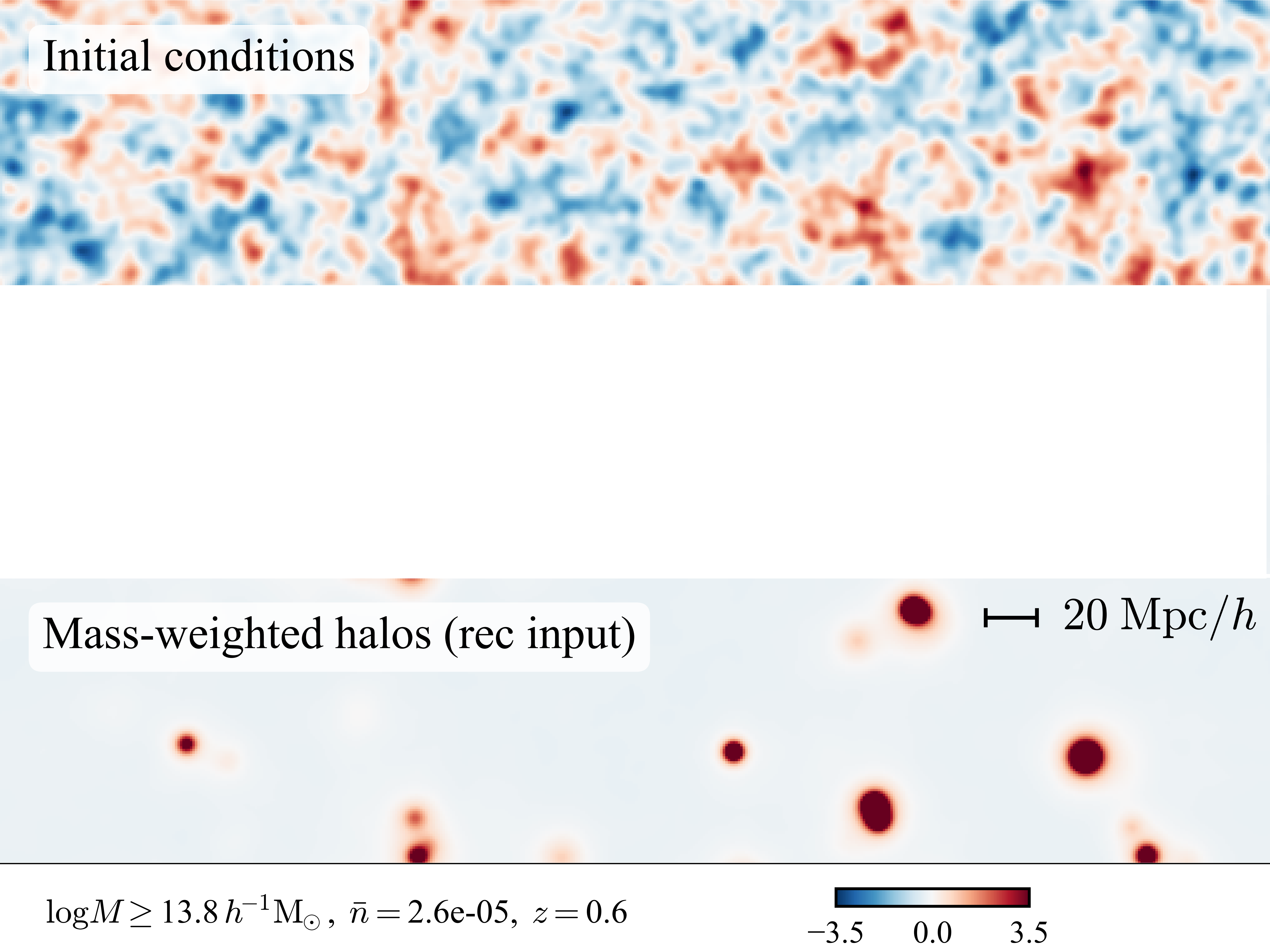
- Shot noise
- Halo/galaxy bias (doing right now)
- Redshift space distortions
- Survey mask & depth variation (inhomogeneous noise)
- What happens to primordial f_{NL} after reconstruction?

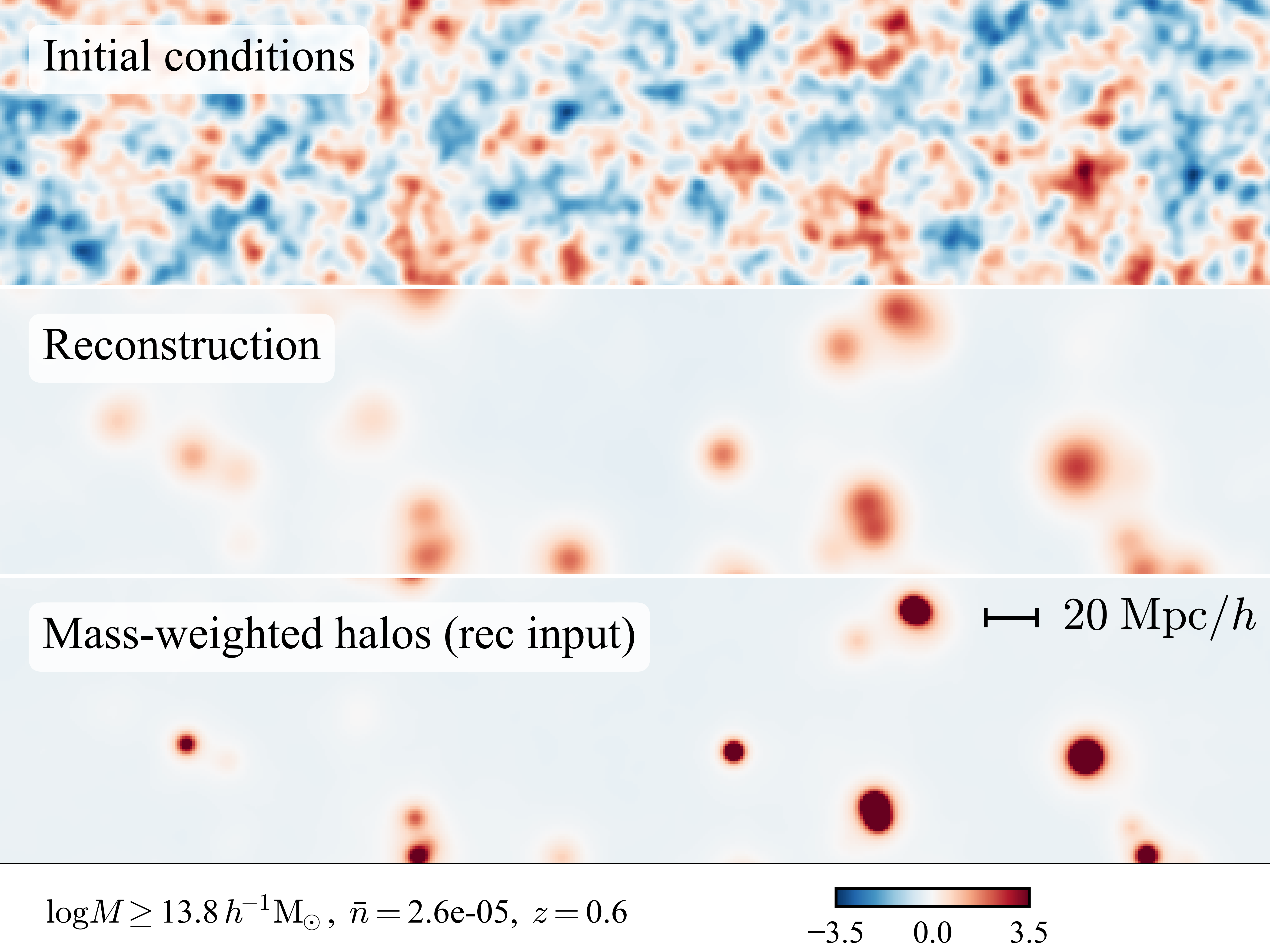
Halo reconstruction: Outlook (preliminary!)

- 4 halo mass bins at z=0.6

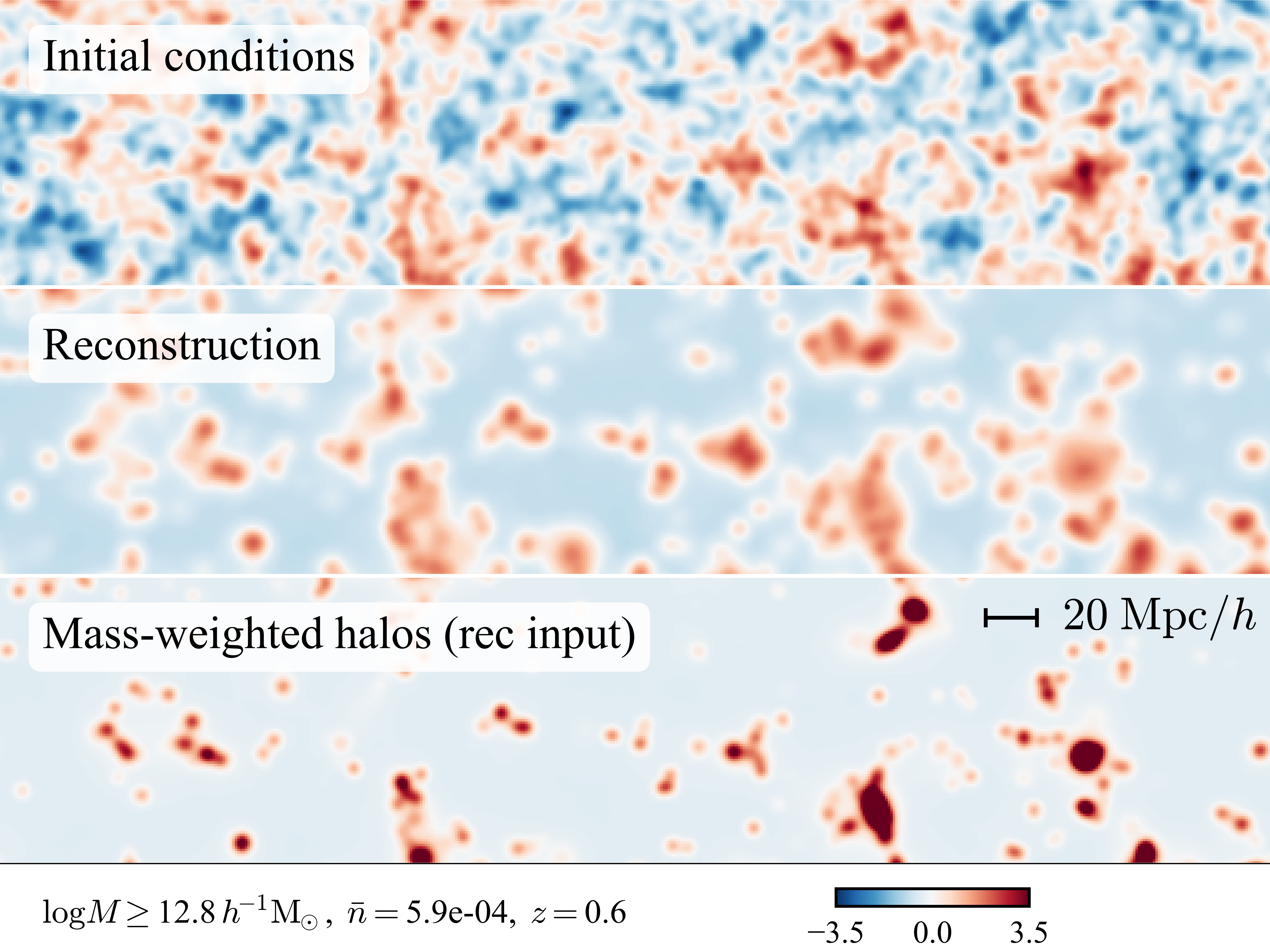


- Weigh by halo mass
- 5 MP-Gadget sims with 1536^3 particles, L=500Mpc/h, FOF halos

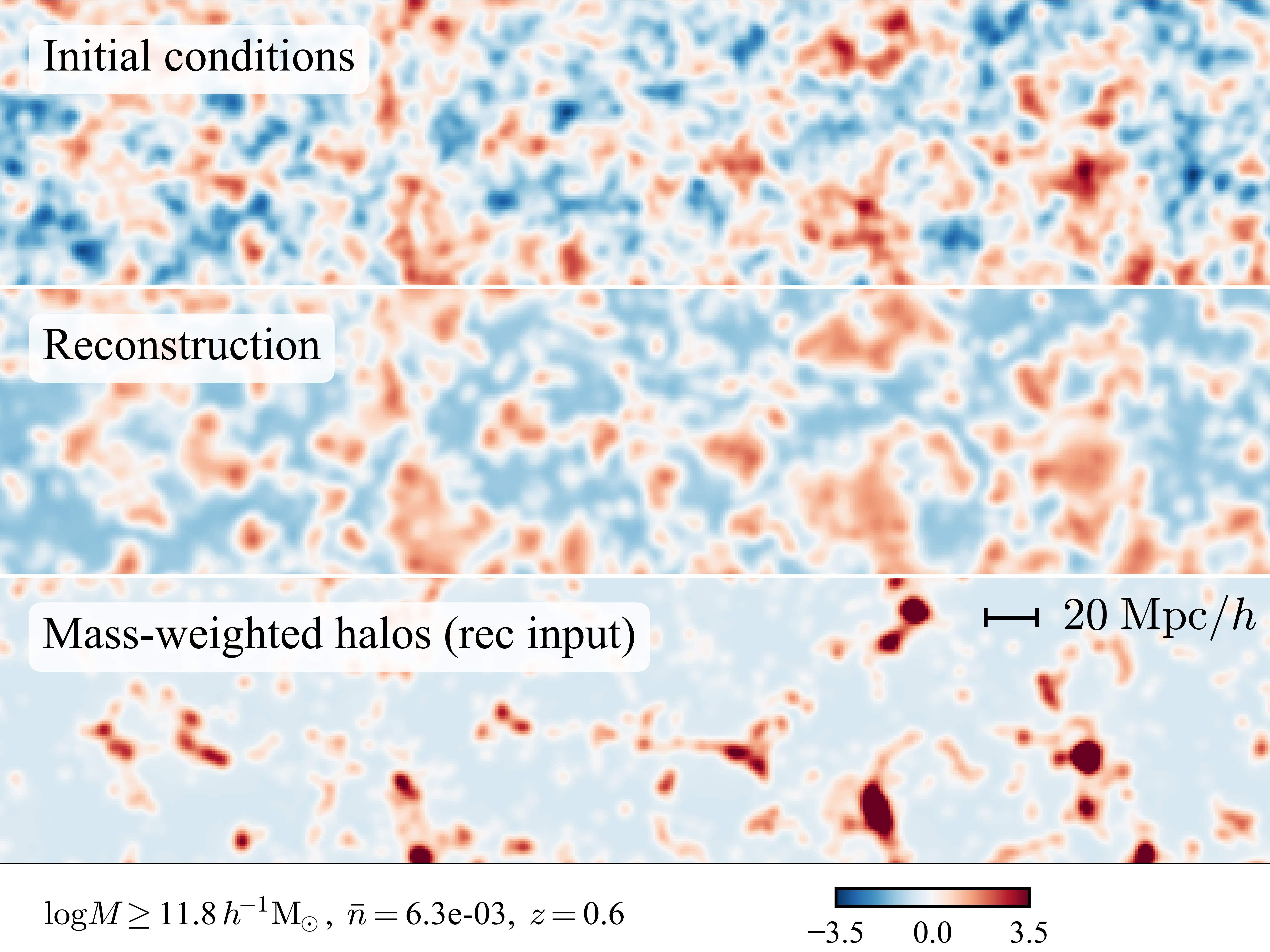

Initial conditions
Mass-weighted halos (rec input)
20 Mpc/h
$\log M \geq 13.8\,h^{-1}\mathrm{M}_\odot$, $\bar{n} = 2.6\text{e-05}$, $z = 0.6$
−3.5
0.0
3.5

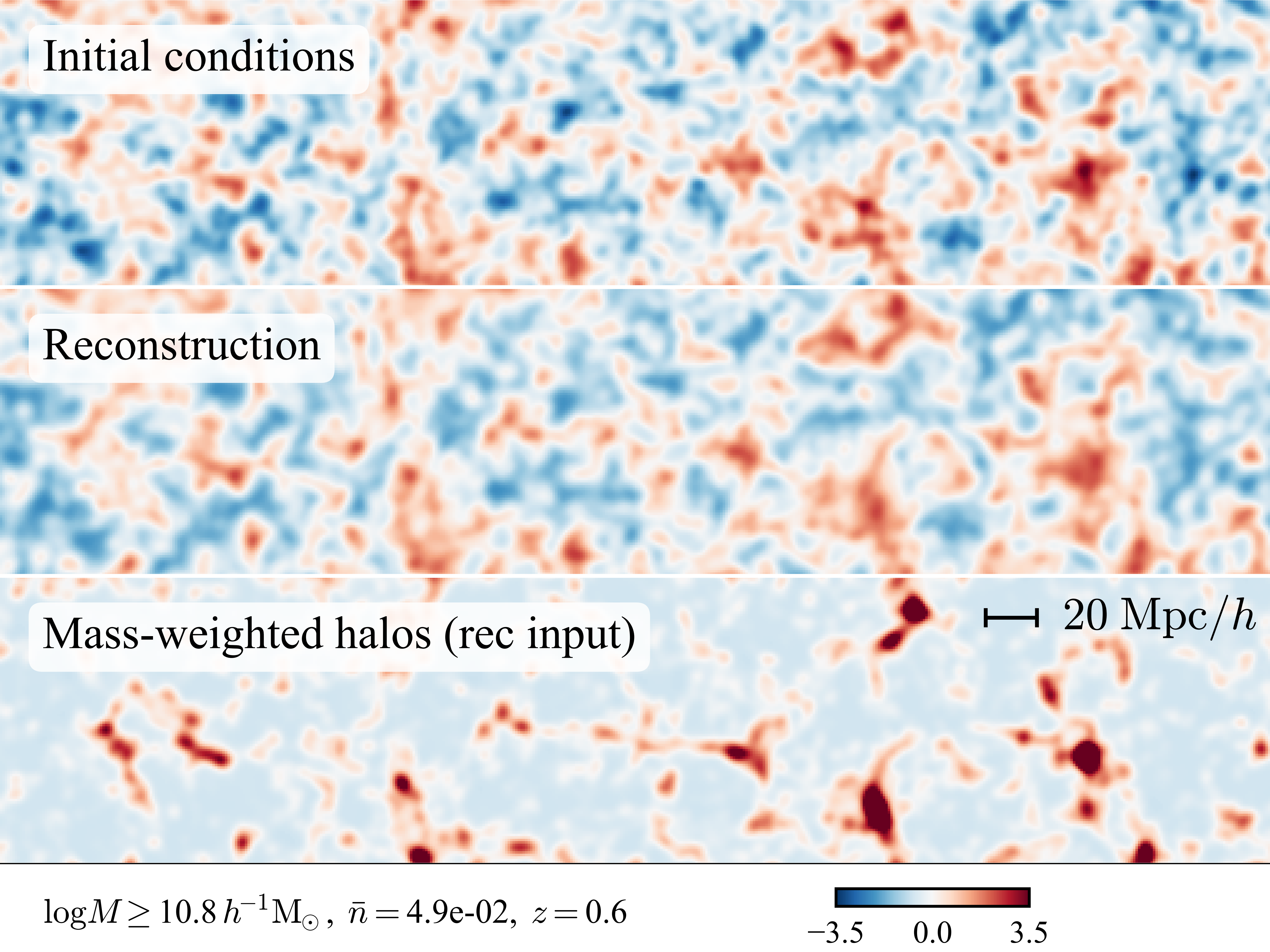


Initial conditions
Reconstruction
Mass-weighted halos (rec input)
20 Mpc/h
$\log M \geq 13.8\, h^{-1}\mathrm{M}_\odot$, $\bar{n} = 2.6\text{e-}05$, $z = 0.6$
−3.5
0.0
3.5



Initial conditions
Reconstruction
Mass-weighted halos (rec input)
20 Mpc/h
log M ≥ 12.8 h^−1 M_⊙ , n̄ = 5.9e-04, z = 0.6
−3.5
0.0
3.5


Initial conditions
Reconstruction
Mass-weighted halos (rec input)
20 Mpc/h
$\log M \geq 11.8\, h^{-1}\mathrm{M}_\odot$, $\bar{n} = 6.3\text{e-}03$, $z = 0.6$
−3.5
0.0
3.5



Initial conditions
Reconstruction
Mass-weighted halos (rec input)
20 Mpc/h
$\log M \geq 10.8\, h^{-1}\mathrm{M}_\odot$, $\bar{n} = 4.9\text{e-}02$, $z = 0.6$
−3.5
0.0
3.5

Conclusions: Part II

Nonlinear physics limits science return of galaxy surveys

Reconstruction can reduce that degradation

At z=0, reconstruction achieves >95% correlation with linear density at $k<0.35\ h\mathrm{Mpc}^{-1}$

Improve BAO signal-to-noise by factor 2.7 (z=0) to 2.5 (z=0.6)

70%-30% improvement over standard BAO reconstruction

Can improve LSS survey science (dark energy, Hubble constant, early universe physics)

Lots of work to be done to apply it to real data

Princeton cosmology seminar

Joint Princeton University/IAS cosmology seminar on Mondays

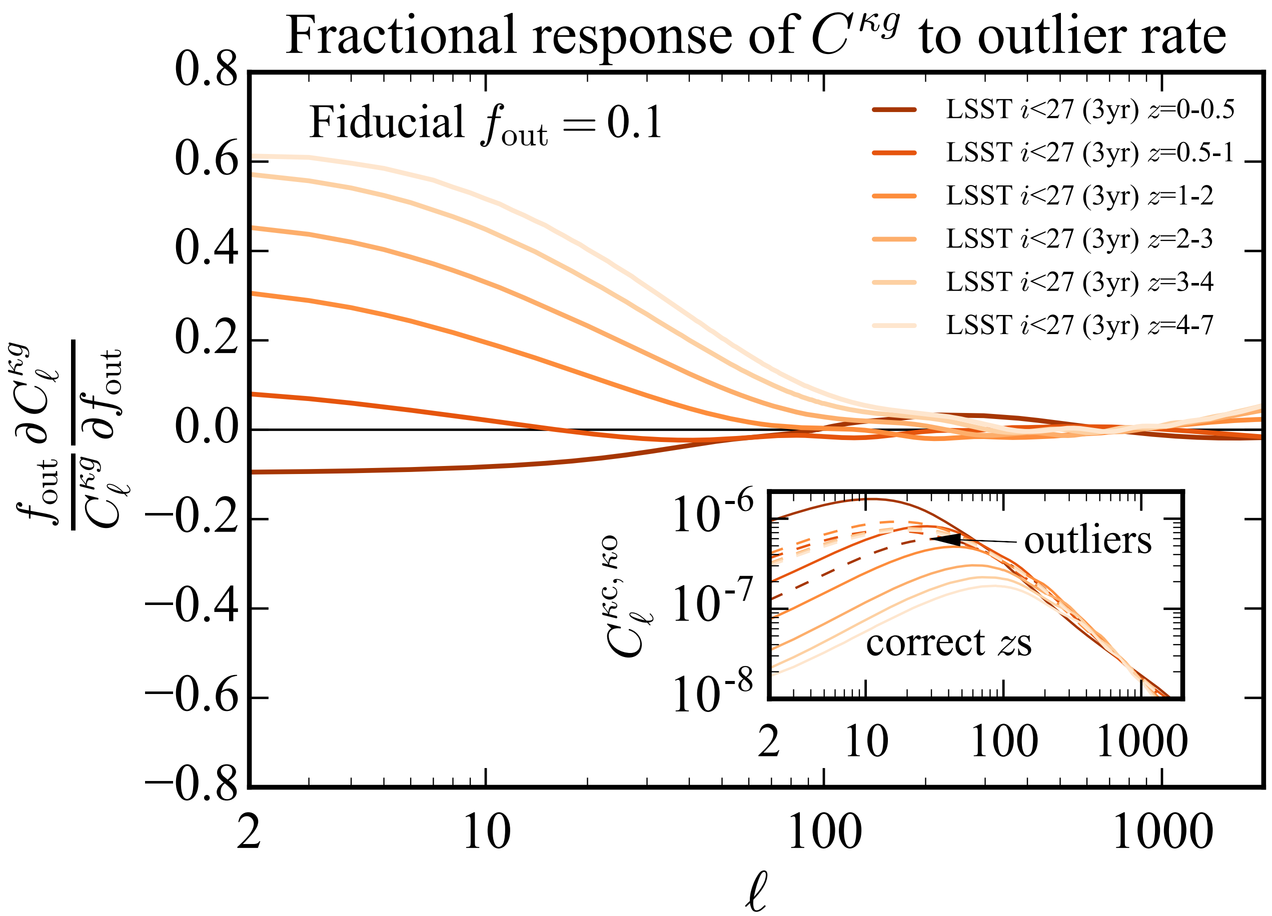
Informal, usually 20-30 minutes talks

If you are around and would like to give a talk please email me

Catastrophic redshift errors

Model catastrophic outliers as

$$\left.\frac{\mathrm{d}n}{\mathrm{d}z}\right|_{i,\mathrm{obs}}(z) = \begin{cases} (1-f^i_{\mathrm{out}})\,\frac{\mathrm{d}n}{\mathrm{d}z}(z) & \text{if } z \in i\text{th bin}, \\ \frac{n_i}{n_{\mathrm{tot}}-n_i}\, f^i_{\mathrm{out}}\,\frac{\mathrm{d}n}{\mathrm{d}z}(z) & \text{else}, \end{cases} \tag{23}$$



MS & Seljak 1710.09465