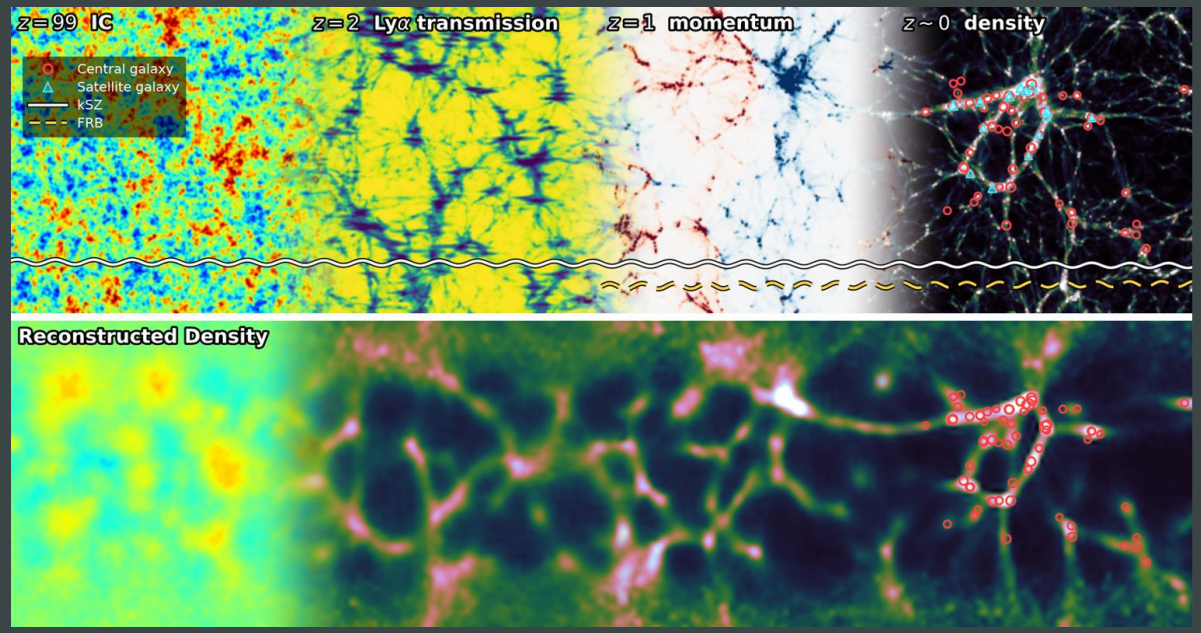
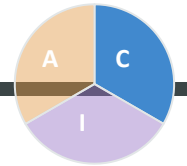
Tracking Baryon Redistribution Beyond Galactic Halos Across Cosmic Time

with Field-Level Environments

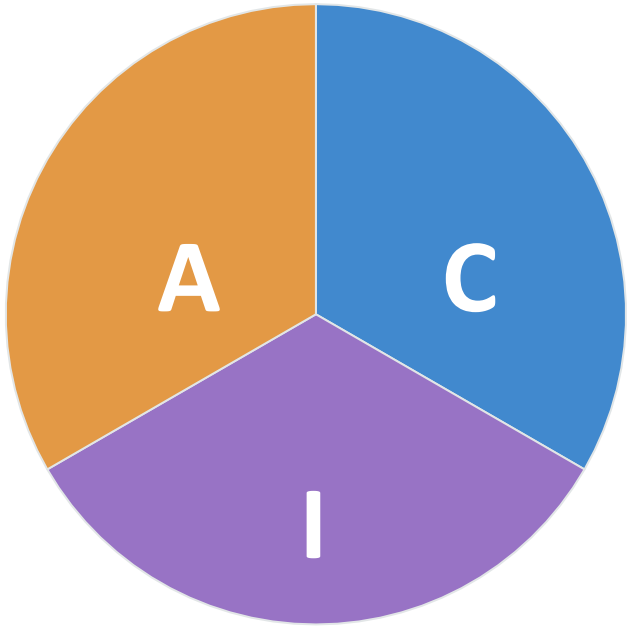


Chenze Dong · Kavli IPMU

In collaboration with Metin Ata, Adrian E. Bayer, Renyue Cen, Weiguang Cui, Romeel Davé, Florian Dedieu, Daniela Galárraga-Espinosa, Benjamin Horowitz, Khee-Gan Lee, Akira Mizuta, Rieko Momose, Kentaro Nagamine, Boon Kiat Oh, Yuri Oku, Daniele Sorini, Abednego Wiliardy



What determines the baryon distribution?



C **Cosmology**

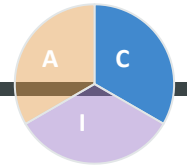
The background model and the statistics of the initial fluctuations, set by a few parameters

A **Astrophysics**

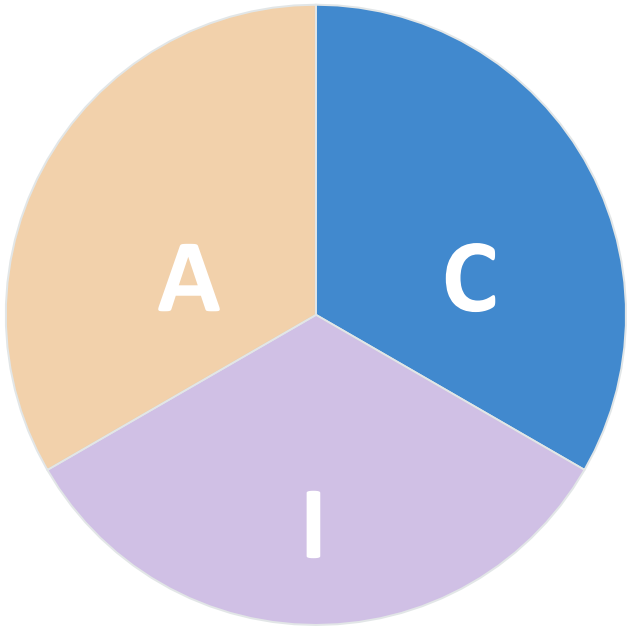
How gas and galaxies evolve: cooling, star formation and feedback; more parameters (and vary among simulation code)

I **Initial realization**

The specific high-dimensional field of amplitudes and phases that decides which structures form where



Three ingredients: Cosmology

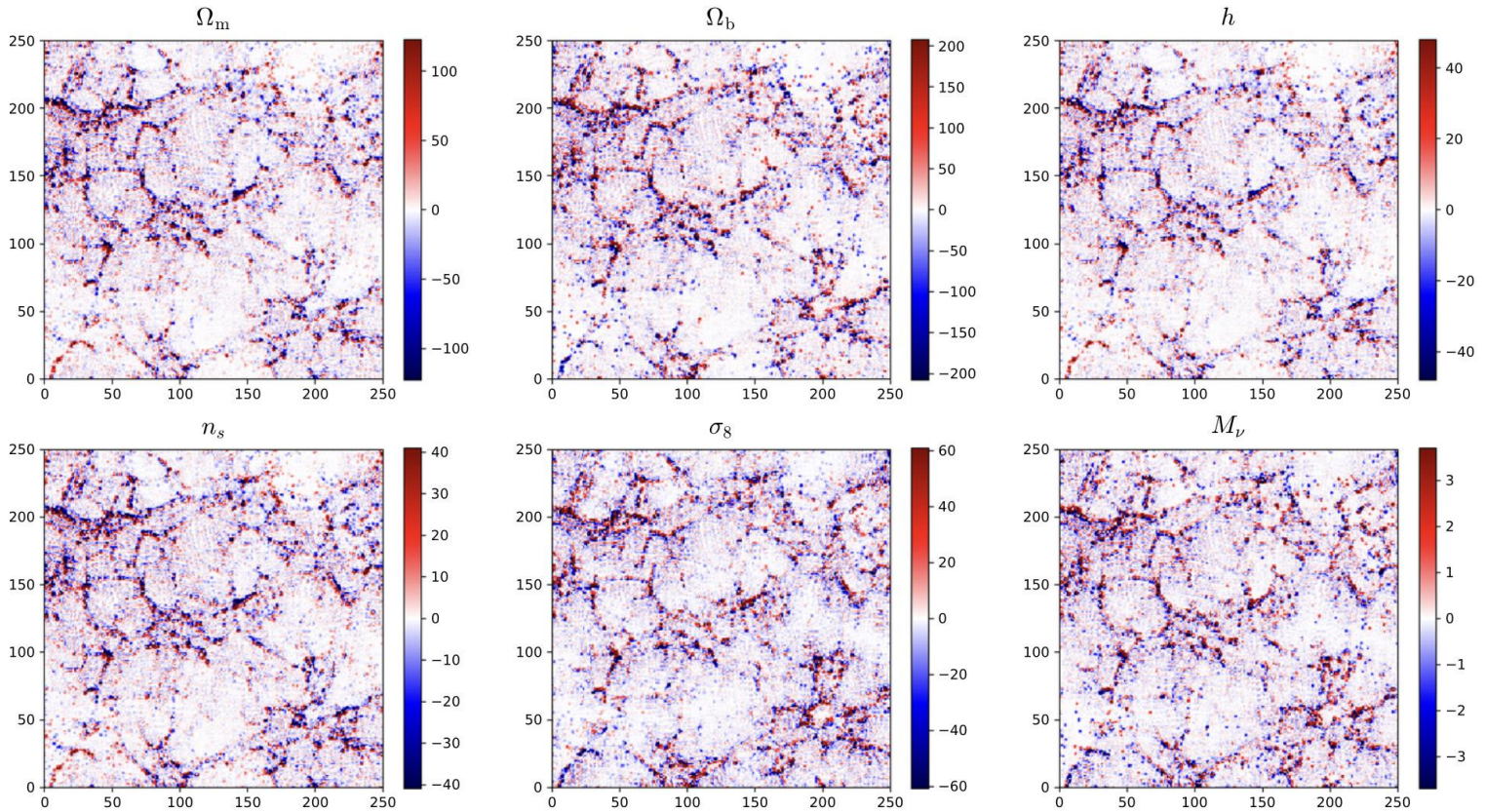


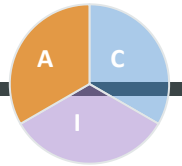
C Cosmology

I Initial realization

A Astrophysics

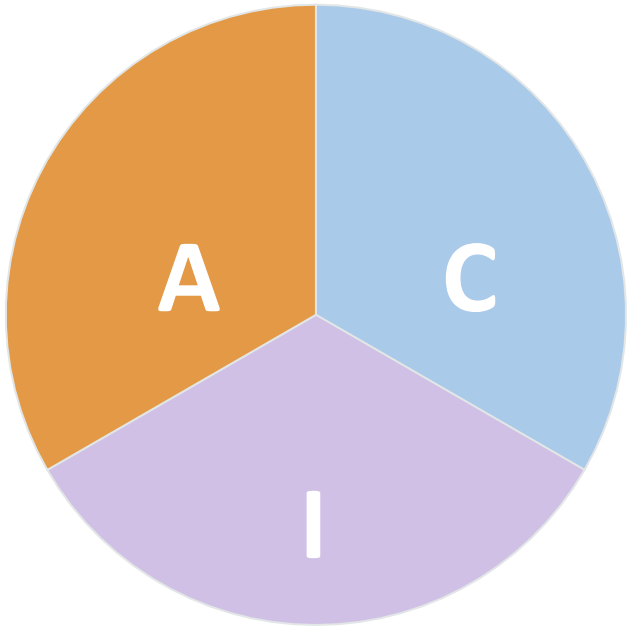
Villaescusa-Navarro+20





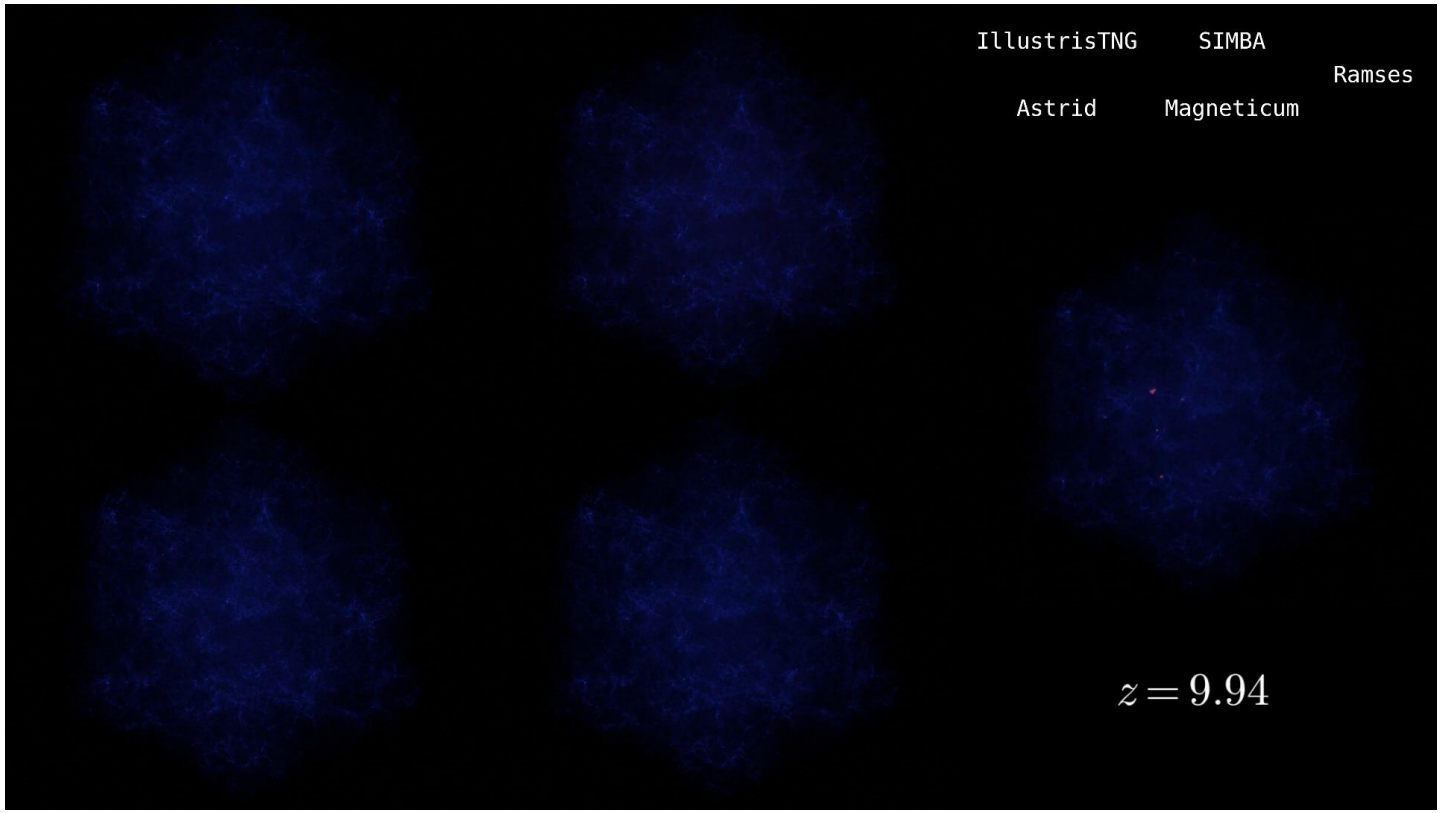
Three ingredients: Astrophysics

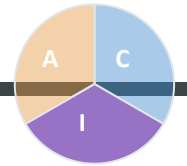
Different feedback models change the distribution and state of the gas



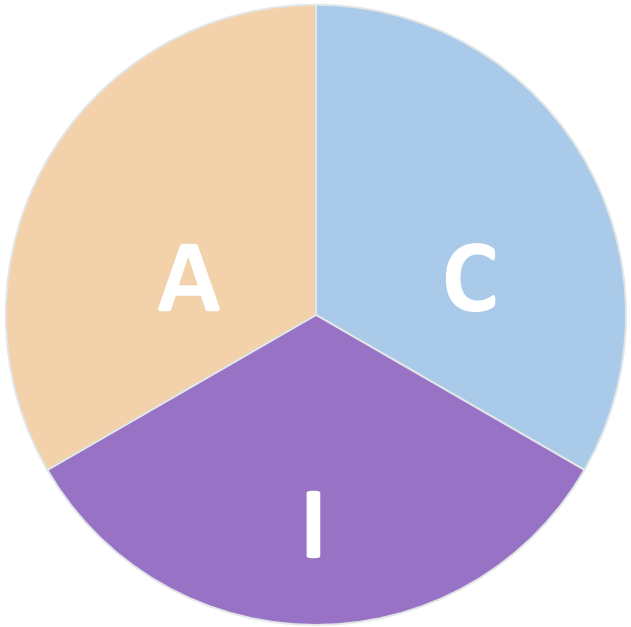
C Cosmology
I Initial realization
A Astrophysics

Villaescusa-Navarro+21





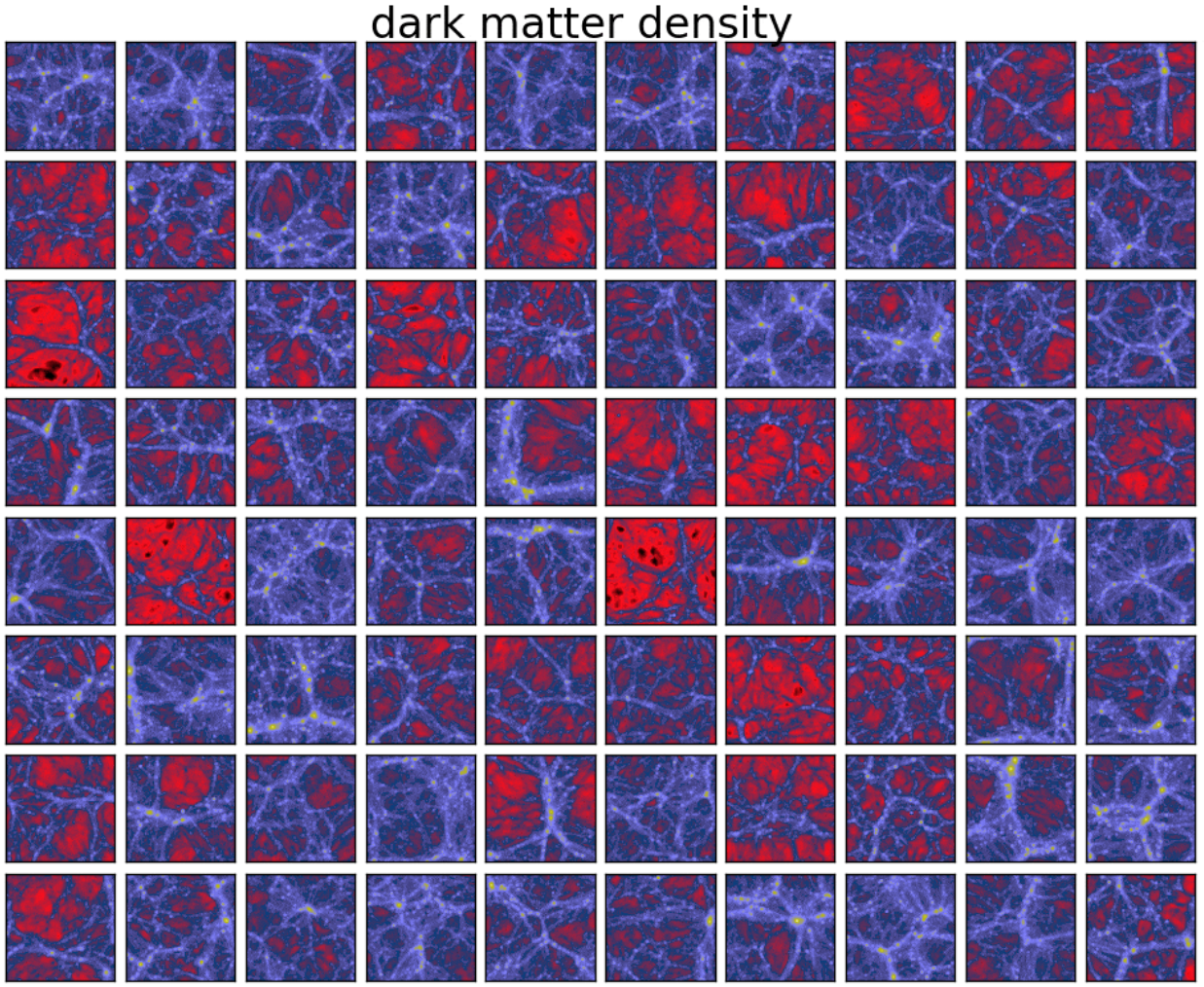
Three ingredients: Initial realization



C Cosmology

I Initial realization

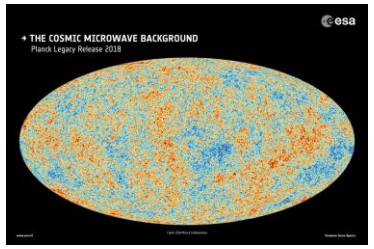
A Astrophysics



Cosmology: Model Initial realization: Realization Astrophysics: Model

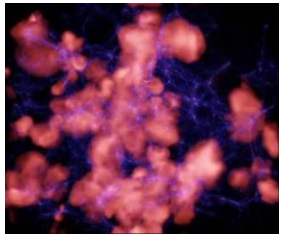
Simulation vs Observation

Cosmology measurement



fit cosmological parameters

Astrophysics calibration

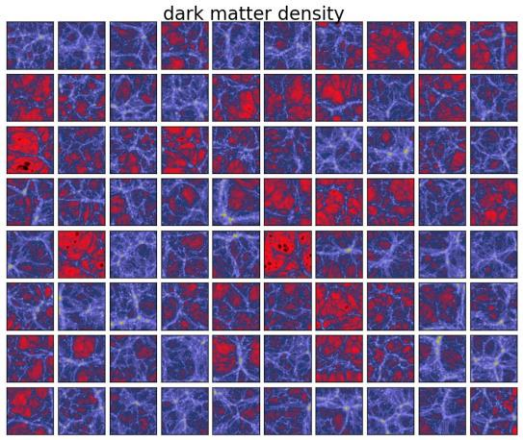


impl. of feedback & tuning

Planck Collaboration+20; Villaescusa-Navarro+21

→

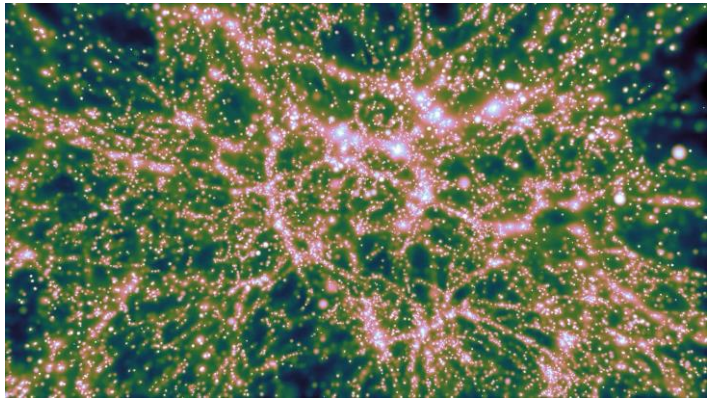
Draw random ICs



Sample an initial realization
at fixed cosmology.

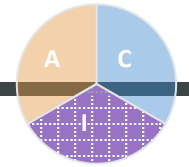
→

Evolve the simulation



Compare with observations

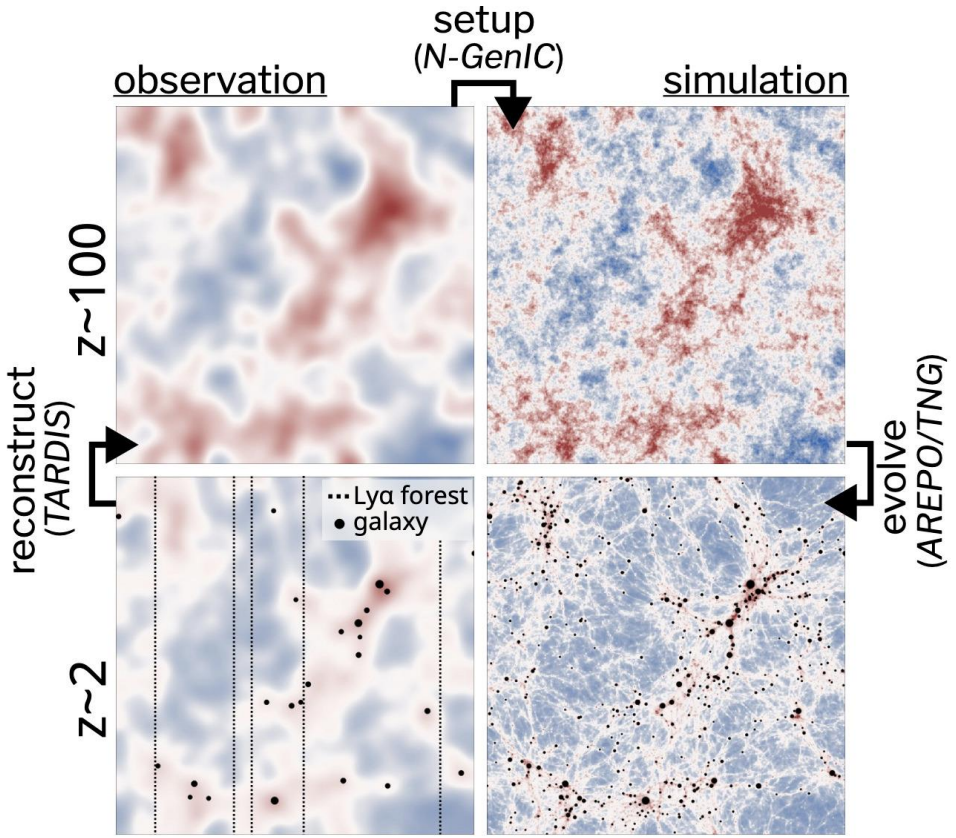
Cosmology: Fixed Initial realization: Sampled Astrophysics: Fixed



Constrained Simulations vs Observations

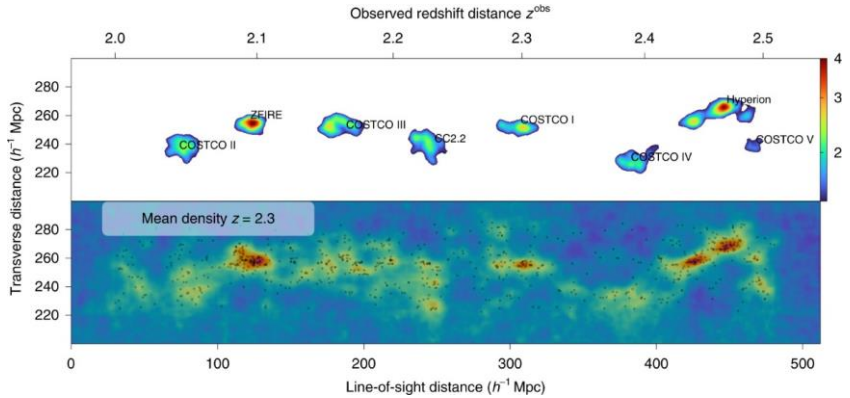
What if we want to simulate *our own Universe?*

Use observations to constrain the initial realization.



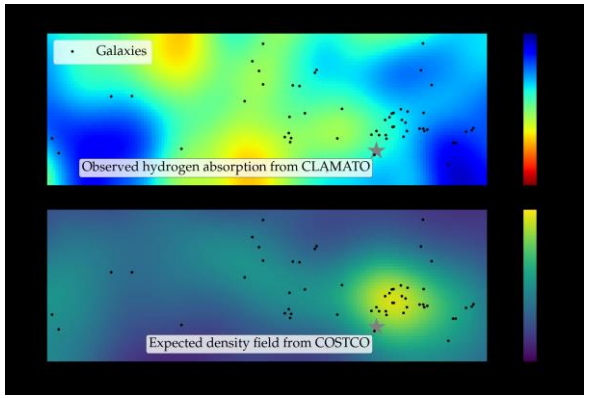
Constrained Simulations vs Baryon Tracers

COSTCO

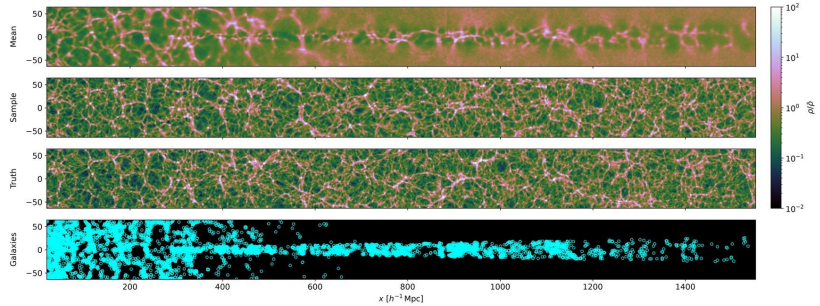


→

CLAMATO · Lyα forest

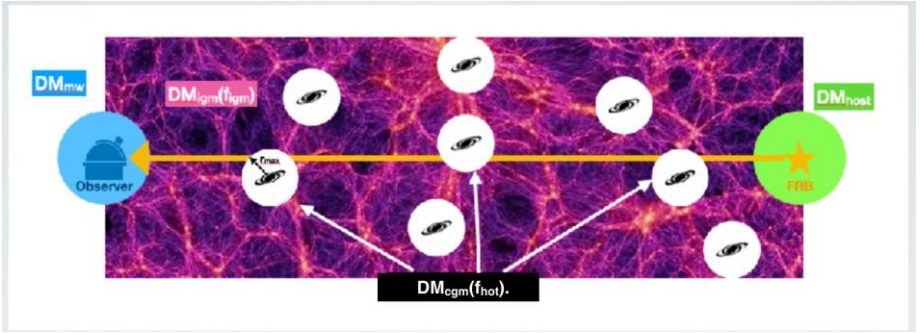


FLIMFLAM · MaiGO



→

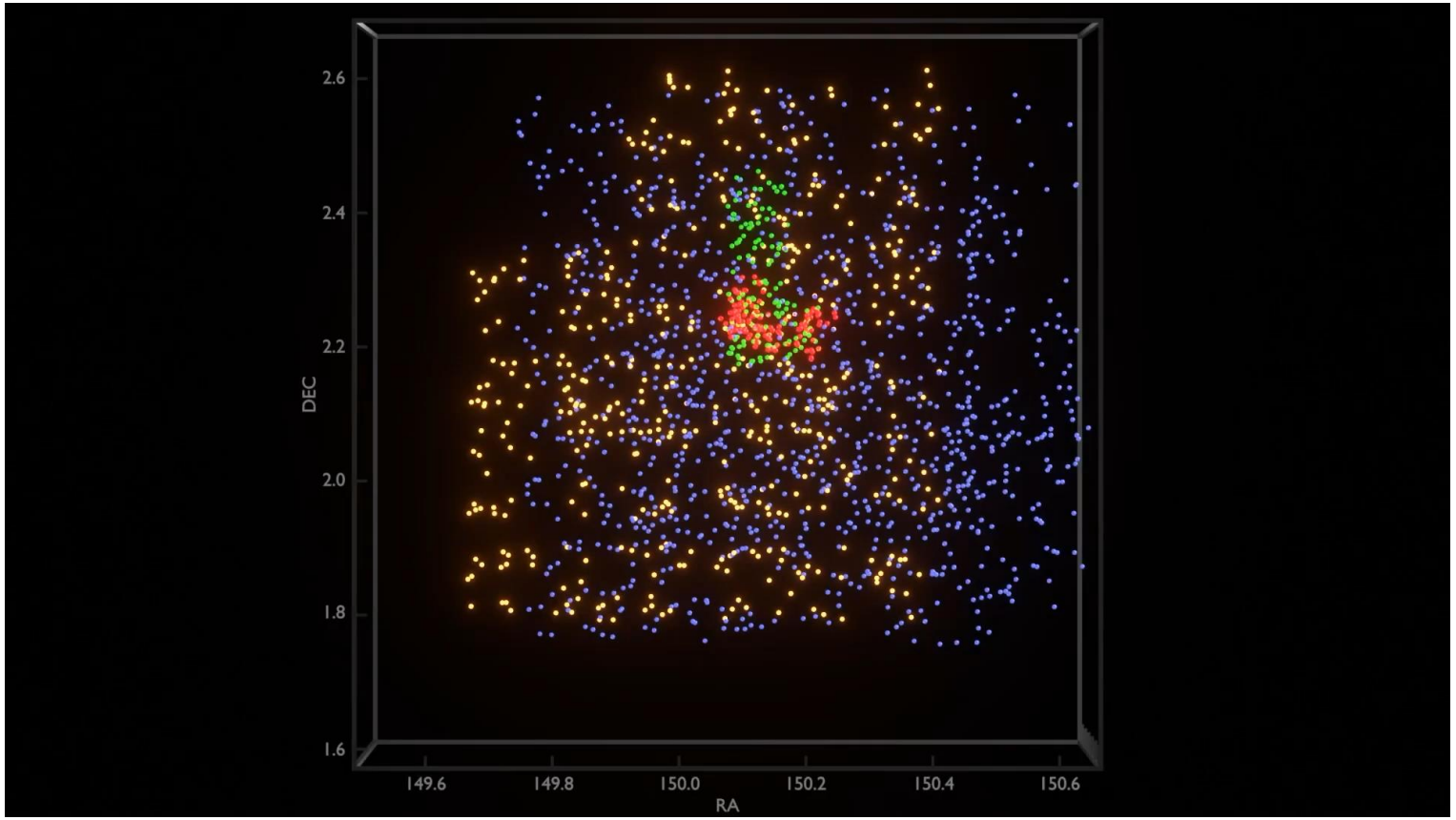
FRB · dispersion measure



Ata+22; Dong+23; Lee+22; Khrykin+24

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

COSTCO: reconstructing the COSMOS field



COSTCO = COnstrained Simulations of The COsmos field

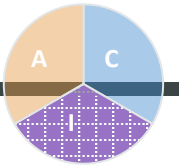
Galaxies constrain ICs

Sample initial realizations compatible
with the COSMOS galaxy distribution.

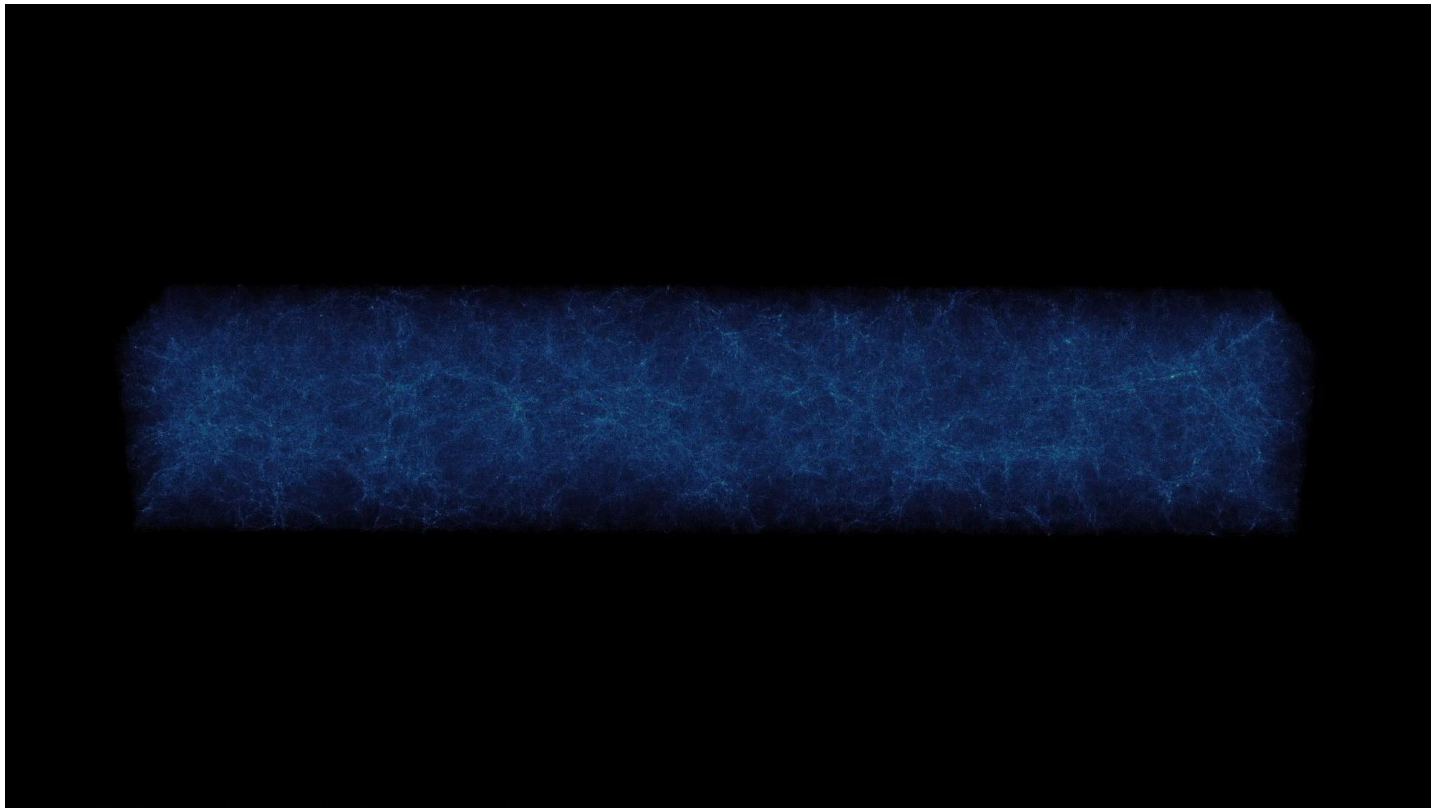
Constrained evolution

Reconstruct the field at $z \simeq 2.3$ and follow its structures to $z = 0$.

Ata+22



Discovering protoclusters with COSTCO



Identify overdensities and follow their descendants across constrained realizations.

Ata+22

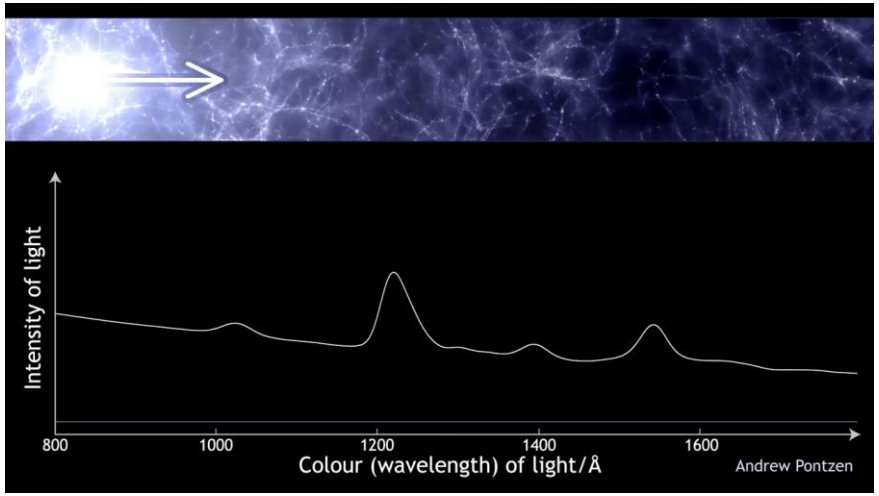
Table 1 | COSMOS field protoclusters in the redshift range of $2 \leq z^{obs} \leq 2.52$

Name	RA (deg)	Dec (deg)	z^{obs}	Final Mass ($h^{-1}M_{\odot}$)
COSTCO J100026.4+020940 (I)	150.110	2.161	2.298	$(4.6 \pm 2.2) \times 10^{14}$
COSTCO J095924.0+021435 (II)	149.871	2.229	2.047	$(6.1 \pm 2.5) \times 10^{14}$
COSTCO J100031.0+021630 (III)	150.129	2.275	2.160	$(5.3 \pm 2.6) \times 10^{14}$
COSTCO J095849.4+020126 (IV)	149.706	2.024	2.391	$(6.6 \pm 2.3) \times 10^{14}$
COSTCO J095945.1+020528 (V)	149.938	2.091	2.283	$(4.3 \pm 2.4) \times 10^{14}$

Confirm protoclusters in observation
&
New protoclusters found!

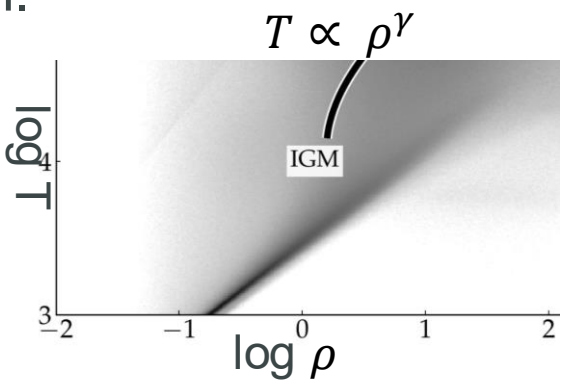
Cosmology: Assumed Initial realization: Constrained Astrophysics: Assumed

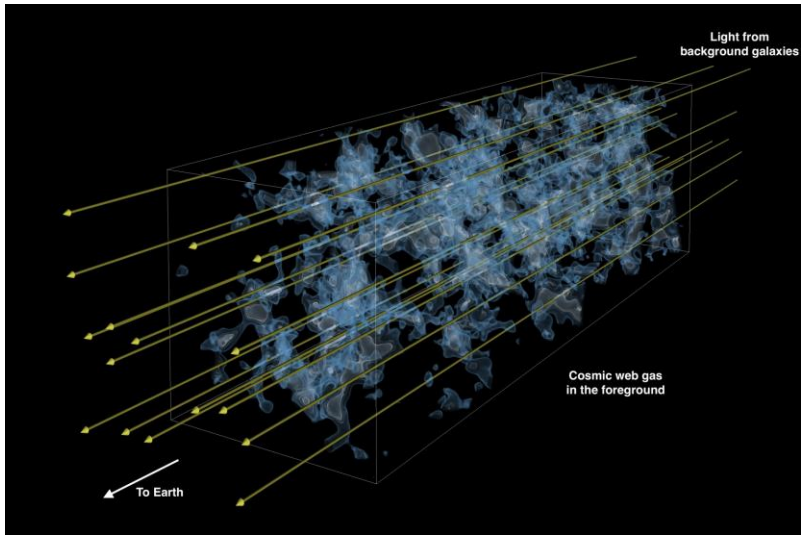
CLAMATO: the Lyα forest as a 3D tracer



Background spectra sample absorption by foreground neutral hydrogen.

$$\tau_{Ly\alpha} \propto \frac{T^{-0.7}(1+\delta_m)^2}{\Gamma_{UV}}$$





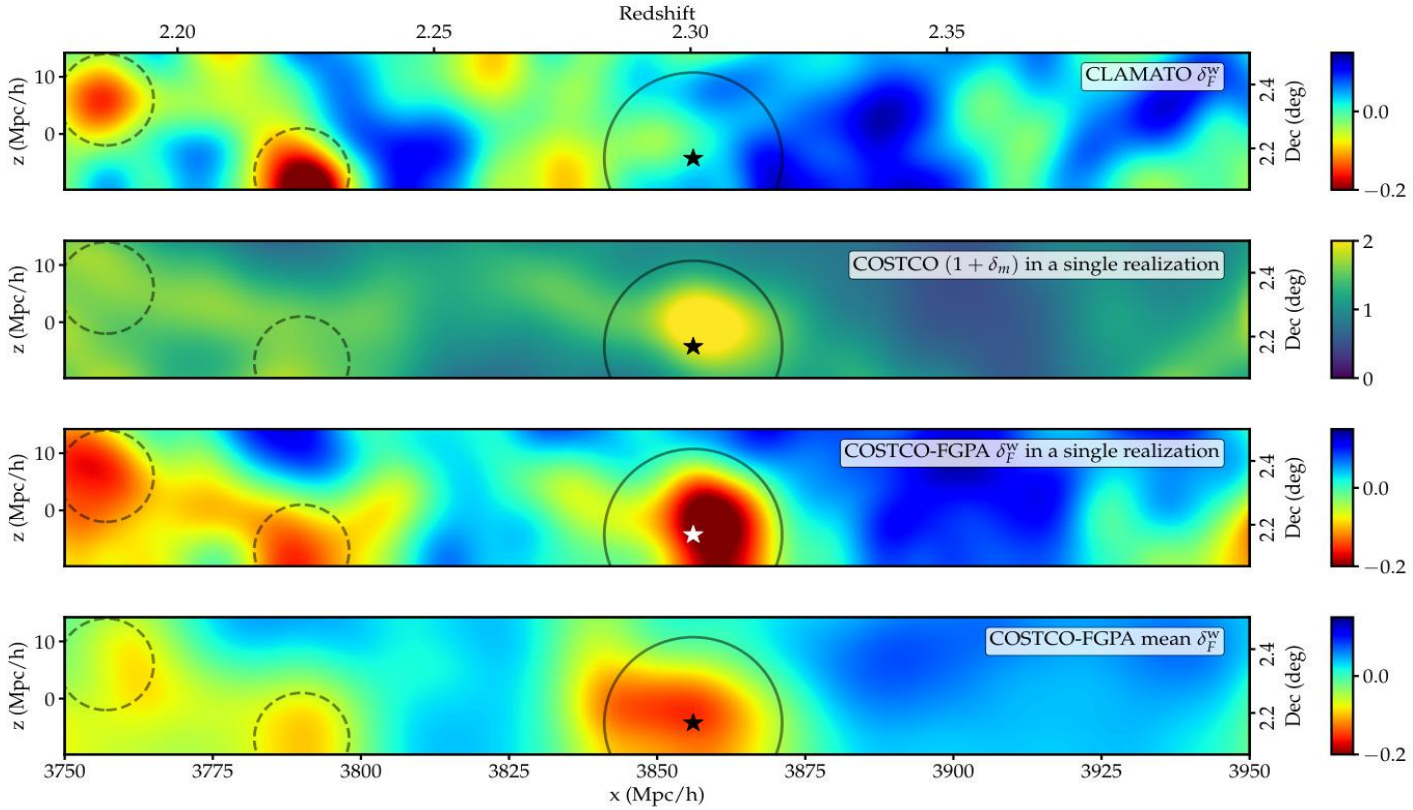
Combine neighboring sightlines to map the IGM in COSMOS.

$$\tau_{Ly\alpha} = -\log F \propto (1+\delta_m)^{2-0.7\gamma}$$

Lee+18; Dong+23

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

COSTCO-I: dense in matter, weak in Lyα absorption

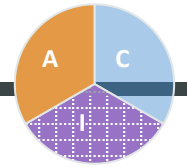


CLAMATO: weak absorption

COSTCO: a dense environment

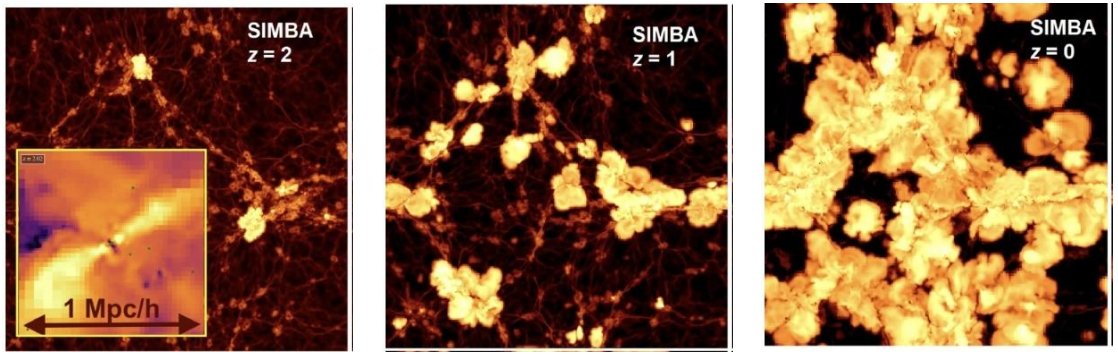
Mock Lyα maps predict stronger absorption after matching the observational sampling and reconstruction.

Dong+23

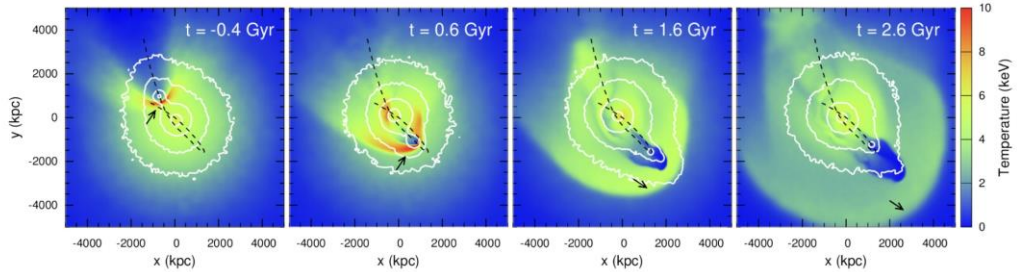


What leads to the missing Lyα protocluster?

Jet-heated bubbles



Gravitational shocks



Dong+23; Zhang+19; Christiansen+20

1 Density is independently constrained

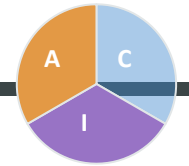
COSTCO predicts an overdensity,
but CLAMATO shows weak absorption.

2 Higher temperature is a plausible cause

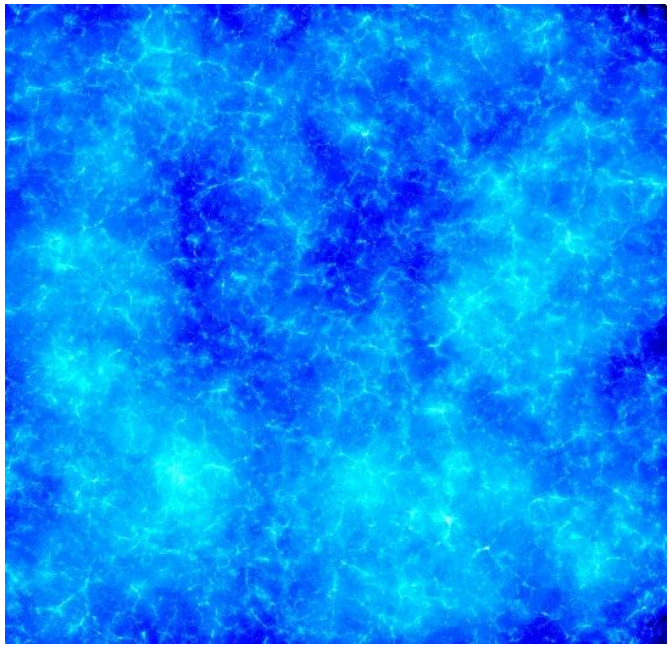
At fixed density and photoionization rate,
hotter gas has less neutral hydrogen.

3 Test gravitational shocks and AGN feedback

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target



Testing heating with The Three Hundred



Related initial environments, different hydrodynamic suites.

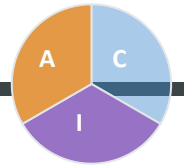
Cui+18; Dong+24

MDPL2 parent simulation

1 h^{-1} Gpc dark-matter-only box
Select regions around massive z = 0 halos.

Trace particles back to the initial conditions.
Split into dark matter and gas, then resimulate.

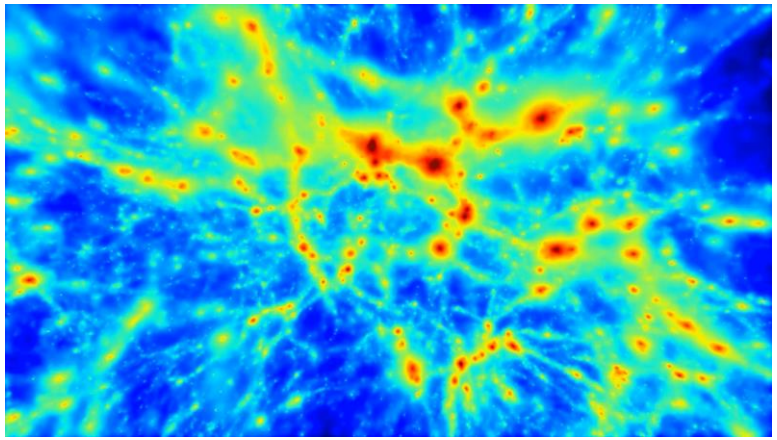
Compare protocluster gas at z = 2.5



Three feedback models, different heating patterns

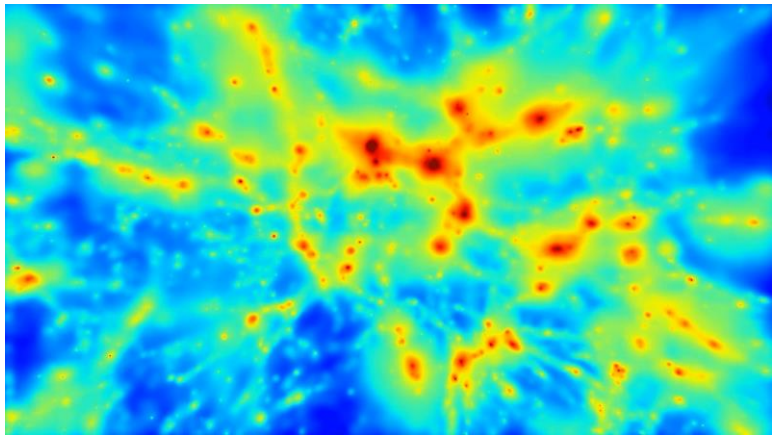
300-GadgetMUSIC

Stellar feedback; no AGN



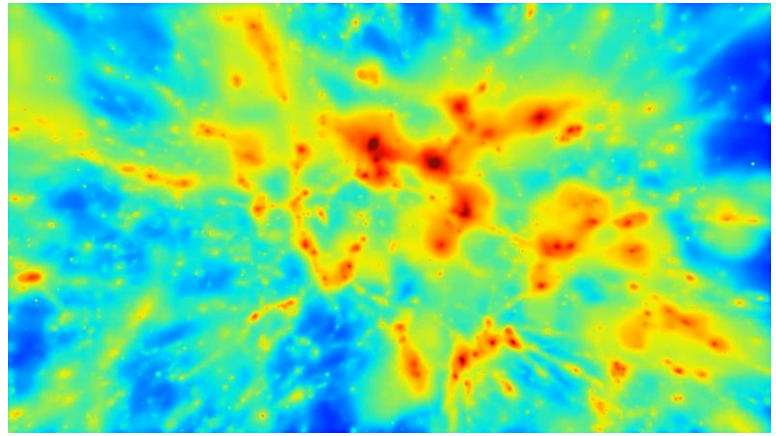
300-GadgetX

Thermal AGN feedback



300-SIMBA

Collimated AGN jets

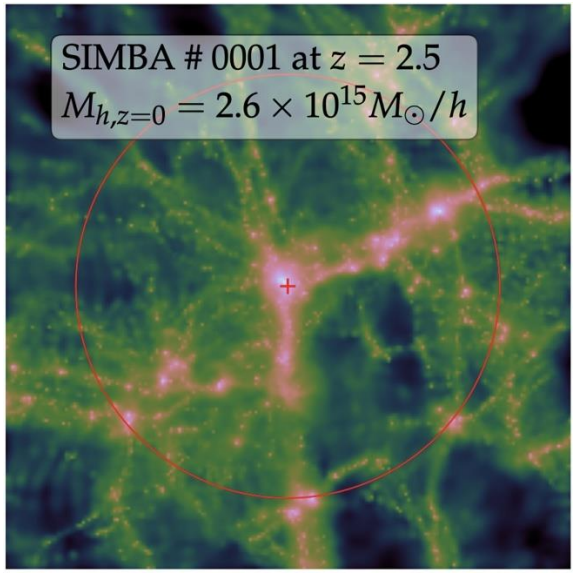


Feedback changes the temperature field in the same protocluster environment.

Dong+24

Matching simulations to COSTCO and CLAMATO

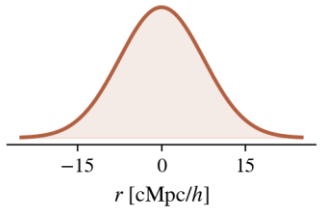
Matter overdensity



$\langle \delta_m \rangle_G$

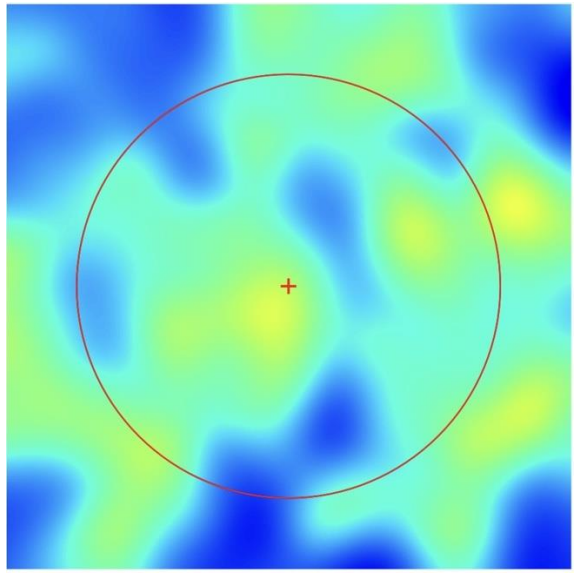
Compare with COSTCO

Same Gaussian weighting



7.5 cMpc/h

Lyα transmission contrast



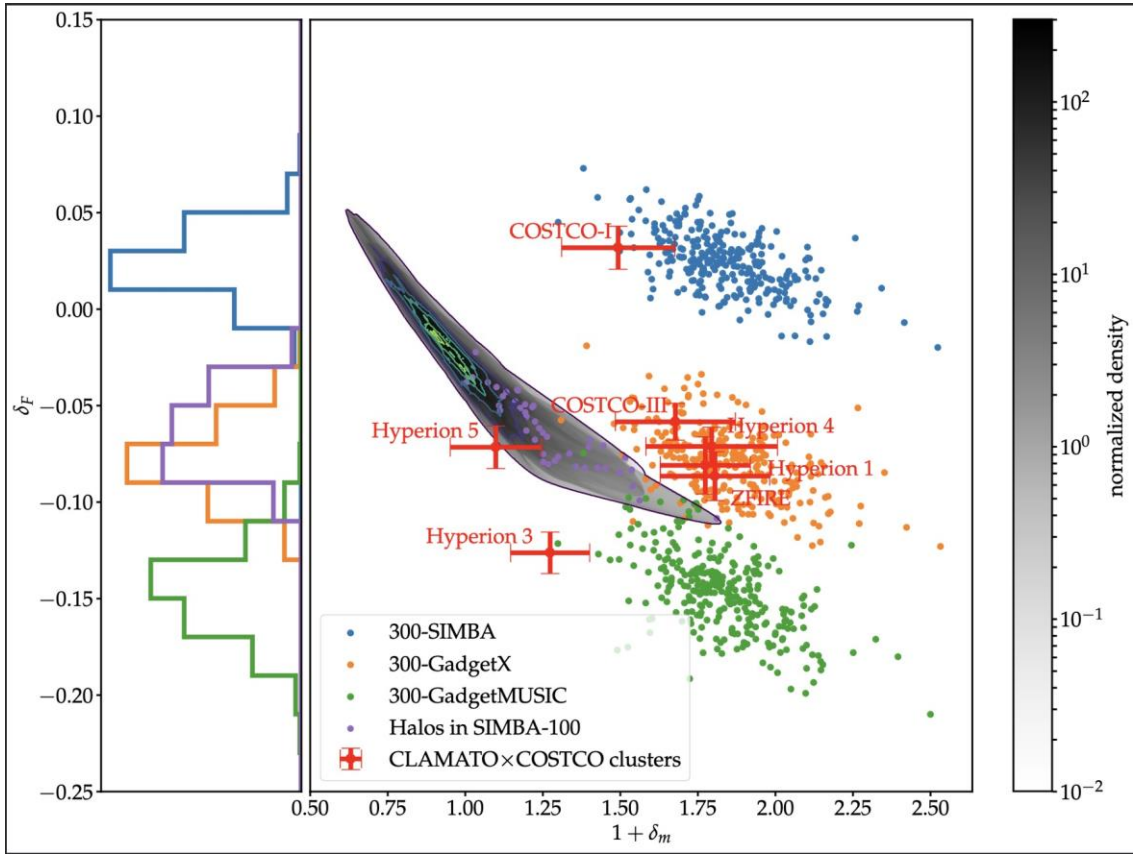
$\langle \delta_F \rangle_G$

Compare with CLAMATO

Dong+24

Cosmology: Fixed Initial realization: Matched Astrophysics: Compared

Feedback models versus observations



Dong+24

1 Lyα responds to feedback

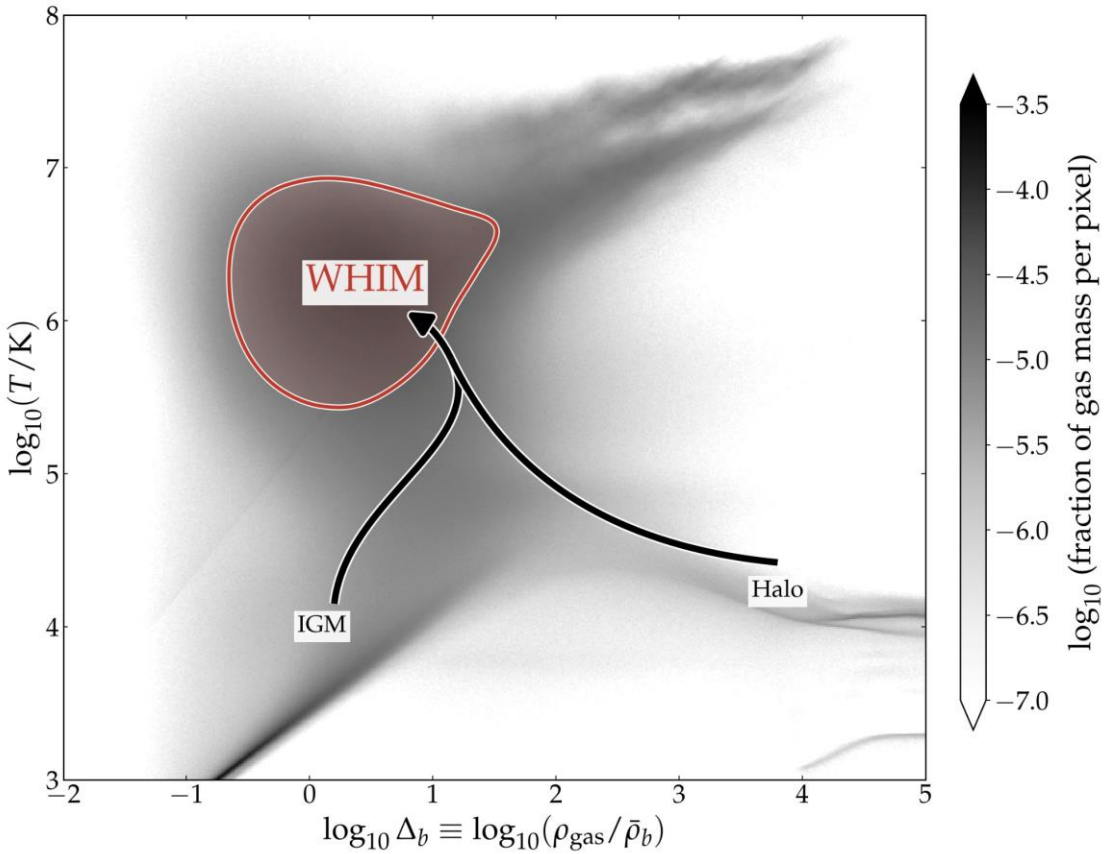
The suites occupy different density–transmission relations.

2 Several favor thermal AGN

Several protoclusters agree with thermal AGN heating.

3 COSTCO-I favors jets

COSTCO-I is close to SIMBA-like simulation.



Dong+24; Lee+22

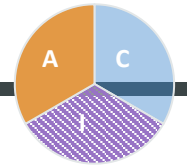
Gas heating & baryon redistribution

COSTCO + Lyα

Why is dense gas so transparent?
– (kinetic) feedback

FLIMFLAM + FRBs

Where are the electrons along a known foreground?
- In or outside halo?

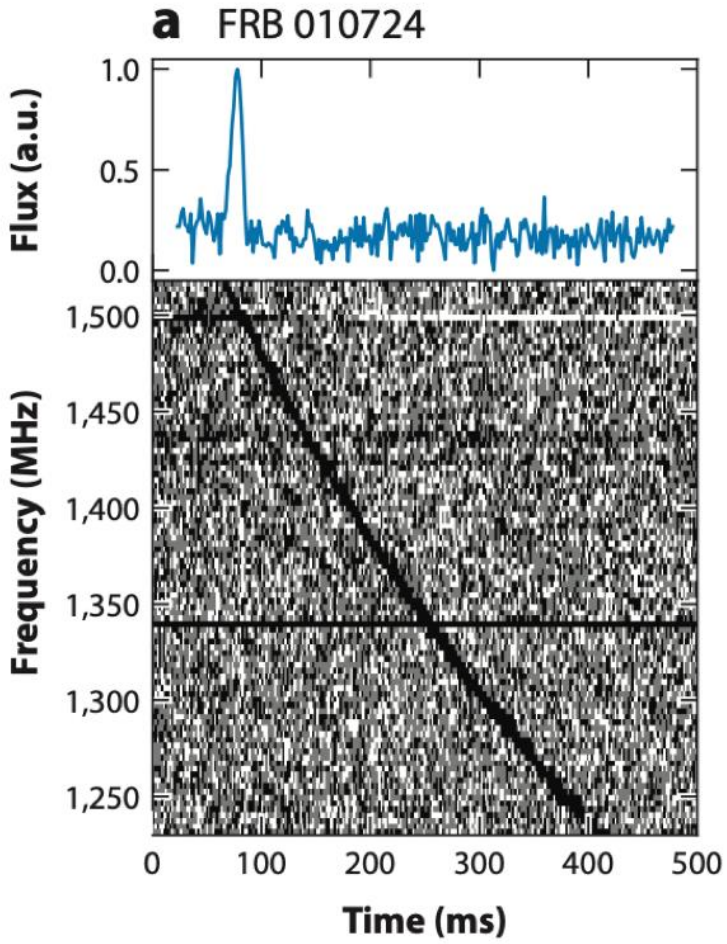


Fast Radio Burst: probe electrons along LOS

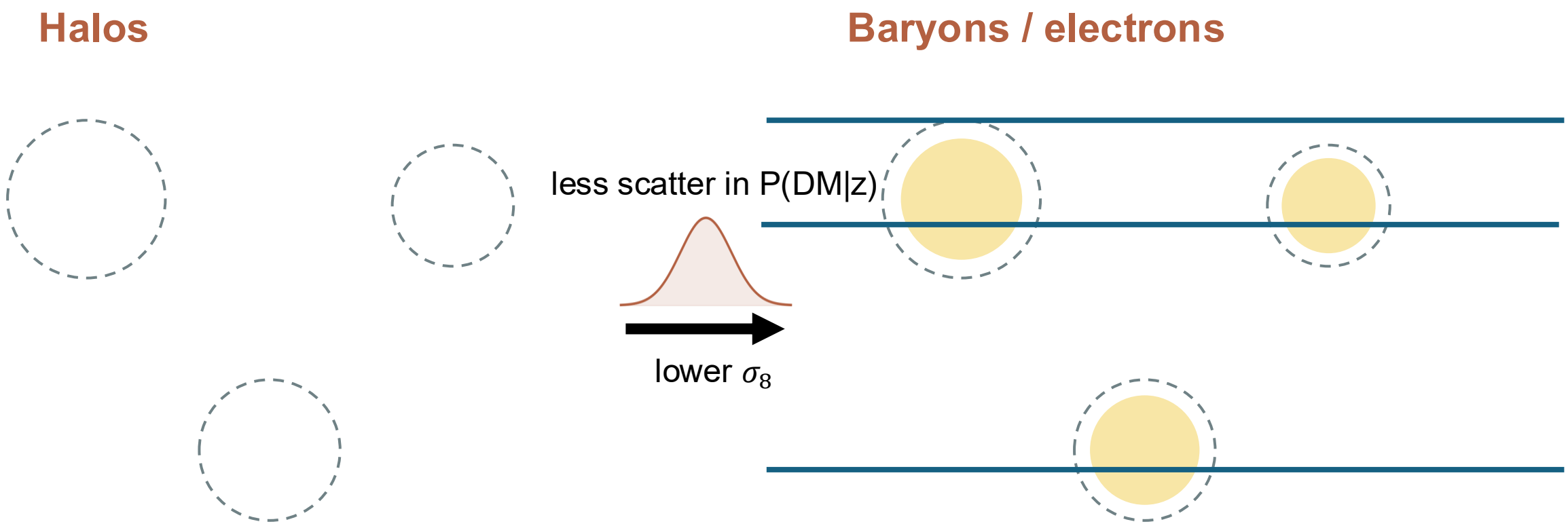
Millisecond-duration radio pulse first identified in 2007

To-date ~10^4 FRBs have been detected; ~10^2 have been localized to specific host galaxies.

Unknown progenitors, but still useful to probe baryons via DM: frequency-dependent time delay



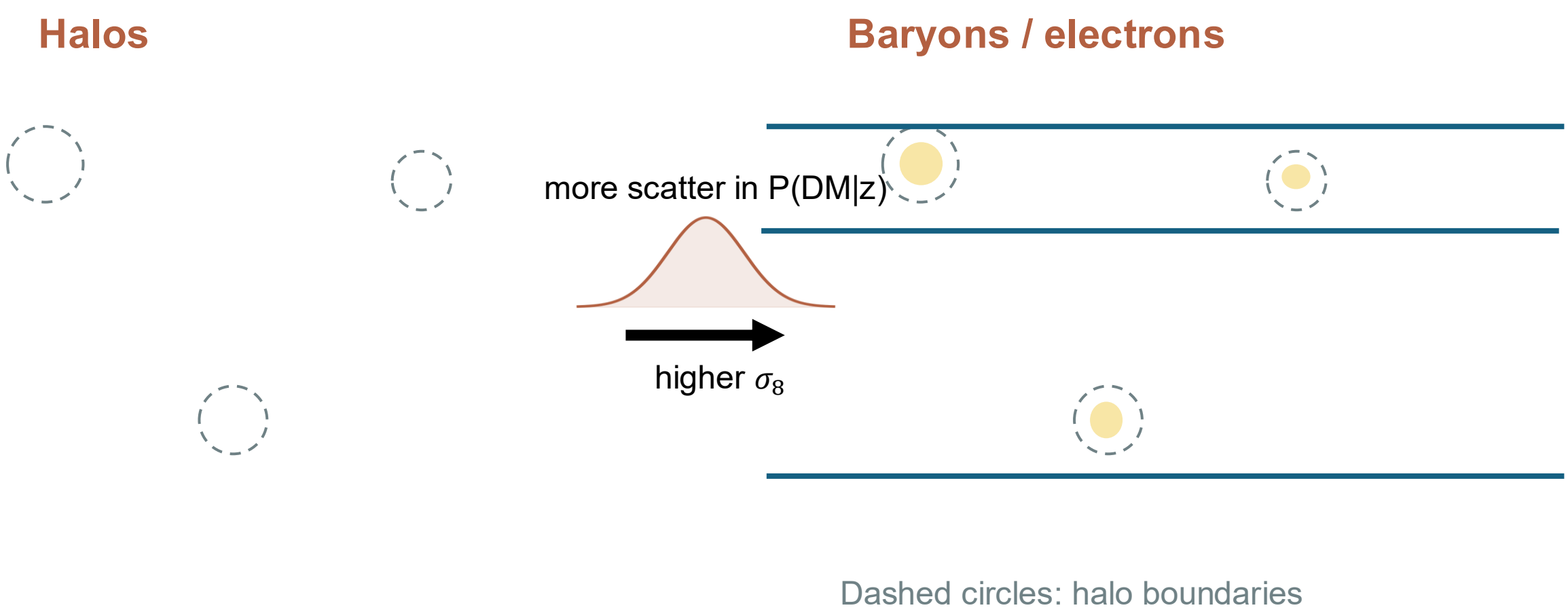
FRB as a probe of cosmology via P(DM|z)



Lee+22; Khrykin+24

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

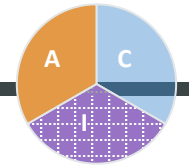
FRB as a probe of cosmology via P(DM|z)



Dashed circles: halo boundaries

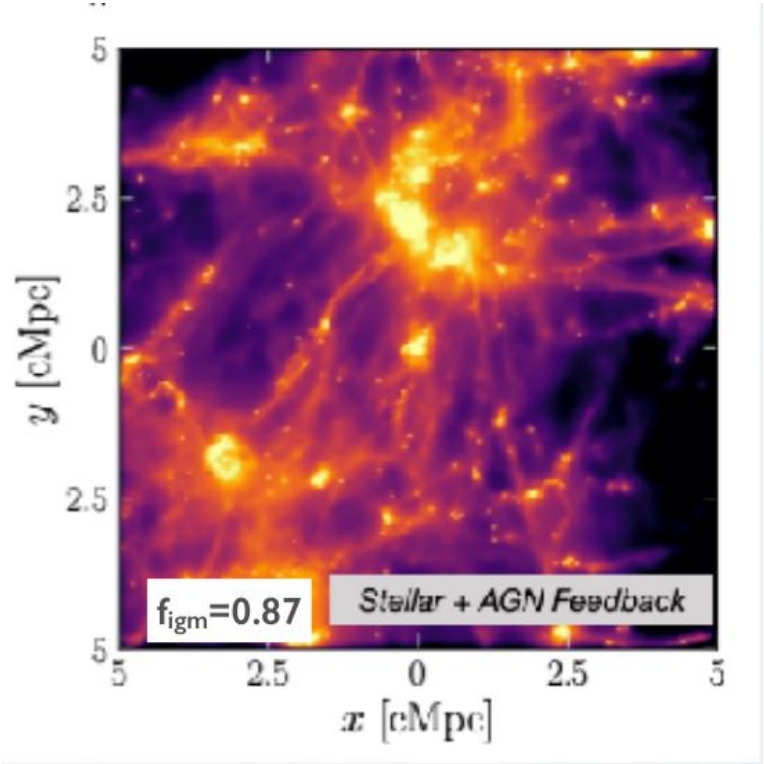
Lee+22; Khrykin+24

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

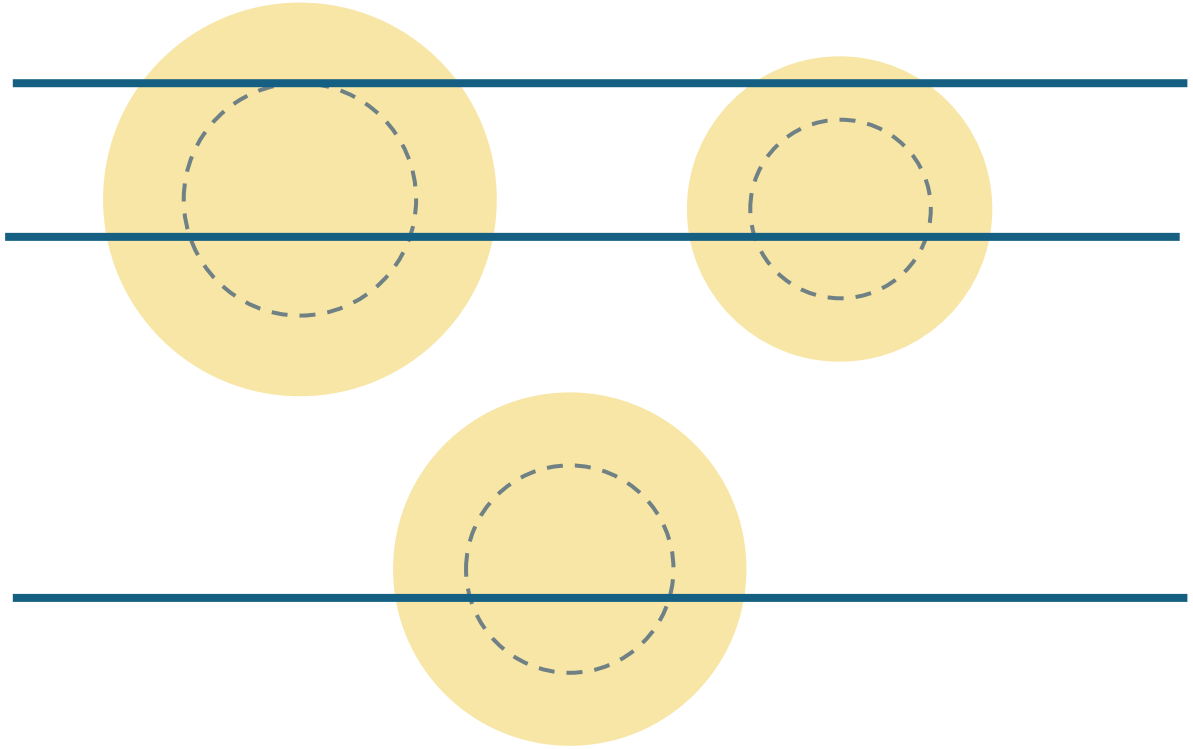


Feedback (also) redistributes gas beyond halos

With feedback



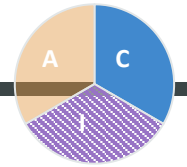
Baryons / electrons



Dashed circles: halo boundaries

Lee+22; Khrykin+24

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

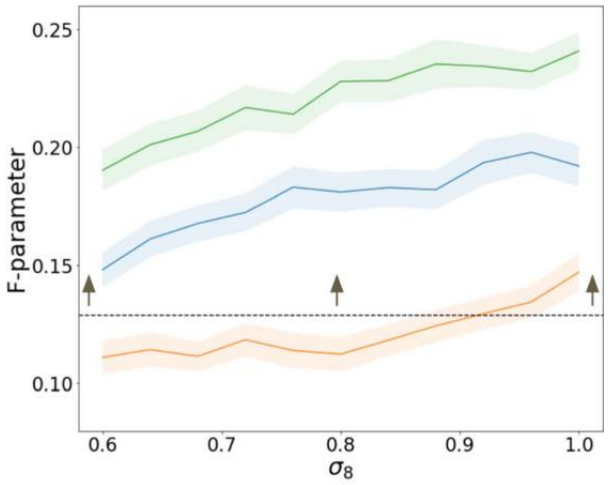


Cosmology and feedback in the DM distribution

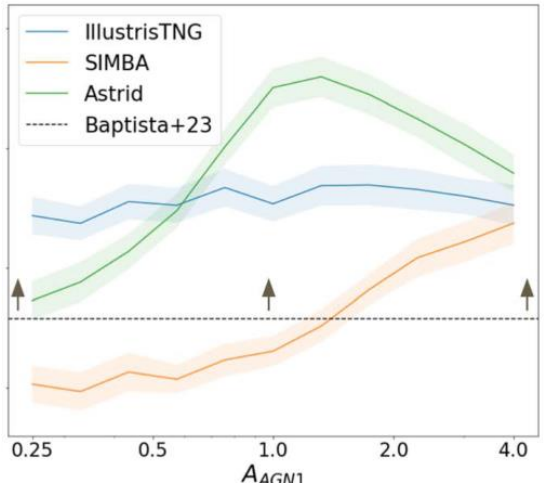
$$p(\mathrm{DM}|z, \theta_{astro}, \theta_{cosmo})$$

$$F = \sigma(\mathrm{DM}/\langle \mathrm{DM} \rangle)\sqrt{z}$$

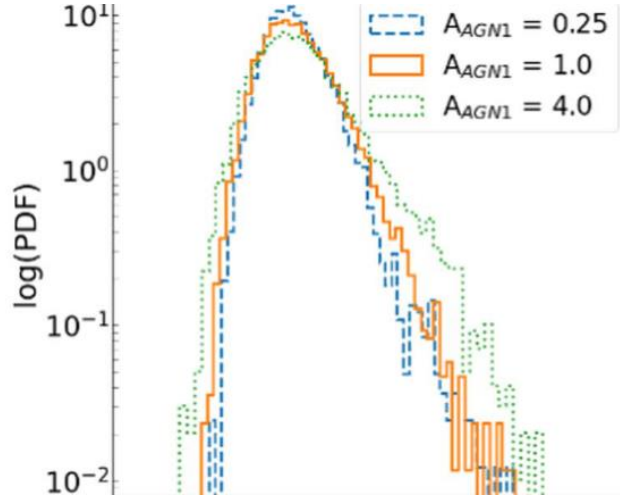
Clustering amplitude



AGN feedback parameter



DM distribution



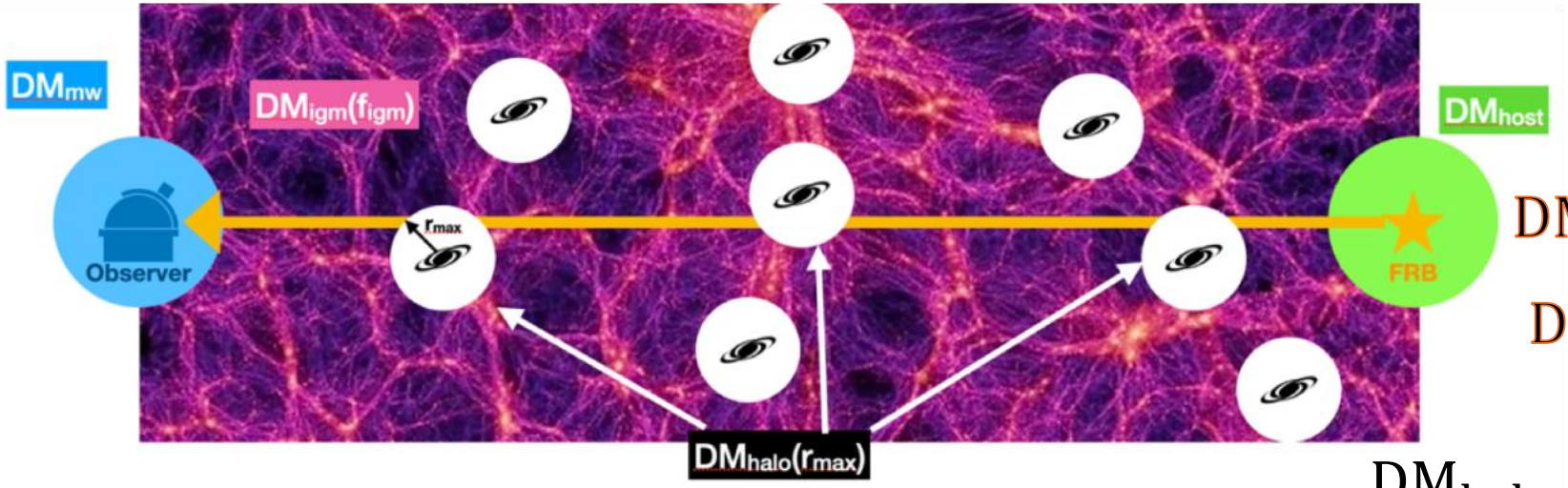
Population inference requires ~10^3 localized FRBs.

Medlock+24

Cosmology: Target Initial realization: Marginalized Astrophysics: Uncertain

FLIMFLAM: stronger constraints from each FRB (on feedback)

FRB Line-of-sight Ionization Measurement From Lightcone AAOmega Mapping



$$\mathrm{DM_{cosmic}} = \mathrm{DM_{IGM}} + \mathrm{DM_{halo}}$$

$$\mathrm{DM_{IGM}} \propto \boldsymbol{f_{IGM}} \int (1+\delta_m) dl$$

$$\mathrm{DM_{halo}} \propto \boldsymbol{f_{CGM}}\ (or\ \boldsymbol{f_{gas}}) \sum DM_{int.halo,i}$$

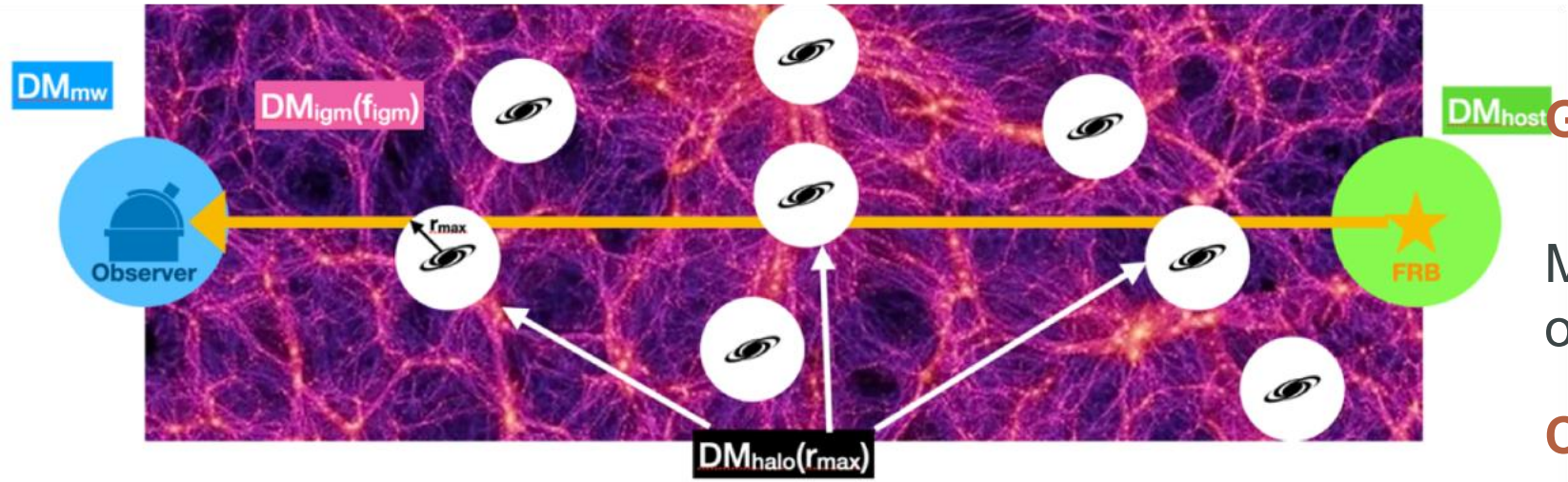
$$\mathrm{DM_{FRB}} = \mathrm{DM_{MW}} + \mathrm{DM_{cosmic}} + \frac{\mathrm{DM_{host}}}{1+z_{\mathrm{FRB}}}$$

Lee+22; Khrykin+24

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

FLIMFLAM: stronger constraints from each FRB

FRB Line-of-sight Ionization Measurement From Lightcone AAOmega Mapping



Galaxy spectroscopy

Measure the structures on the actual sightline.

Conditional DM predictions

$$p(\mathrm{DM}|z, \theta_{astro}, \mathrm{Foreground})$$

$$\mathrm{DM_{FRB}} = \mathrm{DM_{MW}} + \mathrm{DM_{cosmic}} + \frac{\mathrm{DM_{host}}}{1+z_{\mathrm{FRB}}}$$

More constraining power from each FRB.

Lee+22; Khrykin+24

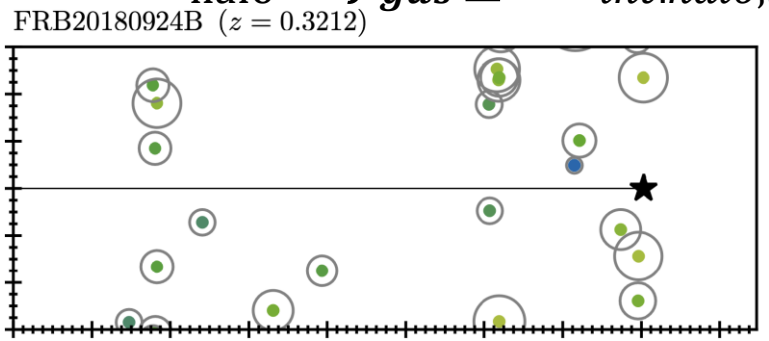
FLIMFLAM DR1: data model

Deep IFU around FRBs

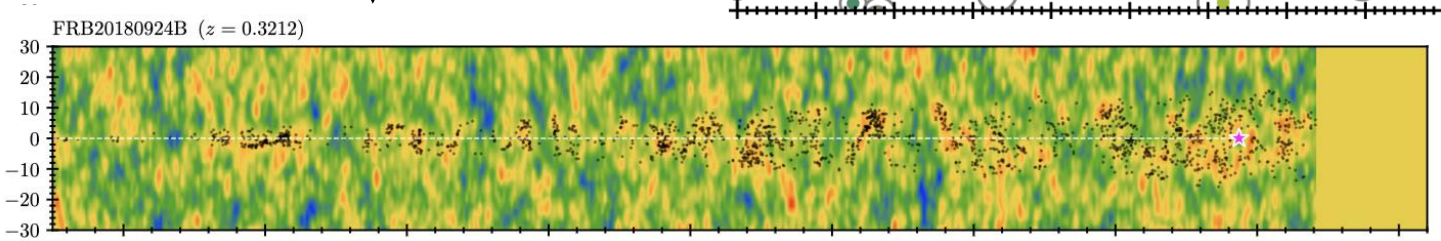
AAOmega foreground observation

FoF group finding

$$\mathrm{DM_{halo}} \propto f_{gas} \sum DM_{int.halo,i}$$



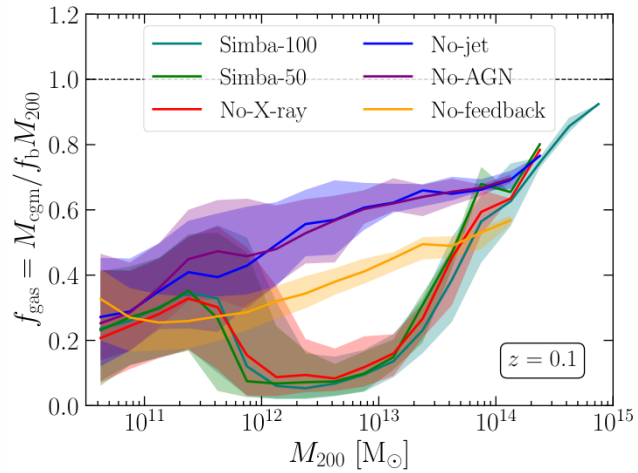
ARGO: HMC sampling



$$DM_{IGM} \propto f_{IGM} \int (1 + \delta_m) dl$$

Huang+25; Khrykin+24

Theoretical expectation

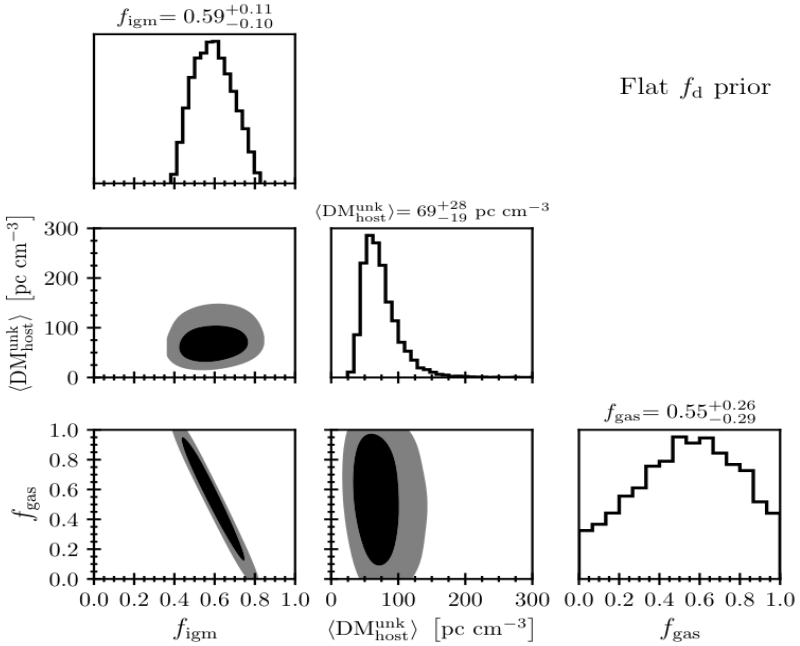


SIMBA predicts a feedback-dependent halo gas fraction f_{gas}.

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

FLIMFLAM DR1: baryon partition from 8 FRBs

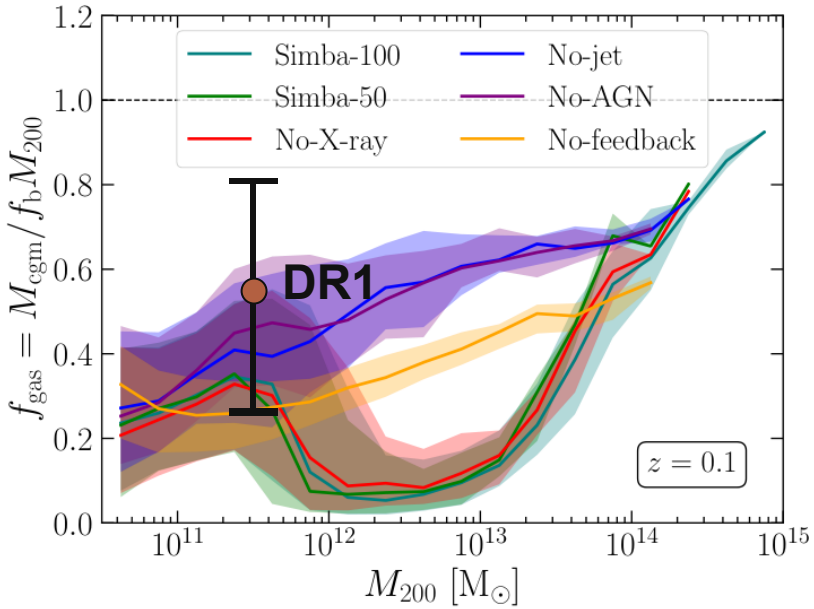
DR1 posterior



$$f_{\rm IGM} = 0.59^{+0.11}_{-0.10}$$

Khrykin+24

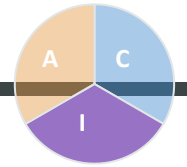
SIMBA + DR1 constraint



$$f_{\rm gas} = 0.55^{+0.26}_{-0.29}$$

Halo-averaged DR1 constraint

Cosmology: Assumed Initial realization: Constrained Astrophysics: Target

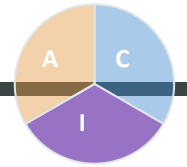


What's new in FLIMFLAM DR2?

DR1	DR2
8 FRB sightlines	**50 FRB sightlines**
AAT / AAOmega Foreground galaxy spectroscopy (r<19.8)	**Subaru PFS + deeper spectroscopy** **More faint foreground galaxies (r<21.3)**
ARGO Evolved-density inference HMC sampling	**MaiGO** **Initial conditions + PM** **RSD + lightcone; MCLMC sampling**

Khrykin+24

Cosmology: Assumed Initial realization: Target Astrophysics: Assumed



MaiGO: the forward model

MCLMC samples

$$p(\mathbf{s} \mid \mathcal{G}) \propto \pi(\mathbf{s})\, \mathcal{L}[\mathcal{G} \mid \lambda_{\rm obs}(\mathbf{s})]$$

New!

White noise

$$\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

→

Initial field

$$\delta_{\mathbf{k}} = \sqrt{P(k)}\, s_{\mathbf{k}}$$

jax-cosmo

→

Gravity

2LPT + Particle-Mesh
w/ JaxPM

→

Lightcone

Interpolate particle positions and momenta

↓

RSD

Apply peculiar velocities

$$1 + \delta_{\rm RSD}$$

←

Galaxy bias

Power-law + cutoff
Predict full count rate

←

Survey selection

$$\lambda_{\rm obs} = R\,\lambda$$

Apply incompleteness

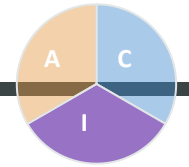
←

Predicted counts

$$\lambda_{\rm obs}$$

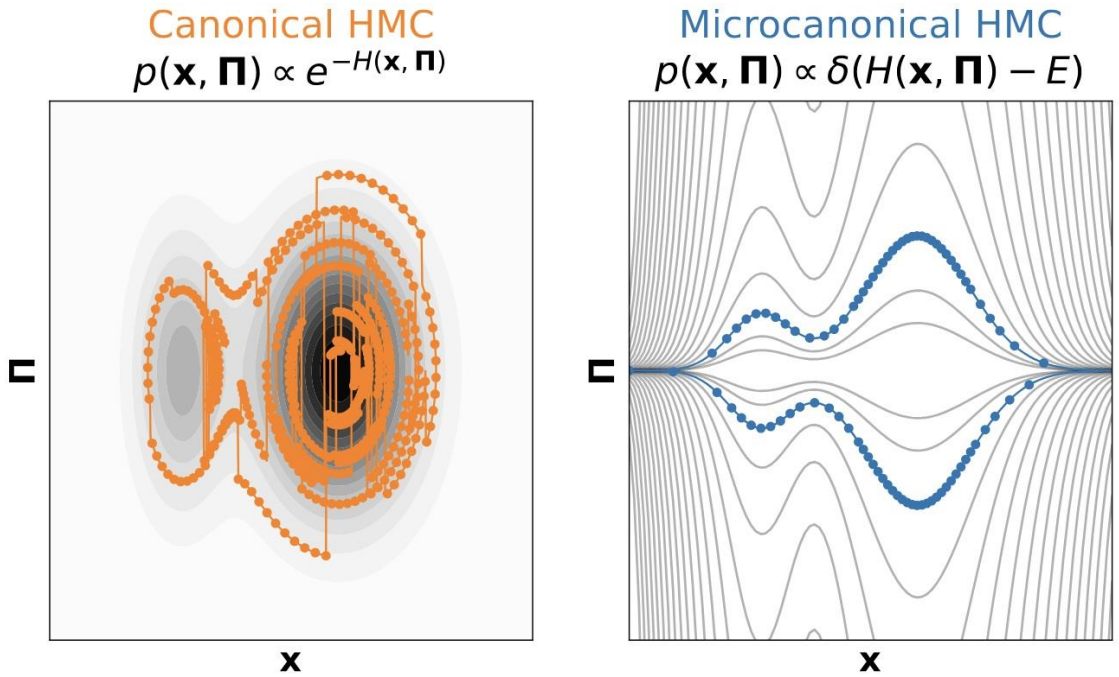
Counts in each cell
Compare with catalog

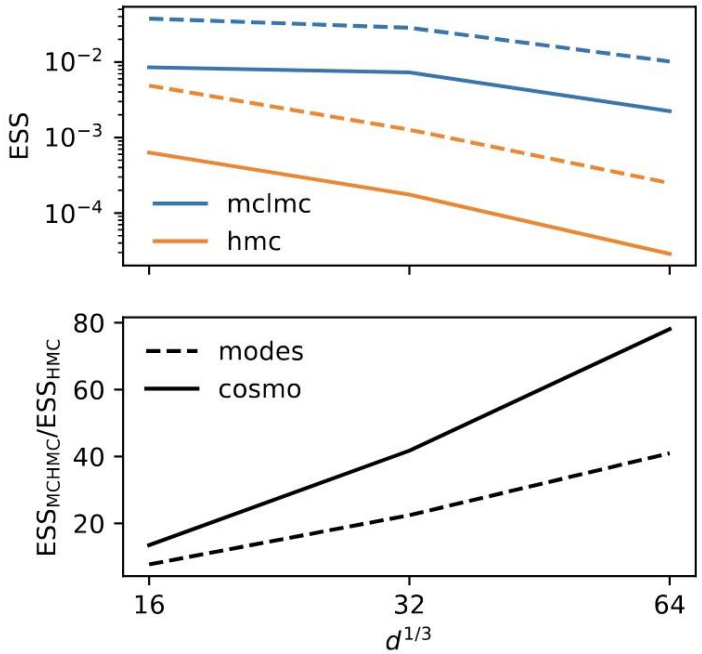
ARGO: evolved-density inference / HMC MaiGO: initial-field inference / MCLMC



MaiGO sampling: HMC and MCLMC

* Happy to discuss on MCLMC usage!





Same target marginal, different dynamics

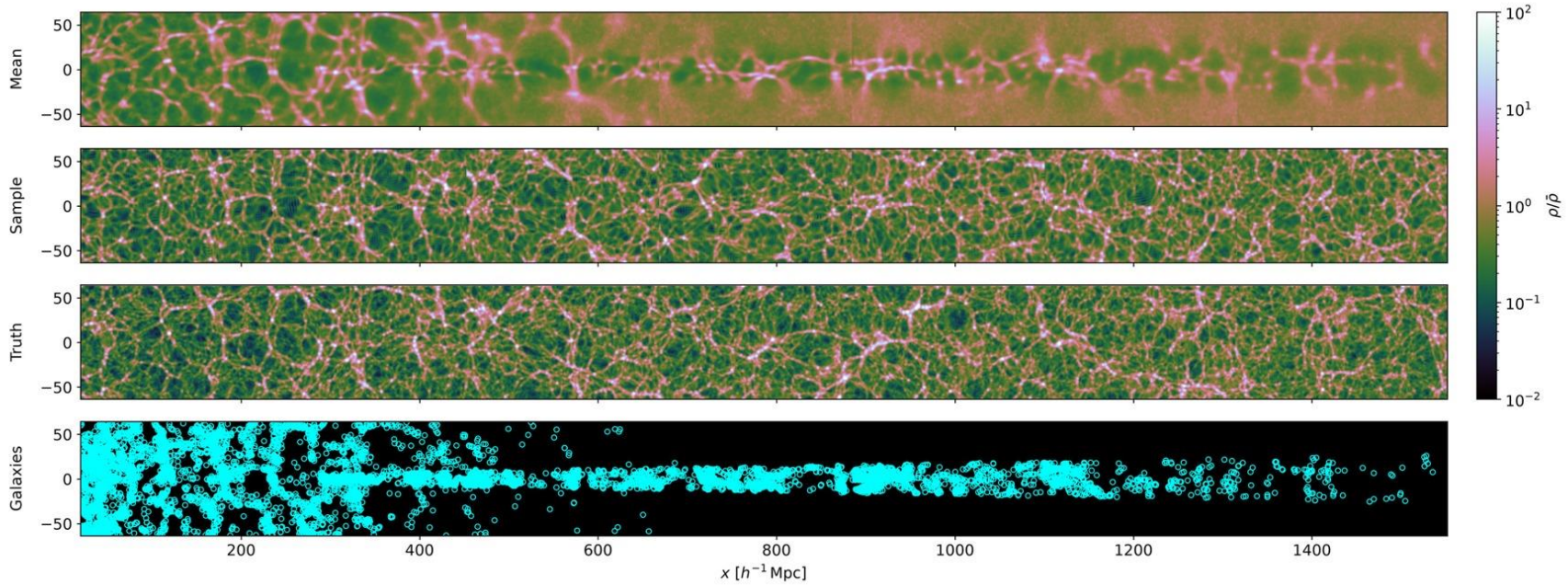
ARGO uses HMC
MaiGO uses MCLMC

Robnik+23; Bayer+23

Cosmology: Assumed Initial realization: Target Astrophysics: Assumed

MaiGO: a showcase on mock data

Mock: Millennium N-body simulation + L-Galaxies SAM



Sample Mean

One sample

Simulation Truth

Tracer Galaxy

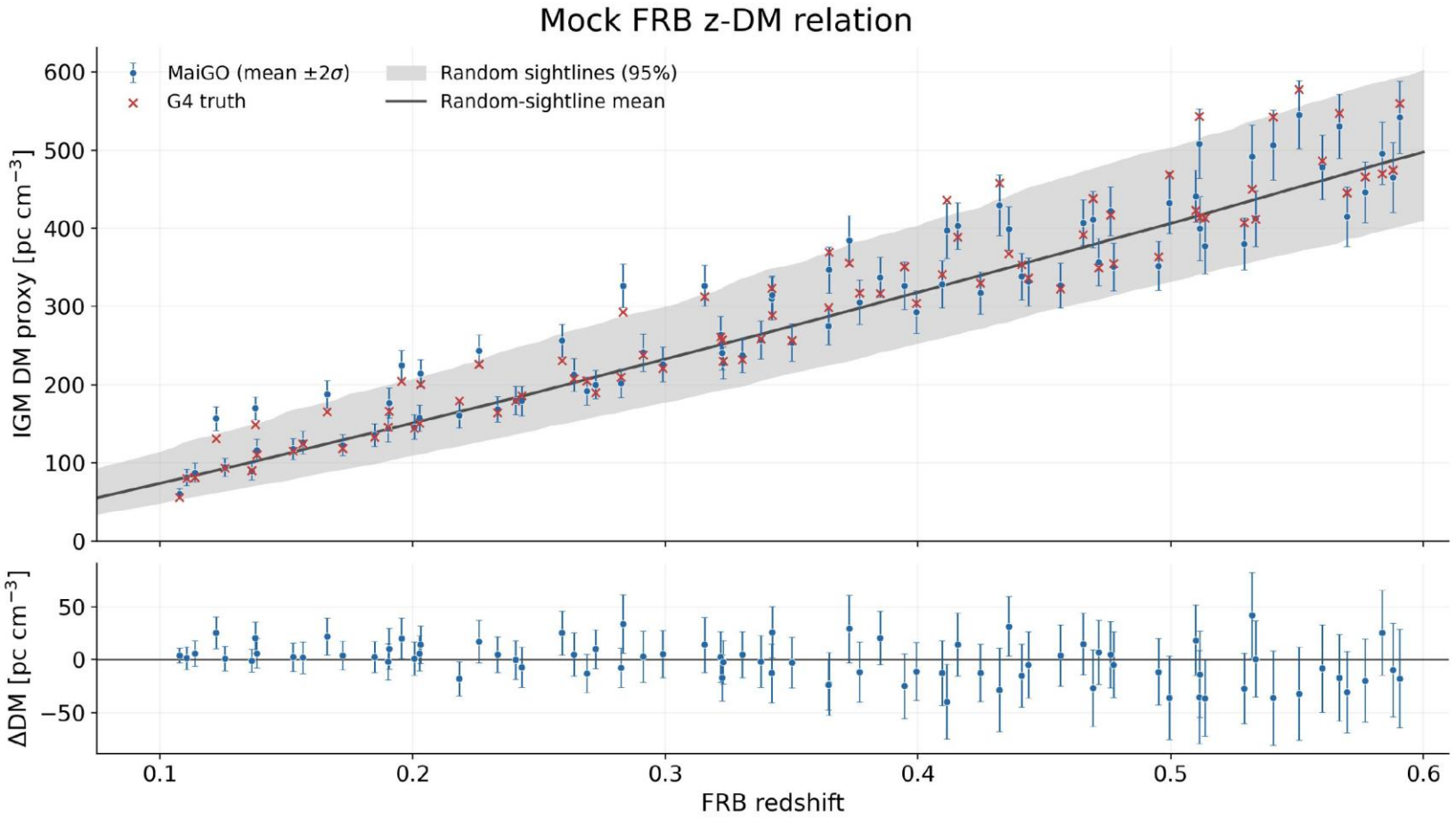
Where galaxies constrain the field

The mean recovers shared structures.

Beyond the observational window

The mean smooths; samples retain fluctuations.

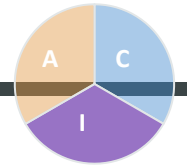
Foreground information sharpens DM predictions



MaiGO

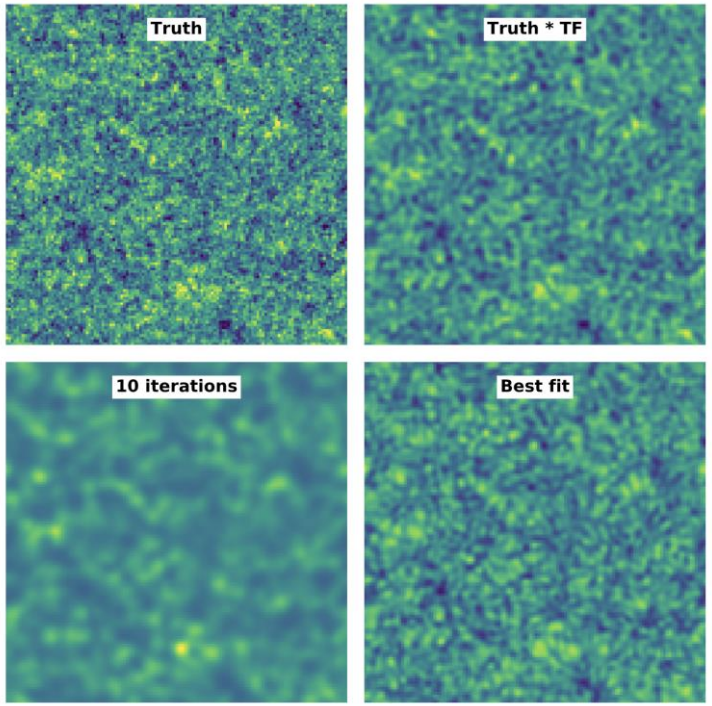
Sightline-specific DM predictions.

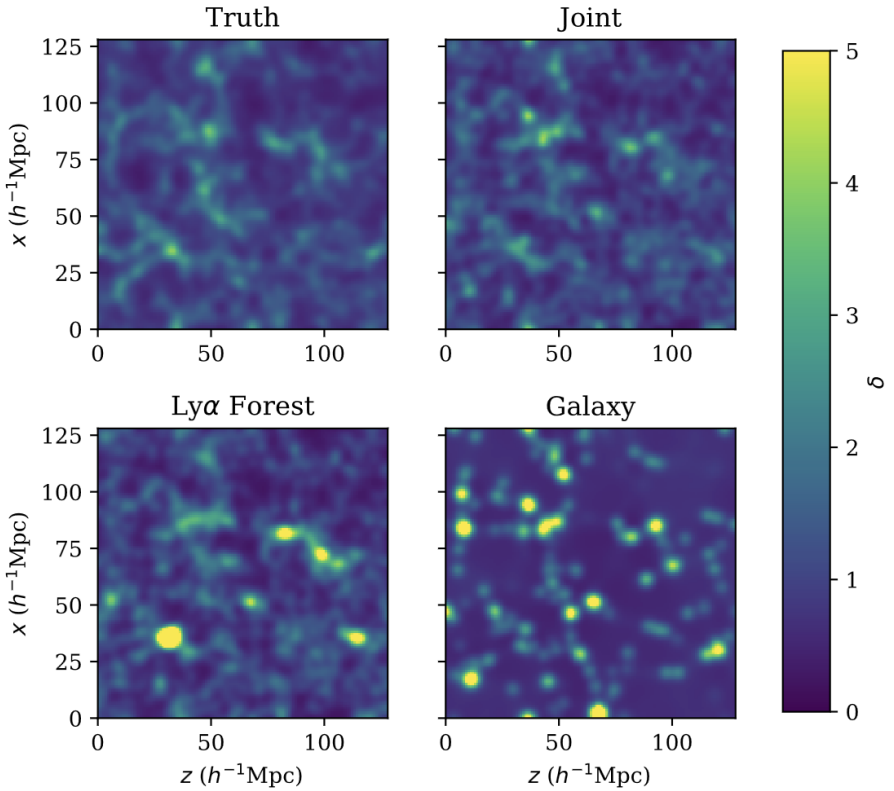
DR2 analysis ongoing; stay tuned!



MujicA: a spin-off with the same forward model

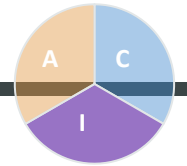
MCLMC autotuning is hard… Can we start with a point estimation?
Point estimations at home (here): on galaxies/halos/Lya





Seljak+17; Modi+18; Horowitz+19; Horowitz+21

Cosmology: Assumed Initial realization: Target Astrophysics: Assumed



MujicA: fit the data but retain unobserved modes

MAP → **MLE** → **MujicA**

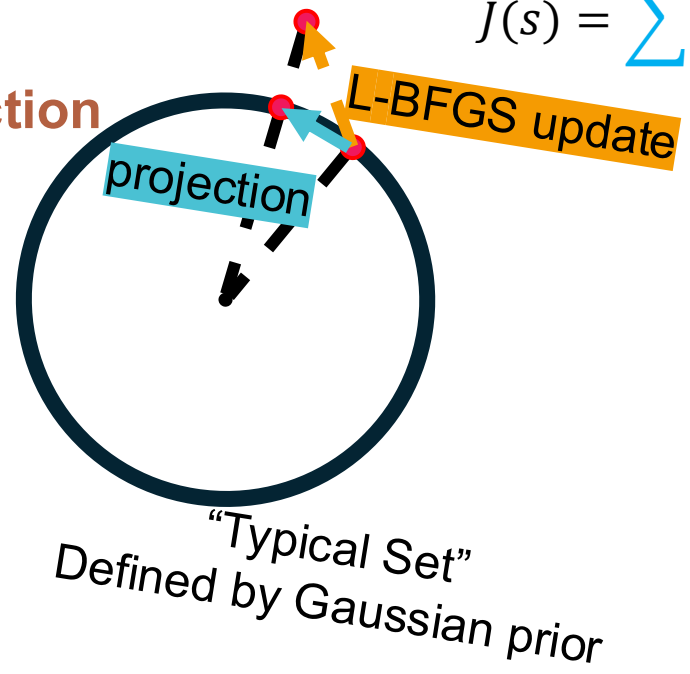
Gaussian prior shrinks weakly constrained modes.

Remove the prior penalty; unobserved modes are free.

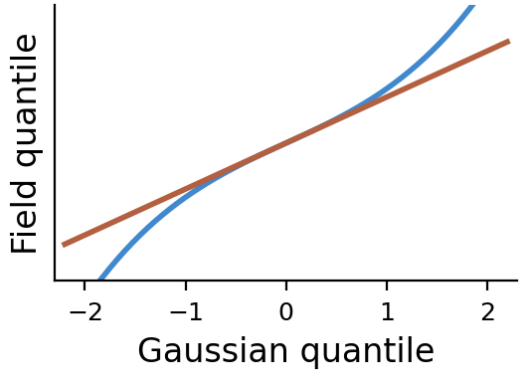
MLE + corrections inspired by Gaussian prior

$$J(s) = \sum [\lambda_G - N_G \log \lambda_G] + \frac{1}{2} s^{\mathrm{T}} s$$

Projection

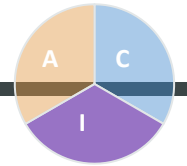


Rank-order matching ("ambundance matching")



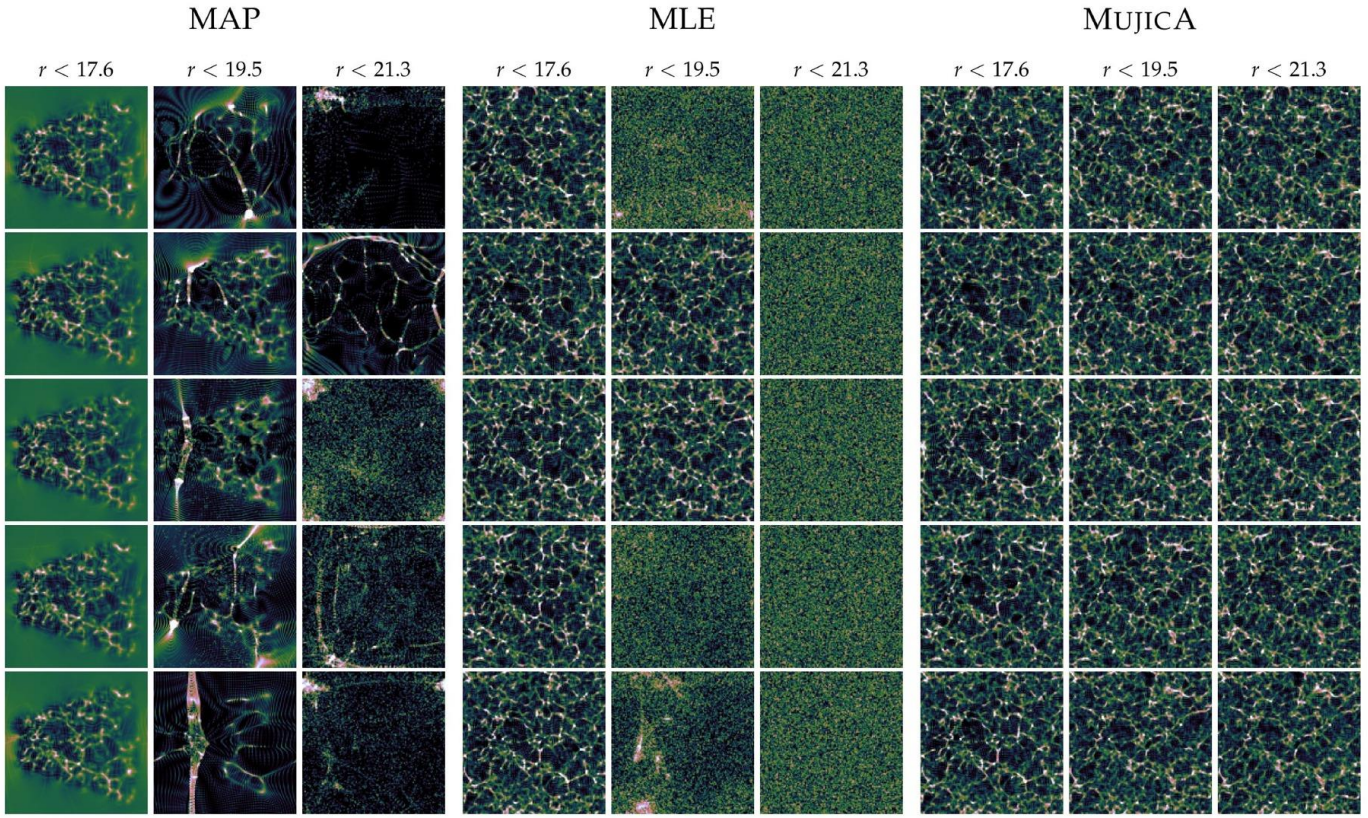
Dong+26

Cosmology: Assumed Initial realization: Target Astrophysics: Assumed



MujicA tests: robustness across initial guesses

Mock: Millennium N-body simulation + L-Galaxies SAM; but with a wide footprint + padding



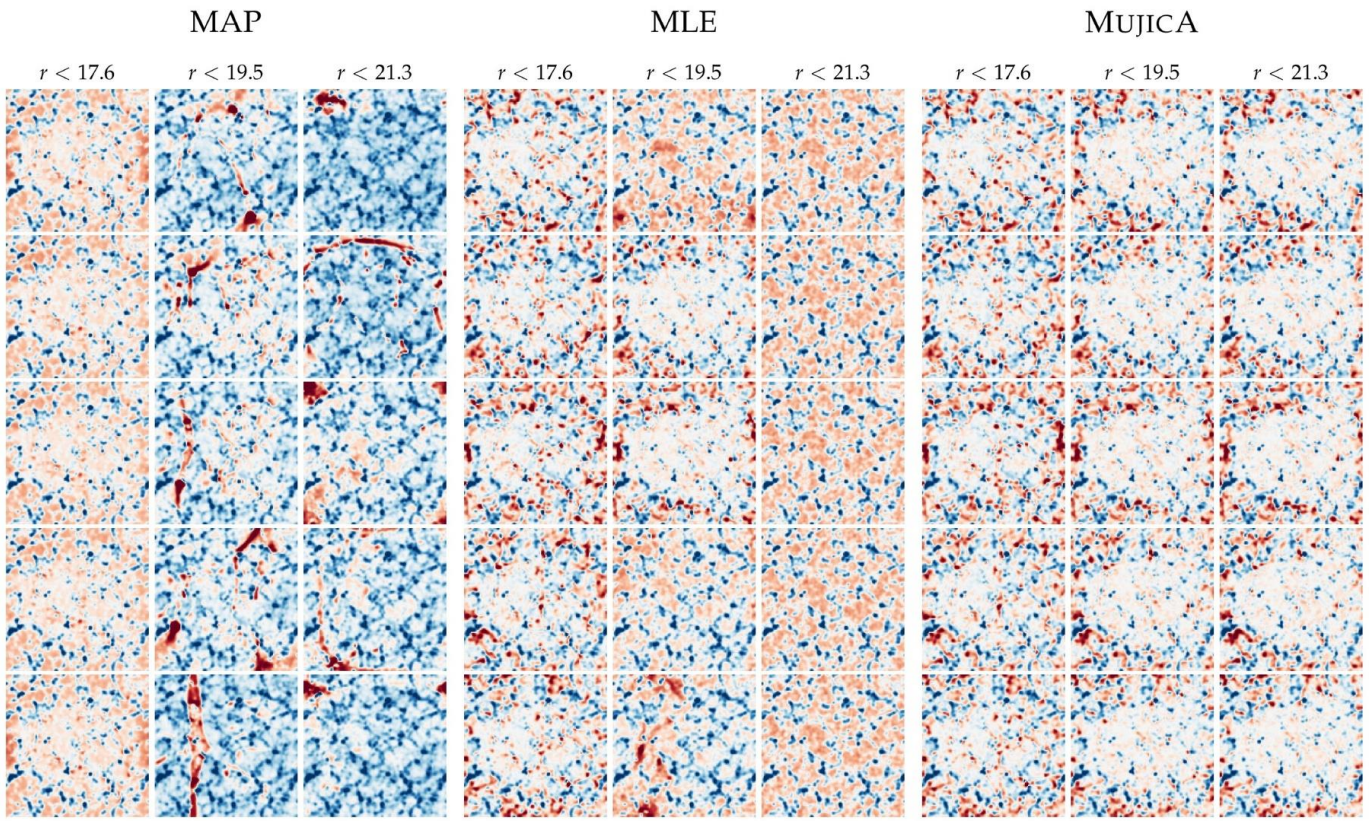
Different initial guesses

Each row starts from a different random field.

Fewer failed reconstructions with MujicA.

Dong+26

MujicA tests: reconstruction residuals



Inside the window

Assess recovery of observed structures.

Outside the window

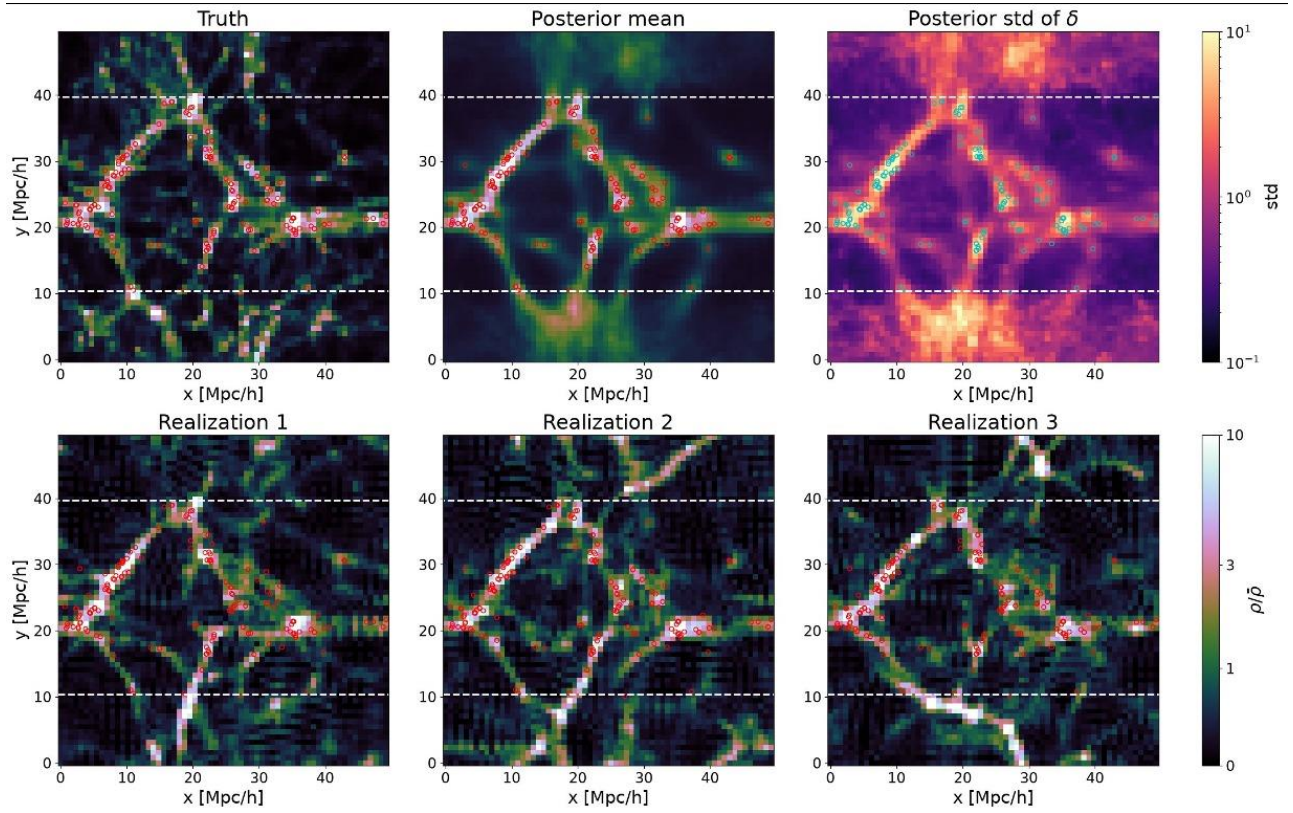
Random modes differ from the mock truth.

Agreement on P(k)!

Ready to use for autotuning

Dong+26

Toward smaller scales with MaiGO: SIMBA-50 (prelim)



50 cMpc/h periodic box

z = 0.3; 64^3 grid

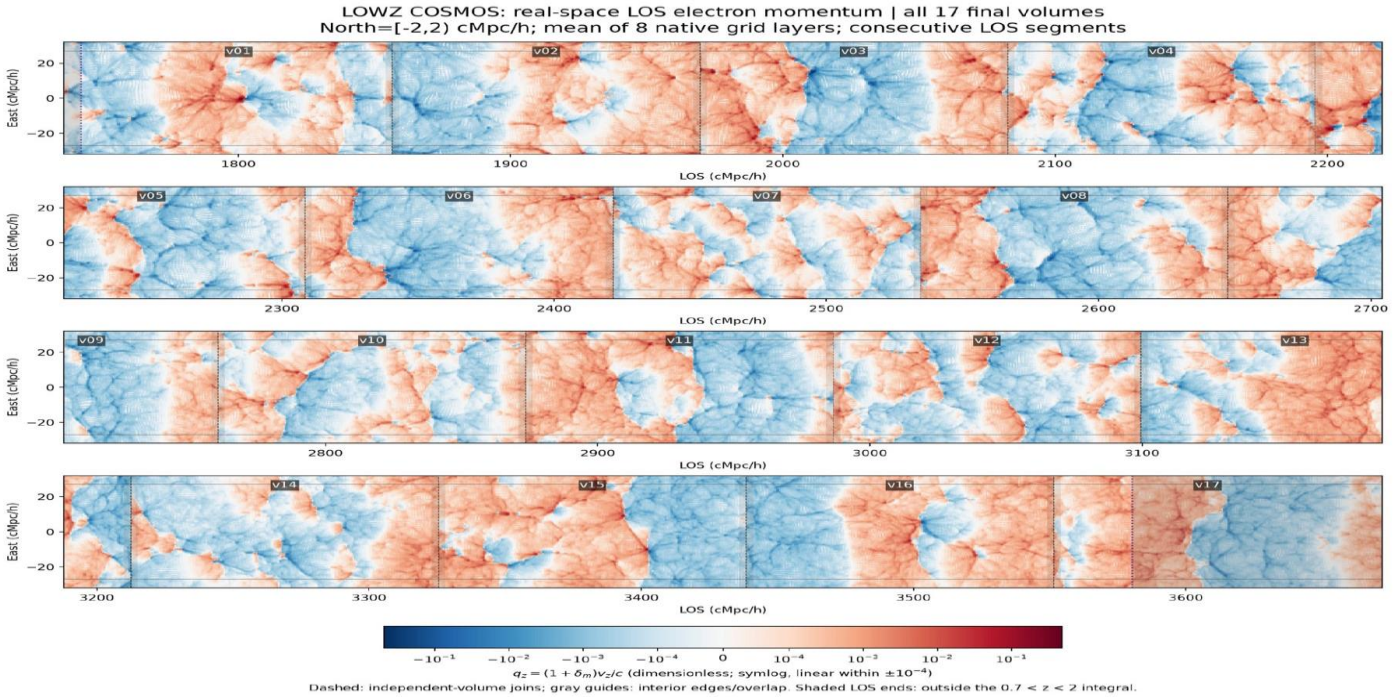
Selection matched to the DESI One-Percent Survey

Inside the dashed lines: observed galaxies

No measurement errors, RSD or lightcone.

MujicA point estimation: COSMOS momentum -> kSZ? (prelim)

zCOSMOS bright/deep + FMOS + DESI: 15,567 records, 0.7 < z < 2



17 LOS volumes

64 × 64 × 128 cMpc/h per box

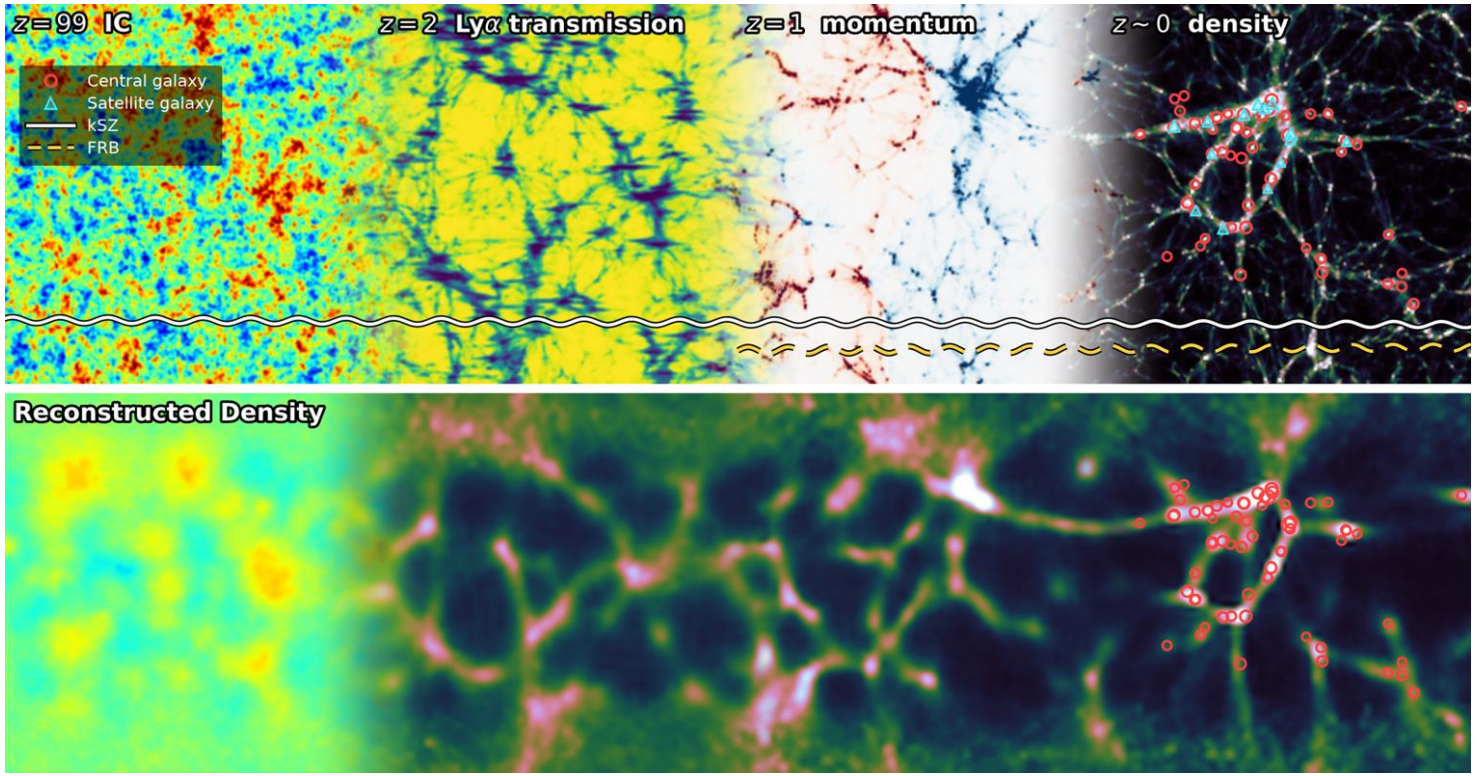
128 × 128 × 256 grid

0.5 cMpc/h cell spacing

5 cMpc/h margins

The narrow transverse footprint leaves bulk motion unconstrained.

Towards joint inference of multi probes at field-level



3D tracers (galaxy; maybe WL/LIM)

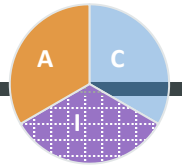
↓ Differentiable LSS sim

IC & structures

x Differentiable (semi-) hydro sim

Baryon probes (Lya, FRB, SZ…)

multi-probe baryon costraint across redshifts

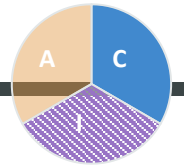


Summary

COSTCO + CLAMATO — A constrained dense environment reveals an unexpected absorption signal and motivates tests of heating.

FLIMFLAM + FRBs — Foreground reconstruction turns the DM distribution into a prediction for individual sightlines.

MaiGO + MujicA — A shared forward model supports posterior sampling and point estimation for different reconstruction goals.



Appendix

COSMIC-BIRTH reconstruction details

Foreground galaxy calibration

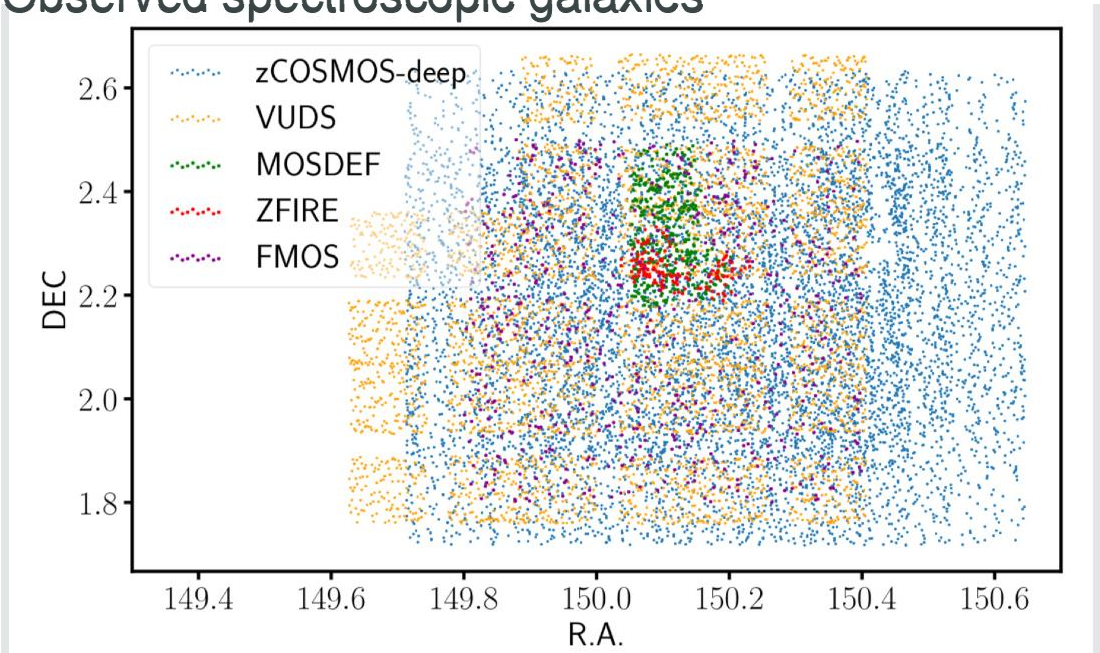
DM statistics and cosmology: A1–A7

References

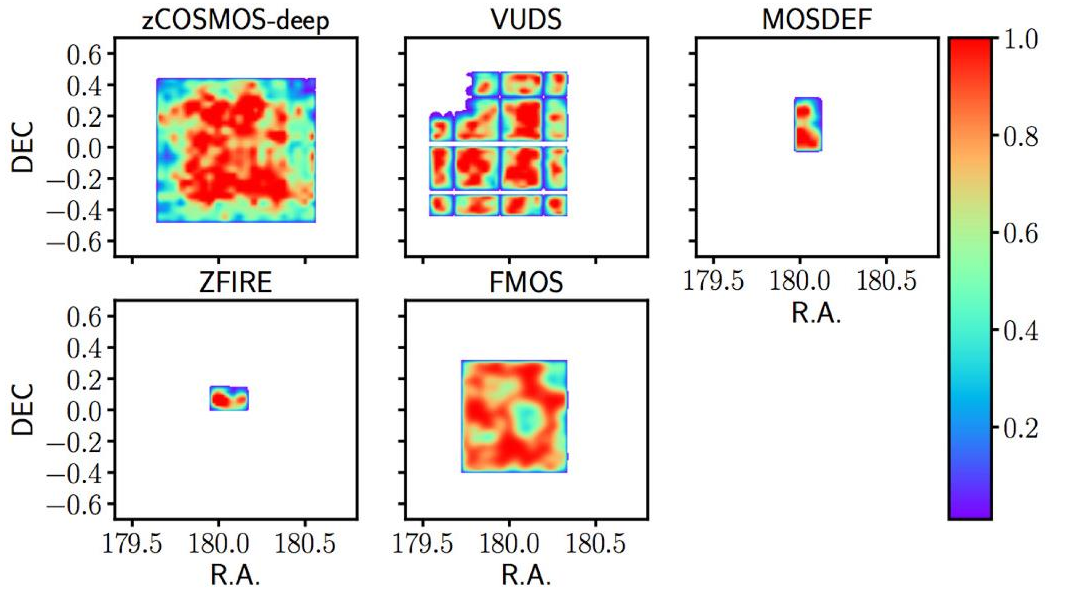
COSTCO: combining the COSMOS surveys

Five galaxy samples, one underlying density field

Observed spectroscopic galaxies



Different angular completeness

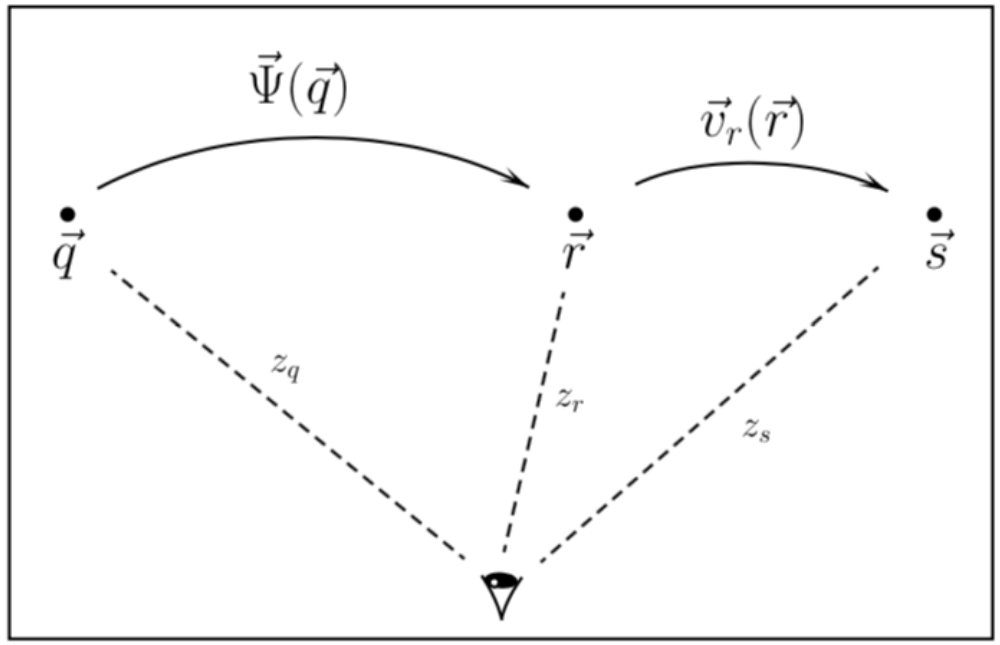


Survey geometry, radial selection and galaxy bias enter separately for each sample.

Kitaura+21; Ata+22



From observed positions to the initial field



q Initial, Lagrangian position

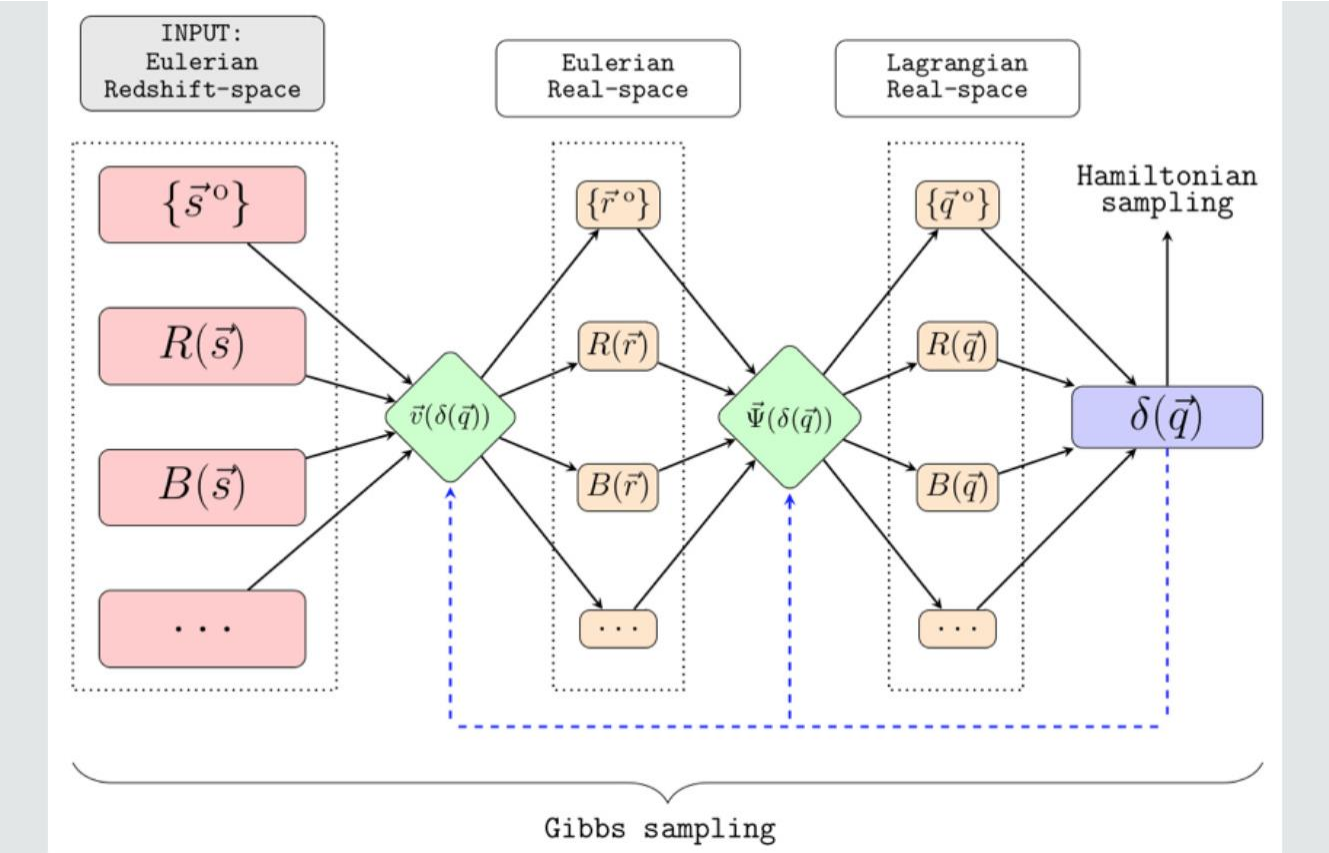
r Evolved, real-space position

s Observed redshift-space position

Model both gravitational displacement and peculiar velocity

Kitaura+21; Ata+22

COSMIC BIRTH: sampling constrained realizations



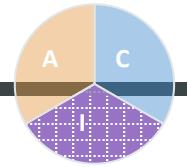
Sample the initial density

HMC conditional on the current tracer positions, response and bias

Update the coordinate mapping

Forward evolution supplies the displacements and velocities

Kitaura+21; Ata+22



COSMIC BIRTH: the inference equations

Joint information from multiple surveys

$$p(\delta \mid \{N_k\}, R, B) \propto \pi(\delta) \prod_k L_k(N_k \mid \lambda_k)$$

$\lambda_k = R_k B_k(\delta)$ R: survey response B: galaxy-bias model

1 Sample the field with HMC

$$\delta(q) \sim p(\delta(q) \mid \{q\}, B(q), R(q))$$

2 Update tracer coordinates

$$\{q\} \sim p(\{q\} \mid \{s^{obs}\}, \delta(q), R(q))$$

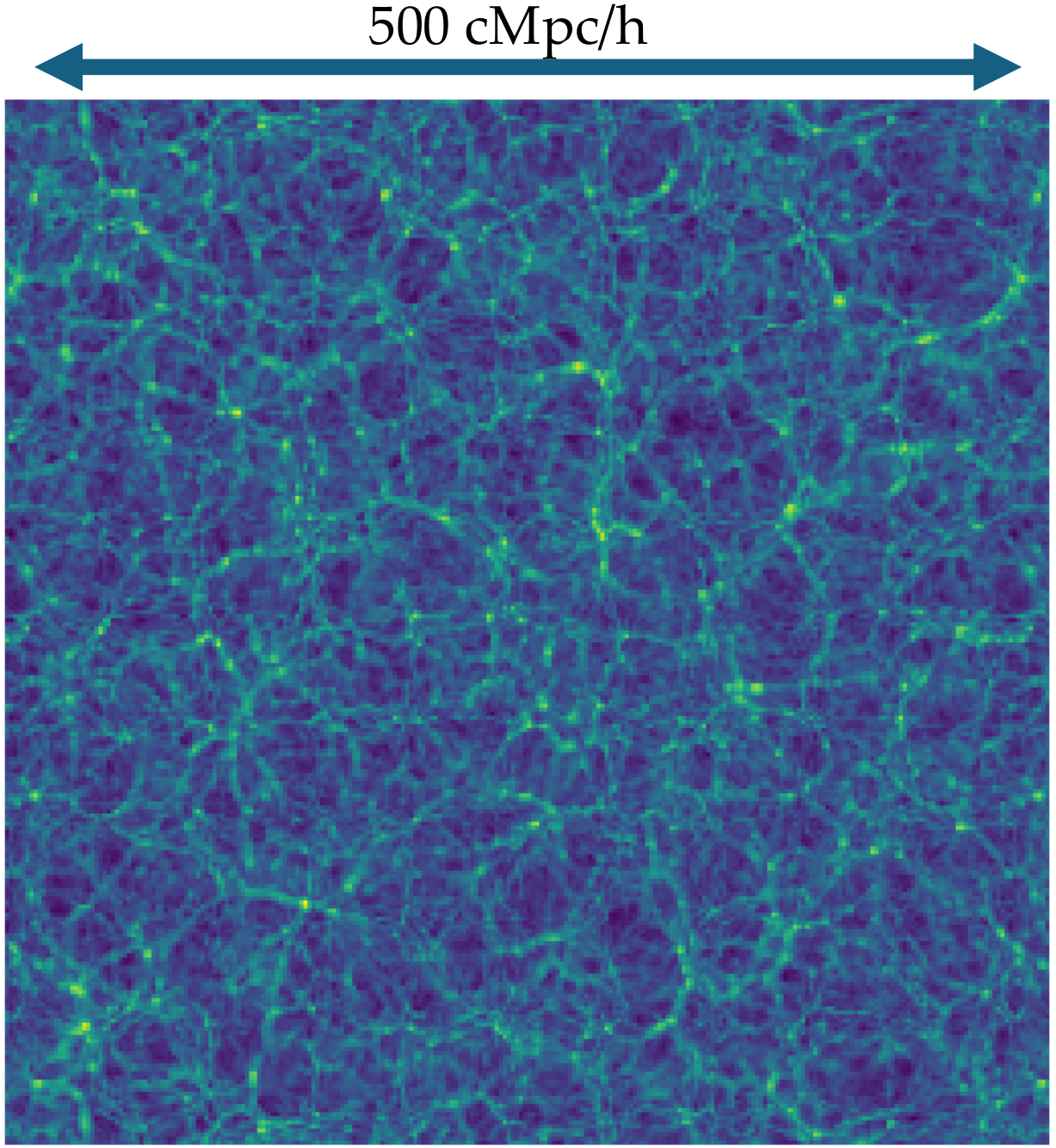
Constrained simulations from sampled initial fields

Initial density samples 2LPT particle initial conditions N-body evolution

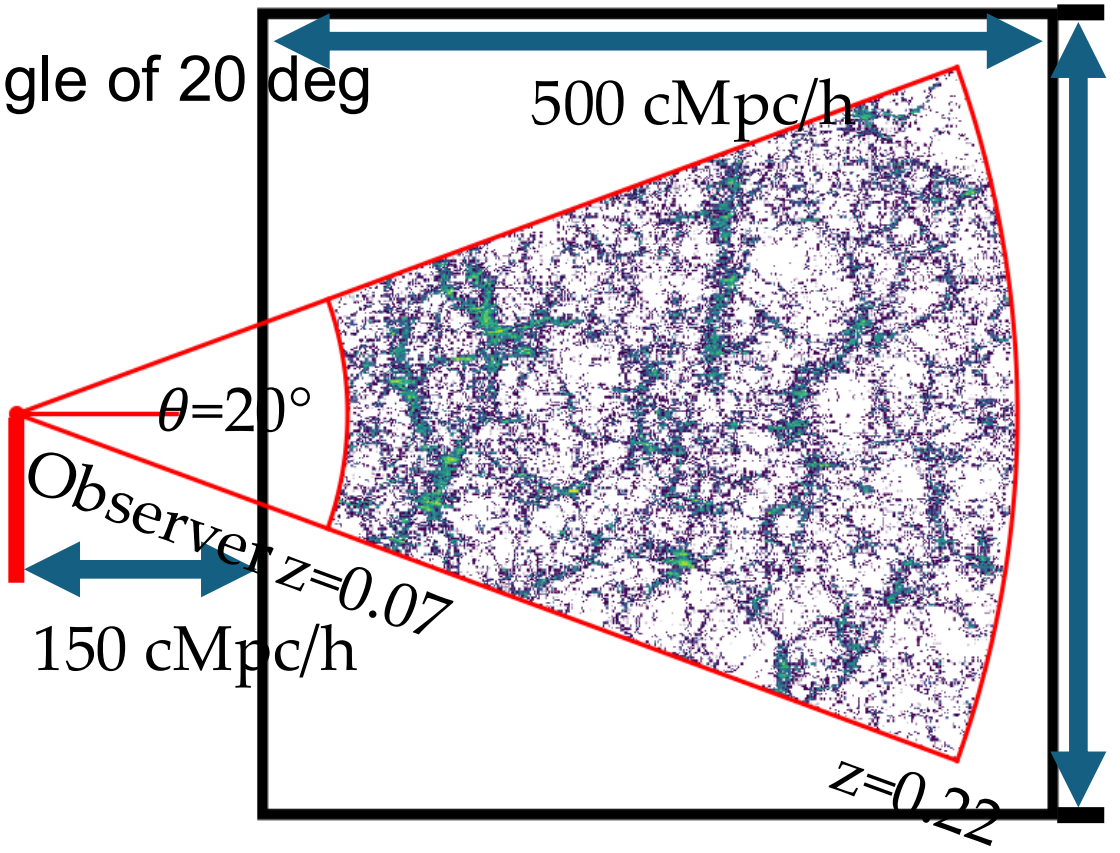
Kitaura+21; Ata+22

Cosmology: Assumed Initial realization: Constrained Astrophysics: Assumed

- Background sim: Milennium simulation (Springel+05)
  - Planck13 cosmology
  - $\sigma_8 = 0.829$
  - $h = 0.673$
  - $\Omega_m = 0.315$
  - Boxsize = $500\ \mathrm{Mpc/h}$
  - N-body sim with Gadget-2



- Mock galaxy catalog: L-Galaxy SAM catalog by Henriques+15
  - With only spec-z galaxies
  - redshift cut: 0.07<z<0.22
  - angular cut: FOV with half opening angle of 20 deg
  - magnitude cut: sdss_r < 21.7



- Survey footprint: R = 20 deg
- Boxsize (500, 500, 500) cMpc
- Particle Mesh Resolution
  - (256, 256, 256) for optimization
  - (128, 128, 128) for sampling

