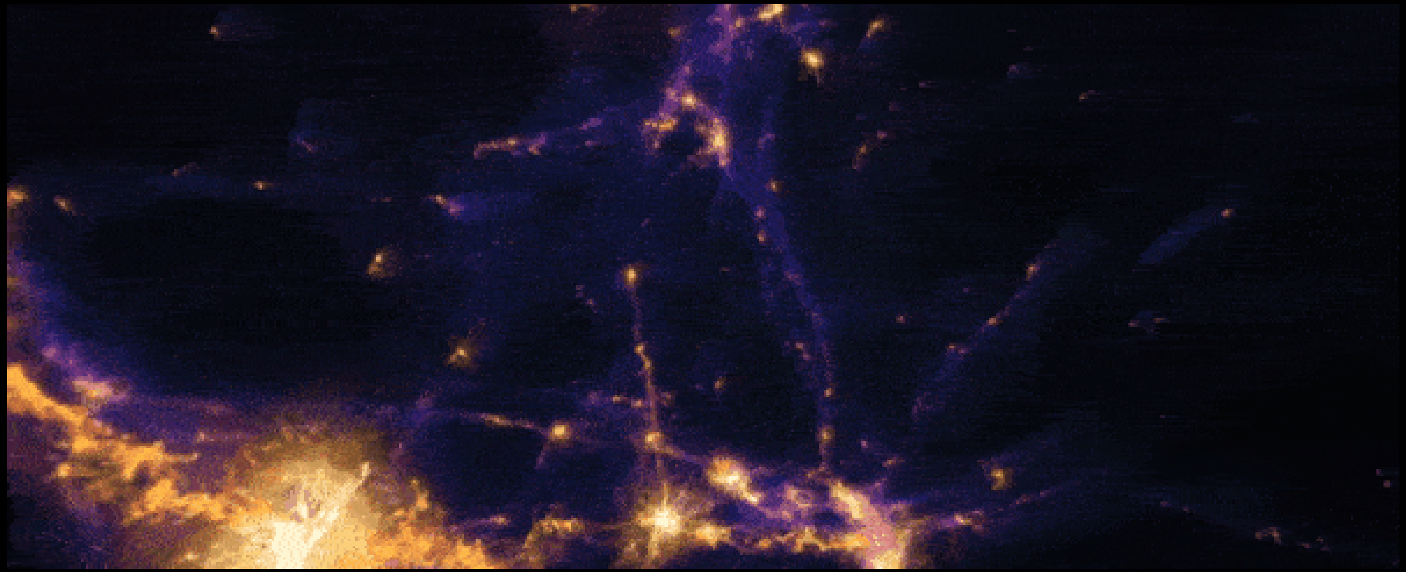
Forward modeling in the era of cosmological surveys



Boryana Hadzhiyska, PhD Candidate

CENTER FOR ASTROPHYSICS
HARVARD & SMITHSONIAN

Berkeley & LBL

10/5/2021

My collaborators

Galaxy-halo connection:

- Lars Hernquist (CfA)
- Sownak Bose (CfA, Durham)
- Sihan Yuan (CfA, Stanford)
- Rachel Somerville (CCA)
- Jay Wadekar (IAS)

Abacus *N*-body simulation:

- Daniel Eisenstein (CfA)
- Lehman Garrison (CfA, CCA)
- Nina Maksimova (CfA)

Analytic approaches (e.g., bias expansion, CMB):

- David Alonso (Oxford)
- Andrina Nicola (Princeton, Washington University)
- Anže Slosar (Brookhaven National Lab)
- Carlos García-García (Oxford)
- Blake Sherwin (Cambridge)

Hadzhiyska+ (2019b), MNRAS.493.5506H
Hadzhiyska+ (2020b), MNRAS.501.1603H
Yuan, Hadzhiyska+ (2020), MNRAS.502.3582Y
Hadzhiyska+ (2020c), MNRAS.502.3599H
Hadzhiyska+ (2021b), MNRAS.508..698H

Hadzhiyska+ (2021a, submitted)
Hadzhiyska+ (2021b, submitted)
Maksimova+ (2021), MNRAS.tmp.2270M
Bose+ (2021, submitted)
Yuan+ (2021, submitted)

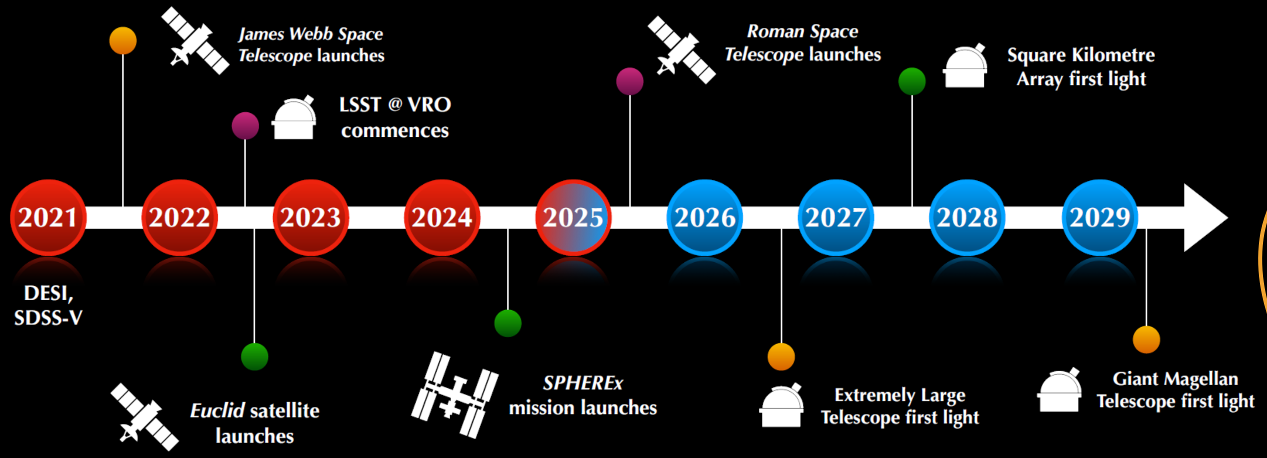
Hadzhiyska+ (2020a), JCAP...10..056H
Hadzhiyska+ (2021a), JCAP...09..020H
Hadzhiyska+ (2019a), PhRvD.100b3547H
Karim+ (2021, in prep.)

Why care about the galaxy-halo connection?

What do we do with all the data?

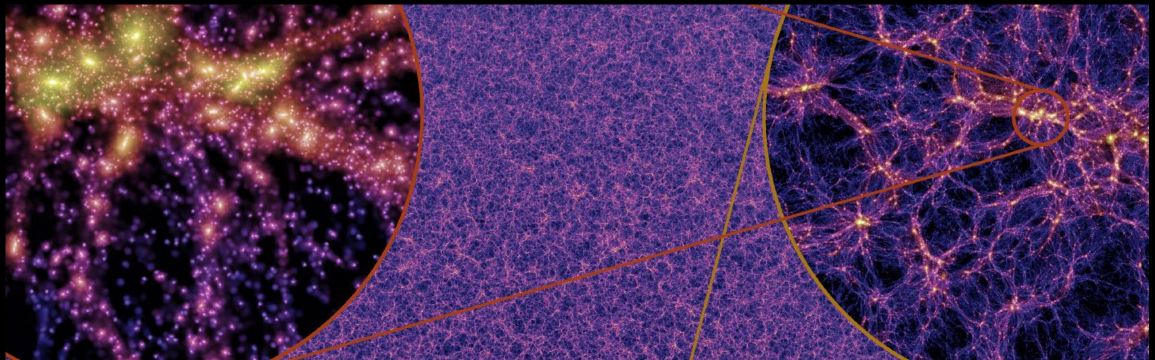
Observations:

- Billion-dollar experiments will measure the galaxy clustering at the subpercent level
- **BUT** without accurate models, we lose valuable information about cosmology and galaxy formation

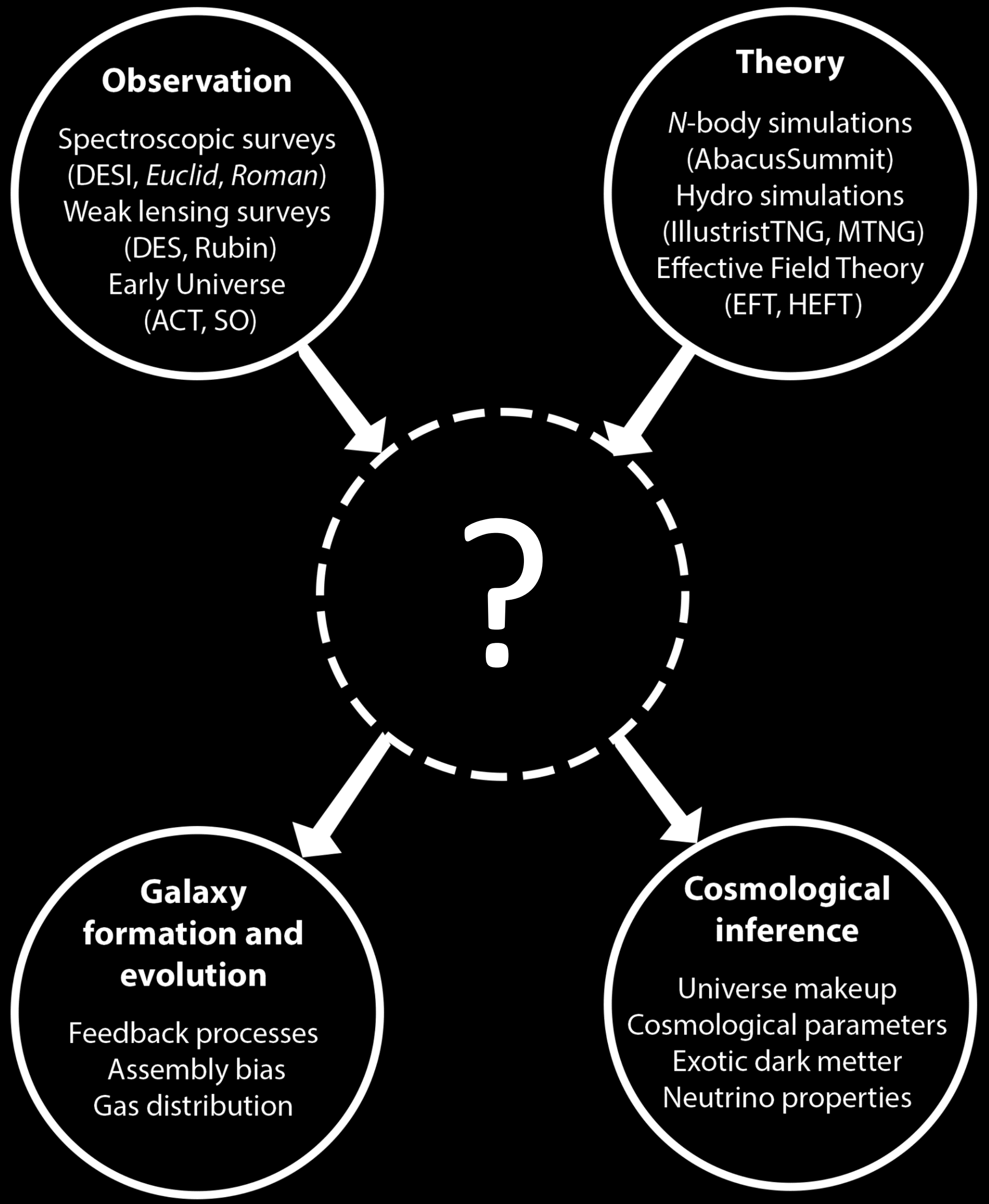


Theory:

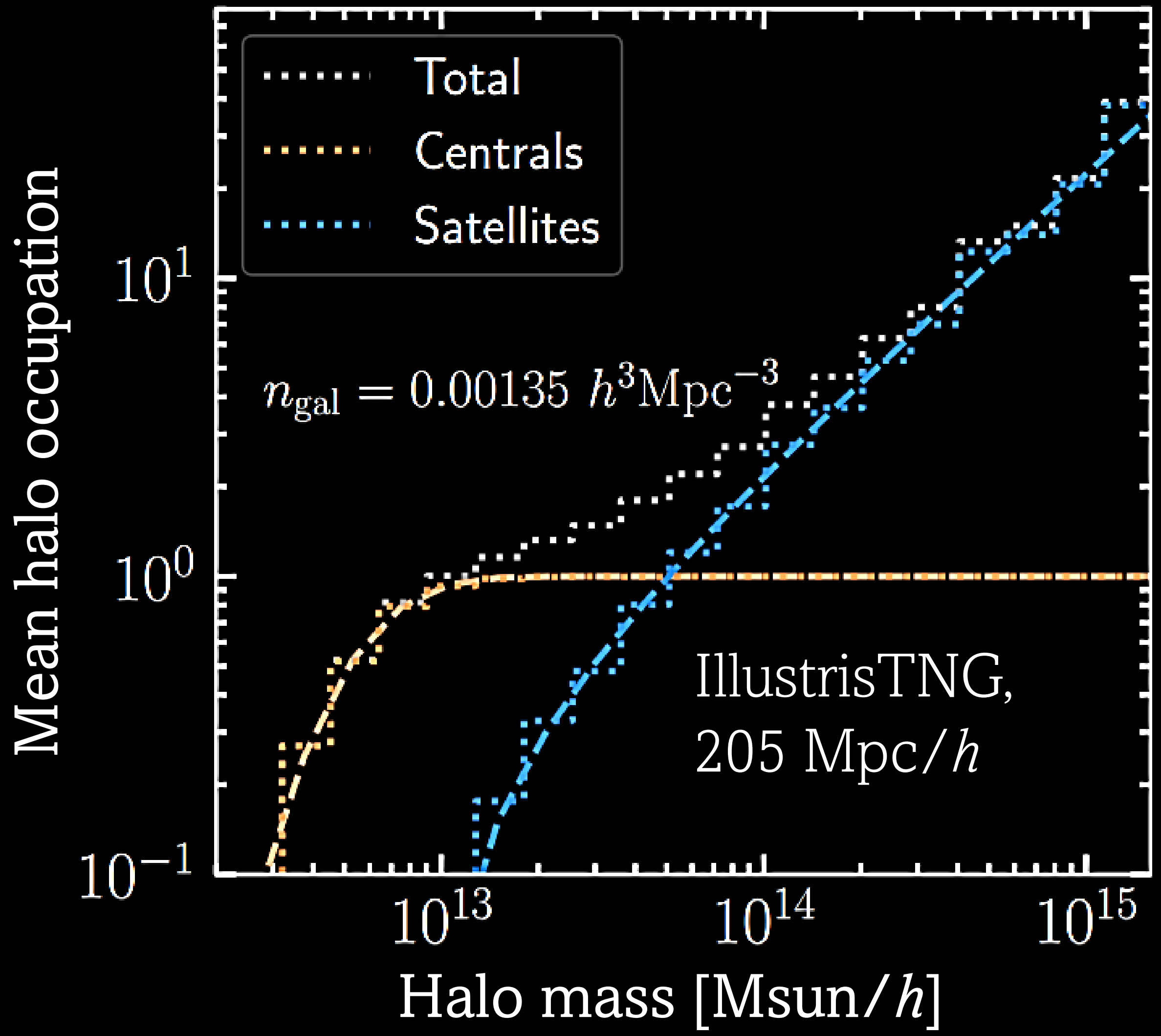
- Bias models discard small scales due to uncertainty in galaxy physics
- **BUT** important cosmological effects are imprinted on small-scales
- *N*-body simulations predict small-scale dark-matter clustering
- **BUT** lack galaxy physics
- Hydro simulations have galaxies
- **BUT** computationally expensive



Conundrum schema

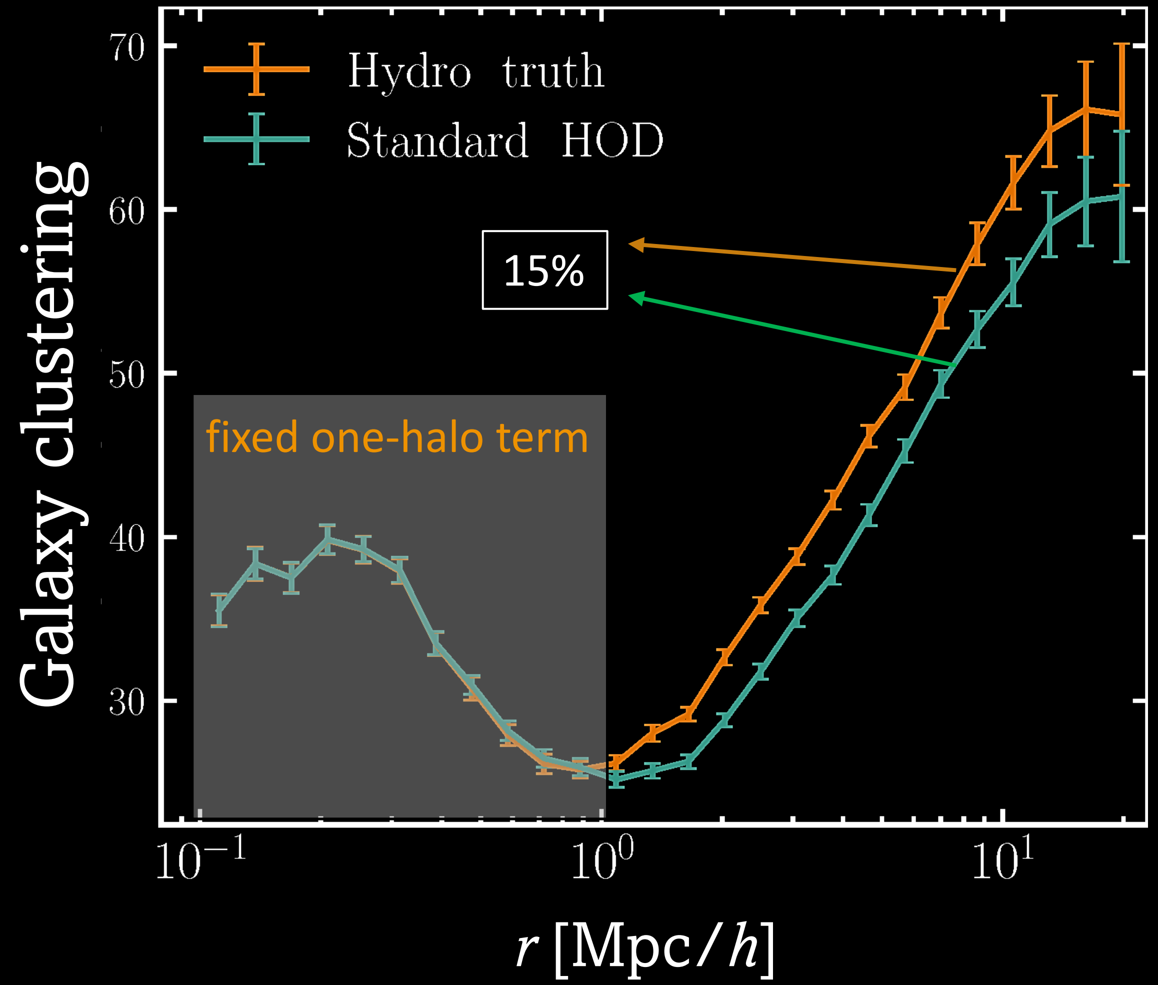


Halo occupation distribution (HOD)



- **HOD theory:** The properties of galaxies are dictated by the properties of the dark-matter halo they reside in.
- **Mass-only HOD:** simplest and most widely used; assumes halo mass alone predicts galaxy occupancy
- **Luminous red galaxies** (LRGs)

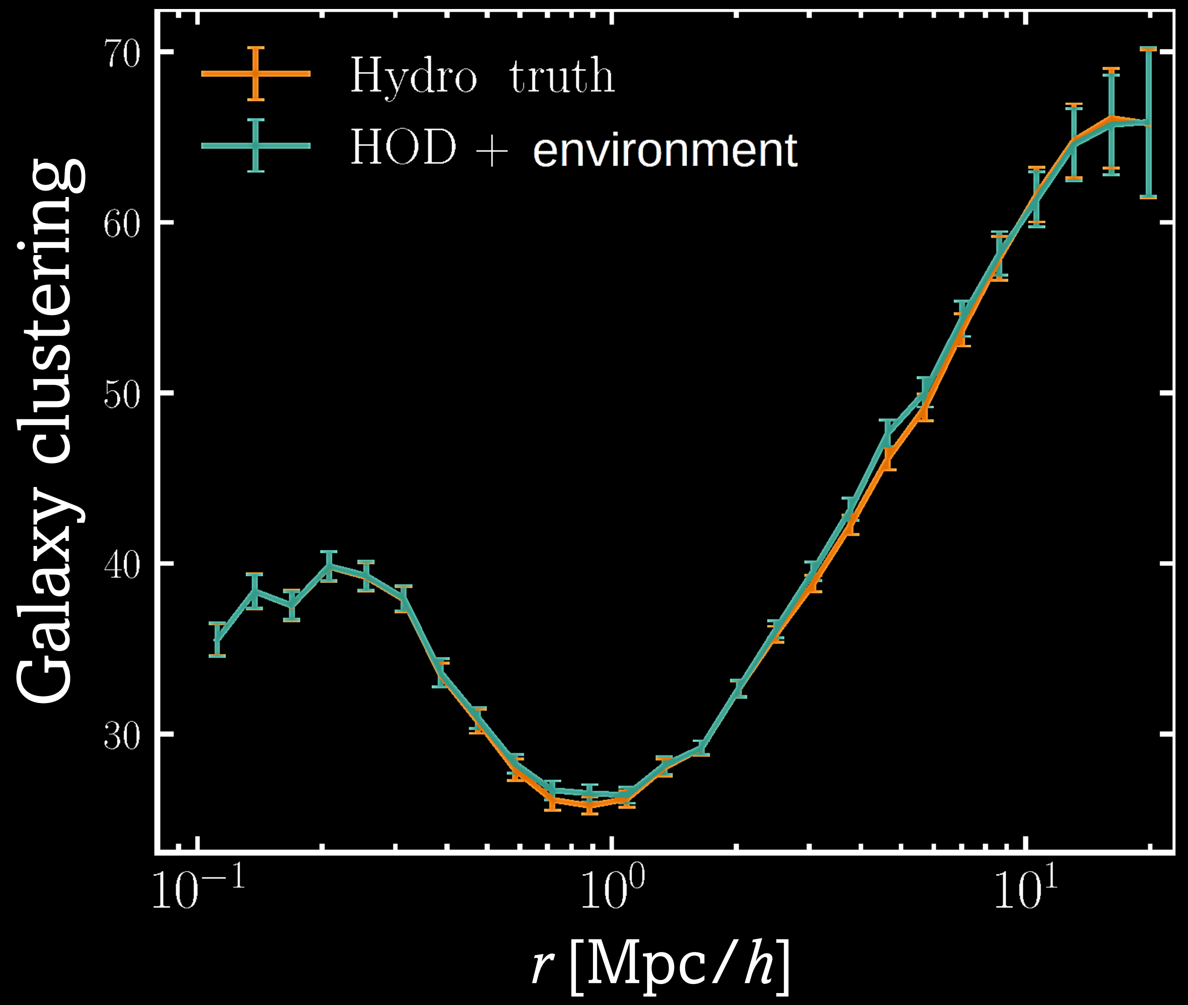
The mass-only HOD does not work well



Hadzhiyska+ (2019), MNRAS.493.5506H

- **Mass-only HOD cannot recover the LRG clustering at the 10-15% level! (see also Beltz-Mohrmann+ (2020), Xu+ (2020))**
- **Well above the subpercent level requirement set by experiments**
- **Proof of "assembly bias": dependence of halo occupancy on additional halo parameters other than mass**

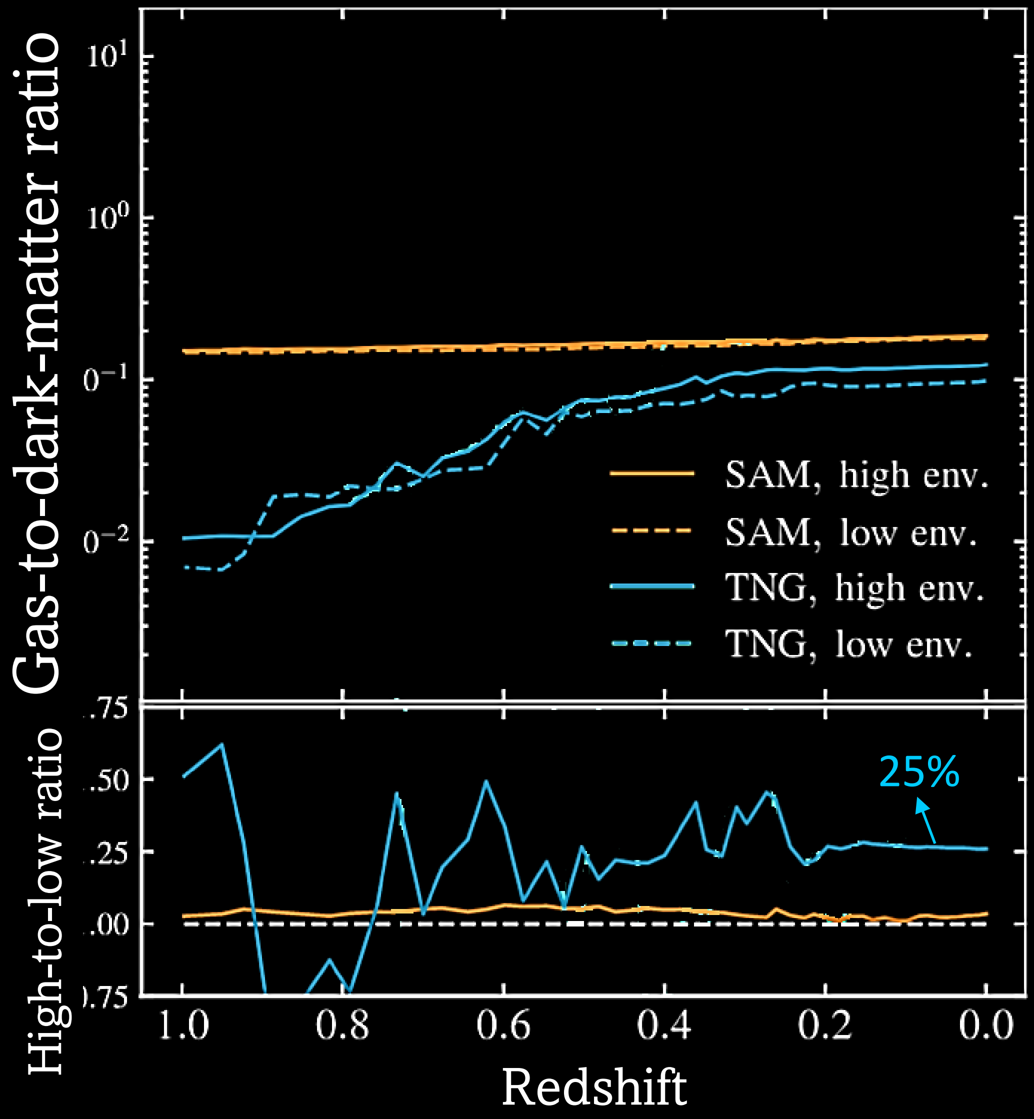
A new kind of "assembly bias" is to blame



Hadzhiyska+ (2019), MNRAS.493.5506H

- Historically studied "assembly bias" parameters: concentration, formation time, spin, velocity dispersion, etc. cannot explain away the difference
- Halo environment can successfully reconcile the difference

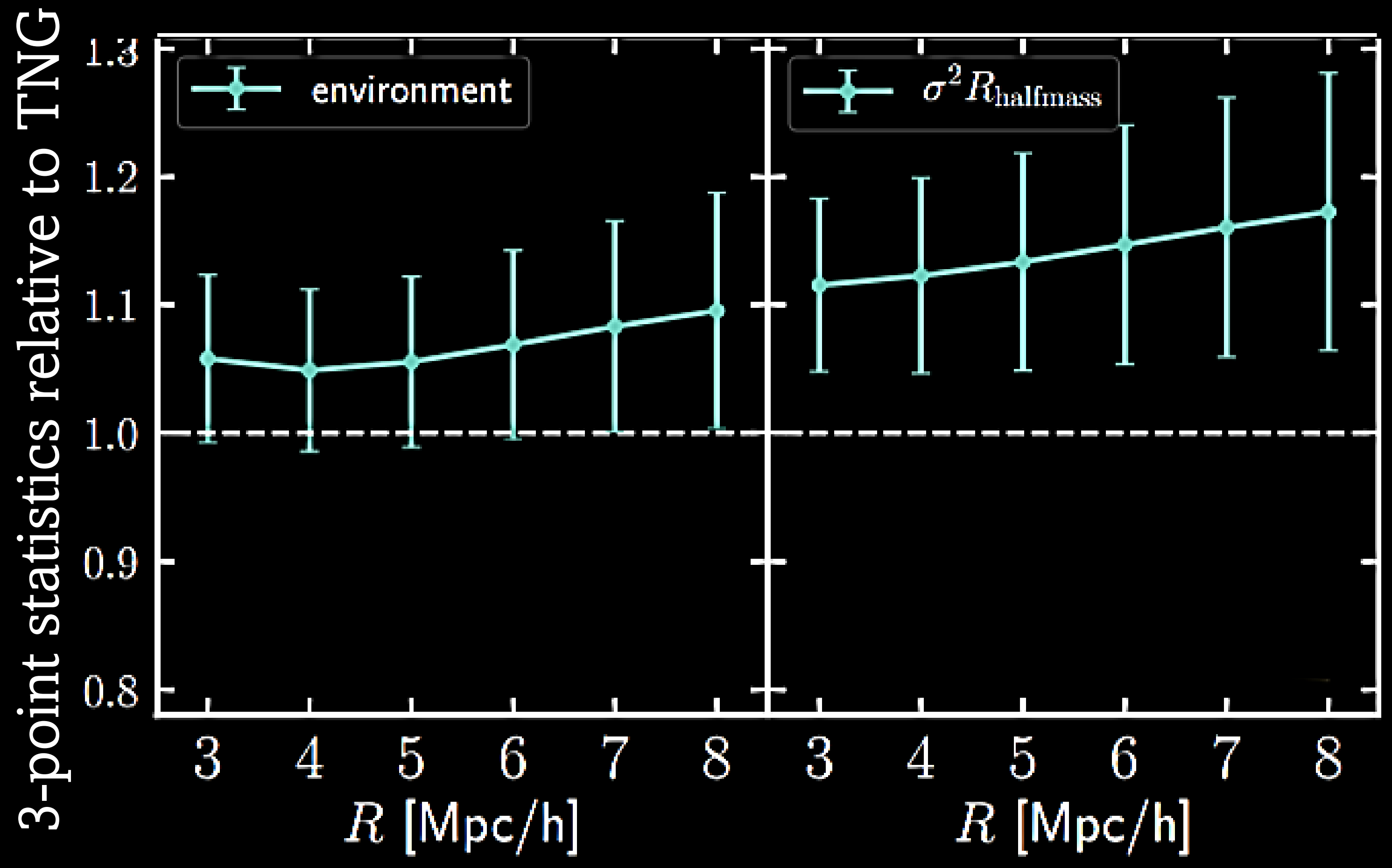
High-density regions supply more gas to the central



- At fixed halo mass, 25% more star-building material available inside high-density TNG halos (backsplash halos, quenching)
- Not the case in most HODs and semi-analytic models (SAMs), which use internal halo properties
- Incorporating environment in HODs and SAMs may be crucial to recovering the galaxy distribution

Hadzhiyska+ (2021b), MNRAS.508..698H

Beyond two-point statistics are valuable!



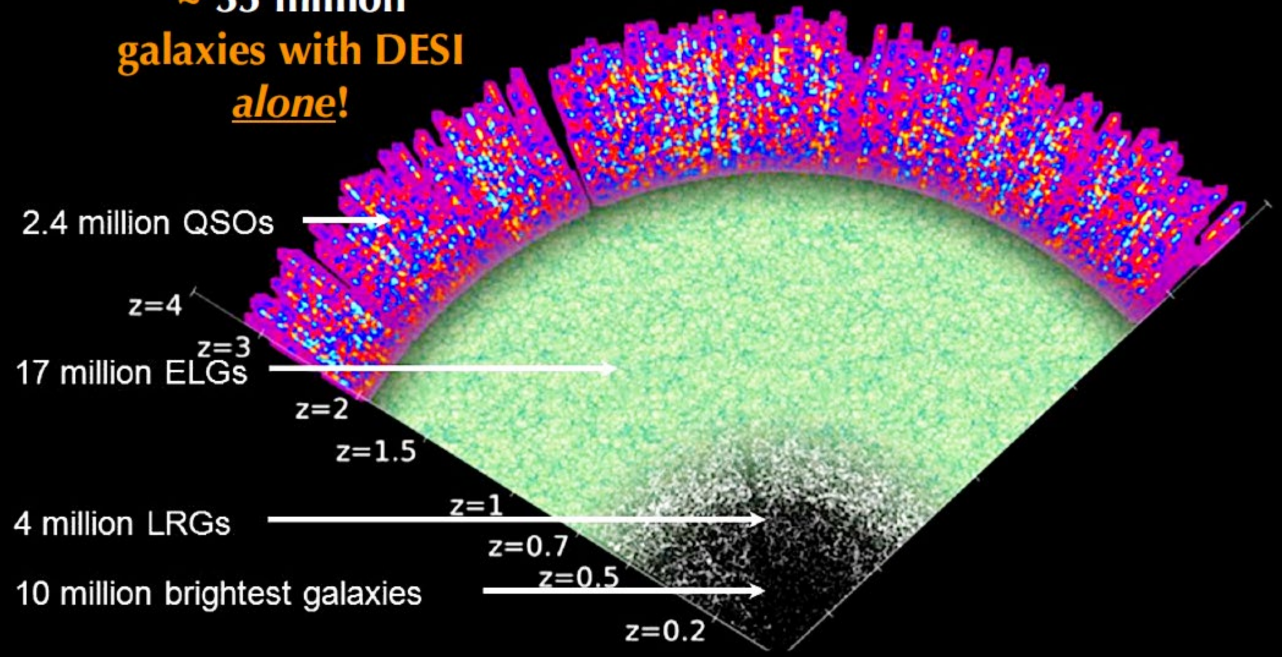
- The two models have matching two-point clustering to the TNG galaxies
- However, noticeable differences appear at higher-order statistics
- Help us differentiate between galaxy-halo models!
- Example: lensing, voids, cumulants, counts-in-cell
- Limited by TNG volume

Hadzhiyska+ (2020b), MNRAS.501.1603H

Emission-line galaxies (ELGs) are understudied

- ELGs: targets of many current and future galaxy surveys (DESI, PFS, *Euclid*)
- Not as well studied as LRGs
- Careful modeling needed to ensure no systematic bias is introduced in the cosmological inference

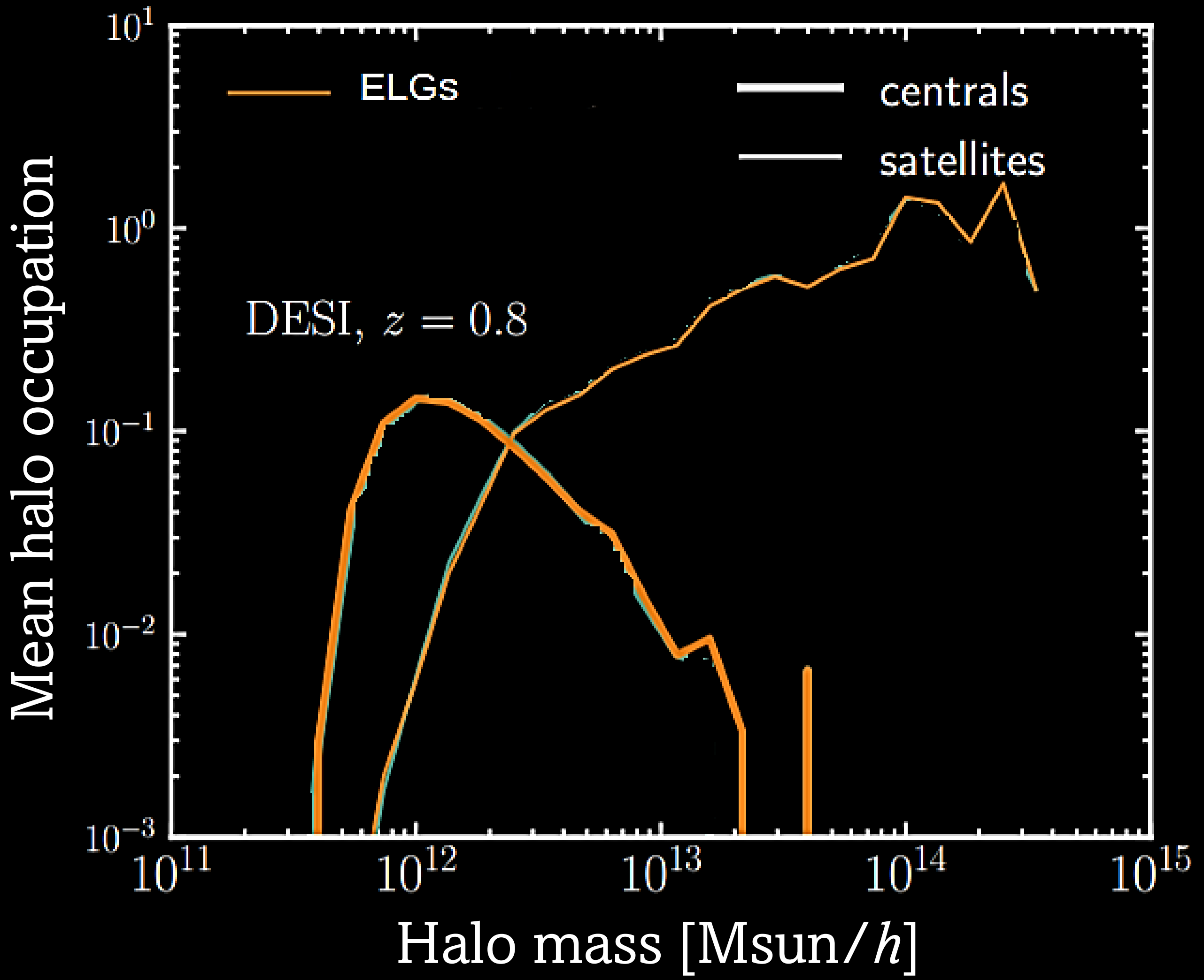
~ 35 million galaxies with DESI *alone*!



+ *Euclid* + LSST @ VRO + *Roman Space Telescope*

Emission-line galaxies (ELGs) behave differently from LRGs

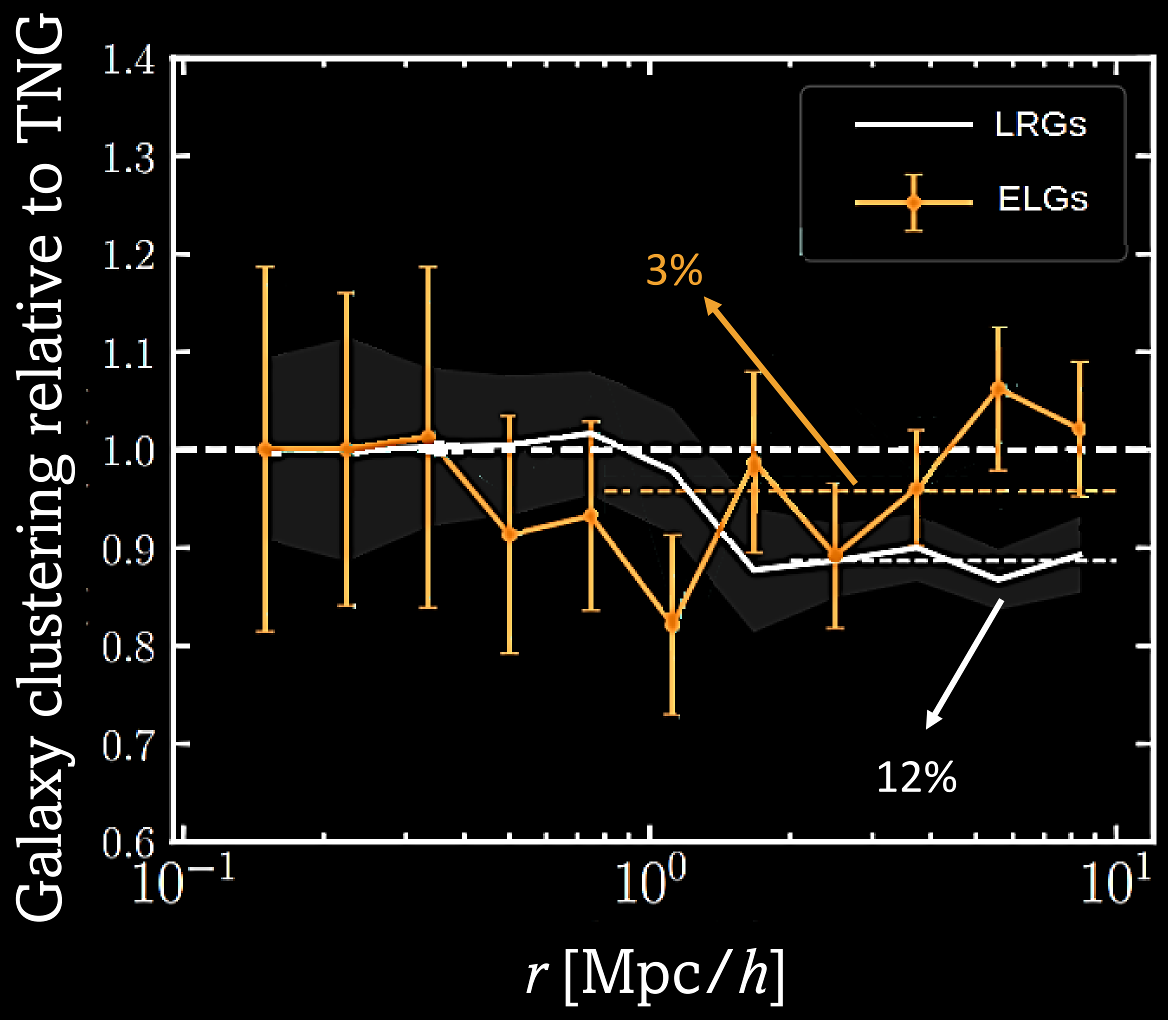
- Created synthetic colors for TNG galaxies at $z \sim 1$
- Extracted ELGs by applying the DESI/eBOSS color cuts
- Halo occupation drastically different from LRGs!
- Need specialized HOD function
- Require higher-resolution N-body simulations



Hadzhiyska+ (2020c), MNRAS.502.3599H

Emission-line galaxies (ELGs) behave differently from LRGs

- ELGs have a much weaker galaxy assembly bias signal (3%) compared with 10% for LRGs ($z \sim 1$)
- Implies surveys targeting ELGs suffer from less systematic effects from assembly bias

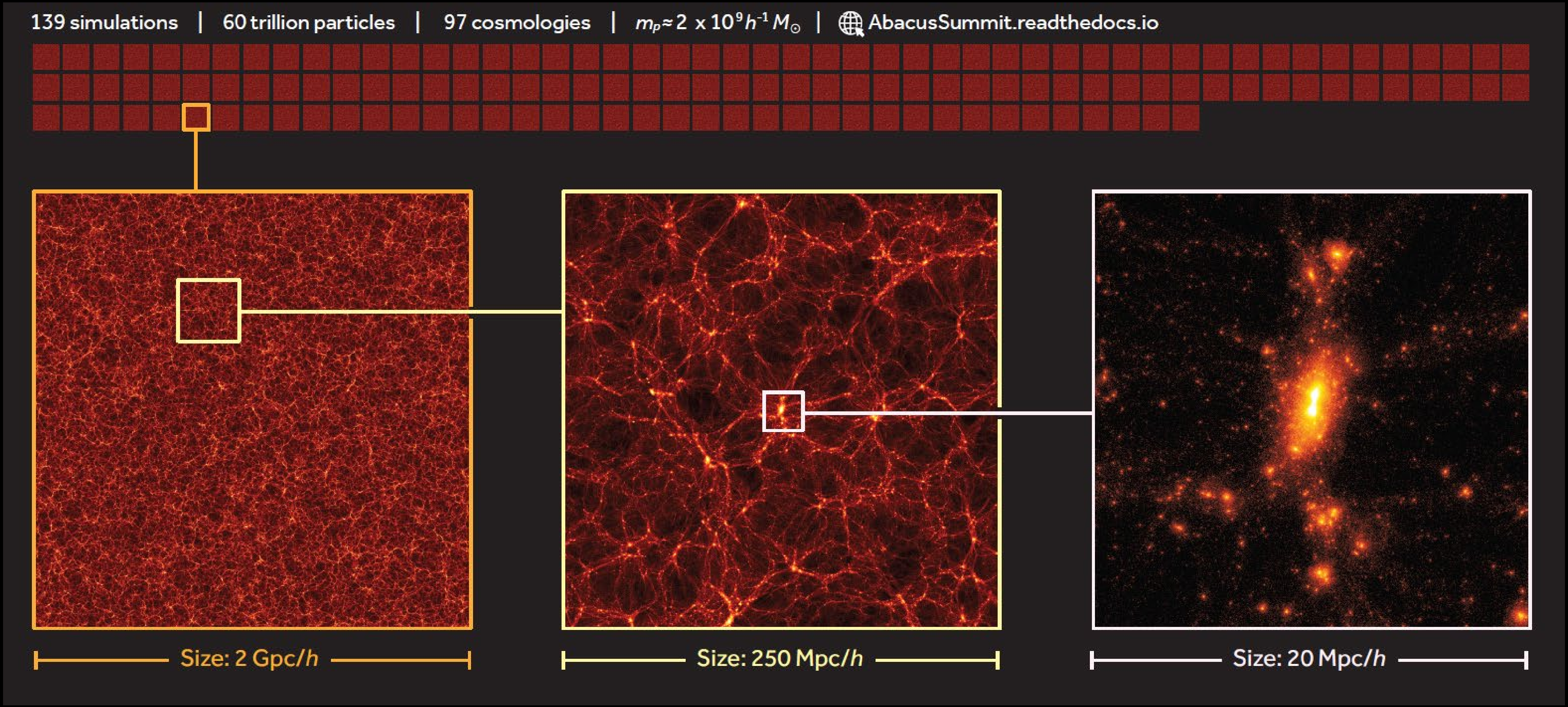


Hadzhiyska+ (2020c), MNRAS.502.3599H

How do we apply that knowledge to the analysis of surveys?

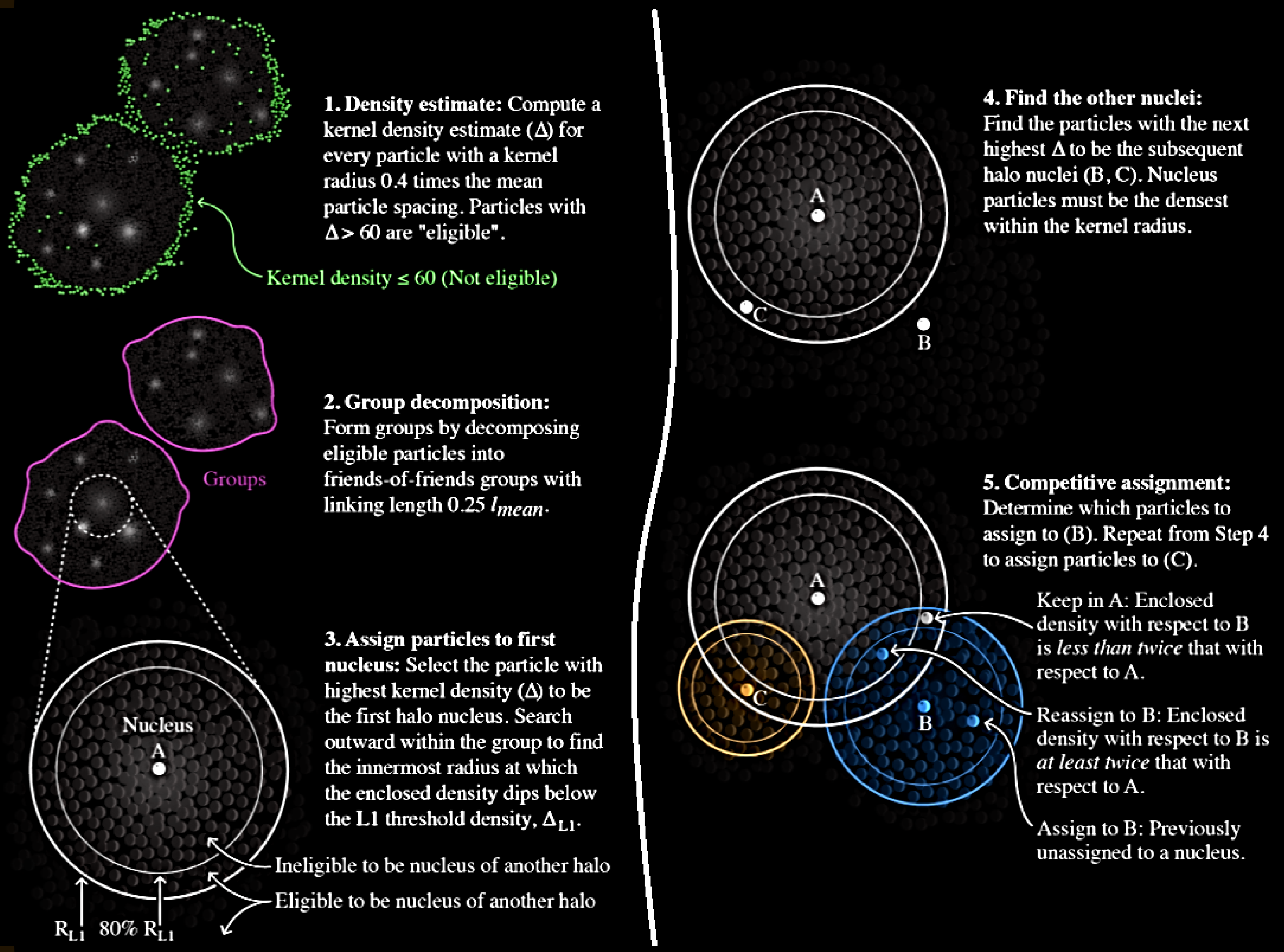
AbacusSummit: largest-yet *N*-body suite

N. Maksimova, L. Garrison, D. Eisenstein, B. Hadzhiyska+, 2021



CompaSO: A new halo finder

- Stands for “competitive assignment to spherical overdensities”
- Highly optimized and efficient for on-the-fly halo finding
- Performs substantially better than other configuration-space finders (is faster and more accurate)
- Comparable to more sophisticated, computationally expensive finders such as ROCKSTAR
- Generate mocks in 0.1 s

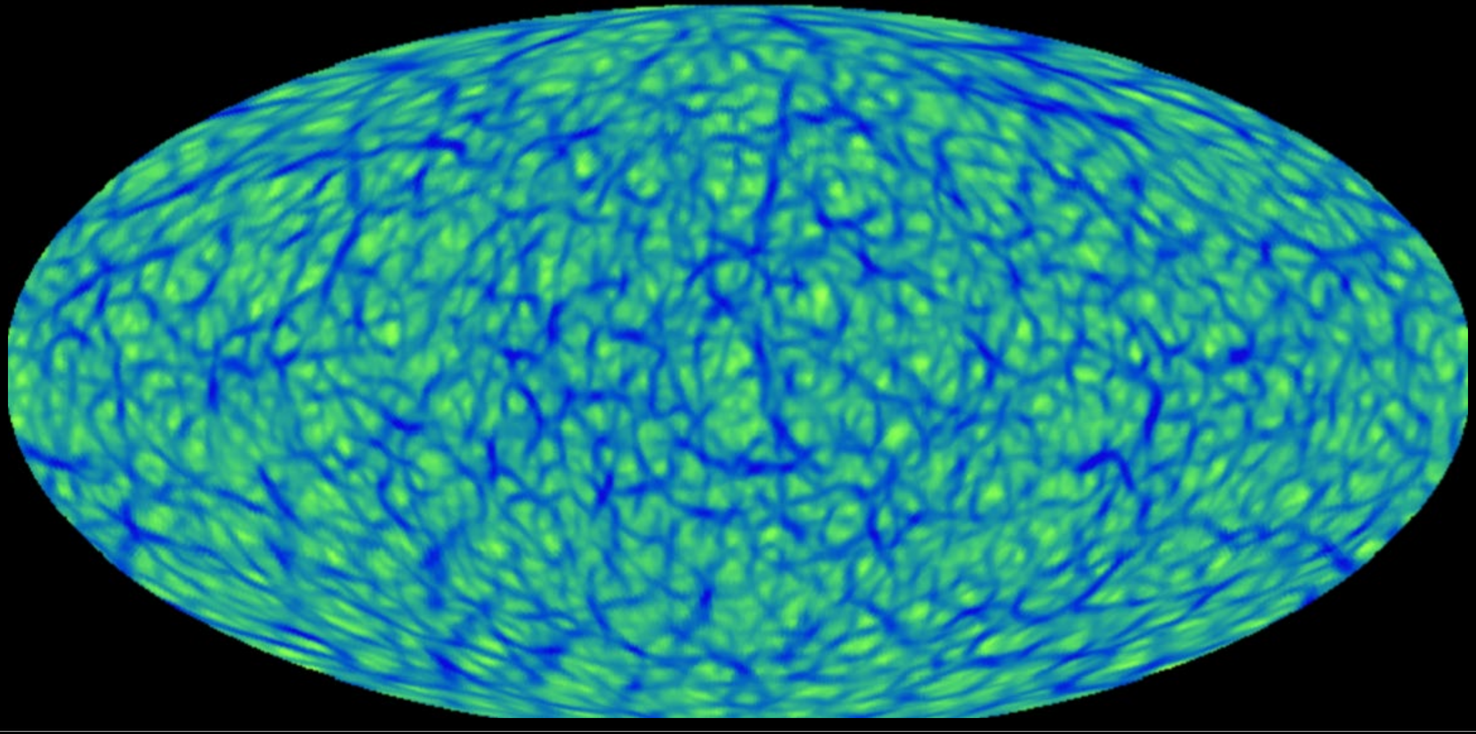


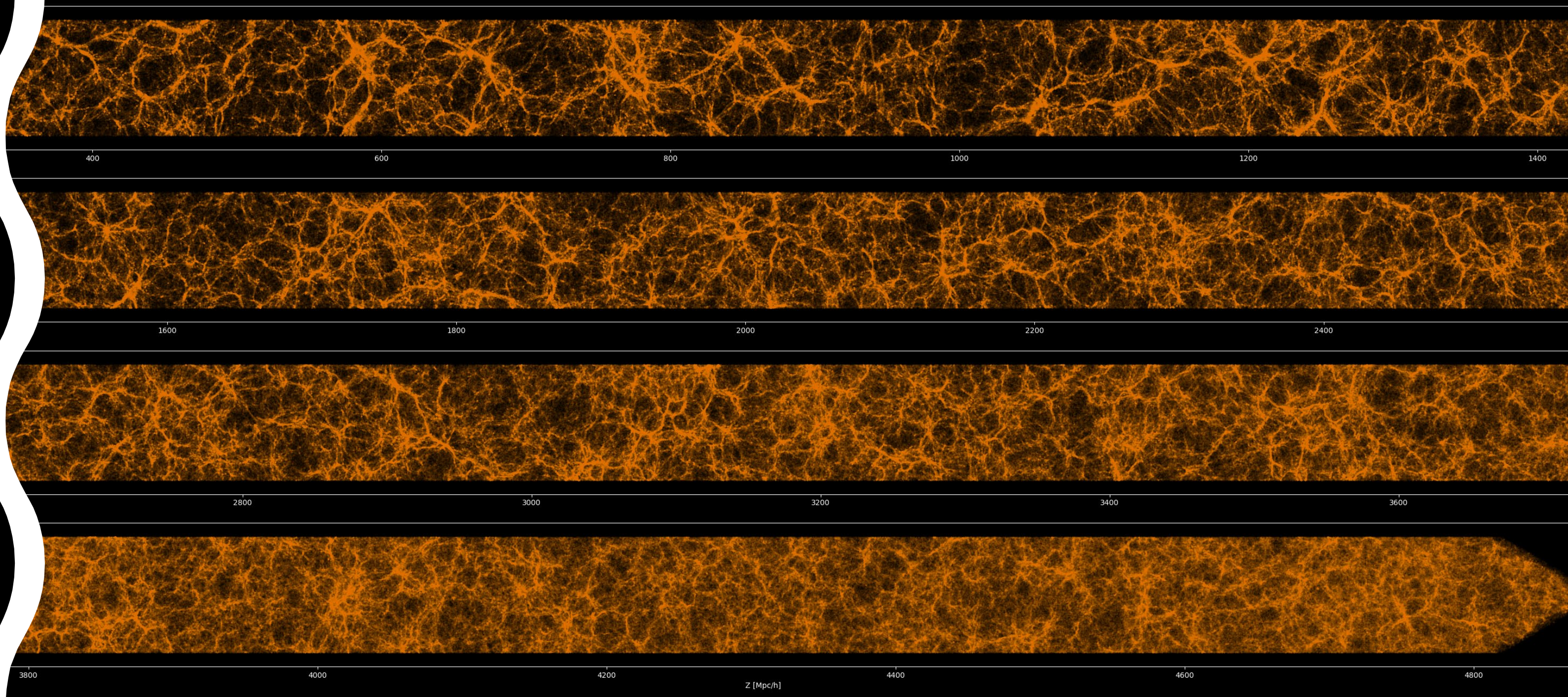
Hadzhiyska+ (2021a, submitted)

The AbacusSummit halo light cone catalogs

- Publicly available!
- 25 simulations cover 2000 sq. deg. until $z \sim 2.4$
- 2 simulations cover the full sky until $z \sim 2.18$
- Can be readily populated with galaxies on the sky
- Produce highly realistic and accurate mock catalogs!

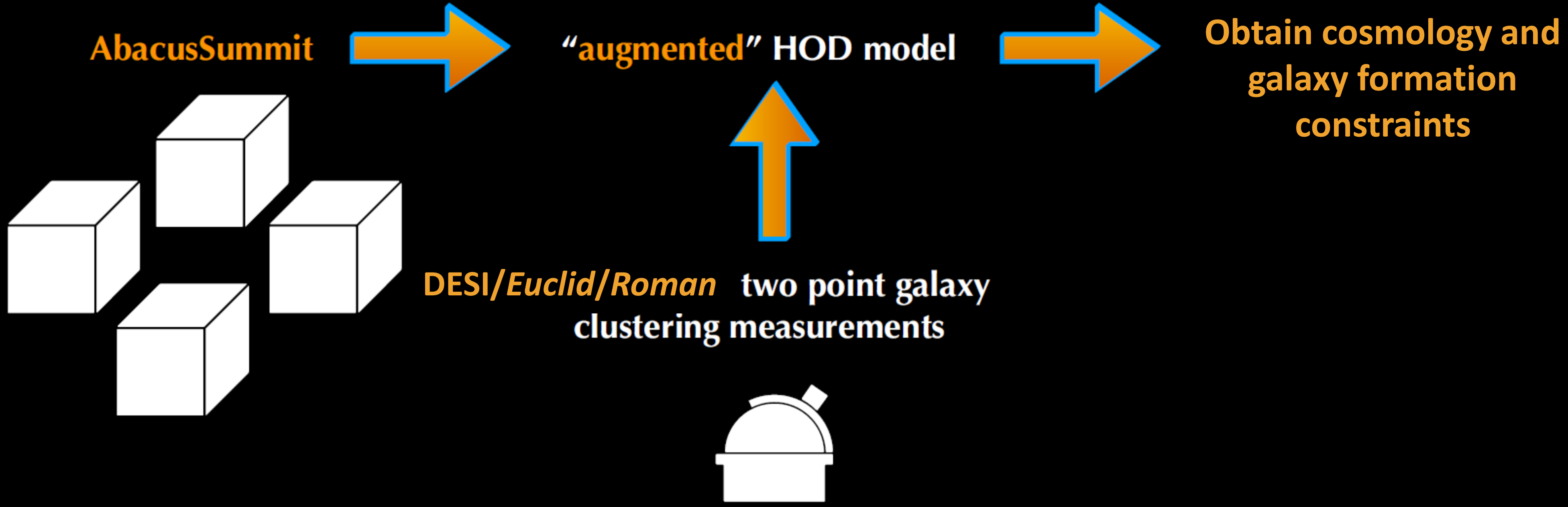
Hadzhiyska+ (2021b, submitted)



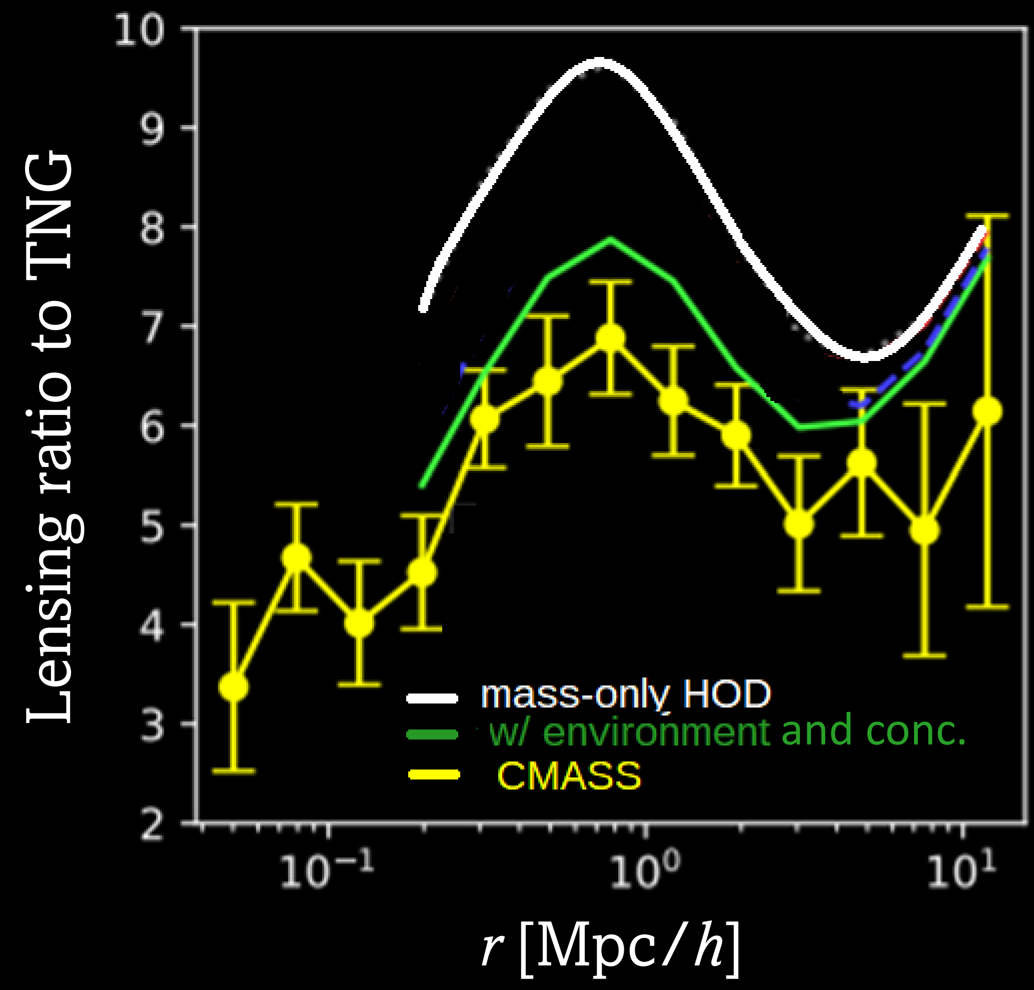


How to analyze observations with accurate models

- Step 0: Study observational effects on halo light cones
- Step 1: Forward model in simulations and compare with observations



Detecting environment assembly bias in CMASS BOSS data



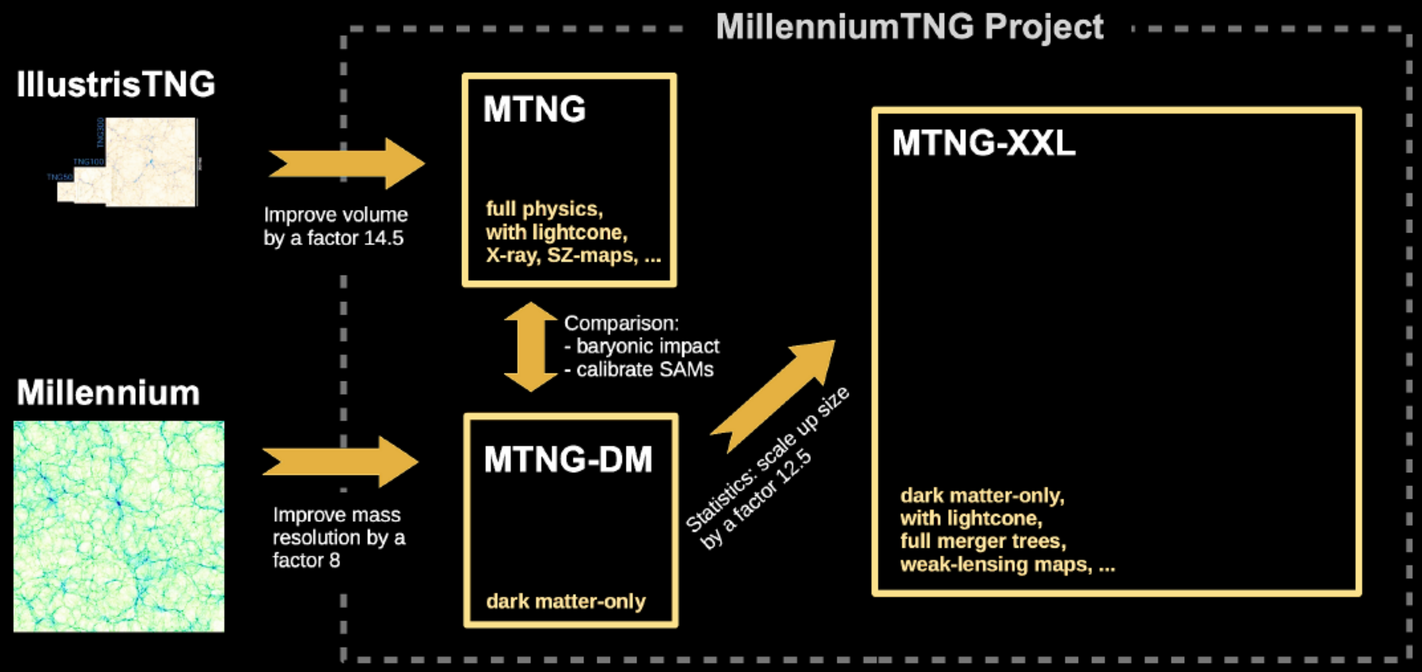
- Simple augmentation of the HOD model with environment and concentration
- Detected positive environment effect with high significance
- Including environment in the analysis reduces the tension by half in the "Lensing is low" tension
- Baryon effects + assembly bias explain it all?

Yuan, Hadzhiyska+ (2020), MNRAS.502.3582Y
Amodeo & ACT Collab. (2020)


Looking forward

MTNG: largest-yet hydro simulation

An effort led by:
Volker Springel, Lars Hernquist, Carlos Frenk, Simon White,
Ruediger Pakmor, Boryana Hadzhiyska, Sownak Bose,



- 15 x volume of TNG
- Better large-scale statistics
- Can study 3-point correlations, void statistics, counts-in-cell
- Various tracers (LRGs, ELGs, X-ray, SZ, CMB lensing)

MTNG: Ongoing and future projects

1. (Ongoing) Is tertiary assembly bias necessary for predicting the large-scale galaxy distn?
2. (Ongoing) How sensitive are void statistics to baryonic and assembly bias effects?
3. (Ongoing) What are the baryonic effects from Sunyaev-Zel'dovich (SZ) maps
4. (Ongoing) What are the intrinsic alignments of galaxies?
5. How to populate larger *N*-body simulations (using e.g., machine learning, SAM)?
6. How do we generalize our conclusions beyond the particulars of TNG physics?

Surveys: Ongoing and future projects

1. (Ongoing) Cross-correlation between DESI ELGs and CMB lensing (*Planck*)
2. (Ongoing) Multi-tracer analysis (LRGs, ELGs, QSOs) of DESI Y1 data
3. (Ongoing) Constraining cosmology from photometric surveys, BAO, CMB, with Hybrid Effective Field Theory (HEFT)
4. Cross-correlation between DESI/*Euclid*/*Roman* and CMB lensing/SZ
5. Evolution of assembly bias effects over time → learn galaxy physics

Backup slides

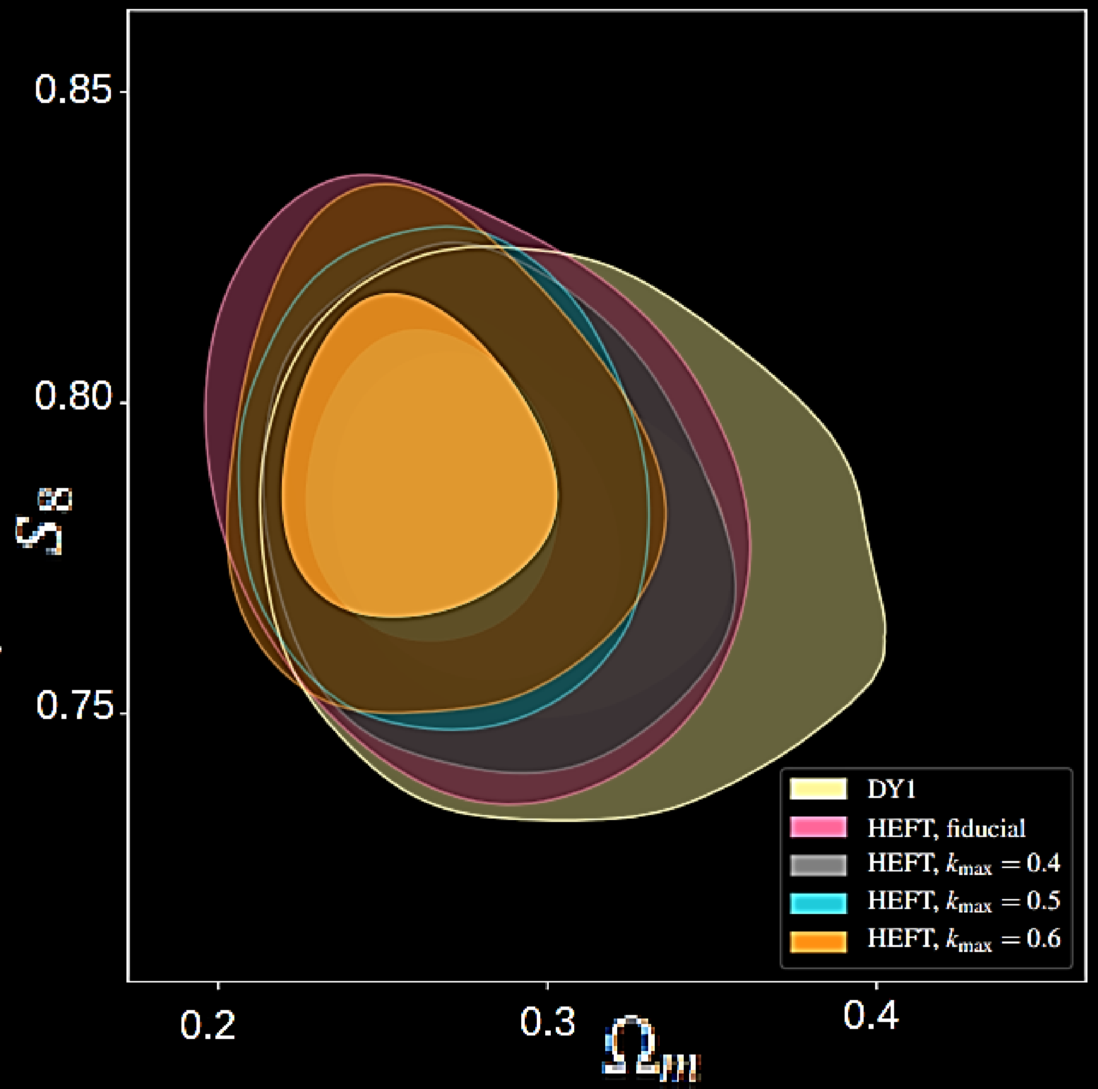
Hybrid Effective Field Theory

- Expansion to second order in Lagrangian Perturbation Theory
- Advection from Lagrangian to Eulerian space done numerically through sims
- Computing 15 basis spectra to fit galaxy power spectrum

$$1 + \Delta_{g,L} = 1 + b_1\delta_L + b_2(\delta_L^2 - \langle\delta_L^2\rangle) + b_s(s_L^2 - \langle s_L^2\rangle) + b_\nabla \nabla^2\delta_L$$

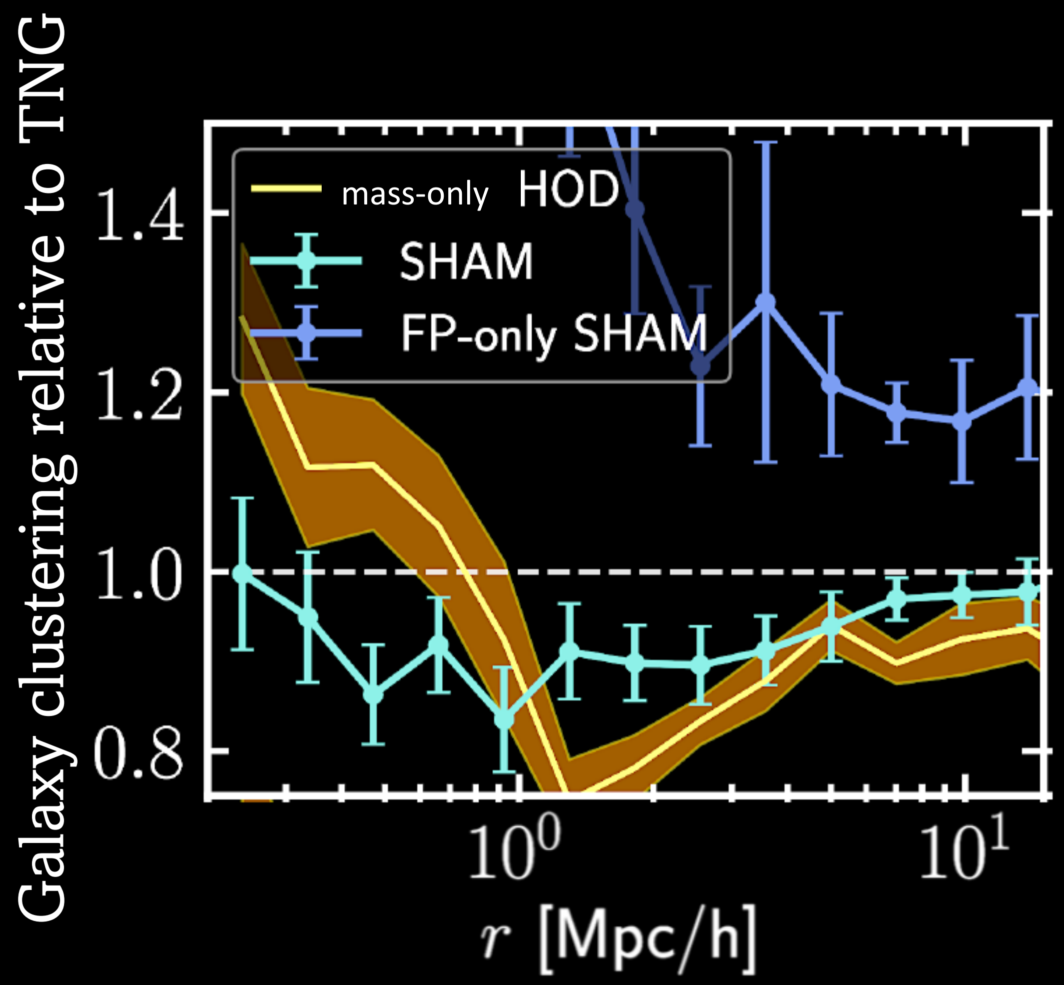
$$1 + \Delta_g(\mathbf{x}) = \int d^3\mathbf{q}\,[1 + \Delta_{g,L}(\mathbf{q})]\,\delta^D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q}))$$

$$P_{gm}(k) = \sum_{\alpha\in\mathcal{O}} b_\alpha P_{1\alpha}(k), \quad P_{gg}(k) = \sum_{\alpha\in\mathcal{O}}\sum_{\beta\in\mathcal{O}} b_\alpha b_\beta P_{\alpha\beta}(k)$$



Modi+ (2020), Hadzhiyska+ (2021a), JCAP...09..020H

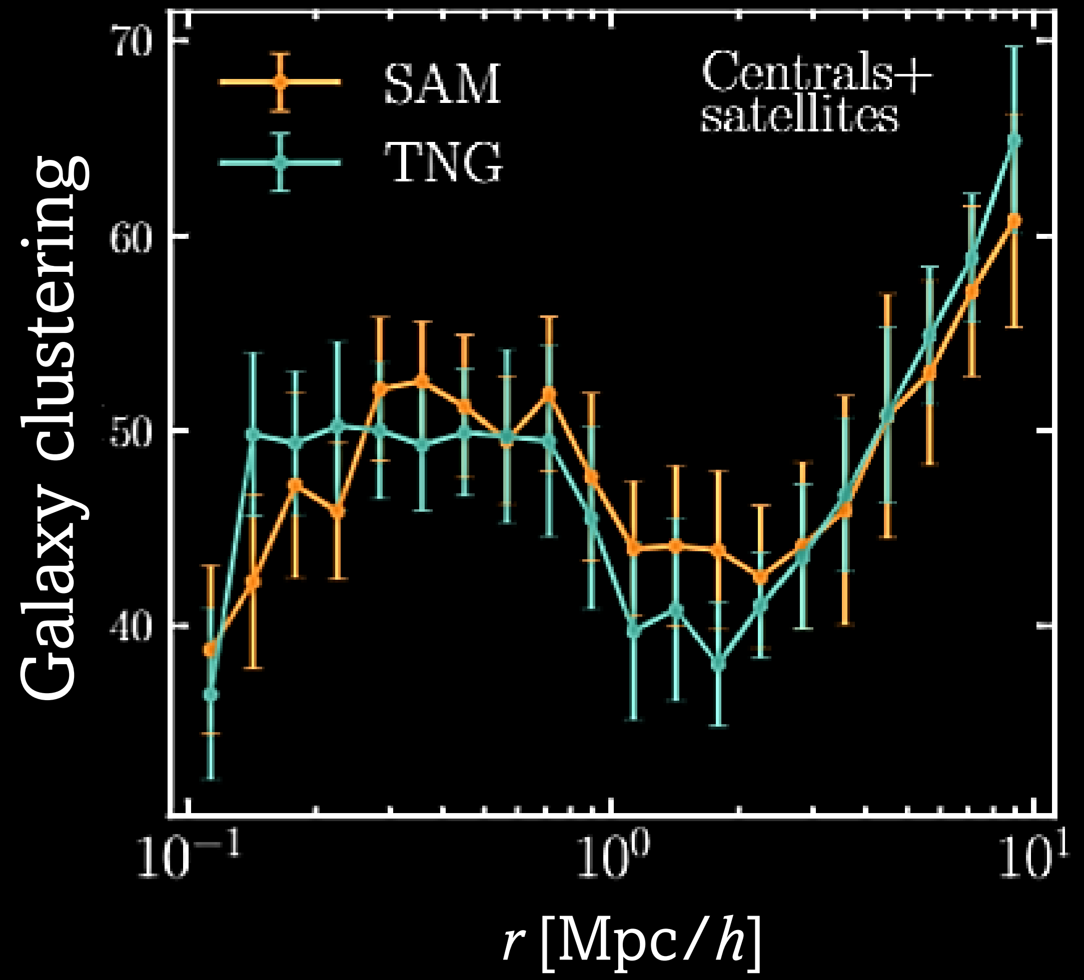
Other empirical methods fail, too!



Hadzhiyska+ (2020b), MNRAS.501.1603H

- Subhalo abundance matching (SHAM): "paints" galaxies onto subhalos after rank-ordering them by a dark-matter property
- Fails as well at > 5% level, requires subhalos, unclear how to treat different galaxy tracers
- More complex models can reproduce the TNG clustering such as HOD models with assembly bias, semi-analytic models (see Hadzhiyska,… Somerville+ 2021)

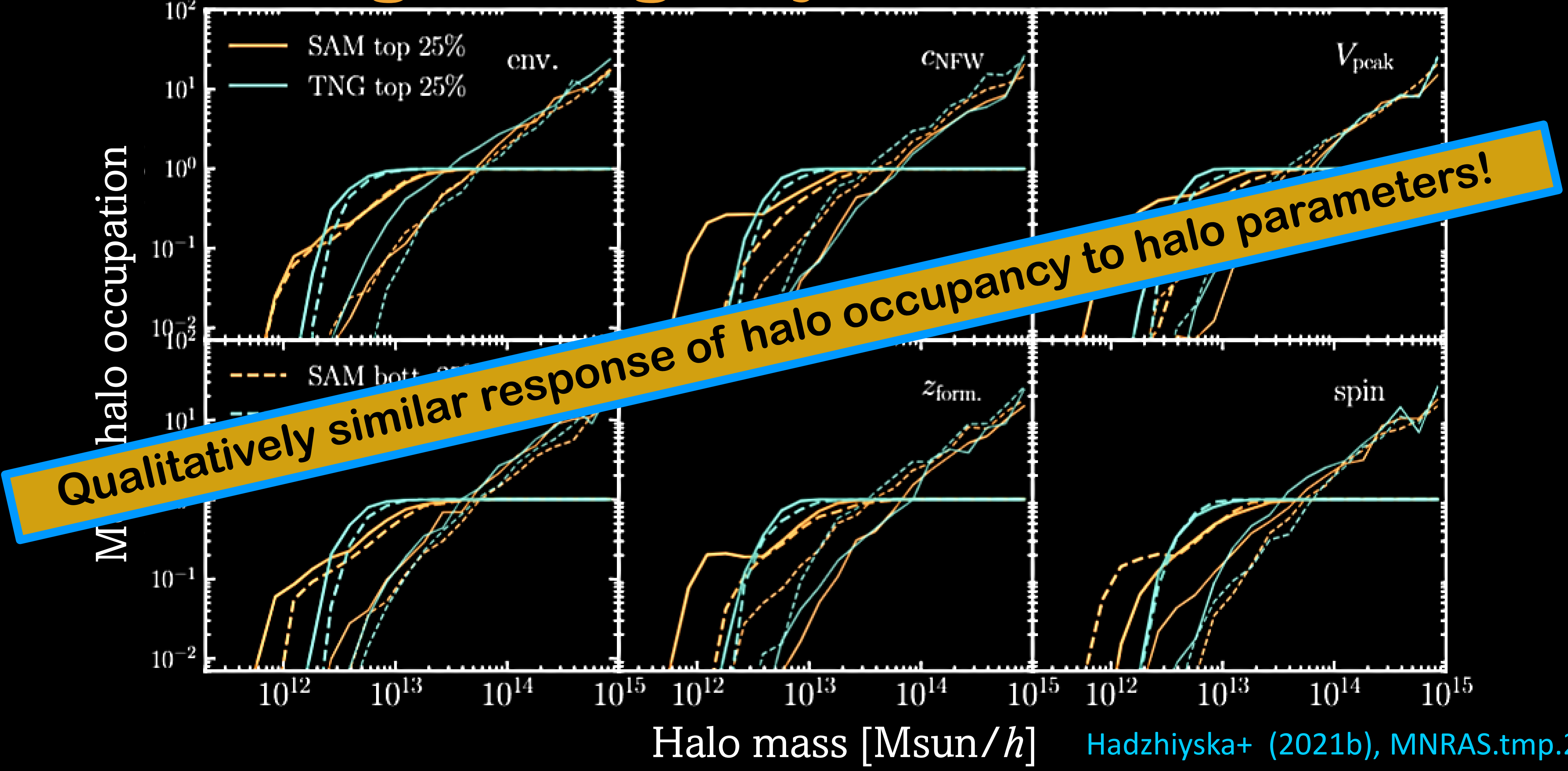
Physically intuitive but inexpensive methods go a long way



- Comparison between R. Somerville's SAM and TNG
- Despite not matched to each other, two-point clustering is well-matched b/n the two
- Suggests clustering can be recovered in cheaper ways
- Analysis needs to be repeated for other tracers and higher redshifts

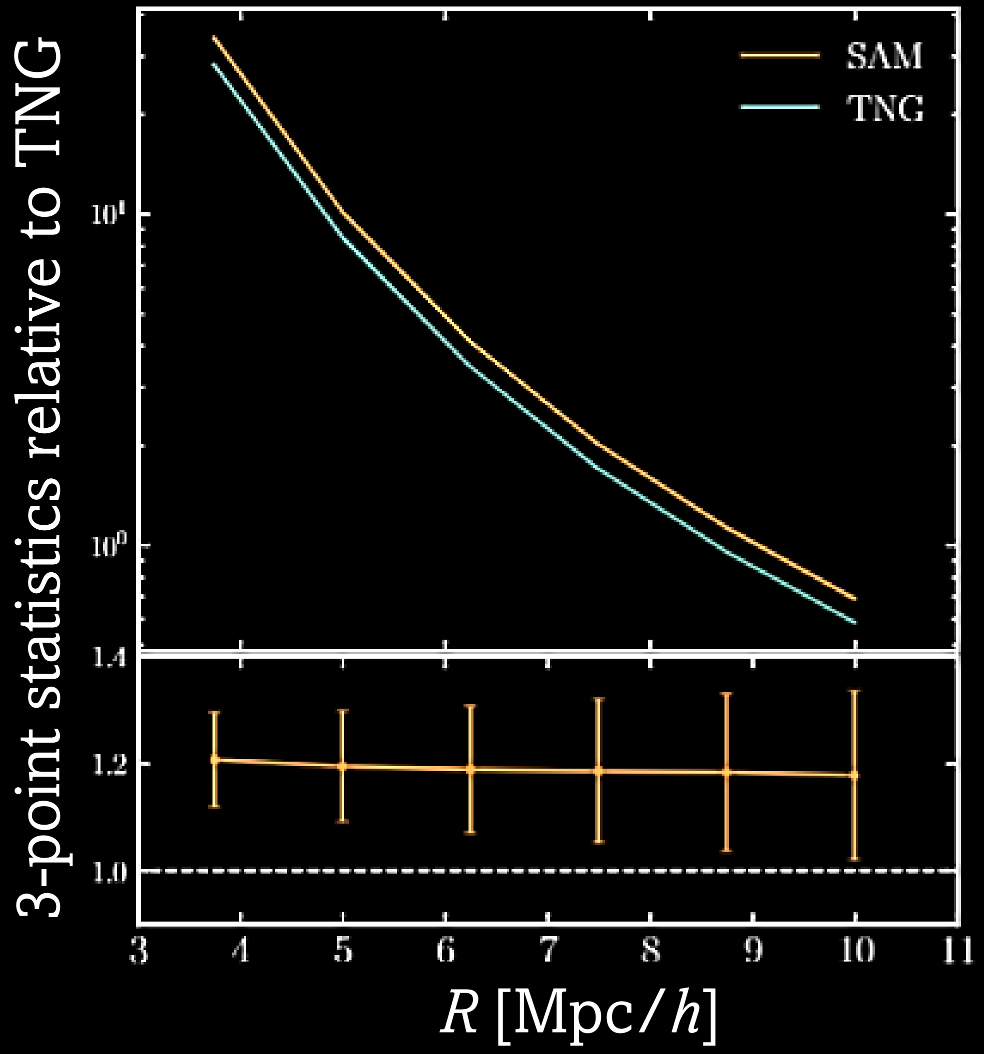
Hadzhiyska+ (2021b), MNRAS.tmp.2334H

Physically intuitive but inexpensive methods go a long way



Hadzhiyska+ (2021b), MNRAS.tmp.2334H

Physically intuitive but inexpensive methods go a long way

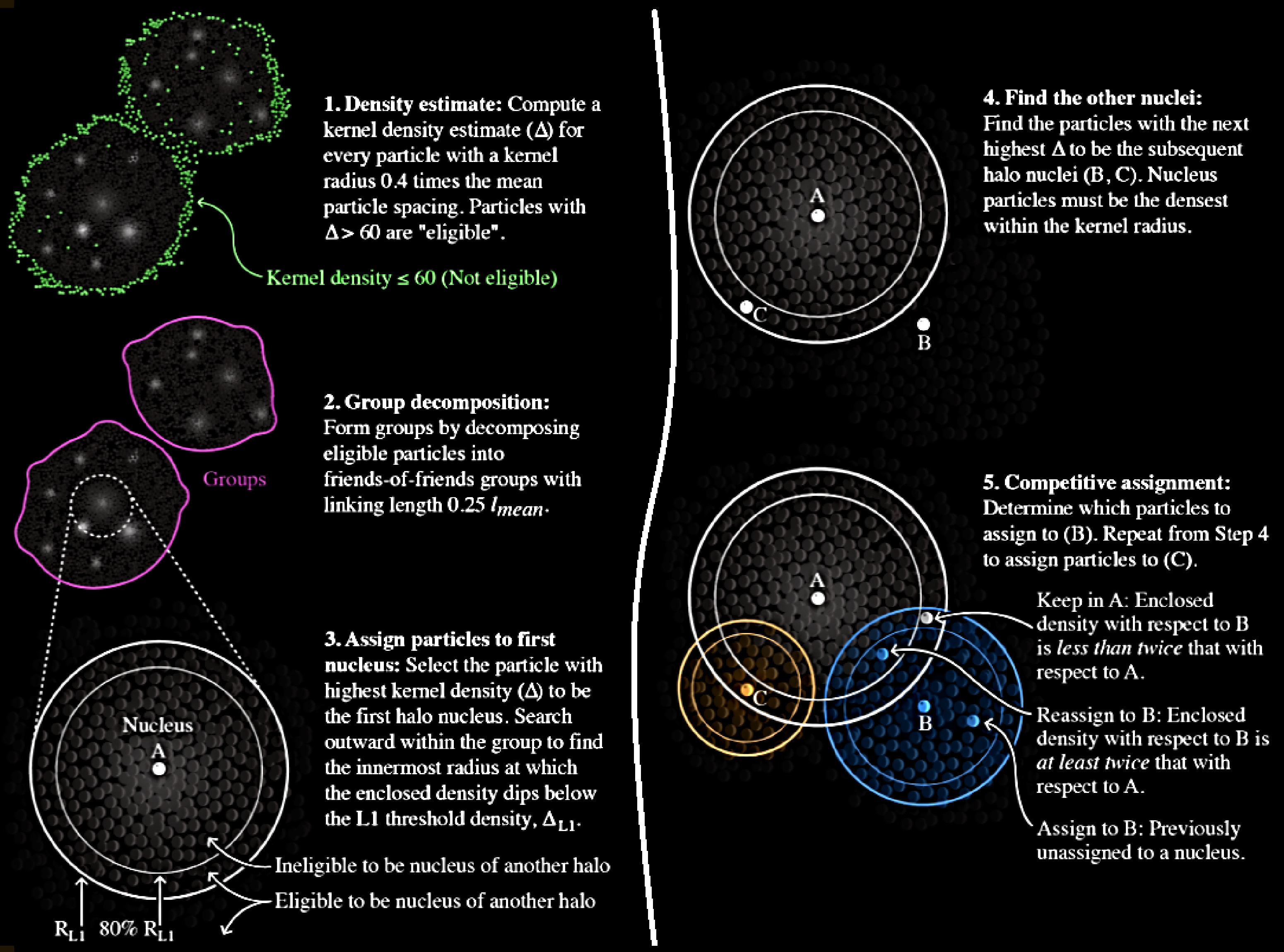


- Despite having well-matched two-point clustering, SAM and TNG display different higher-order statistics
- Suggests including higher-order statistics in the calibration of SAMs
- Including more observables for calibration (e.g., cross-correlations with early Universe probes, alternative stats, wide range of redshifts)

Hadzhiyska+ (2021b), MNRAS.tmp.2334H

CompaSO: A new halo finder

- Stands for "competitive assignment to spherical overdensity"
- Highly optimized and efficient for on-the-fly halo finding
- Faster and more accurate than other position-based finders
- Comparable to more computationally expensive finders (ROCKSTAR)



Hadzhiyska+ (2021a, in prep.)