



# Galaxy Clustering With Fewer Galaxies

---

Oscar Hickman, Carlton Baugh, Baojiu Li, Sownak Bose, Kai Wang

Durham ICC, 08/05/2026

*Euclid Deep Field South 2025*

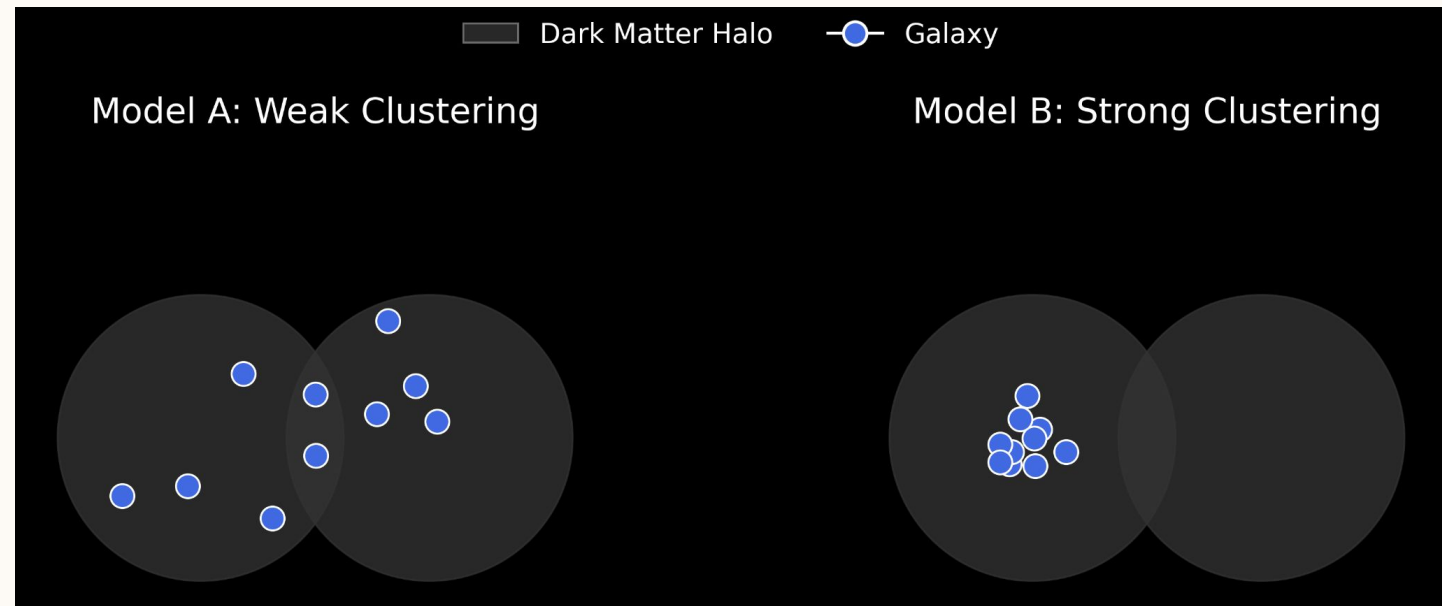
# The Power of Galaxy Clustering

Luminosity functions tell us how many galaxies exist.

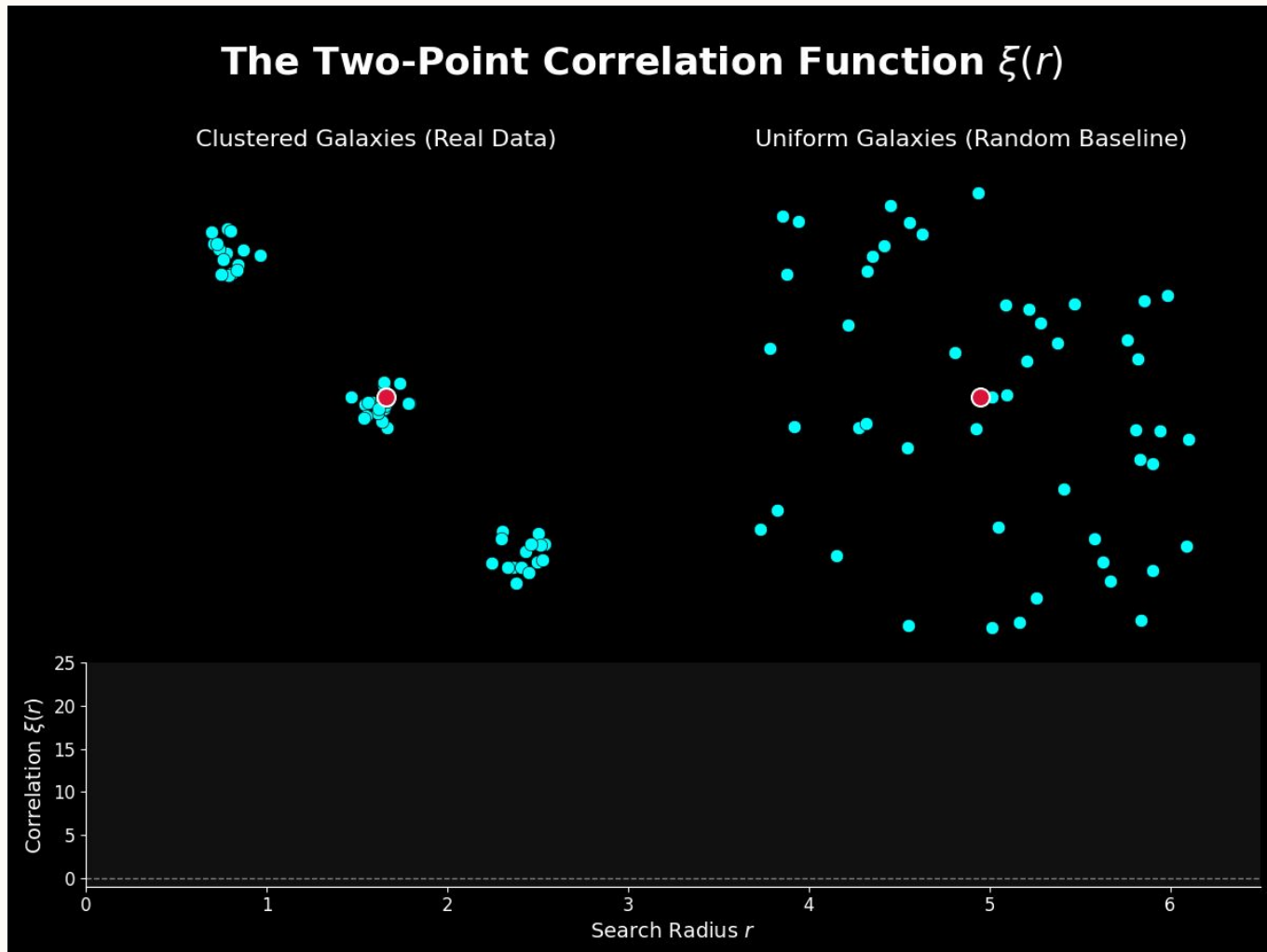
Spatial clustering tells us where those galaxies live.

Different physical models (e.g. varying dynamical friction or feedback) can produce the exact same luminosity function.

Clustering helps break this degeneracy.



# The Two Point Correlation Function



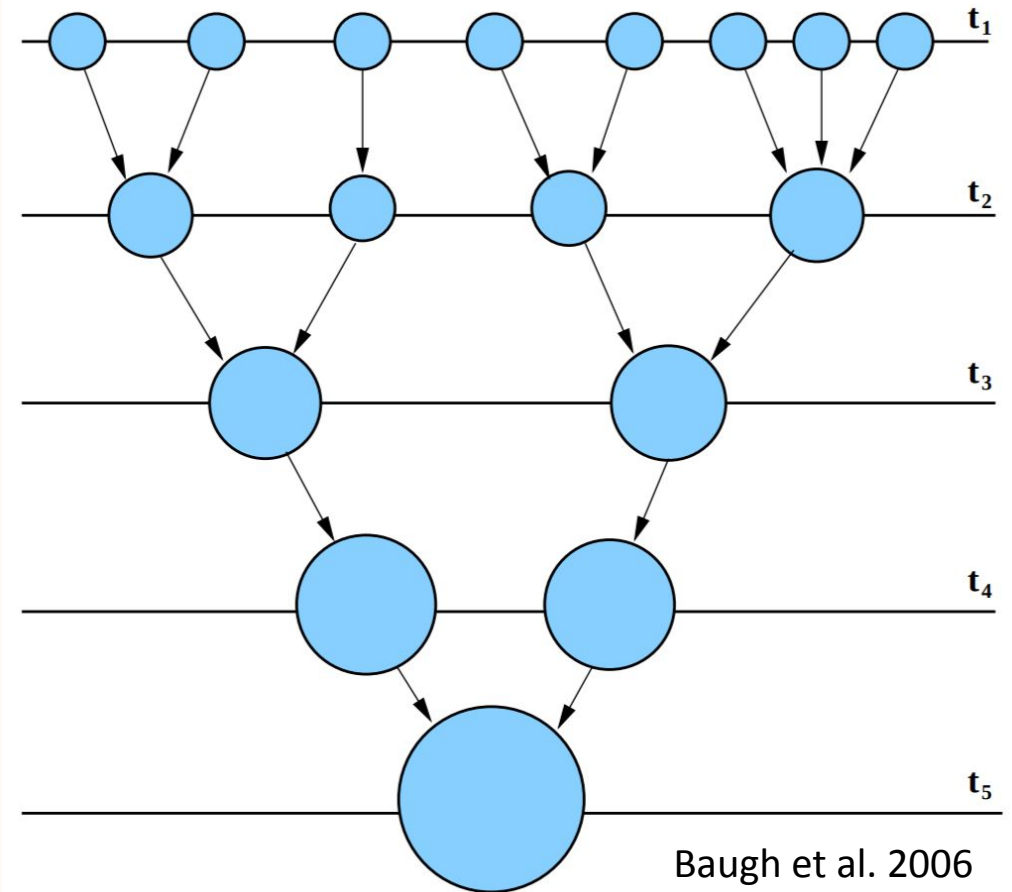
It measures the excess probability of finding a pair of galaxies separated by a distance  $r$ , compared to a completely random universe.

# Semi Analytic Modelling

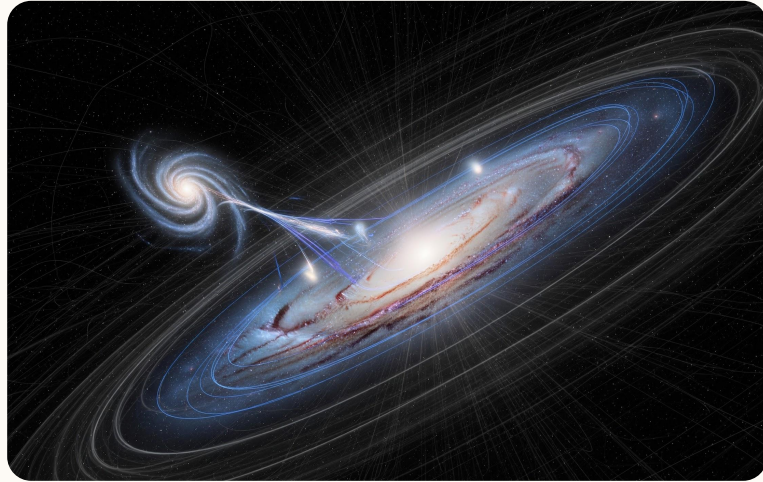
**Empirical Models:** Fast, but use statistical shortcuts instead of physical laws.

**Hydrodynamical Simulations:** are highly physical, but computationally expensive for large survey volumes.

**Semi Analytic Models:** Populates dark matter merger trees with baryons by solving differential equations for gas cooling, star formation, feedback etc. across time (traces the halo tree history).

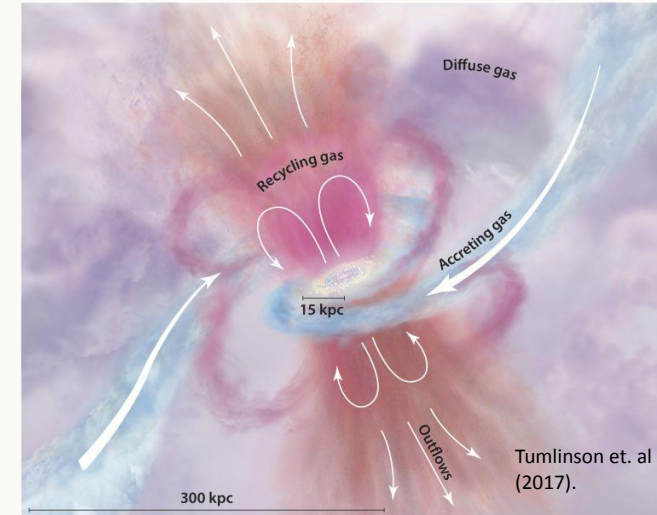


# GALFORM Physics: Shaping Galaxy Clustering



## Dynamical Friction:

- Drag force on orbiting satellites causes them to lose momentum and spiral inward.
- Stronger friction = faster merging, which removes galaxies, lowering clustering amplitude; slower merging boosts the signal.



## Feedback:

- Supernova feedback: suppresses star formation in low-mass halos, preventing dwarf galaxies from diluting the clustering signal.
- AGN Feedback: Quenches massive halos

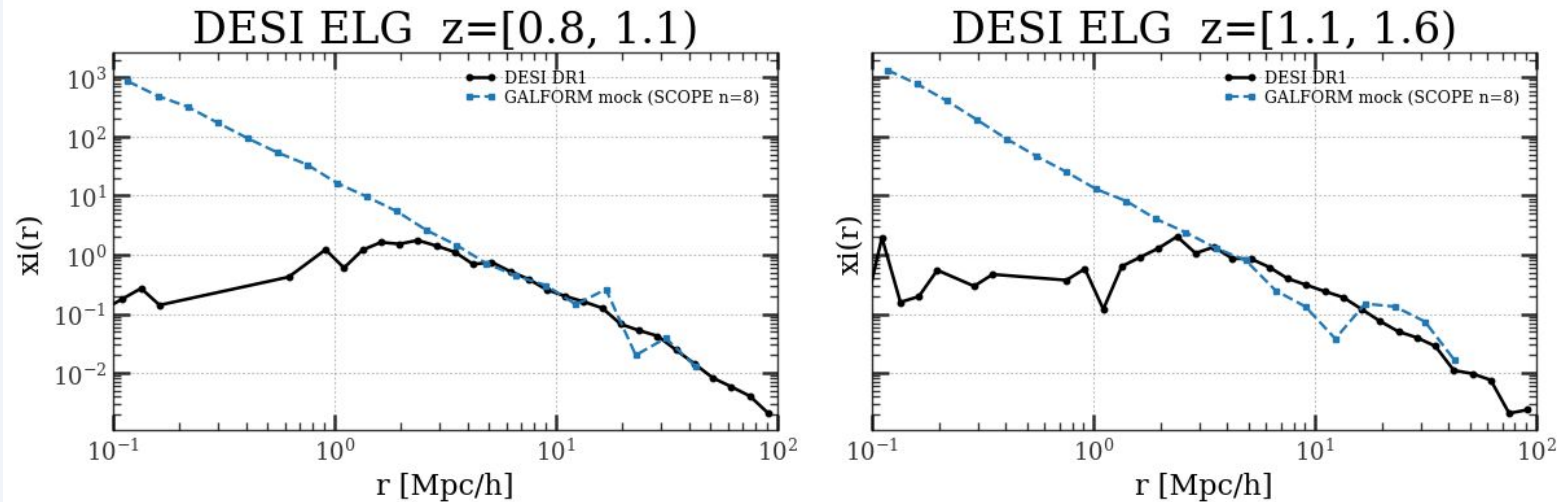
# Dark Energy Spectroscopic Instrument (DESI)

- Constructing the largest, most precise 3D map of the universe to date by measuring the optical spectra of over 40 million galaxies and quasars
- Spectrograph on the 4-meter Mayall Telescope.
- 5,000 optical fibers capturing simultaneous spectra - fiber technology developed right here at Durham University!



# DESI DR1 2PCF

DESI DR1 ELG vs GALFORM mock (SCOPE, n=8)



**Mass Cut,  $M_* > 10^{9.5} M_\odot$ :**  
Removes unobservable dwarfs.

**sSFR Cut,  $> 10^{-10.5} \text{ yr}^{-1}$ :**  
Isolates actively star-forming galaxies (the source of emission lines).

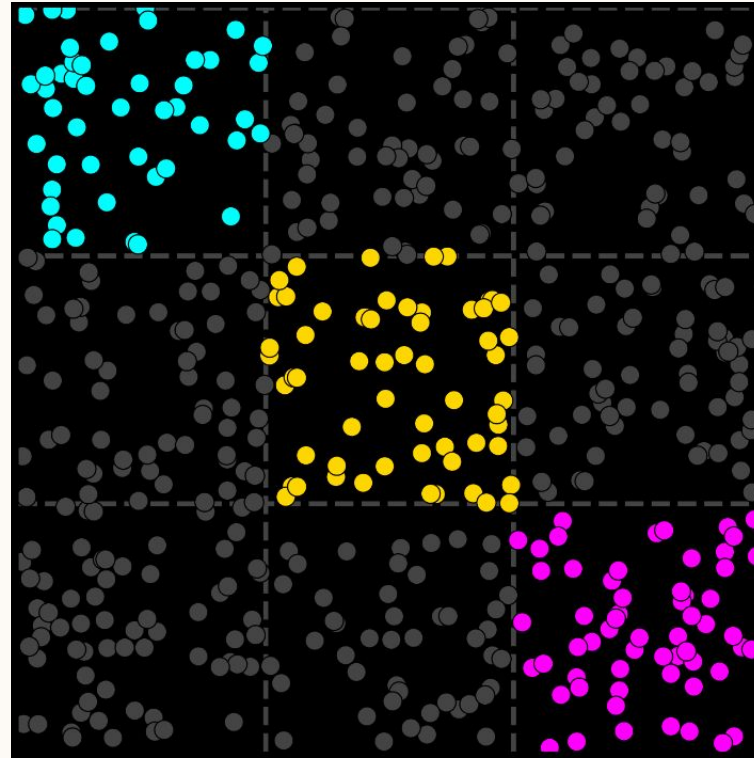
**Rest-frame Color,  $g - r \leq 0.7$ :** Matches the optical "blue cloud" selection of the DESI survey.

# ~~Sub-Volume~~ Independent Realisation Architecture

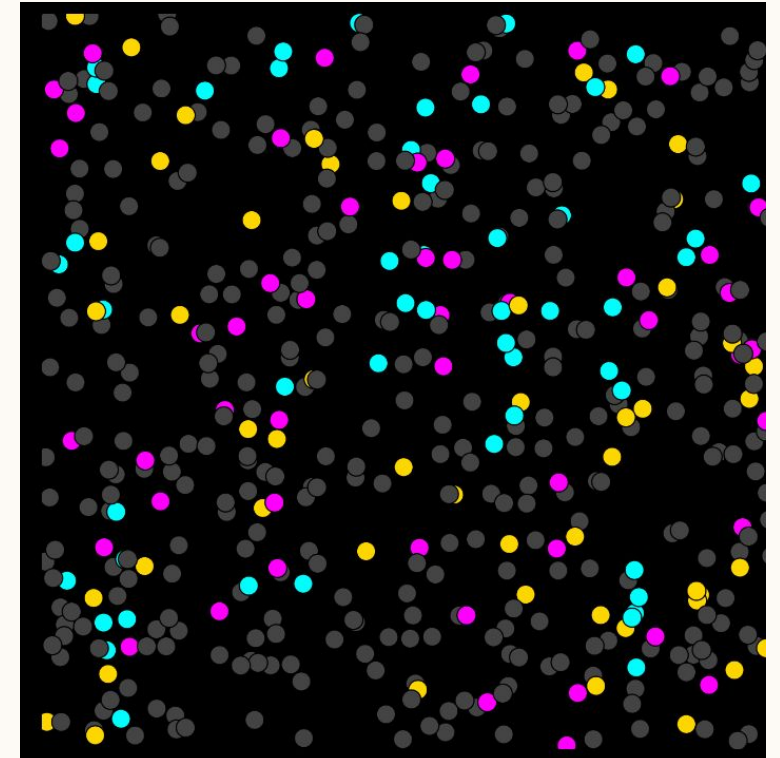
To make the simulation computationally tractable, the halos are partitioned into  $k$  sub-volumes

These are interlaced halo groups, not geometric chunks. Galaxies from sub-volume 1 are physically scattered throughout the entire simulation box

spatial subvolume



independent realisation



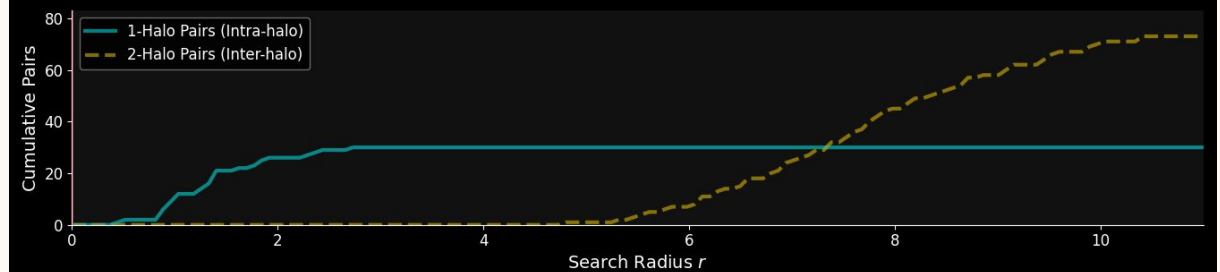
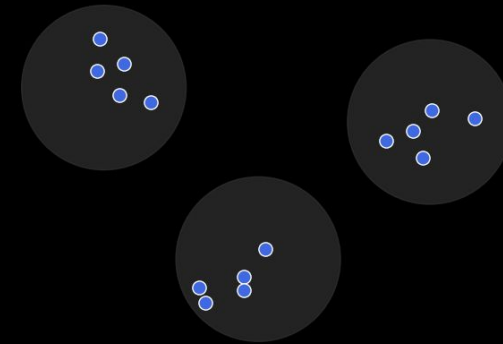
# One and Two Halo Terms

**1-halo term:** Pairs of galaxies that live within the same dark matter halo. This probes satellite survival and galaxy merger timescales

**2-halo term:** Pairs of galaxies across different halos. This probes the large-scale linear bias and the underlying dark matter web.

## Mapping Large-Scale Structure (1 vs 2-Halo Terms)

Physical Space (Larger Halos & Internal Separations)



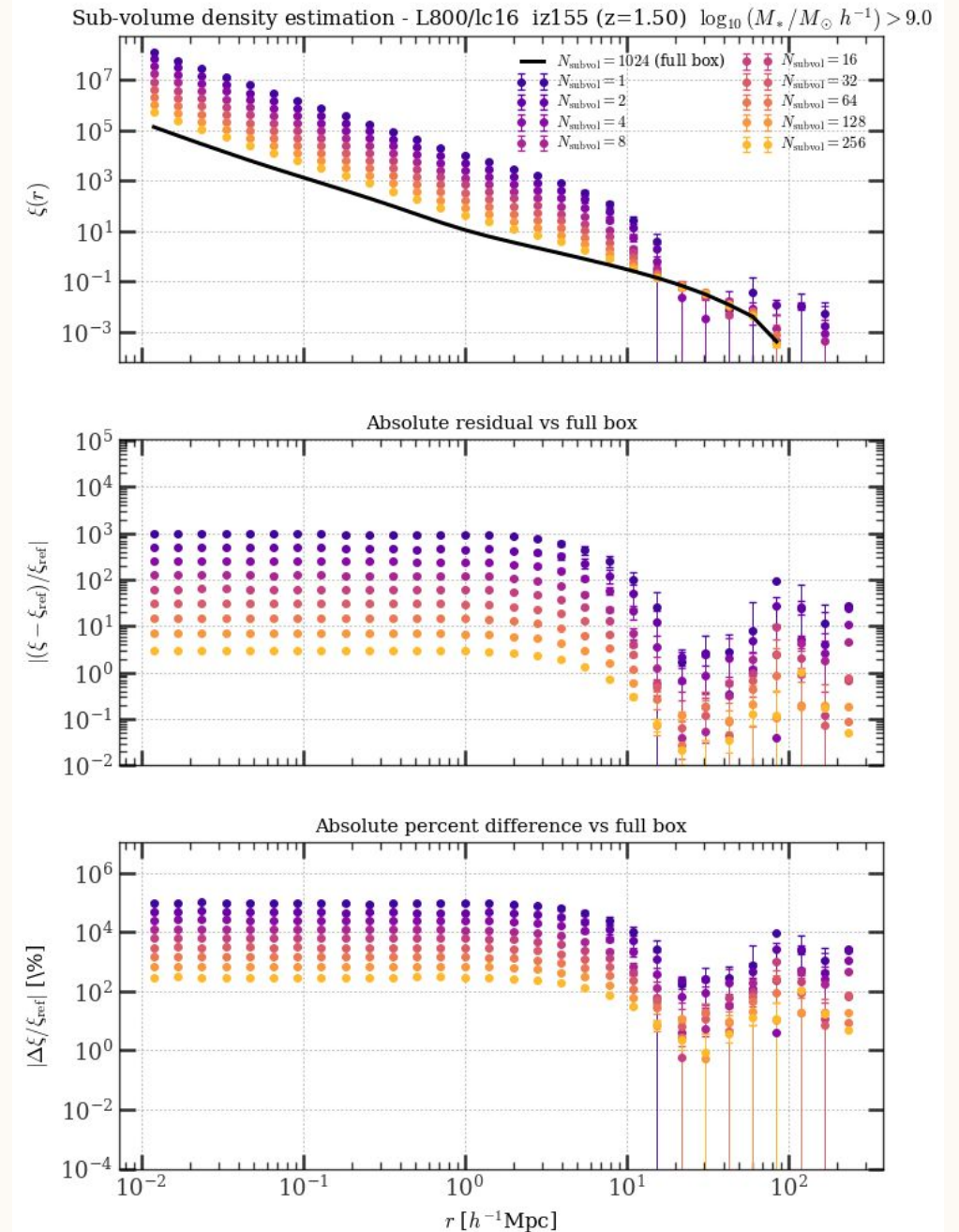
# The estimating problem

The small-scale (1-halo) signal is artificially inflated:

1-halo pairs scale as  $= \frac{m}{k}$

2-halo pairs scale as  $= \frac{m(m-1)}{k(k-1)} \approx \left(\frac{m}{k}\right)^2$

Scale differently, stacking them breaks the estimation



# SCOPE correction

Instead of dumping all galaxies into one pool, we preserve their sub-volume labels

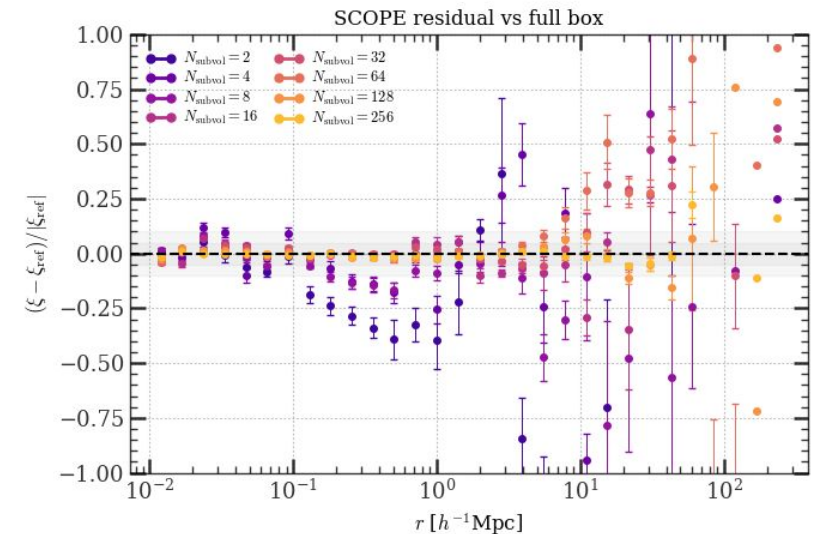
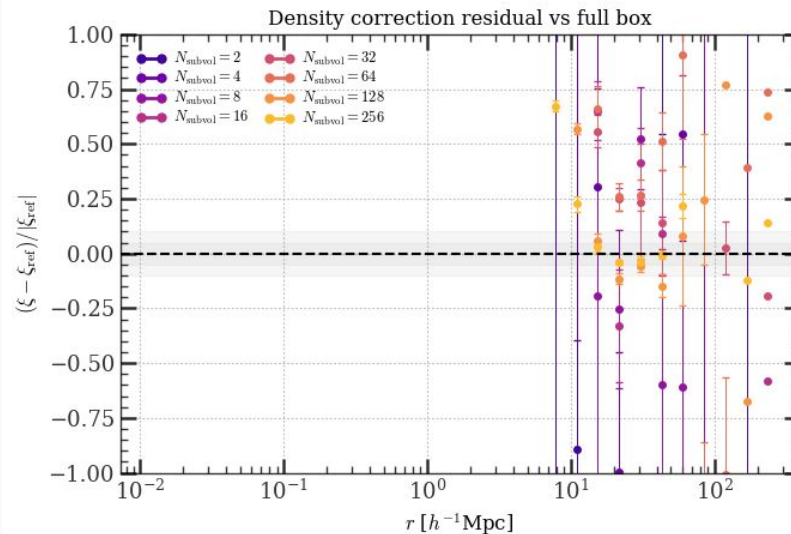
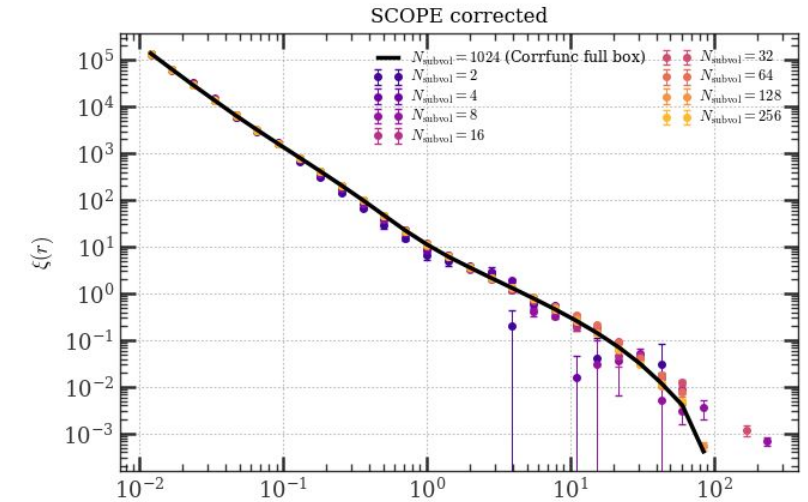
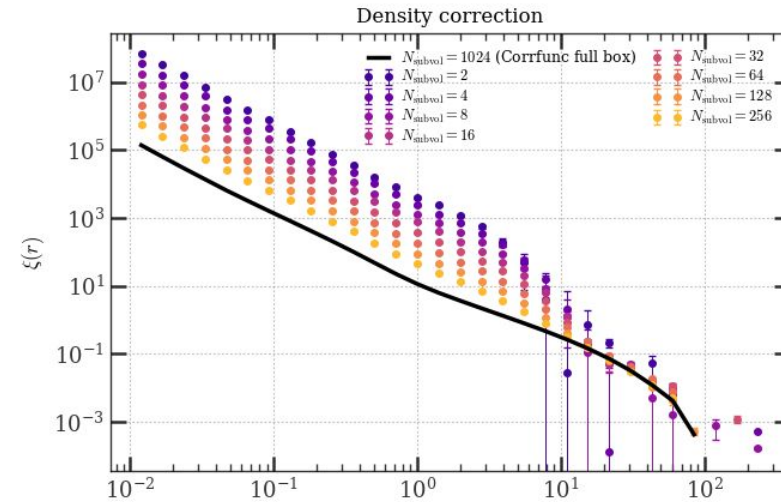
Auto-Pairs (1-halo): Multiplied by exactly

$$\alpha = \frac{k}{m}$$

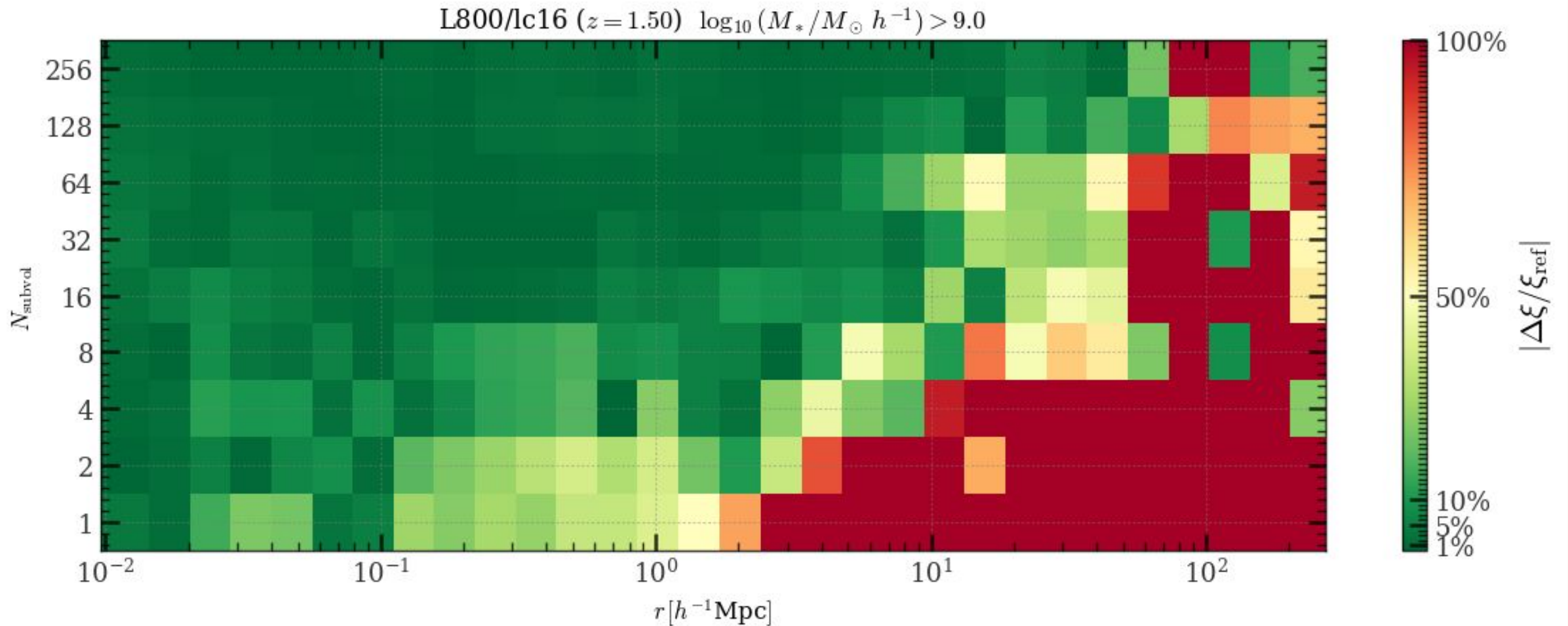
Cross-Pairs (2-halo): Multiplied by exactly:

$$\beta = \frac{k(k-1)}{m(m-1)}$$

$\xi(r)$  convergence — L800/lc16 iz155 ( $z=1.50$ )  $\log_{10}(M_*/M_\odot h^{-1}) > 9.0$



# Convergence



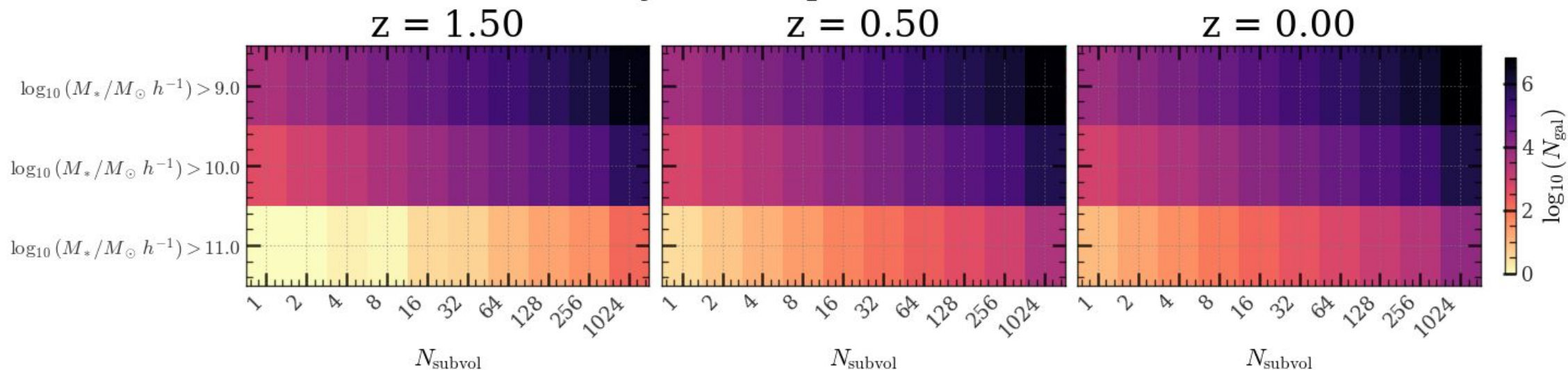
# SCOPE Performance

Built entirely in Rust for maximum memory efficiency and speed.

Utilises the Rayon library for extreme data parallelism.

Converts sequential pair-counting tasks into parallel iterators, safely processing multiple sub-volumes simultaneously without race conditions.

Galaxy counts per selection



# Summary

**Clustering:** The 2PCF is fundamentally more constraining than luminosity functions for testing the galaxy-halo connection.

**Physics drives 2PCF:** GALFORM's dynamical friction dictates the 1-halo term, while AGN/Supernova feedback drives the 2-halo bias.

**Sub-sampling breaks the 1-halo term:** Naively combining independent realisations mathematically inflates intra-halo pairs by a factor of  $\frac{k}{m}$ .

**SCOPE reconstructs the 2PCF:** By separating pairs and applying geometric weights, we recover the full-box clustering signal with fewer galaxies.

**Computational Efficiency:** Clustering measurements are now feasible even with sparse data.