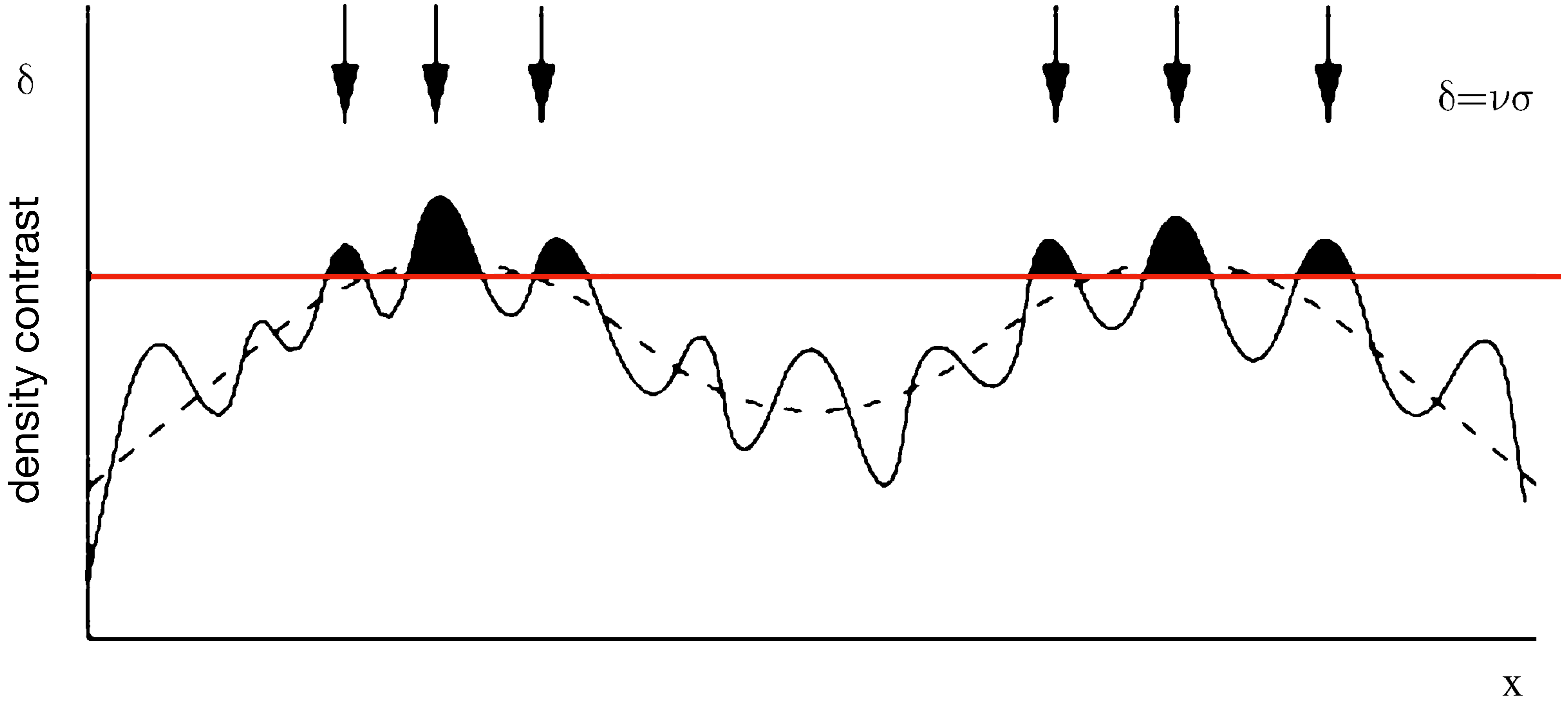
The large-scale structure of the Universe

Cristina García Vergara

AstroTwinCoLo (Virtual) School 2020

Previous lecture: Dark matter halos and galaxies are biased tracers of the dark matter



Threshold at which a massive DM halo can be formed

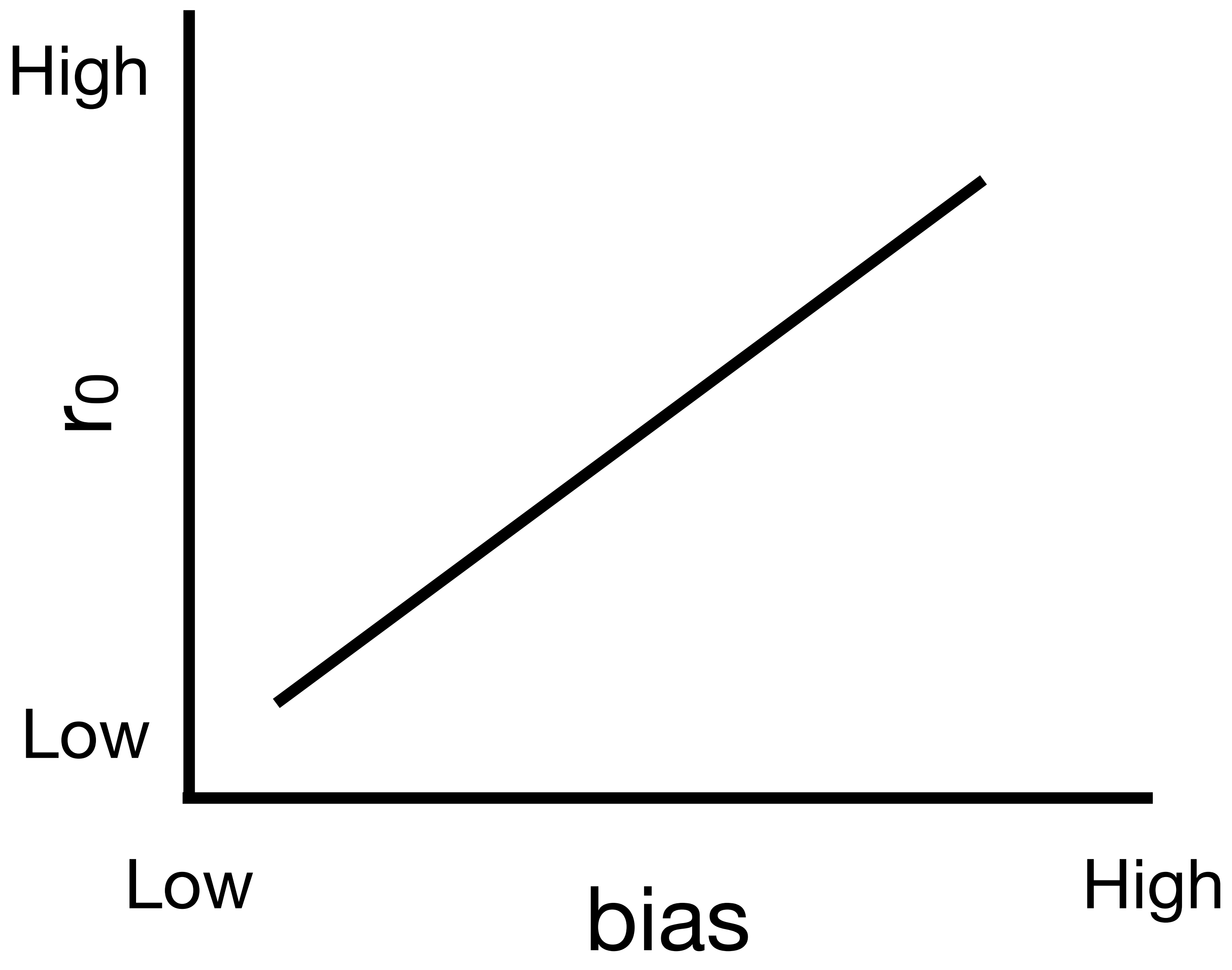
The combination of small- and large- scale fluctuation of dark matter together with the condition required for a halo to be formed, make dark matter halos (and therefore galaxies) to be biased tracers of the dark matter distribution, with more massive halos being more biased tracers.

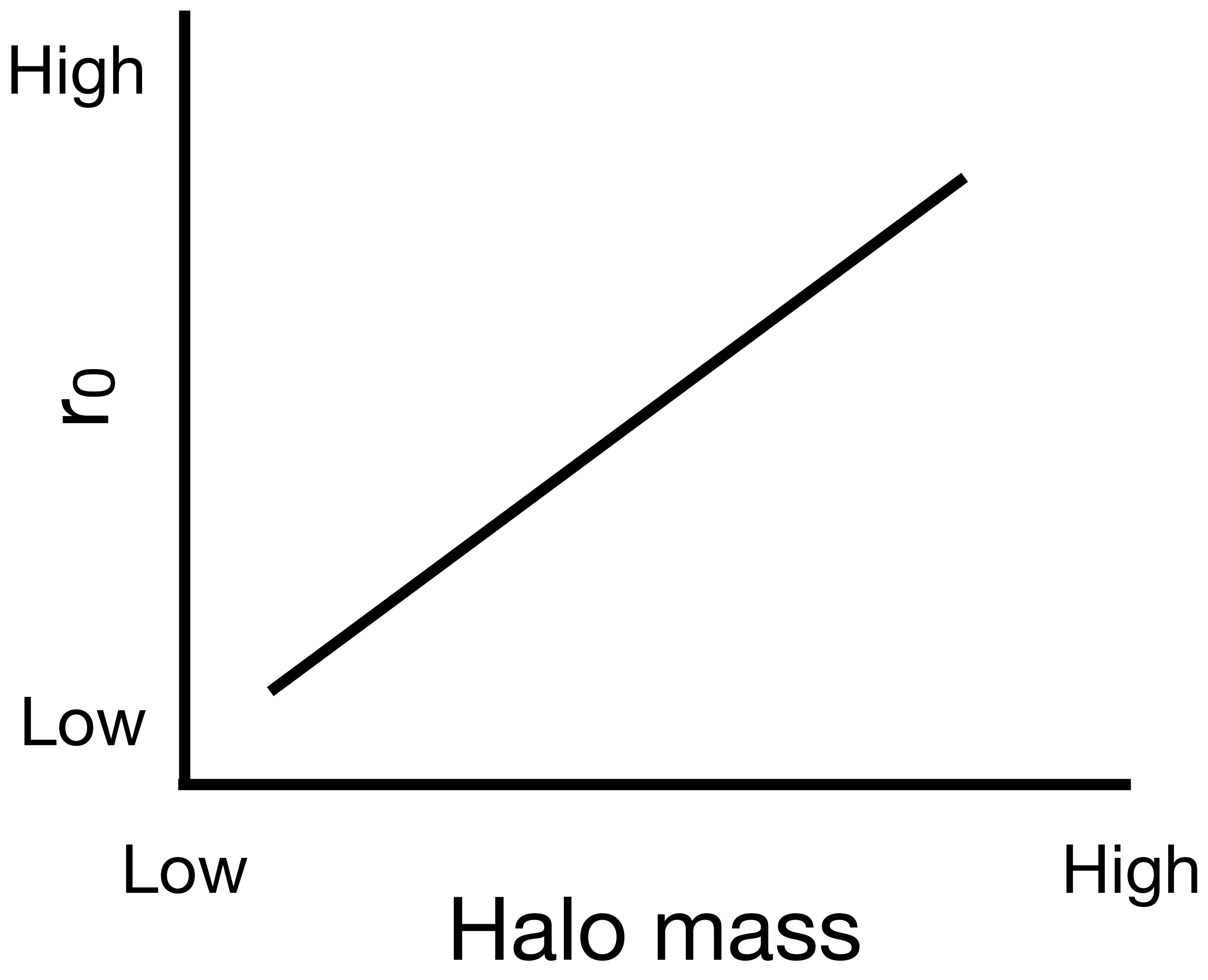
Previous lecture: Relation between correlation length, bias and dark matter halo mass

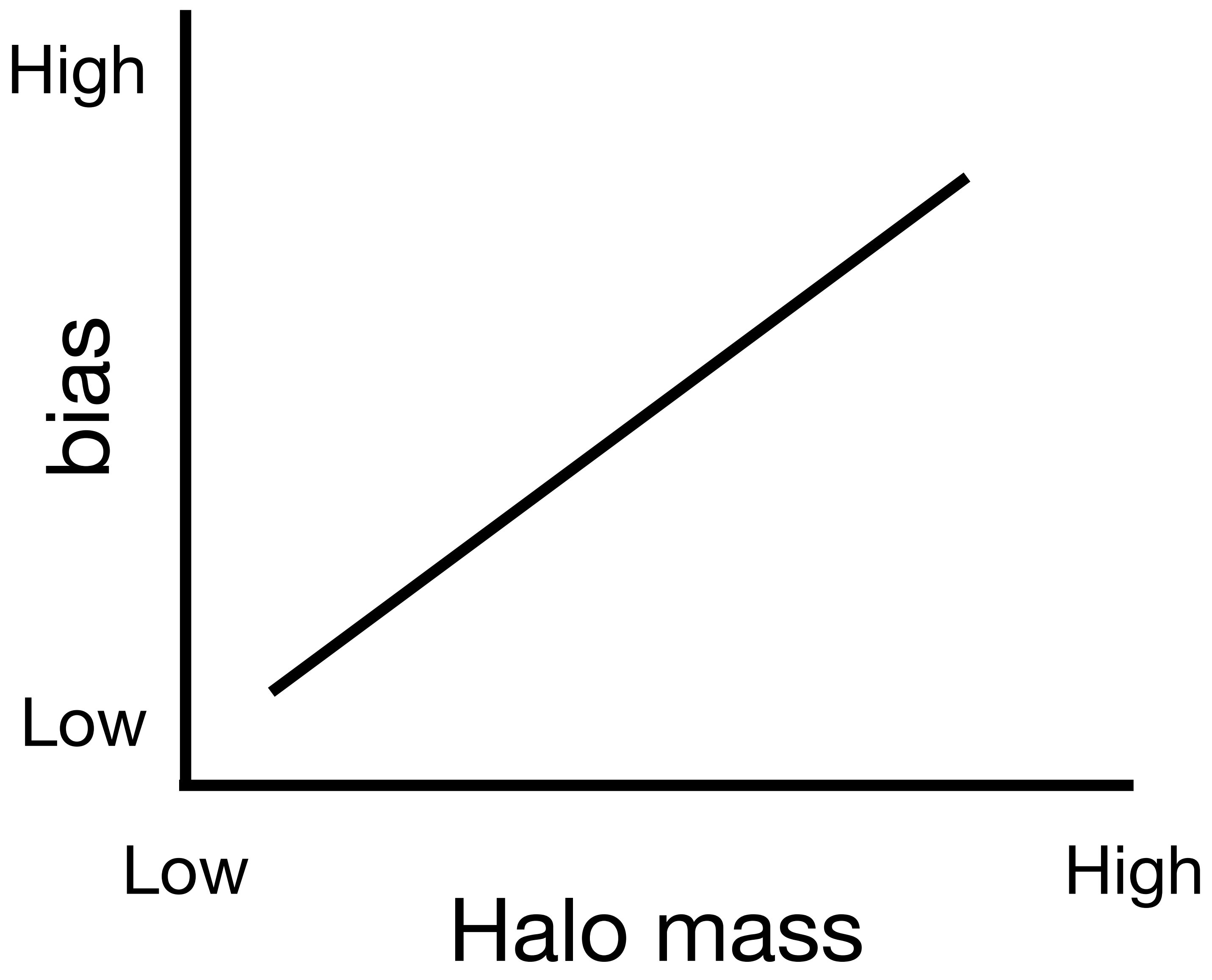
$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$$

$$b_g = \sqrt{\xi_g/\xi_{DM}}$$

At a fixed redshift:

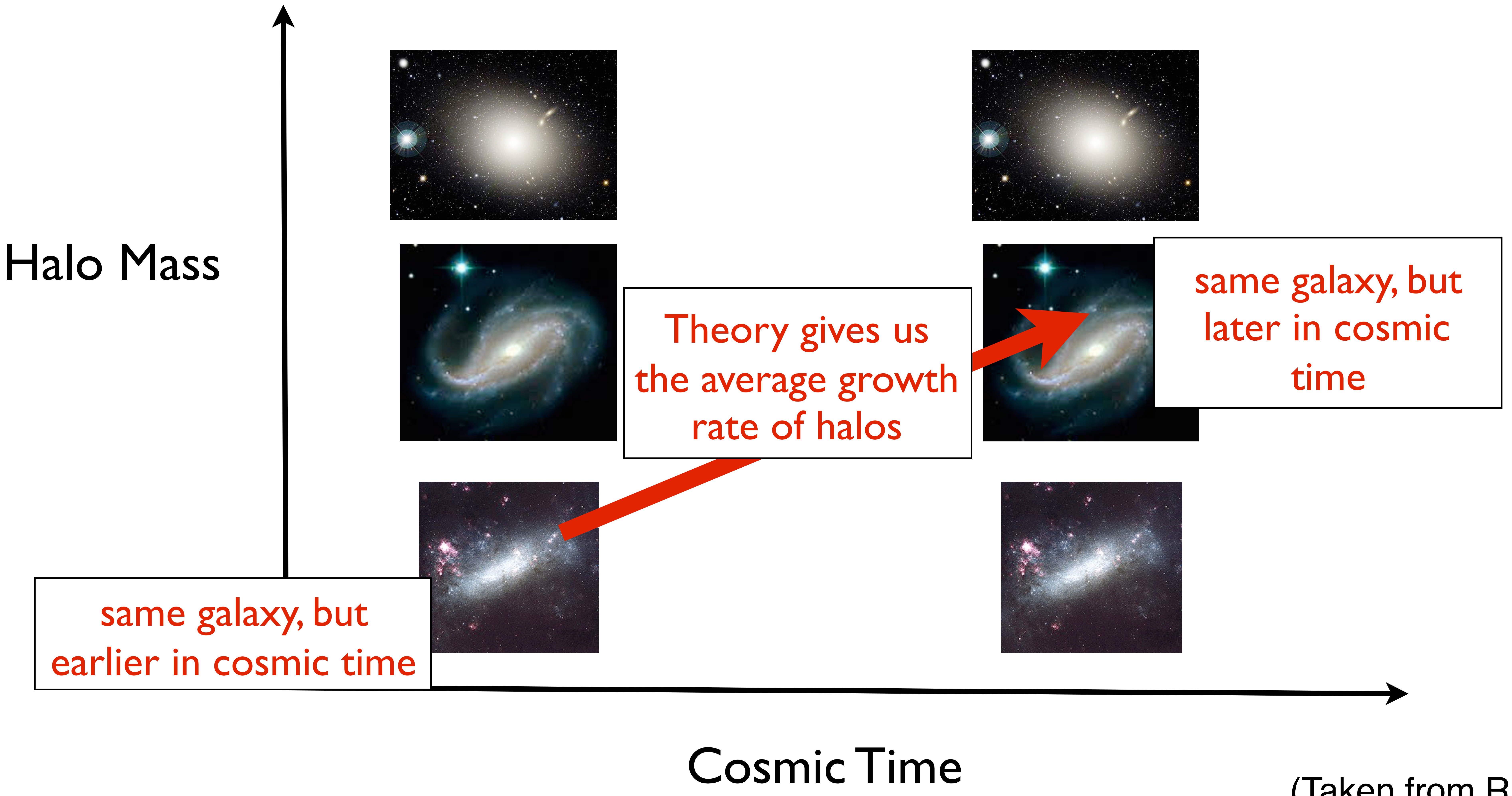






Previous lecture: Why is it useful to learn about the dark matter halos in which galaxies live?

It provides us a powerful tool for tracing the same population of galaxies through cosmic time.



Halo Mass

Cosmic Time

Theory gives us the average growth rate of halos

same galaxy, but earlier in cosmic time

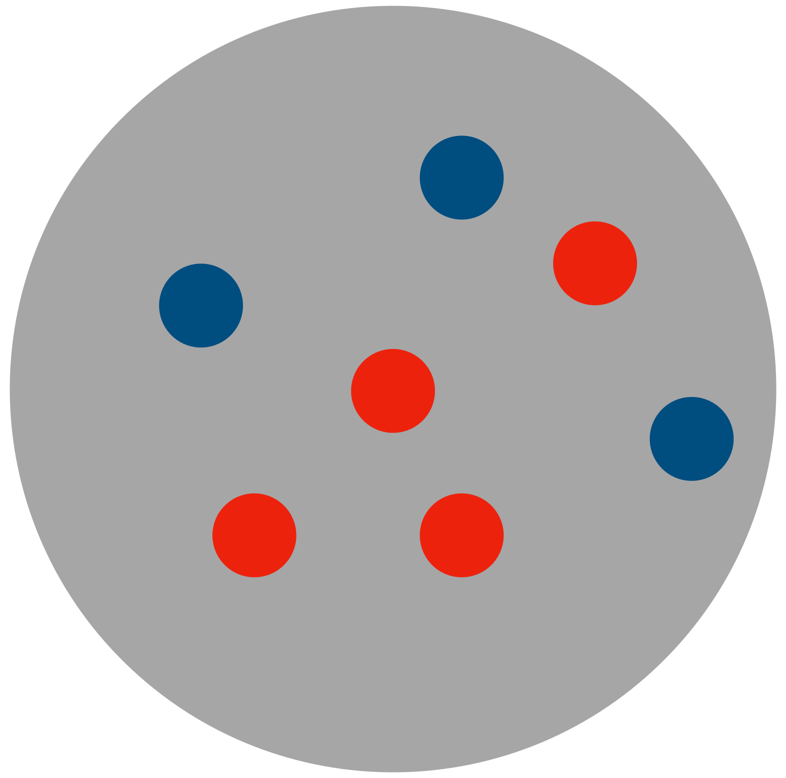
same galaxy, but later in cosmic time

(Taken from R. Bouwens' Lectures)

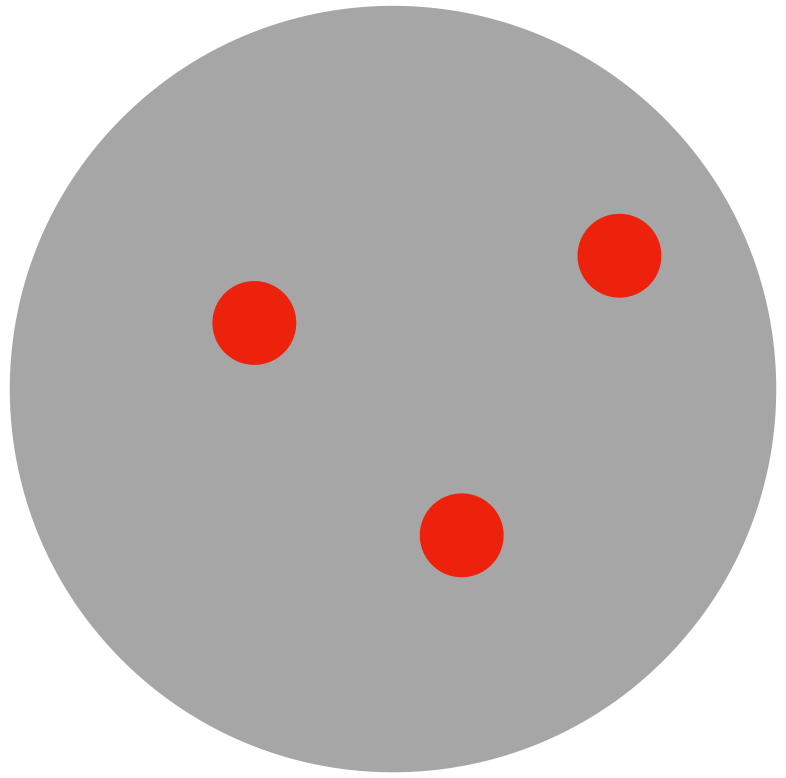
Previous lecture: Why is it useful to learn about the dark matter halos in which galaxies live?

It provides us a powerful insights about physical processes involved in the formation or evolution of galaxies.

For example, if high-z quasars inhabit massive halos, that means that they live in more massive environments with more galaxies in their neighborhood. In these environments merger are more common, and then we could believe that quasar may be originated, or triggered by mergers.



More massive halos
More galaxies
Higher merger rate



Less massive halos
Less galaxies
Less merger rates

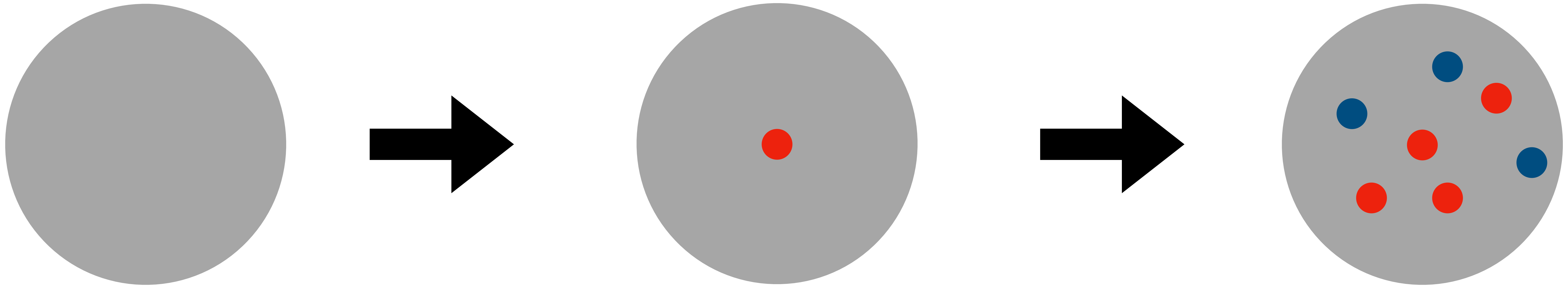
Why at high redshift quasars are only found in extremely massive dark matter halos? One (possible) answer: because quasars are originated from mergers

This class

- How can correlation function measurements provide insights about physical processes involved in the formation and evolution of galaxies?.
- Using clustering to identify and characterize the most prominent peaks of the dark matter density field in the early universe (tracers of matter overdensities).
- Final summary.

How can correlation function measurements provide insights about physical processes involved in the formation and evolution of galaxies?

Remember from Lecture 1: Galaxies form in dark matter halos.



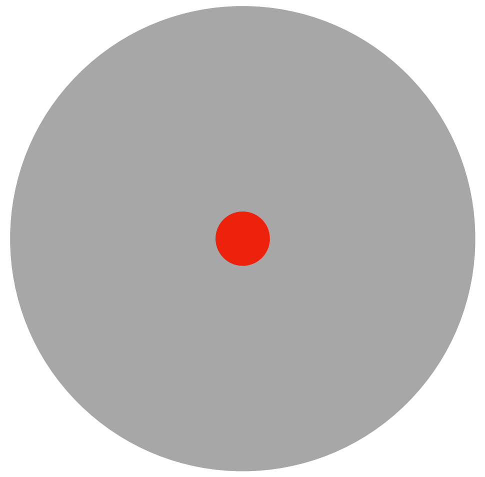
Almost all the dark matter halos, contain a galaxy in their center (central galaxy), and dark matter halos can also contain more than one galaxy (satellite galaxies).

This implies that the exact shape of the galaxy correlation function depends on two things:

1) How dark matter halos are distributed in space.
2) How galaxies populate dark matter halos.

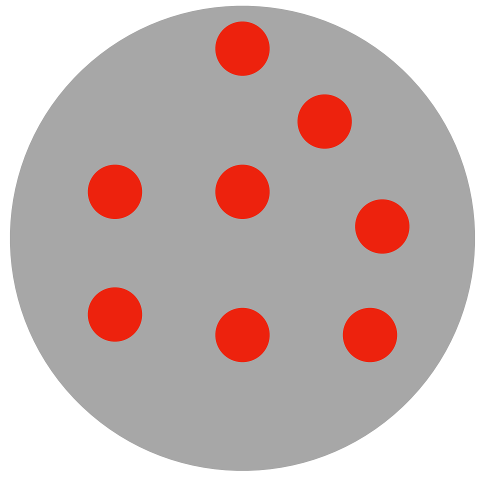
How galaxies populate a dark matter halo?

If we consider different models about how galaxies populate dark matter halos (the so-called halo occupation distribution models):



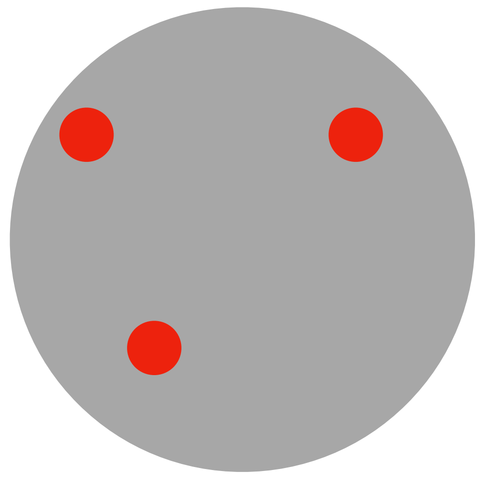
All the halos contain 1 galaxy in their center?

or



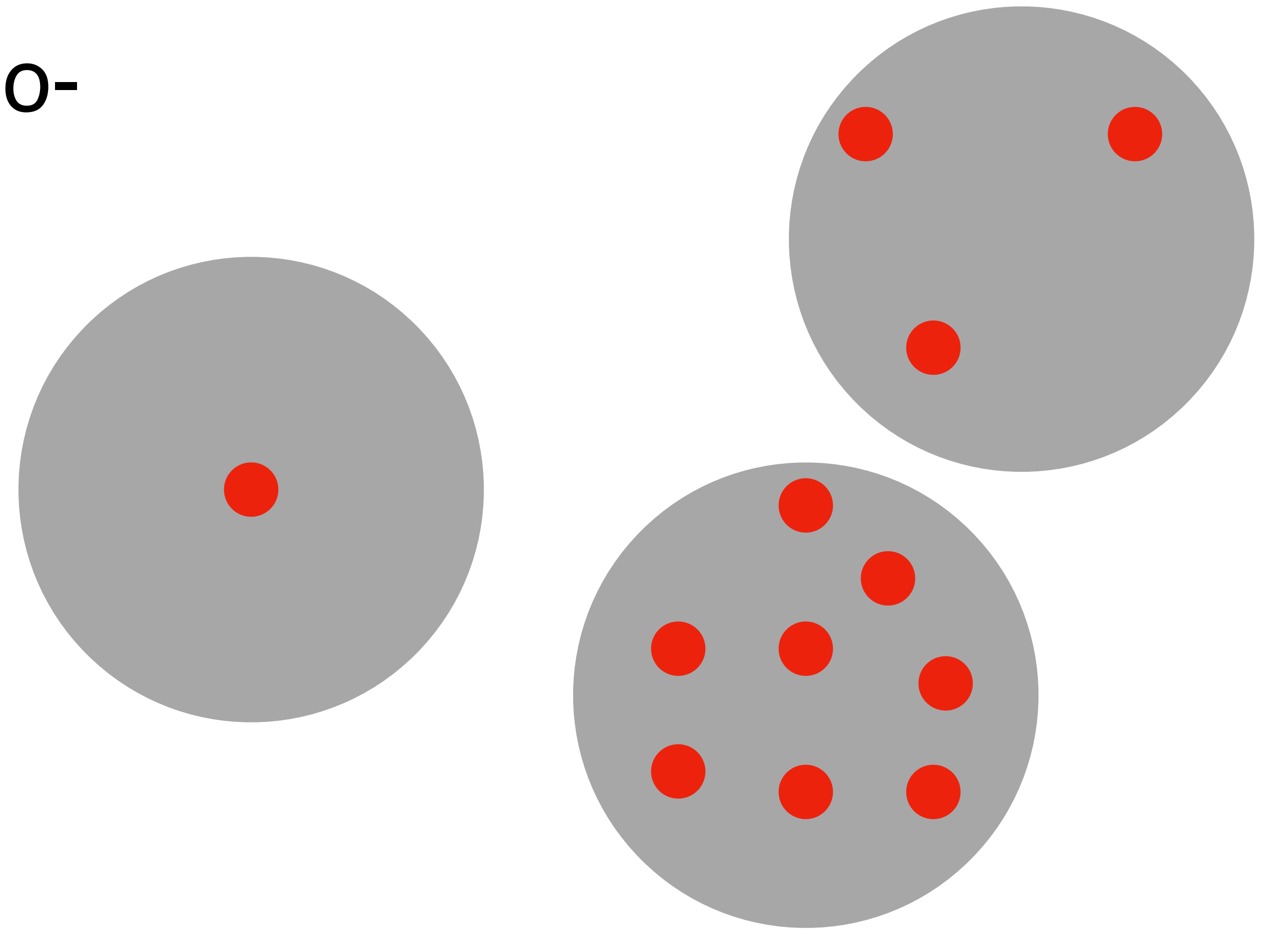
All the halos contain several galaxies randomly distributed within the halo?

or



All the halos contain few galaxies but only in the outskirts?

or



All the halos are populated differently?

We would measure a different galaxy correlation function for each different model.

Therefore, we can use the observed galaxy clustering to put constraints on halo occupation distribution models, and then understand which physical processes are involved in galaxy formation and evolution.

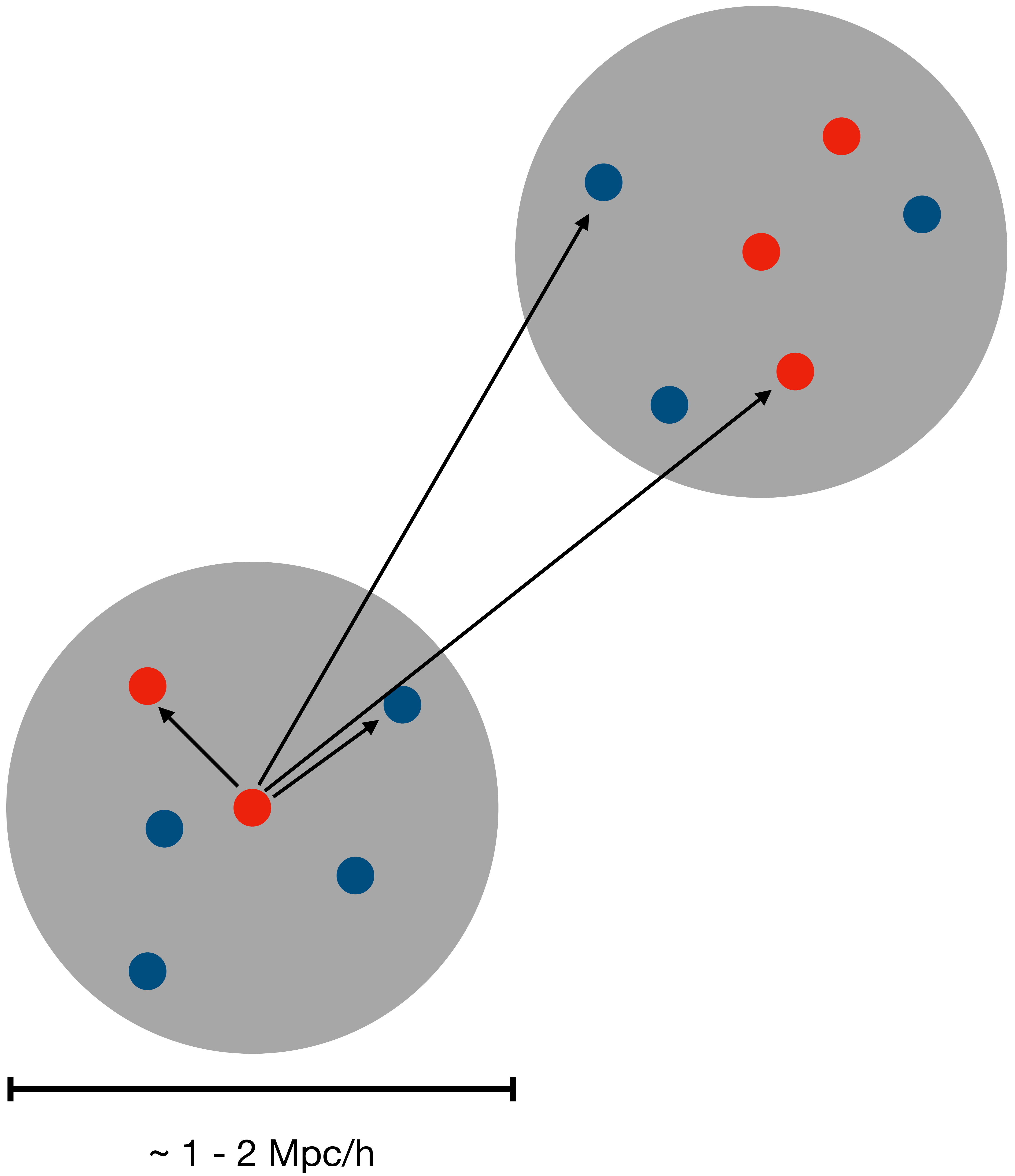
The power law-shape for the correlation function of galaxies

Clustering on smaller scales is a result of galaxy pairs within the same halo. The number pairs is dominated by the radial profile of galaxies within an halo: **one-halo term.**

The clustering of galaxies on scales larger than a typical halo, results from pairs of galaxies in separate halos. The number of pairs is more dominated by the distribution of halos on space: **two-halo term.**

The observed clustering of galaxies is the result of a combination of the large and small scales number counts, and it is observed to have a power-law shape.

Why a power-law shape?



The power law-shape for the correlation function of galaxies

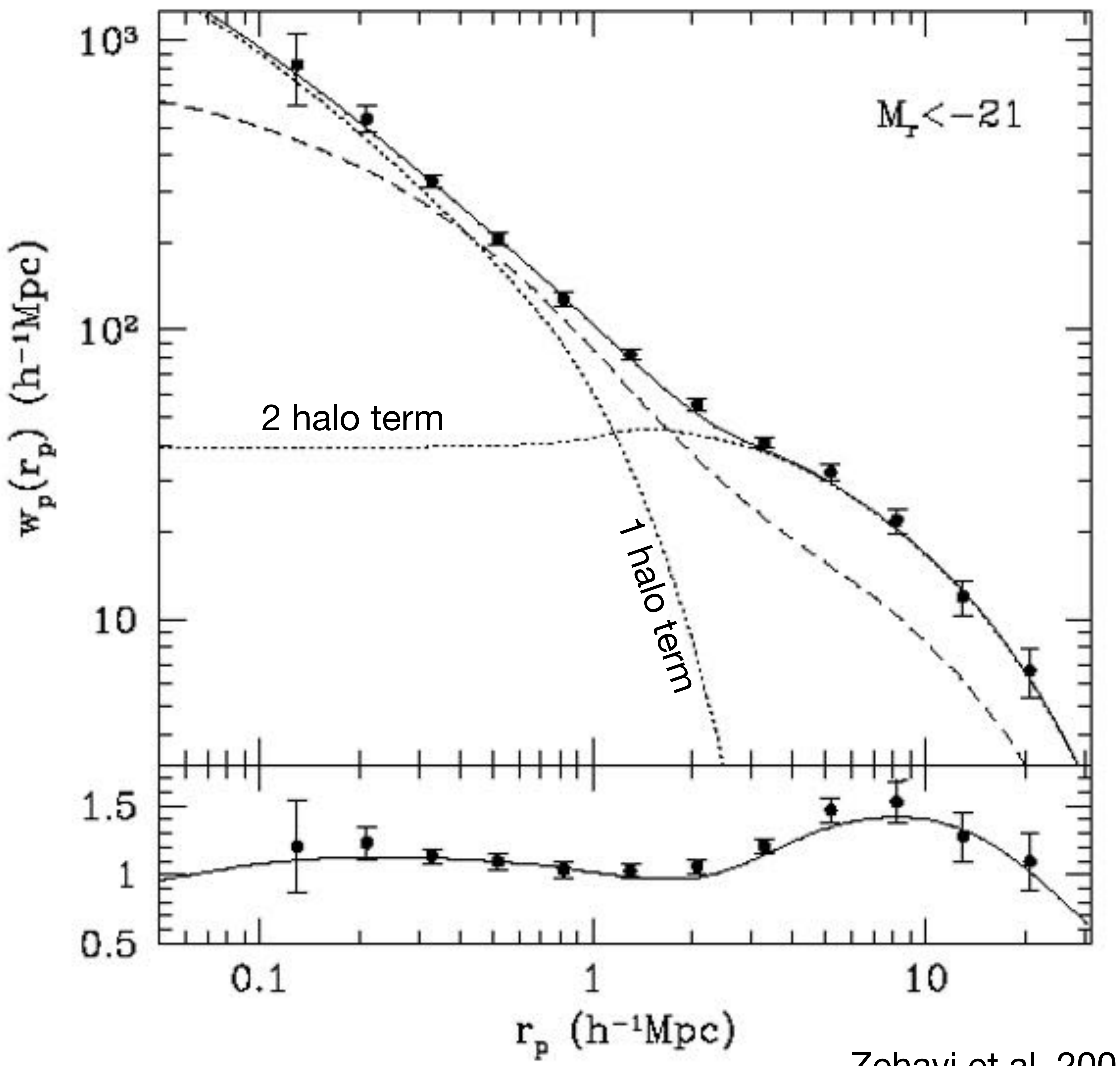
The 2-halo term: come from the number of pairs at large scales, dominated by the distribution of halos on space.

This is relatively well understood and can be determined from simulations, since it depends mostly on properties of dark matter, on expansion history of the universe, on the formation of halos in the density field.

The 1-halo term: come from the number of pairs at small scales, dominated by the radial profile of galaxies within an halo. This depend on how galaxies populate dark matter halos.

This is more difficult to understand, because it depends on complex baryonic physics (for example: star formation, feedback, cooling, mergers, etc). We need models to describe it.

The projected correlation function of galaxies from SDSS



Zehavi et al. 2004

Halo occupation distribution (HOD) modeling

- Several works have proposed different models to describe how galaxies populate dark matter halos (e.g. Peacock & Smith 2000, Benson et al. 2000, Berlind & Weinberg 2002, Kravtsov et al. 2004, etc).

- Most of the models include at least two parameters:
    - The number of galaxies (brighter than some limit) in a halo of a given mass (the halo occupation number).
    - The distribution of galaxies within an halo (density profile).

And models also includes different physical processes involved in the formation and evolution of galaxies as well as interaction between galaxies.

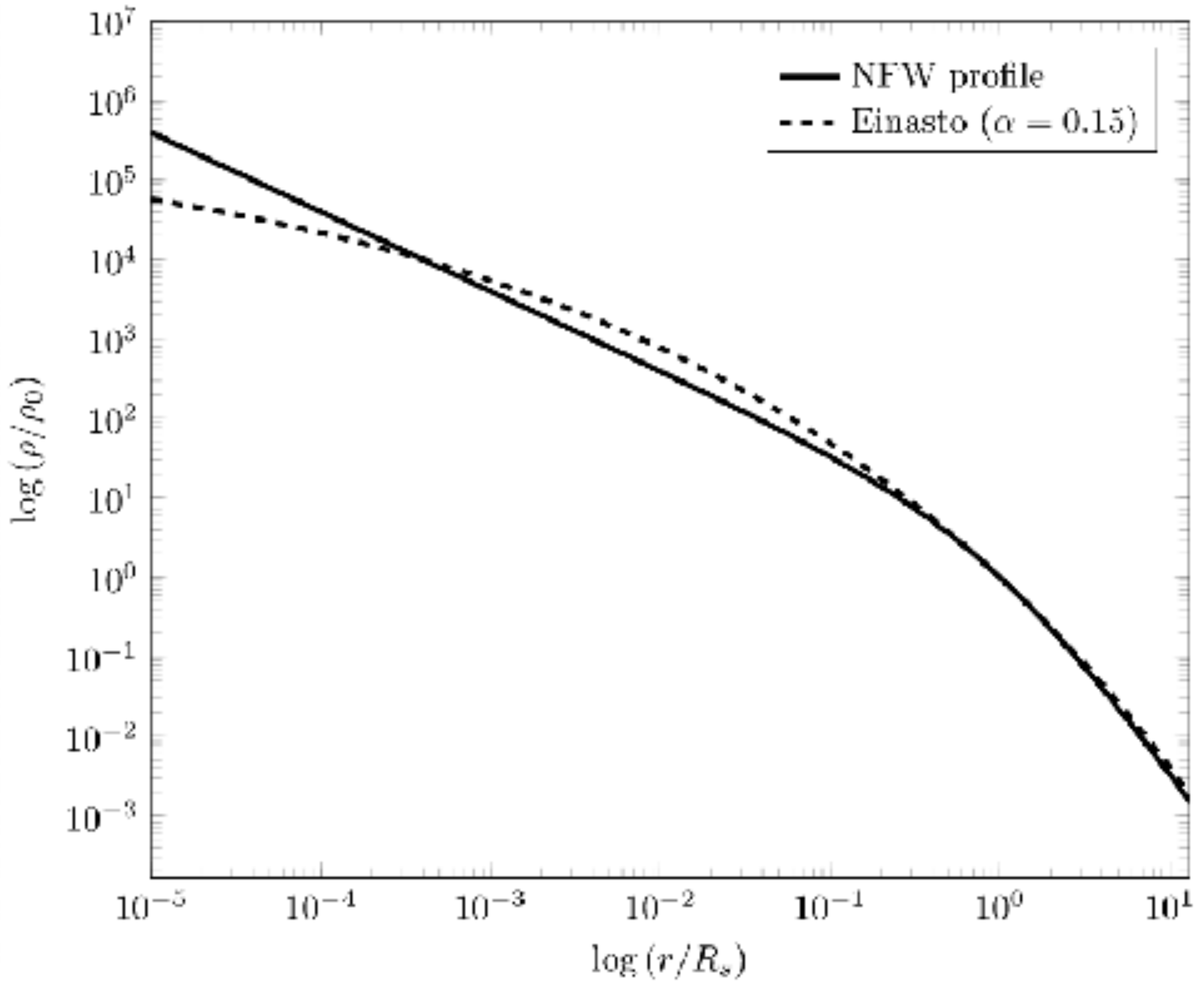
Halo occupation distribution (HOD) modeling

- For the radial distribution of galaxies within an halo, it is typically assumed that the distribution of galaxies follow the distribution of dark matter within the halo.

- The distribution of dark matter within an halo is well known to roughly follow a Navarro-Frenk-White (NFW) profile:

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}$$

where r_s is a characteristic radius and ρ_s represent the amplitude of the density profile



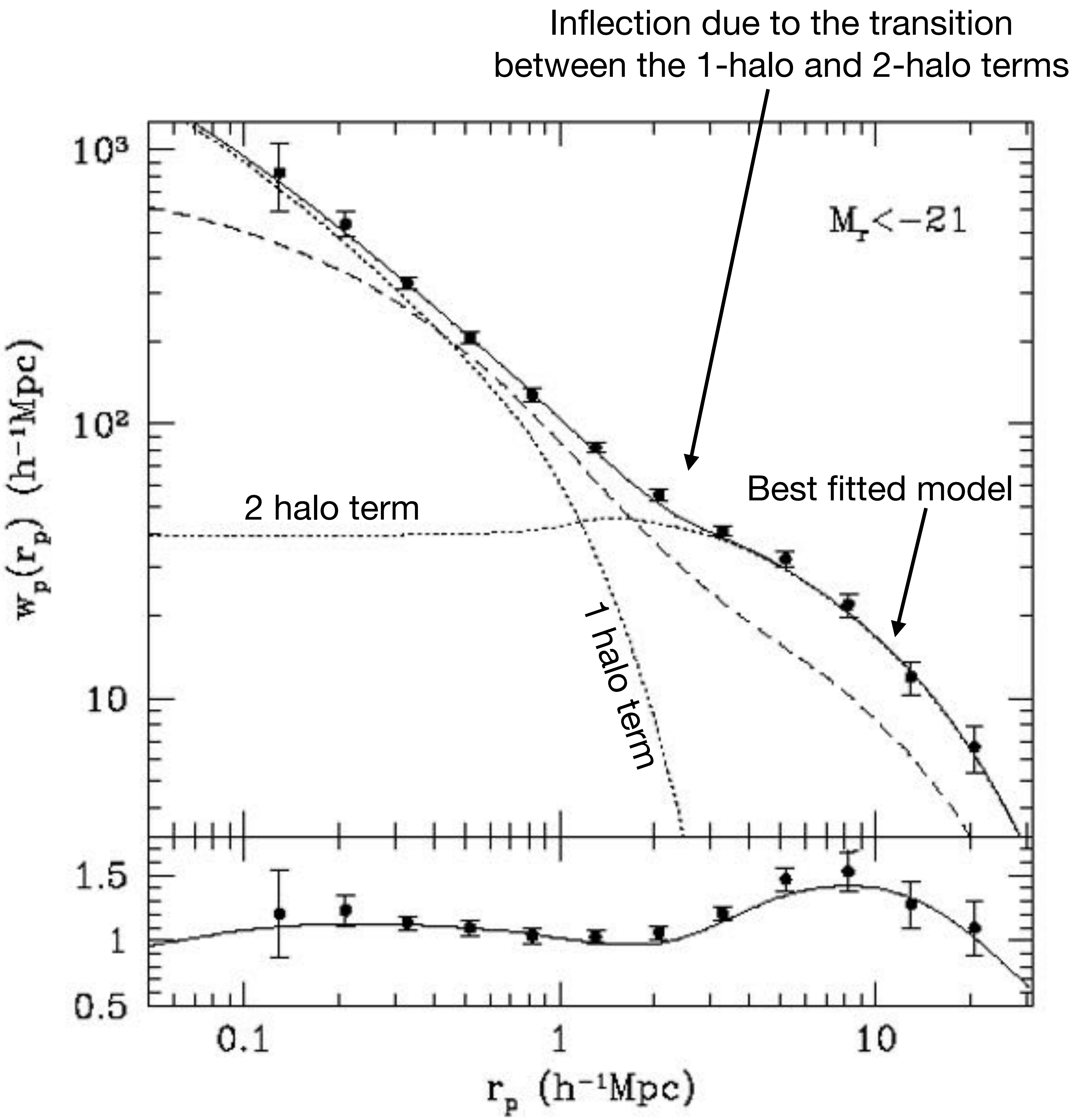
- This assumption determine the shape of the 1-halo term of the correlation function of galaxies.

The power law-shape for the correlation function of galaxies

The combination of the 1-halo and 2-halo terms result in a power law-shape for the galaxy correlation function.

The observed correlation function can be fitted to find the best model parameters.

We can find constraints for the halo occupation number, the density profile and dominant physical processes.



Zehavi et al. 2004

How can correlation function measurements provide insights about physical processes involved in the formation and evolution of galaxies?

But remember: clustering depends on properties of galaxies.

We have different correlation function for different populations, which implies that different models will be best fitted for different populations. This can constrain which physical processes are dominant for the different populations.

Example: HOD modeling provide a clear explanation for the increased clustering observed for faint red galaxies.

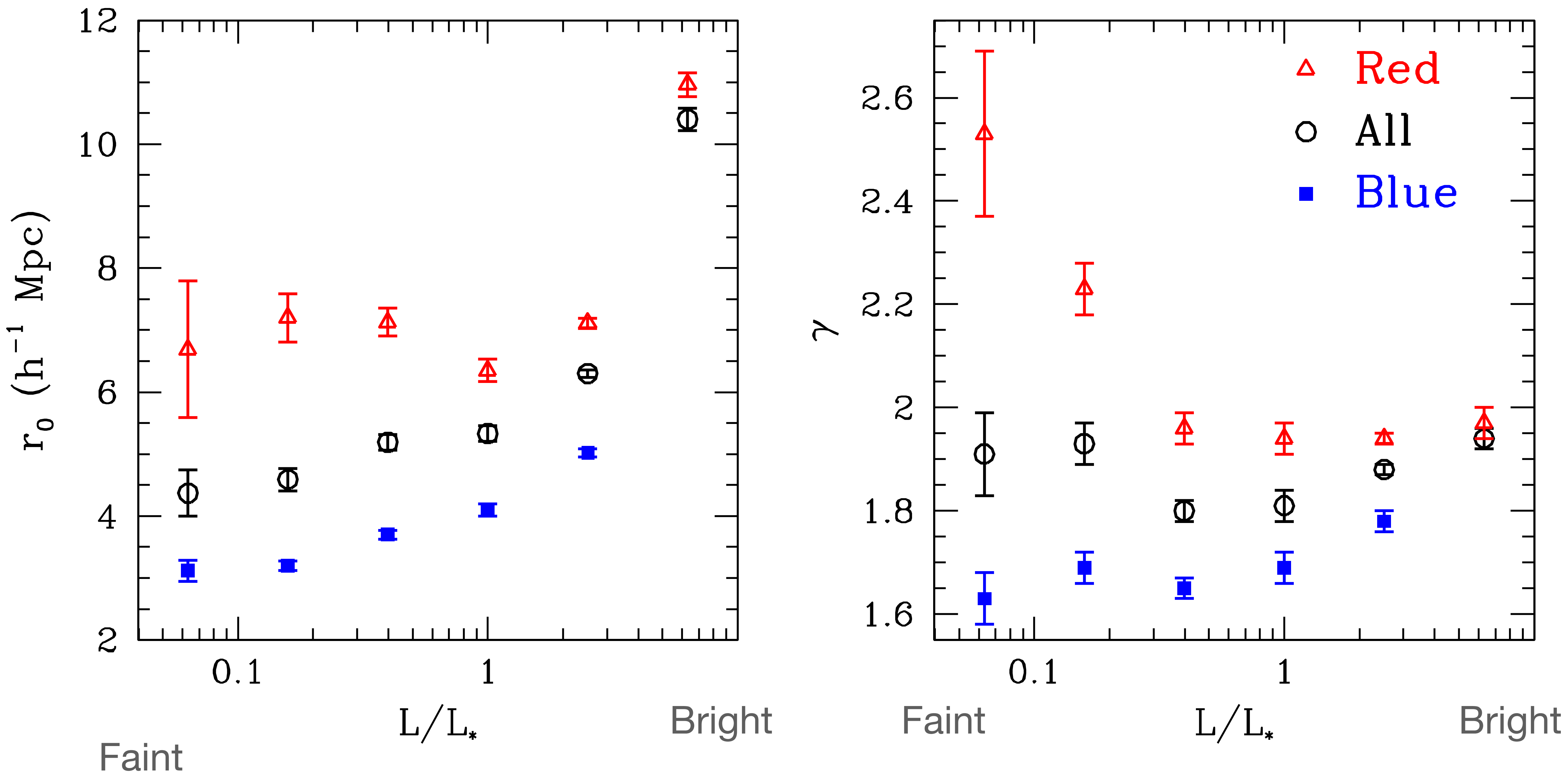
How can correlation function measurements provide insights about physical processes involved in the formation and evolution of galaxies?

Example: HOD modeling provide a clear explanation for the increased clustering observed for faint red galaxies.

HOD fitting for this correlation function, suggest that red faint galaxies are satellite galaxies in massive halos. The fraction of galaxies that are satellite is much higher for red galaxies (60%) than for blue galaxies (25%).

This also make the slope of red galaxies steeper in the correlation function (resulting from the increased fraction of satellite galaxies)

The best fitting power-law model for the correlation function of red and blue galaxies.



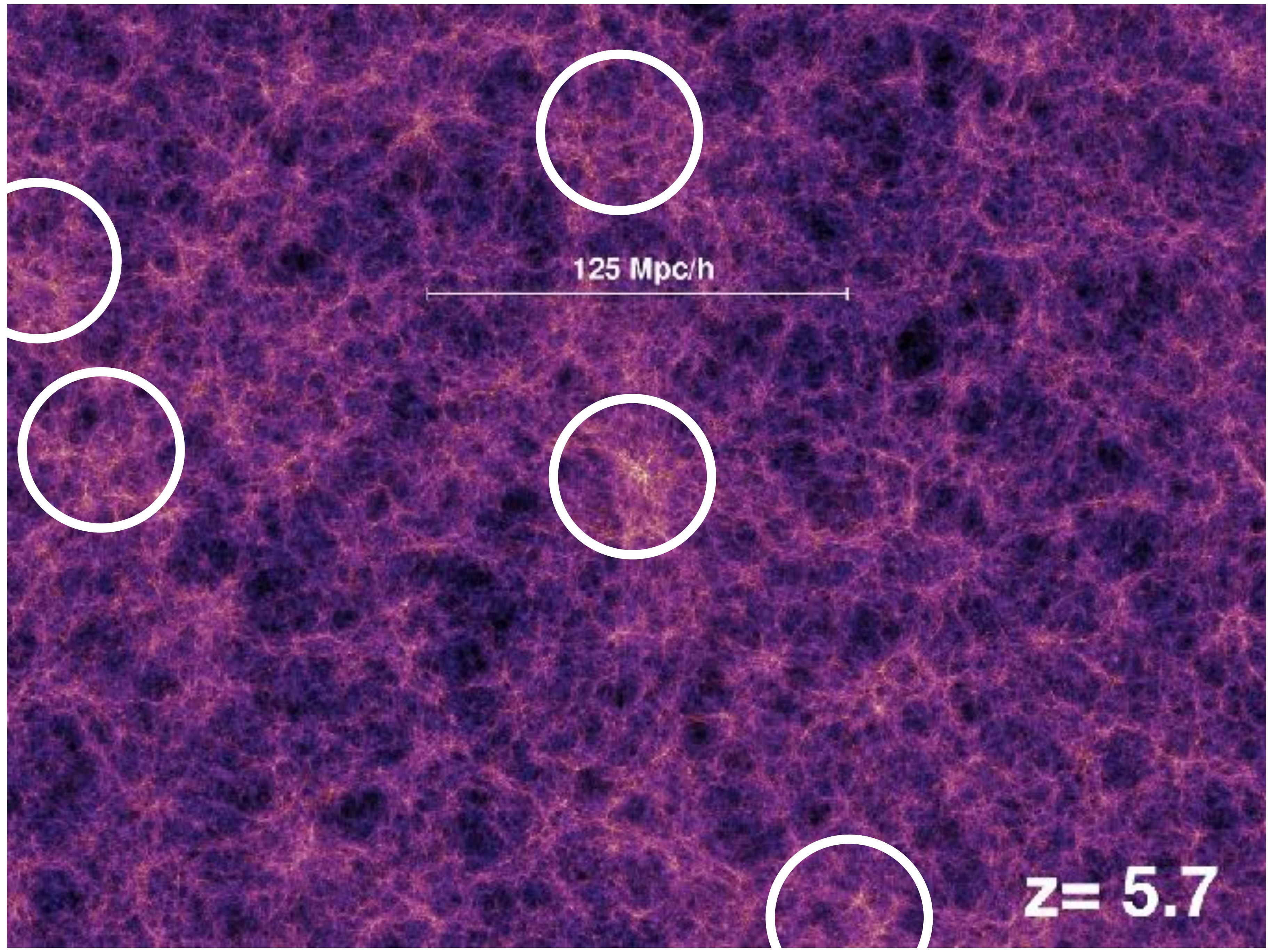
(Zehavi et al. 2011)

Other Example: Determining the number of galaxies in a halo of a given mass allow us to estimate merger rates in a halo, which could be one trigger mechanism for high star formation rate in galaxies.

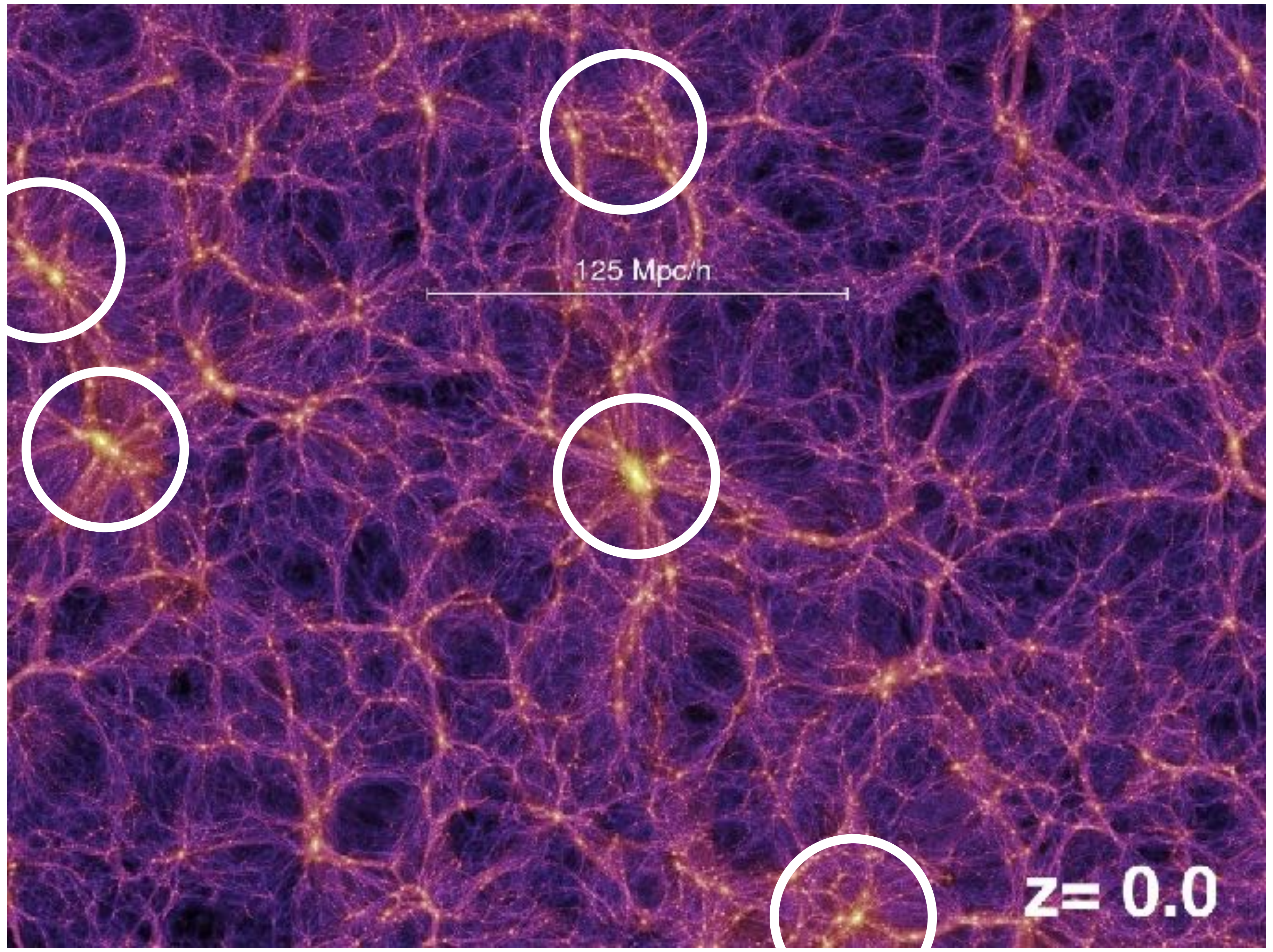
HOD model facilites the interpretation of the observed CF and provides constraints on models of gal formation and evolution within halos.

Using clustering to identify and characterize the most prominent peaks of the dark matter density field in the early universe

How can we identify the most prominent peaks of the dark matter density field in the early universe?



Protoclusters



Clusters of galaxies

As clusters of galaxies are the most prominent overdensities at z=0, they must have formed in the highest peaks in the early Universe. These regions can be identified as protoclusters of galaxies in the early universe.

How can we identify protoclusters?

- Identification of protoclusters in surveys is hard, because the contrast density is still small at high-z, and we need large and deep surveys to detect concentrations of galaxies over sky. Additionally we would need redshift information to check if galaxies are in the same gravitationally bound structure.

- An alternative: We can use tracers of matter overdensities: Objects executed to be located on massive halos in the early universe.

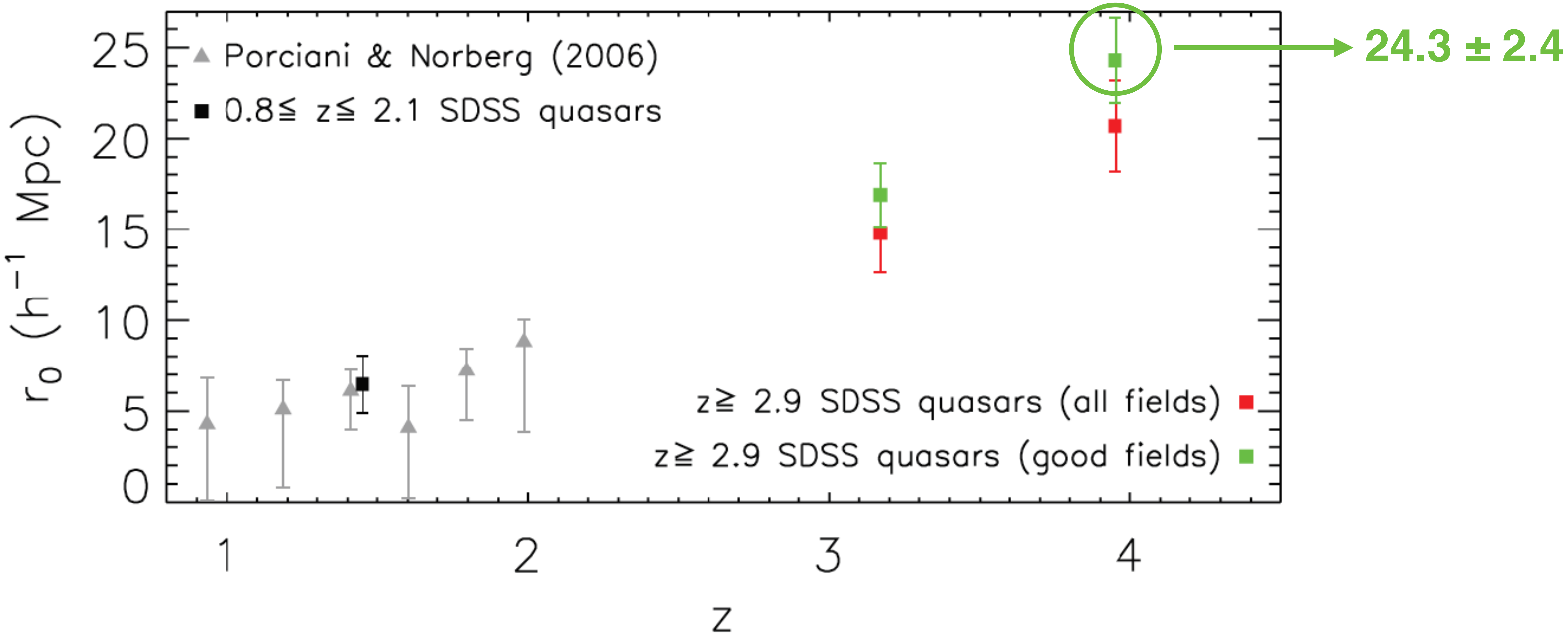
- From clustering measurements, we know that the most massive galaxies at high-z should be located in the most massive dark matter halos, then we can use massive galaxies as tracers of protoclusters in the early universe.

- Why we want to do that? To test the standard picture of structure formation. Specifically, to figure out if massive galaxies are situated in massive dark matter halos in the early Universe as the theoretical predictions suggest. Also because we can use protoclsuters as laboratories for the study of galaxy properties in the early universe.

Quasars should trace massive dark matter halos in the early Universe

Based on clustering measurements

Shen et al. 2007

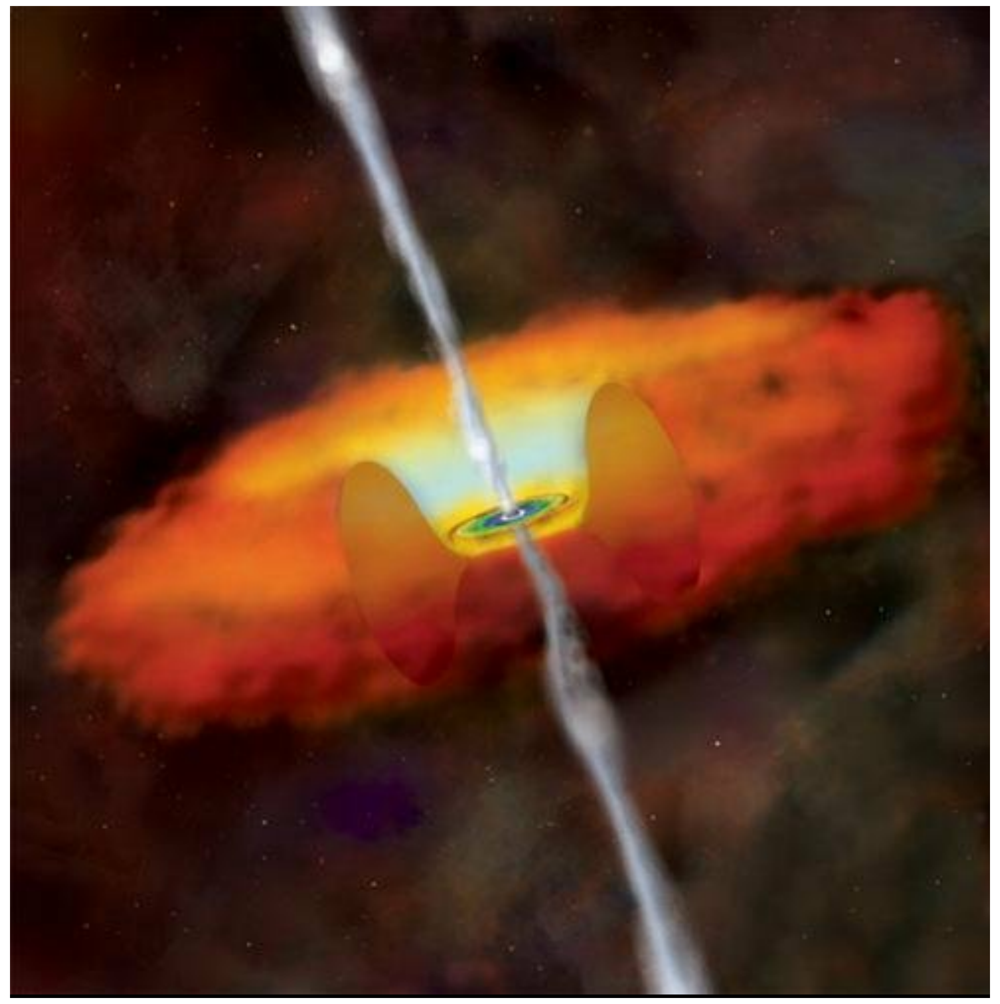


Extremely strong quasar clustering at z > 3.5

$M_{halo} > (4 - 6) \times 10^{12}\ M_\odot/h$

If this is true, we expect a large concentration of galaxies around quasars

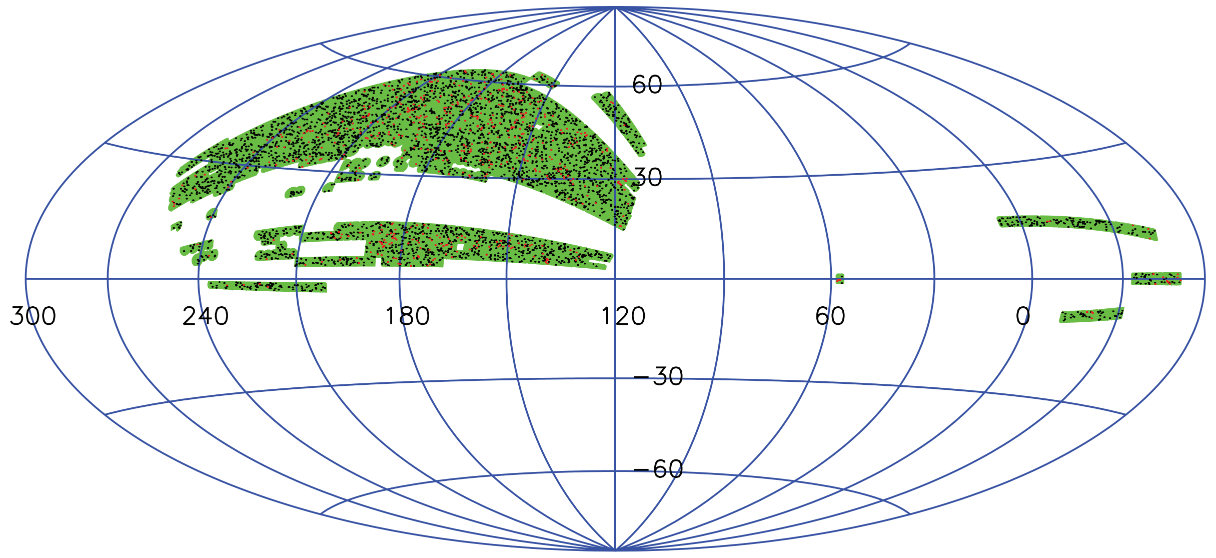
Why do we use quasars and not other massive galaxies?



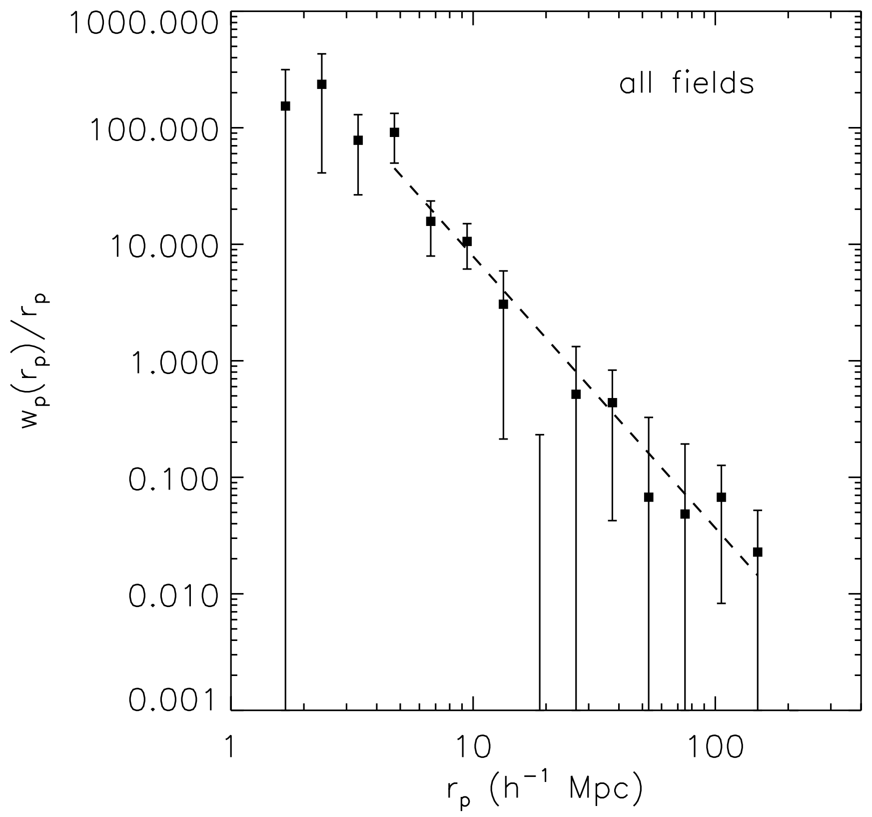
- Other candidates: SMGs, although their clustering is still not precisely constrained over cosmic time, and then we don’t really know if they are really that massive.
- Other candidates: Radio galaxies at z~2 (They are strongly clustered and have been actually successfully used as tracers of overdensities).
- If we want to explore highest redshift, quasars are ideal because this is the most clustered population in the universe at z~4 (that is already an evidence that they trace massive structures).
- Conveniently they are also very luminous then “easy” to find.

How to test if quasars trace protoclusters?

1. By measuring their clustering

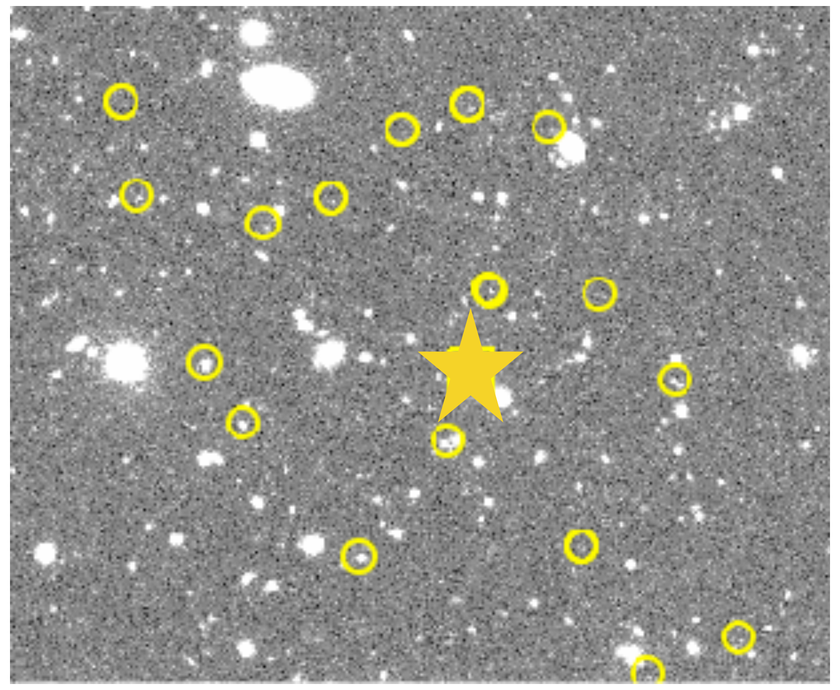


Massive galaxy survey

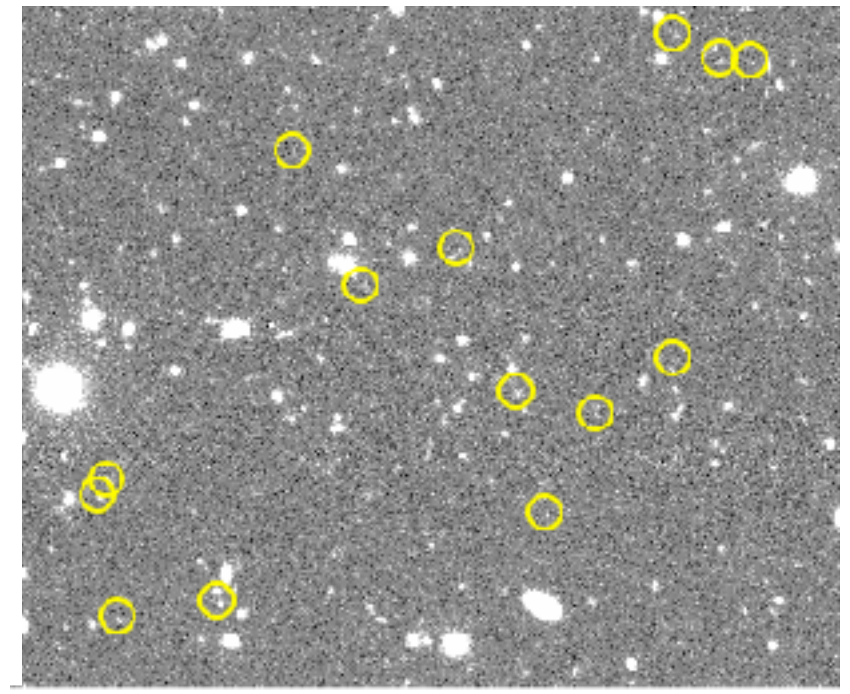


Already done for quasars

2. By directly detecting overdensities of galaxies around quasars.



Quasar field



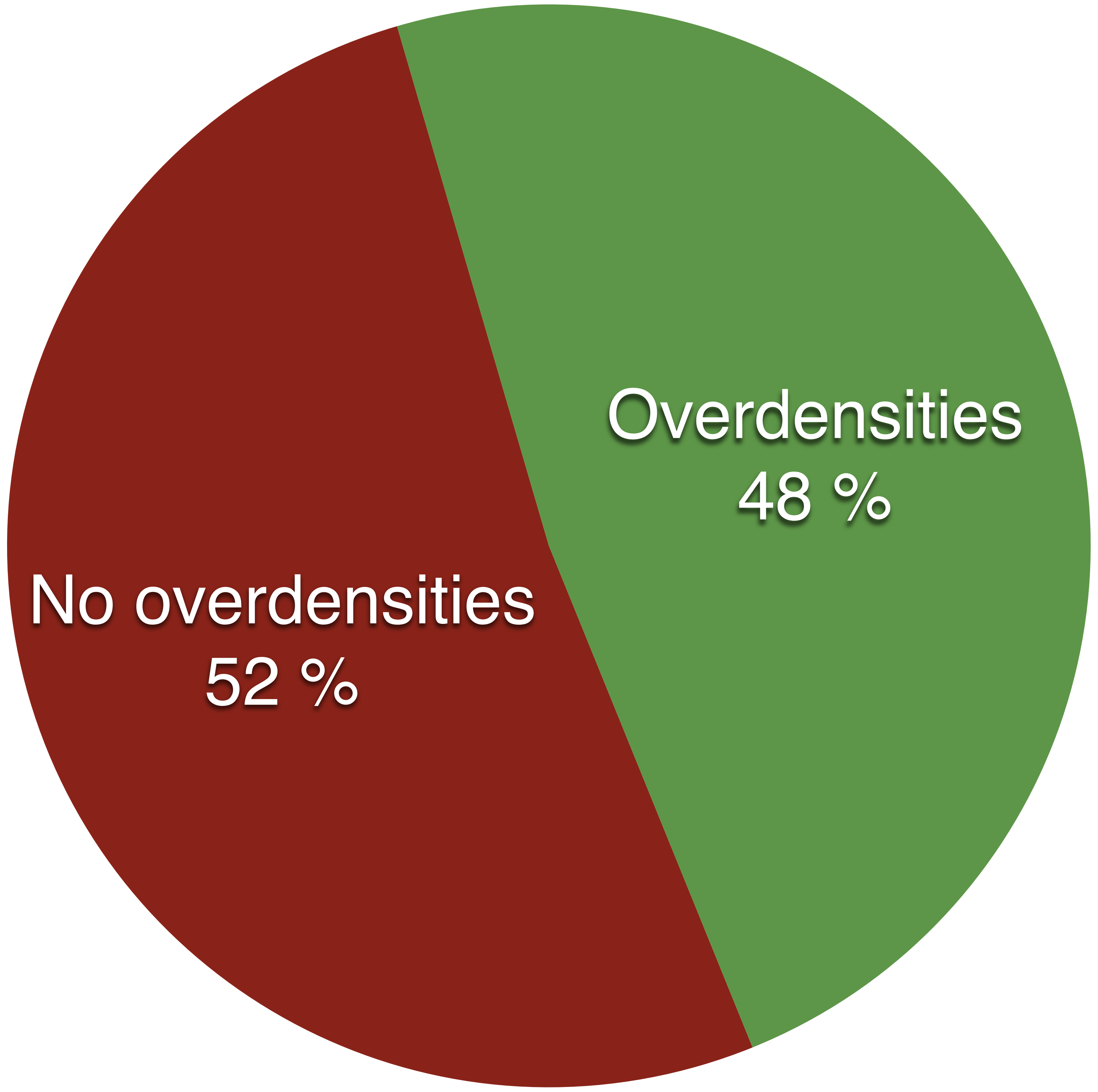
Blank field

$$\delta = \rho_{\rm qso}/\rho_{\rm blank}$$

We expect a large concentration of galaxies around quasars at z~4

Detection of galaxy overdensities around $z \gtrsim 4$ quasars: contradictory results

~30 quasar fields studied so far:



Why are these results inconclusive?

- Only individual quasar fields studied, then large cosmic variance effects and small statistics for each one.

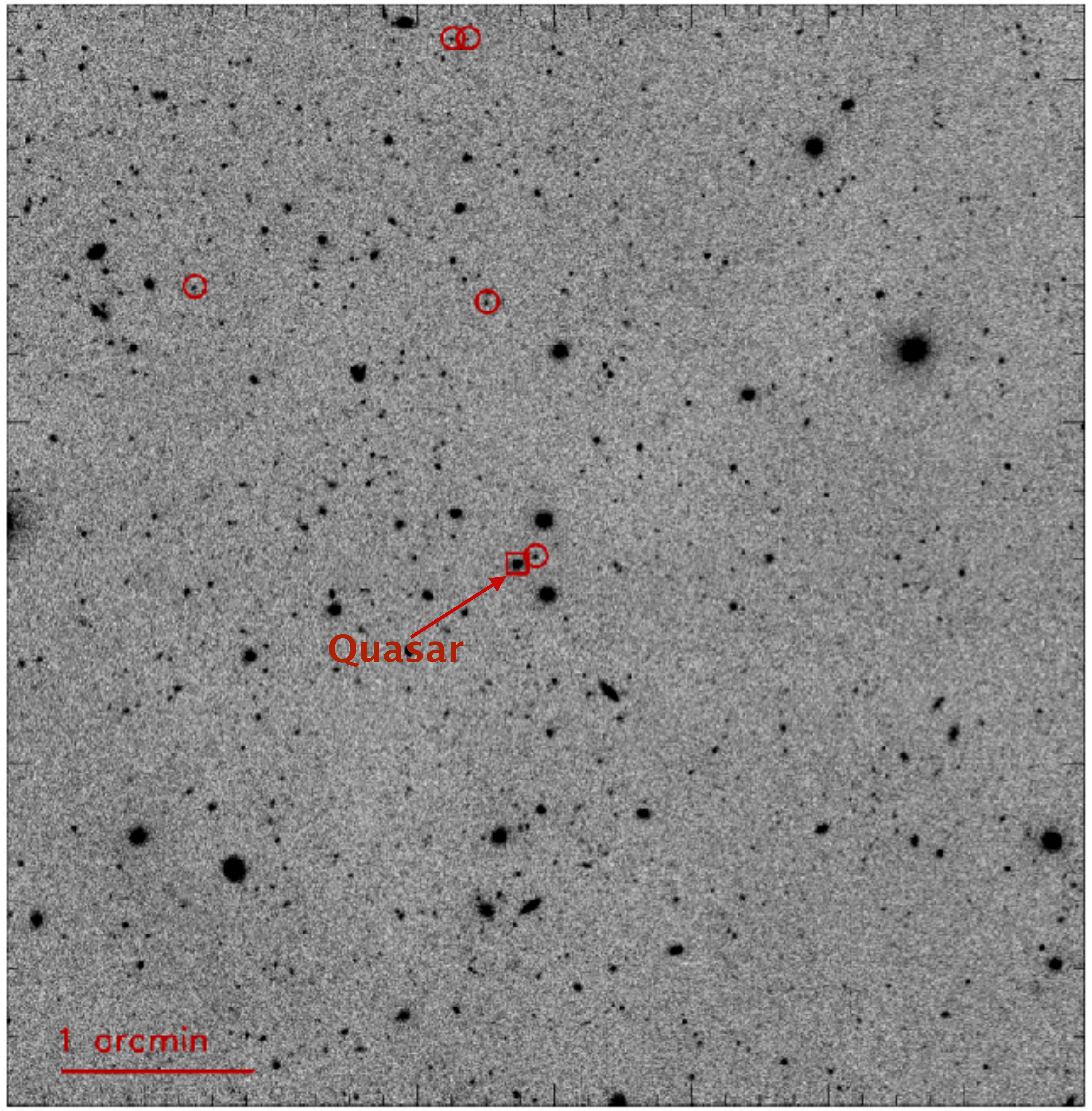
Willott+05, Kim+09,
Bañados+13, Husband+13,
Simpson+14, Mazzucchelli+17,
Kikuta+17, Goto+17, Ota+18.

Adams+05, Stiavelli+05, Zheng+06,
Kashikawa+07, , Kim+09, Utsumi+10,
Capak+11, Swinbank+12, Morselli+14,
Balmaverde+17, Ota+18

How can clustering solve this problem: the quasar-galaxy cross-correlation function

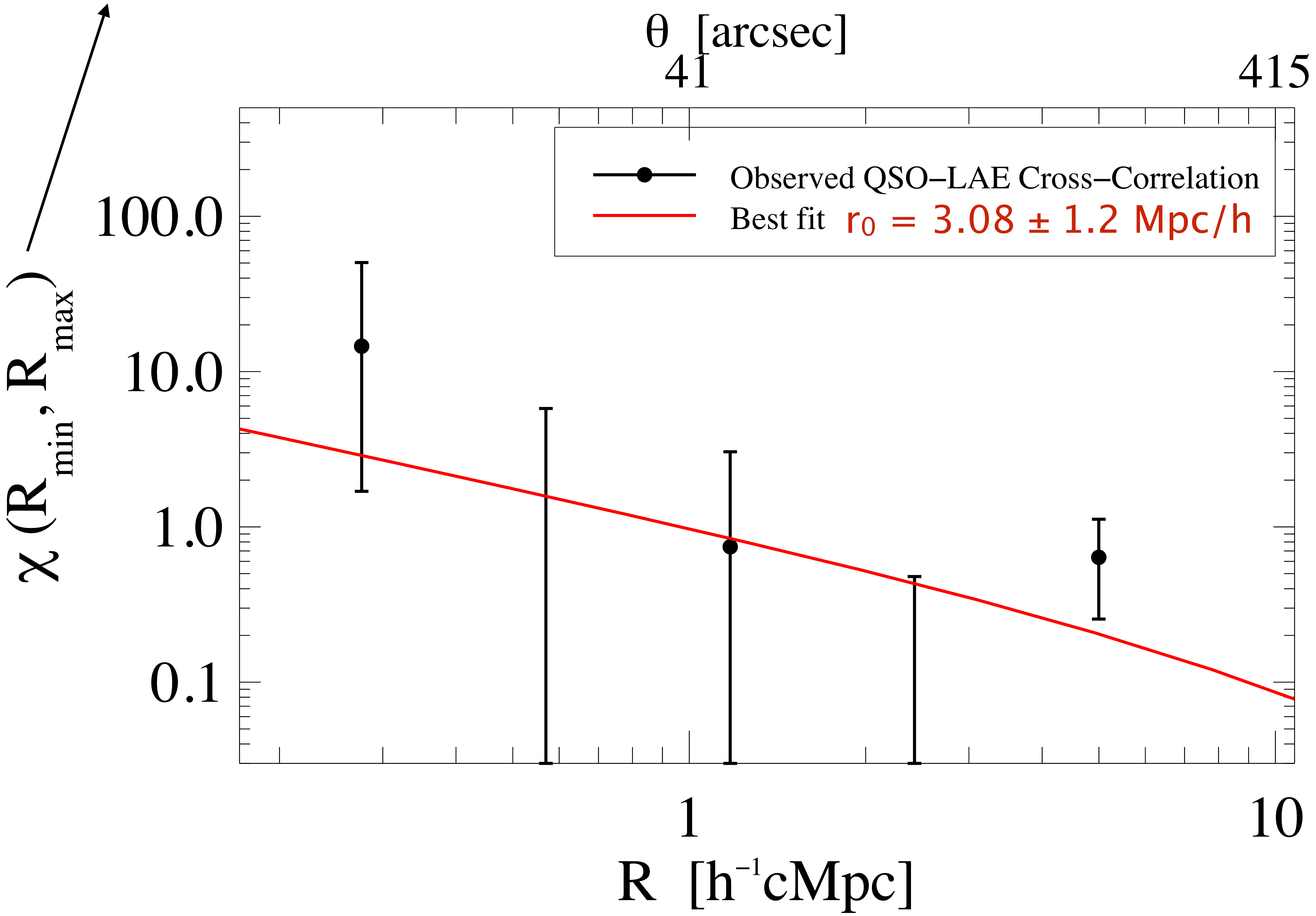
Expectation: Galaxies should be accumulated around quasar implying a large quasar-galaxy cross-correlation function.

Example of one quasar field imaged with VLT/FORS2



Galaxies are strongly clustered around quasars at z~4

$$\chi = \frac{\langle QG \rangle}{\langle QR \rangle} - 1$$

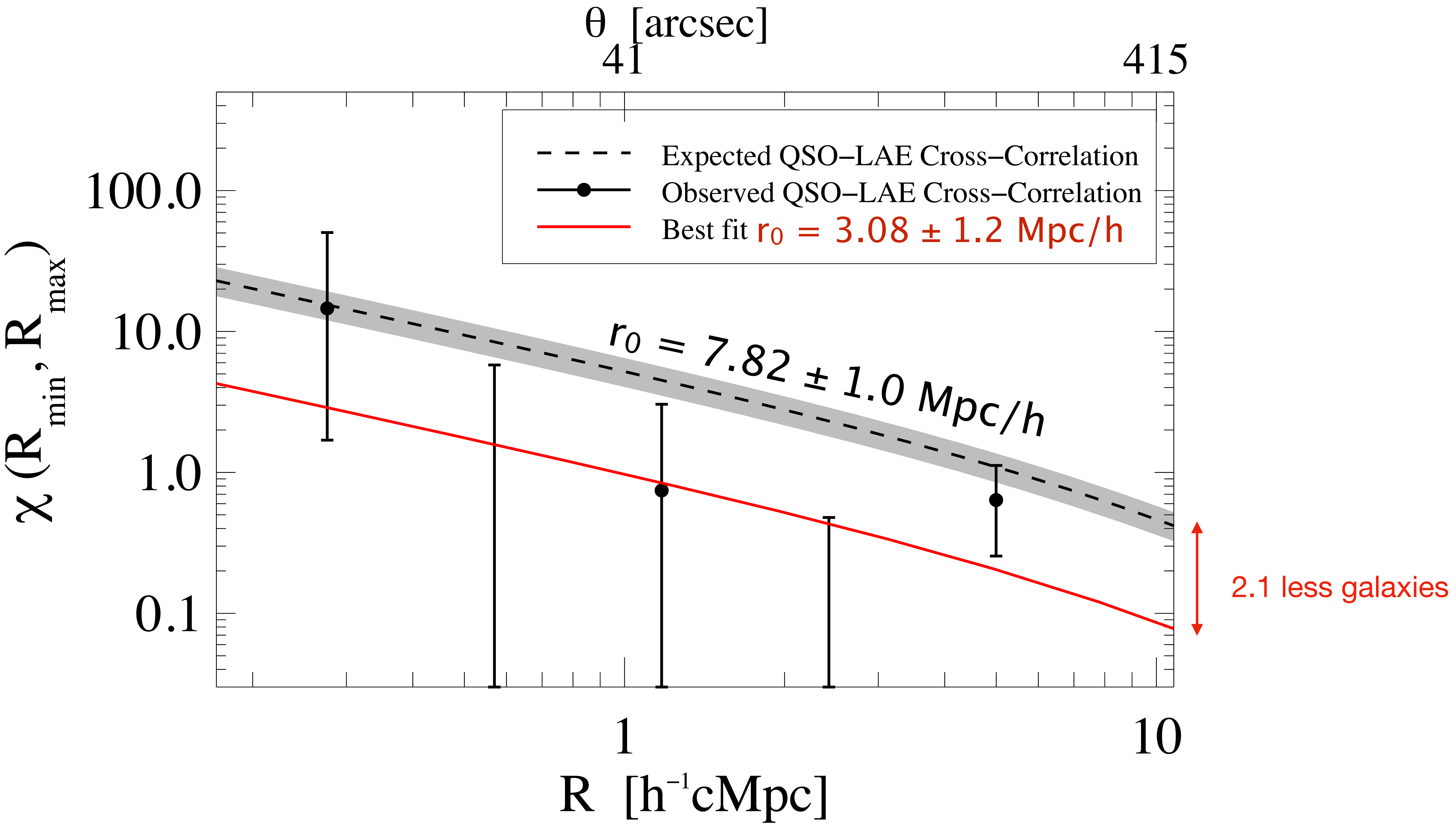


Projected galaxy-quasar cross-correlation function based imaging of 17 quasar fields at z~4

A positive Quasar-galaxy cross-correlation function, consistent with a power-law shape indicative of a concentration of LAEs centered on quasars.

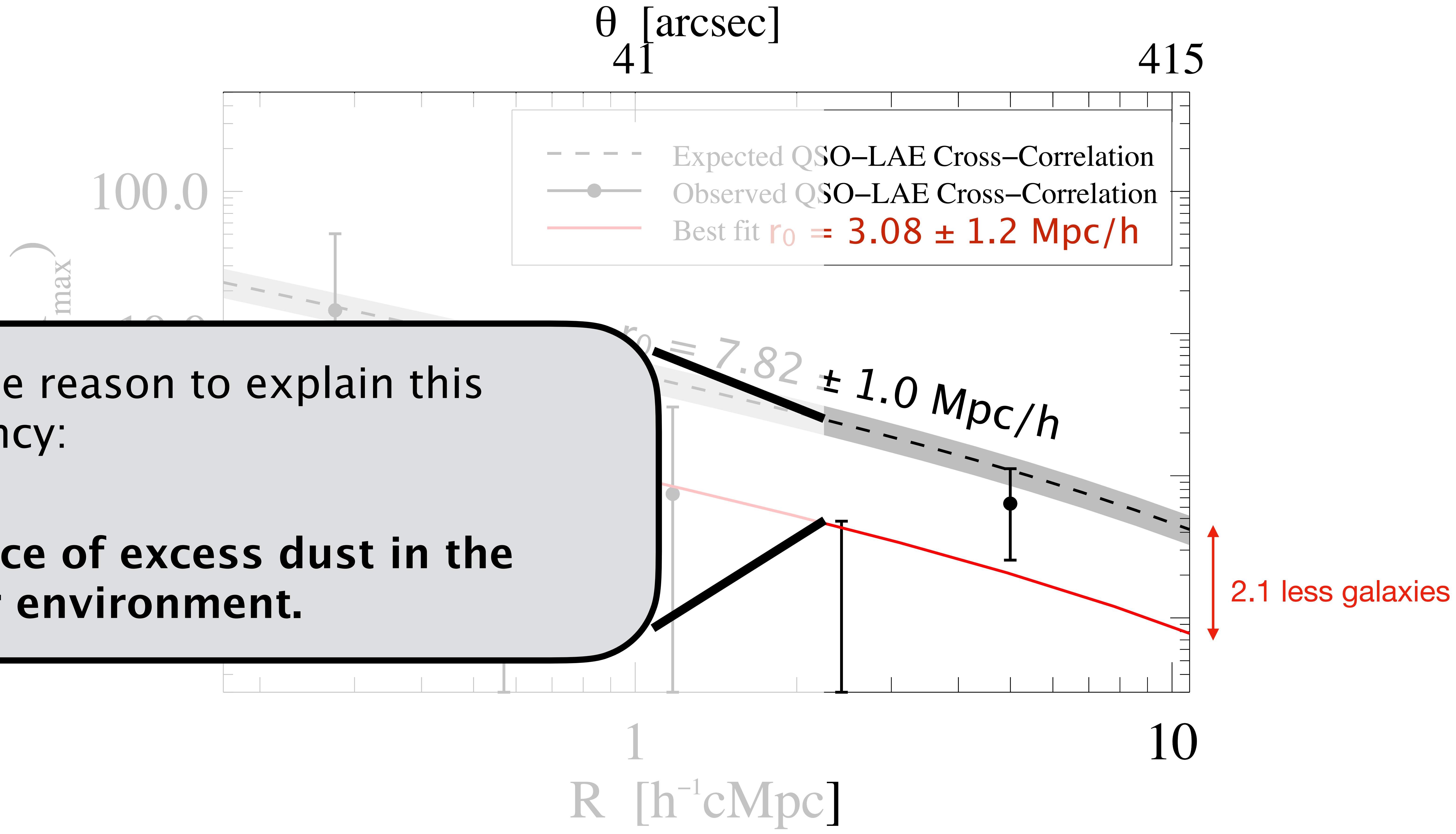
Garcia-Vergara et al. 2019

Comparison with the expectation from a linear bias model



Detections of slightly less galaxies than the expected

Comparison with the expectation from a linear bias model



A possible reason to explain this discrepancy:

- **Presence of excess dust in the quasar environment.**

Clustering also provide insights about physical processes associated with different populations

Observations of 17 QSO fields at z~4 using ALMA to detect dusty galaxies

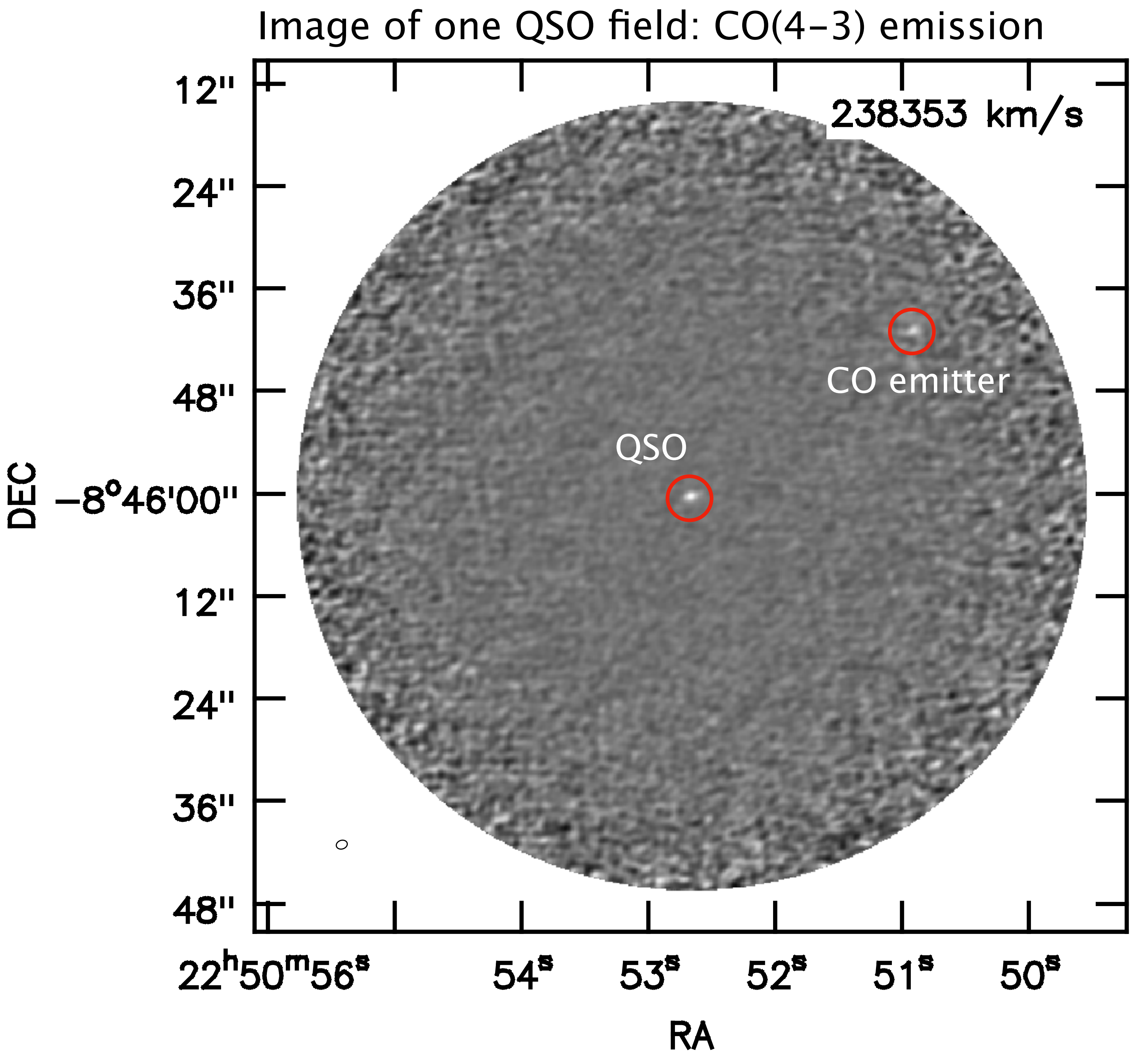
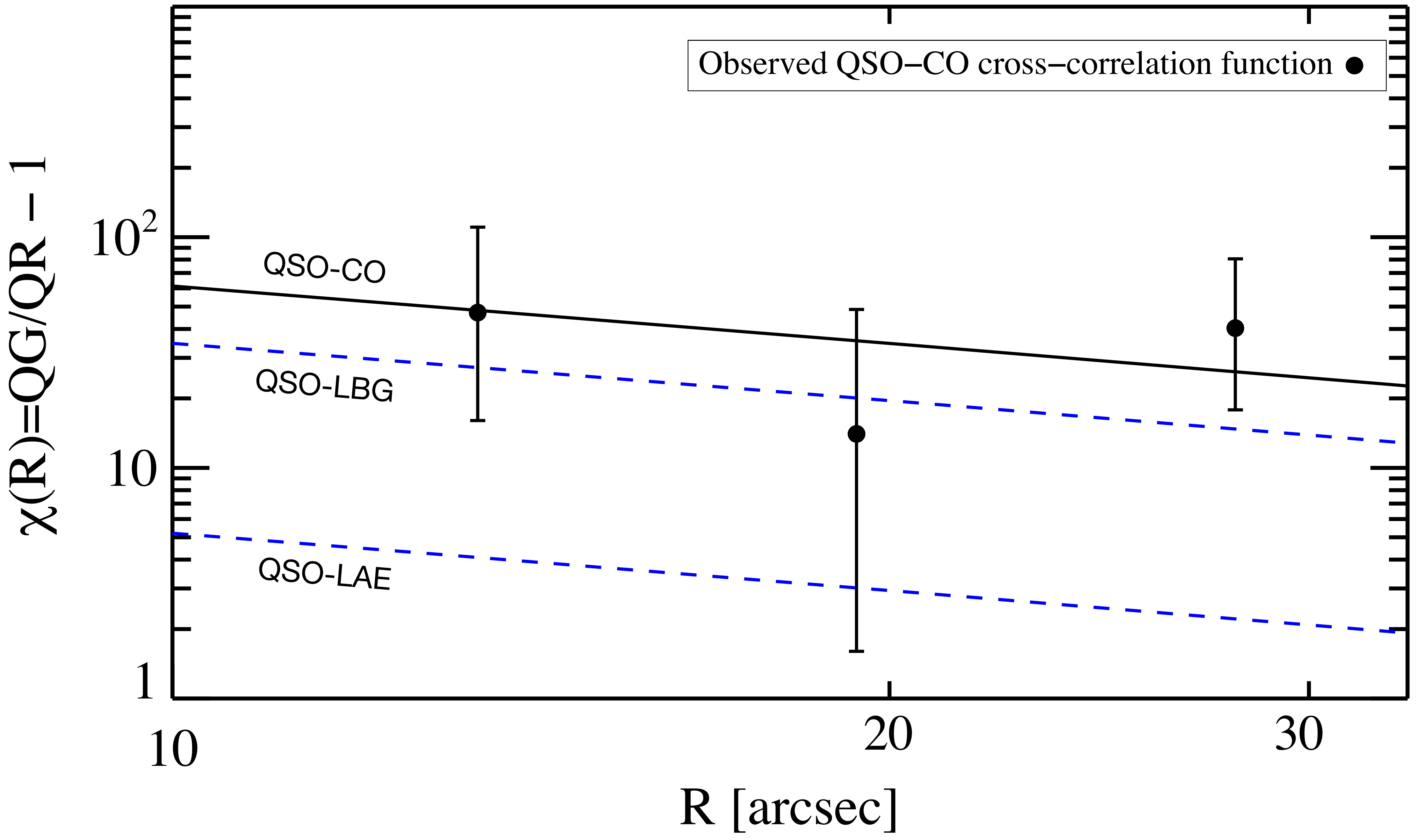


Image of one QSO field: CO(4-3) emission

Projected galaxy-quasar cross-correlation function based imaging of 17 quasar fields at z~4, for dusty galaxies



García-Vergara+in prep.

Clustering also provide insights about physical processes associated with different populations

- QSOs trace massive structures in the early universe (based on the QSO auto-correlation function and on the galaxy-quasar cross-correlation function: two independent methods).

- Clustering techniques overcome complications that lead to small statistics results.

- Overdensities of dusty galaxies around them are significantly higher than what we find for optical galaxies. This may indicate:
  - Since dusty galaxies are more strongly concentrated around quasars, they may be more massive than optically detected galaxies.
  - Presence of excess dust in galaxies near QSOs.

From Lecture 1

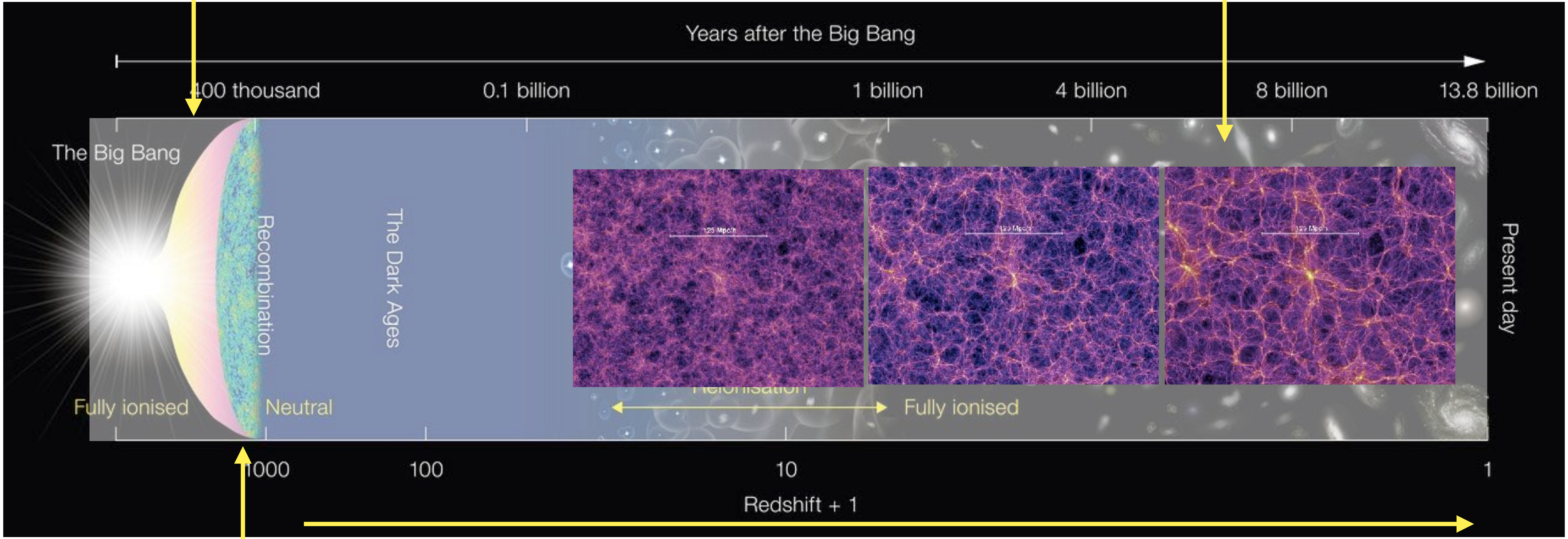
Some interesting questions

- What is large-scale structure?
- How the matter is distributed in the universe?
- Have some regions in the universe more matter than other regions?
- Does the matter distribution evolve over time?
- How can we “measure” and quantify the distribution of matter over a volume?
- Why is it important to quantify?
- What information can we get from studying the matter distribution?
- Can we really learn about galaxy evolution with this?

Large-scale structure in the universe

Inflation

Dark matter distribution

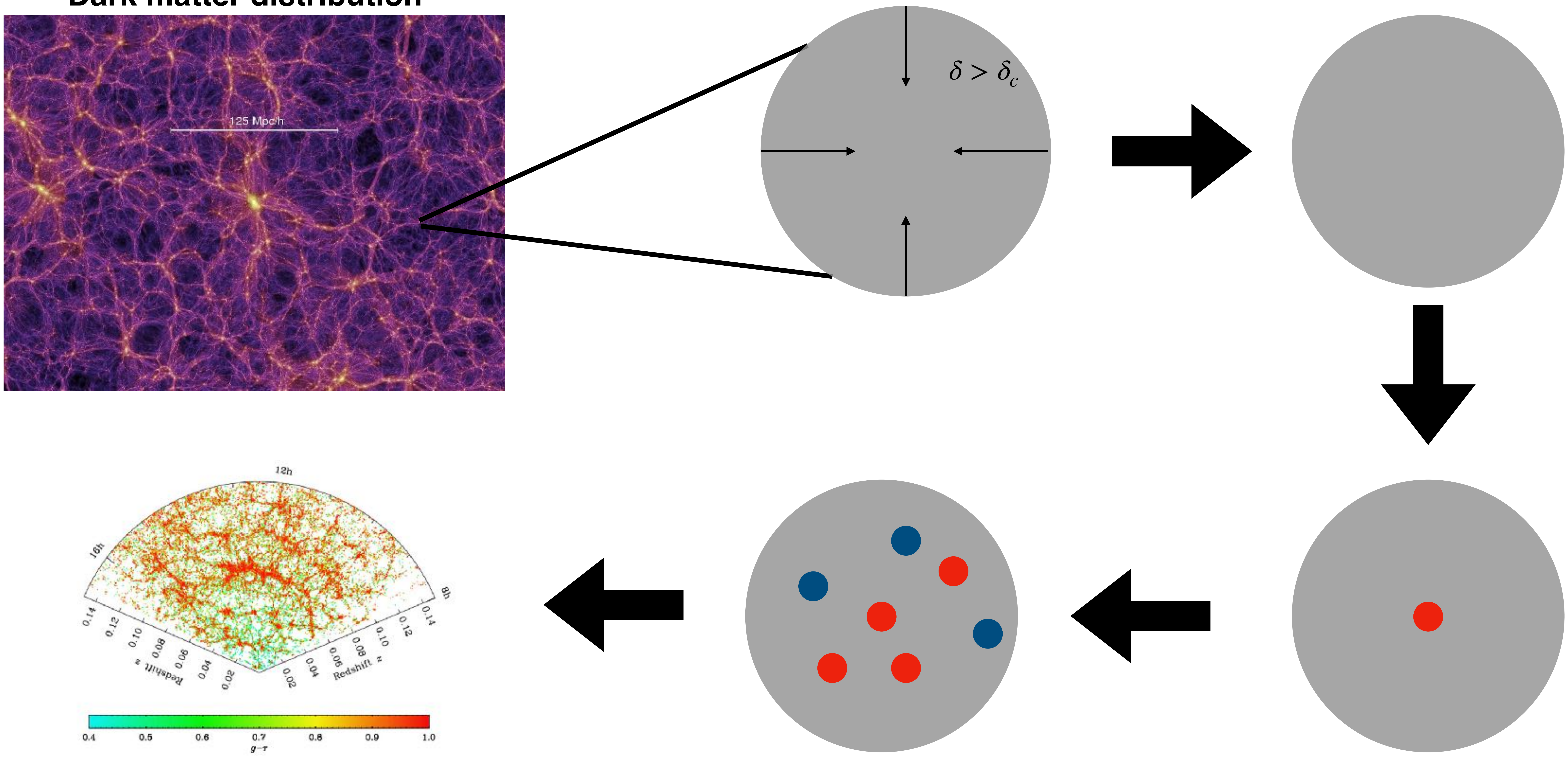


CMB

Universe is expanding and Density fluctuations evolve into structures we observe: galaxies, clusters, super-clusters.

Large-scale structure in the universe

Dark matter distribution



Measuring the large-scale structure

The two-point correlation function ξ(r) is defined as a measure of the excess probability dP, over a random occurrence of finding a galaxy in a volume element dV at a separation r from another galaxy.

$$\boxed{dP = \bar{n}[1 + \xi(r)]\mathrm{dV}}$$

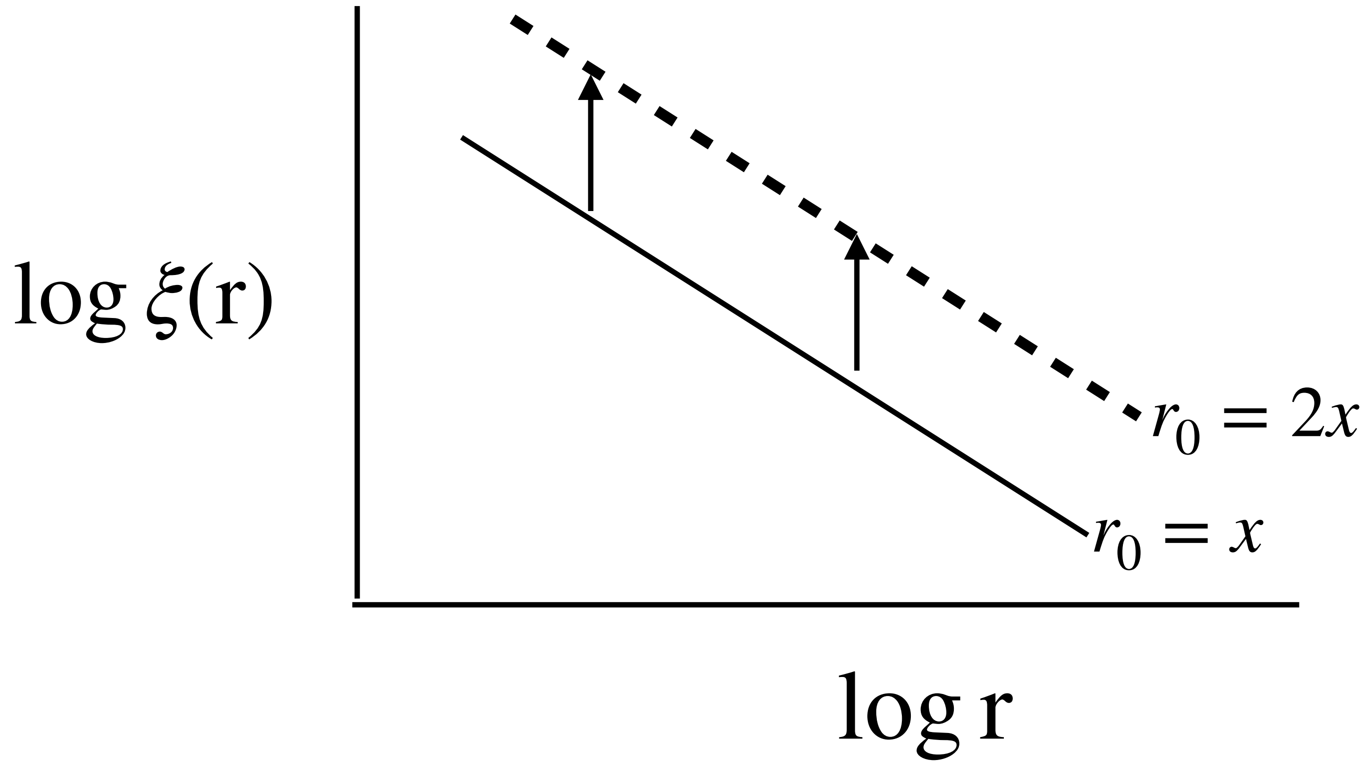
where n is the mean number density of the galaxy sample in question.

Observations indicate that ξ(r) is well described by a power-law: $\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$

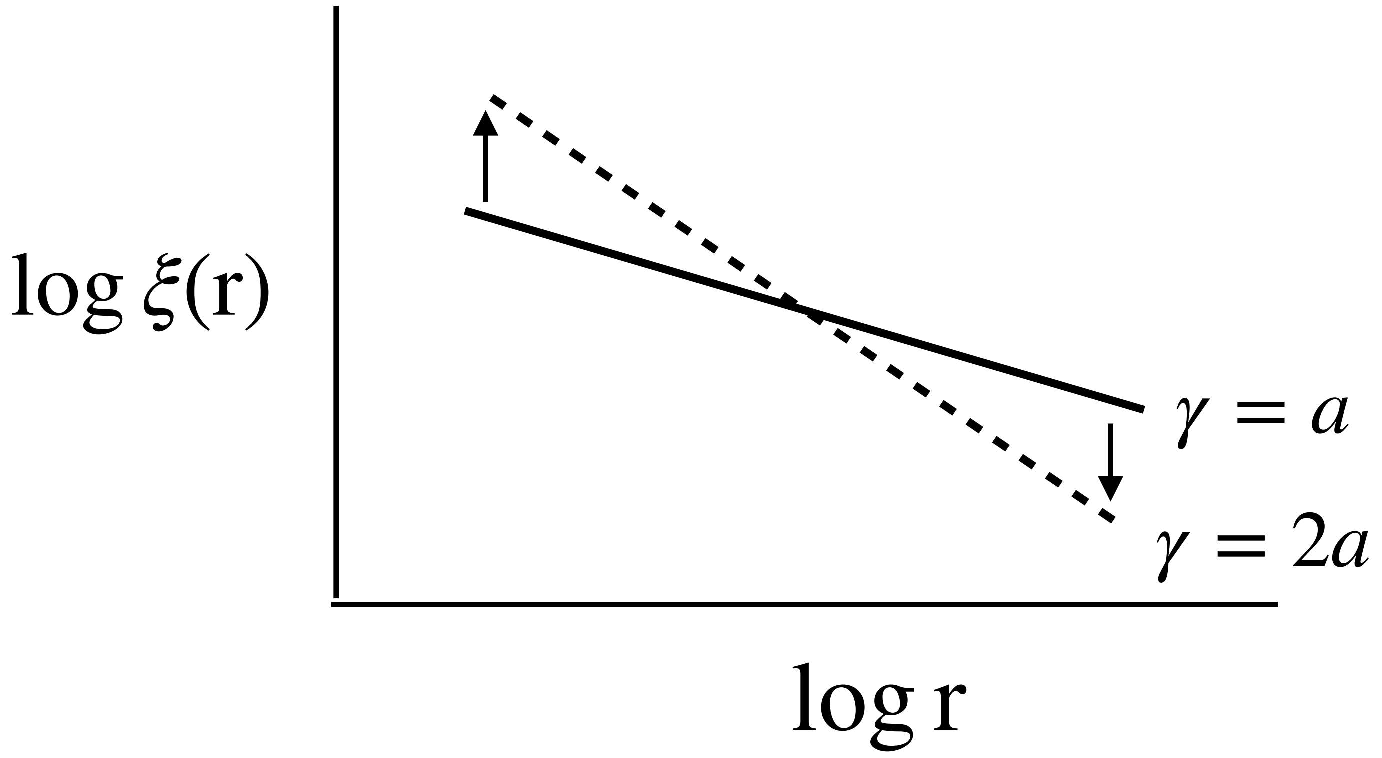
r_0 correlation length
γ slope (typically 1.8)

→ Strong clustering at small scales and weak clustering at large scales.

Effect of the correlation length:



Effect of the slope:



Measuring the large-scale structure

If you have
2D positions

↓

Angular correlation
function $\omega(\theta)$

(Integrated over all the redshift space)

↓

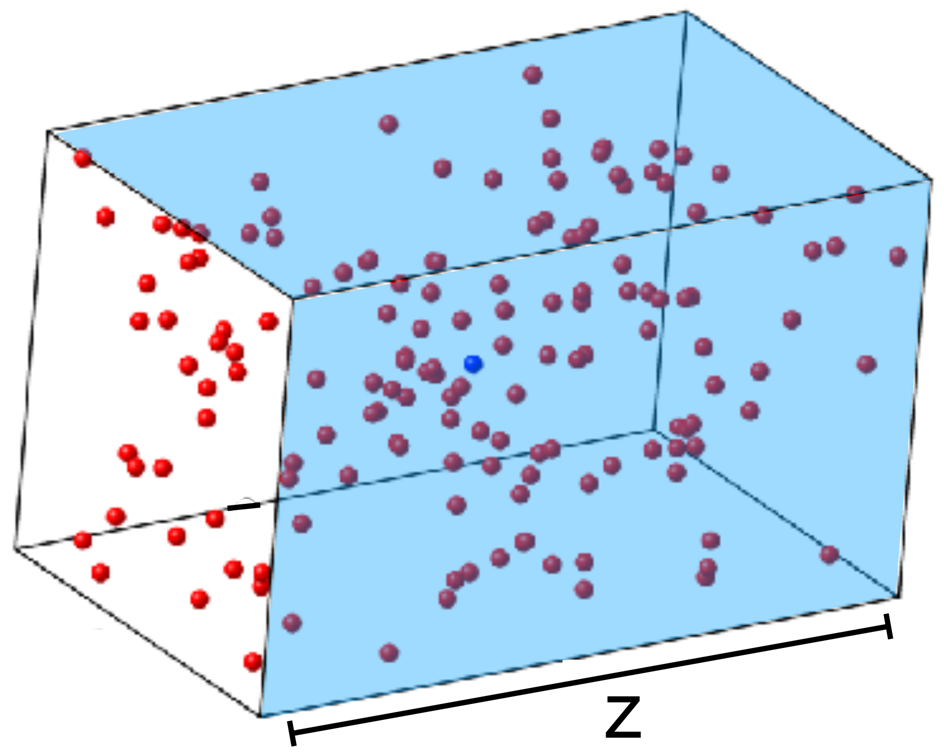
Fit the measurement
to get A and β

↓

Assumptions about
the z distribution

↓

Get r_0, γ



If you have
3D positions

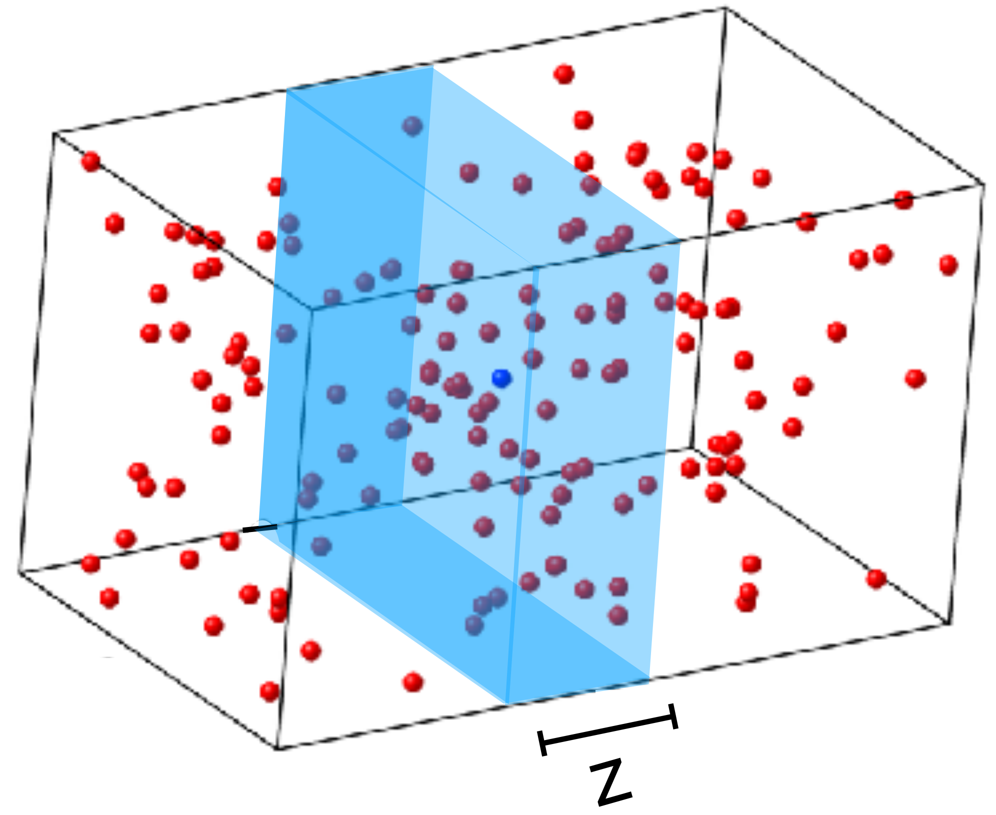
↓

Projected correlation
function $\omega(R)$

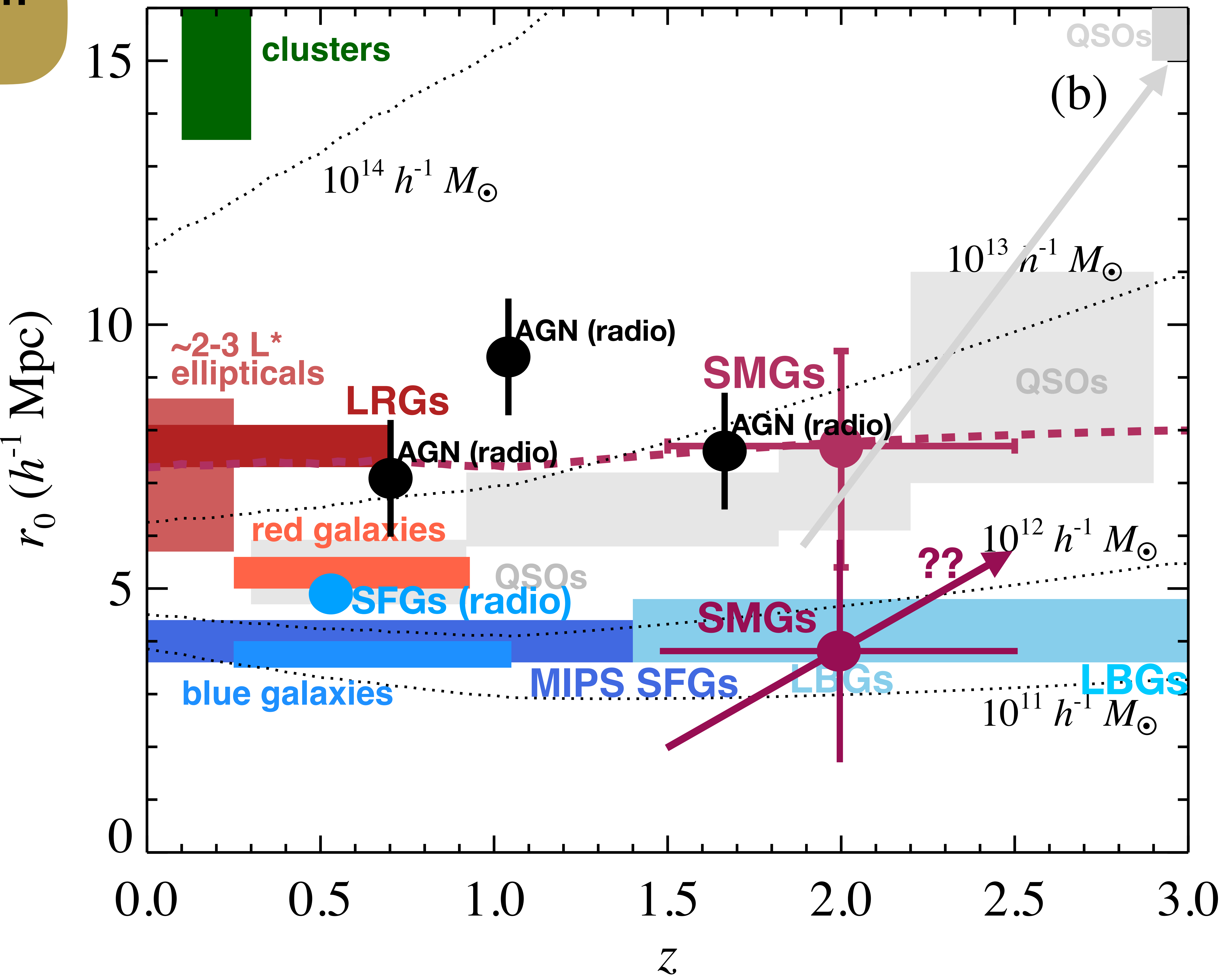
(Integrated over a narrow redshift space)

↓

Fit the measurement to
get r_0, γ

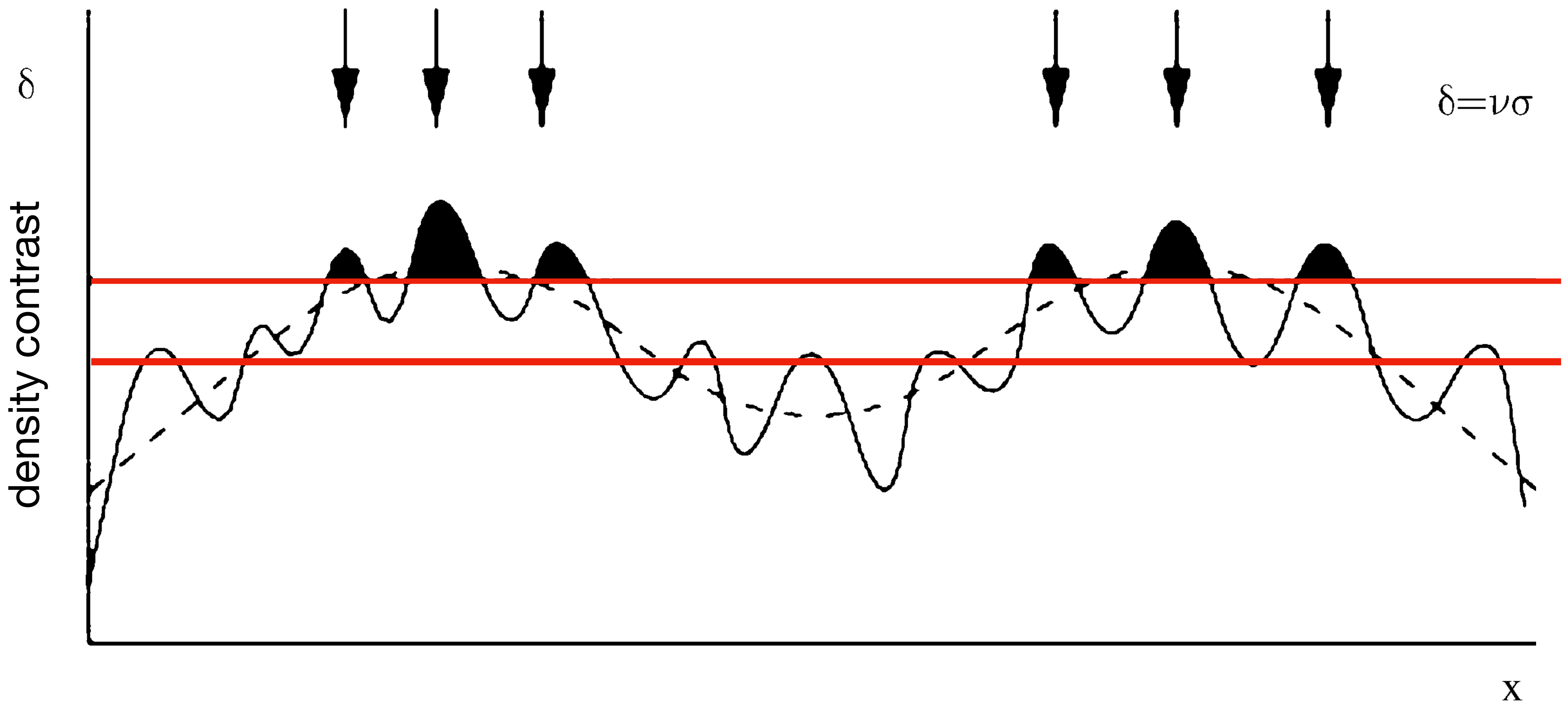


Clustering dependence on galaxy properties



(Adapted from Hickox et al. 2012)

Dark matter halo and galaxies are biased tracers of dark matter



Threshold at which a massive DM halo can be formed

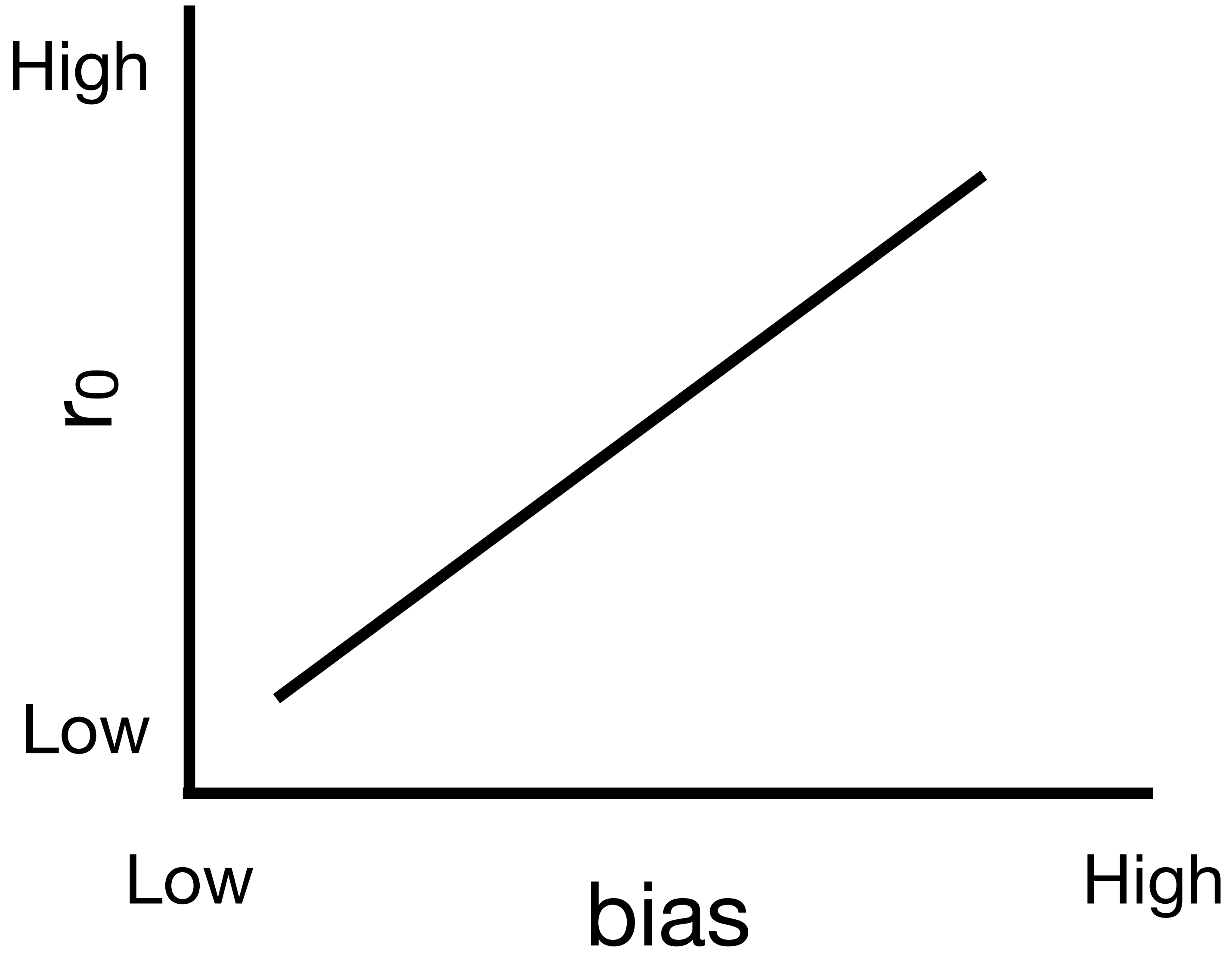
Threshold at which a less massive DM halo can be formed

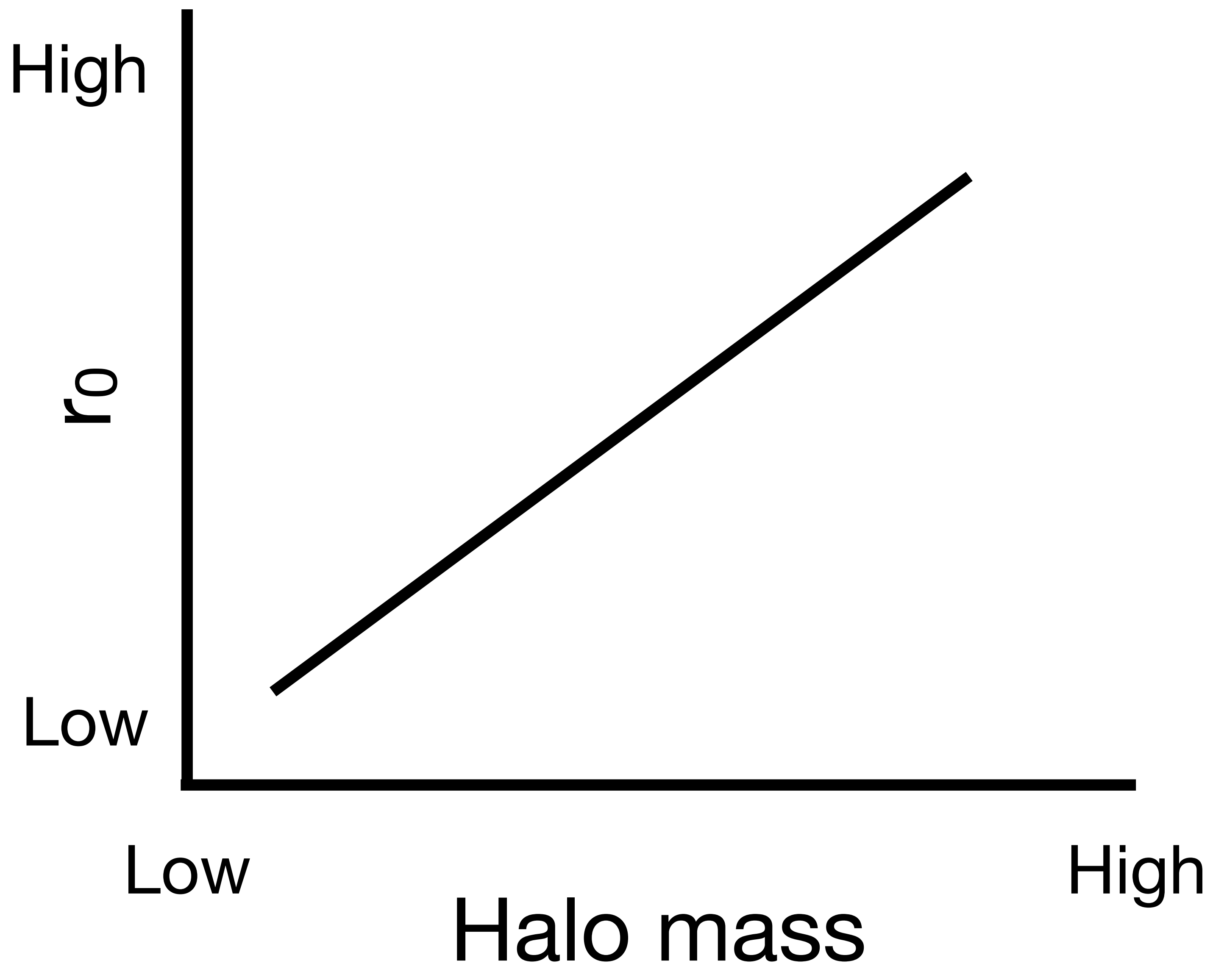
The most massive dark matter halos are expected to trace worst the dark matter density distribution (because they only form where large scale fluctuations are high) and be the most biased. Less massive halos will trace better the dark matter distribution, and they are less biased.

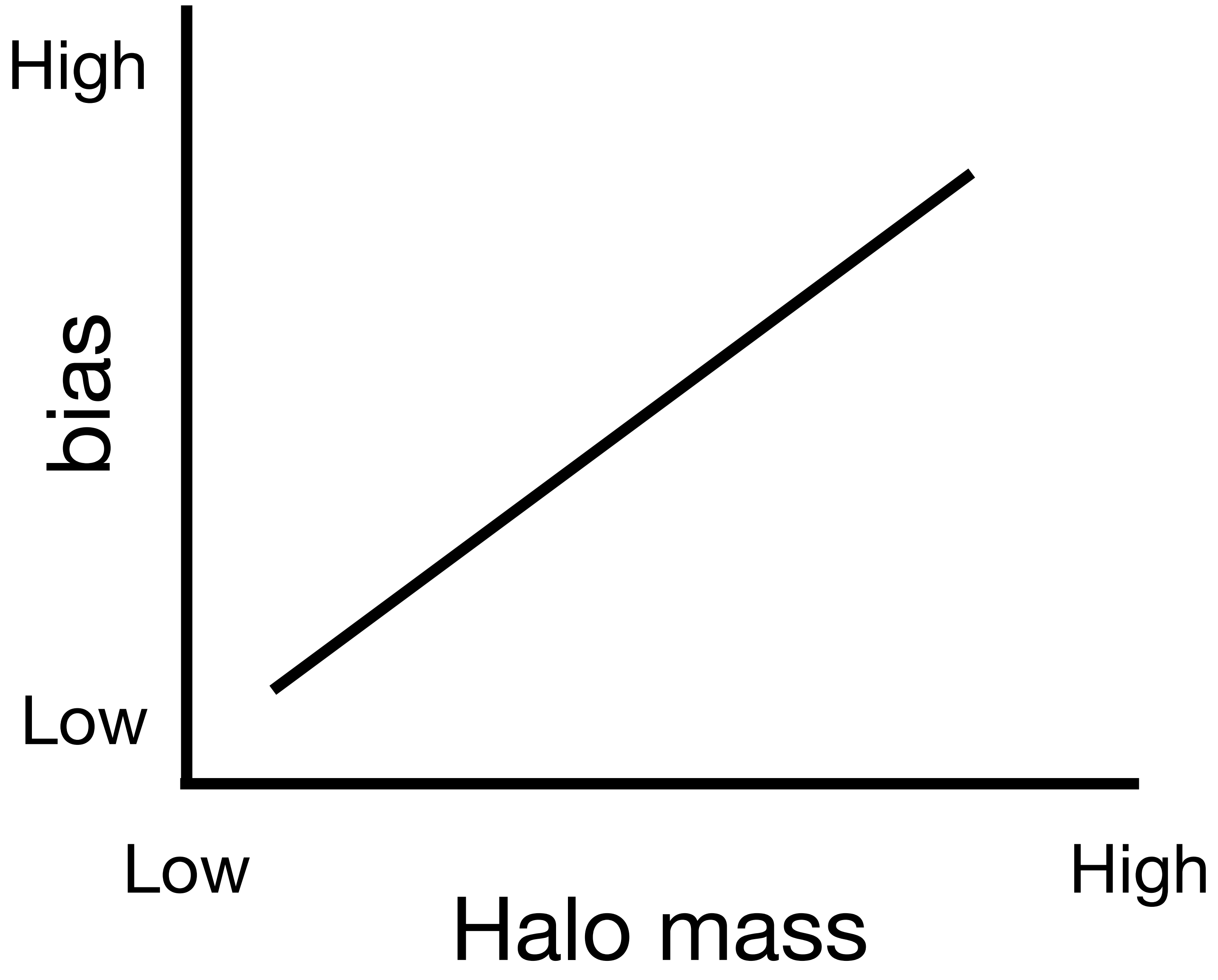
Relation between correlation length, bias and dark matter halo mass

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \qquad b_g = \sqrt{\xi_g/\xi_{DM}}$$

- **At a fixed time (redshift)**

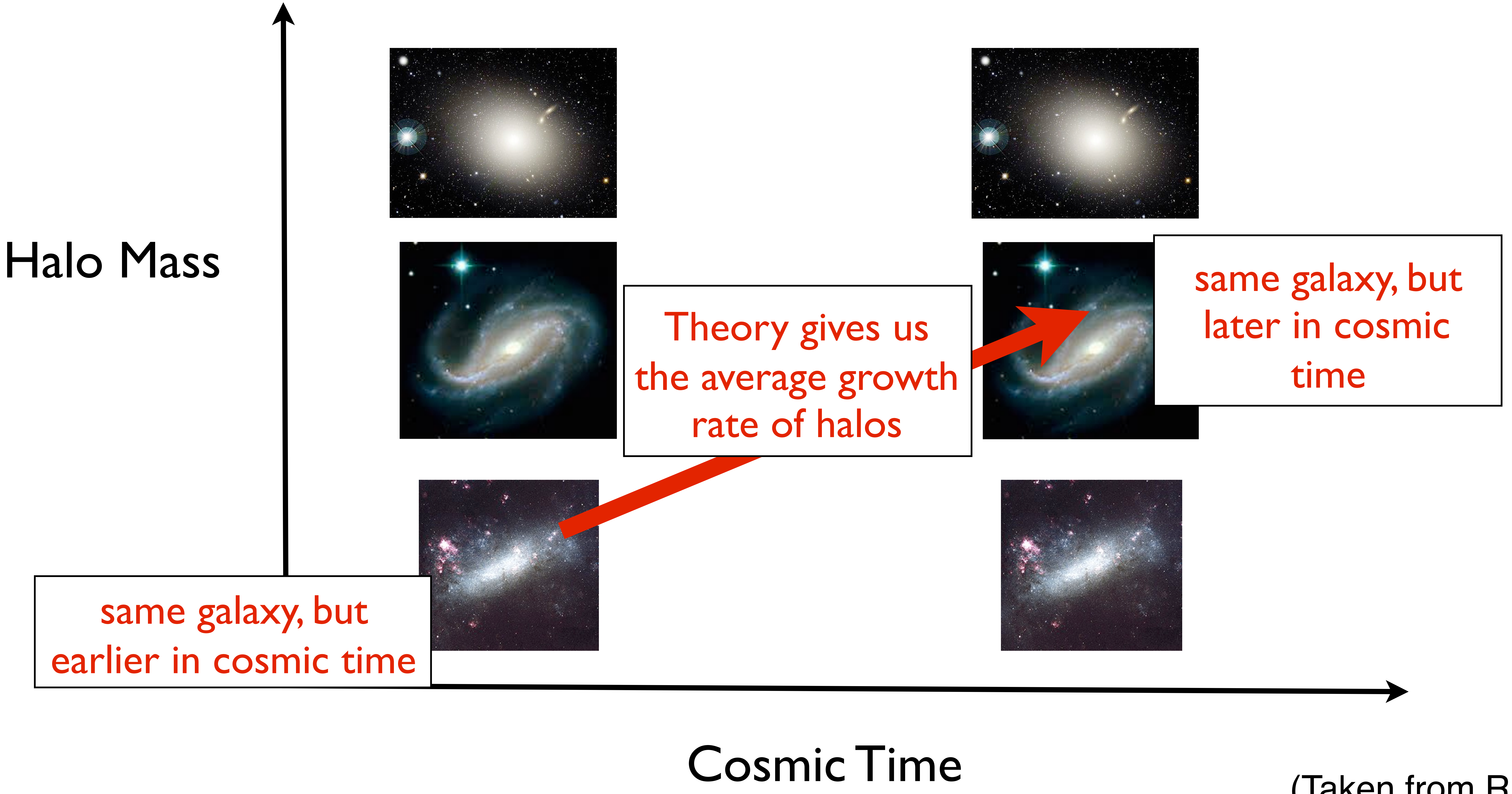






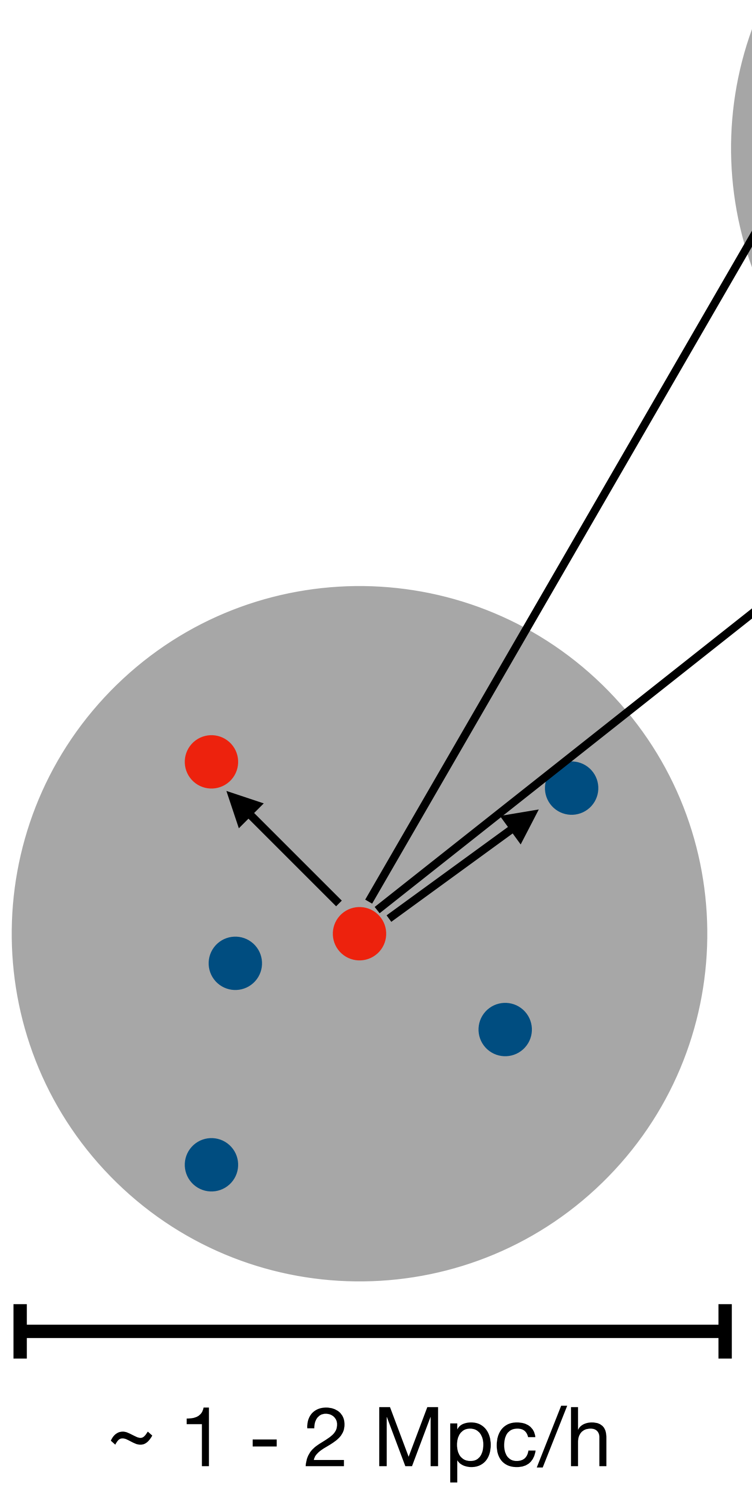
We really can learn about galaxy evolution using clustering measurements

It provides us a powerful tool for tracing the same population of galaxies through cosmic time.

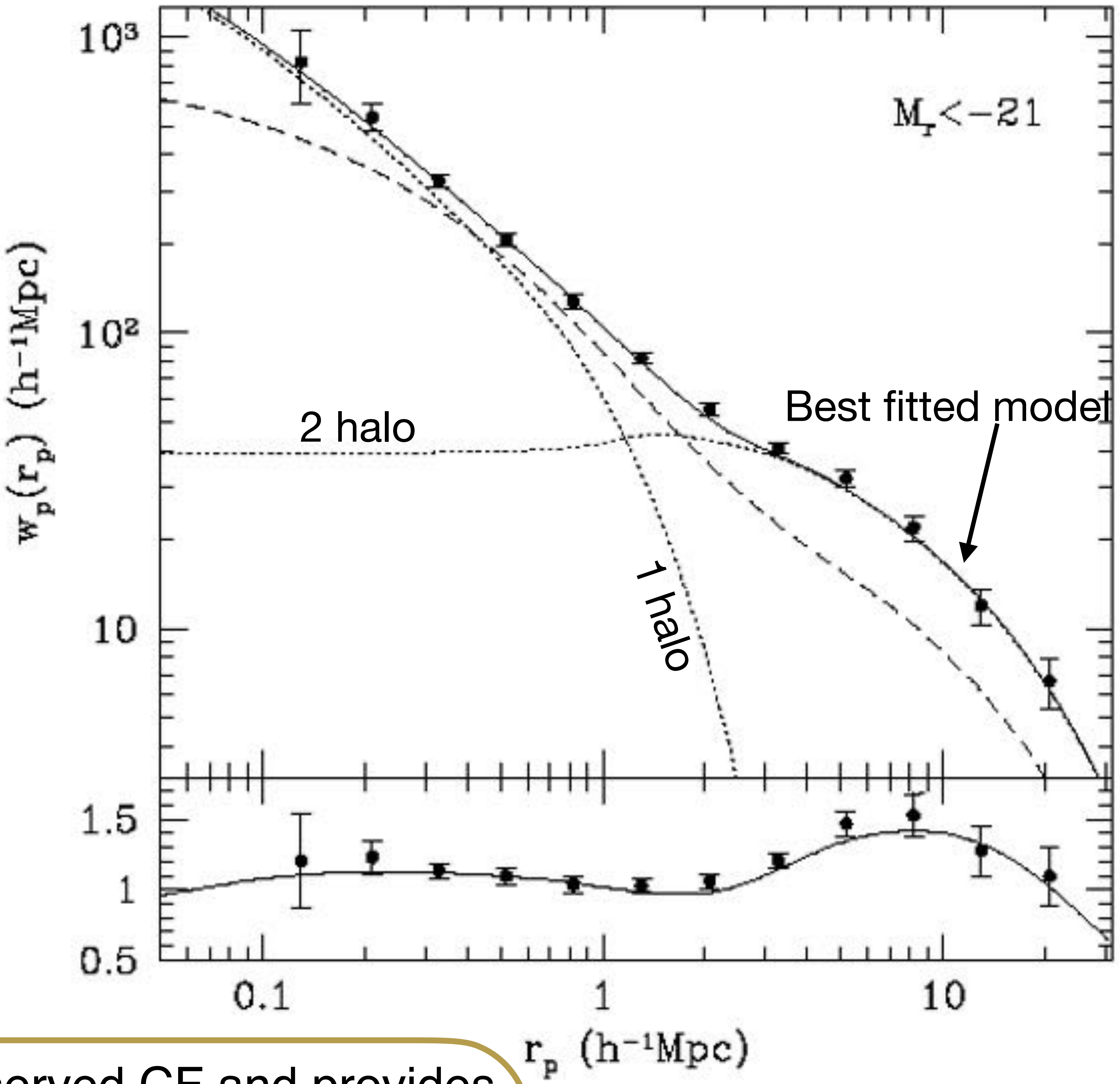


(Taken from R. Bouwens' Lectures)

We really can learn about physical processes in galaxies using clustering measurements



The combination of the 1-halo and 2-halo terms result in a power law-shape for the galaxy correlation function.



HOD model facilites the interpretation of the observed CF and provides constraints on models of gal formation and evolution within halos.

We really can learn about properties of galaxies in different environments using clustering measurements

- Presence of excess dust in galaxies in QSOs environments?