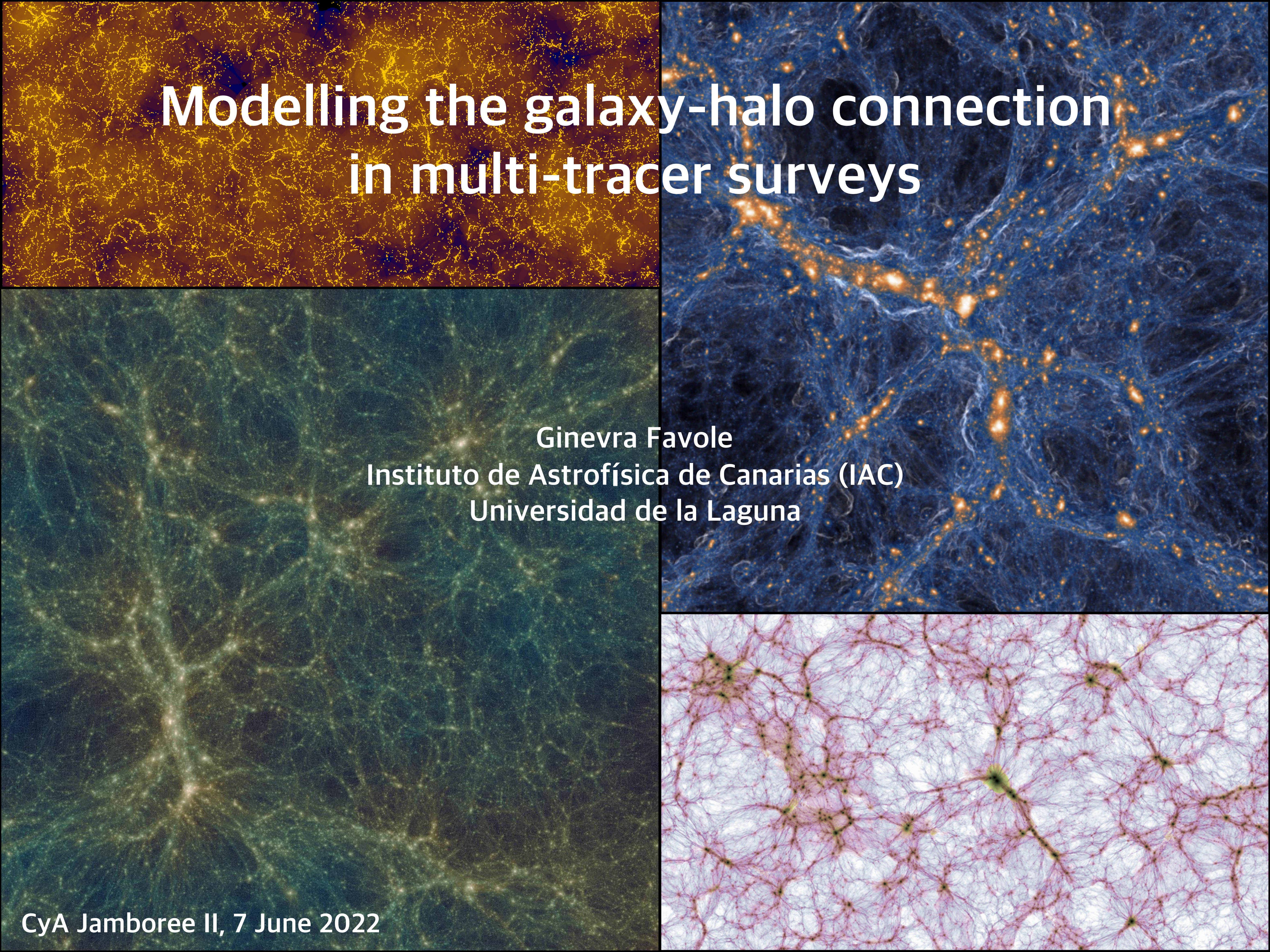


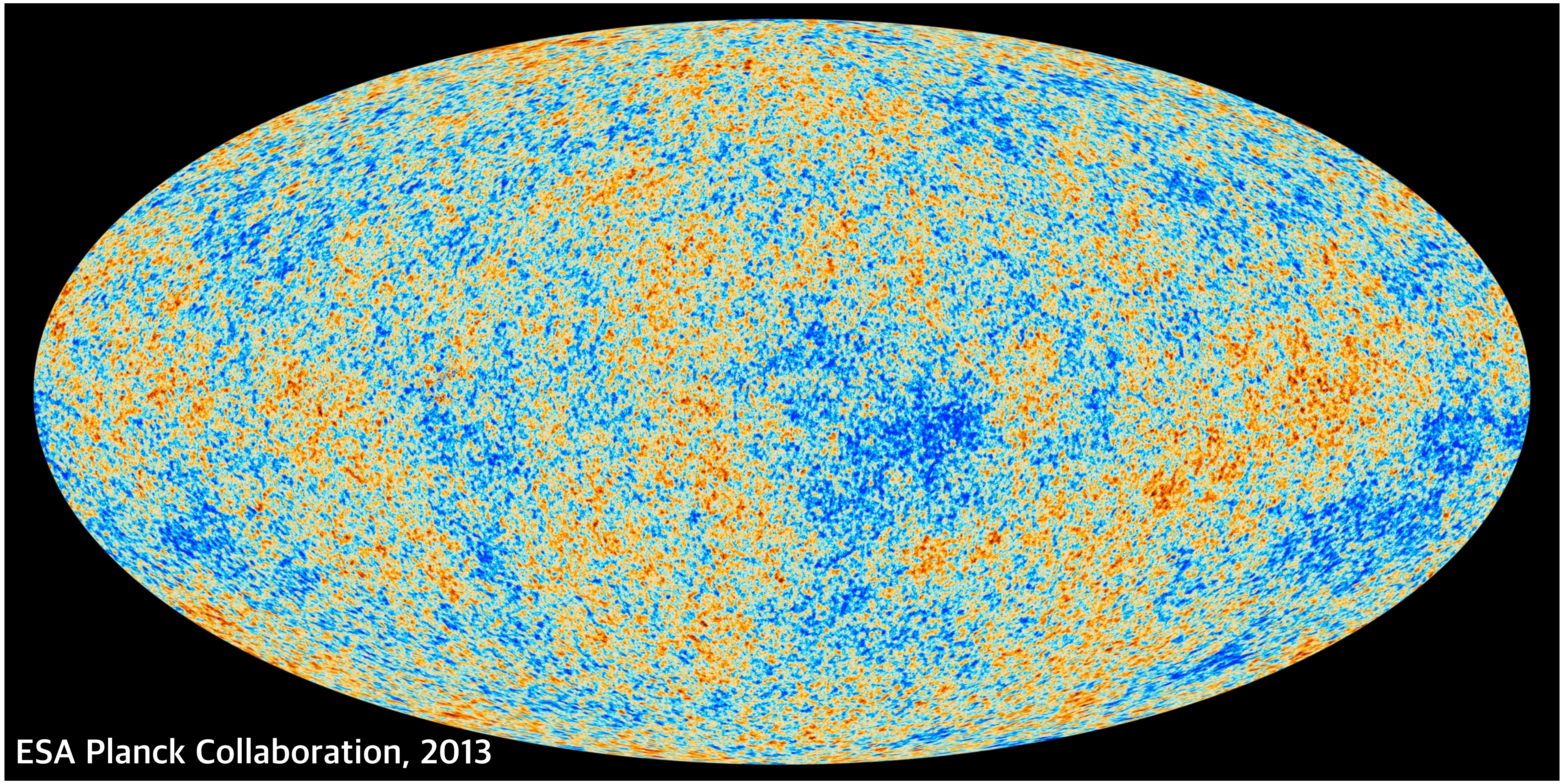
Modelling the galaxy-halo connection in multi-tracer surveys

Ginevra Favole
Instituto de Astrofísica de Canarias (IAC)
Universidad de la Laguna

CyA Jamboree II, 7 June 2022

Motivations

Primordial quantum fluctuations propagate as sound waves leaving their imprint in the CMB. They are the seeds of the **large-scale structure**


ESA Planck Collaboration, 2013

2 cosmological probes: **Galaxy clustering + Weak gravitational lensing**

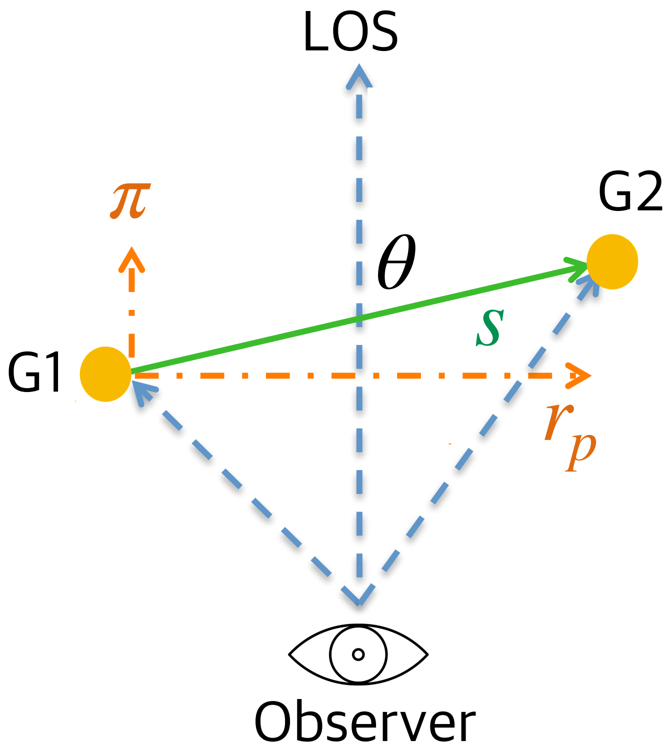
Goals:

- Understand how different populations of galaxies populate their host haloes (**HOD**)
- Probe the stellar-to-halo-mass relation (**SHMR**)
- Constrain **growth of structure f(z)**, galaxy **bias**, **$\sigma_8(z)$, breaking the mutual degeneracies**

Galaxy clustering

2-point correlation function (2PCF)

$$dP = n^2[1 + \xi(s)]dV_1 dV_2$$

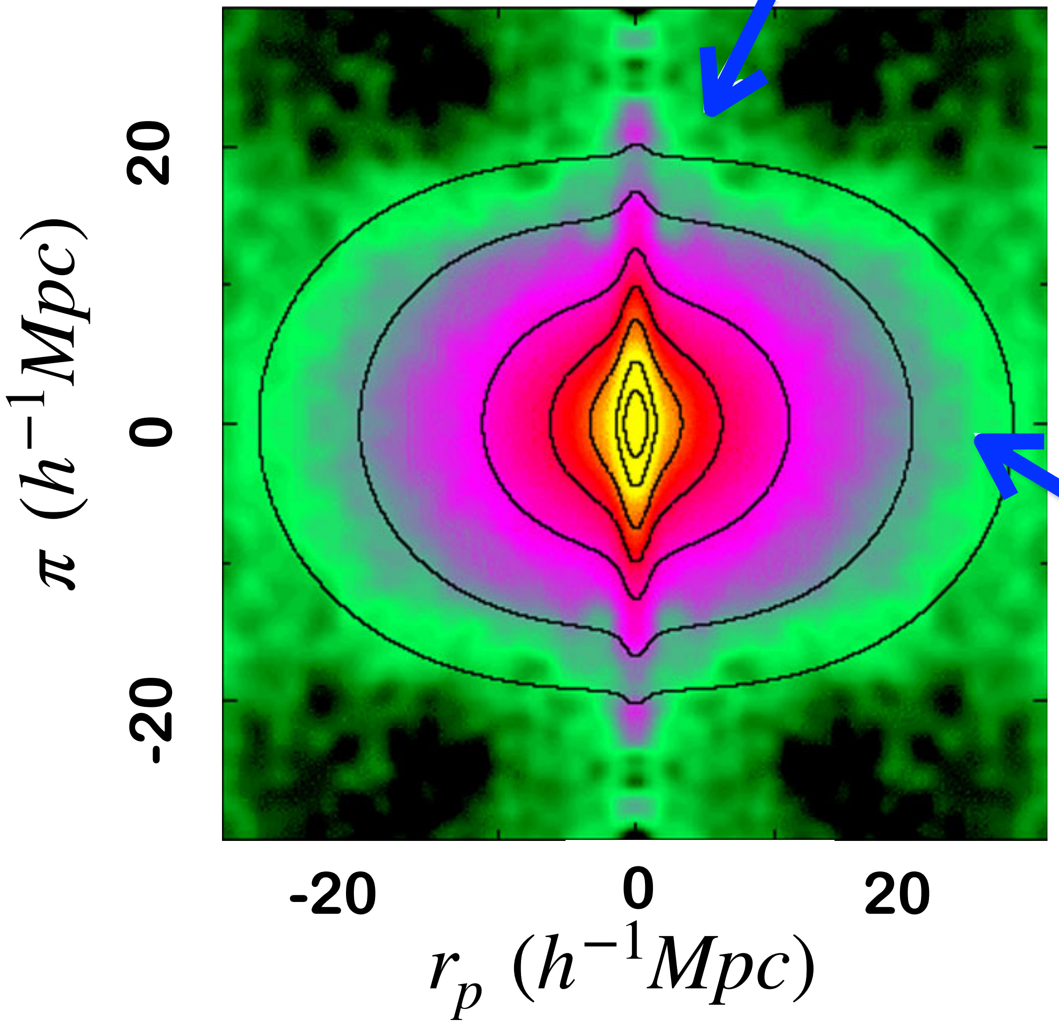


$$s = \sqrt{r_p^2 + \pi^2} \qquad \mu = \cos\theta$$

Landy & Szalay 1993 estimator:

$$\boxed{\xi(s,\mu) = \frac{DD(s,\mu) - 2DR(s,\mu)}{RR(s,\mu)} + 1}$$

z-space distortions (**RSD**)
peculiar velocities of galaxies



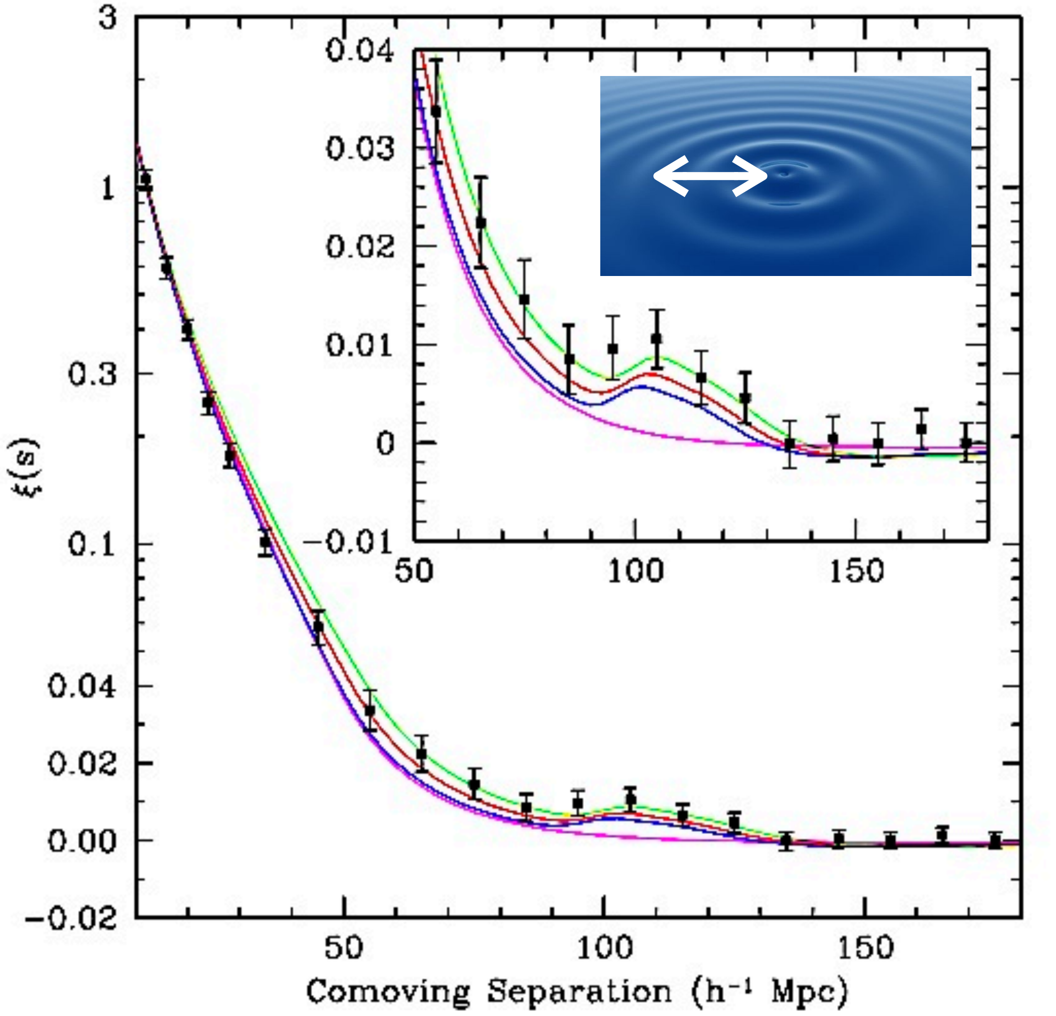
Kaiser squashing,
gravitational infall
of galaxies
towards over
densities

2PCF multipoles by expanding in Legendre polynomials:

BAO~110 Mpc/h standar ruler

$$\xi_l(s) = \frac{2l+1}{2}\int_{-1}^{+1} \xi(s,\mu)P_l(\mu)d\mu$$

l=0 monopole, spherical average
l=2 quadrupole traces satellites, peculiar velocities
l=4 hexadecapole traces peculiar velocities



Eisenstein et al. 2005

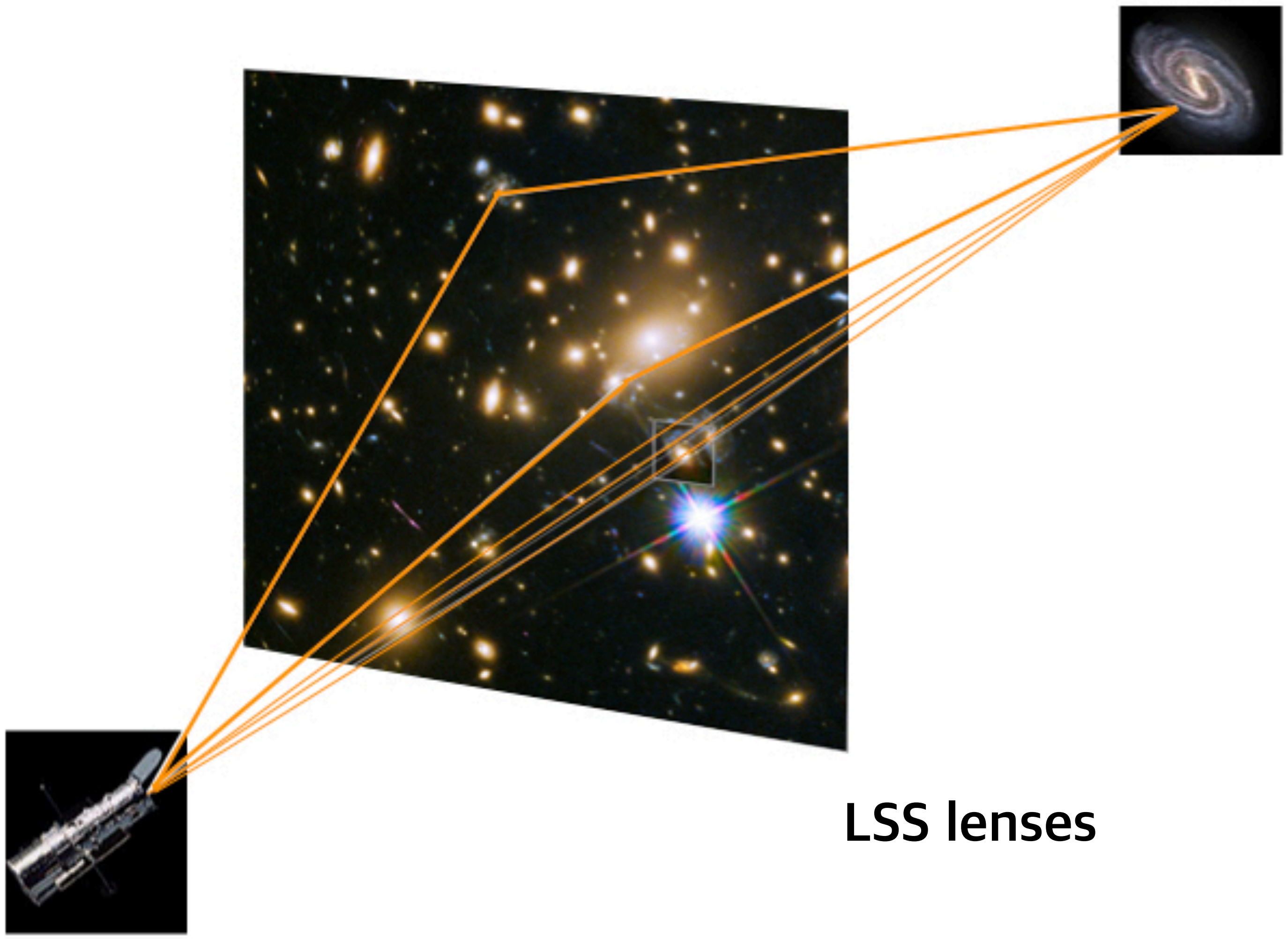
Projected 2PCF mitigates RSD:

$$w_p(r_p) = 2\int_0^\infty \xi(r_p,\pi)d\pi \longrightarrow b(r_p) = \sqrt{\frac{w_p^{gal}(r_p)}{w_p^m(r_p)}}$$

Each 2PCF is sensitive to a physical process/effect happening on a particular scale

Weak gravitational lensing

Cosmic shear: distortions produced by massive objects of the LSS on the light path of distant galaxies

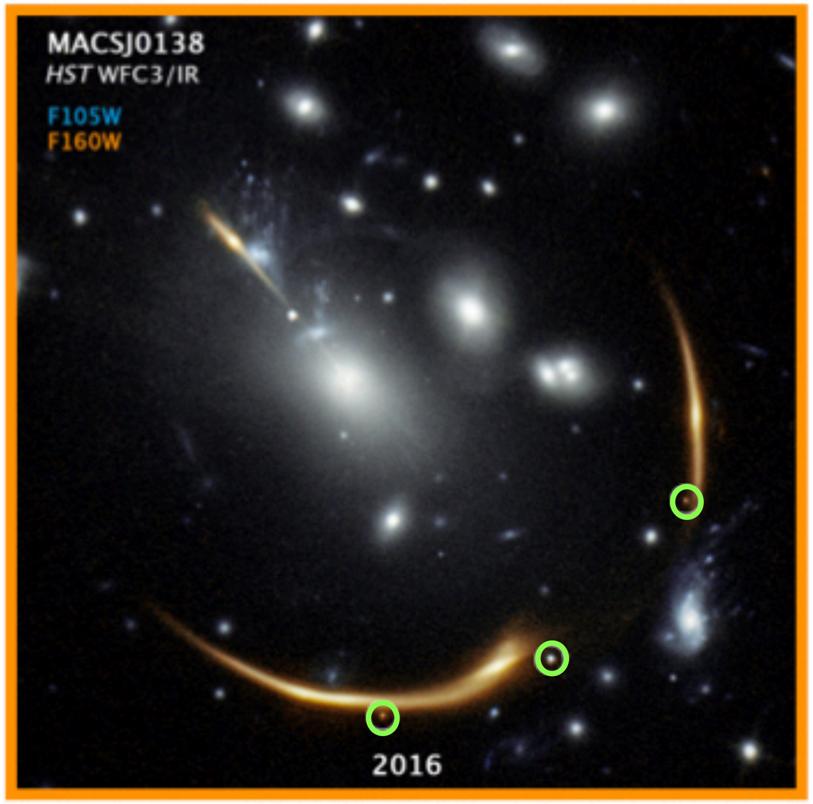


Source

LSS lenses

Observer

Lensed image



Supernova lensed 3 times by massive cluster

Weak effect, average over a **large N of sources** —> Euclid >30 sources/arcmin^2

Differential surface density as a function of the transverse separation between lenses and sources:

Mandelbaum et al. 2006 estimator:

$$\boxed{\Delta\Sigma_{\rm gm}(r_p) = \frac{\sum_{\rm l,s} w_{\rm s} e_t \Sigma^{-1}_{\rm crit}(z_{\rm l}, z_{\rm s})}{2\mathrm{R}\sum_{\rm l,s} w_{\rm l,s}}}$$

l=lenses; s=sources

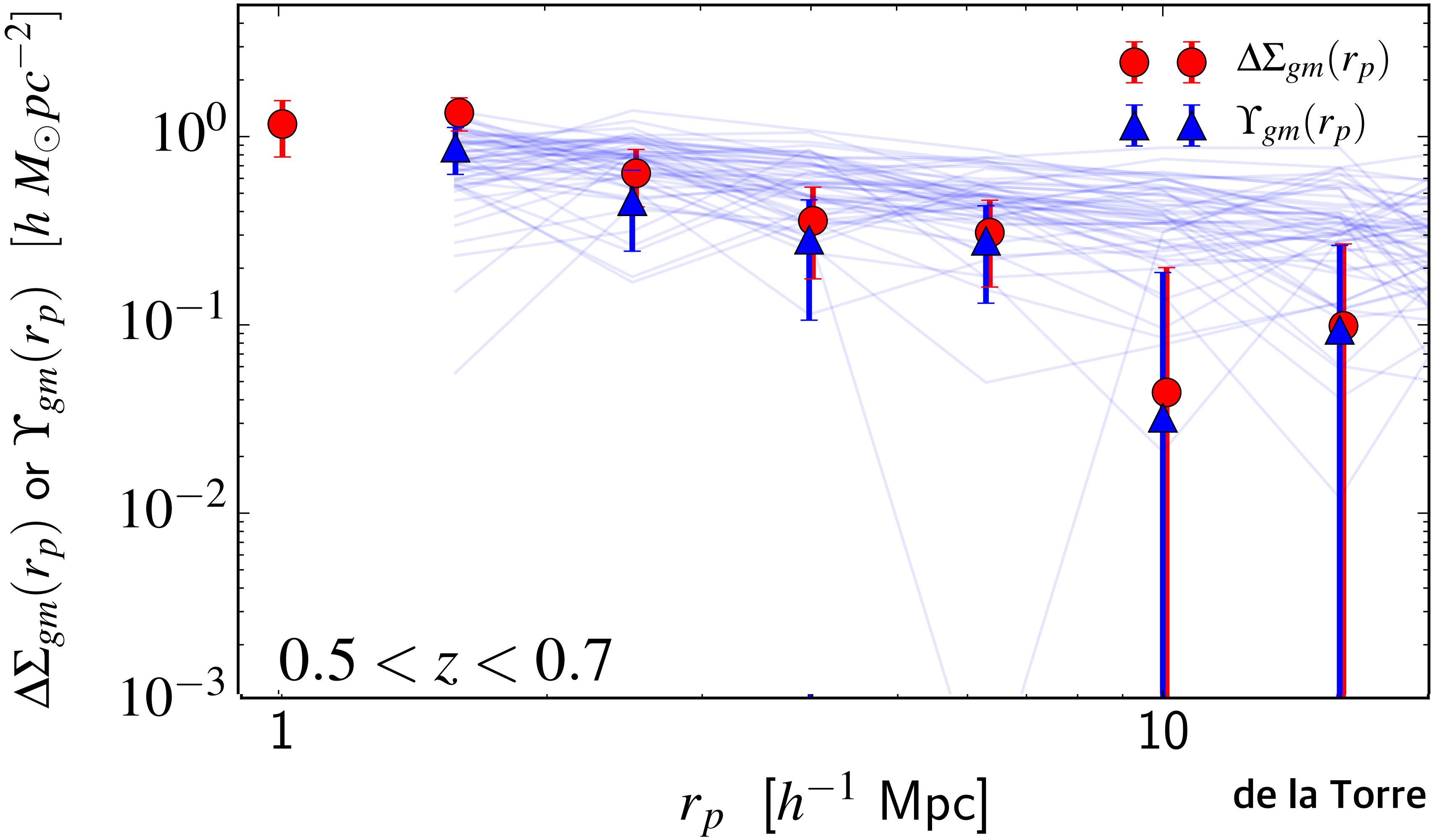
e_t tangential shear (from source ellipticities)

$w_{\rm s}$ source shape weights

$$\Sigma_{\rm crit}(z_{\rm l}, z_{\rm s}) = \frac{c^2 D_{\rm s}}{4\pi G D_{\rm l} D_{\rm ls}}$$

$w_{\rm l,s} = w_{\rm s}\Sigma^{-2}_{\rm crit}(z_{\rm l}, z_{\rm s})$ source-lens weights

R~0.87 shear responsivity



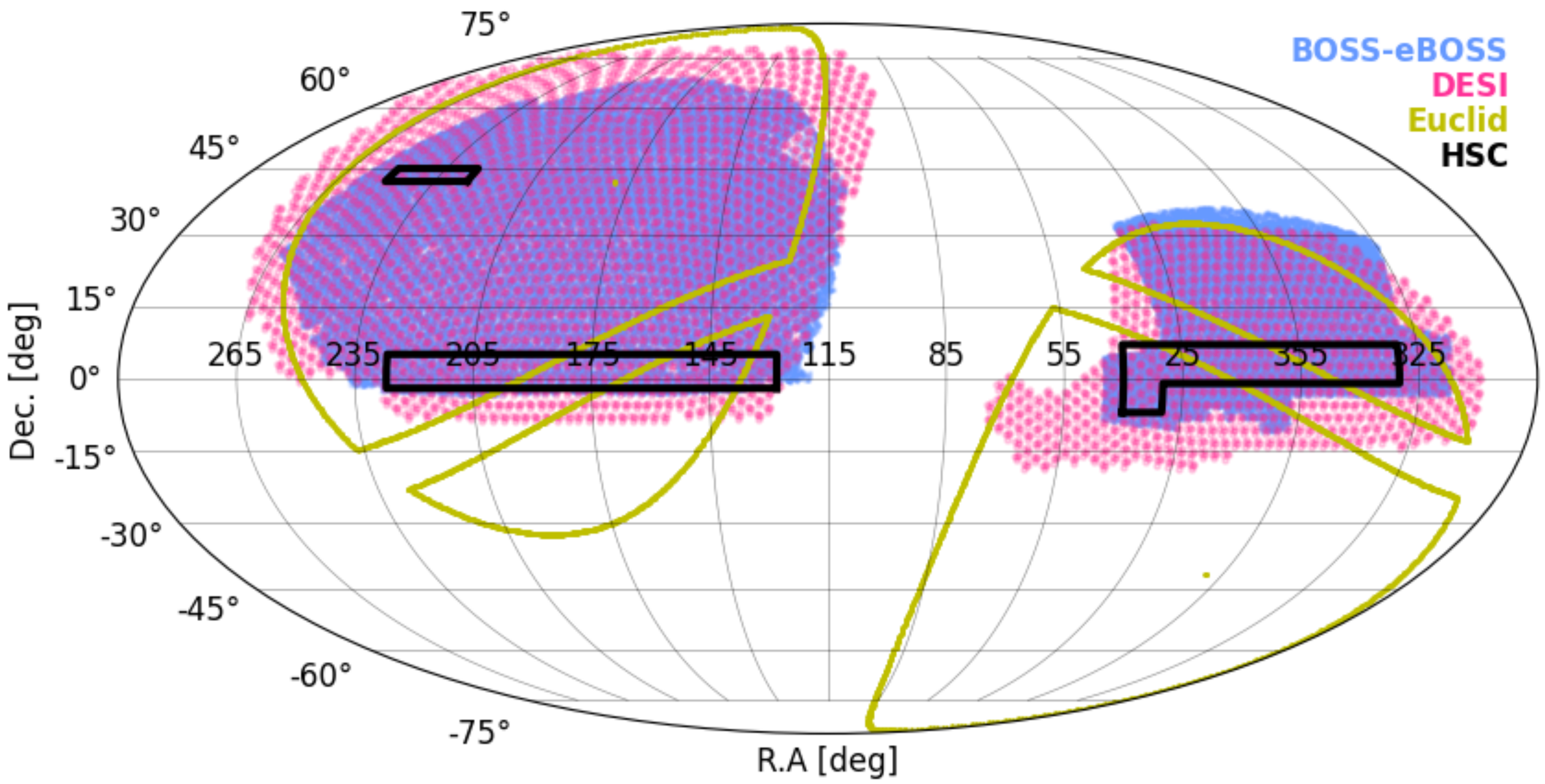
de la Torre et al. 2017

Annular measurement:

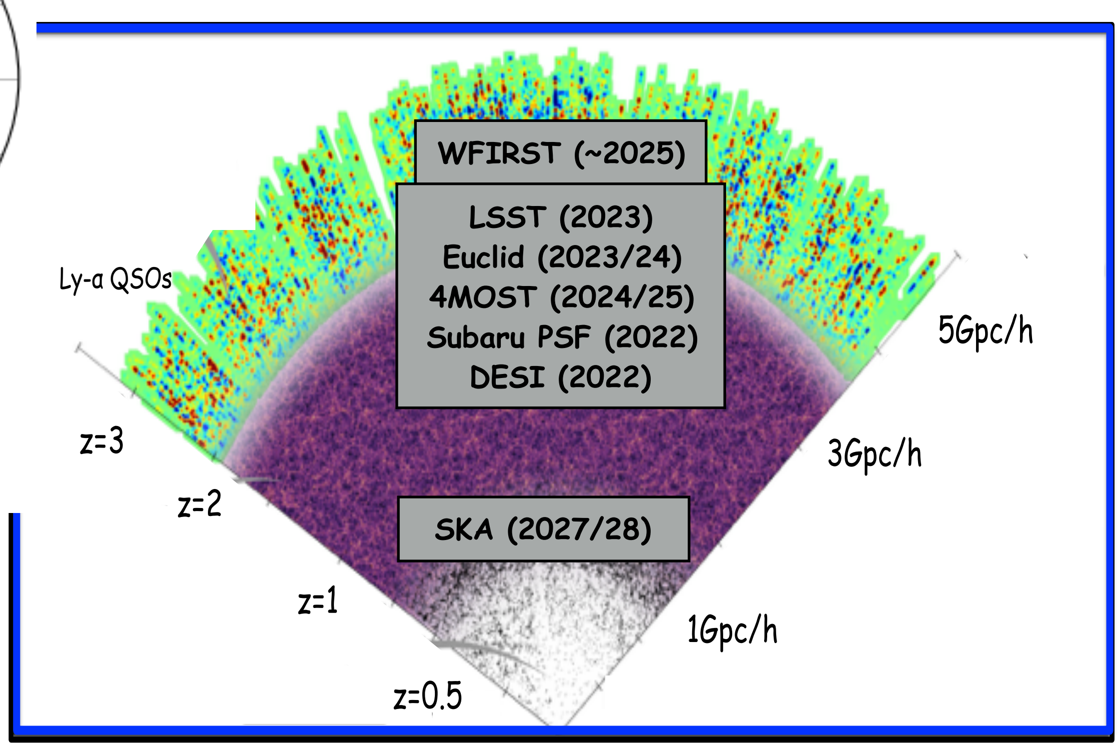
$$Y_{\rm gm}(r_p) = \Delta\Sigma(r_p) - \frac{r_0^2}{r_p^2}\Delta\Sigma(r_0)$$

r_0=1Mpc/h cut-off radius

Cosmological surveys



Courtesy of A. Raichoor



SDSS-III/BOSS (2009-14): 1.5M galaxies over 10,000 deg^2, mostly **LRGs** z<0.7

SDSS-IV/eBOSS (2014-20): 375k **LRGs** z<0.8, 260k **[OII] ELGs** z<1, 740k **QSOs** over 7500 deg^2

DESI (2020-): 10M **[OII] ELGs z<1.7**, **LRGs<1, QSOs z>2** over 14,000 deg^2

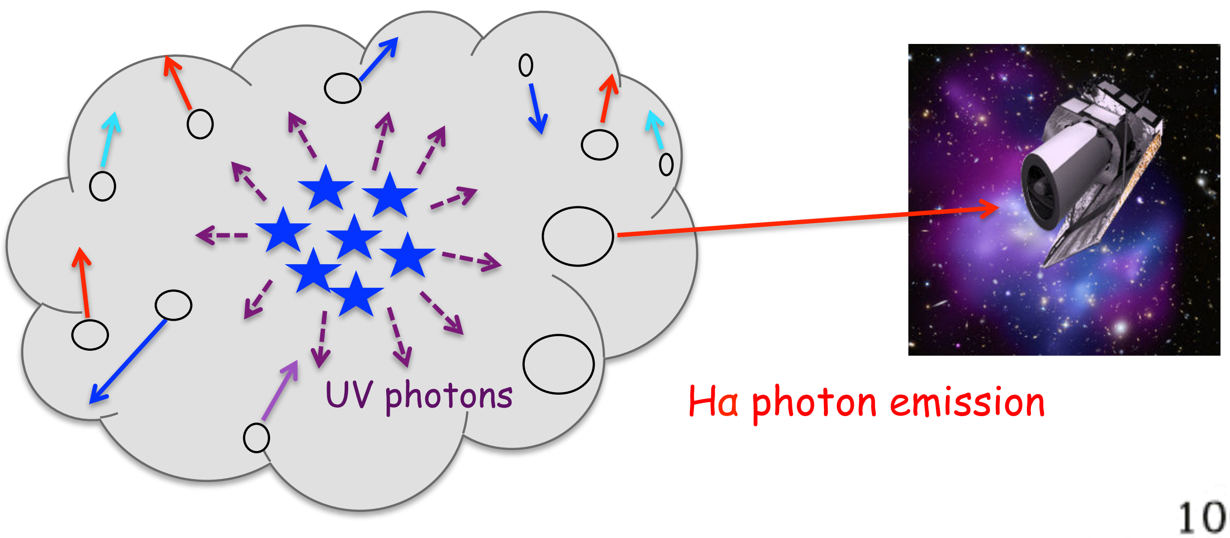
EUCLID (2022-): 50M **Hα ELGs z<2** with F> 2x10^{-16}erg/s, **QSOs** at z>2 over 15,000 deg^2, **clustering+lensing,** exquisite photo-z's key for WL

multi-tracer surveys

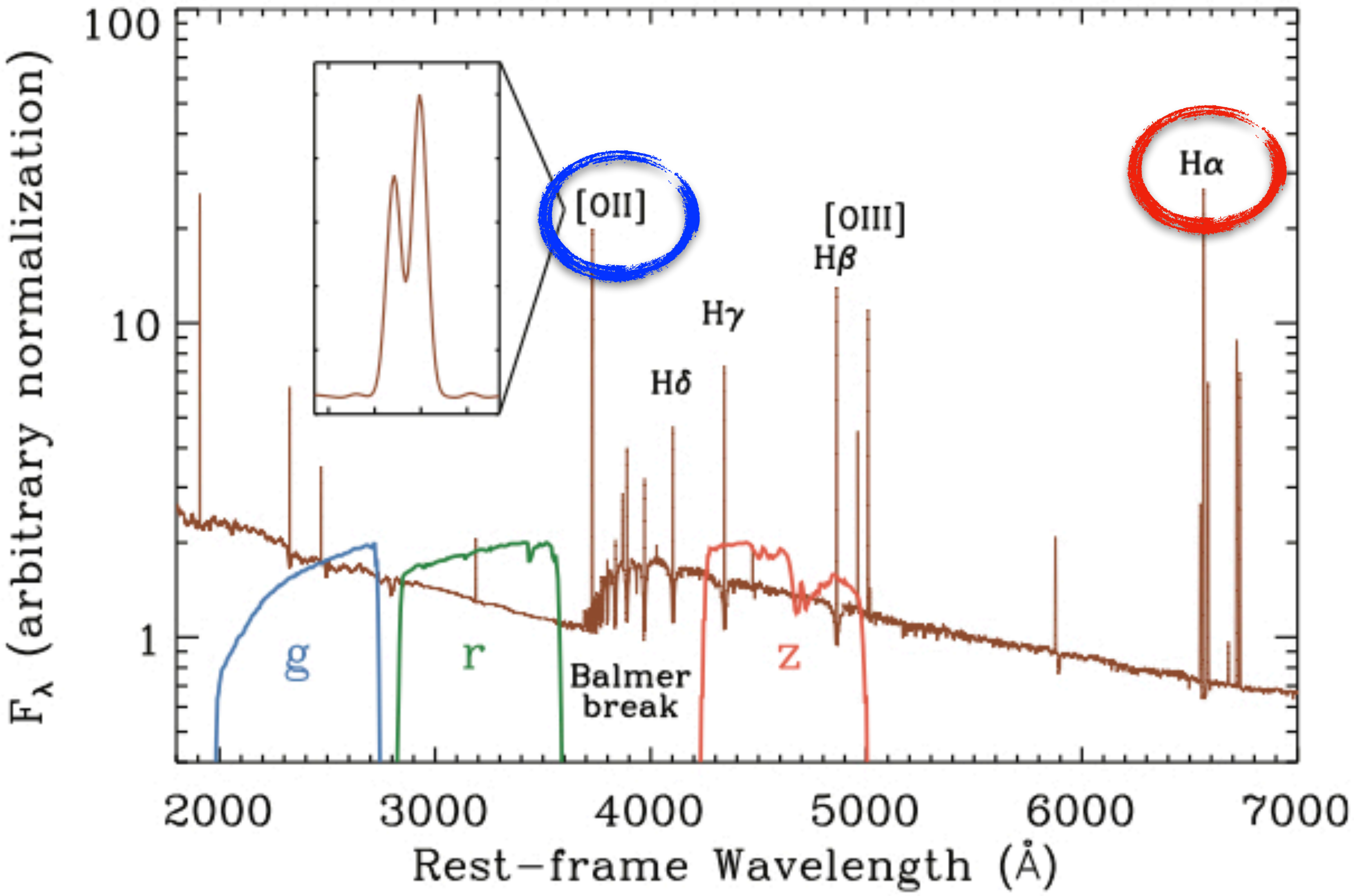
Emission line galaxies (ELGs)
Quasars (QSOs)
Luminous Red Galaxies (LRGs)

Emission line galaxies (ELGs)

Bright nebular emission generated when young massive stars in HII regions ionise the surrounding gas



http://desi.lbl.gov.tdr/

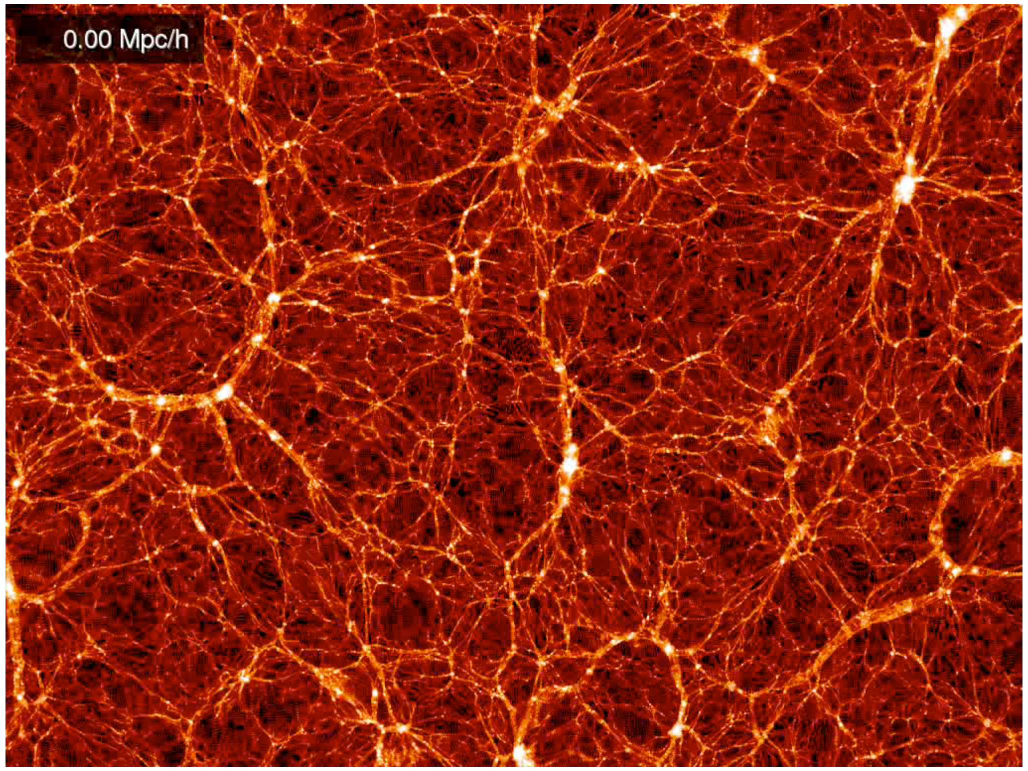


The upcoming surveys will target **ELGs** out to **z~2 to trace** the **BAO, growth of structure, star formation history** and deliver the **most precise 3D maps** of the Universe to date.

N-body DM-only cosmological simulations

Past:

MultiDark



	N particles	Lbox (Mpc/h)	mass resolution (M_{sun}/h)
BigMD	3840^3	2500	2.36×10^{10}
MDPL2	3840^3	1000	1.5×10^{9}
SMD	3840^3	400	9.63×10^{7}

cosmosim.org

Klypin et al. 2016

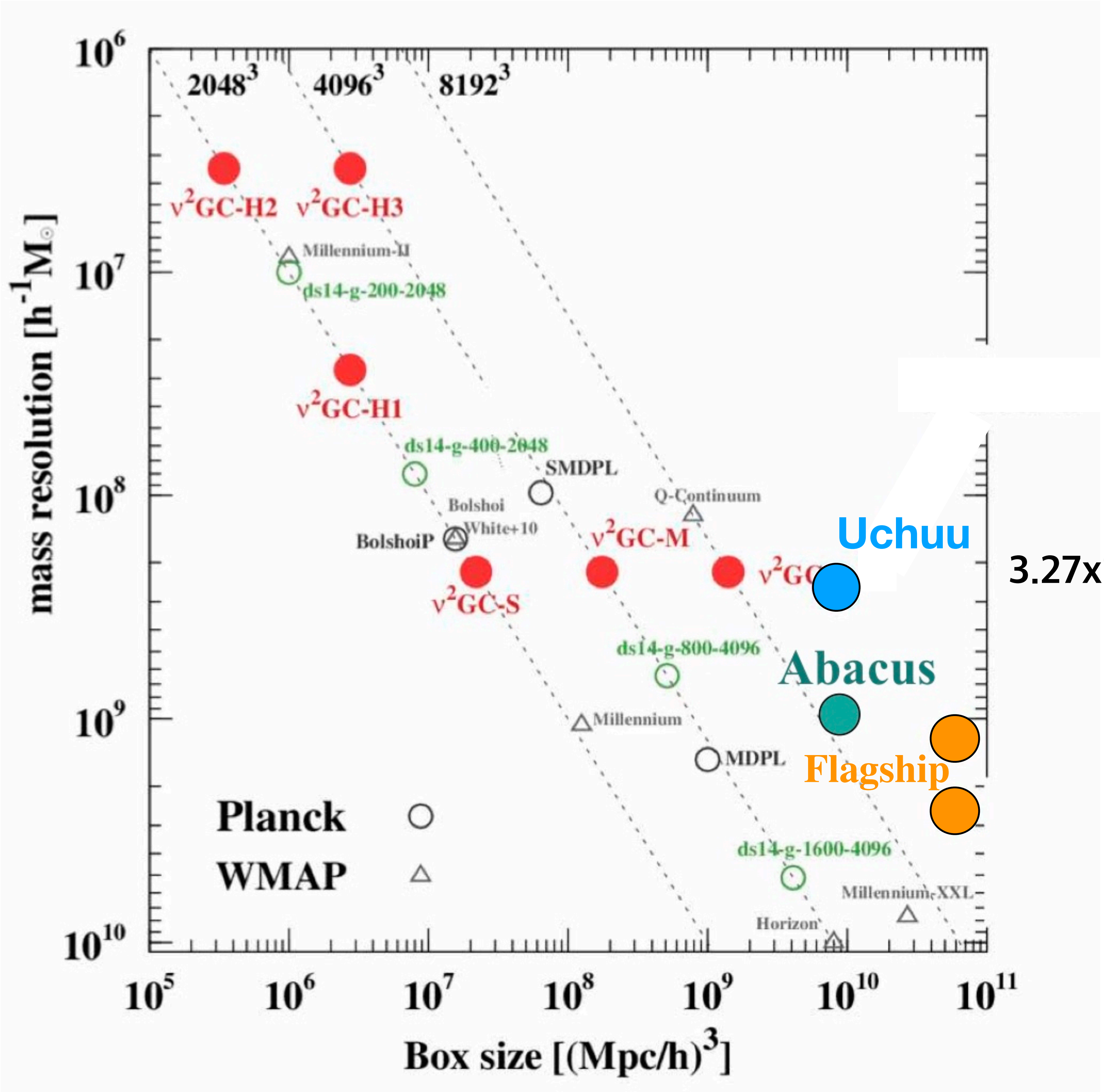
Ongoing:

N particles	Lbox (Mpc/h)	mass resolution (M_{sun}/h)
12800^3	2000	3.27×10^{8}

skiesanduniverses.org

Ishiyama et al. 2020

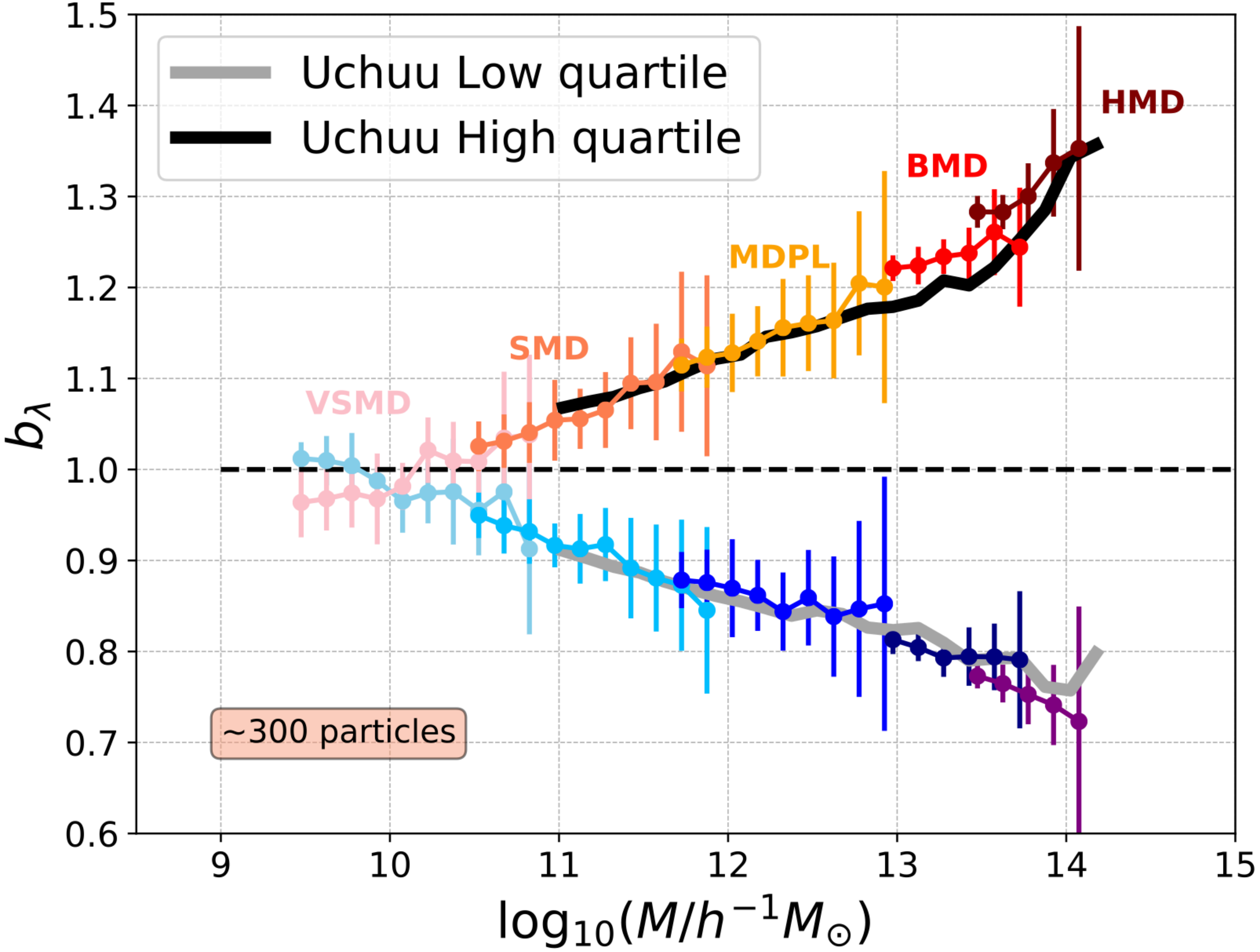
High resolution will be key for **resolving the smallest haloes** hosting ELGs



$3.27\text{x}10^8\ M_{sun}/h$

Courtesy of T. Ishiyama (Chiba University)

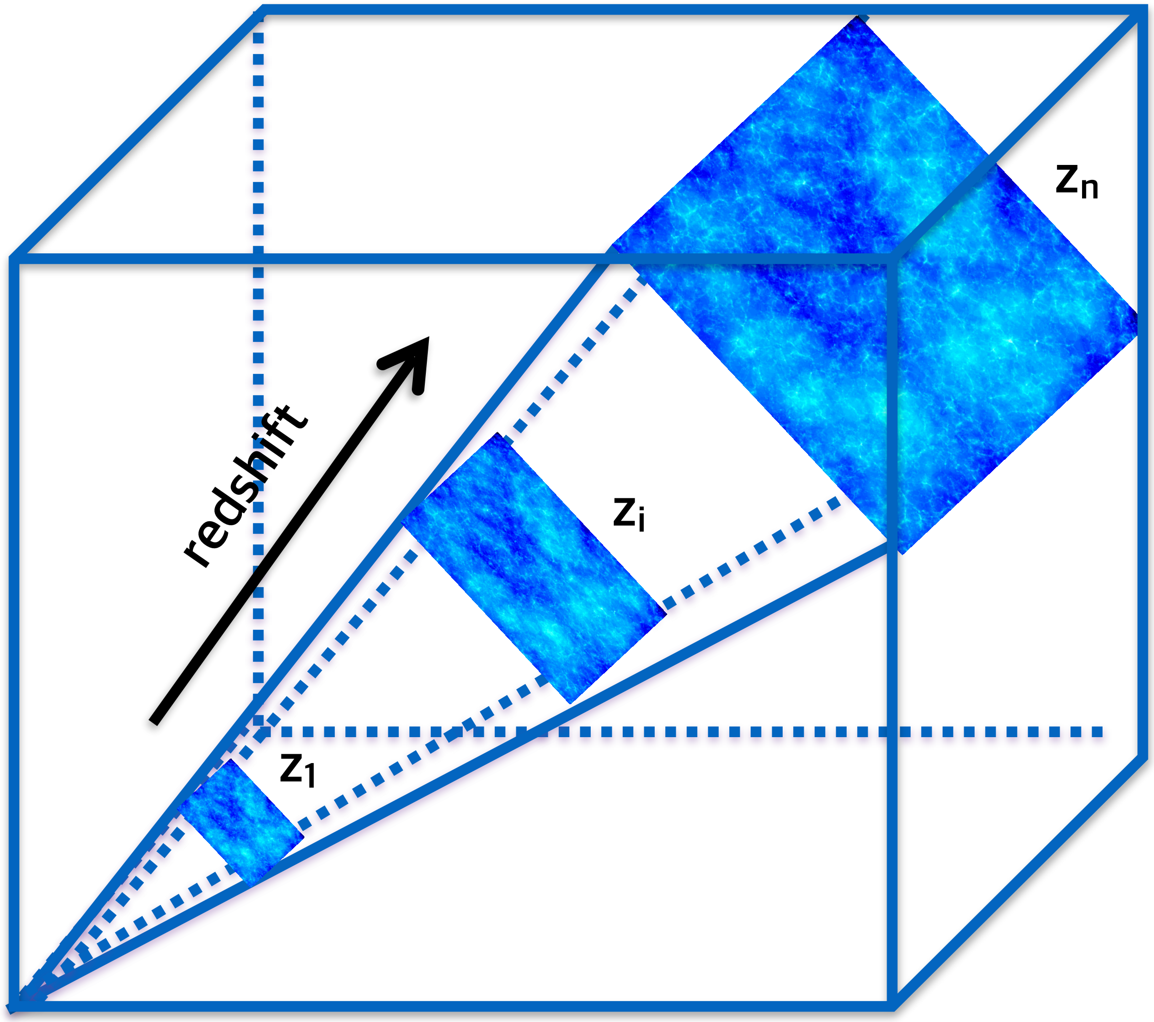
Large volume will allow us to **constrain halo assembly bias** over **4 orders of magnitude**



Courtesy of A. D. Montero-Dorta (USM, Chile)

Light-cones

Concatenate the simulation snapshots in the observed z range using the **Survey Generator Algorithm (SUGAR)**:



Rodríguez-Torres et al. 2016

LC more realistic than single snapshot —> includes **full z evolution** and **n(z) fluctuations**

The galaxy-halo connection

Galaxies are **biased tracers** of the underlying **dark matter distribution**, therefore we populate DM haloes with galaxies using their spatial properties using two main methods:

I. Halo Occupation Distribution (HOD)

Cooray+02; Berlin & Weinberg+02; Kravtsov+04; Zheng+05,07

II. SubHalo Abundance Matching (SHAM)

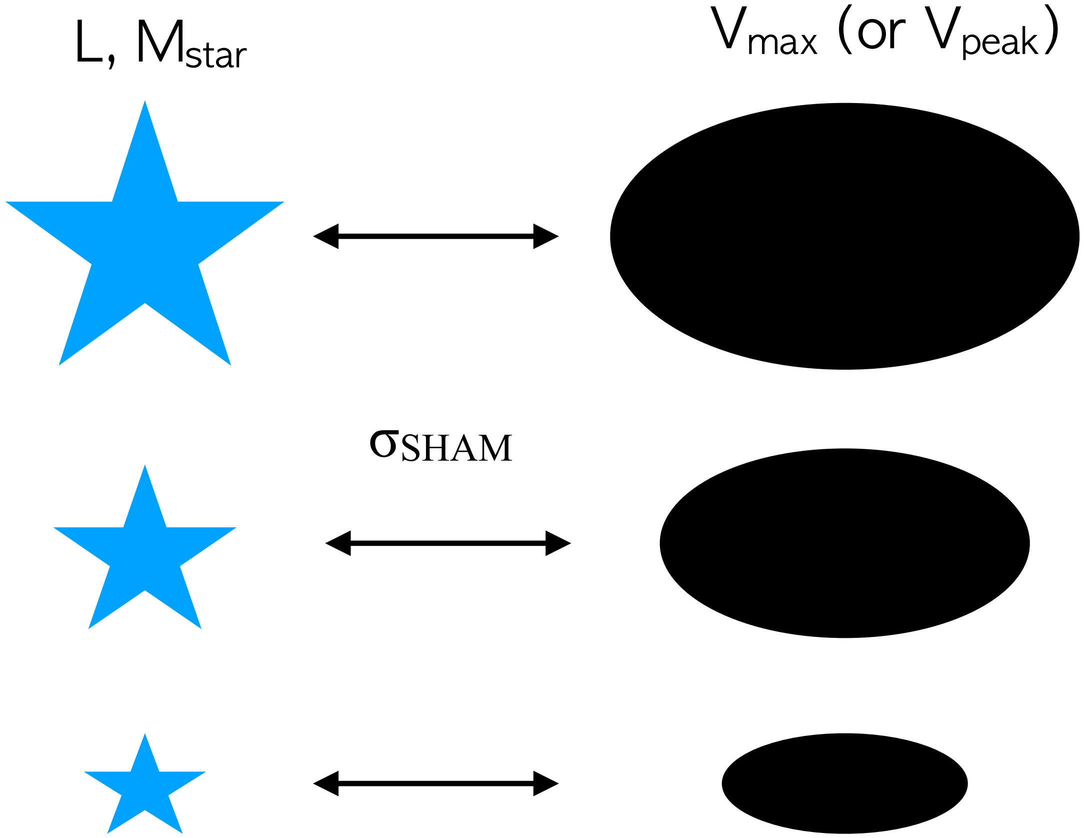
Conroy+06, 09; Behroozi+10; Trujillo-Gomez+11

Standard SHAM

Basic assumption: more massive/luminous galaxies live in more massive haloes

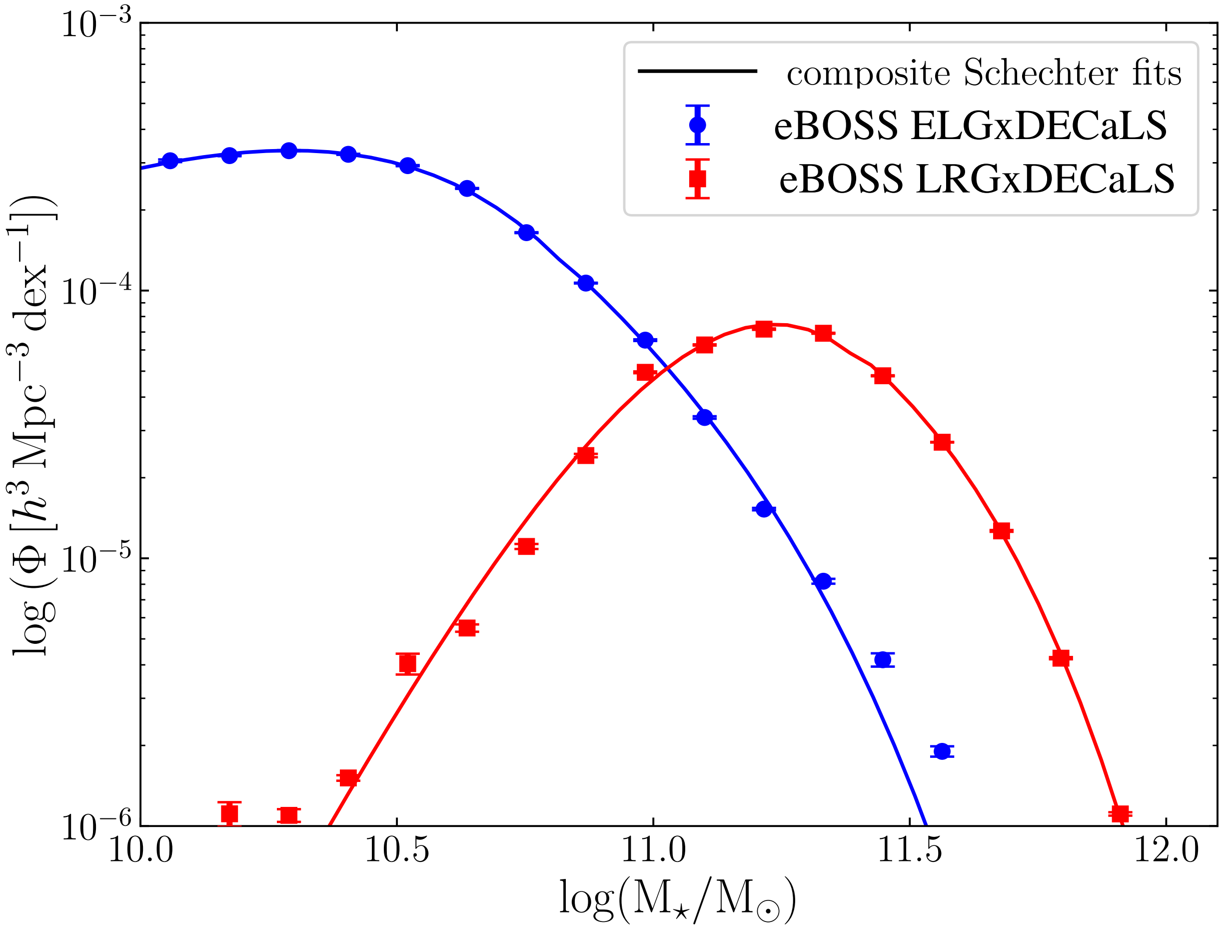
$$V_{\max}^{\text{scattered}} = V_{\max}\,[1 + \mathcal{G}(0, \sigma_{\text{SHAM}})]$$

Rank order haloes/galaxies allowing some scatter:



Rank order V_{peak} and assign galaxy L/M_{star} sampled from the observed L/M_{star} function until the observed n(z) is reached.

$$n_{\text{gal}}(> M_{\star}) = n_{\text{h}}(> V_{\text{peak}})$$



GF et al. 2022c in prep.

Standard SHAM works well only for complete galaxy samples

Modified SHAM for incomplete samples

galaxy multi-tracers at high z are often incomplete —> SHAM needs modification

$$\boxed{\mathrm{PDF}(V_{\mathrm{peak}}^{\mathrm{mean}}, \sigma_V, f_{\mathrm{sat}}) = f_{\mathrm{sat}} G_{\mathrm{s}}(V_{\mathrm{peak}}^{\mathrm{mean}}, \sigma_V, f_{\mathrm{sat}}) + (1 - f_{\mathrm{sat}}) G_{\mathrm{c}}(V_{\mathrm{peak}}^{\mathrm{mean}}, \sigma_V, f_{\mathrm{sat}})}$$

GF et al. 2017; GF et al. 2022b, 2022c in prep.

Gaussian realisations normalised to match the observed number of galaxies per z bin:

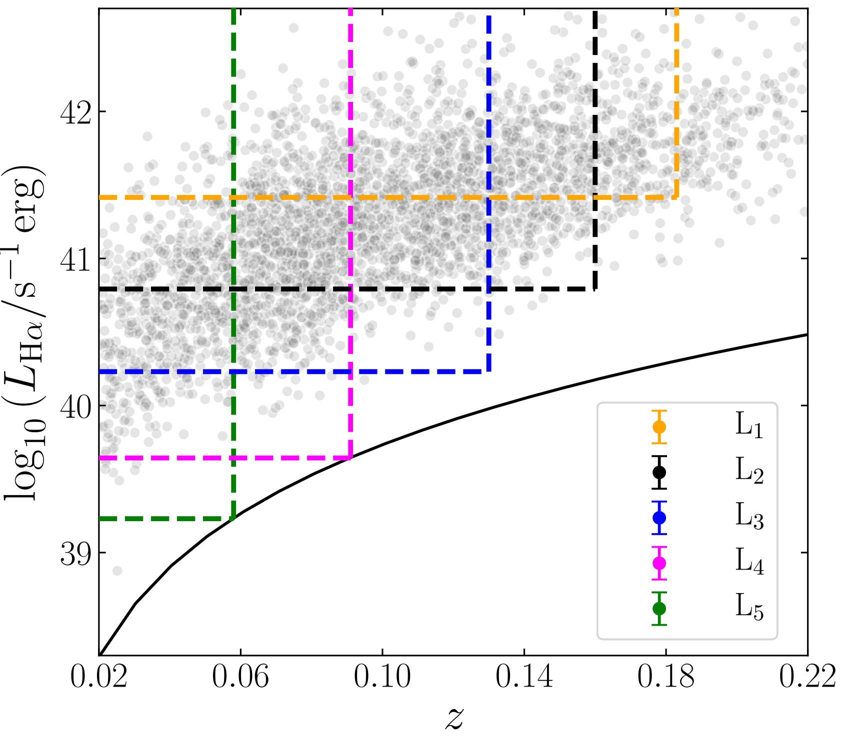
$$\int G_{\mathrm{s}}(V_{\mathrm{peak}}^{\mathrm{mean}}, \sigma_V, f_{\mathrm{sat}})\, dV_{\mathrm{peak}}^{\mathrm{mean}} = N_{\mathrm{tot}}(z) f_{\mathrm{sat}} \qquad \int G_{\mathrm{c}}(V_{\mathrm{peak}}^{\mathrm{mean}}, \sigma_V, f_{\mathrm{sat}})\, dV_{\mathrm{peak}}^{\mathrm{mean}} = N_{\mathrm{tot}}(z)(1 - f_{\mathrm{sat}})$$

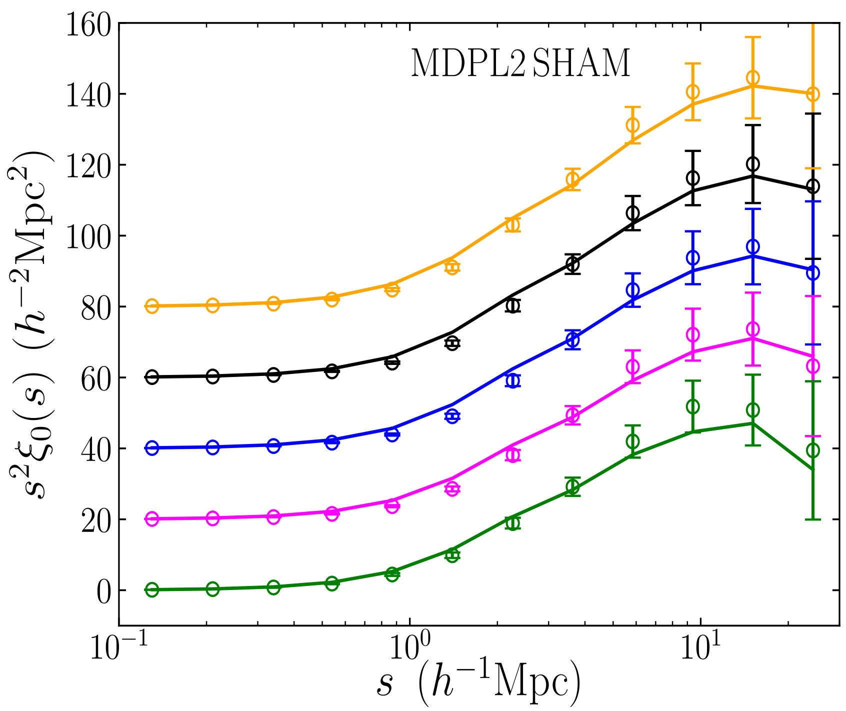
In practice:

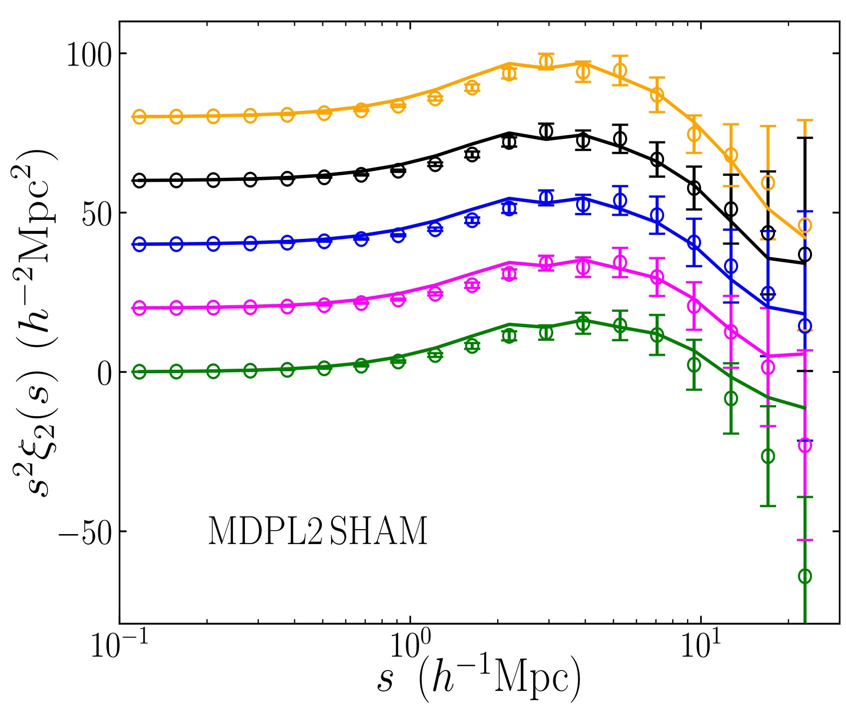
1. Compute the cen/sat halo velocity functions. In each (z, V$_{\mathrm{peak}}$) bin we have $N_{\mathrm{tot}}^{c/s}$ haloes.
2. Using the PDF above, draw $N_{\mathrm{gauss}}^{c/s}$ haloes with $V_{\mathrm{peak}}^{\mathrm{mean}} \pm \sigma_V$
3. Force the halo distribution to match the Gaussian shape by downsampling using:

$$P_{\mathrm{c/s}}(z, V_{\mathrm{peak}}^{\mathrm{mean}}) = \frac{N_{\mathrm{gauss}}^{\mathrm{c/s}}(z, V_{\mathrm{peak}}^{\mathrm{mean}})}{N_{\mathrm{tot}}^{\mathrm{c/s}}(z, V_{\mathrm{peak}}^{\mathrm{mean}})}$$

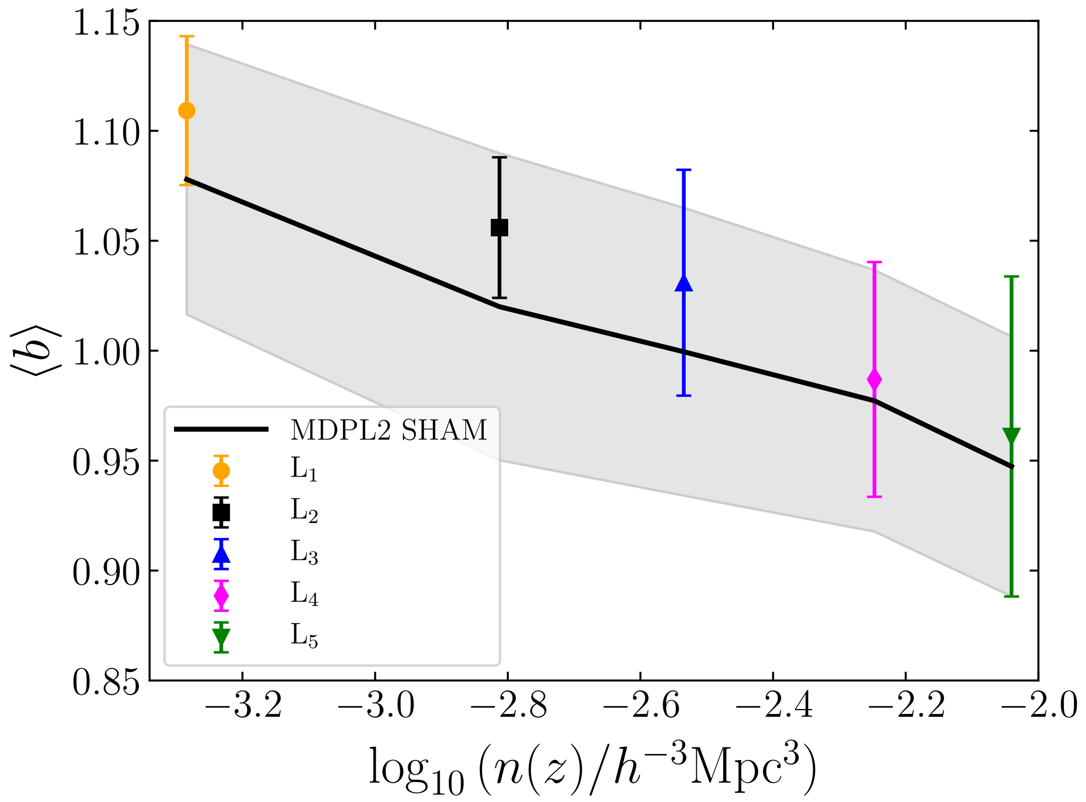
SDSS DR7 Main Hα ELGs at $0.02<z<0.22$ limited at $r_{mag}<17.77$ and $F>2\times10^{-16}$ erg/cm^2/s (Euclid-like)







sample name	$f_{\rm sat}$ [%]	$V_{\rm peak}^{\rm mean}$ [km s^{-1}]	$\log({\rm M_h}/h^{-1}{\rm M_\odot})$
L_1	25.4±0.1	194±81	11.71±0.69
L_2	22.5±0.4	215±89	11.86±0.71
L_3	19.6±0.7	244±97	12.06±0.70
L_4	20.2±0.4	242±101	12.09±0.74
L_5	20.9±0.5	267±97	12.18±0.68



From SHAM on average we find:

$$M_{halo} = (1.0 \pm 0.2) \times 10^{12}\, h^{-1} M_\odot$$

$$f_{sat} = (21.7 \pm 0.4)\,\%$$

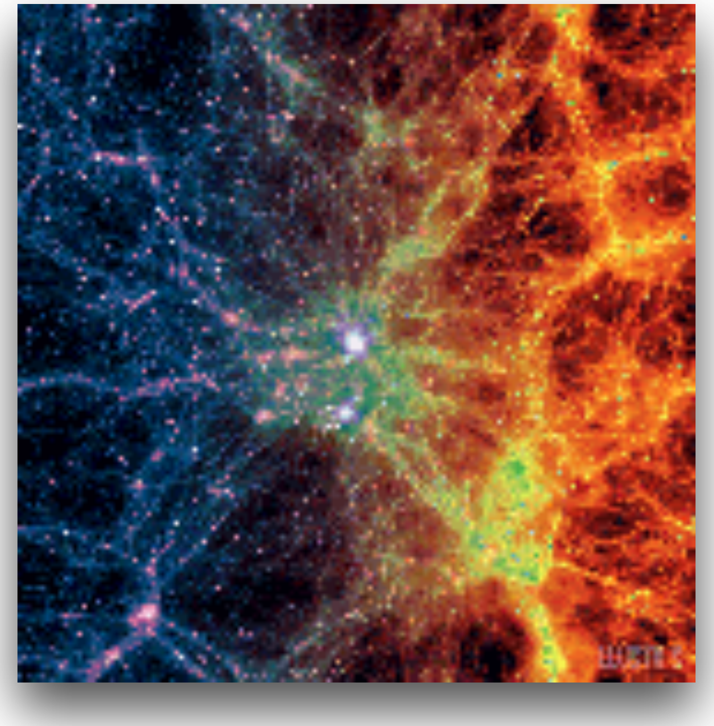
$$bias \sim 1$$

GF et al. 2022b in prep.

Decorated SHAM including secondary properties

AGE MATCHING model (Hearing et al. 2013) extended to build more predictive mocks including primary and **secondary halo/galaxy properties**

IllustrisTNG100



L_{box} = 75 Mpc/h

N_{DM} = 1820^3

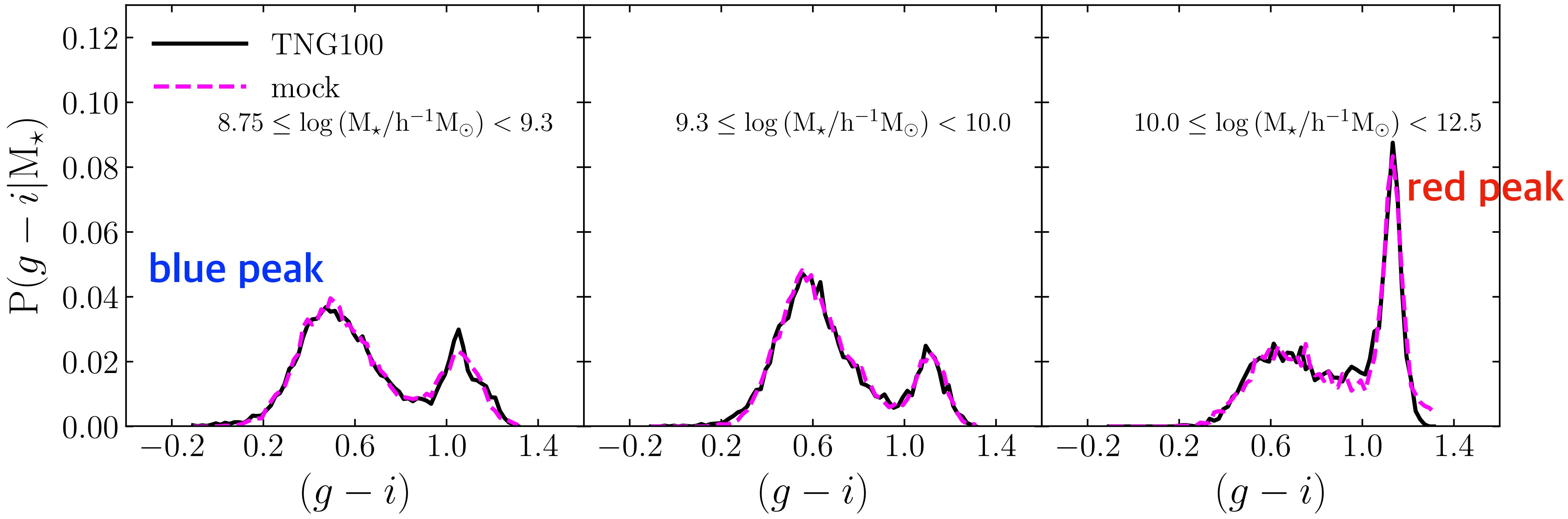
M_{DM} = 5.1 x $10^6 M_{sun}$

M_{gas} = 9.4 x $10^5 M_{sun}$

	Galaxies		Haloes
I	$L, M_\star$	⟷	$V_{\rm max}, V_{\rm peak}$
II	$(g-i)$, SFR	⟷	$z_{\rm starve}, c_{\rm infall}, \delta_{\rm R}^{\rm env}, \alpha_{\rm R}, \delta_{\rm R}$

↓ **Halo tidal properties**

Secondary matching using **conditional PDFs in bins of stellar mass**:



GF et al. 2022a, arXiv: 2101.10733

Halo tidal tensor computed on a 1024^3 cubic lattice using **SPIDER** code (**Martizzi et al. 2019**) and interpolated at each halo location and R_{vir}. By diagonalising the tensor we define:

Halo tidal overdensity

$$\boxed{\delta_R = \lambda_1 + \lambda_2 + \lambda_3}$$

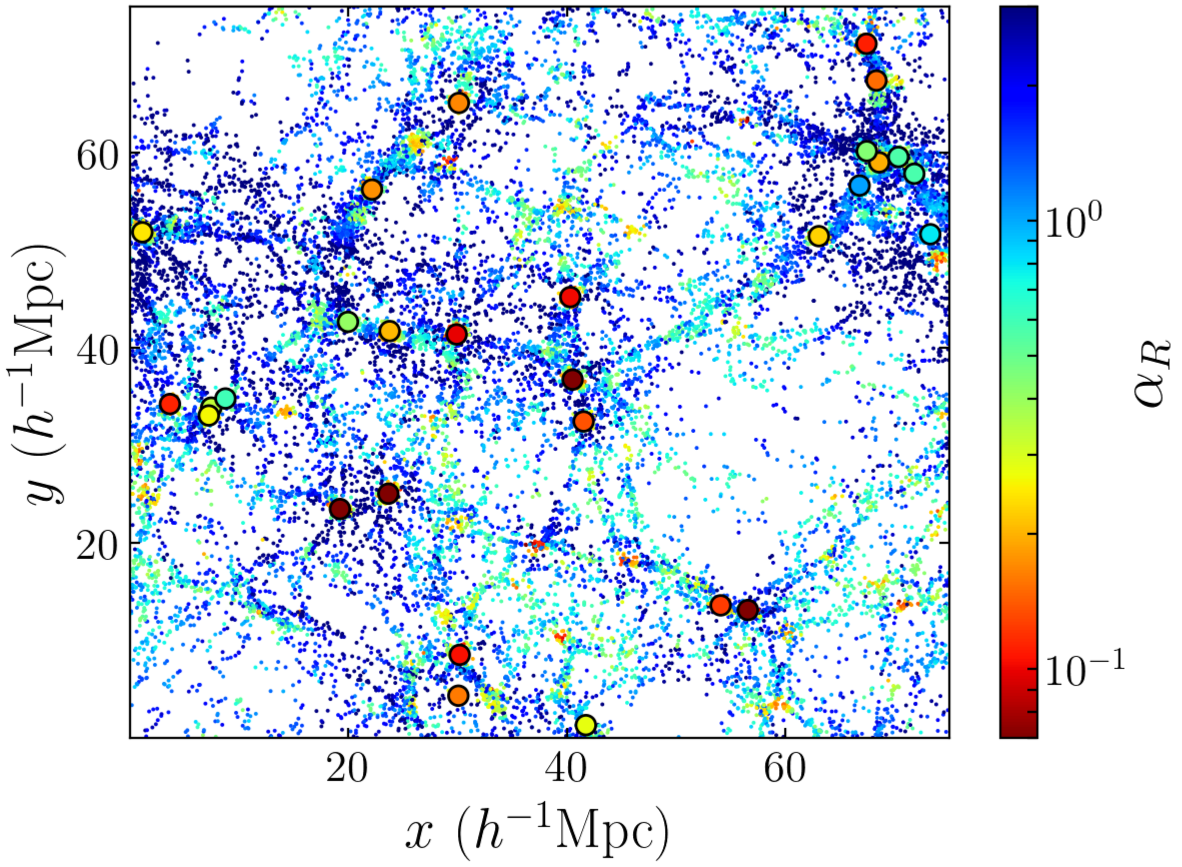
Paranjape et al. 2018
Ramakrishnan et al. 2019

Tidal anisotropy parameter

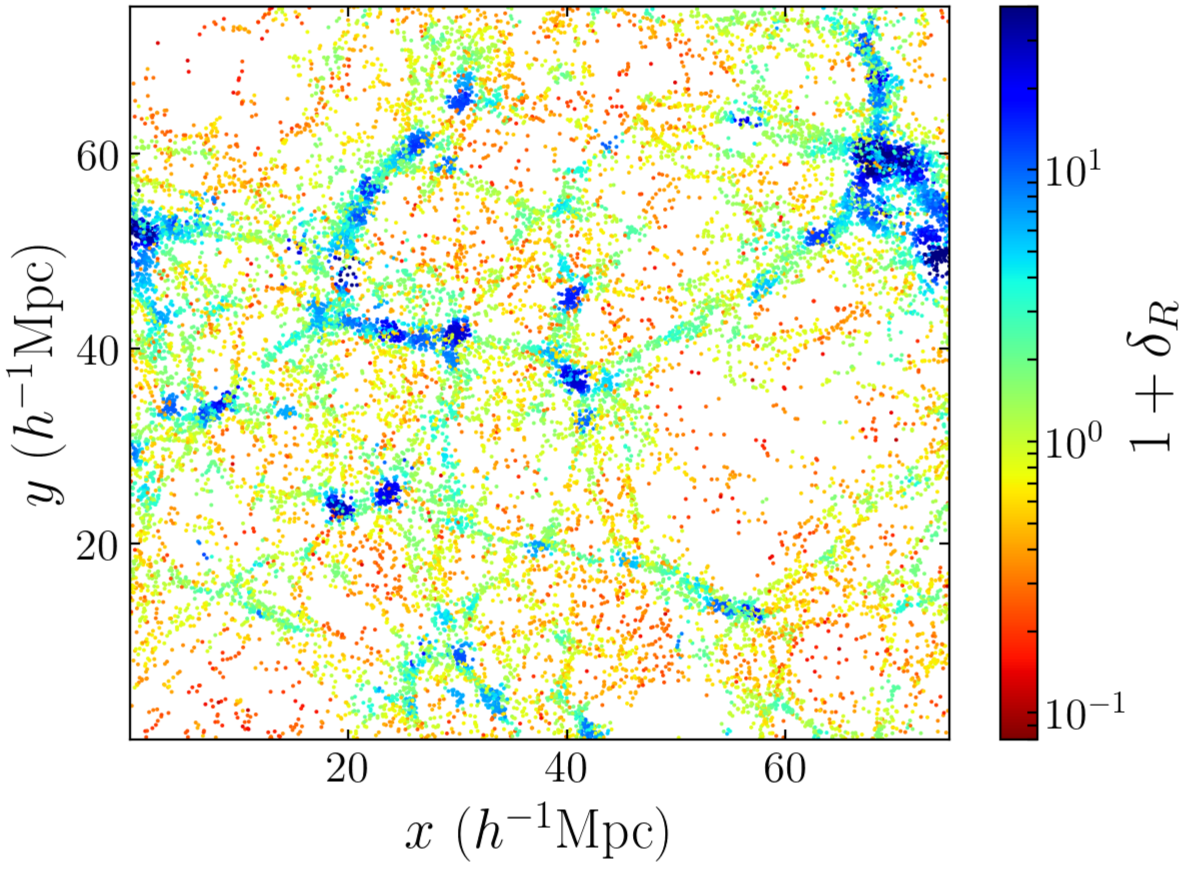
$$\boxed{\alpha_R \equiv (1+\delta_R)^{-1}\sqrt{q^2}}$$

$$q^2 = \frac{1}{2}\left[(\lambda_2-\lambda_1)^2 + (\lambda_3-\lambda_1)^2 + (\lambda_3-\lambda_2)^2\right]$$

The anisotropy parameter is a **mediator between the internal and large-scale properties of haloes**



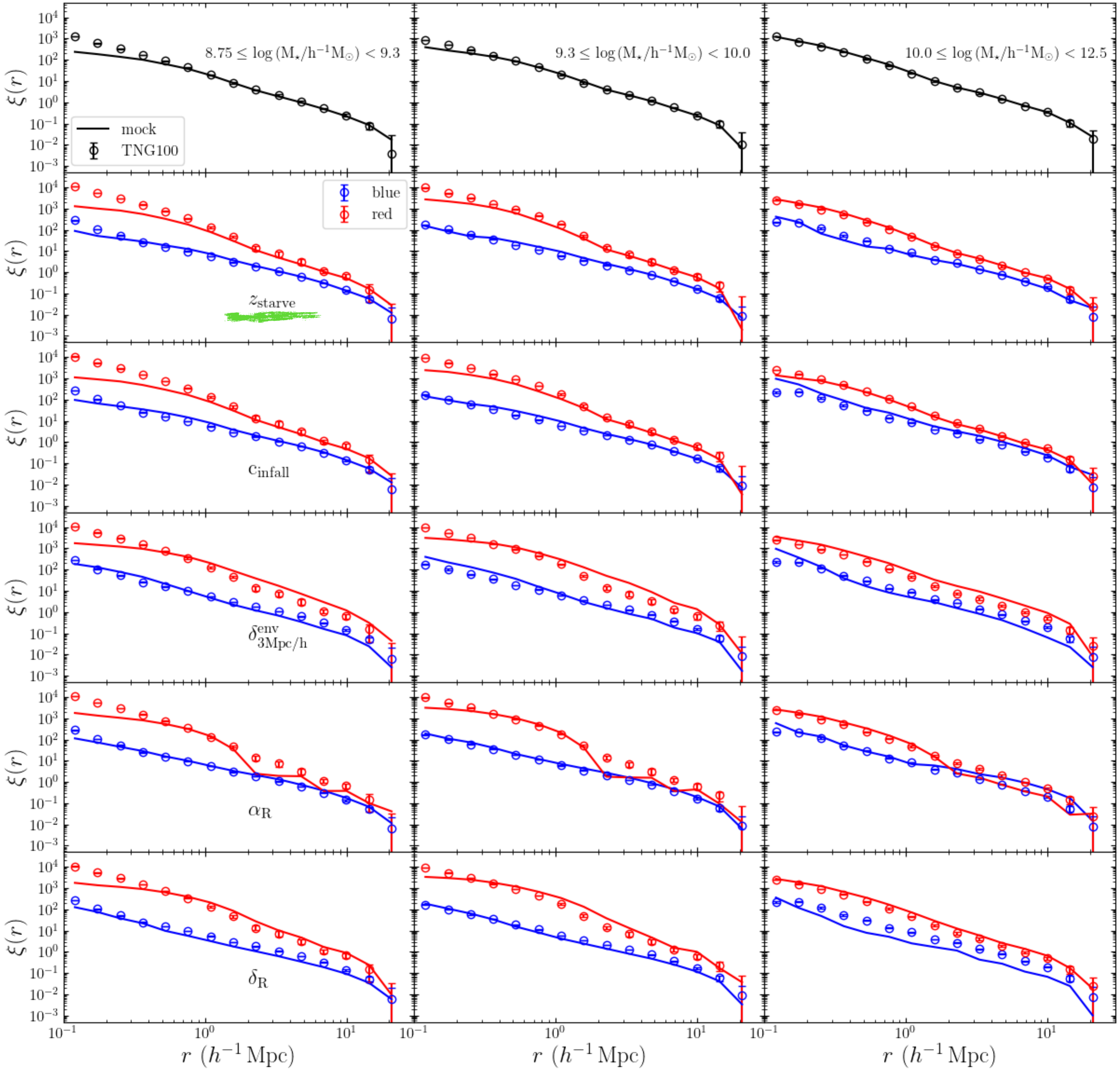
Anisotropy is higher in filaments an lower in knots



δ_R

GF et al. 2022a, arXiv: 2101.10733

Red : $(g - i) \geq 0.85$ Blue : $(g - i) < 0.85$



The more correlated the secondary props the better they perform

GF et al. 2022a, arXiv: 2101.10733

Summary

- SHAM is a simple, yet powerful, prescription able to link galaxies to their host DM haloes reproducing the clustering signal of complete samples.

- The upcoming surveys will target millions of galaxy **multi-tracers at high z**, most of them very **incomplete in luminosity/M_{star}**. Therefore, a **modified SHAM approach is needed** to accurately model the clustering (and lensing) signal on all scales.

- **SHAM+Age Matching links the inner and large-scale galaxy/halo properties**, properly including the **secondary halo bias** and the physics of **galaxy formation/evolution**. However, the accuracy of the method strongly depends on the secondary properties chosen and their mutual correlations.