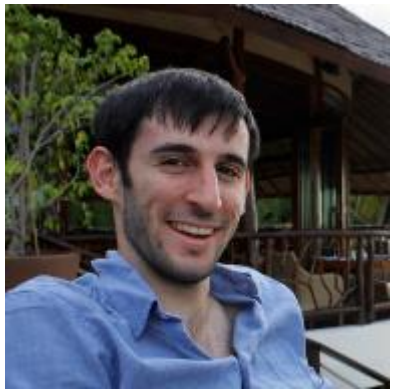
Simulations and the Galaxy–Halo Connection

Yao-Yuan Mao (Stanford/SLAC ➡ PITT PACC)

@yaoyuanmao | yymao.github.io

SCMA6 @ CMU 6/10/16

My collaborators at Stanford/SLAC





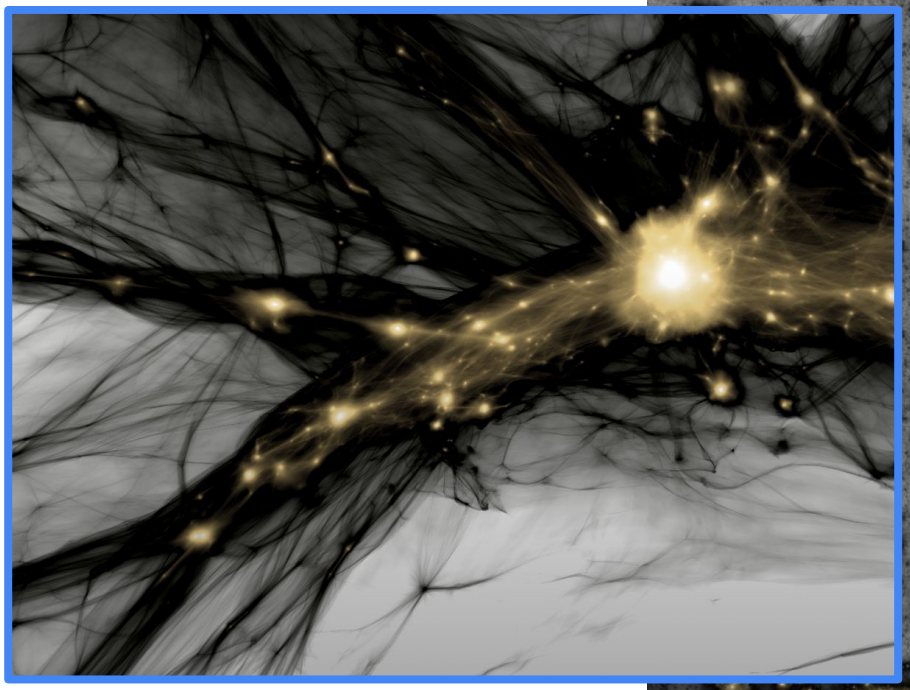


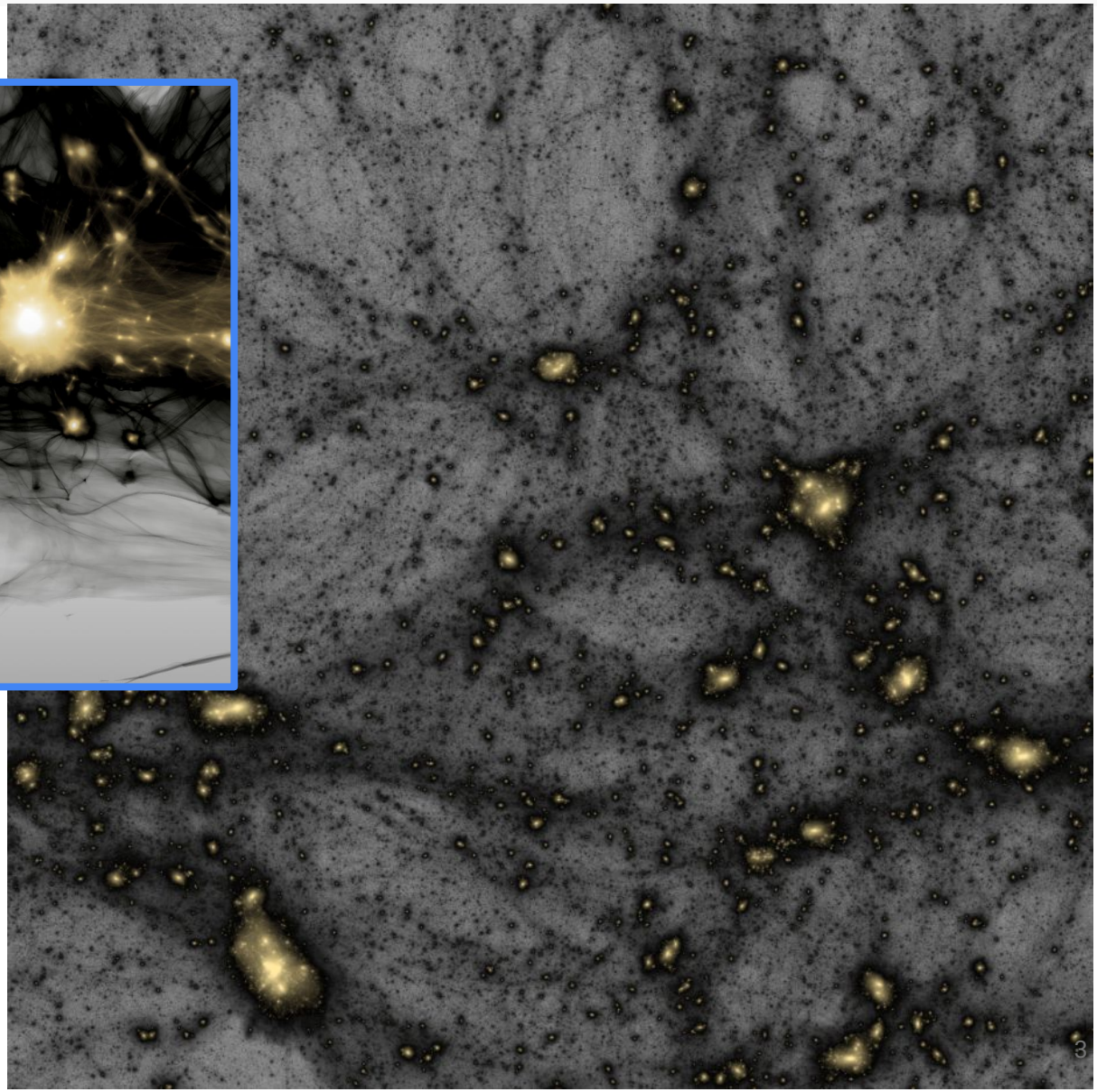

Joe DeRose
Ben Lehmann (➡ UCSC)
Vincent Su

Marc Williamson (➡ NYU)
Hung-I Eric Yang

Risa Wechsler
Phil Marshall
Matthew Becker

Cosmological Dark Matter Simulations



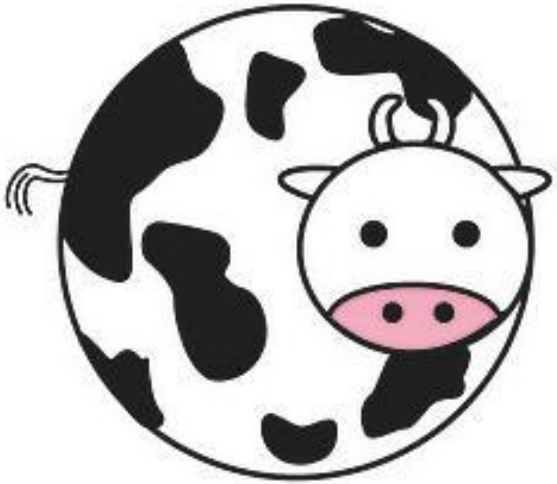


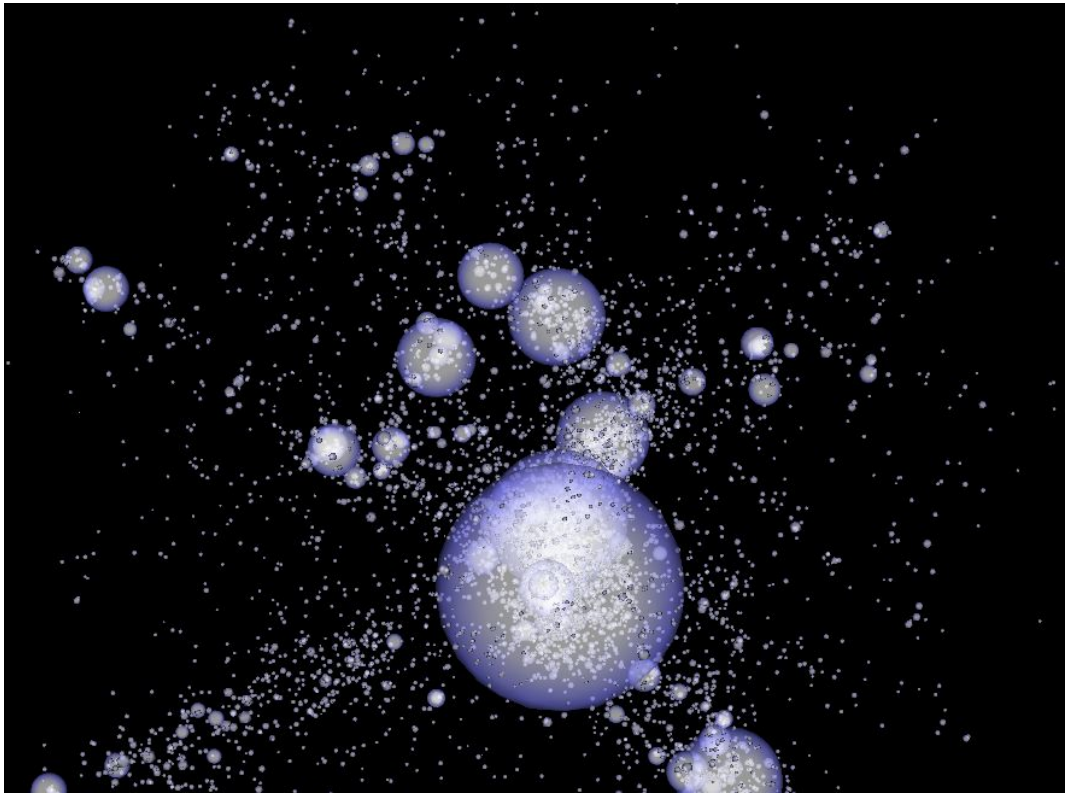
MW-size zoom-in simulations
[YYM+ 2015]

Dark Sky Simulations
[Skillman+ 2014]

Dark Matter Halos

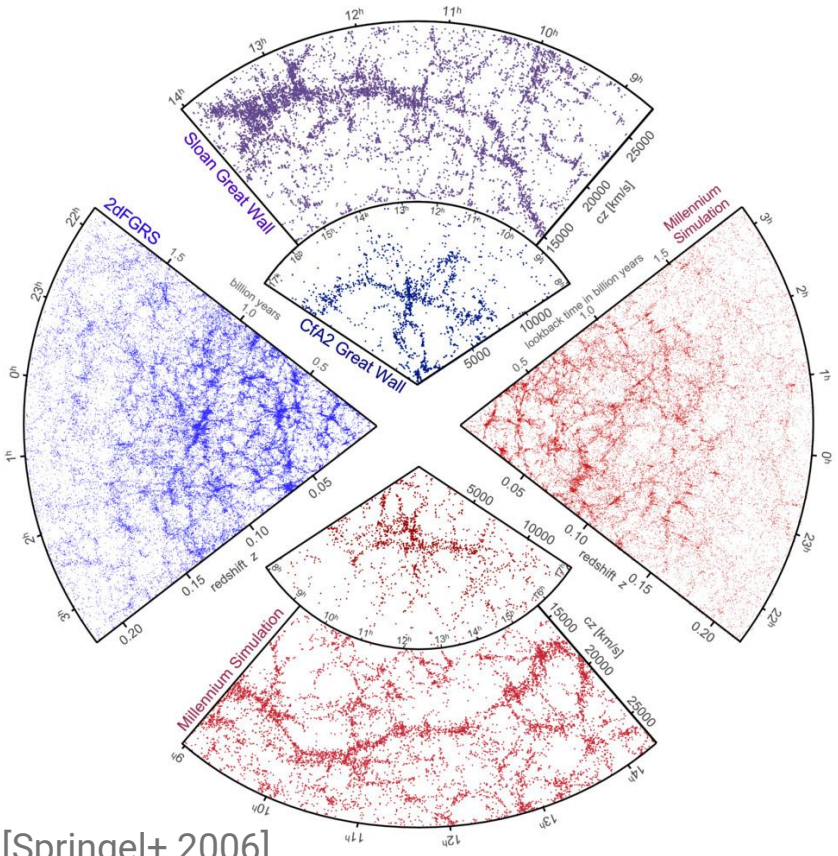
- High-density regions of dark matter
- Nests of visible galaxies
- Has hierarchical forming/merging processes
- Identified by a “halo finder” (e.g. Rockstar)





(Yes, this actually happens in conferences.)

Why do we model the Galaxy–Halo Connection?



[Springel+ 2006]

➔ **Cosmology and large-scale structure**
- Galaxy clustering
- Lensing
- Intensity maps

To marginalize over the GHC

➔ **Galaxy formation physics**
- Feedback, cooling, and many processes
- Dwarf galaxies

To predict and interpret the GHC

Galaxy–Halo Connection

➔ Dark matter-only simulations provide information about halos.
How about galaxies?

➔ How do galaxy properties depend on halo properties?

➔ What are the difficulties?

Hydrodynamical simulations [cf. Snyder's talk today]

➔ computationally expensive
(difficult to marginalize over)

➔ include most physics
(but still some subgrid models)

Semi-analytical models

➔ build galaxies along the merger histories

➔ controlled by "physical" parameters

Empirical models:

- Halo Occupation Distribution
- (Subhalo) Abundance Matching

➔ fewer parameters, fits to statistics directly ("data-driven")

➔ minimalist's approach

Example:
Connecting galaxies and halos with the Abundance Matching framework

- ➔ What are the “parameters” in abundance matching?

- ➔ Can we associate physical meanings to these parameters?
- ➔ Can we marginalize over these parameters?
- ➔ Can we do the two simultaneously?

To marginalize over the GHC

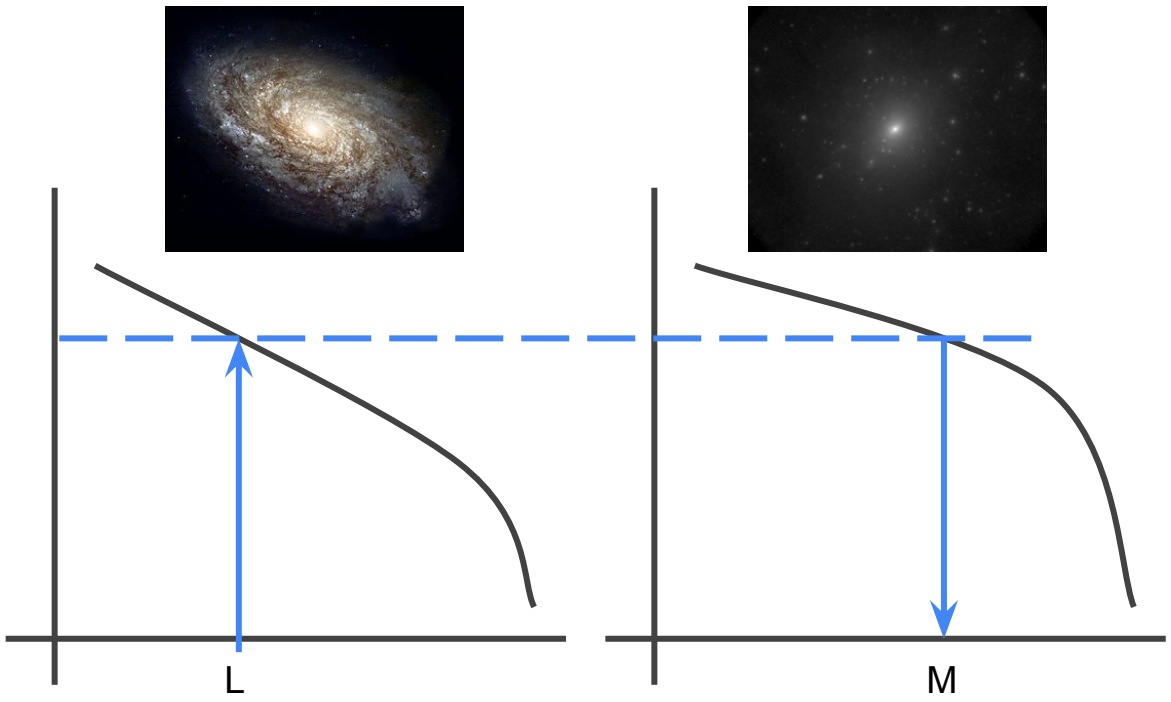
To predict and interpret the GHC

[cf. Wolfgang’s talk on Thur]

(Subhalo) Abundance Matching

Connects halos and galaxies assuming:

- ➔ galaxies live in halos and subhalos
- ➔ one halo property is strongly correlated with one galaxy property

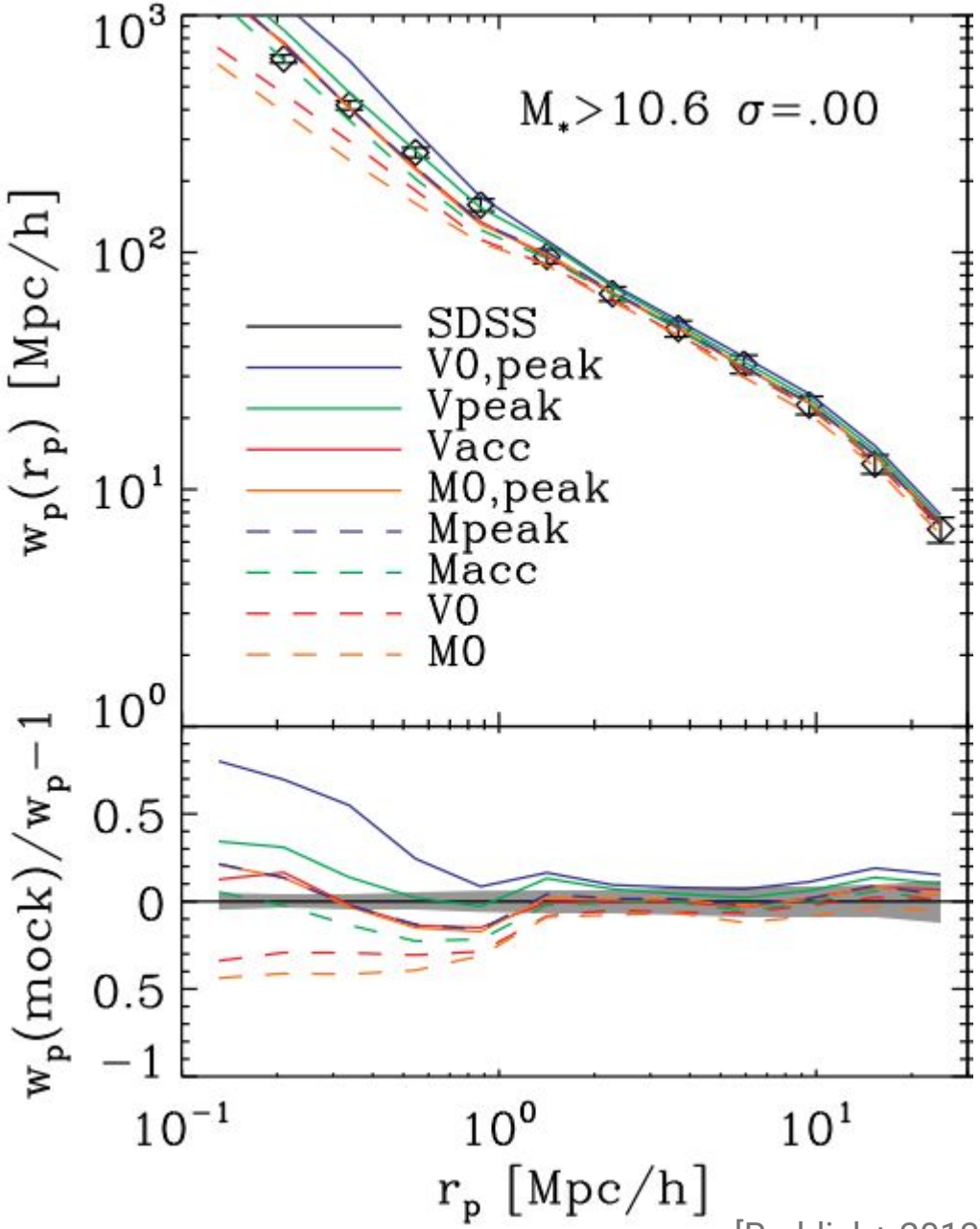


- ➔ Matching the halo proxy (e.g. mass) with galaxy luminosity (or stellar mass) at the same **number density**.
- ➔ Only the **rank** of the halo proxy matters.
- ➔ 1-point function is matched by construction, but other statistics are *not*

Kravtsov+ 2004
Vale & Ostriker 2004, 2006
Conroy+ 2006
Reddick+ 2013

Choice of the Matching Proxy

- ➔ Agreement b/w observed and predicted 2-point correlation functions (or other statistics) provides an evaluation on the proxy.
- ➔ It was found that “V-peak” or “M-peak” work the best, among others. And “V-peak” is better than “M-peak”.



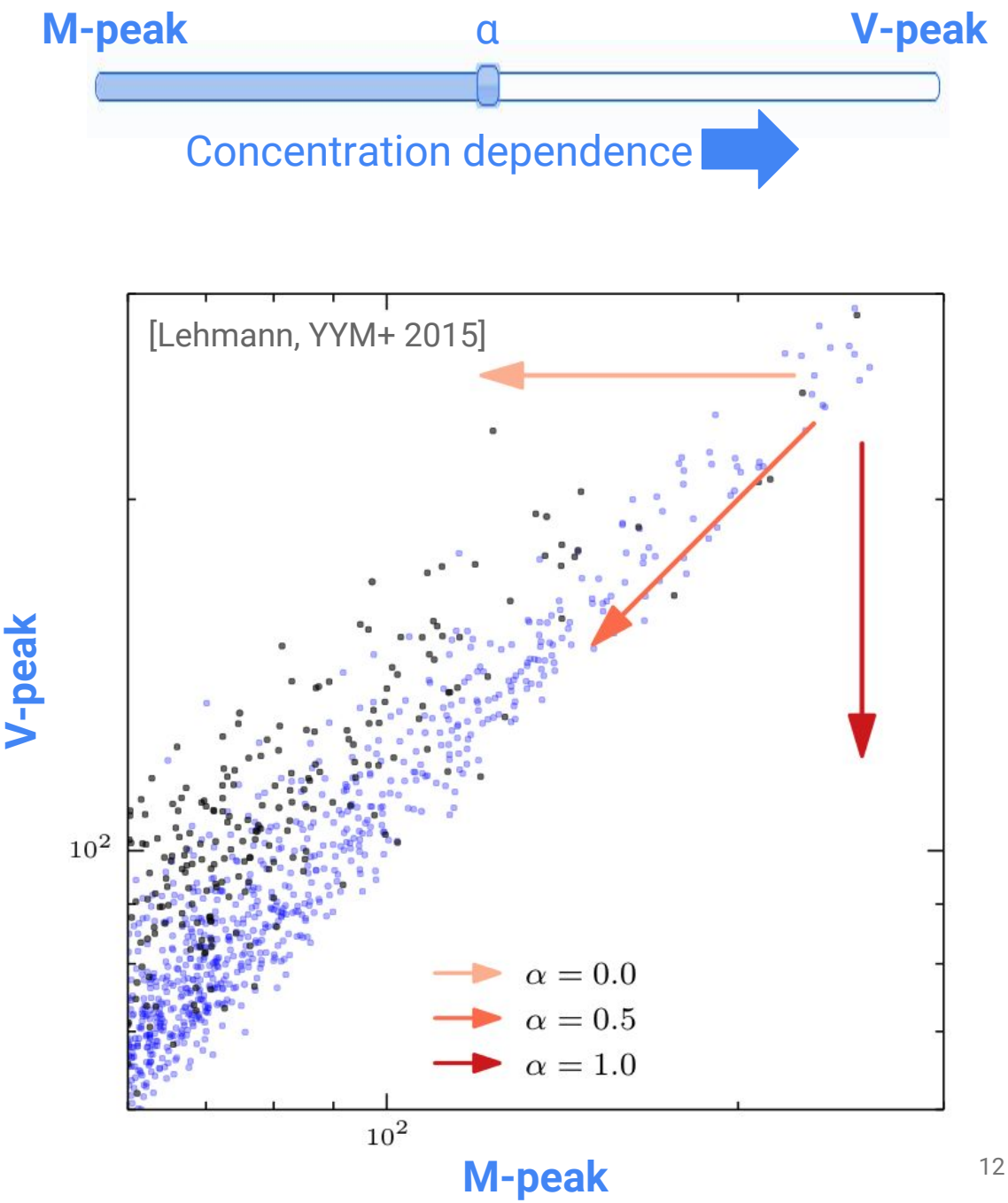
[Reddick+ 2013]

Problems with discrete choices

- ➔ Attaching too much physical meaning to a specific choice
- ➔ Cannot calculate inference, nor marginalize over the uncertainty associated with proxy choice

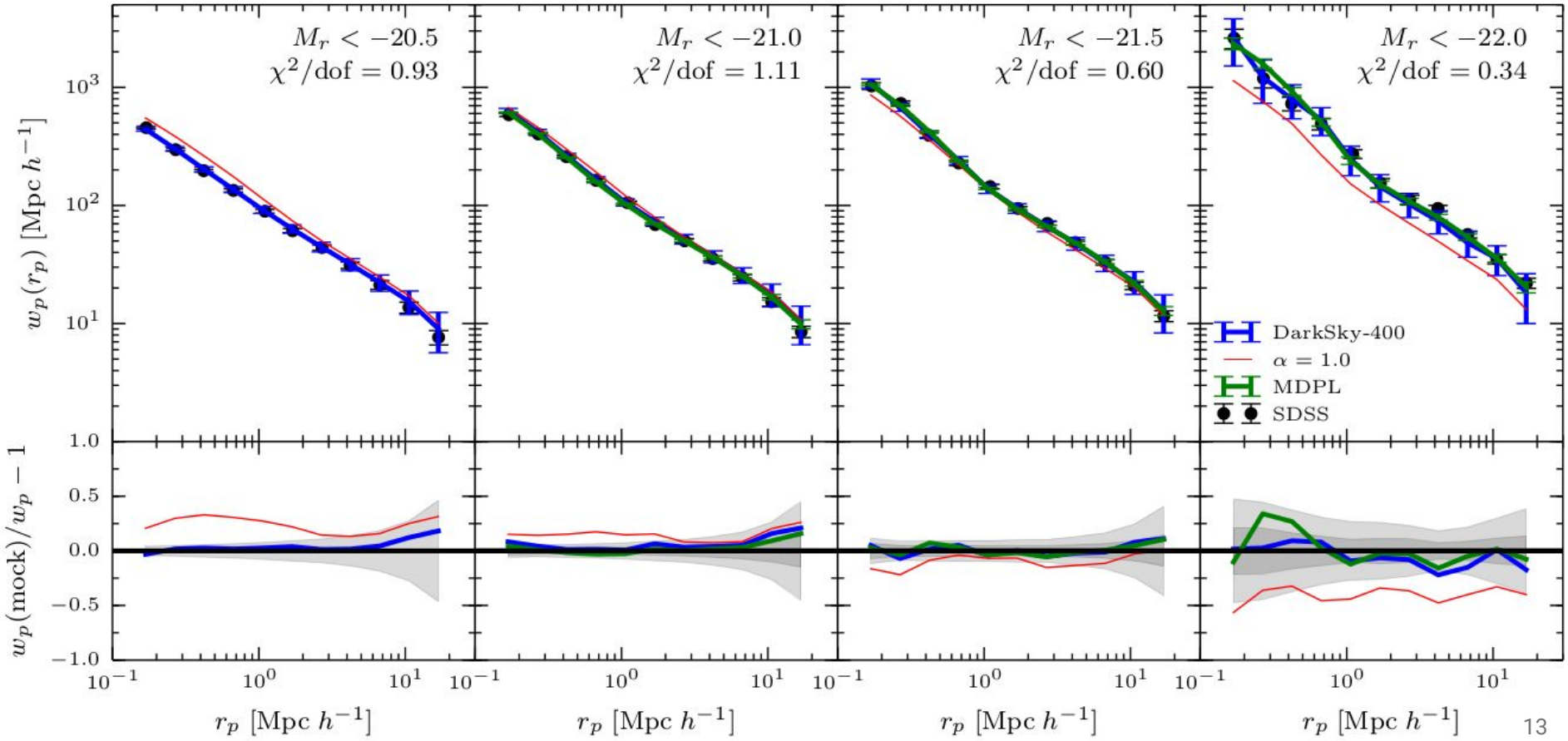
Interpolating matching proxies

- ➔ choosing between “mass” or “maximal circular velocity” *is* changing the concentration dependence
- ➔ parametrize this freedom by “α”
- ➔ changing “α” changes how we rank halos



Constraints from large-scale structure

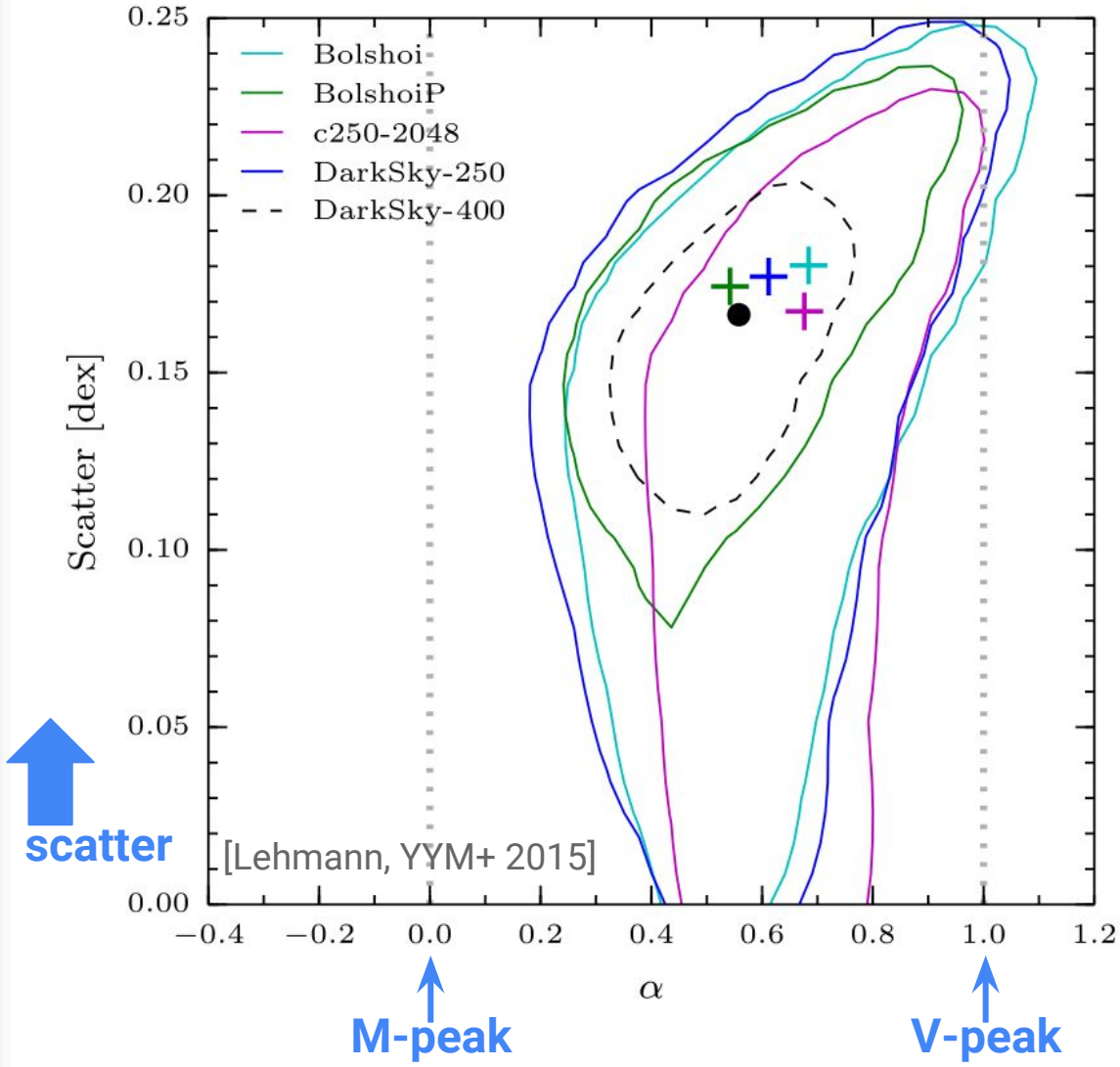
- ➔ Luminosity-selected sample built from SDSS DR7 [Reddick+ 2013]
- ➔ Fitting projected 2-point correlation function



Constraints on α and scatter

- “M-peak” and “V-peak” are both excluded
- α > 0 implies that data requires some concentration dependence (assembly bias)
- previous studies on AM commonly used 250 Mpc/*h* boxes

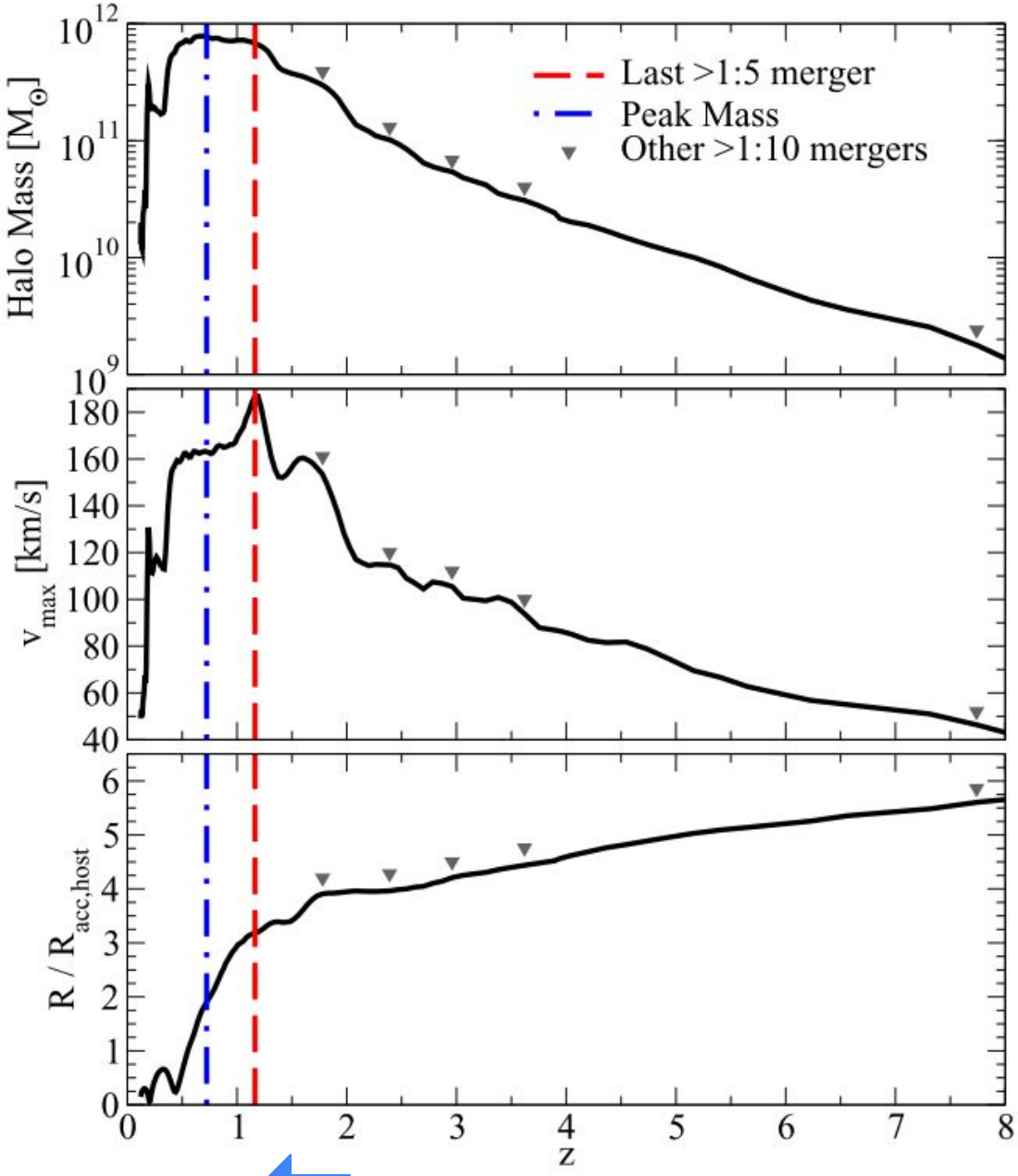
Joint constraint on α and scatter from four luminosity-selected samples.



(p = 0.05)

Choosing an epoch for the halo proxy

- ➔ Subhalos are stripped when/after entering the host halo
- ➔ Stellar content lives in a deeper potential and is less affected

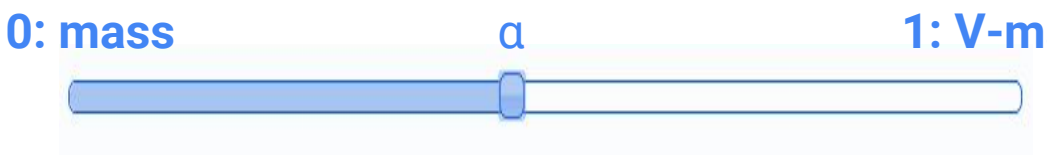


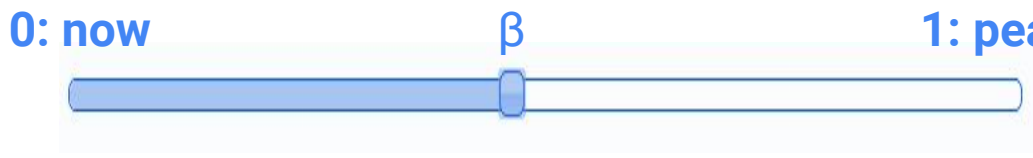
Later in time

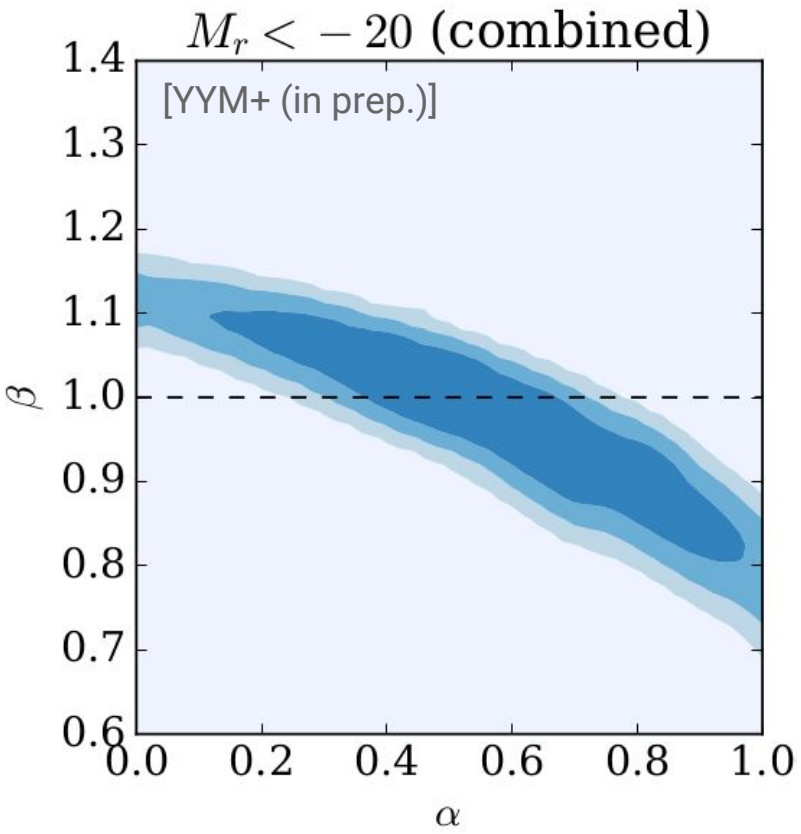
[Behroozi+ 2013]

Interpolating among multiple matching proxies

- ➔ β controls the choice b/w using proxy at peak or now.
- ➔ Constraints from 2-pt correlation function, with a fixed scatter of 0.15 dex.
- ➔ Allowed region of α increases significantly
- ➔ β is well constrained to be close to 1 (peak)







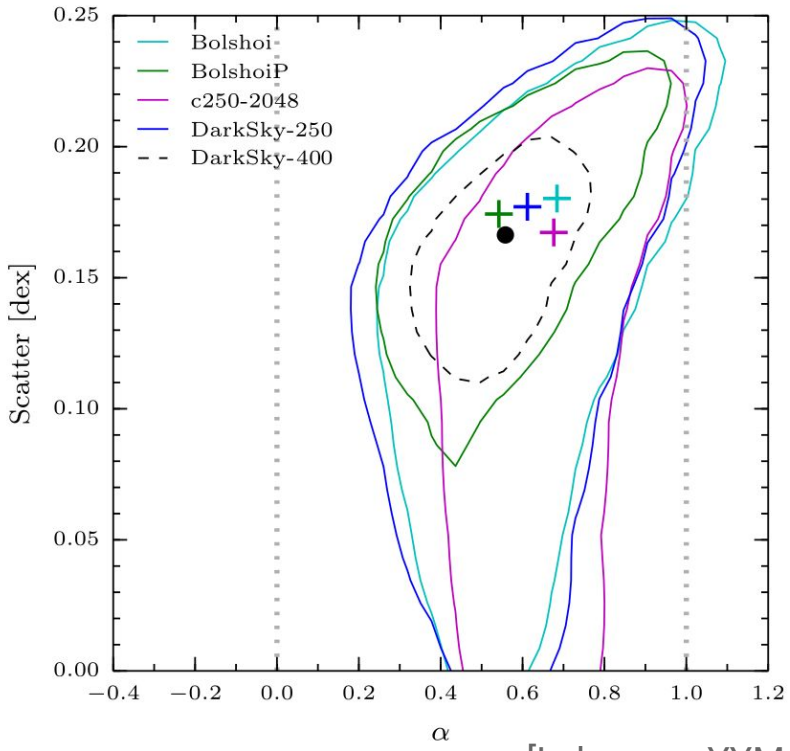
What have we learned?

➔Even with abundance matching, which has strong constraints from the framework itself (when compared to HOD, SAM, or hydro sims), there is still much freedom when mapping galaxies to dark matter halos [for HOD, cf. Hahn's poster on Thur]

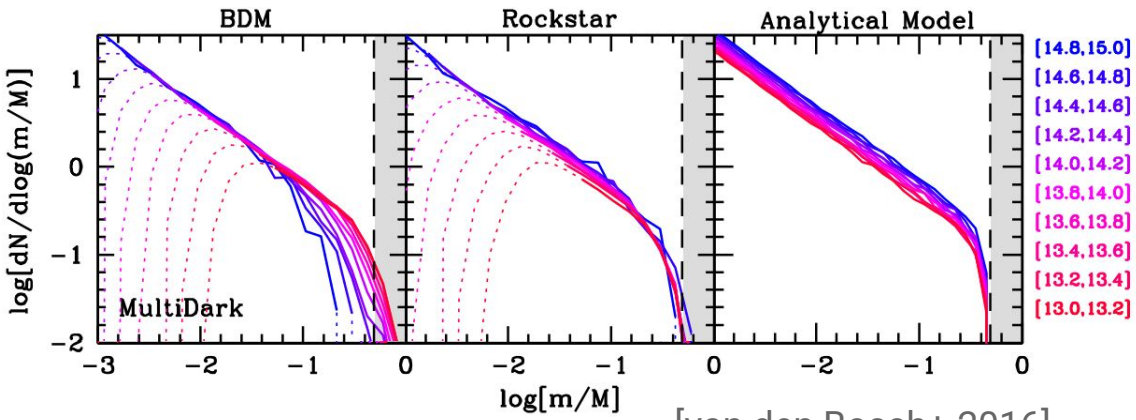
- ➔ Should avoid over-interpreting the physical meaning of a particular choice of matching proxy.
- ➔ This "freedom" in the galaxy-halo connection, when constrained by data, provides insights into galaxy formation physics:
  - do galaxy properties depend on halo formation history (at fixed mass)?
  - does initial stripping of subhalo strongly affect the galaxy within?
- ➔ The "freedom" should be also marginalized for other inferences.

Challenges for simulations

- ➔ Finite volume of simulations results in stochastic errors (sample variance)
- ➔ Finite resolution of simulations results in systematic errors (bias)



[Lehmann, YYM+ 2015]

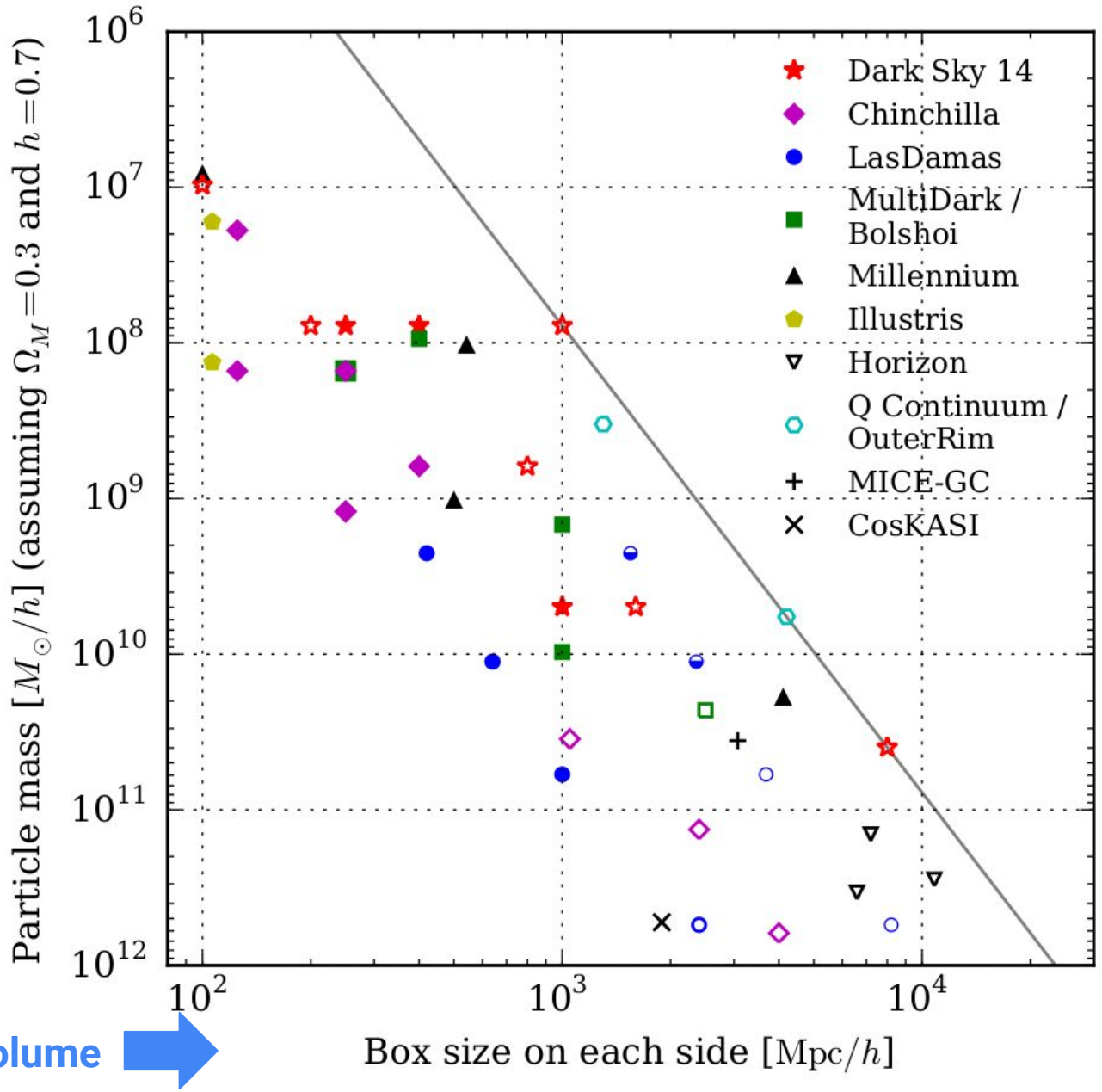


[van den Bosch+ 2016]

The computational cost of a dark matter-only simulation is ~ $N \log N$

Given limited resources, it is always a trade-off between:

- a higher resolution
- a larger volume



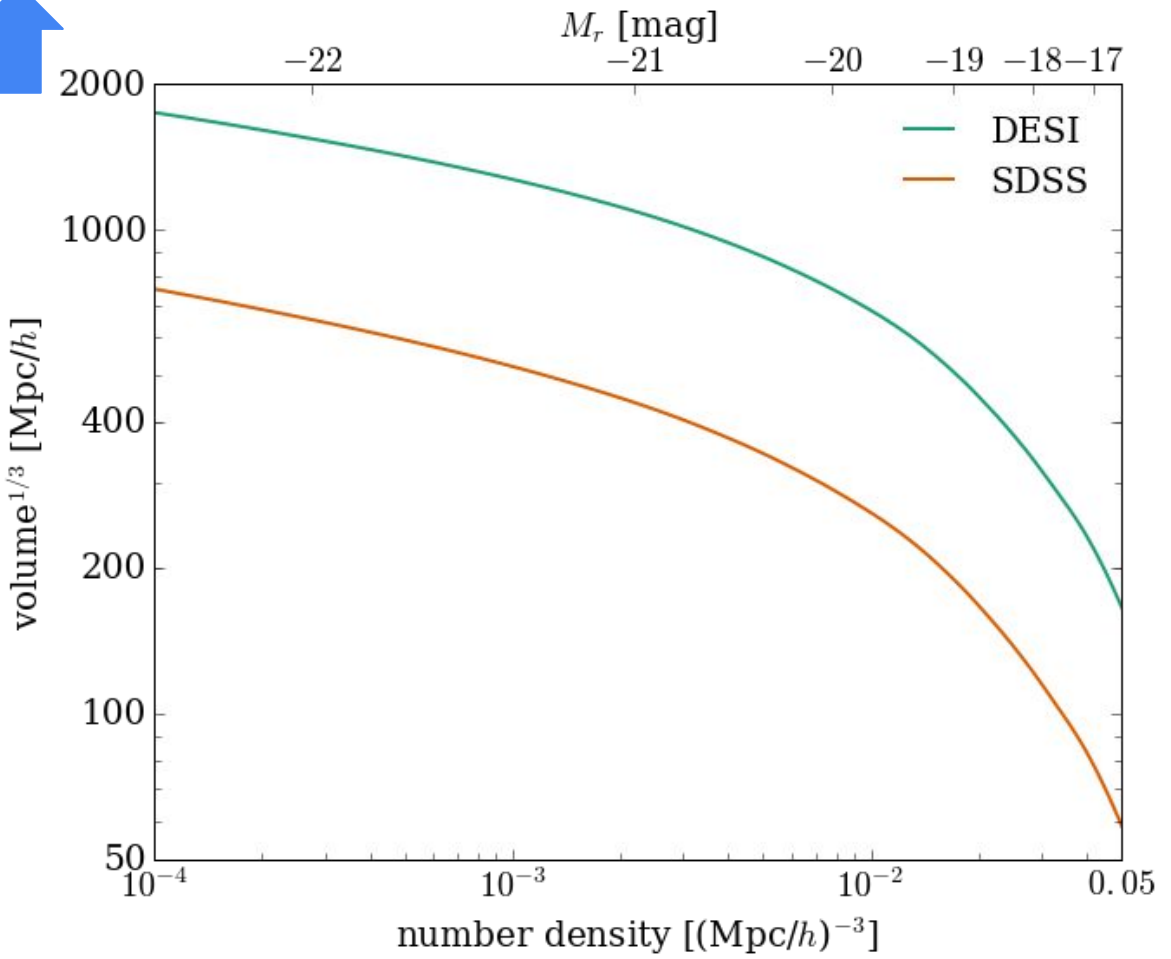
Trade-off in cosmological simulations

Resolution requirements

- ➔ To construct a volume-limited samples from a magnitude-limited survey, the dimmer galaxies we look at, the smaller the total volume is.
- ➔ To model dimmer galaxies, we need to resolve smaller halos

Larger volume ↑

Dimmer galaxies →



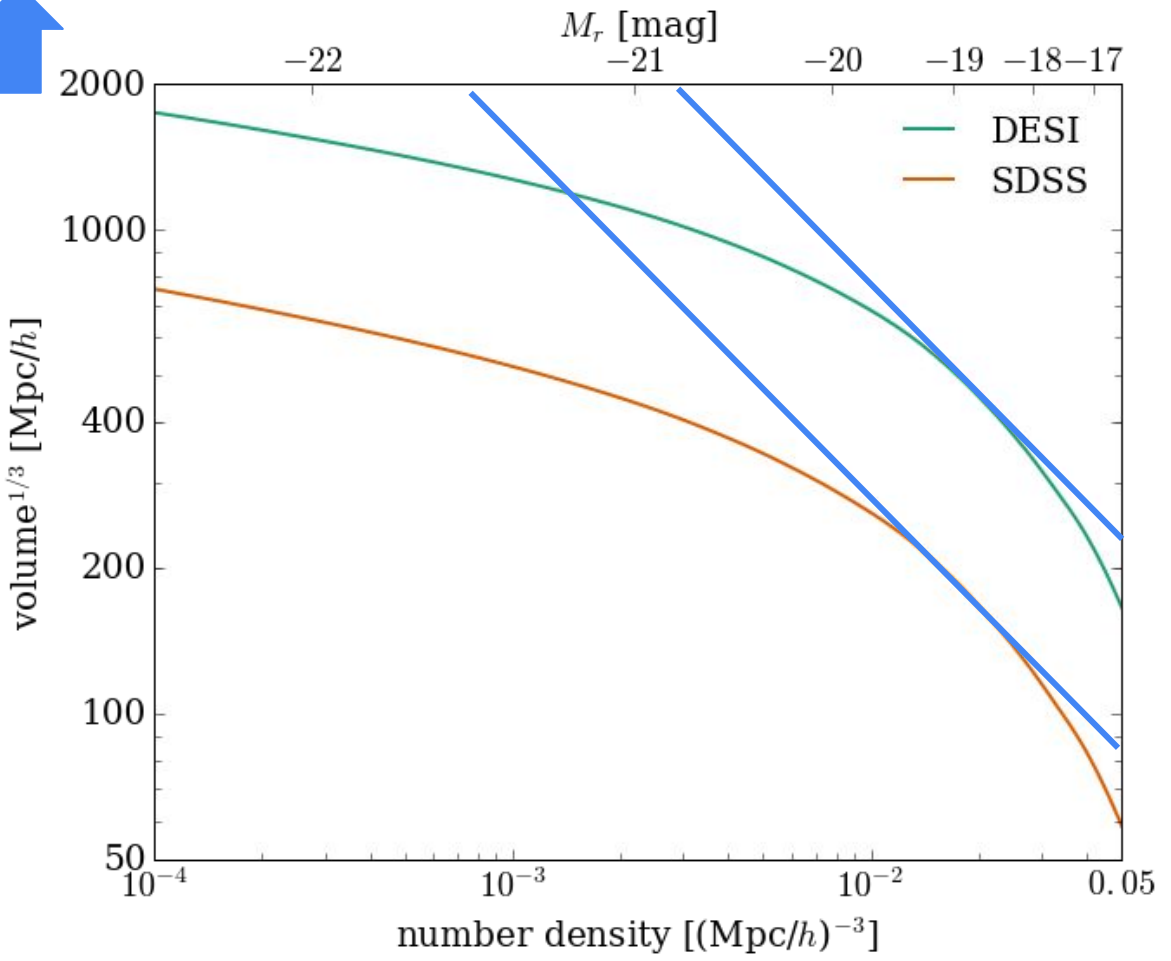
[YYM, Yang+ (in prep.)]

Resolution requirements

- ➔ The sample variance (stochastic error) increases as the volume decreases
- ➔ Hence, the most informative sample is ***not*** the dimmest sample
- ➔ Simulations should balance b/w volume and resolution

larger volume (less stochastic error)

dimmer galaxies (more systematic error)



[YYM, Yang+ (in prep.)]

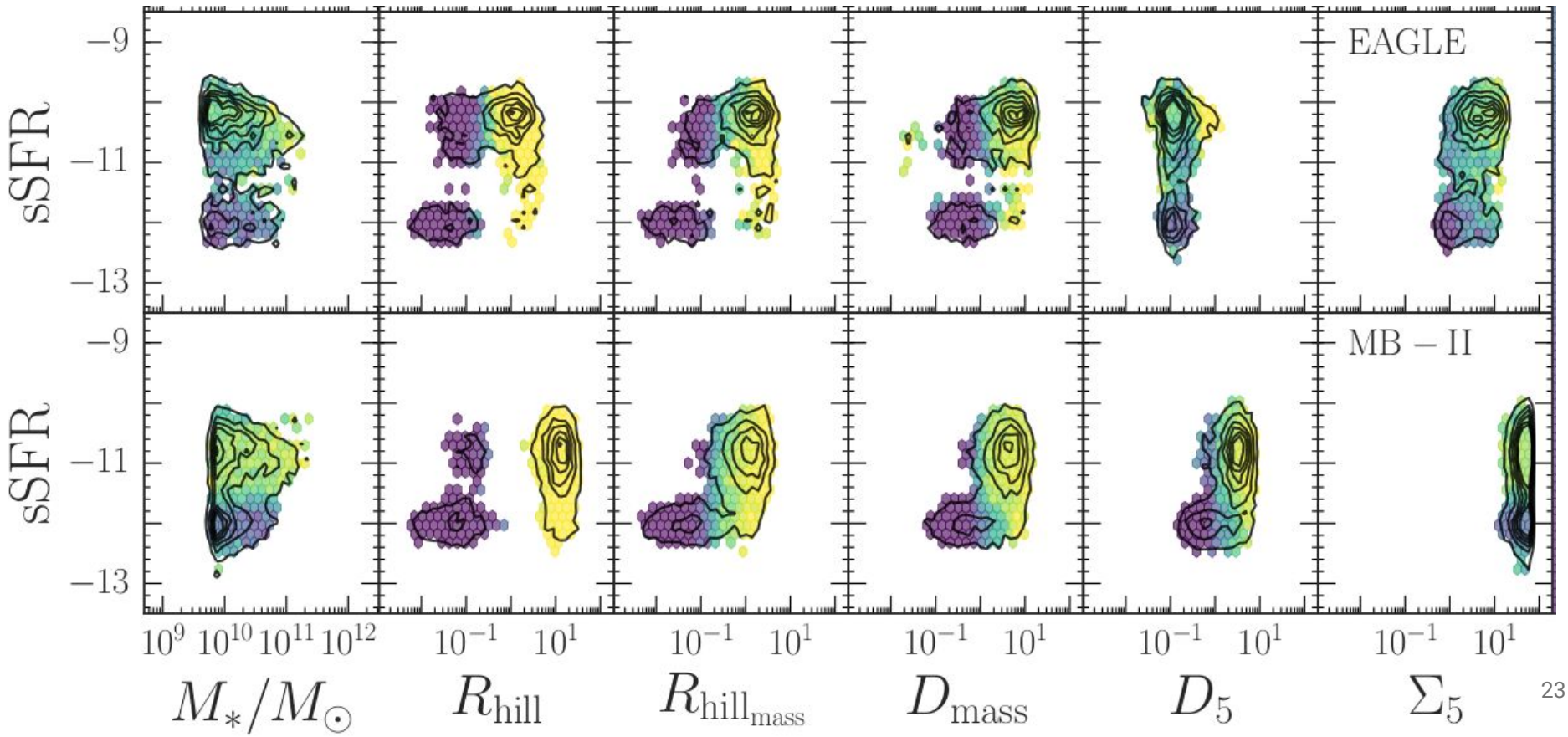
Despite the challenges, there are still other ways to tackle the galaxy–halo connection for specific questions

- ➔ Can we infer other galaxy properties if its luminosity is already known? Color? Morphology?
- ➔ Can we infer halo properties for specific galaxies from their observed properties?

Inferring galaxy color from large-scale density

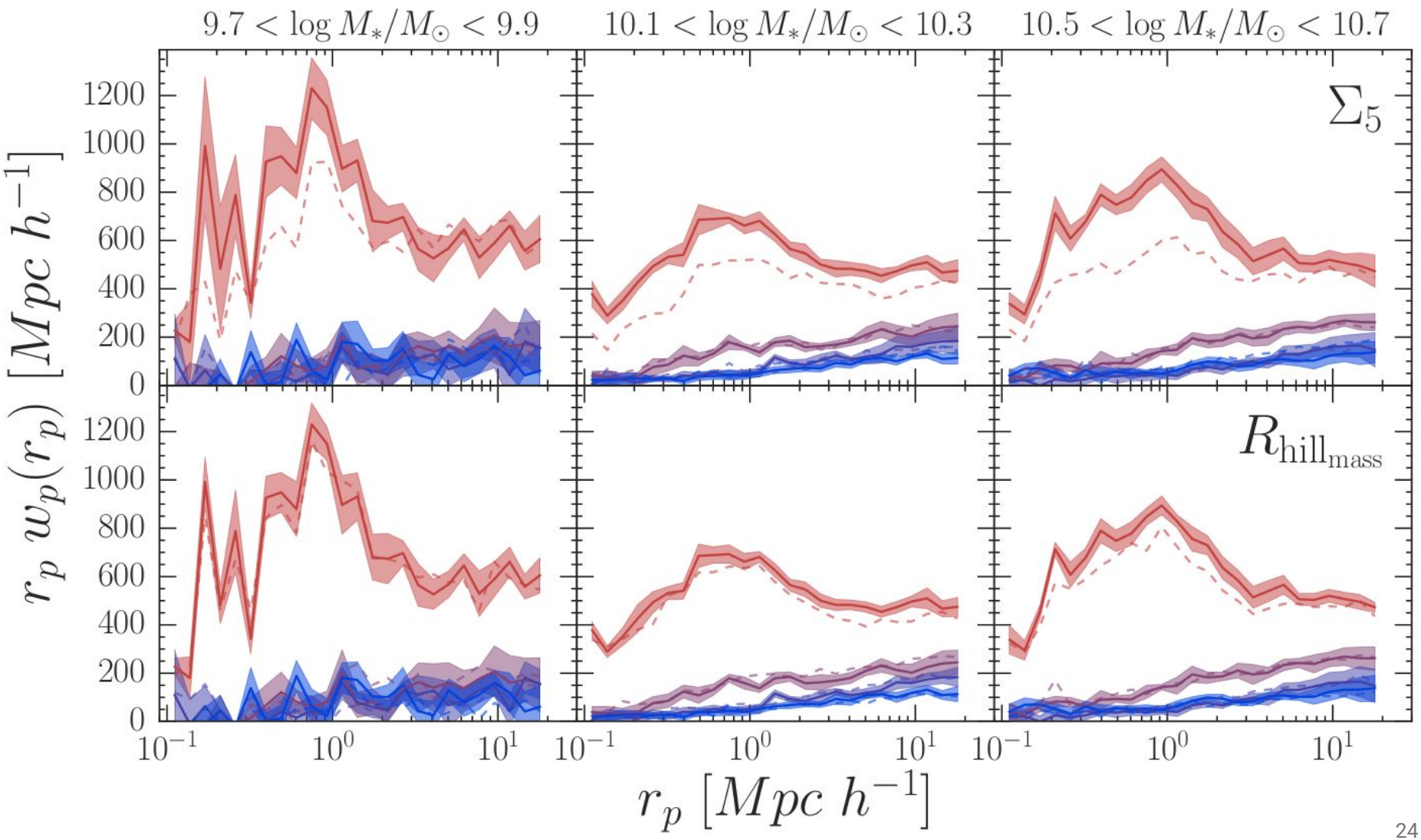
- ➔ Galaxy color and its environment are correlated [cf. Chen's talk on Wed]
- ➔ Can we find informative "features" that are predictive for galaxy color?
- ➔ Can we find "features" that are less sensitive to resolution limits?

[Su, DeRose, YYM+ (in prep.)]



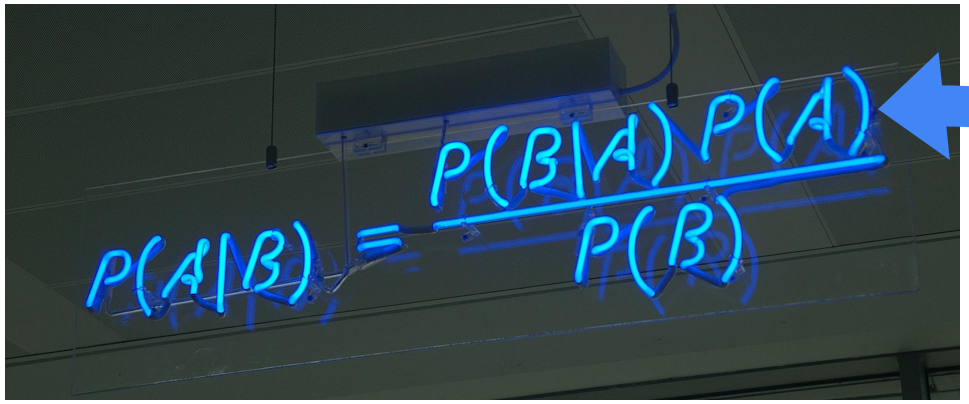
Predicting clustering signals for galaxies of different colors (sSFR)

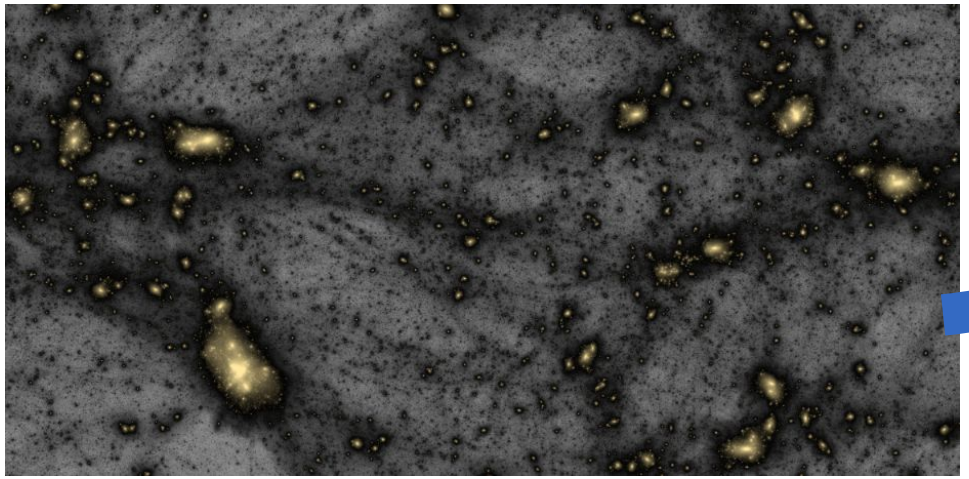
[Su, DeRose, YYM+ (in prep.)]

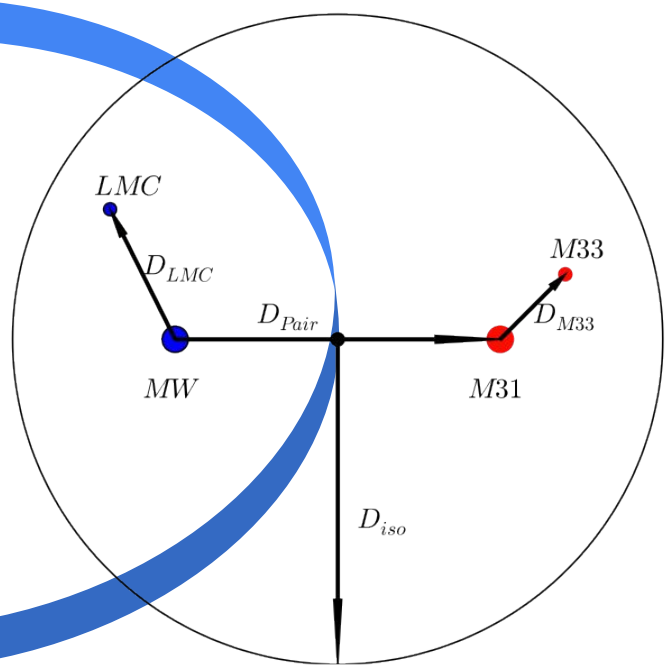


Inferring halo properties from kinematics

- ➔ The kinematics of our neighbor galaxies provide information on the mass profile of the dark matter halo [cf. Eadie's talk today]
- ➔ How to construct a proper prior? Can simulations help?



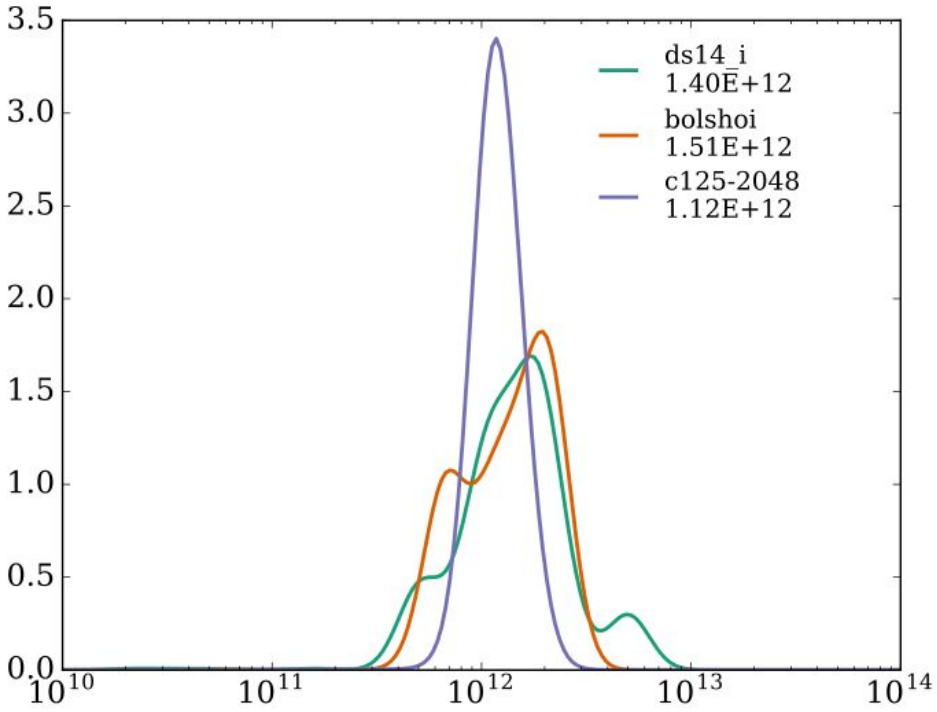




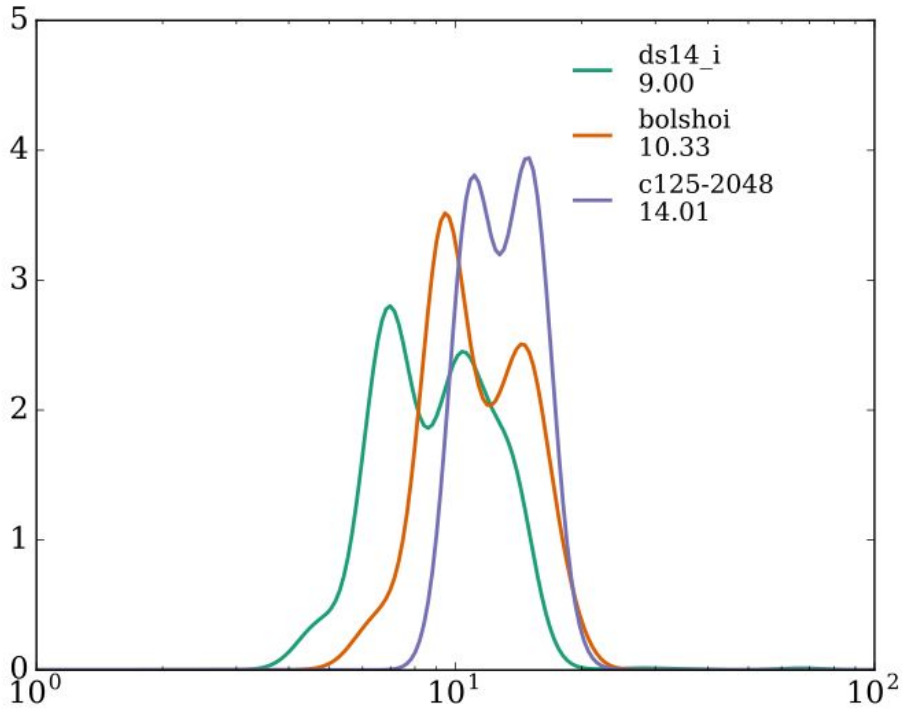
[Williamson, YYM+ (in prep.)]

Posterior distributions of halo mass (left) and concentration (right)
Using LMC and SMC kinematics and maximal circular velocities
(update to Busha+ 2011)

[Williamson, YYM+ (in prep.)]



halo mass



concentration

Summary

and thank you!

Yao-Yuan Mao (Stanford ➡ PITT)
@yaoyuanmao | **yymao.github.io**

- ➔ Dark matter simulations provide robust and scalable predictions of the matter distribution for a given cosmology
- ➔ Need to model or to marginalize over the uncertainties in the galaxy–halo connection
- ➔ Empirical models such as abundance matching provide a framework for the above (but we're not doing it yet)
- ➔ Trade-off b/w resolution and volume is always challenging. Need to balance different sources of error
- ➔ Other statistical/ML methods can be applied to simulations to probe the galaxy–halo connection