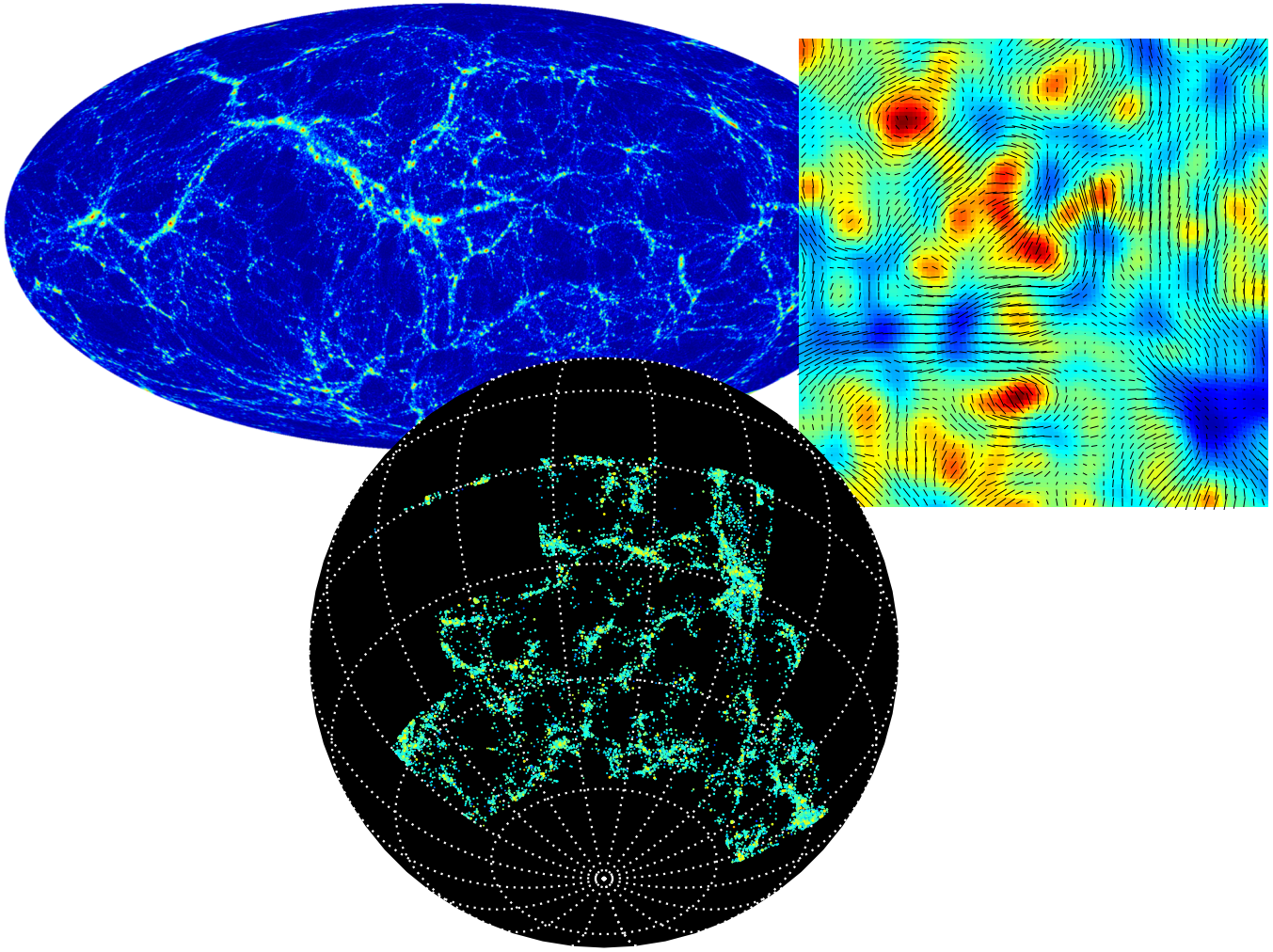
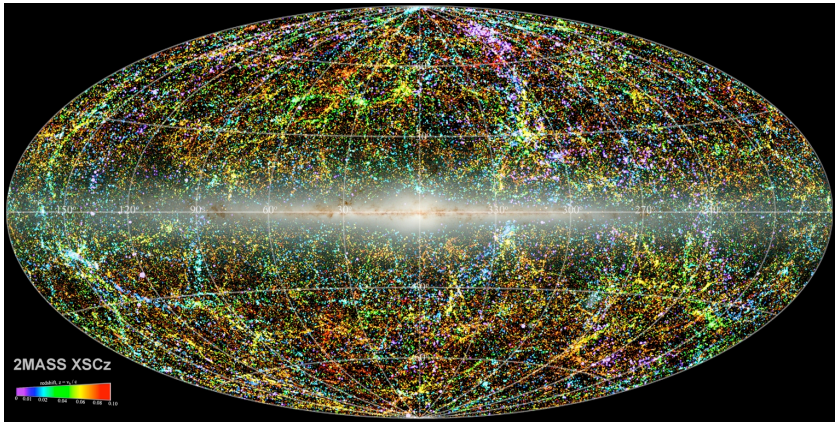
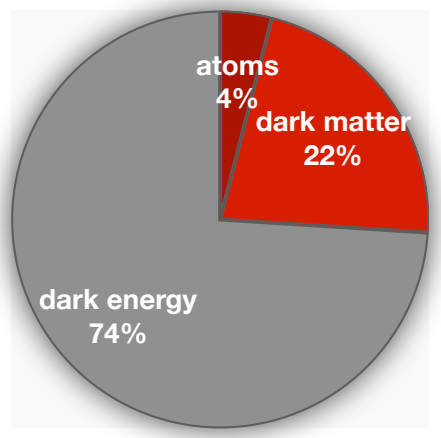
Simulated Galaxy Catalogs for Deep Wide Surveys
methods, prospects, challenges



Risa Wechsler
Stanford/SLAC/KIPAC

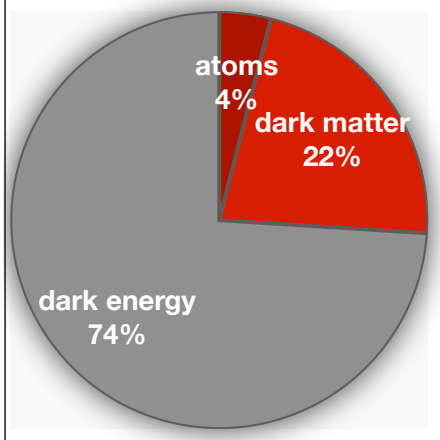
probing fundamental physics with galaxy surveys



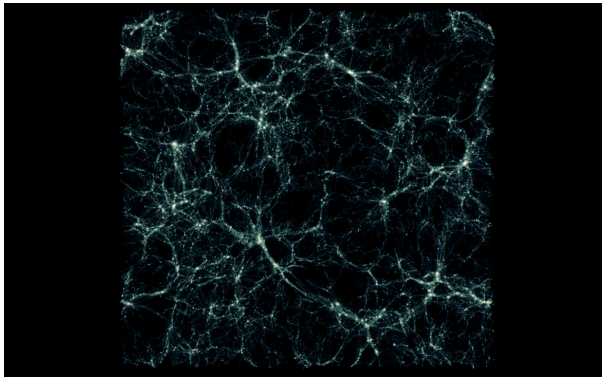


we will map out the formation of structure with next generation surveys

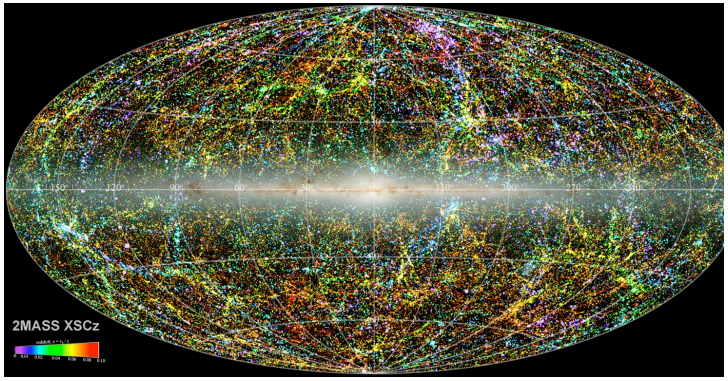
infer the cosmological parameters and test the standard cosmological model



theory



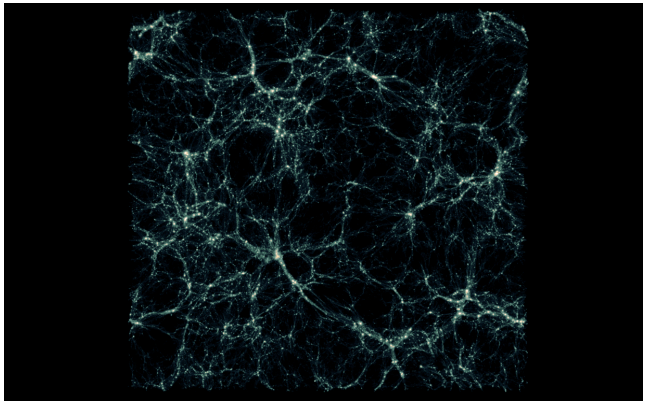
predict the formation of structure

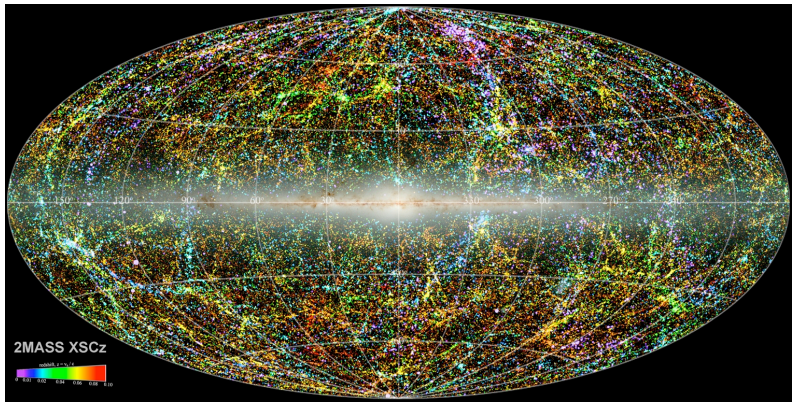


and its relationship to observations of galaxies

- What is the state of the art?
- What are the dominant systematics that limit our ability to predict the key observables?
- What new observations could improve/inform these predictions?

the galaxy - dark matter connection





- **ideal:** *predict* the galaxy population for a given cosmological model e.g. P(k | L, SFR, color, etc) from physical principles

- **practical:** describe the galaxy population for a given cosmological model with a flexible parameterization; marginalize over this parameterization for the possible galaxy population when constraining cosmology

Simulated Sky Surveys

Want simulations that allow a realistic cosmology analysis for the main dark energy probes

- galaxy clustering & baryon acoustic oscillations
- cluster abundance and clustering
- lensing: shear-shear correlations; galaxy-galaxy lensing; cluster mass calibration
- cross-correlation between galaxies and the CMB
- etc

Want to produce a realistic simulated sky

- observed properties of galaxies
- large-scale structure of galaxies
- realistic impact of lensing shear on galaxies
- as many relevant observational systematics as possible

Want to produce *many* full area and depth sky surveys; need lightweight simulations

- many cosmological models
- a variety of galaxy models for a given cosmology
- multiple skies for covariance

the galaxy - dark matter connection: methods

hydrodynamical simulations

resolve all histories, with baryonic physics
typically ~ factor of 10-100 more expensive than dark matter only
essential! but not yet predictive, and unlikely to be for cosmological volumes.

semi-analytic models

based on merger trees from dm simulations + simple prescriptions for galaxy formation physics
"physical models", but only somewhat predictive

empirical models

what galaxy population is consistent with observables, given the dark matter distribution?
halo occupation / conditional luminosity function: resolve all host halos
(sub)halo abundance matching: resolve subhalos; merger trees
dark matter density based (to extend resolution)

the galaxy - dark matter connection: key issues in modeling 1 - 100 billion galaxies

- how much volume?
  - surveys probing few to ~hundred of Gpc^3. would like to simulate a given survey hundreds (thousands?) of times.
- how faint? do you need a full magnitude limited survey?
  - typical $L > 0.1\text{-}1L^*$ galaxies are used for cosmology. however, correlations / projections with dimmer galaxies can be important for understanding systematics
  - modeling *all* galaxies requires ~ 1e7 M_{sun} halos; galaxies used for cosmology live in > few e 9 M_{sun} halos and subhalos; centrals of ~ few e11-12 M_{sun} halos
  - what galaxy properties are important?
  - luminosity / stellar mass, color / star formation rate, morphology / internal structure / surface brightness?
- what correlations are important?
  - significant power in joint probes and in smaller scales. requires understanding e.g. correlations between halo occupation & environment; substructure properties and accretion histories;

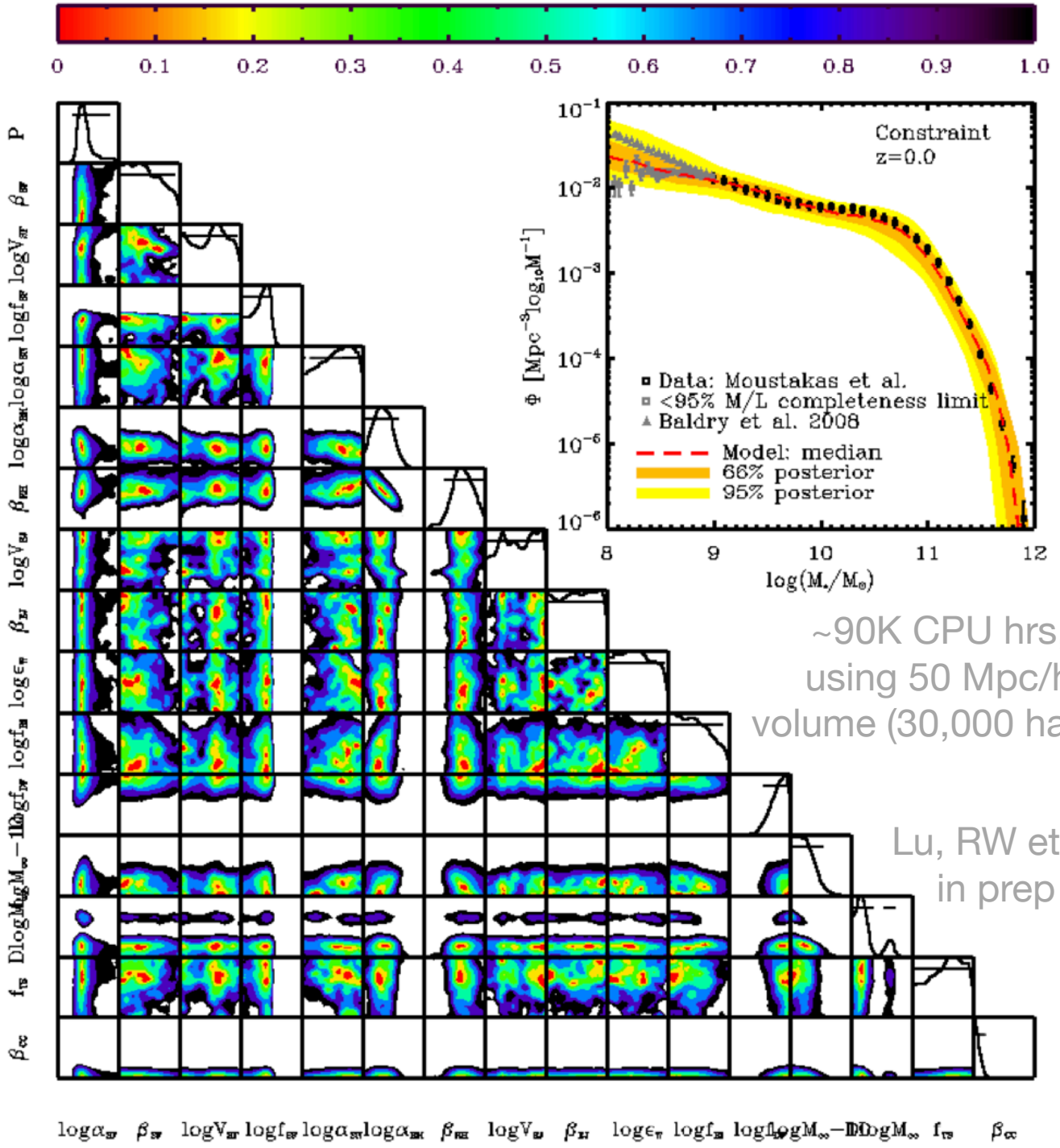
Bayesian SAMs

e.g. Lu et al 2011, 2012
Henriques et al 2012

substantial progress but still significantly behind existing data.
e.g., not yet done with clustering constraints, evolution difficult, color distribution difficult, etc.

need to constrain models with a large range of existing data. requires substantial computation!

balance between flexibility and how physical the model is

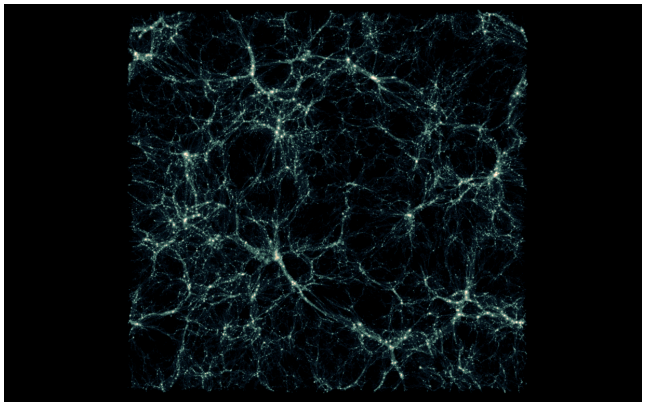


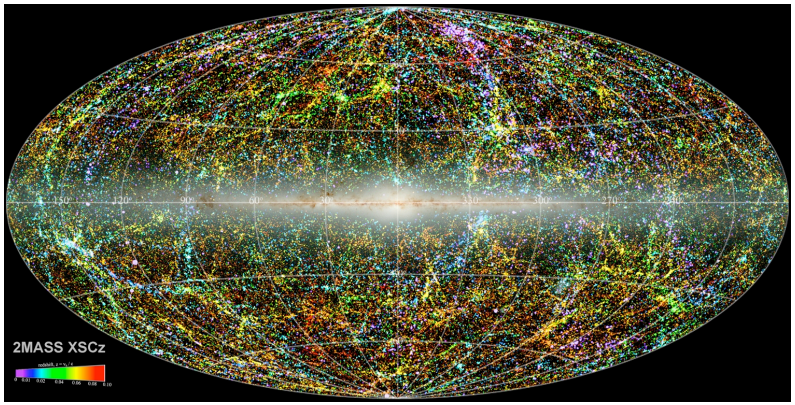
~90K CPU hrs, using 50 Mpc/h volume (30,000 halos)

Lu, RW et al in prep

these data *can* be modeled accurately with simple empirical models that use a small number of parameters, given simulations that resolve all halos and subhalos

empirical models





what is the statistical connection between what we predict for large-scale structure and what we measure for the galaxy population?

the distribution of matter & galaxies can be described by:

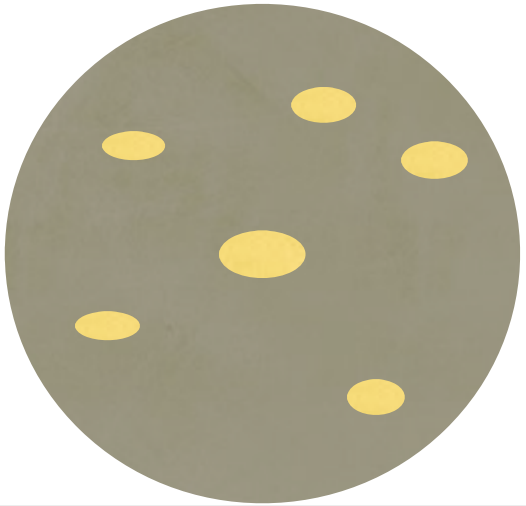
- the mass function of dark matter halos
- the clustering of dark matter halos
- the statistical relationship of matter / galaxies to the dark matter halos
  - P(N|M) **"halo occupation distribution"**
  - radial distribution of matter / galaxies in halos
  - velocity distribution of matter / galaxies in halos

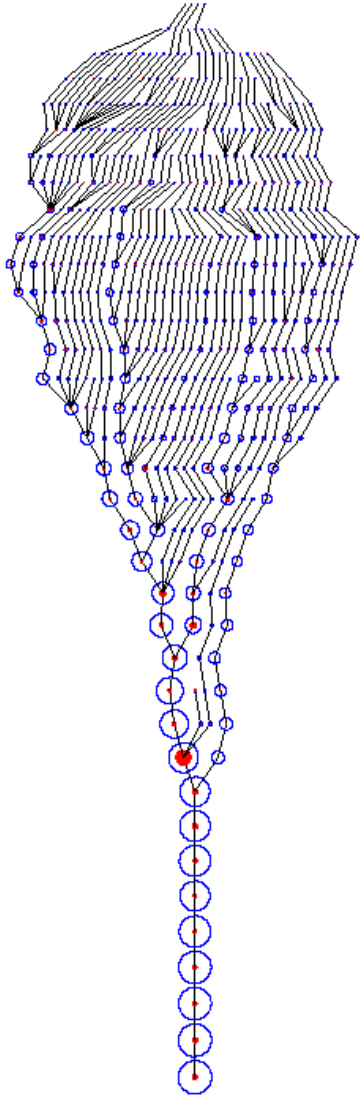
the basic assumption: stuff (matter, galaxies, etc) lives in halos

" the halo model"

The simple part of CDM Galaxy formation theory

- in collapsed dark matter halos, where the density is high, gas cools and sinks to the center to form a galaxy
- expect a galaxy at the center of each density peak massive enough to form stars
- dark matter halos and the galaxies within them merge
- expect one galaxy in every dm halo *and subhalo* massive enough to form stars

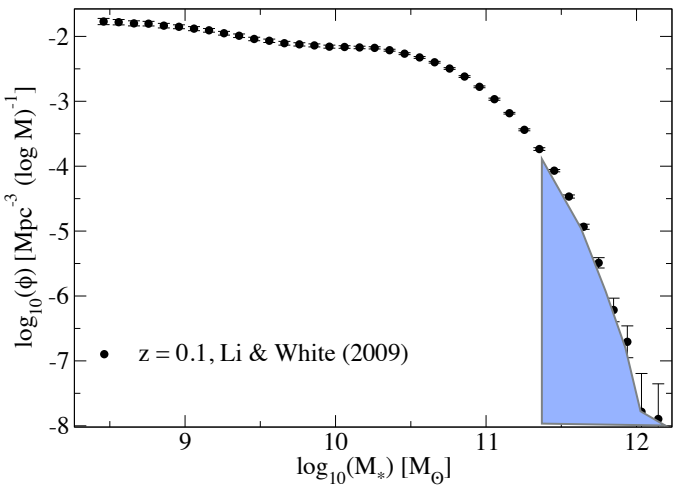


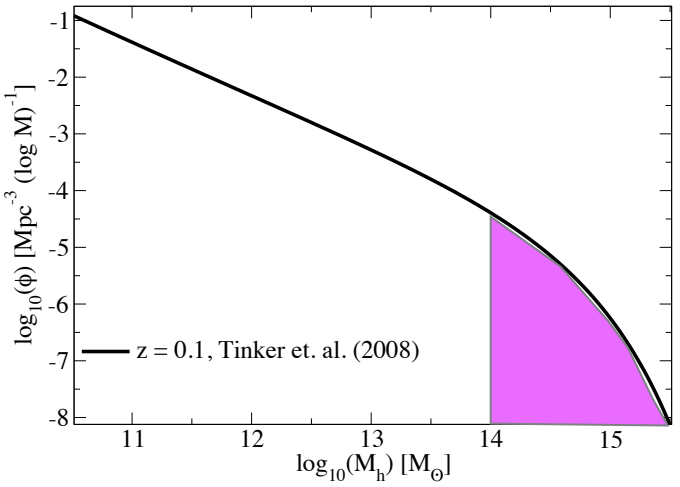


abundance matching technique

luminosity/stellar mass function

velocity/halo mass function



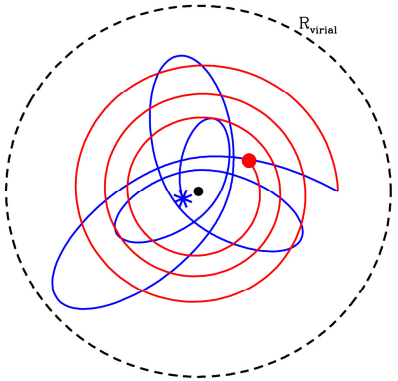


assign galaxies to halos by matching n(>M*) to n(>M)
(assume the most massive galaxy lives in the most massive halo)

key assumptions: one galaxy per dark matter clump (above some mass threshold); galaxy mass/luminosity tightly correlated with halo mass/velocity
requires ~ kpc res cosmological simulations

e.g. Kravtsov, Berlind, RW, et al 2004, Conroy, RW & Kravtsov 2006, Conroy & RW 2009, Behroozi, Conroy & RW 2010, Stewart et al 2012; also Vale & Ostriker 2006, 2007, Moster et al 2010, Guo & White 2010, Simha et al 2010, Wetzel & White 2010, Neistein et al 2011, Reddick et al 2012...

matching galaxies to halos and subhalos

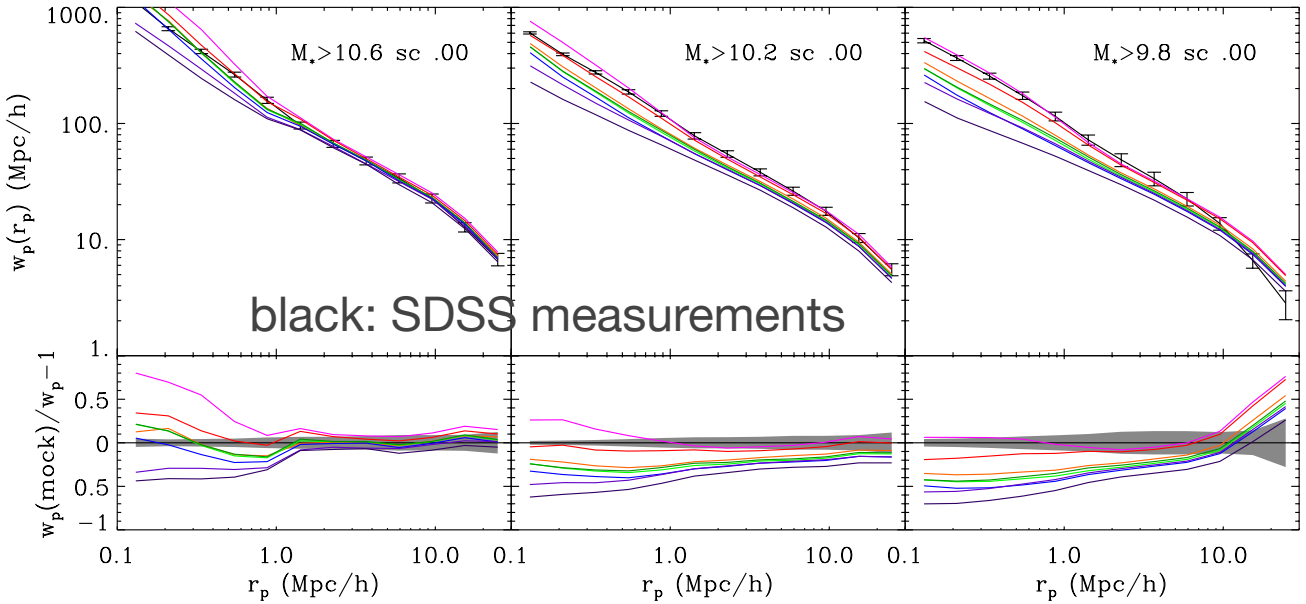


how much scatter in galaxy properties, at given halo property?

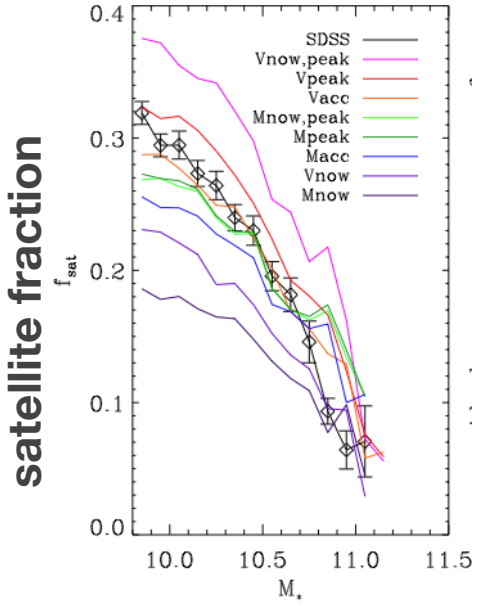
what is the relationship between the stellar mass of central galaxies vs. satellite galaxies?

how much can satellite halos be stripped by their host before they lose their galaxy?

these can be constrained with data!



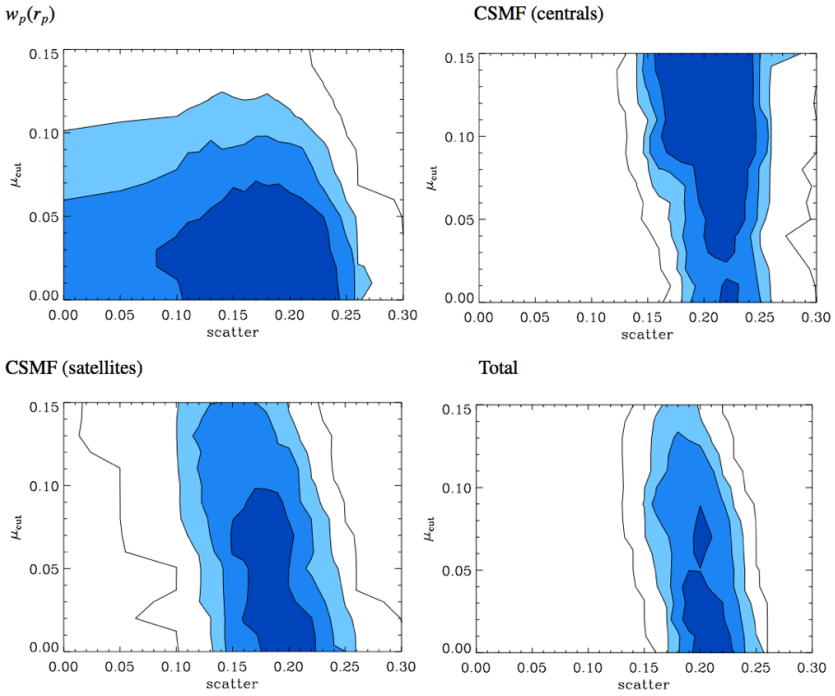
Reddick, RW, Tinker & Behroozi 2012



satellite fraction

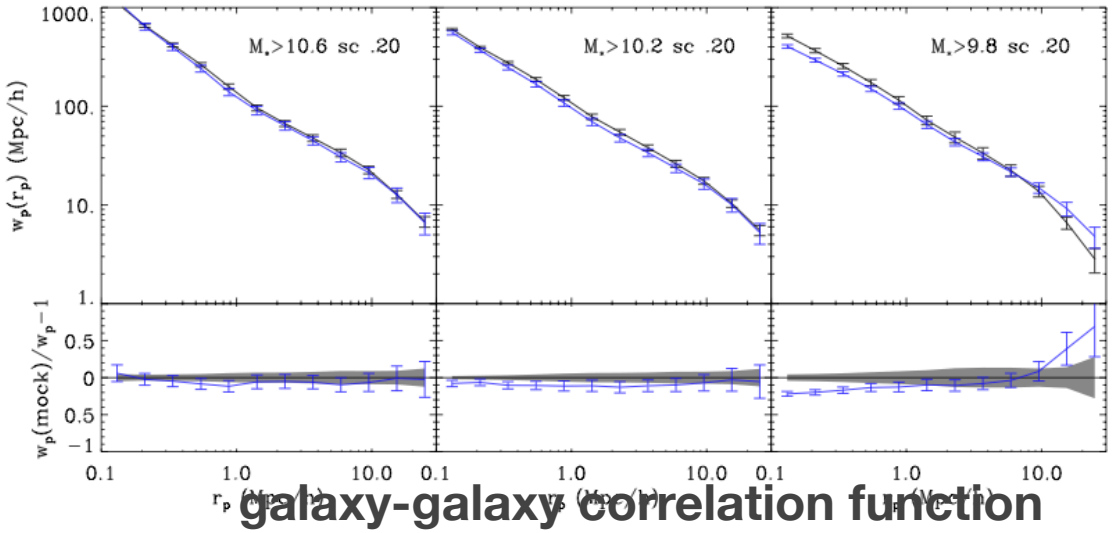
stellar mass

in high resolution simulations with merger trees: accurate models for galaxy-halo connection in the local Universe assigning galaxies to subhalos

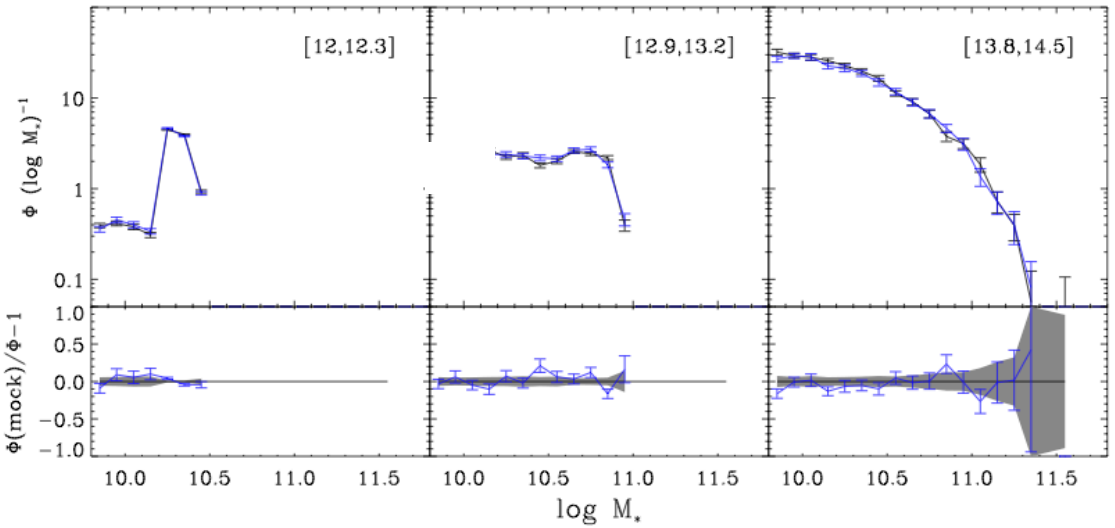


Reddick, RW, Tinker & Behroozi 2012
black: SDSS measurements
blue: best fit model

uses just two parameters for the galaxy model
to predict all clustering properties between galaxies and mass as a function of galaxy luminosity

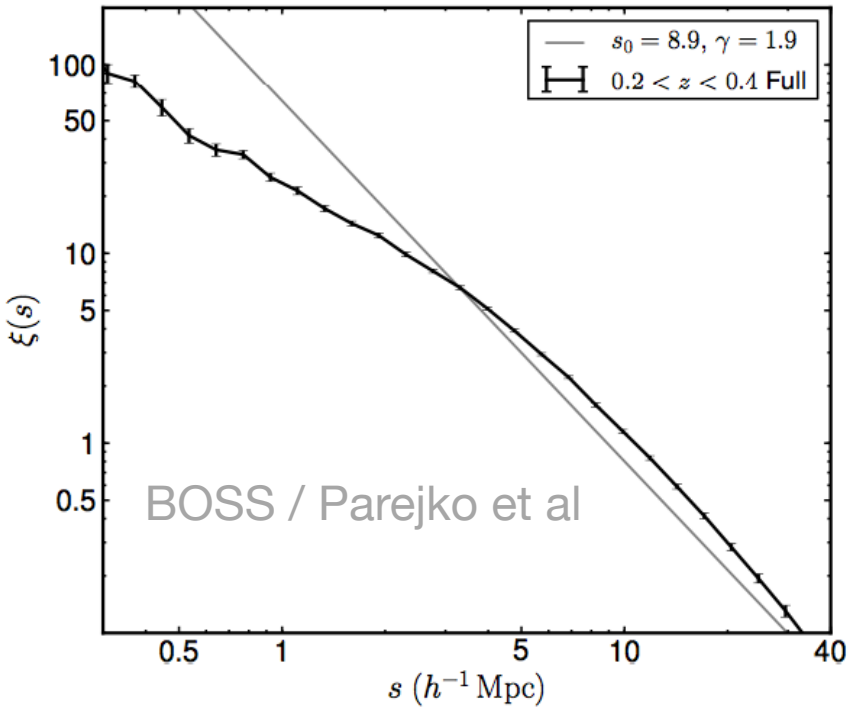


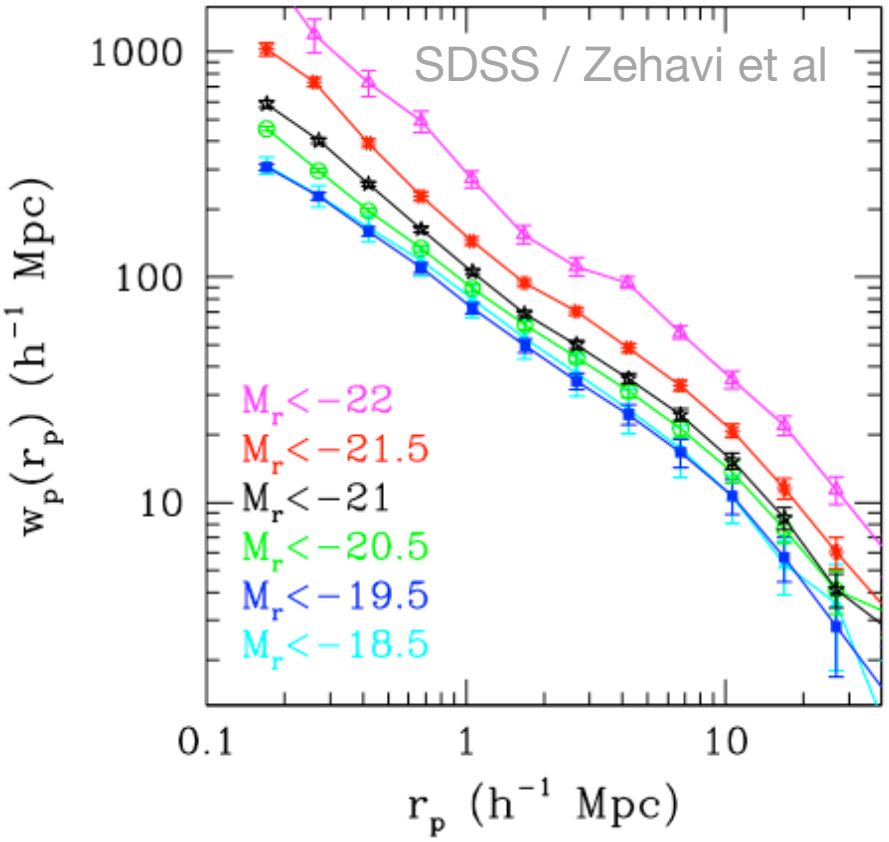
galaxy-galaxy correlation function



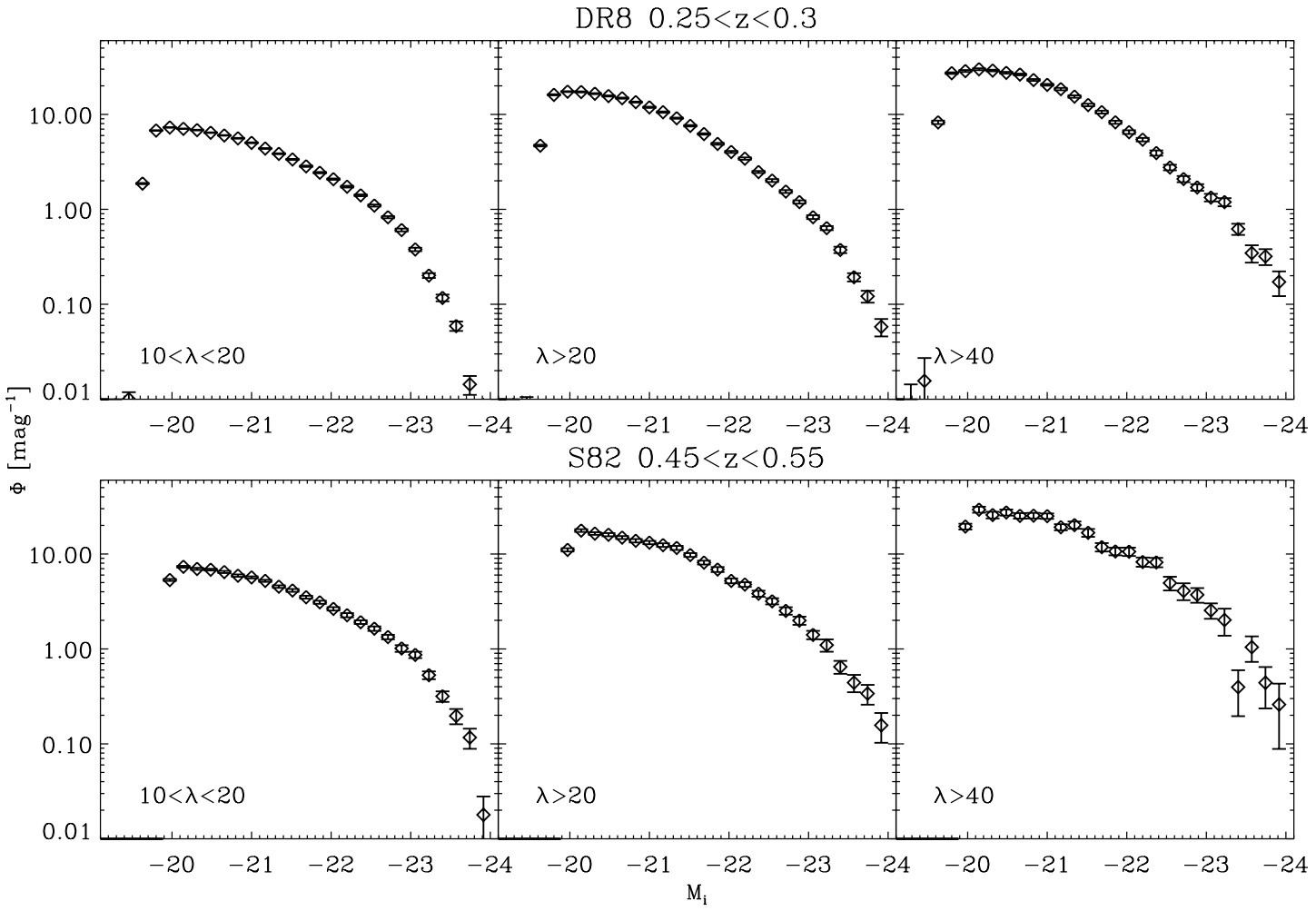
Conditional Stellar Mass Function

precise measurements of galaxy clustering





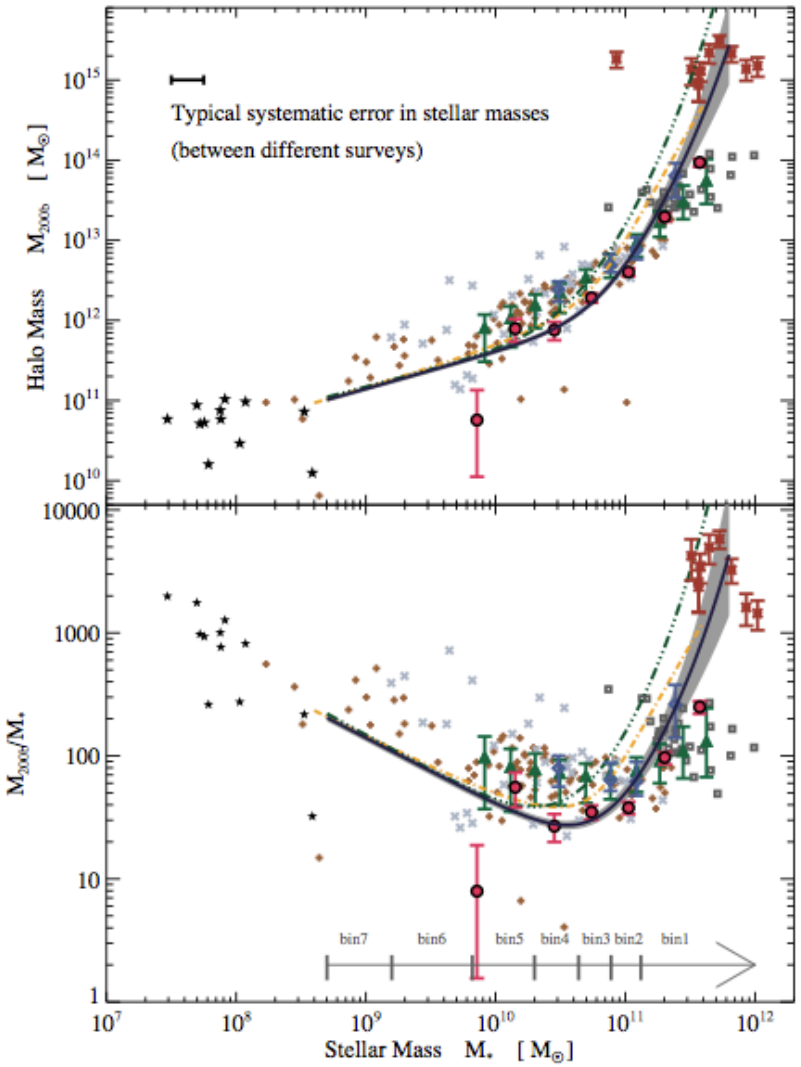
distribution of galaxies in clusters (observed conditional luminosity function)



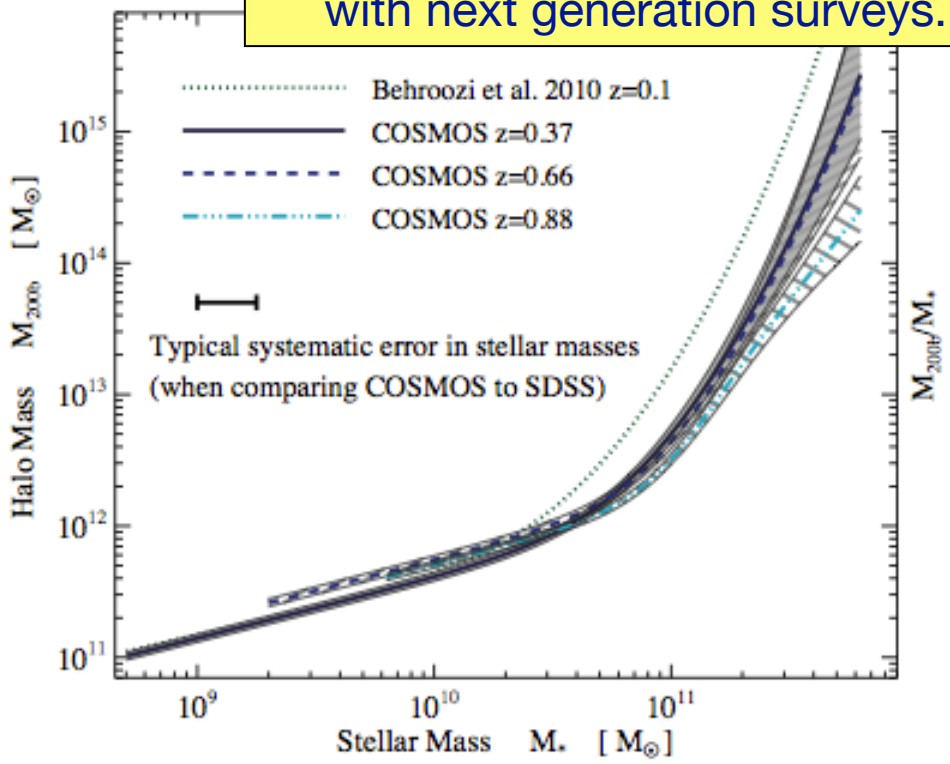
Reddick et al in prep
red galaxies in redmapper clusters, using new SDSS cluster catalogs (Rykoff, Rozo et al)

galaxy-galaxy lensing:
joint constraints on M*/M from stellar mass function, galaxy clustering, and galaxy-galaxy lensing from z=0.2-0.9

Leauthaud et al 2012
using data from COSMOS survey



this analysis is for 2 sq. degrees!
will be able to make very precise with next generation surveys.



uncertainty dominated by stellar mass functions

combinations and cross-correlations

- galaxy A - galaxy B cross-correlations
- galaxy-cluster cross correlations
- galaxy-CMB correlations
- data is really rich!
- predictions are really rich!

these data *can* be modeled accurately with a small number of parameters, given simulations that resolve all halos and subhalos...

it doesn't require hydro (for clustering), but it does require ~ kpc force resolution

the two parameter model I described applies to luminosity-selected samples at a given redshift.

it doesn't describe (by itself) arbitrary galaxy populations; need to think about missing correlations.

several empirical approaches to deal with galaxy formation *histories* are under development.
e.g. Conroy & Wechsler 2009;
Moster et al 2012, Behroozi, RW, Conroy 2013; Yang et al 2013
progress but a lot more work needed here and/or in flexible sams.

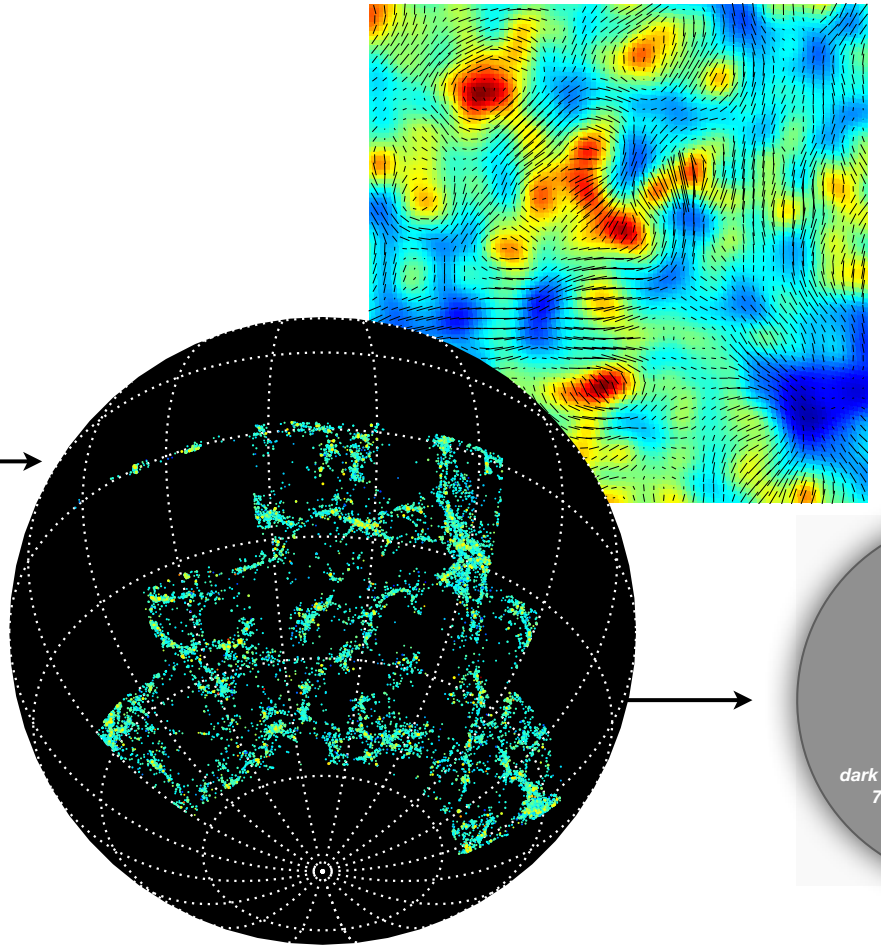
we are not yet at this resolution for large volumes!

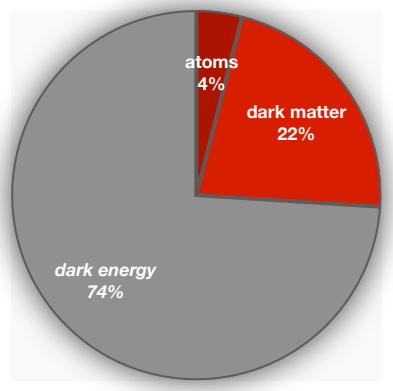
(requires ~trillion particle simulations, e.g. 30000^3 particles in 3Gpc)

Simulating full surveys

DES Mock Pipeline / R. Wechsler

- simulation
- halo finding
- merger trees
- assign galaxies
- calculate shear
- lens galaxies
- photometric errors
- photometric redshifts
- create sky images





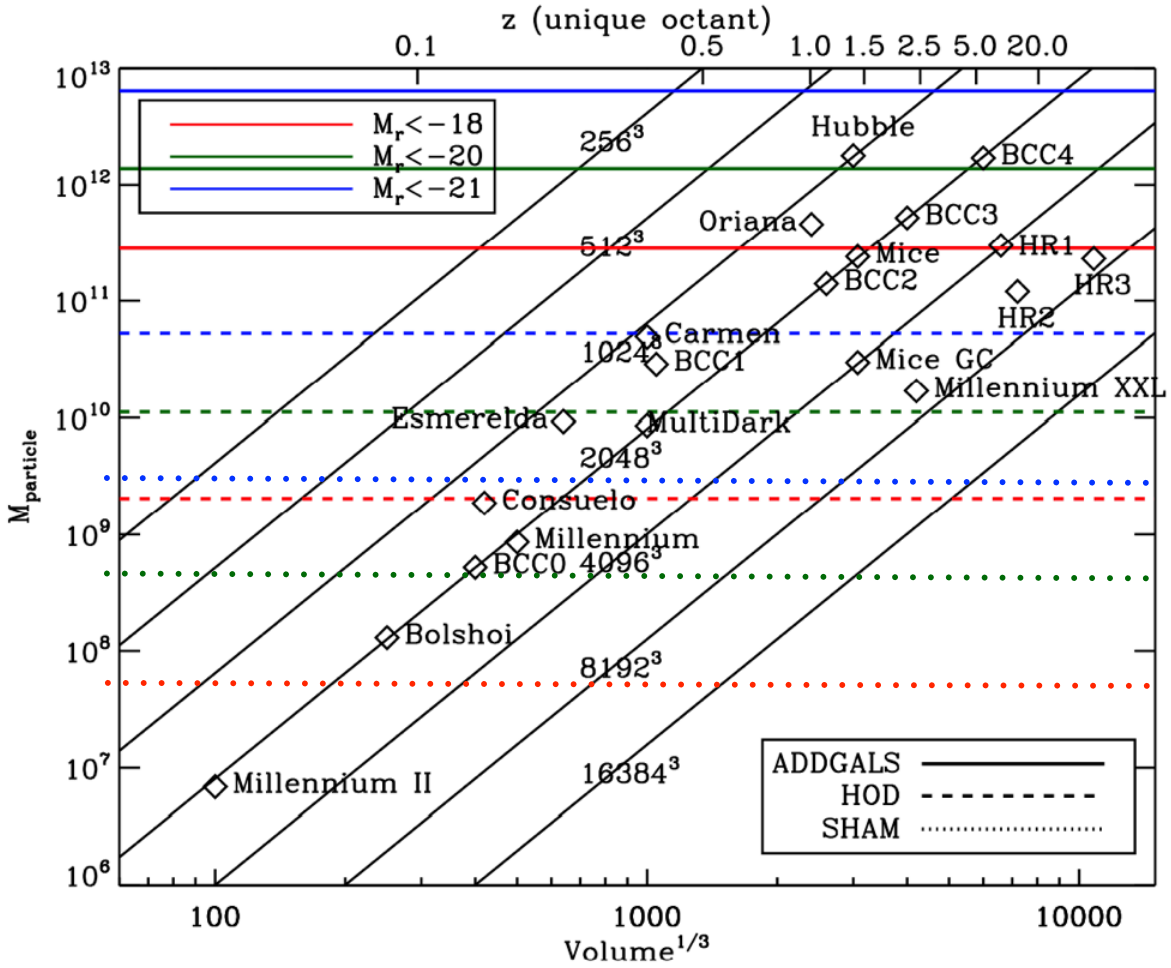
Wechsler, Busha et al 2013;
Busha, Wechsler, Becker et al 2013
with DES simulation working group

currently have a simulated catalog of ~1 billion galaxies (i ~ 25) over 1/4 of the sky, on a lightcone out to z=2, which includes galaxy photometry in many bands. galaxy magnitudes and shapes fully lensed along the lightcone. extensively tested with SDSS and other data.

Large area simulations “DES-BCC”

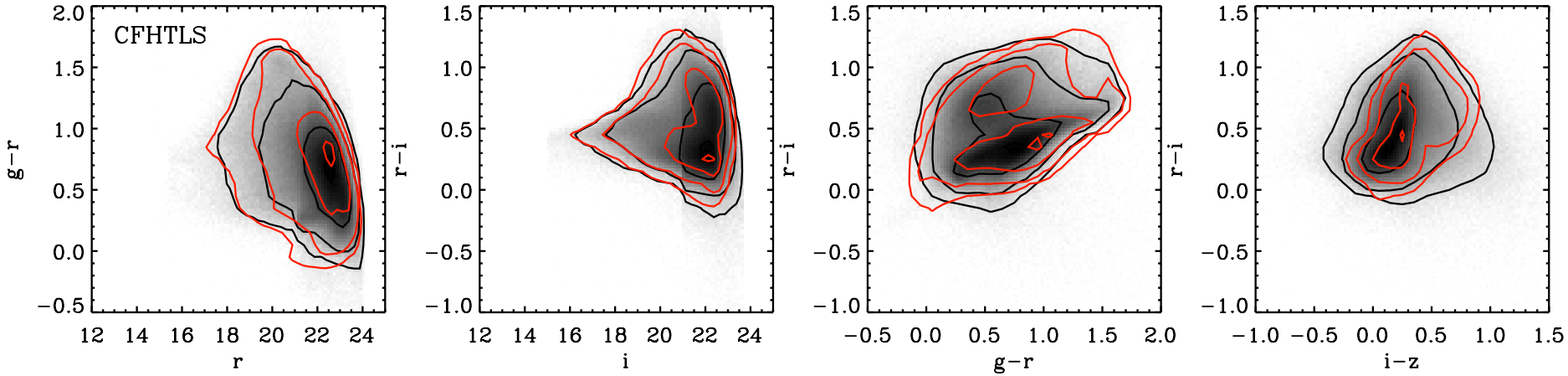
- current set up:
  - N-body lightcones to z~6 (based on 4 sim boxes with 2048^3 particles)
  - halo finding from rockstar, includes multiple mass def., concentrations, etc.
  - ~ 1 billion galaxies added using addgals, over 1/4 sky (10313 sq. degrees), complete to i ~ 24
  - photometry in many bands, including LSST bands and DES, SDSS (DR8+S82), VISTA (VHS +VIKING), CFHTLS, NDWFS, DEEP, WISE, IRAC
- shear on the full quarter of sky using CALCLENS; currently with 6.2” resolution
- extensive development and testing with SDSS data and other higher redshift data; designed to go to full DES depth (completely to a bit deeper than i ~ 24)
- stars included now, currently working on quasars
- planning ~ 10 cosmologies this year for DES with full galaxy and shear catalogs (sims already done for a few of them)
- allows science analysis related to clusters, weak lensing, LSS, photometric redshifts, spectroscopic followup design, etc.

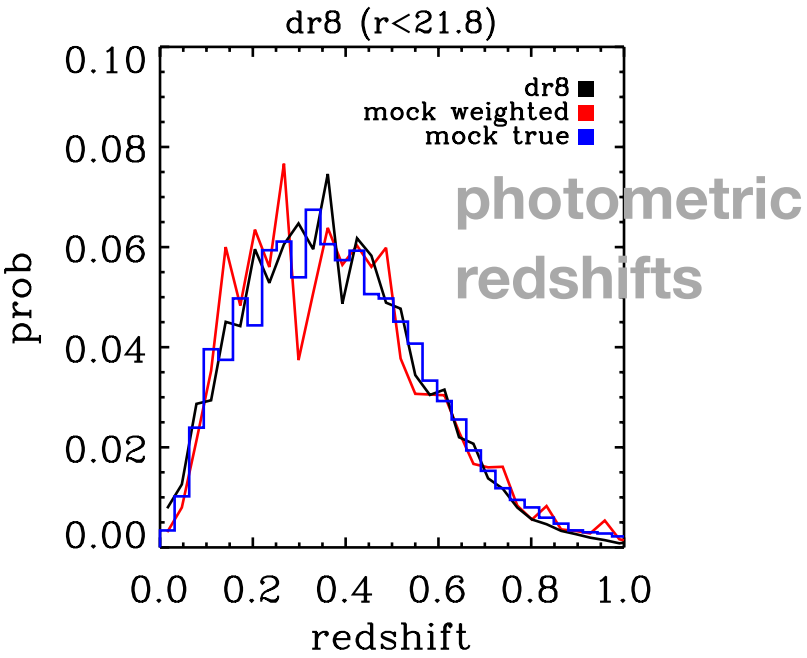
what is the maximally low resolution for the problem?



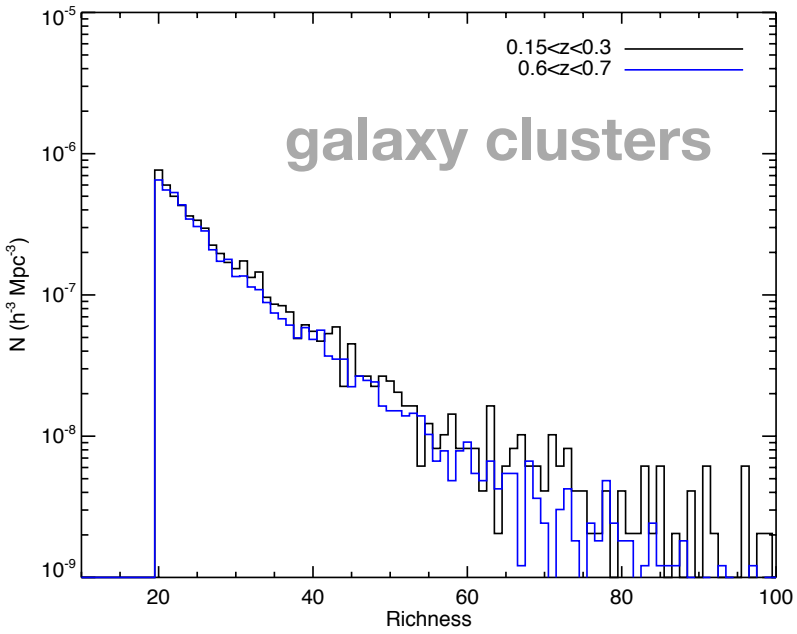
- ADDGALS model (Wechsler, Busha et al 2013) based on smoothed DM density. central / satellite difference is important.
- other recent approaches to push resolution, eg. de la Torre & Peacock 2012

galaxy colors and luminosities





photometric redshifts

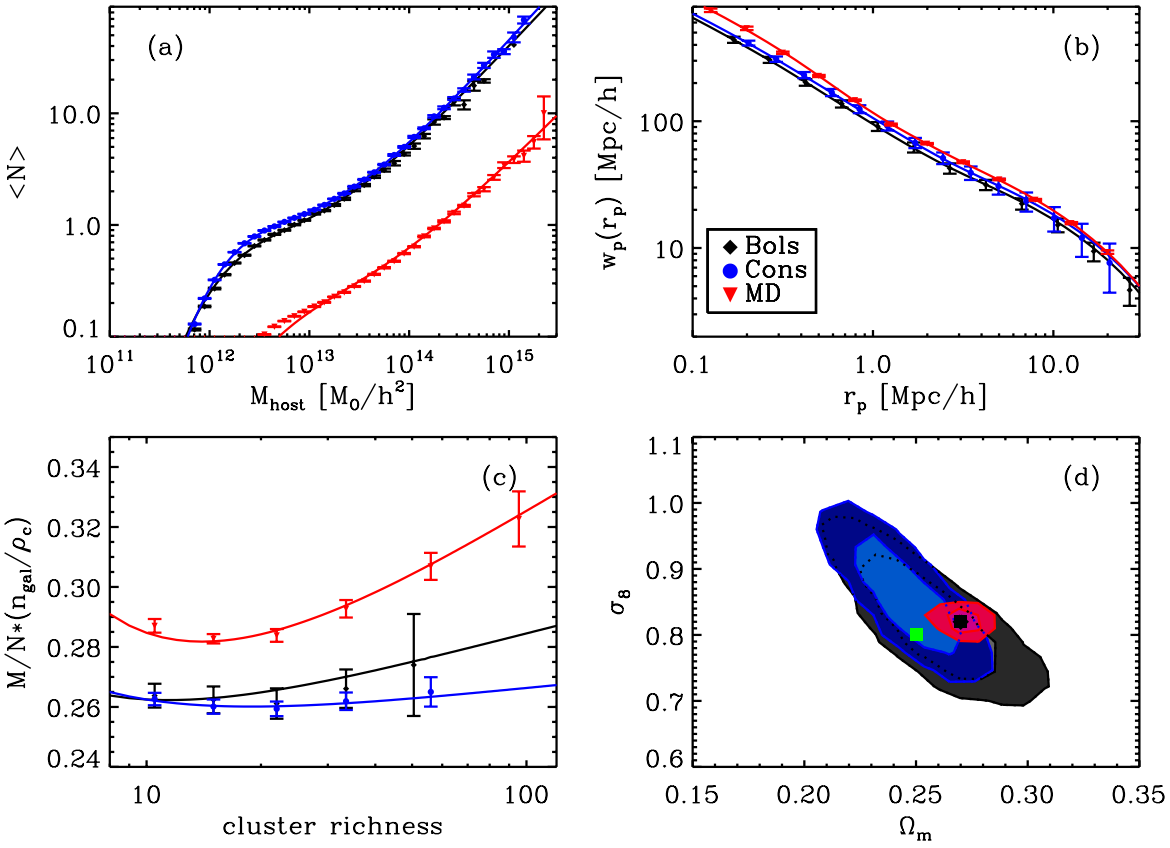


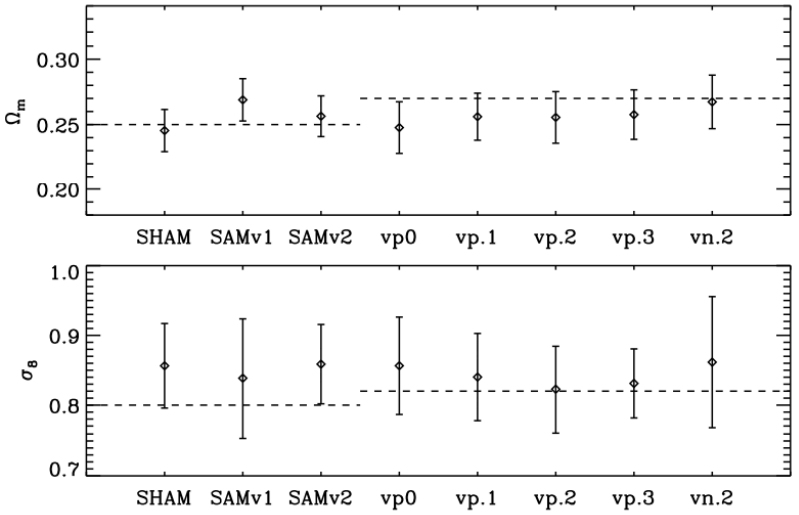
galaxy clusters

Do you get the same cosmological results for different galaxy formation models?

Reddick et al in prep

can be done, but for this probe (which includes small scales) requires very careful modeling of the halo occupation.





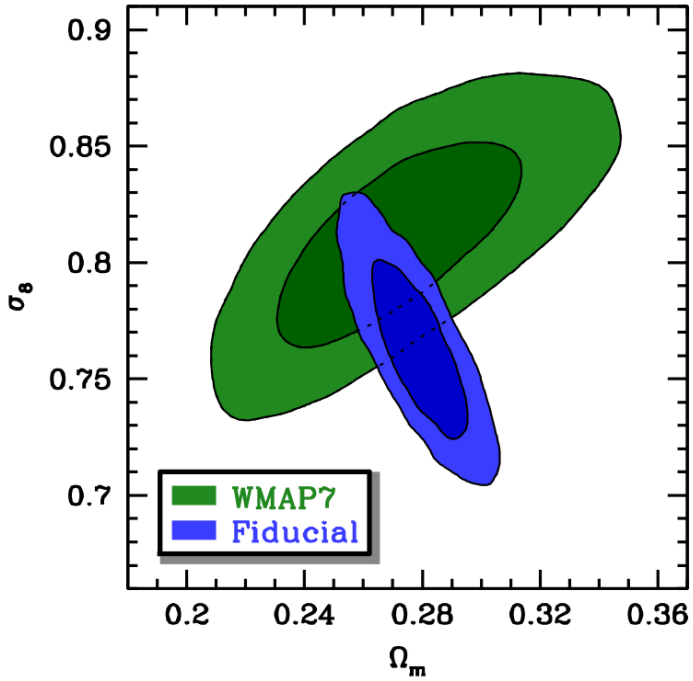
one example: combining galaxy clustering on small scales with mass to number ratio of galaxies in clusters

need to be doing this for every cosmological probe, for a range of galaxy prescriptions to understand which systematics matter.

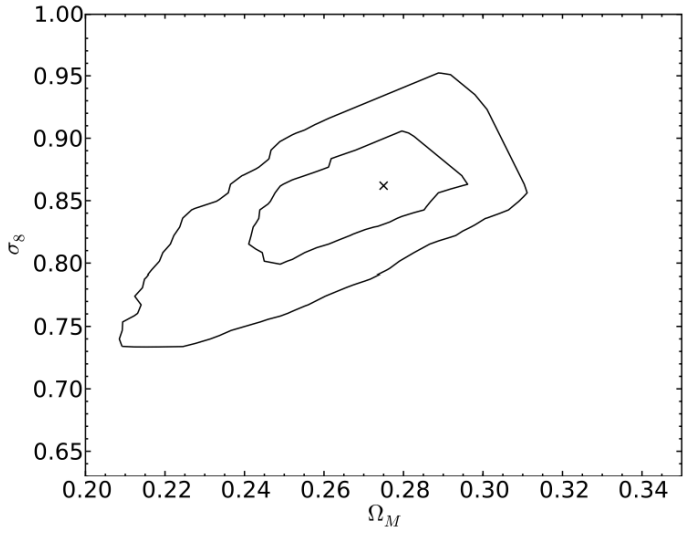
status and outlook

- current models
  - approaches that associate galaxies all with resolved dark matter structures work very well with few parameters, but requires high resolution
  - approximate approaches have been developed that require less resolution, can (just barely) model full surveys but lots of room for improvement!
  - over the next decade these approaches should be merged. with trillion particle simulations this is possible!
- observations (combination of clustering, lensing, galaxy properties in groups/clusters) are very constraining, and there is still a lot to be done with current data. models will be keeping up with data as it comes in from the surveys.
- next generation models should model full evolution histories of galaxies within halos/subhalos, but need to be flexible enough to fully explore parameters space
- next generation cosmology measurements using large-scale structure should be run in parallel on simulations to understand both theory and data systematics.
- there is a lot of work to do
  - accuracy matters!
  - needs substantial computing time
  - needs a different kind of workflow

other “small-scale” probes, just two examples



Yang et al; Cacciato et al 2013
combining galaxy clustering and galaxy-galaxy lensing



Simha & Cole 2013
just combining galaxy clustering and an assumption that galaxies live in halos and subhalos (maybe too specific in this case)

not yet mature, but have a lot of potential.