Lighting up the Dark:
Connecting Galaxies and Mass

Risa Wechsler
Stanford University & Stanford Linear Accelerator Center

U of A
March 19, 2007
A “Standard” Cosmological Model

- We know what dark matter does, but very little about what it is.
- We know what the baryons are, but their physics is much more complex to understand.
- We almost nothing about dark energy, except how much there is.
cosmic microwave background tells us how clumpy the early universe was.

we can observe the how clumpy the galaxies are today. but what about the matter distribution?
dark matter simulations: evolving the initial power spectrum into the nonlinear regime
can simulate the distribution of mass...

and can observe the distribution of light with large surveys
the era of large surveys

- Sloan Digital Sky Survey (SDSS)  
  - 8500 square degrees  
  - u, g, r, i, z to r = 21: 100 million galaxies  
  - spectroscopy for r < 17.7: 800,000 galaxies

- Dark Energy Survey (DES)  
  - 5000 square degrees  
  - g, r, i, z photometry to i = 25  
  - dark energy from 4 methods: supernovae, cluster abundance, weak lensing & galaxy clustering
  - 4000 sq. degrees overlap with the South Pole Telescope S-Z survey

- LSST, SNAP, ...
  - 201?

- dark energy can be probed by measuring the growth of structure in the universe and the evolution of the volume element

- require an understanding of the relationship between dark matter and galaxies
how does mass connect to light?

- **the big questions:**
  - how can measurements of the visible universe (galaxies and galaxy clusters) teach us about the dark part of it?
  - how can constraints on the relationship between dark matter and galaxies contribute to our understanding of the physical processes that govern galaxy formation?

- **talk outline**
  - galaxy clustering, the halo occupation distribution, and the halo model
  - connecting galaxies to dark matter halos/subhalos & their evolution
  - galaxy clusters as probes of cosmology & halo occupation: new constraints from the SDSS maxBCG catalog
Why not just predict the mass-light connection ab initio?

Hydrodynamic Simulations of Dark Matter + Gas

lots of recent progress, but very computationally expensive, can’t resolve much of the important physics

Semi-Analytic Models

Parameterize the physics, with prescriptions for:
gas distribution, gas cooling, star formation, feedback, dust
galaxy mergers + more
in the framework of merging DM halos

lots of free parameters, may miss some of the important physics
galaxies are not randomly distributed in space... they are clustered

galaxy-galaxy
two-point correlation function

animation of SDSS DR2 galaxies by M. Blanton

SDSS data
z=0.1

physical separation

\[ \xi(r) = \left( \frac{r}{5.59 \ h^{-1}\text{Mpc}} \right)^{-1.84} \]
galaxy clustering is a function of luminosity, color & type: galaxies must have a non-trivial relation to the matter distribution

\[ b = \sqrt{\xi_{gg}/\xi_{mm}} \]

Zehavi et al 2004
evolution and scale dependence of clustering

large-scale galaxy clustering is not a strong function of redshift

clustering is more strongly scale dependent at high redshift
dark matter clustering looks somewhat different

evolves rapidly with redshift

2PCF not a power law

evolution is a strong function of matter density and dark energy

Colin et al 1999
galaxies are biased tracers of the dark matter distribution

\[ b = \sqrt{\frac{\xi_{gg}}{\xi_{mm}}} \]

more snow on mountaintops

more light in deeper potential wells
dark matter halos are also biased

VIRGO consortium

Nagai & Kravtsov

first sites of halo formation

Peak-Background Split
• Schematic Picture:
3
2
1
0
x
δc
Large Scale "Background"
Enhanced "Peaks"

δc

0
1
2
3
x
W. Hu
abundance

Clustering

halo mass

e.g., Sheth & Tormen 1999
Jenkins et al 2001
Warren et al 2005

Mo & White 1996
Seljak & Warren 2004

Zehavi et al 2004

brightness

galaxy abundance

$\Phi(M_{\text{vir}} - 5 \log_{10} h)$

$M_{\text{vir}} - 5 \log_{10} h = -19.40 \pm 0.43$

$P$ = 0.184 ± 0.057

$Q$ = 0.624 ± 0.30

$\alpha$ = -1.05 ± 0.01

$M_{\text{vir}}$ = 5 $\log_{10} h$

$n$ = 2.5 $\log_{10} h$

$P_{\text{vir}} = 2.5 \log_{10} h$

$\Phi(M_{\text{vir}} - 5 \log_{10} h)$

galaxy magnitude

halo abundance

$\log_{10} \phi(M_{\text{halo}})$

$M_{\text{halo}}$ = 10 $\log_{10} h$

halo mass

galaxy bias

$\frac{b}{h}$

galaxy luminosity

$\log_{10} L$
the missing piece:
how galaxies populate halos

"halo occupation distribution"
probability \( P(N|M) \) for a halo of mass \( M \) to host \( N \) galaxies

many ways to constrain this:
encodes what clustering and abundance can tell you about galaxy formation

\[
\frac{\langle N(M) \rangle}{\langle N(M) \rangle} < \frac{\langle N(M_1) \rangle}{\langle N(M_2) \rangle}
\]
dark matter substructure

massive host halo

galactic subhalo

simulation by A. Kravtsov

simulation by B. Allgood

simulation by VIRGO consortium
the missing piece: how galaxies populate halos

natural assumption: galaxies live in dark matter subhalos
halo occupation of galactic halos

\[ \alpha^2 \equiv \frac{\langle N(N-1) \rangle}{\langle N \rangle^2} \]

\[ N_{\text{sub}} \sim M \]

- a physically motivated way of characterizing non-linear bias of galaxies/subhalos
- no smoothing scale; naturally incorporates stochasticity, naturally brings about scale dependence

Kravtsov, Berlind, Wechsler, et al 2004

Average number of galactic (sub)halos

Host halo mass
The halo occupation approach: using clustering to connect galaxies and dark matter

1. Parameterize the halo occupation $P(N|M)$ for a given set of galaxies, with 3-5 parameters.

2. Assume that you know $b(M)$ and $n(M)$ for halos, plus the distribution of $dm$ and/or galaxies within halos (either directly measure in simulations, or use analytic approximations. the later is typically referred to as “The Halo Model”)

3. Assume that halo mass is the only thing that impacts clustering.


5. Need to do this as a function of luminosity: “conditional luminosity function”: use clustering + global LF to constrain LF as a function of cluster mass (e.g. van den Bosch et al 2003, Yang et al 2005, etc.)

Zehavi et al 2005

---

**Figure:**

The halo occupation approach: using clustering to connect galaxies and dark matter. Parameters are estimated for different luminosity bins, showing how the clustering properties ($w_p$) change with halo mass ($M$). The plots illustrate the dependence of clustering on halo mass and provide evidence for the validity of the Halo Model.
The subhalo counting approach: instead, go from subhalos directly to galaxies using some mass connected property

assign luminosities to subhalo circular velocities by matching $n(>v_{\text{max}})$ to $n(>L)$

Conroy, Wechsler & Kravtsov 2006;
see also Vale & Ostriker 2006, 2007
Distinct Halo Evolution

Subhalo Evolution

Accretion epoch

Wechsler et al 2002
galaxy-galaxy correlation function at z=0

SDSS, z=0
data: Zehavi et al 2004

Conroy, Wechsler & Kravtsov 2006
galaxy-galaxy correlation function at higher redshift

DEEP, $z=1$

- Data: Coil et al 2005
  - astro-ph/0512233

- $M_B - 5 \log(h) < 20.5$
- $M_B - 5 \log(h) < 20.0$

Subaru, $z=4-5$

- Data: Ouchi et al 2005

- $M_B - 5 \log(h) < 19.5$
- $M_B - 5 \log(h) < 19.0$

- The scale of typical halos

Conroy, Wechsler, & Kravtsov 2006
Galaxies have a tight correspondence to their dark matter subhalos: a simple model which relates galaxy luminosity to subhalo velocity with no free parameters matches galaxy clustering from $z=0$ to $z=5$ (including the powerlaw nature and departures from it, luminosity and redshift dependence, scale dependence).

The hierarchical clustering predicted by LCDM ($\sim30\%$ matter, $\sim70\%$ dark energy) has been directly observed. Gravity and the dynamical evolution of dark matter halos and subhalos are the primary drivers of galaxy clustering.

we have used this model to understand:

1. galaxy-mass correlations (Tasitsiomi et al 2004)
2. close pair statistics & evolution (Berrier et al 2006)
4. the scaling of mass-to-light ratios (Tasitsiomi, RW et al in prep)

how degenerate is this model with scatter in $v$-$L$ and with cosmological parameters? (work in progress with Conroy & Kravtsov)

can we come up with a galaxy formation recipe that produces this $M$-$L$ relation? (work in progress with Somerville & Hutcheson)

what does this imply about galaxy evolution?
what does this imply about galaxy evolution?

Is the observed weak evolution in the massive end of the galaxy stellar mass function consistent with the amount of merging in LCDM?

use the m-L connection at z=1, track buildup of stars as halos merge and are disrupted

confront models with amount of evolution in the stellar mass function, luminosity of the central galaxy (BCG) as a function of halo mass, amount of light in the ICL as a function of halo mass

matches several observations only if most of merging stars go into ICL and not into the central BCG.

model is purely dissipationless: comparison between halo mass & stellar mass at different redshifts has implications for star formation.
The evolving efficiency of galaxy formation

- Match stellar MF to halo MF at various epochs
- As universe evolves, the peak conversion efficiency evolves to lower halo masses ("downsizing")
- Combine z=1 dissipationless evolution with z=0 to get implied star formation
Ways to connect galaxies to dark matter halos

- ab-initio modeling (hydrodynamical simulations & semi-analytic modeling)
- use measured clustering and abundance for a galaxy population to connect to halos
- use abundance of galaxies as a function of some mass proxy (luminosity, stellar mass) to connect to individual dark matter clumps (halos + subhalos) by mass/velocity
- try to go out and measure things that relate to dark matter halos (groups and clusters)
Cluster abundance: a probe of the dark matter halo mass function

- One way to probe the growth of structure is by measuring the growth of the most massive objects.

- Because the early clumpiness of the universe is well measured, the current clumpiness of the matter power spectrum, $\sigma_8$, constrains this growth.

- The high end of the halo mass function is particularly sensitive to this more clumpiness (higher $\sigma_8$) --> more clusters.

To do cosmology:

1. How many clusters (and how far away can they be identified)?

2. How good is the mass estimator (how well measured, how much scatter with $M$)?

3. How well can the mass estimator be understood?
the game plan:

- build a large cluster catalog from observational data
- measure cluster properties, including redshift & mass proxies
- determine the selection function of the cluster finder (purity, completeness, connection between clusters and halos)
- use a variety of mass indicators to study mass-observable relations and their scatter
- test these techniques with realistic simulations
- use counts + mass estimates to constrain cosmology
What about optical clusters?

**cosmology**

1. optical clusters are plentiful and have regular properties; can get a large sample of clusters, with accurate photometric redshifts, for free from photometric surveys

2. larger scatter in the mass-observable relation than other methods, but still reasonably tight relation between optical properties and mass (number of galaxies, luminosity), quite complete and pure; projection effects are no longer severe

3. several complementary mass probes:
   1. using spectroscopic data, velocity dispersion -- richness relation & scatter
   2. by stacking clusters, measure weak lensing signal, mass--richness rltln+profiles
   3. by stacking clusters, measure X-ray lum & profiles in ROSAT data, Lx-richness

4. key: need to understand the selection function, including the full statistical mass-observable relation very well.

**halo occupation and galaxy evolution**

clusters connect directly to halos: if the selection function and mass scale can be understood, clusters provide a sensitive probe of the relationship between galaxies and halos ("the halo occupation distribution", HOD) --> constrains gf
new sample of SDSS clusters

- ~7500 square degrees of photometric data
- new cluster finder, maxBCG: uses photo data to identify BCGs and red sequence galaxies
- optimized to detect clusters from $z=0.1-0.3$ in SDSS
- ~180,000 with $\geq 2$ bright red galaxies;
  ~14,000 with $\geq 10$ bright red galaxies (now public!)


Ben Koester, Tim McKay, Gus Evrard, Eduardo Rozo, Jim Annis
Sarah Hansen, Matt Becker, Erin Sheldon, Dave Johnston, Eli Rykoff

cluster abundance as a function of richness
Cluster luminosity functions

direct test of the link between galaxies and halos

Hansen, RW, Sheldon et al 2007 in prep
Hansen, McKay, RW et al 2005

i-band absolute magnitude [mag - 5log_{10}h]

surface density of cluster galaxies [h^2 Mpc^{-2}]

clear presence of a central galaxy
key thing is to make a robust connect between cleanly predicted quantities (like the halo mass function) and the observables (e.g., the cluster abundance as a function of the number of galaxies)

need realistic simulations of galaxy populations in clusters to do this

**good**: use simulations to connecting observed properties to physical ones, e.g.:

1. \( P(N_{\text{obs}}|N) \) or \( P(N_{\text{obs}}|M) \) --- the cluster finder “selection function”, including contamination and incompleteness

2. \( P(\sigma_{\text{gal}}|\sigma_{\text{dm}}) \) or \( P(\sigma_{\text{gal}}|M) \)

3. connection between \( <M_{\text{lens}}(N_{\text{obs}})> \) and \( <M_{\text{halo}}(N_{\text{obs}})> \)

4. connect properties of galaxies in clusters to properties of galaxies in halos

**even better**: use simulations to develop and test **observational constraints** on the selection function \( P(\text{observable}|\text{theoretical}) \), e.g.:

1. develop a formalism for parameterizing the selection function and determine cosmological constraints from abundances

2. test techniques to measure the mass-richness scatter with velocity dispersions

3. test techniques to measure the amount of miscentering in the data
Realistic simulations of galaxy populations in clusters: what is required of them?

Ideally, you want a suite of dark matter simulations (with a range in cosmological parameter space) connected with a galaxy population that reproduces all relevant statistical properties of the observed universe.

They should reproduce the observed properties that are used for cluster finding, e.g.:

1. BCG and red sequence colors and luminosities and their connection to background galaxies; the joint luminosity-color-density relation

2. Cluster profiles, as a function of galaxy color

3. Galaxy velocities

For current and future photometric surveys, need to model fairly dim galaxies in large volumes:

1. For SDSS maxBCG, need at least -19.5 galaxies in a ~ 1Gpc/h^3 volume

2. Several future surveys in planning/construction phases, e.g. the Dark Energy Survey (DES), with i~24 galaxies over 5000 sq. degrees (~3 mags deeper in ~8 times larger volume), and LSST
the recipe (ADDGaLS) Wechsler et al 2007 “Adding Density Determined Galaxies to Lightcone Simulations”

1. **empirical**: populate the simulation with a known luminosity function
2. constrain the luminosity-dependent bias to match the data by tuning \( P(\delta_{dm}|Mr) \) for galaxies
3. add galaxy colors from real galaxies, based on the luminosity and local galaxy density

several tests:
1. luminosity function and evolution
2. color distribution and evolution
3. correlation function with lum & color
4. E-S0 ridgeline (cluster galaxies)
5. density profiles

now: run cluster finder on mocks, perform all measurements on mocks as they are performed on the data

measured halo occupation is in good agreement with other methods (constrained HOD & subhalo method)
the cluster selection function in simulations

\[
P(N_{\text{obs}}|m) = \sum P(N_{\text{obs}}|N_{\text{true}})P(N_{\text{true}}|m)
\]

Rozo, Wechsler et al 2007a
• The maxBCG sample is both highly pure and complete.
Constraints with a known selection function

1. we recover our input parameters
2. with perfect knowledge of the selection function, get percent level accuracy in parameter estimation including traditional systematics like projection, scatter, and incompleteness

however, we don’t yet have confidence that the normalization of the selection function is robust to changes in cosmology and HOD
Results: constraints from 14000 maxBCG clusters

\[ \sigma_8 = 0.92 \pm 0.10 \]
\[ \sigma_8 > 0.68 \pm 0.10, \ 99\% \ confidence \]

Priors:
- \( \Omega_m h^2 = 0.128 \pm 0.010 \)
- \( h = 0.73 \pm 0.05 \)
- \( \alpha = 1.0 \pm 0.15 \)

Assumes that simulations are reliable to calculate form of selection function, but only weak constraints on absolute normalization.

Includes substantially more complete systematics than any previous analysis.

Rozo, Wechsler et al 2007b (astro-ph/this week)
Current analysis uses ONLY the counts, and very weak selection function priors. Primary constraint actually comes from the shape of the richness function. We are now working on adding:

- Weak lensing data measured from clusters stacked on richness provides a calibration of the mass-richness relation (Erin Sheldon, David Johnston)
- Velocity Dispersion of galaxies with spectra stacked on richness provides a measurement of both the normalization and the scatter in the velocity dispersion- richness relation (Matthew Becker, Tim McKay)
- X-ray data: stacked ROSAT data as a function of richness + some high res data for massive clusters (Eli Rykoff)
- Cluster correlation function (Juan Estrada)

In addition, there is substantial room for improvement by finding a richness estimator that can be matched to true halo richness in a robust way with changing cosmology.
Cluster mass profiles from weak lensing (the galaxy mass correlation function)

- Light from background galaxies is distorted by the cluster potential.
- Low signal to noise for individual clusters; stack to get a signal (as in galaxy-galaxy lensing).
- Measures the cluster-mass correlation function: can get profiles and mass estimates.

Based on a sample of \(~100000\) groups & clusters

Sheldon et al 2007
Provides mass profiles, mass calibration, and mass-to-light profiles

Sheldon et al 2007, Johnston et al in prep, Sheldon et al in prep

Preliminary Mass calibration - Luminosity

\[ X = \ln(i\text{Lum}_{200}) \]
\[ Y = \ln(M_{200}) \]
\[ Y = A + B X + C X^2 \]
\[ A = 25.9 \pm 0.26 \]
\[ B = 1.76 \pm 0.16 \]
\[ C = -0.059 \pm 0.025 \]

\[ M_{200} = K r_{200}^3 \]
\[ K = 2.92326e+14 \]
preliminary indications for lower normalization?

slope of the mass-richness relation

power spectrum amplitude

value from X-ray?
advances in dark matter simulations now identify and track the merging history of the halos and substructures that host galaxies over volumes large enough to measure large-scale clustering.

several new tools to connect galaxies to dark matter halos. combining the observed number density and clustering of galaxies with simulations can provide powerful constraints on the link between galaxies and halos.

this empirically constrained link provides the bridge between galaxy formation models and cosmology.

the clustering of galaxies as a function of redshift, luminosity and scale can be understood from the dynamics of dark matter substructures; implications for evolution and star formation as well.

optically-selected clusters provide excellent statistics and can now be very complete and uncontaminated. careful understanding of the selection function + combination of mass estimators will allow powerful constraints on cosmology & HOD.

\[ \sigma_8 = 0.92 \pm 0.10 \]
modeling and understanding subhalos and the subhalo-galaxy connection:
Charlie Conroy (Princeton) Andrey Kravtsov (Chicago)
also
Iro Tasitsiomi (Princeton) Felipe Marin (Chicago), Andreas Berlind (NYU), James Bullock (Irvine), Andrew Zentner (Chicago)

constraints from SDSS clusters:
Eduardo Rozo (OSU)
Ben Koester, Sarah Hansen (Chicago)
Tim McKay, Gus Evrard, Matt Becker, Eli Rykoff (Michigan)
Jim Annis, Josh Frieman (Fermilab), Erin Sheldon (NYU), Dave Johnston (JPL)