

Phenomenology of dark matter structure formation

The halo model: Theory

Halo abundances, clustering, profiles

In practice: HOD, CLF, SHAM

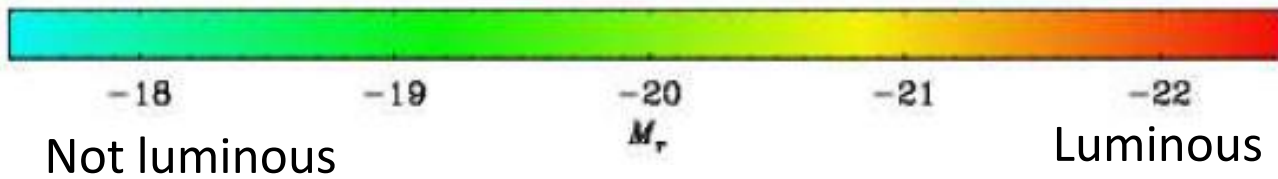
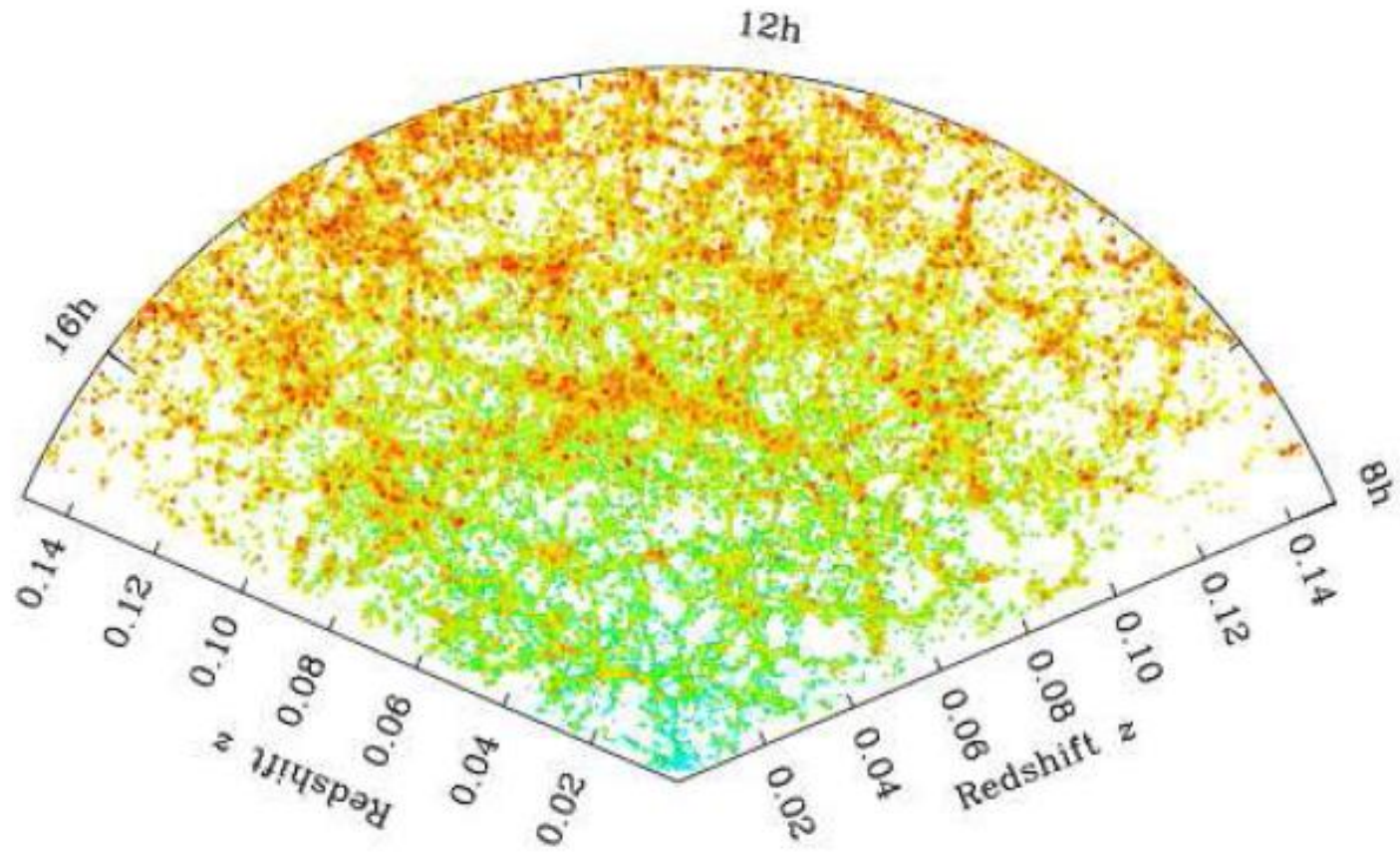
Assembly bias

Halo-
model

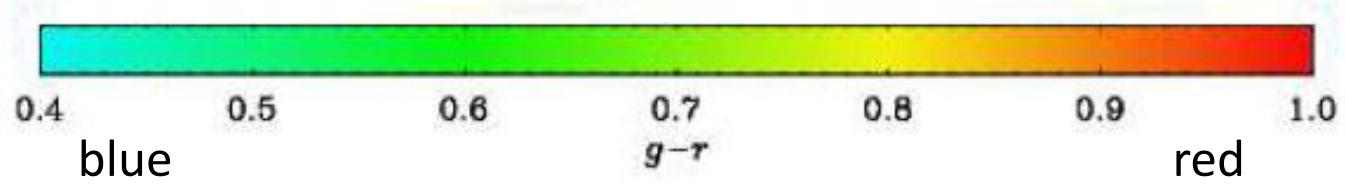
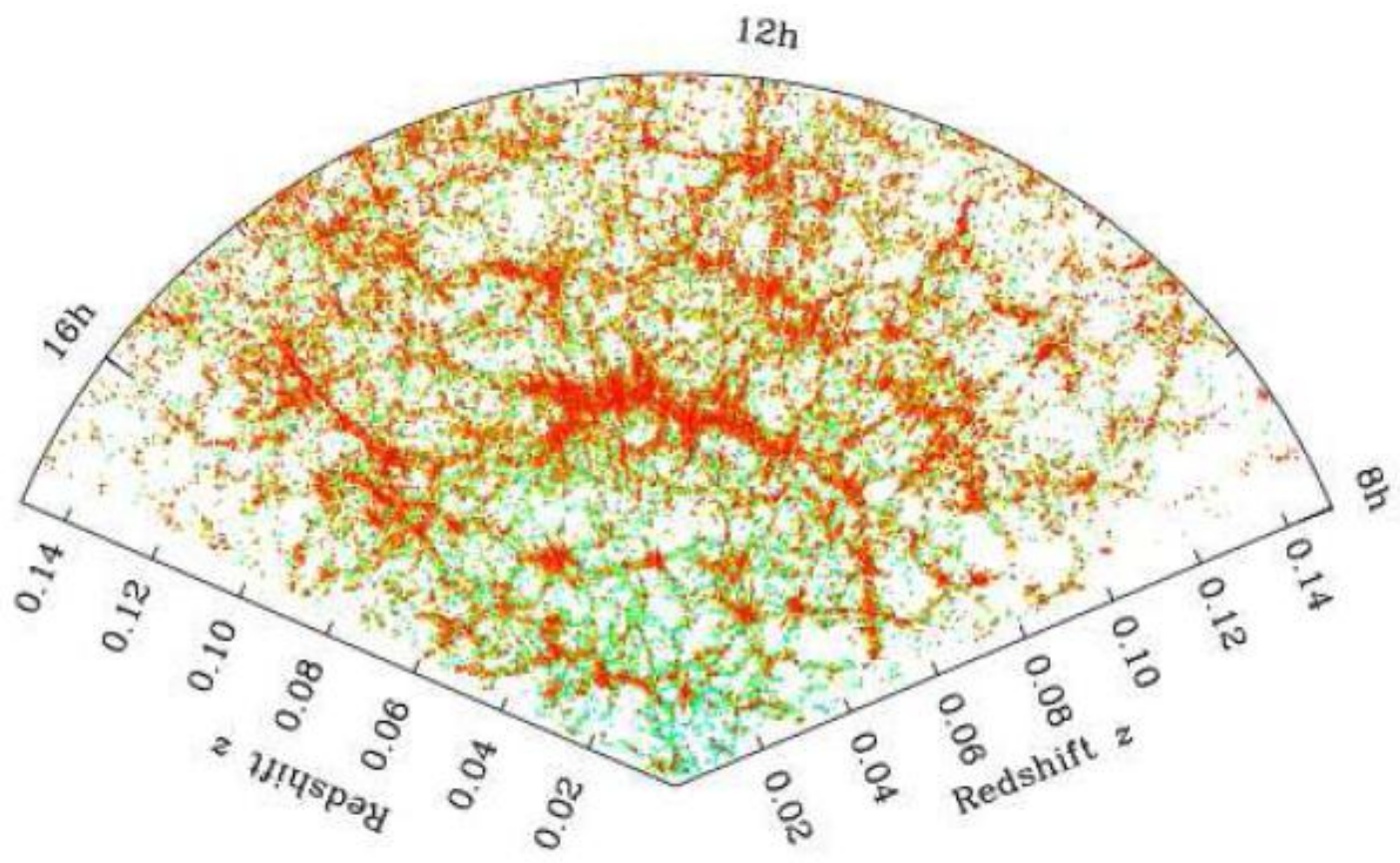
\approx

Circles in
circles

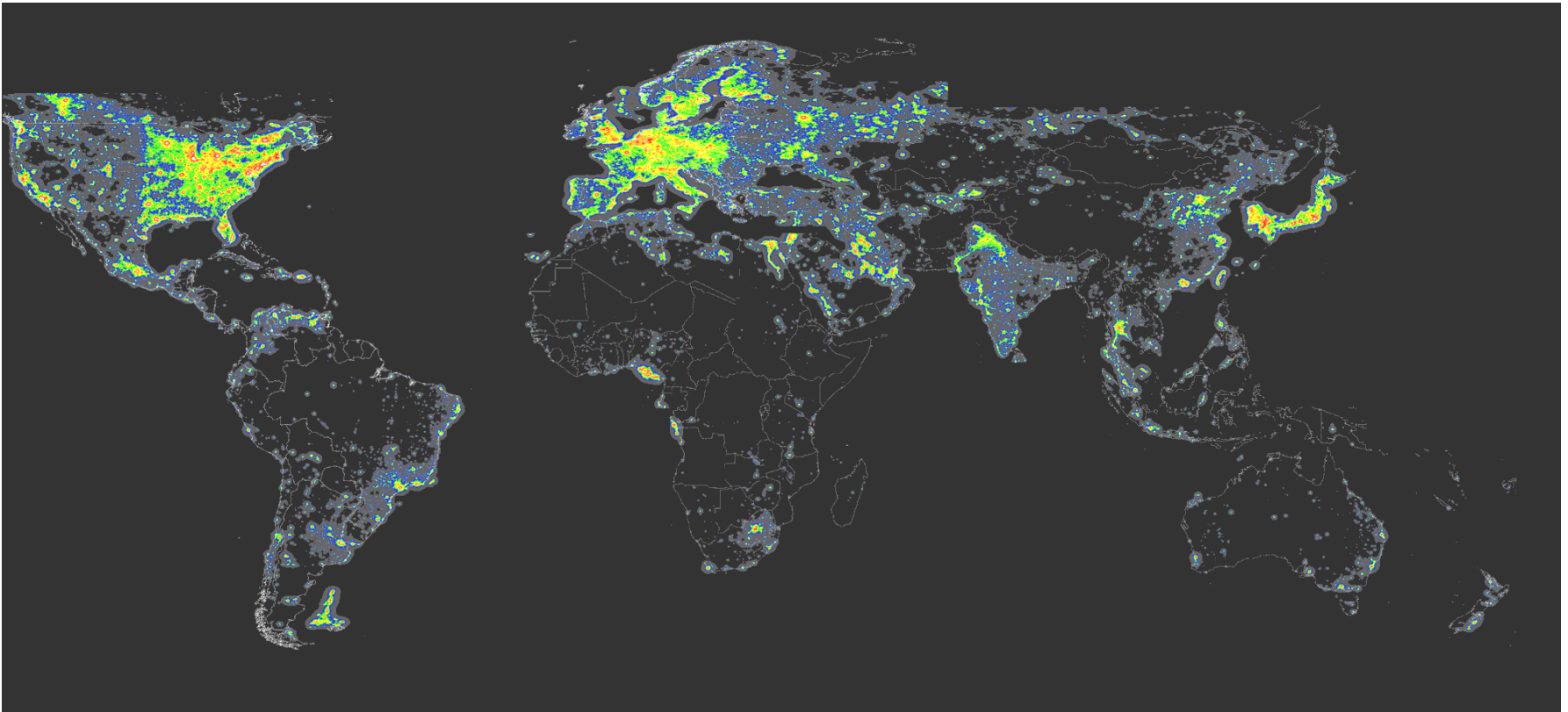




Zehavi et al. 2010 (SDSS)



Complication: Light is a biased tracer



Not all galaxies are fair tracers of dark matter;
To use galaxies as probes of underlying dark matter
distribution, must understand 'bias'

You can observe a lot
just by watching

How to describe different point processes which are all built from the same underlying density field?

THE HALO MODEL

Review in Physics Reports (Cooray & Sheth 2002)

A THEORY OF THE SPATIAL DISTRIBUTION OF GALAXIES*

J. NEYMAN AND E. L. SCOTT

Statistical Laboratory, University of California

Received February 18, 1952

ABSTRACT

A theory of the spatial distribution of galaxies is built, based on the following four main assumptions: (i) galaxies occur only in clusters; (ii) the number of galaxies varies from cluster to cluster, subject to a probabilistic law; (iii) the distribution of galaxies within a cluster is also subject to a probabilistic law; and (iv) the distribution of cluster centers in space is subject to a probabilistic law described as quasi-uniform. The main result obtained is the joint probability generating function $G_{N_1, N_2}(t_1, t_2)$ of numbers N_1 and N_2 of galaxies visible on photographs from two arbitrarily placed regions ω_1 and ω_2 , taken with fixed limiting magnitudes m_1 and m_2 , respectively. The theory ignores the possibility of light-absorbing clouds. The function $G_{N_1, N_2}(t_1, t_2)$ is expressed in terms of four functions left unspecified, which govern the details of the structure contemplated. Methods are indicated whereby approximations to these functions can be obtained and whereby the general validity of the hypotheses can be tested.

Center-satellite process requires knowledge of how

- 1) halo abundance;
 - 2) halo clustering;
 - 3) halo profiles;
 - 4) number of galaxies per halo;
- all depend on halo mass (+ ...)

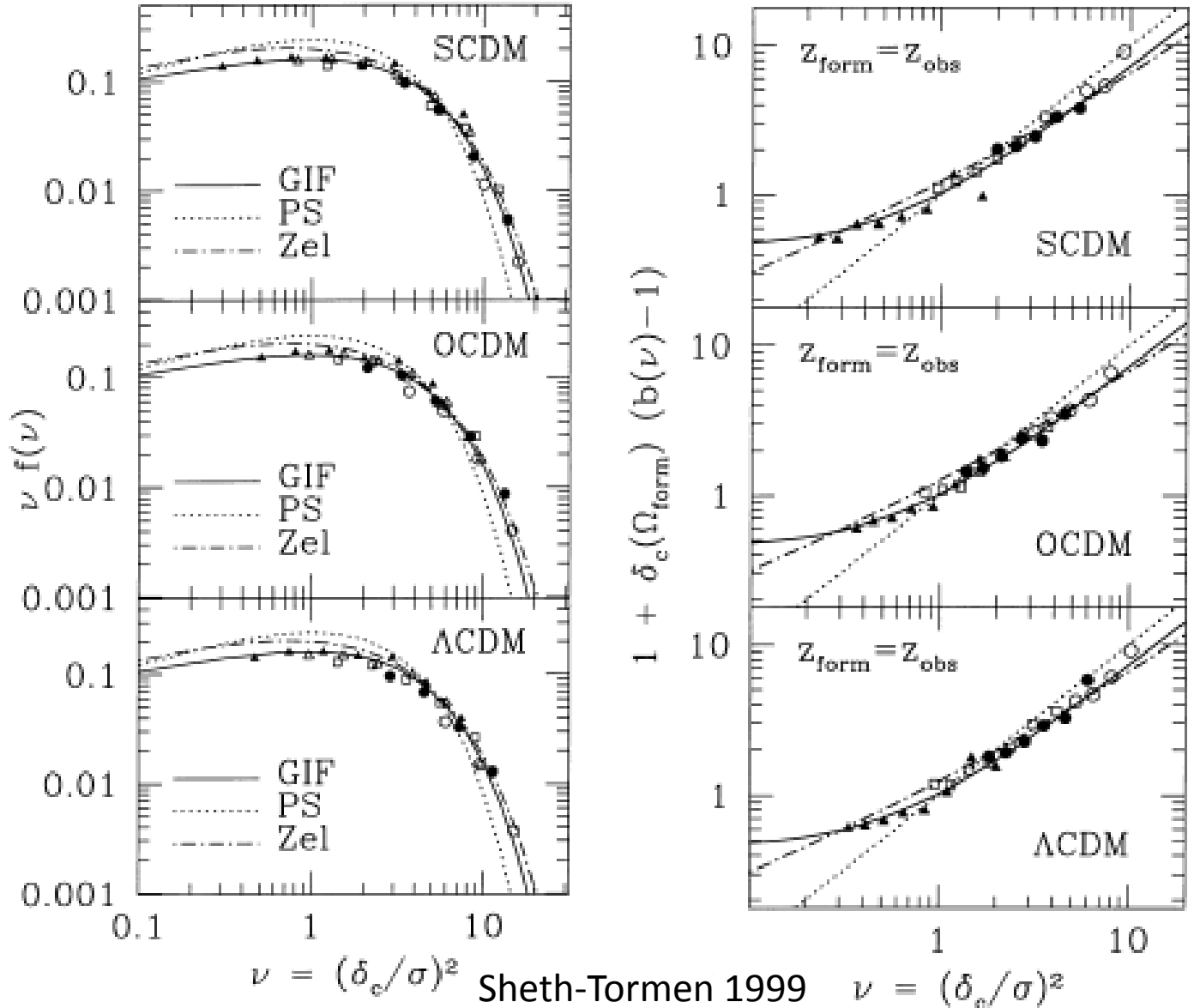
(Revived, then discarded in 1970s by Peebles, McClelland & Silk)

(Almost) universal mass function

$(m/\rho) (dn/d\ln v) = \nu f(\nu) = A [1 + (q\nu)^{-p}] \sqrt{q\nu/2\pi} \exp(-q\nu/2)$
 where all $\nu = (\delta_c/\sigma)^2$
 and A ensures integral over all ν is unity

and halo bias

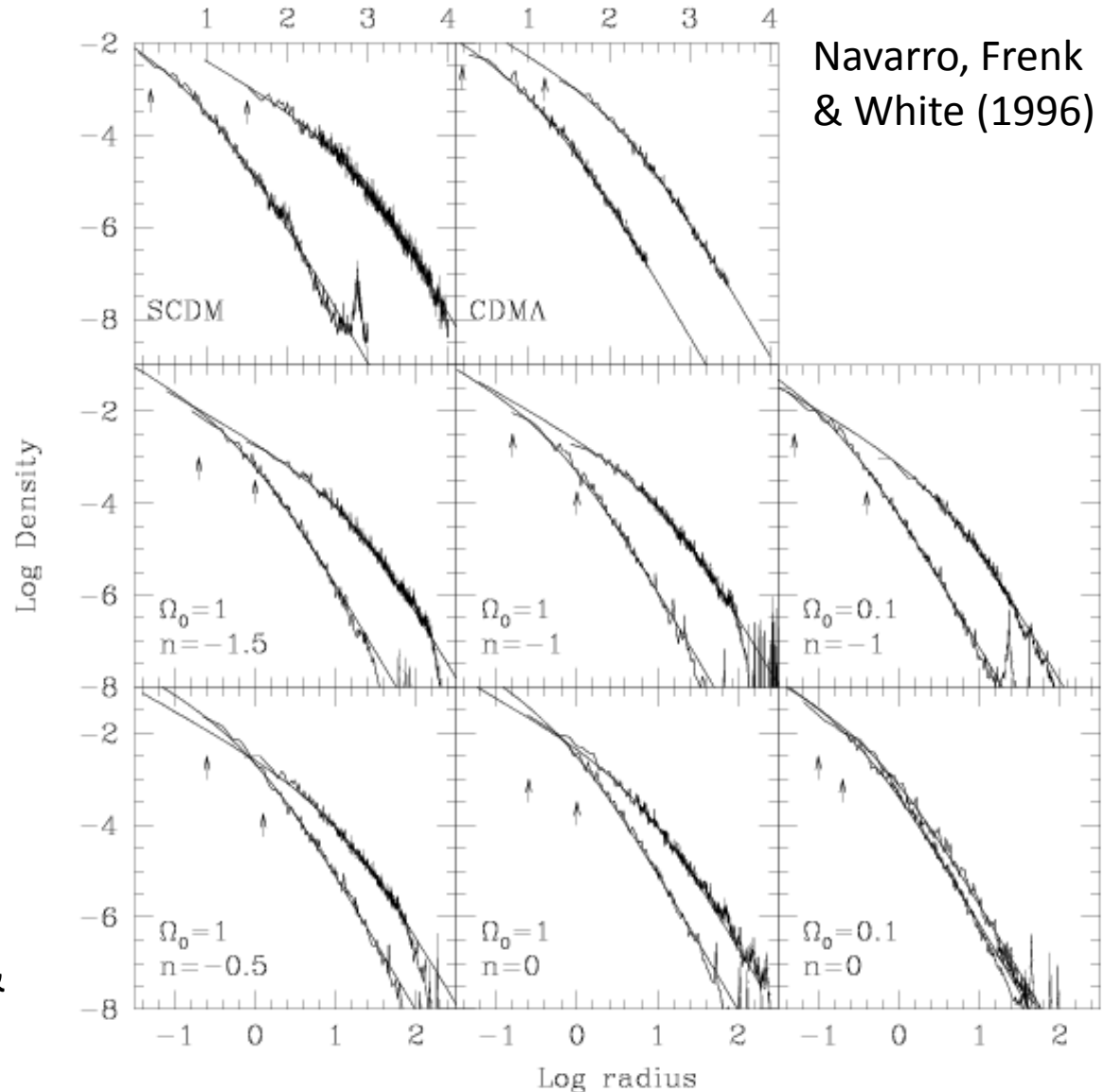
$b(\nu) = 1 + d\ln f/d\delta_c$



Universal Halo Profiles

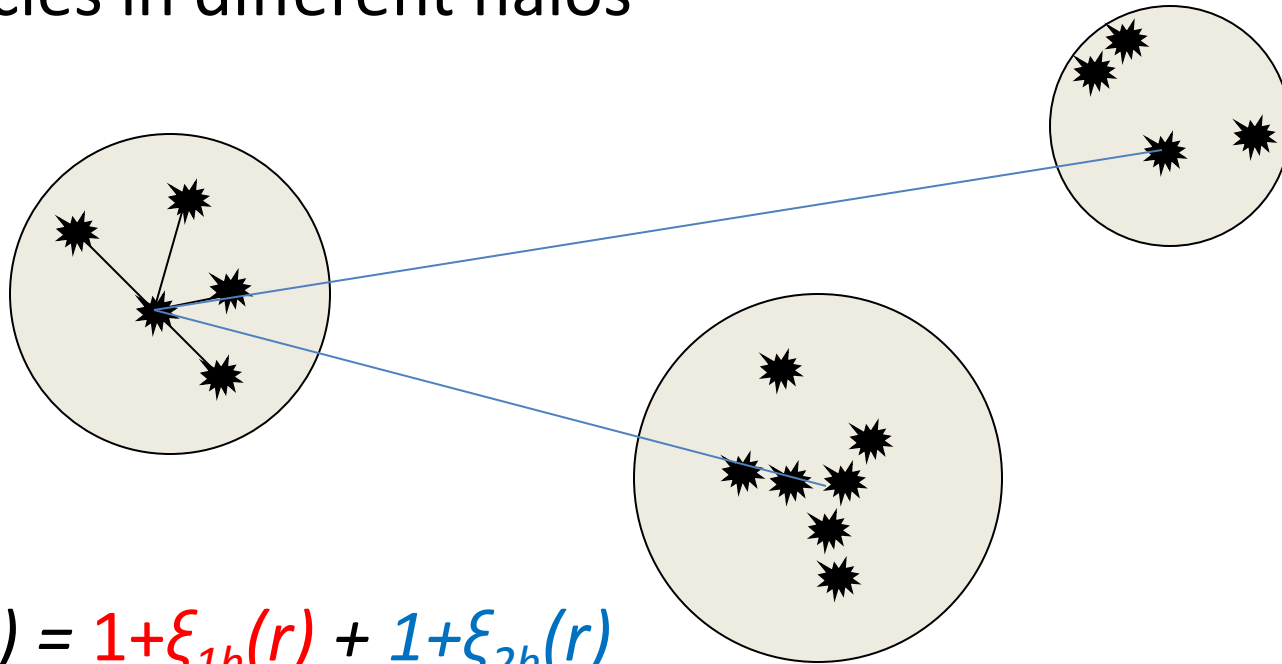
$$\rho(r) = 4\rho_s/(r/r_s)/(1+r/r_s)^2$$

- Not quite isothermal
- Depend on halo mass, formation time
- Massive halos less concentrated (partially built-in from GRF initial conditions)
- Distribution of shapes (axis-ratios) known (Jing & Suto 2001)



The halo-model of clustering

- Two types of pairs: both particles in same halo, or particles in different halos



- $1+\xi(r) = 1+\xi_{1h}(r) + 1+\xi_{2h}(r)$
- All physics can be decomposed similarly: ‘nonlinear’ effects from within halo, ‘linear’ from outside

The dark-matter correlation function

$$\xi_{dm}(r) = 1 + \xi_{1h}(r) + \xi_{2h}(r)$$

- $1 + \xi_{1h}(r) \sim \int dm n(m) m^2 \xi_{dm}(r/m) / \rho^2$
- $n(m)$: comoving number density of m -halos
- Comoving mass density: $\rho = \int dm n(m) m$
- $\xi_{dm}(r/m)$: fraction of total pairs, m^2 , in an m -halo which have separation r ; depends on (convolution of) density profile within m -halos
- This term only matters on scales smaller than the virial radius of a typical M_* halo (\sim Mpc)
 - Need not know spatial distribution of halos!

$$\xi_{dm}(r) = 1 + \xi_{1h}(r) + \xi_{2h}(r)$$

- $\xi_{2h}(r) \approx \int dm_1 \frac{m_1 n(m_1)}{\rho} \int dm_2 \frac{m_2 n(m_2)}{\rho} \xi_{2h}(r|m_1, m_2)$
- Two-halo term dominates on large scales, where peak-background split estimate of halo clustering should be accurate: $\delta_h \sim b(m) \delta_{dm}$
- $\xi_{2h}(r|m_1, m_2) \sim \langle \delta_h^2 \rangle \sim b(m_1) b(m_2) \langle \delta_{dm}^2 \rangle$
- $\xi_{2h}(r) \approx [\int dm m n(m) b(m) / \rho]^2 \xi_{dm}(r)$
- On large scales, linear theory is accurate:
 $\xi_{dm}(r) \approx \xi_{Lin}(r)$ so $\xi_{2h}(r) \approx b_{eff}^2 \xi_{Lin}(r)$

Dark matter power spectrum

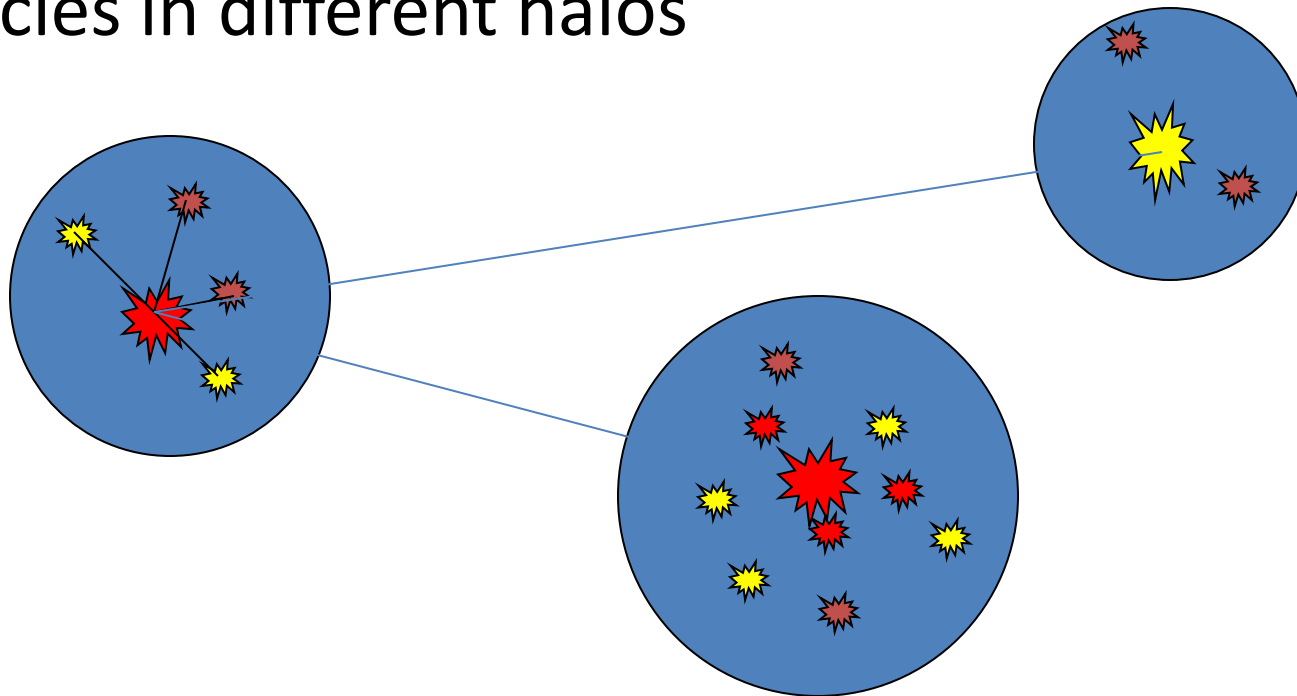
- Convolutions in real space are products in k-space, so $P(k)$ is easier than $\xi_{1h}(r)$

$$P(k) = P_{1h}(k) + P_{2h}(k)$$

- $P_{1h}(k) = \int dm n(m) m^2 |u_{dm}(k|m)|^2 / \rho^2$
- $P_{2h}(k) \approx [\int dm n(m) b(m) m u_{dm}(k|m) / \rho]^2 P_{dm}(k)$

The halo-model of galaxy clustering

- Two types of particles: central + 'satellite'
- Two types of pairs: both particles in same halo, or particles in different halos



- $1+\xi_{\text{obs}}(r) = 1+\xi_{1h}(r) + 1+\xi_{2h}(r)$
 $1+\xi_{1h}(r) = 1+\xi_{\text{cs}}(r) + 1+\xi_{\text{ss}}(r)$

The halo-model of galaxy clustering

- Write as sum of two components:
 - $1 + \xi_{1\text{gal}}(r) = \int dm n(m) g_2(m) \xi_{\text{dm}}(m|r) / \rho_{\text{gal}}^2$
 - $\xi_{2\text{gal}}(r) \approx [\int dm n(m) g_1(m) b(m) / \rho_{\text{gal}}]^2 \xi_{\text{dm}}(r)$
 - $\rho_{\text{gal}} = \int dm n(m) g_1(m)$: number density of galaxies
 - $\xi_{\text{dm}}(m|r)$: fraction of pairs in m -halos at separation r
- Think of mean number of galaxies, $g_1(m)$, as a weight applied to each dark matter halo
 - Galaxies ‘biased’ if $g_1(m)$ not proportional to m , ..., $g_n(m)$ not proportional to m^n (Jing, Mo & Boerner 1998; Benson et al. 2000; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001)
 - Central + Poisson satellites model (see later) works well
- Similarly, Y_{SZ} or T_x are just a weight applied to halos, so same formalism can model cluster clustering

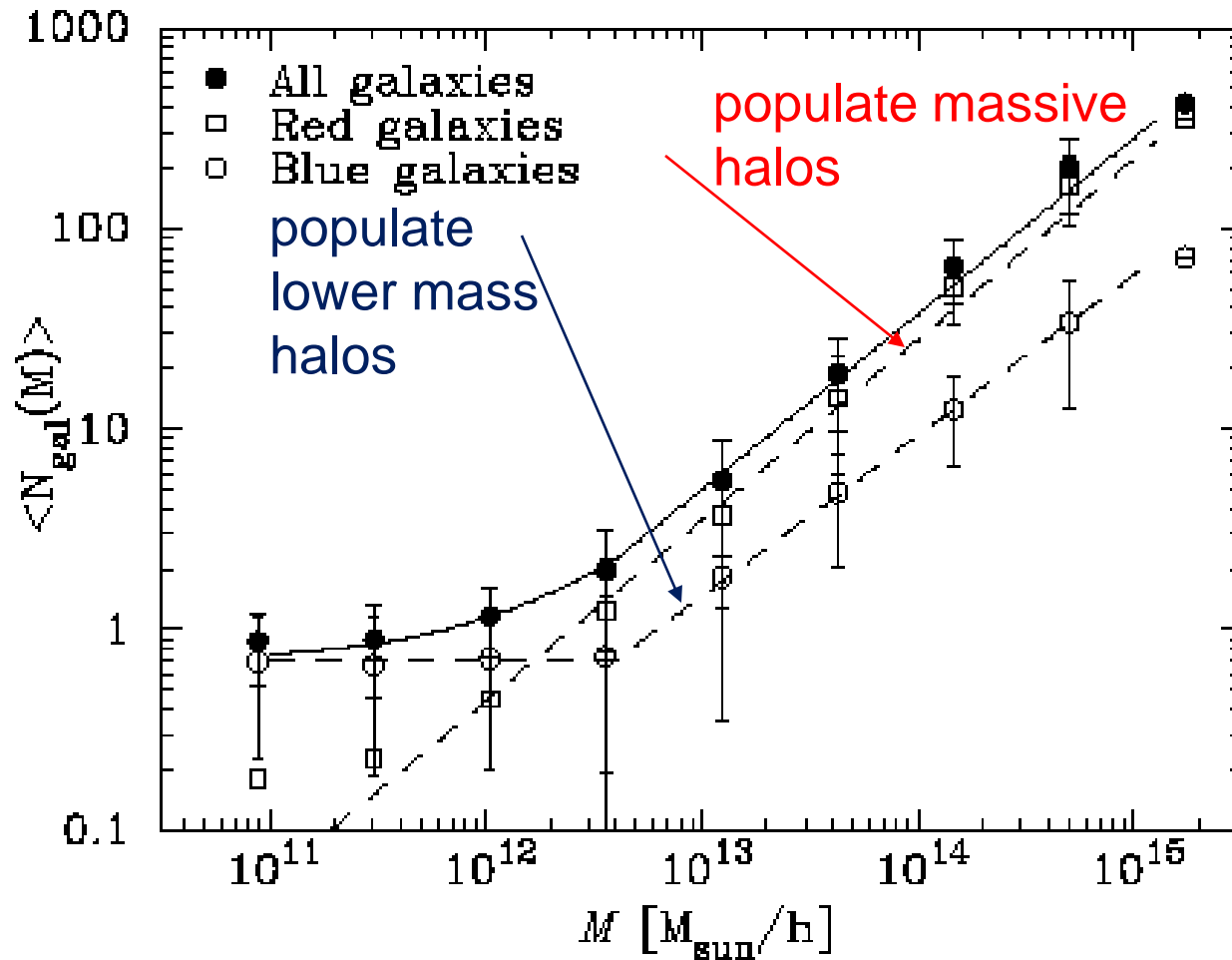
Power spectrum

- Convolutions in real space are products in k-space, so $P(k)$ is easier than $\xi(r)$:

$$P(k) = P_{1h}(k) + P_{2h}(k)$$

- $P_{1h}(k) = \int dm n(m) g_2(m) |u_{dm}(k|m)|^2 / \rho^2$
- $P_{2h}(k) \approx [\int dm n(m) b(m) g_1(m) u_{dm}(k|m) / \rho]^2 P_{dm}(k)$
- Galaxies 'biased' if $g_n(m)$ not proportional to m^n

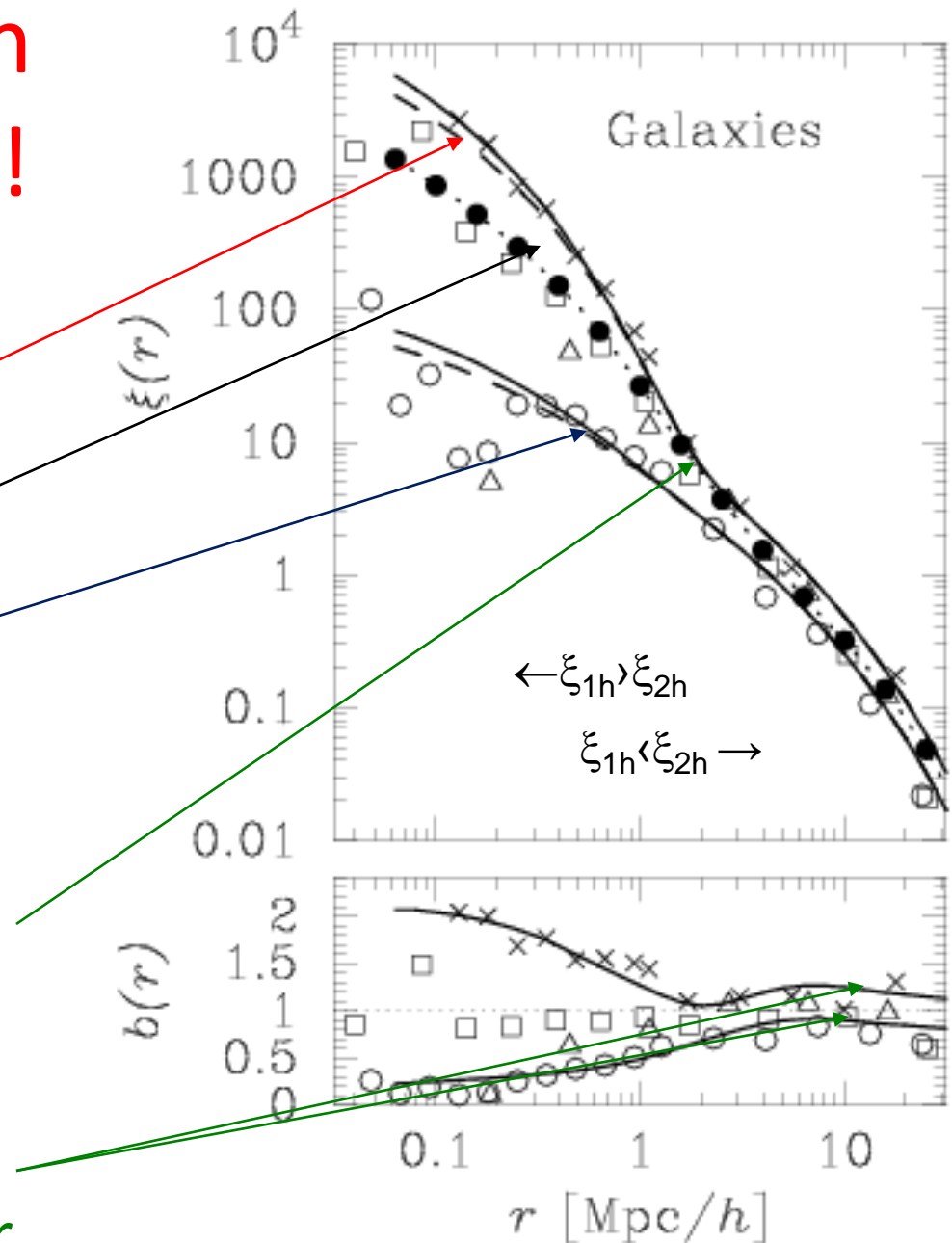
Type-dependent clustering: Why?



Spatial distribution within halos second order effect (on >100 kpc)

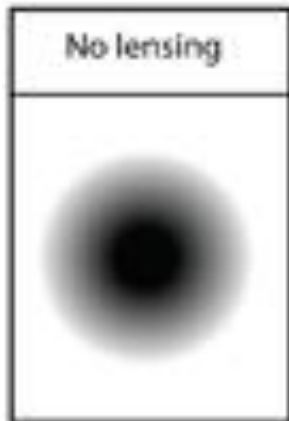
Comparison with simulations: OK!



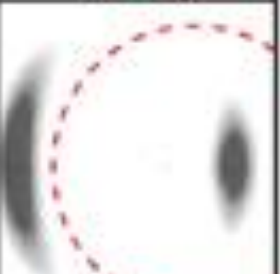
- Halo model calculation of $\xi(r)$
- Red galaxies
- Dark matter
- Blue galaxies
- Note inflection at scale of transition from 1halo term to 2-halo term (\sim virial radius)
- Bias constant at large r

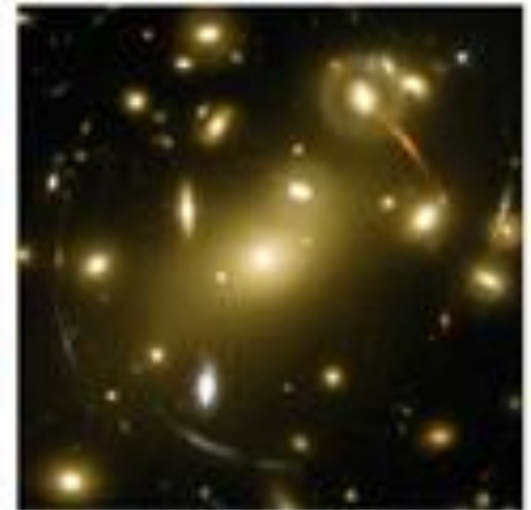


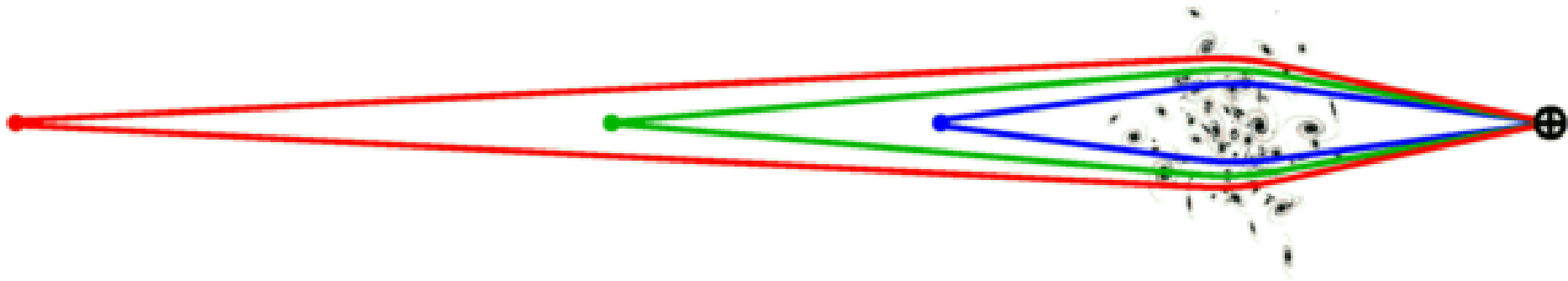
Cosmology from Gravitational Lensing

Volume as function of redshift
Growth of fluctuations with time



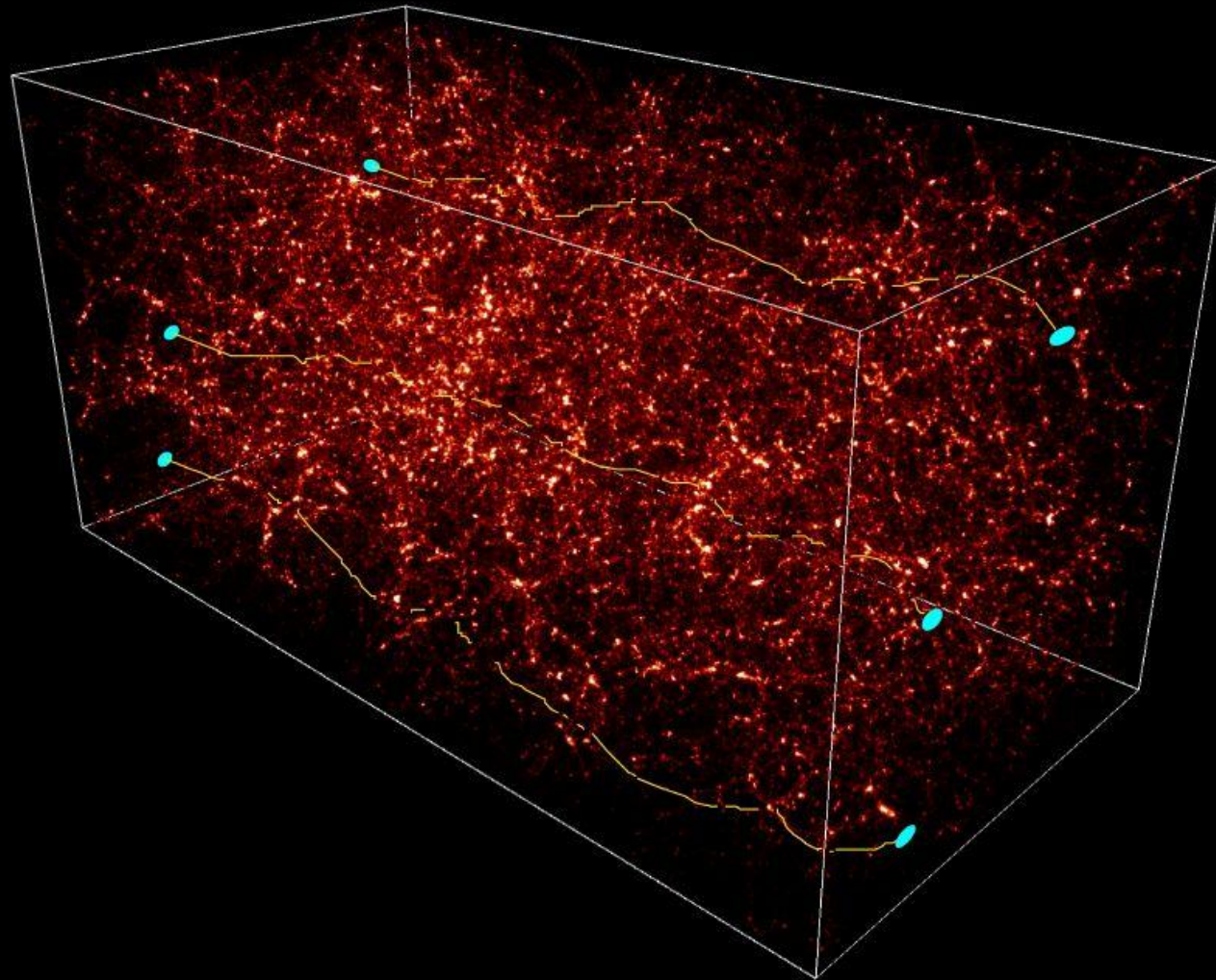
Weak lensing	Flexion	Strong lensing
		
Large-scale structure	Substructure, outskirts of halos	Cluster and galaxy cores





- Focal length strong function of cluster-centric distance; highly distorted images possible
- Strong lensing if source lies close to lens-observer axis; weaker effects if impact parameter large
- Strong lensing: Cosmology from distribution of image separations, magnification ratios, time delays; but these are rare events, so require large dataset
- Weak lensing: Cosmology from correlations (shapes or magnifications); small signal requires large dataset

DEFLECTION OF LIGHT RAYS CROSSING THE UNIVERSE, EMITTED BY DISTANT GALAXIES

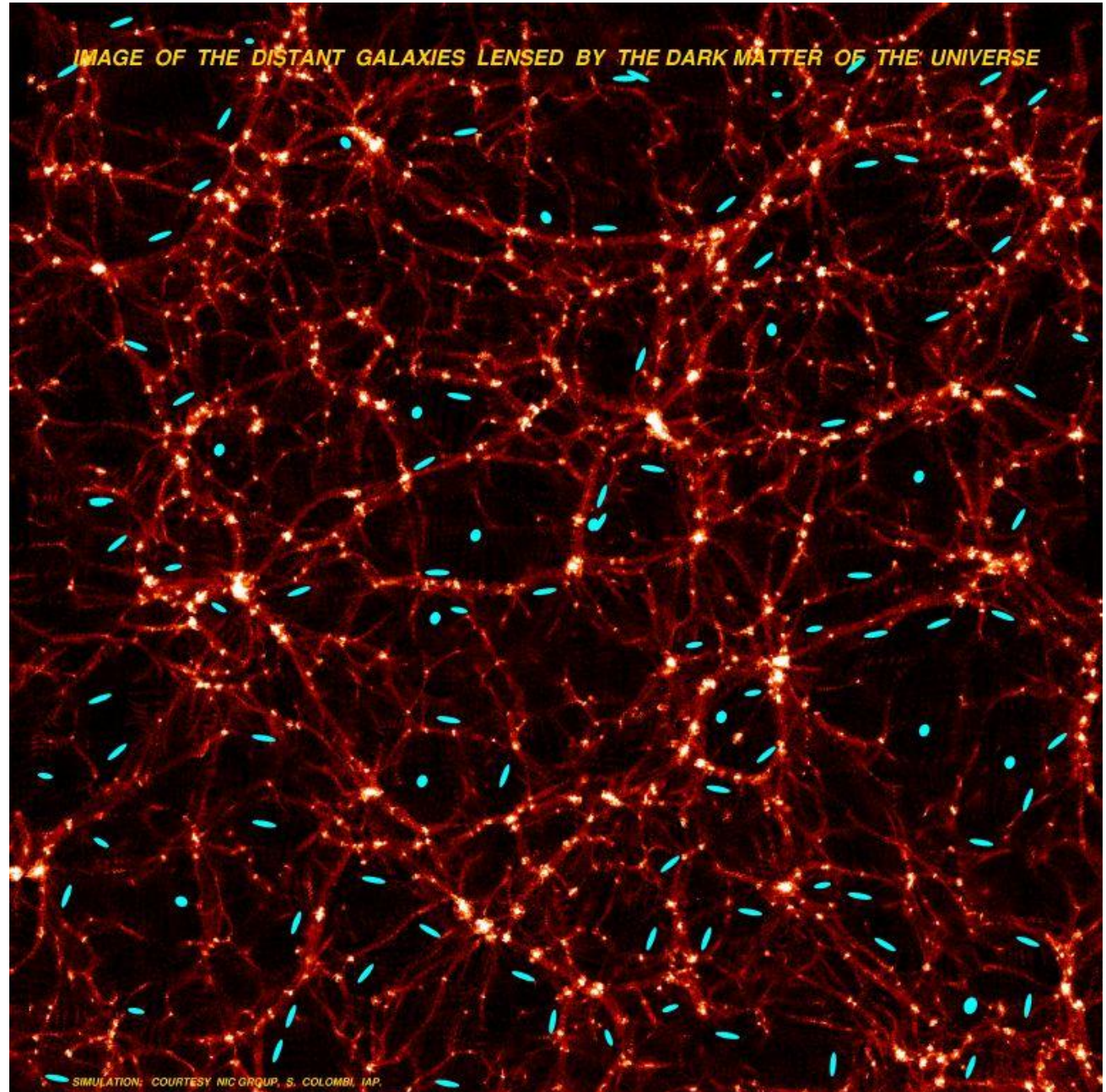


SIMULATION: COURTESY NIC GROUP, S. COLOMBI, IAP.

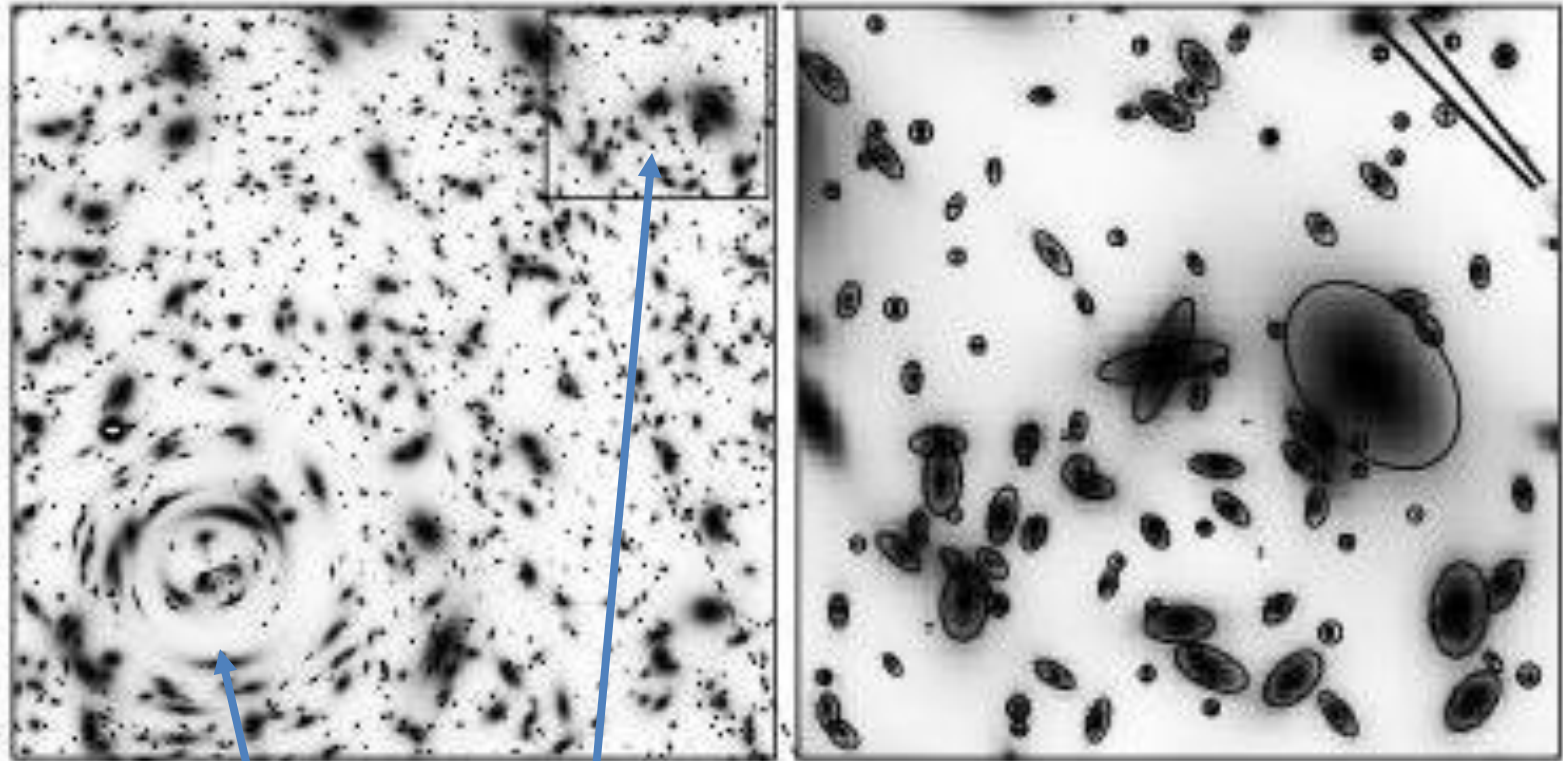
Lensing provides a measure of dark matter along line of sight

Weak lensing:
Image
distortions
correlated with
dark matter
distribution

E.g., lensed
image
ellipticities
aligned parallel
to filaments,
tangential to
knots (clusters)



The shear power of lensing



stronger

weaker

Cosmology from measurements of correlated shapes; better constraints if finer bins in source or lens positions possible



Galaxy-lensing power spectrum

$$P(k) = P_{1h}(k) + P_{2h}(k)$$

- $P_{1h}(k) = \int dm n(m) m u(k|m) g_1(m) u_g(k|m) / n_g \rho$
- $P_{2h}(k) \approx [\int dm n(m) b(m) m u(k|m) / \rho]$
x $[\int dm n(m) b(m) g_1(m) u_g(k|m) / n_g] P_{dm}(k)$

The other half of phase-space: Velocities

Just as statistics can be split into two regimes, so too can the physics: linear + nonlinear

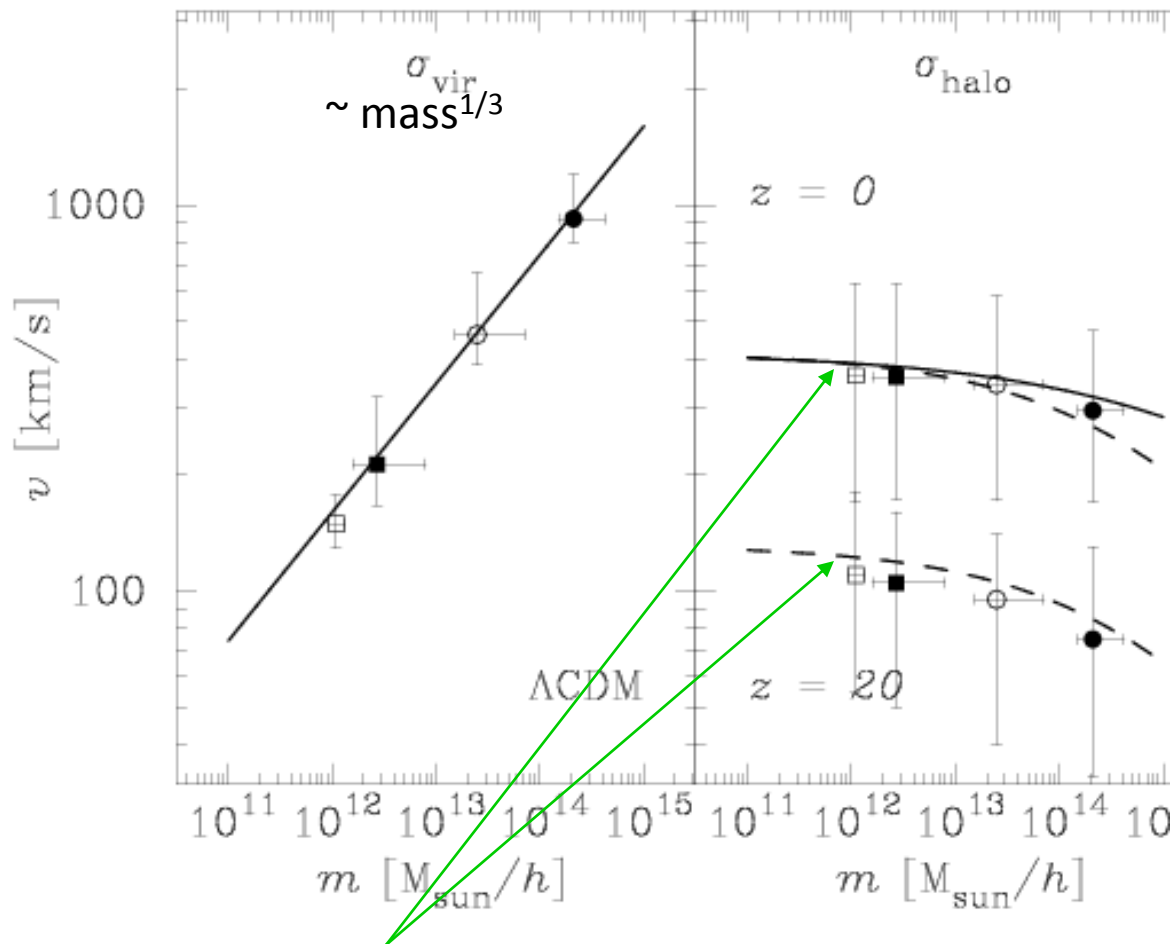
Non-Maxwellian Velocities?

- $v = v_{vir} + v_{halo}$
- Maxwellian/Gaussian velocity within halo (dispersion depends on parent halo mass, because $v^2 \sim GM/r_{vir} \sim M^{2/3}$)
+ Gaussian velocity of parent halo (from linear theory \approx independent of m)
- Hence, at fixed m , distribution of v is convolution of two Gaussians, i.e.,

$p(v/m)$ is Gaussian, with dispersion

$$\sigma_{vir}^2(m) + \sigma_{Lin}^2 = (m/m_*)^{2/3} \sigma_{vir}^2(m_*) + \sigma_{Lin}^2$$

Two contributions to velocities



- Virial motions (i.e., nonlinear theory terms) dominate for particles in massive halos
- Halo motions (linear theory) dominate for particles in low mass halos

Growth rate of halo motions \sim consistent with linear theory;
Zeldovich should be good approximation for halo motions

Exponential tails are generic

- $p(v) = \int dm m n(m) G(v|m)$

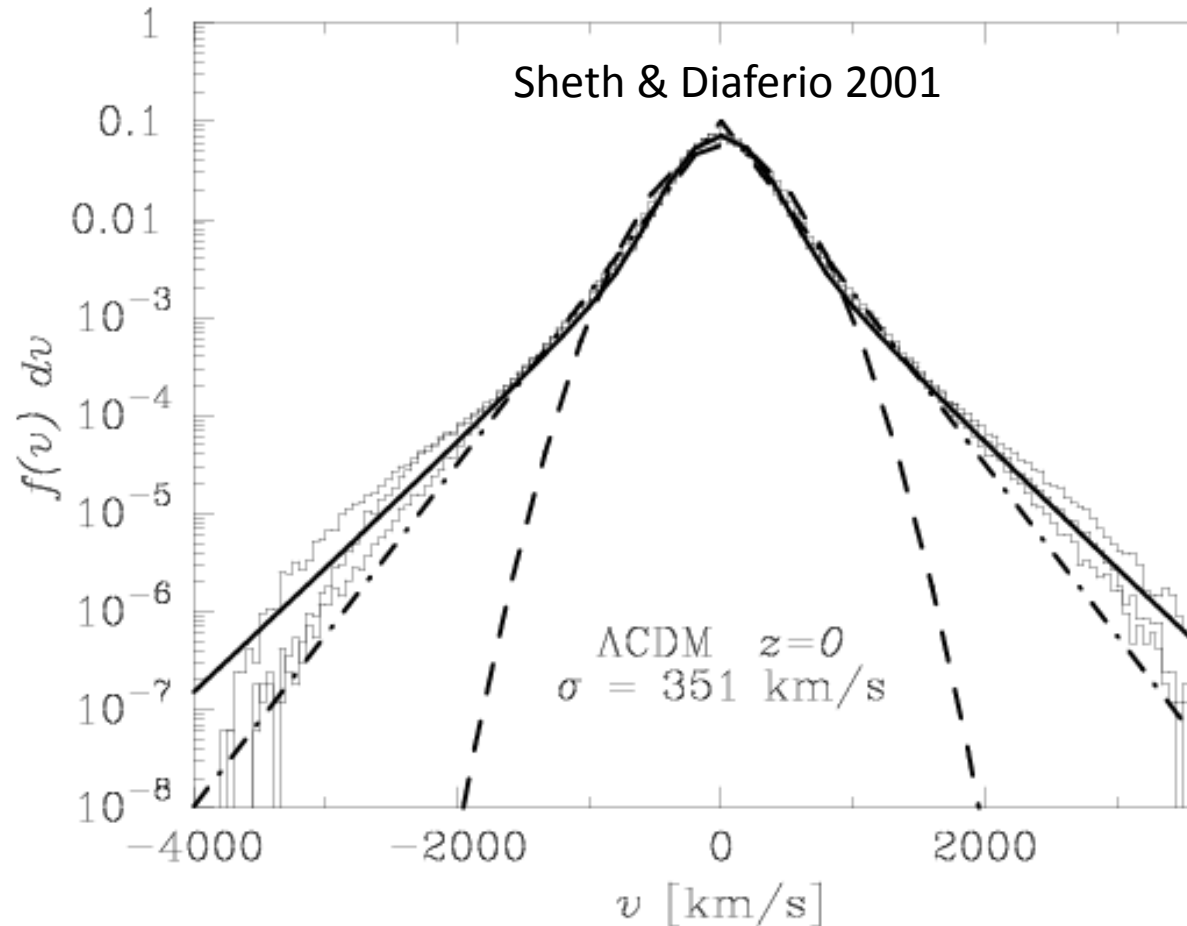
$$F(t) = \int dv e^{ivt} p(v) = \int dm n(m) m e^{-t^2 \sigma_{\text{vir}}^2(m)/2} e^{-t^2 \sigma_{\text{Lin}}^2/2}$$

- For $P(k) \sim k^{-1}$, mass function $n(m) \sim$ power-law times $\exp[-(m/m_*)^{2/3}/2]$, so integral is:

$$F(t) = e^{-t^2 \sigma_{\text{Lin}}^2/2} [1 + t^2 \sigma_{\text{vir}}^2(m_*)]^{-1/2}$$

- Fourier transform is product of Gaussian and FT of K_0 Bessel function, so $p(v)$ is convolution of $G(v)$ with $K_0(v)$
- Since $\sigma_{\text{vir}}(m_*) \sim \sigma_{\text{Lin}}$, $p(v) \sim$ Gaussian at $|v| < \sigma_{\text{Lin}}$ but exponential-like tails extend to large v

Comparison with simulations



Gaussian core with exponential tails as expected

Redshift space power spectrum

$$P_s(k) = P_{1h}(k) + P_{2h}(k)$$

$$u_s(k|m) = u(k|m) e^{-k^2 \mu^2 \sigma_{\text{vir}}^2(m)/2}$$

- $P_{1h}(k) = (1 + f\mu^2)^2 \int dm n(m) g_2(m) |u_s(k|m)|^2 / n_g^2$
- $P_{2h}(k) \approx [\int dm n(m) b(m) g_1(m) u_s(k|m) / n_g]^2 \times (1 + f\mu^2)^2 P_{\text{dm}}(k)$

Bells and whistles

(which matter for CDM \rightarrow WDM)

- Mass-concentration and scatter
 - Different profiles for red vs blue
- Distribution of halo shapes
 - Correlation of shapes with surrounding large scale structure
 - Projection effects matter for conc-m relation!
- Substructure = galaxies? Correlations with concentration/formation, time/environment
 - Correlation of substructure with large scale structure

- Handle ‘assembly bias’ easily by treating m as vector (m , conc, formation time, spin, ...)
 - See Musso et al. (2012, 2014), Dalal et al. (2008)
 - Statements that halo model cannot treat this bias are based on common but NOT essential assumption that m = halo mass only
 - Of course, now need moments of central and satellite distributions as a function of \mathbf{m} rather than just m .

Halo Model: HOD, CLF, SHAM

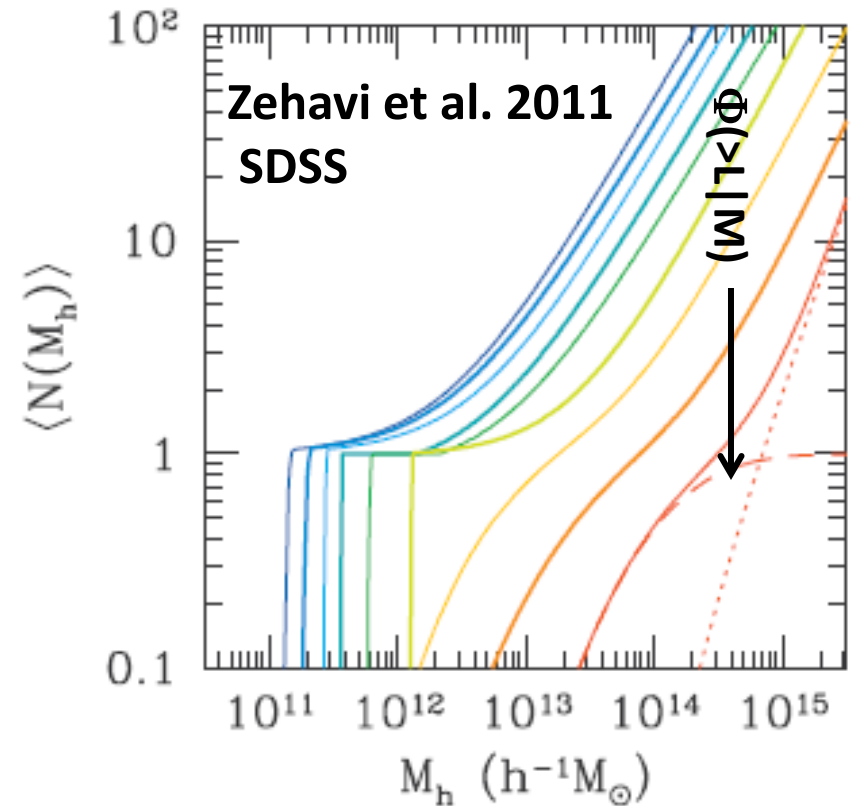
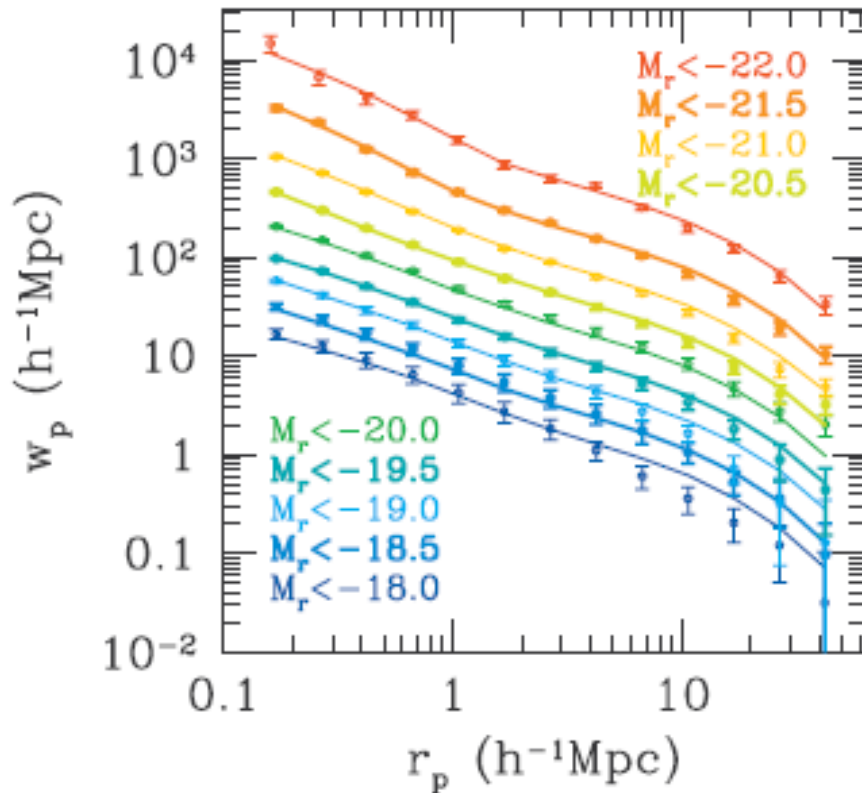
- Goal is to infer $p(N | m)$ from measurements of abundance and clustering
 - Abundance constrains $\langle N | m \rangle = g_1(m)$
 - 1-halo term of n-pt clustering constrains $g_n(m)$
- HOD uses abundance and 2pt statistics to constrain $p(N | m)$ from different samples (Zehavi et al. 2011; Skibba et al. 2014)
- CLF now does too, to constrain $\phi(L | m)$ (Lu et al. 2014)
- Since $\langle N(>L) | m \rangle = \phi(>L | m)$, HOD~CLF but with different systematics
- SHAM (Klypin+ 1999; Sheth-Jain 2003; Conroy+ 2006) uses abundance only, but gets 2pt stats quite well anyway (Moster et al. 2013)
 - Problematic for color selected samples

Halo model in practice:

Central + Poisson satellites

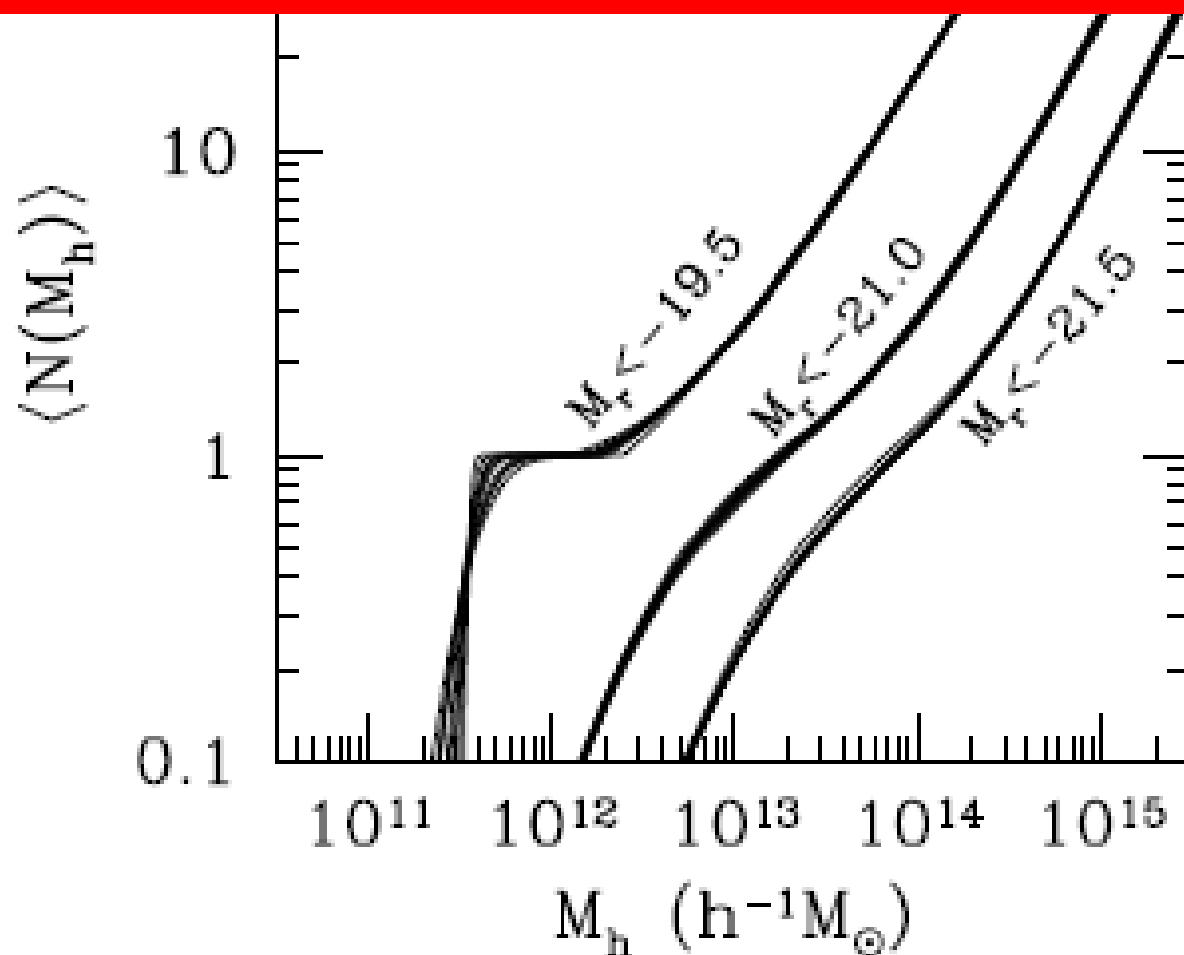
- In this model we want to place one galaxy close to (at!) the halo center, and the others with an \sim NFW profile around it. So, if we define $u_s(m|k) = u(k|m) e^{-k^2\mu^2\sigma_{\text{vir}}(m)^2/2}$ then we can write this model, with z-space distortions, as (real space is $\sigma_{\text{vir}}=0$ and $f=0$):
- $g_1(m) u(k|m)$
 - $f_{\text{cen}}(m) [1 + \langle N_{\text{sat}} | m \rangle u_s(k|m)] (1 + f\mu^2)$
 - (1 instead of u, because the central galaxy is at center, so the relevant ‘density profile’ is a delta function)
- $g_2(m) u^2(k|m)$
 - $f_{\text{cen}}(m) [2\langle N_{\text{sat}} | m \rangle u_s(k|m) + \langle N_{\text{sat}}(N_{\text{sat}} - 1) | m \rangle u_s^2(k|m)] (1 + f\mu^2)^2$
 - = $f_{\text{cen}}(m) [2\langle N_{\text{sat}} | m \rangle u_s(k|m) + \langle N_{\text{sat}} | m \rangle^2 u_s^2(k|m)] (1 + f\mu^2)^2$
 - cen-sat pairs
 - sat-sat pairs

Luminosity dependence of clustering

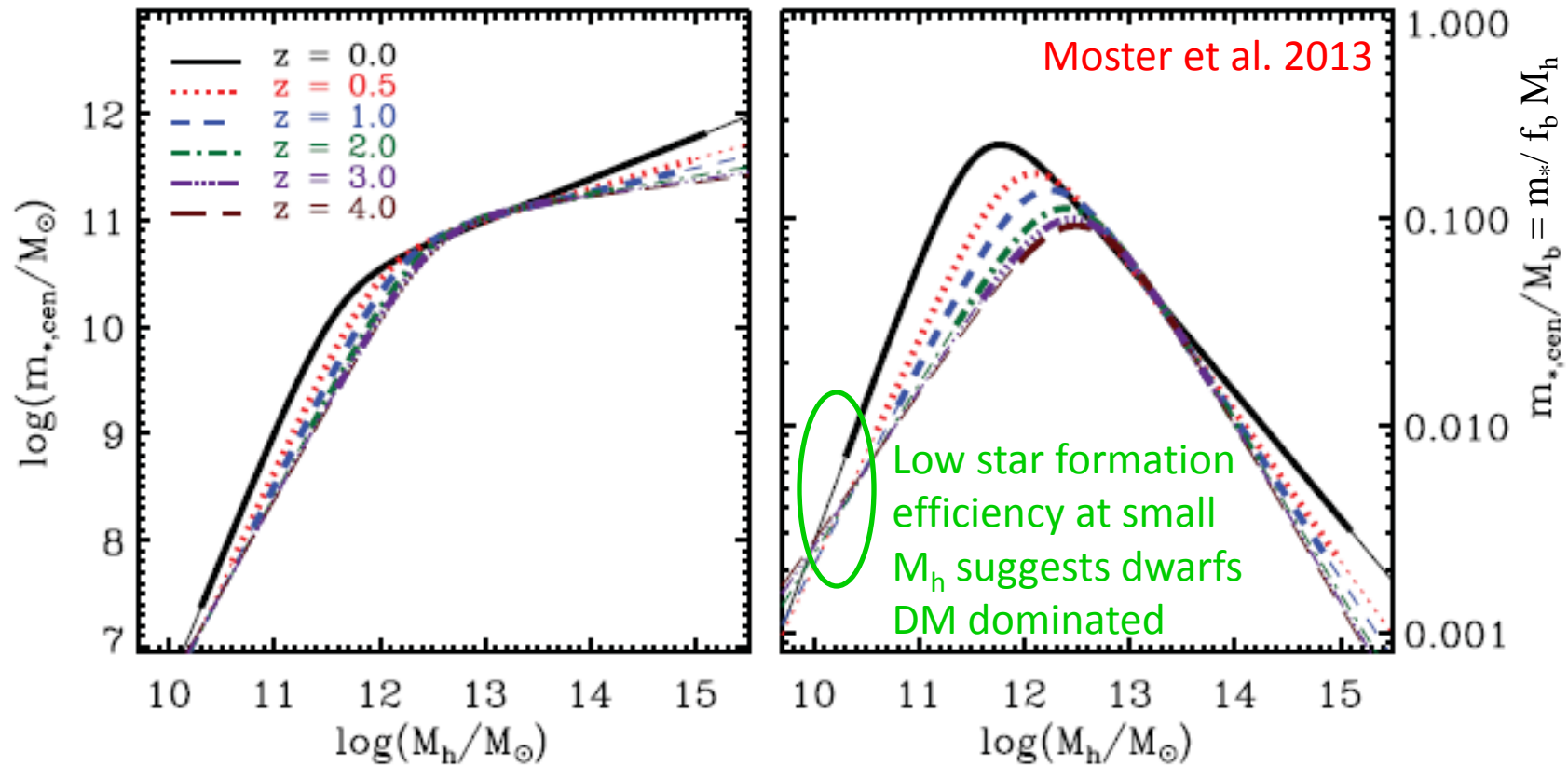


$$\langle N_{\text{gal}} | m \rangle = f_{\text{cen}}(m) [1 + \langle N_{\text{sat}} | m \rangle]$$

$$\langle N(M_h) \rangle = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\log M_h - \log M_{\min}}{\sigma_{\log M}} \right) \right] \left[1 + \left(\frac{M_h - M_0}{M'_1} \right)^\alpha \right], \quad (6)$$

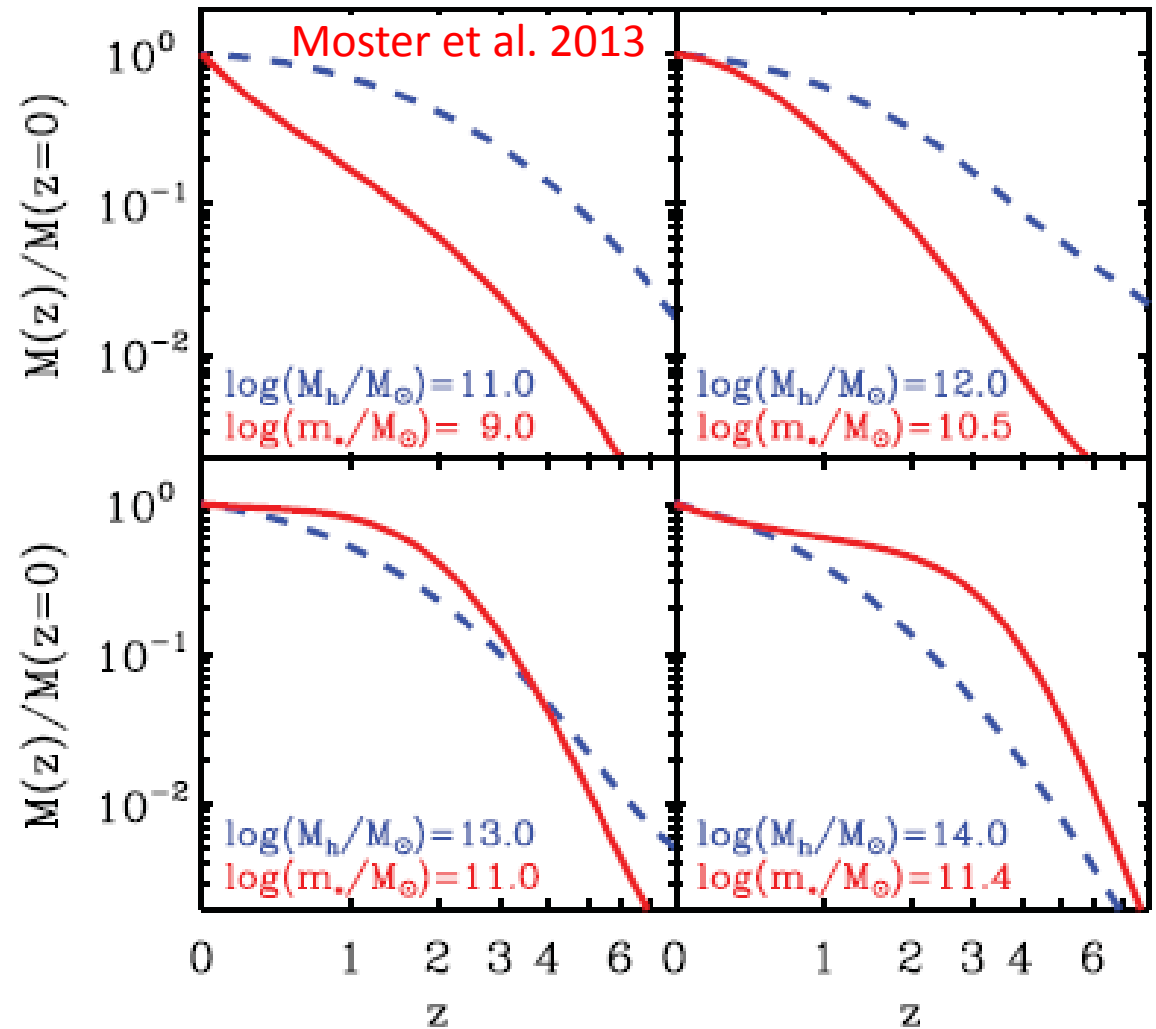


From $\phi(L|M)$ or $\phi(M^*|M)$ can determine $\langle M^*|M \rangle$; i.e. star formation efficiency as function of halo mass



Knowing $\langle M_* | M_h \rangle$ at each z yields estimates of $\text{SFR}(M_h, z)$ for the population (i.e., not object by object)

- Knowing M_* - M_h at each z yields $M_*(z)$ given $M_*(0)$ and $M_h(0)$
- Since $M_h(z)$ also known, can compare growth in situ vs mergers
- Hence, can deduce $SFR(M_h, z)$ for the population (but not object by object)
- Clustering also predicted - OK



Assembly bias

On the environmental dependence of halo formation

Ravi K. Sheth^{1★} and Giuseppe Tormen^{2★}

¹*Department of Physics & Astronomy, University of Pittsburgh, 3941 O'Hara Street, PA 15260, USA*

²*Dipartimento di Astronomia, Vicolo dell'Osservatorio 2, 35122 Padova, Italy*

Accepted 2004 February 12. Received 2004 February 10; in original form 2003 November 11

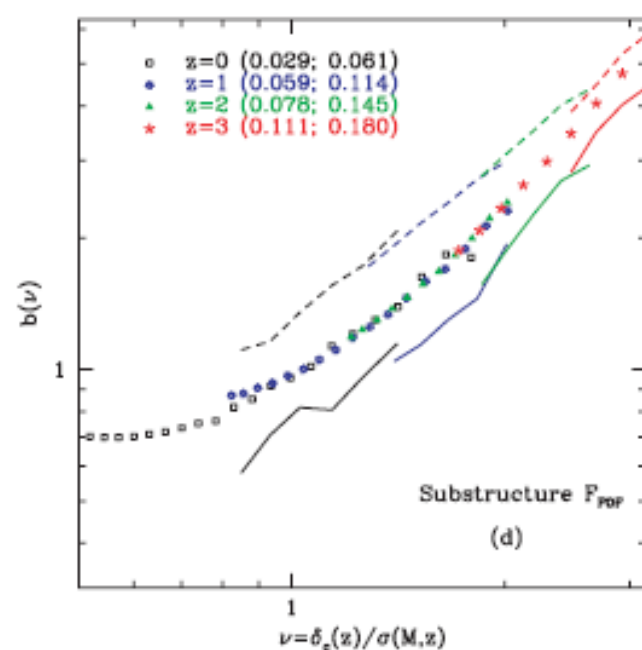
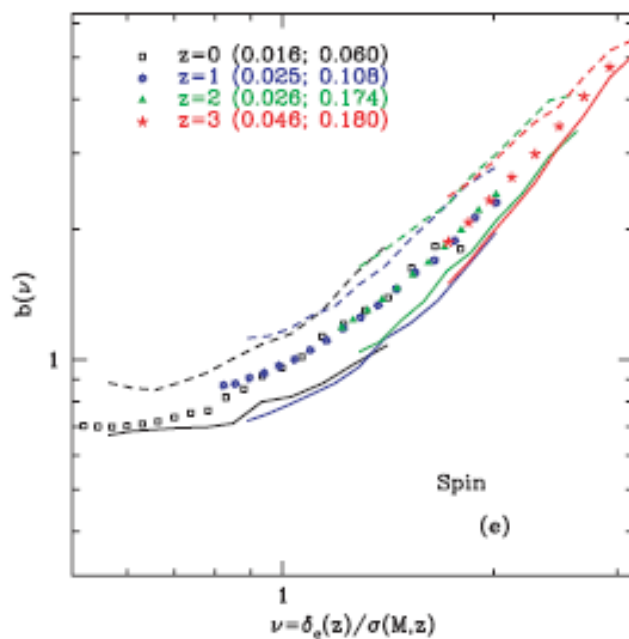
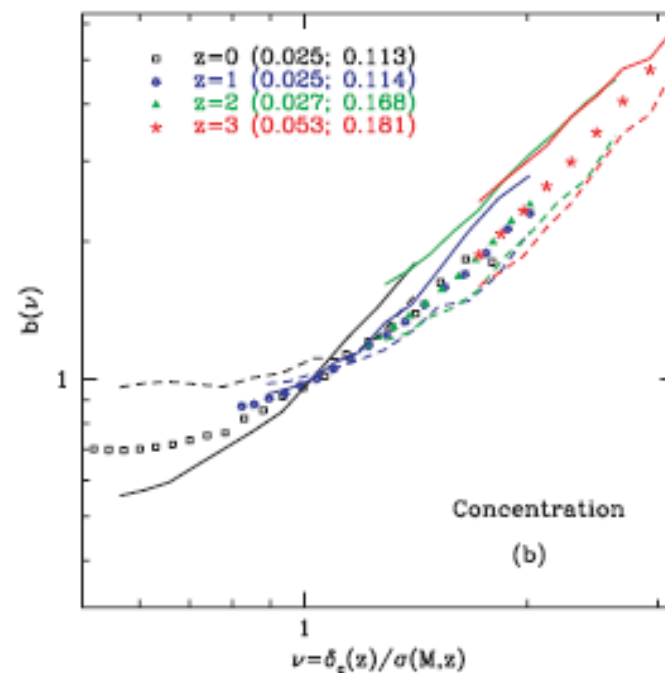
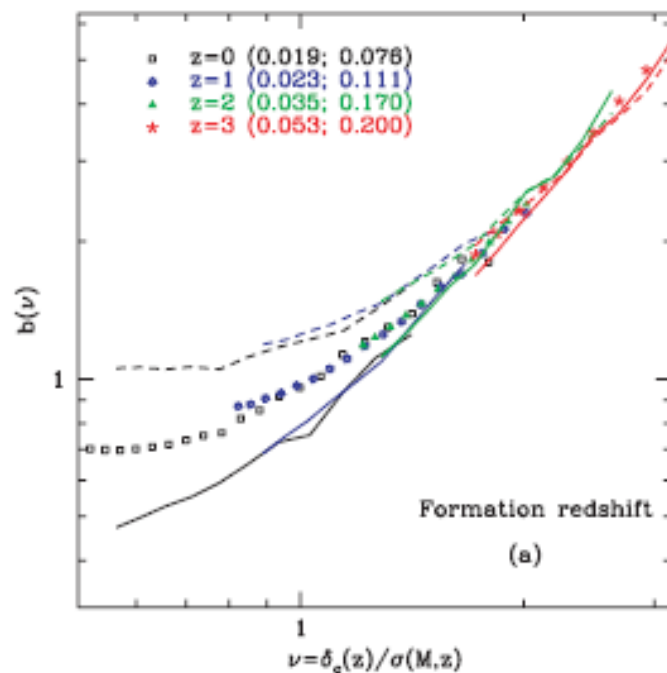
$p(N | M_h, x) \neq p(N | M_h)$
E.g. $x = z_{\text{form}},$
conc, spin,
etc.

ABSTRACT

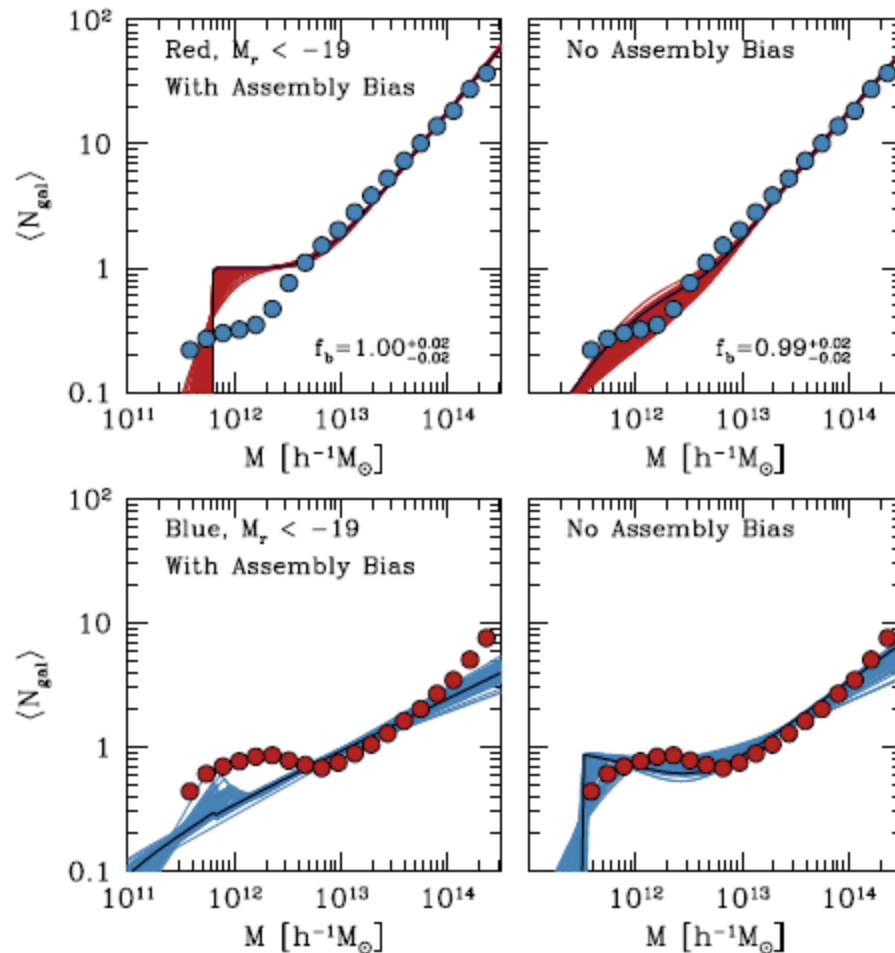
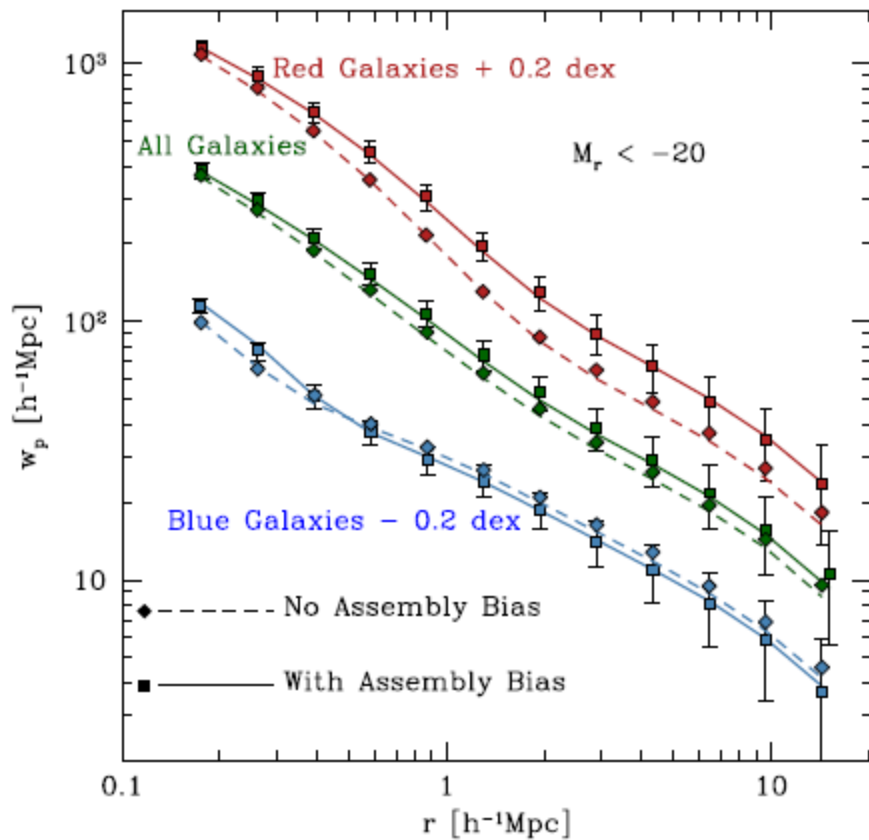
A generic prediction of hierarchical gravitational clustering models is that the distribution of halo formation times should depend relatively strongly on halo mass, massive haloes forming more recently, and depend only weakly, if at all, on the large-scale environment of the haloes. We present a novel test of this assumption, which uses the statistics of weighted or ‘marked’ correlations, which prove to be particularly well-suited to detecting and quantifying weak correlations with environment. We find that close pairs of haloes form at slightly higher redshifts than more widely separated halo pairs, suggesting that haloes in dense regions form at slightly earlier times than haloes of the same mass in less dense regions. The environmental trends we find are useful for models that relate the properties of galaxies to the formation histories of the haloes that surround them.

Key words: galaxies: clusters: general – cosmology: theory – dark matter.

Because M_{sat}^* depends on more than M_h , MEAM will exhibit assembly bias
Would be interesting to correlate MEAM scatter with $x = z_{\text{form}}, c$

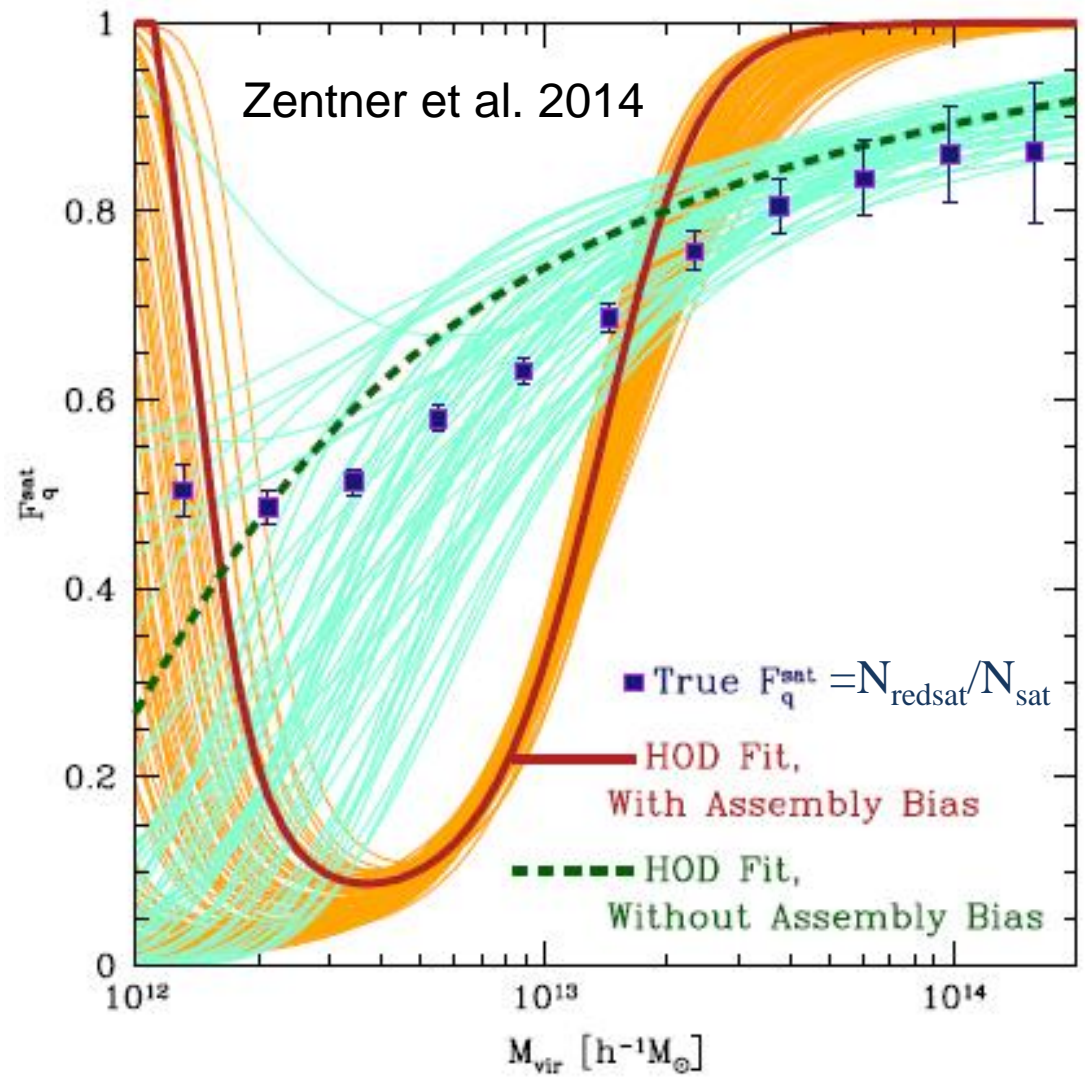


Inferred HOD may be biased



Erase assembly bias by scrambling centrals and satellites at fixed M_h ; this insures $p(N|M_h)$ same

This can lead to incorrect conclusions about galaxy formation and cosmology



- In early days Halo Model was touted by some as being the end of SAMs; SAMs argued Assembly bias was end of Halo Model
- Increased complexity means SHAM, MEAN not far from SAM

- In early days Halo Model was touted by some as being the end of SAMs; SAMs argued Assembly bias was end of Halo Model
- Increased complexity means SHAM, MEAN not far from SAM (though still simpler)

You should always go to other people's funerals; otherwise they won't go to yours.

Halo Model based approaches attractive because they interpret observations in language which is easy to relate to simulations, semi-analytic models

Increased complexity is blurring difference between SHAMs and SAMs

Observational and Assembly biases matter!

Halo Model based approaches attractive because they interpret observations in language which is easy to relate to simulations, semi-analytic models

Increased complexity is blurring difference between SHAMs and SAMs

Observational and Assembly biases matter!

**You had better know
where you're going,
or you might not get there**

Halo Model is simplistic ...

- Nonlinear physics on small scales from virial theorem
- Linear perturbation theory on scales larger than virial radius (exploits 20 years of hard work between 1970-1990)
- Halo mass is more efficient language (than e.g., dark matter density) for describing nonlinear field

...but quite accurate!

Useful for cosmology and galaxy formation from Large Scale Structure Sky Surveys

- Baryon Acoustic Oscillations
- Cluster counts and clustering
- Weak gravitational lensing
- Redshift space distortions
- Supernovae IA
- **Your name here!**