The Astrophysical Journal, 633:791–809, 2005 November 10

THEORETICAL MODELS OF THE HALO OCCUPATION DISTRIBUTION: SEPARATING CENTRAL AND SATELLITE GALAXIES

Zheng Zheng,[1,2,3] Andreas A. Berlind,[4] David H. Weinberg,[1] Andrew J. Benson,[5] Carlton M. Baugh,[6] Shaun Cole,[6] Romeel Davé,[7] Carlos S. Frenk,[6] Neal Katz,[8] and Cedric G. Lacey[6]

Received 2004 September 8; accepted 2005 July 18

ABSTRACT

The halo occupation distribution (HOD) describes the relation between galaxies and dark matter at the level of individual dark matter halos. The properties of galaxies residing at the centers of halos differ from those of satellite galaxies because of differences in their formation histories. Using a smoothed particle hydrodynamics (SPH) simulation and a semianalytic (SA) galaxy formation model, we examine the separate contributions of central and satellite galaxies to the HOD, more specifically to the probability $P(N|M)$ that a halo of virial mass M contains N galaxies of a particular class. In agreement with earlier results for dark matter subhalos, we find that the mean occupation function $\langle N \rangle_M$ for galaxies above a baryonic mass threshold can be approximated by a step function for central galaxies plus a power law for satellites and that the distribution of satellite numbers is close to Poisson at fixed halo mass. Since the number of central galaxies is always zero or one, the width of $P(N|M)$ is narrower than a Poisson distribution at low N and approaches Poisson at high N. For galaxy samples defined by different baryonic mass thresholds, there is a nearly linear relation between the minimum halo mass $M_{\rm min}$ required to host a central galaxy and the mass M_1 at which an average halo hosts one satellite, with $M_1 \approx 14M_{\rm min}$ (SPH) or $M_1 \approx 18M_{\rm min}$ (SA). The stellar population age of central galaxies correlates with halo mass, and this correlation explains much of the age dependence of the galaxy HOD. The mean occupation number of young galaxies exhibits a local minimum at $M \sim 10M_{\rm min}$ where halos are too massive to host a young central galaxy but not massive enough to host satellites. Using the SA model, we show that the conditional galaxy mass function at fixed halo mass cannot be described by a Schechter function because central galaxies produce a "bump" at high masses. We suggest parameterizations for the HOD and the conditional luminosity function that can be used to model observed galaxy clustering. Many of our predictions are in good agreement with recent results inferred from clustering in the Sloan Digital Sky Survey.

Subject headings: cosmology: theory — galaxies: formation — galaxies: halos — large-scale structure of universe

1. INTRODUCTION

The halo occupation distribution (HOD) has emerged as a powerful framework for describing galaxy bias and modeling galaxy clustering (e.g., Ma & Fry 2000; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002). It characterizes the bias between galaxies and mass in terms of the probability distribution $P(N|M)$ that a halo of virial mass M contains N galaxies of a given type, together with relative spatial and velocity distributions of galaxies and dark matter within halos. The HOD as a function of galaxy type (defined by luminosity, color, morphology, etc.) is a fundamental prediction of galaxy formation theory (e.g., Kauffmann et al. 1997, 1999; Benson et al. 2000; White et al. 2001; Yoshikawa et al. 2001; Berlind et al. 2003, hereafter B03; Kravtsov et al. 2004, hereafter K04), which can be tested by deriving the HOD empirically from observed clustering. B03 showed good agreement between the predictions of smoothed particle hydrodynamics (SPH) simulations and semianalytic (SA) calculations for the same cosmological model, and K04 showed that the HOD of substructures in high-resolution N-body simulations has many of the same features. In this paper we extend the analysis of B03 by separately examining the contributions of central and satellite galaxies to the HOD, following the approach of K04.

The central galaxies of halos are treated distinctly in SA models of galaxy formation (e.g., White & Frenk 1991; Kauffmann et al. 1993; Cole et al. 1994; Avila-Reese et al. 1998; van Kampen et al. 1999; Somerville & Primack 1999). Central galaxies accrete most or all of a halo's cooling gas. In a merger between two halos, the most massive progenitor galaxy becomes the central galaxy of the merged halo, and other, satellite galaxies may merge with it after being dragged in by dynamical friction. In hydrodynamic simulations (e.g., Cen & Ostriker 1992; Katz et al. 1992; Evrard et al. 1994; Pearce et al. 1999; White et al. 2001; Yoshikawa et al. 2001) there is no explicitly separate treatment of central and satellite galaxies, but the central objects of halos emerge as a distinct class, more massive and usually older than other galaxies in the same halo (B03), in qualitative agreement with SA predictions. The central galaxies in SPH simulations lie close to the most strongly bound dark matter particles and move slowly relative to the halo center of mass, while the radial profiles and velocity dispersions of satellites are similar to those of dark matter. B03 show that SPH and SA predictions of $P(N|M)$ are remarkably similar provided that one chooses mass

[1] Department of Astronomy, Ohio State University, 4055 McPherson Laboratory, 140 West 18th Avenue, Columbus, OH 43210; zhengz@astronomy.ohio-state.edu, dhw@astronomy.ohio-state.edu.

[2] Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540; zhengz@ias.edu.

[3] Hubble Fellow.

[4] Center for Cosmology and Particle Physics, New York University, 4 Washington Place, New York, NY 10003; aberlind@cosmo.nyu.edu.

[5] Department of Astronomy, California Institute of Technology, MS 105-24, Pasadena, CA 91125; abenson@astro.caltech.edu.

[6] Institute for Computational Cosmology, University of Durham, South Road, Durham DH1 3LE, UK; c.m.baugh@durham.ac.uk, shaun.cole@durham.ac.uk, c.s.frenk@durham.ac.uk, cedric.lacey@durham.ac.uk.

[7] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721; rad@as.arizona.edu.

[8] Department of Physics and Astronomy, University of Massachusetts, 710 North Pleasant Street, Amherst, MA 01003; nsk@kaka.phast.umass.edu.

thresholds that yield the same galaxy number density. They further show that the two methods predict similar dependencies of $P(N|M)$ on galaxy baryonic mass and stellar population age.

K04 show that the description of $P(N|M)$ for subhalos in N-body simulations simplifies considerably if one distinguishes the contributions of central and satellite substructures. (We use the terms substructure and subhalo interchangeably.) For a subhalo sample limited by maximum circular velocity, the mean occupation function $\langle N\rangle_M \equiv \sum_{N=0}^{\infty} NP(N|M)$ is well approximated by the sum of a step function for the central substructure, $\langle N_{\rm cen}\rangle_M = \Theta(M - M_{\rm min})$, where $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$, and a power law for the satellites, $\langle N_{\rm sat}\rangle_M = (M/M_1)^{\alpha}$, with $\alpha \approx 1$. The resulting shape, with a cutoff, a plateau at low M, and a power law at high M, is similar to that found for SPH and SA model galaxies (B03). More importantly, if one assumes that the distribution of $N_{\rm sat}$ with respect to the mean $\langle N_{\rm sat}\rangle_M$ is a Poisson distribution, then the model naturally explains the transition from a sub-Poisson width at low $\langle N\rangle_M$ (where the contribution of the central galaxy dominates) to a Poisson width at high $\langle N\rangle_M$ (where the satellites dominate). This transition is a common feature in all of the SA and SPH calculations mentioned above, and the sub-Poisson fluctuations at low $\langle N\rangle_M$ play a crucial role in shaping the galaxy two-point correlation function (Benson et al. 2000). Guzik & Seljak (2002) also distinguish central and satellite galaxies in their HOD modeling of galaxy-galaxy lensing, but they focus on samples defined by bins of luminosity (rather than thresholds) and therefore adopt a different parameterization. They show that the SA models of Kauffmann et al. (1999) predict a $\langle N\rangle_M$ for a narrow luminosity bin that is well approximated by a Gaussian in $\log M$ for central galaxies and a power law for satellites.

Here we apply the K04 approach to the SPH and SA galaxy populations studied by B03. We consider samples limited by thresholds in baryonic mass, which should be similar to observational samples limited by luminosity thresholds, particularly for observations at longer wavelengths. We also examine samples divided on the basis of stellar population age, which should be comparable to observational samples divided by color or spectral type. Distinguishing central and satellite galaxies proves especially valuable in understanding the behavior of these "red" and "blue" galaxy HODs. We present simple parameterizations of the predicted HODs, which can guide efforts to infer the HOD from observed galaxy clustering. We also present results for the conditional luminosity function (CLF) of galaxies at fixed halo mass (Yang et al. 2003; van den Bosch et al. 2003a). Our predictions can be tested by analyses of large galaxy redshift surveys like the Two Degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000). Several of our qualitative predictions are in good agreement with recent analyses of these surveys, in particular with the Zehavi et al. (2005) study of the luminosity and color dependence of the SDSS galaxy correlation function, as we discuss in § 5.

2. THEORETICAL MODELS

We use the same SPH simulation and SA galaxy formation model as B03. We review these calculations briefly here and refer the reader to B03 for more details.

The SPH simulation assumes a cosmological model with $\Omega_m = 0.4$, $\Omega_\Lambda = 0.6$, $\Omega_b = 0.02\,h^{-2}$, $h \equiv H_0/(100\ {\rm km\ s^{-1}\ Mpc^{-1}}) = 0.65$, $n = 0.95$, and $\sigma_8 = 0.8$ (see descriptions by Murali et al. 2002; Davé et al. 2002; Weinberg et al. 2004). It follows the evolution of 144^3 gas and 144^3 dark matter particles in a box of $50\,h^{-1}$ Mpc on each side from $z = 49$ to 0. The softening length of the gravitational force is $7\ h^{-1}$ kpc comoving. The simulation incorporates radiative and Compton cooling and phenomenological prescriptions for star formation and supernova feedback. Galaxies are identified as gravitationally bound groups of star and cold gas particles that are associated with a common local maximum in the baryon density. Dark matter halos are identified using a friends-of-friends algorithm (Davis et al. 1985) with a linking length of 0.173 times the mean interparticle separation. Each galaxy is assigned to the halo that contains the dark matter particle closest to the galaxy center of mass. In each halo, the galaxy whose center is closest to the position of the most bound dark matter particle (defined as the halo "center") is tagged as the "central" galaxy, while others are regarded as satellite galaxies.

The SA model used by B03 is GALFORM (Cole et al. 2000). For a given halo, the model generates a "merger tree" using a Monte Carlo method, starting at $z = 0$ and branching into progenitor halos until it reaches a starting redshift. Then the model works forward in time to follow the formation and evolution of galaxies in each progenitor halo. Phenomenological prescriptions are used to model star formation and feedback, dynamical friction within halos, and mergers of galaxies. There is always a galaxy residing at the center of each halo. Before a halo experiences a major merger, cooling gas is assumed to accrete onto the disk of this galaxy and form stars. If two halos merge, the most massive galaxy is set to become the central galaxy of the merged halo, and any other galaxies become satellites. If two galaxies of comparable mass merge, then all their stars (disk+spheroid) form the spheroid of the remnant, which may regrow a new disk by subsequent gas accretion. Some adjustable parameters of this model are chosen on the basis of observed properties of the local galaxy population, most notably the galaxy luminosity function (LF). However, no parameters are adjusted to reproduce observed galaxy clustering or the SPH results. The model uses the same cosmological parameters as the SPH simulation, and it is supplied with the same halo population identified in the SPH simulation so that the HOD predictions of these two models can be compared halo by halo. To improve statistics, 10 SA realizations are conducted for each of the 70 most massive SPH halos, and 40 realizations are done for less massive halos in each mass bin of width $\Delta\log M = 0.1$.

3. HALO OCCUPATION DISTRIBUTIONS

Galaxy samples of different space densities are constructed by choosing galaxies above different baryonic mass thresholds. The baryonic mass resolution limit of the SPH simulation (corresponding to 64 SPH particles) is $5.42\times10^{10}\ M_\odot$, and the space density of simulated galaxies above this threshold is $\bar{n}_g = 0.02\,h^3\,{\rm Mpc}^{-3}$. The baryonic mass threshold that yields the same space density in the SA model is $1.45\times10^{10}\ M_\odot$, lower because of the suppressed gas cooling and enhanced supernova feedback (see B03 for further discussion). In § 3.1 we focus on this sample with $\bar{n}_g = 0.02\ h^3\ {\rm Mpc}^{-3}$, which is the largest one we can create from the SPH simulation. It should correspond roughly to an observational sample defined by a luminosity threshold $M_r = -18.6$ ($0.18L_*$), where we have used the Blanton et al. (2003) LF to find the luminosity threshold that yields the same comoving space density. Scatter in stellar mass-to-light ratios makes mass-threshold and luminosity-threshold samples different, but the substantial variations of these mass-to-light ratios across the galaxy population are largely a systematic function of luminosity (Kauffmann et al. 2003), with only moderate scatter at fixed L.

In § 3.2 we divide this sample into two equal parts on the basis of stellar population age: the median look-back time (when half the stars had formed) in the case of the SPH simulation and the

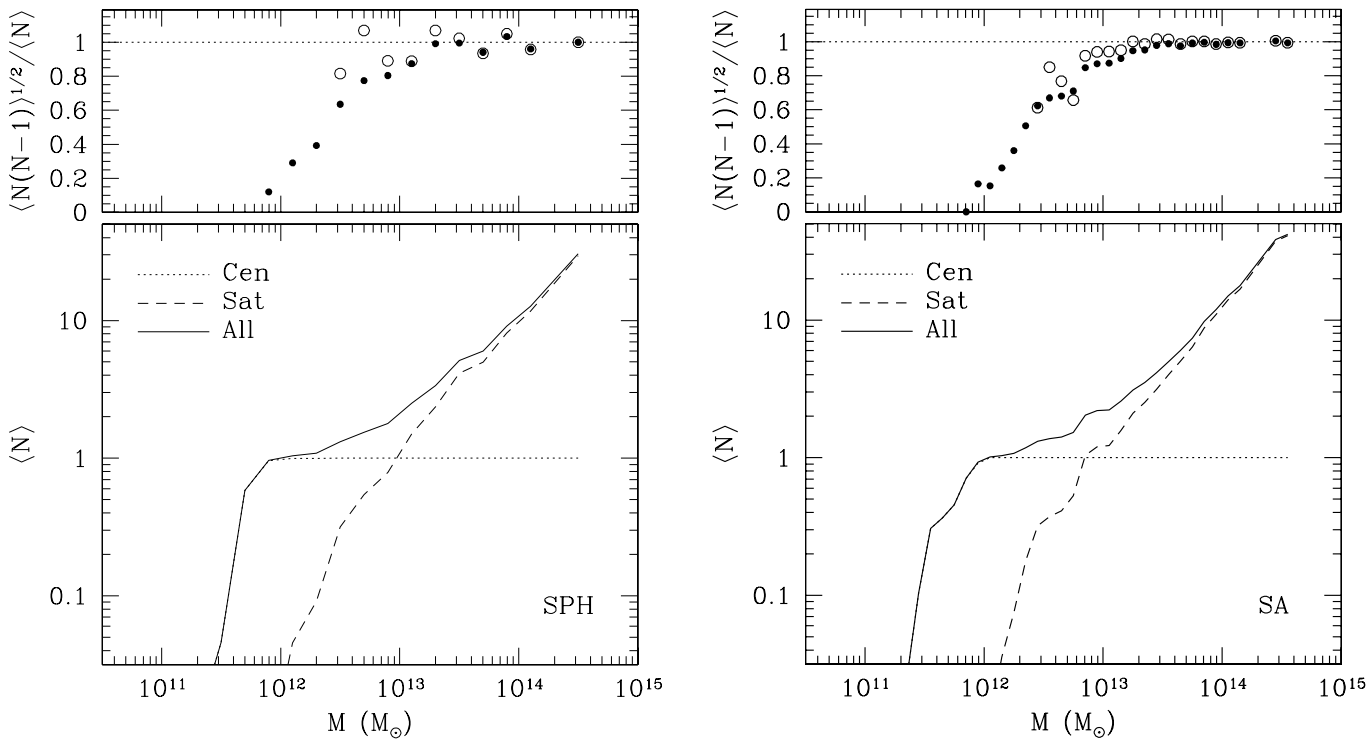

FIG. 1.—Mean occupation number and scatter as a function of halo mass, separated into central and satellite galaxies. Predictions are shown for the $\bar{n}_g = 0.02\ h^3\ \mathrm{Mpc}^{-3}$ samples from the SPH simulation (*left*) and from the SA model (*right*). The bottom panels plot the mean occupation numbers of central, satellite, and all galaxies. In the top panels, circles show $\langle N(N-1)\rangle^{1/2}/\langle N\rangle$, indicating the width of the probability distribution, for all galaxies (*filled circles*) and satellite galaxies (*open circles*). For Poisson $P(N|M)$, this ratio would be 1 (*dotted curve*). This figure can be compared to Fig. 4 of K04.

mean look-back time in the case of the SA model. The SA code automatically computes the mass-weighted mean stellar age, but it does not produce the step-by-step stellar mass record needed to compute the median stellar age a posteriori. The SPH code does produce this record, making the median age easier to compute; this quantity is probably more robust than the mean age in the SPH simulations because star formation rates are underestimated in a galaxy's early history, when it is near the mass resolution threshold. Our analysis uses only the rank order of ages, not the absolute values, so we expect the difference in age definitions to have minimal impact on our results. Population age should be a fairly good proxy for galaxy color or spectral type, so this division mimics the red/blue or early/late divisions studied by Zehavi et al. (2002, 2005) in the SDSS and Norberg et al. (2002) and Madgwick et al. (2003) in the 2dFGRS. In § 3.3 we consider samples of progressively higher baryonic mass threshold and lower mean space density, to investigate the predicted luminosity dependence of the HOD. Although we present both SPH and SA results throughout this section, the SA model has better statistics because of the multiple realizations of each halo, and it thus allows us to investigate some finer points of the HOD predictions.

3.1. *HOD for All Galaxies*

Figure 1 shows mean occupation numbers as a function of halo mass predicted by the SPH simulation and by the SA model, for the $\bar{n}_g = 0.02\ h^3\ \mathrm{Mpc}^{-3}$ samples. Mean occupation functions for central and satellite galaxies are similar to those found for subhalos by K04 (see their Fig. 4). The total mean occupation function is the sum of a steplike function representing the contribution of central galaxies and a power-law–like function representing the contribution of satellite galaxies.

If each sample is constructed by first selecting dark matter halos above a minimum mass, the mean occupation function $\langle N_{\mathrm{cen}}\rangle_M$ of central "subhalos" (really the halos themselves) would be a strict step function, since $N_{\mathrm{cen}} = 0$ below the minimum mass and $N_{\mathrm{cen}} = 1$ above it. However, our samples are based on minimum galaxy *baryonic* masses, so scatter in the relation between baryonic mass of the central galaxy and virial mass of the halo smooths the step, with $\langle N_{\mathrm{cen}}\rangle_M$ increasing from 0.1 to 0.9 over a factor of ~2–3 in halo mass. Since a halo necessarily contains either zero or one central galaxies, the probability distribution $P(N_{\mathrm{cen}}|\langle N_{\mathrm{cen}}\rangle_M)$ is a nearest-integer distribution (more technically, a Bernoulli distribution), with $P(1) = 1 - P(0) = \langle N_{\mathrm{cen}}\rangle_M$.

The mean occupation function of satellite galaxies in both the SPH and SA calculations is roughly a power law, $\langle N_{\mathrm{sat}}\rangle_M \propto M^{\alpha}$, although it tails off more rapidly at masses where $\langle N_{\mathrm{sat}}\rangle_M < 1$. In fact, the power-law index is $\alpha \approx 1$ in both cases, implying a simple proportionality between halo mass and satellite number. The mean number of galaxy pairs in a halo, $\langle N(N-1)\rangle$, is important for the small-scale behavior of the galaxy two-point correlation function. The top panels of Figure 1 plot the quantity $\langle N(N-1)\rangle^{1/2}/\langle N\rangle$, as a function of halo mass. If N is Poisson distributed, then this quantity is unity. Satellite galaxies have $\langle N(N-1)\rangle^{1/2}/\langle N\rangle \approx 1$ for all masses where $\langle N_{\mathrm{sat}}\rangle_M > 1$ (*open circles*). In the regime where $\langle N_{\mathrm{sat}}\rangle_M \sim 1$, the SA model seems to predict a value of $\langle N(N-1)\rangle^{1/2}/\langle N\rangle$ slightly lower than unity. The width of the total galaxy $P(N|M)$ (*filled circles*) is close to Poisson at high masses, where satellite galaxies dominate the occupation number, but it is substantially sub-Poisson $[\langle N(N-1)\rangle^{1/2}/\langle N\rangle < 1]$ at low masses because the nearest-integer distribution for central galaxies is narrower, with $N_{\mathrm{cen}}(N_{\mathrm{cen}} - 1) \equiv 0$. Our findings that satellite galaxies have a power-law–like $\langle N_{\mathrm{sat}}\rangle_M$ with $\alpha \approx 1$, that satellite numbers follow a roughly Poisson distribution with respect to their mean, and that the sub-Poisson fluctuations of the total galaxy $P(N|M)$ are thus a consequence of central galaxies are all in agreement with the findings of K04 for dark matter subhalos.

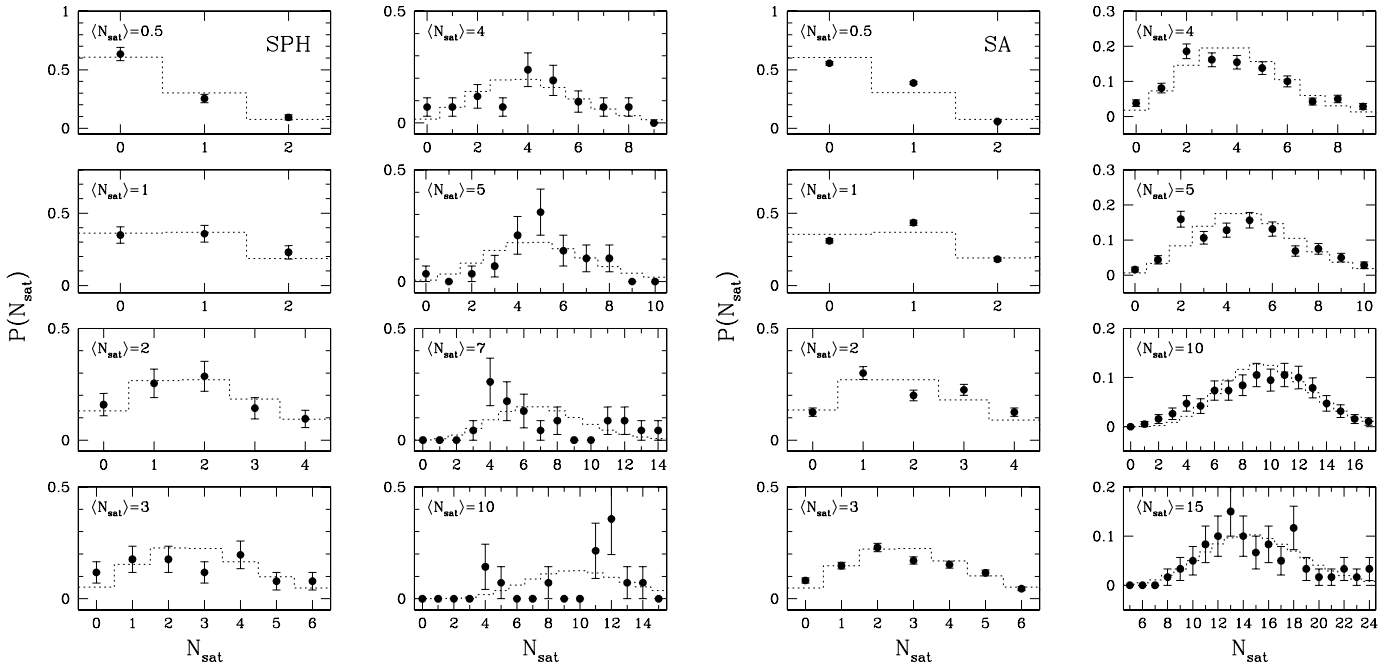

FIG. 2.—Probability distributions of satellite numbers as a function of the mean occupation number of satellites, predicted by the SPH simulation (*left*) and by the SA model (*right*). Filled circles are predictions of the models, and the Poisson error in each bin is assigned as the error bar. The dotted histogram in each panel shows the Poisson distribution of the same mean.

To test how closely the probability distribution of satellite occupation number follows a Poisson distribution, Figure 2 plots $P(N_{\rm sat}|\langle N_{\rm sat}\rangle_M)$ at different values of $\langle N_{\rm sat}\rangle_M$ (*filled circles*), in comparison to a Poisson distribution that has the same mean (*dotted histogram*). For both the SPH and the SA models, the typical bin width of halo mass is $\Delta\log M = 0.4$, but it is set to be 0.1 for the SA model if $\langle N_{\rm sat}\rangle_M \leq 2$. As a result of multiple realizations, the SA model gives better statistics than the SPH simulation, especially for high occupation numbers that correspond to rare, massive halos. Nevertheless, both models predict that $P(N_{\rm sat}|\langle N_{\rm sat}\rangle_M)$ is impressively close to a Poisson distribution over the full range $\langle N_{\rm sat}\rangle_M = 0.5-15$ where we have adequate statistics. While visual inspection of Figure 2 suggests that the match is not exact (we have not carried out a formal statistical test), it shows that the Poisson approximation for $P(N_{\rm sat}|\langle N_{\rm sat}\rangle_M)$ is likely to be adequate for most predictions of galaxy clustering statistics, such as counts-in-cells distributions and higher order correlation functions.

To model two-point correlation functions of SDSS galaxies, Zehavi et al. (2005) parameterize the mean occupation function for galaxies brighter than a luminosity threshold as a step function for $\langle N_{\rm cen}\rangle_M$ plus a truncated power law $\langle N_{\rm sat}\rangle_M = (M/M_1)^\alpha$ for satellites. This simple model has three free parameters: the minimum halo mass $M_{\rm min}$ below which $N_{\rm cen} = N_{\rm sat} \equiv 0$ and the slope α and normalization M_1 of the power law. For a given cosmology, two parameters can be adjusted to fit the observed correlation function while the third (usually $M_{\rm min}$) is set by matching the mean galaxy number density of the sample. Figure 1 shows that this parameterization captures the main features of the theoretically predicted HOD. It has the benefit of having the same number of free parameters as a power law (since one parameter is fixed by the galaxy number density), allowing a fair comparison of goodness of fit to observed correlation functions. However, as measurements of complementary clustering statistics become available, we can afford to fit HOD models with more free parameters, and they may even become necessary to match the data. For example, with the high-mass end of $\langle N\rangle_M$ constrained by the group multiplicity function (e.g., Peacock & Smith 2000; Marinoni & Hudson 2002; Kochanek et al. 2003; Yang et al. 2005a; A. Berlind, et al. 2005, in preparation), we can use the correlation function and other clustering statistics to investigate the cutoff profile of the mean occupation function. Therefore, it is useful to more accurately parameterize the results from the theoretical models with slightly more complicated prescriptions. Ultimately, we would like to fit observations using a model that does not rely on a theoretically predicted form, but in the near term we can use the predicted form and compare observationally inferred and theoretically predicted parameter values. It is worth noting that if the adopted cosmological model is substantially wrong, then *no* choice of HOD can match the full range of galaxy clustering statistics (Z. Zheng & D. Weinberg 2005, in preparation; see van den Bosch et al. [2003a] for closely related arguments using the CLF).

For both the SPH and SA models, we find that the mean occupation function for central galaxies can be well represented by

$$\langle N_{\rm cen}\rangle_M = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log M_{\rm min}}{\sigma_{\log M}}\right)\right], \qquad (1)$$

where erf is the error function

$${\rm erf}\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt. \qquad (2)$$

The parameters are $M_{\rm min}$, the characteristic minimum mass of halos that can host such central galaxies, and $\sigma_{\log M}$, the characteristic transition width. This functional form corresponds to a Gaussian distribution of $\log M_{\rm gal}$, at fixed halo mass M. If the mean baryonic mass is $\langle M_{\rm gal}\rangle \propto M^\mu$ near $M_{\rm min}$, then $\sigma_{\log M} = \sigma_{\log M_g}/\mu$, where $\sigma_{\log M_g}$ is the scatter in the logarithm of galaxy baryonic mass at fixed halo mass. For a sample defined by a luminosity threshold instead of a baryonic mass threshold, the halo mass dispersion $\sigma_{\log M}$ will be somewhat larger because of the scatter in stellar mass-to-light ratios, and the scatter may deviate

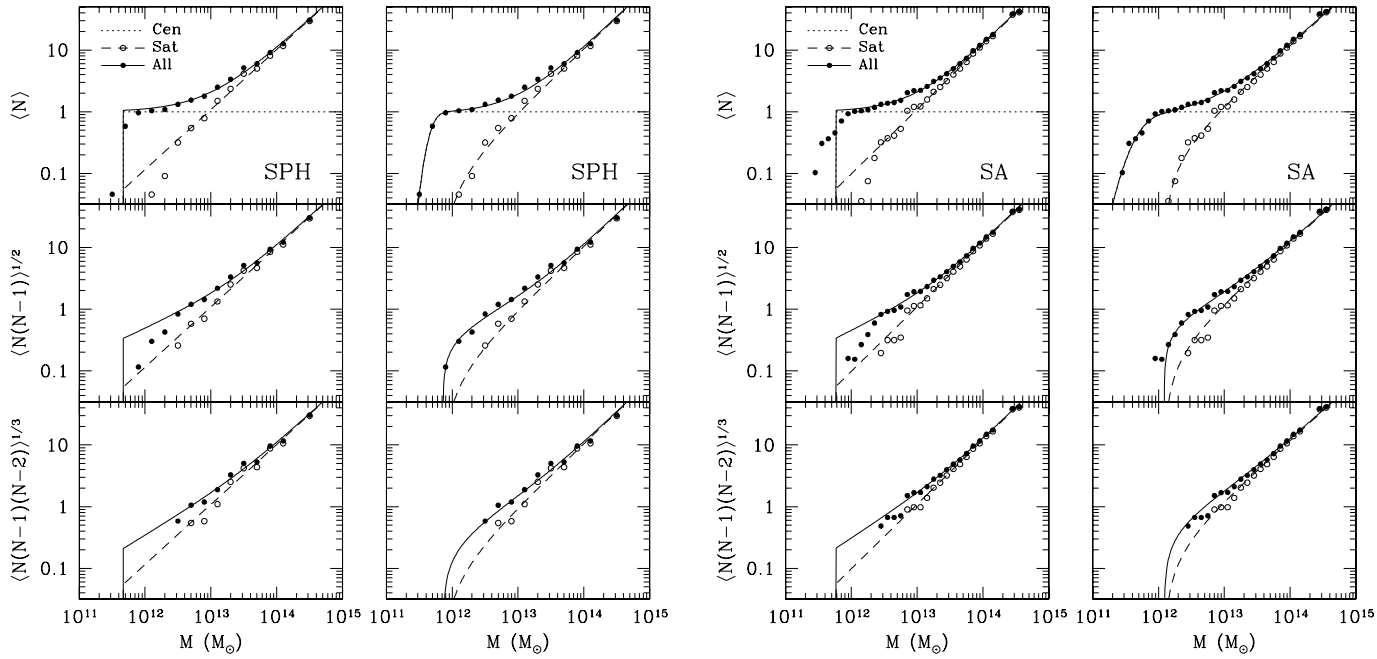

FIG. 3.—Parameterized fits to mean occupation functions (*top panels*) and predicted numbers of galaxy pairs and triplets (*middle and bottom panels*) for the SPH simulation (*left*) and the SA model (*right*). For each model, left panels show results based on three-parameter fits, which assume sharp cutoff profiles of $\langle N_{\rm cen}\rangle_M$ and $\langle N_{\rm sat}\rangle_M$, and right panels show results of fits with more parameters to model the cutoff profiles (see eqs. [1] and [3]). Fits and predictions are plotted as curves, and circles are measurements from the models.

from the Gaussian distribution. The observed Tully-Fisher (Tully & Fisher 1977) relation (roughly $\langle L\rangle \propto M$, $\sigma_{\log L} \sim 0.15$) suggests $\sigma_{\log M} \sim 0.15$ in the mass range $M \sim 10^{12}\ M_\odot$ corresponding to typical bright spiral galaxies. (All logarithms in this paper are base-10.) A detailed discussion of the scatter in the mass-luminosity relation and its effects can be found in Tasitsiomi et al. (2004).

At low masses, the mean occupation function for satellite galaxies drops below a power-law extrapolation of $\langle N_{\rm sat}\rangle_M$ from high masses. Similar to K04, we find that the full range of $\langle N_{\rm sat}\rangle_M$ can be well approximated by the form

$$\langle N_{\rm sat}\rangle_M = \left[(M - M_0)/M_1'\right]^\alpha \qquad (3)$$

for $M > M_0$, where the truncation mass M_0 for satellites may differ from the truncation mass $M_{\rm min}$ for central galaxies. Note that with this parameterization, M_1' is *not* the mass at which $\langle N_{\rm sat}\rangle_M = 1$. The top panels of Figure 3 show fits of the three-parameter ($M_{\rm min}$, M_1, α) and five-parameter ($M_{\rm min}$, $\sigma_{\log M}$, M_1', M_0, α) models to the SPH and SA mean occupation functions. The parameter values are listed in Table 1. Note that these values are likely to depend on the cosmological model as well as the galaxy space density and the physics encoded in the SPH and SA calculations. The three-parameter form (used by Zehavi et al. 2005) captures the results of both models well but not perfectly, while the five-parameter form gives a nearly perfect fit in both cases.

If we assume that central and satellite galaxies have nearest-integer and Poisson distributions, respectively, then we can calculate all the higher moments of the occupation number based on the fits to the mean occupation. In particular, we can predict average numbers of galaxy pairs and triplets inside halos, which are relevant to the one-halo terms of the galaxy two-point and three-point correlation functions, respectively. The total occupation number N is the sum of $N_{\rm cen}$ and $N_{\rm sat}$, and it is easy to show that

$$\langle N(N-1)\rangle = \langle N_{\rm cen}(N_{\rm cen}-1)\rangle + 2\langle N_{\rm cen}N_{\rm sat}\rangle + \langle N_{\rm sat}(N_{\rm sat}-1)\rangle \qquad (4)$$

TABLE 1
HOD PARAMETERS FOR GALAXY SAMPLES WITH DIFFERENT THRESHOLDS OF BARYONIC MASS

		THREE-PARAMETER MODEL			FIVE-PARAMETER MODEL				
$\bar{n}_g$	MODEL	$\log M_{\rm min}$	$\log M_1$	α	$\log M_{\rm min}$	$\sigma_{\log M}$	M_0	$\log M_1'$	α
0.02	SPH	11.67	12.96	0.97	11.68	0.15	11.86	13.00	1.02
	SA	11.77	12.96	1.04	11.73	0.32	12.09	12.87	0.96
0.01	SPH	12.04	13.26	1.03	12.07	0.18	12.28	13.19	0.94
	SA	12.03	13.30	1.02	12.02	0.26	12.28	13.32	1.07
0.005	SPH	12.38	13.55	1.18	12.36	0.15	12.63	13.45	1.00
	SA	12.34	13.64	1.09	12.36	0.42	12.28	13.62	1.04
0.0025	SPH	12.68	13.85	1.24	12.69	0.15	12.94	13.82	1.08
	SA	12.58	13.91	1.16	12.60	0.28	12.77	13.86	1.03

NOTES.—Number density and mass are in units of h^3 Mpc^{-3} and $M_\odot$, respectively. For the three-parameter model, $M_{\rm min}$ is simply set to be the halo mass at which $\langle N\rangle_M = 0.5$, and M_1 and α are obtained through a power-law fit to data points with $\langle N_{\rm sat}\rangle_M > 0.1$.

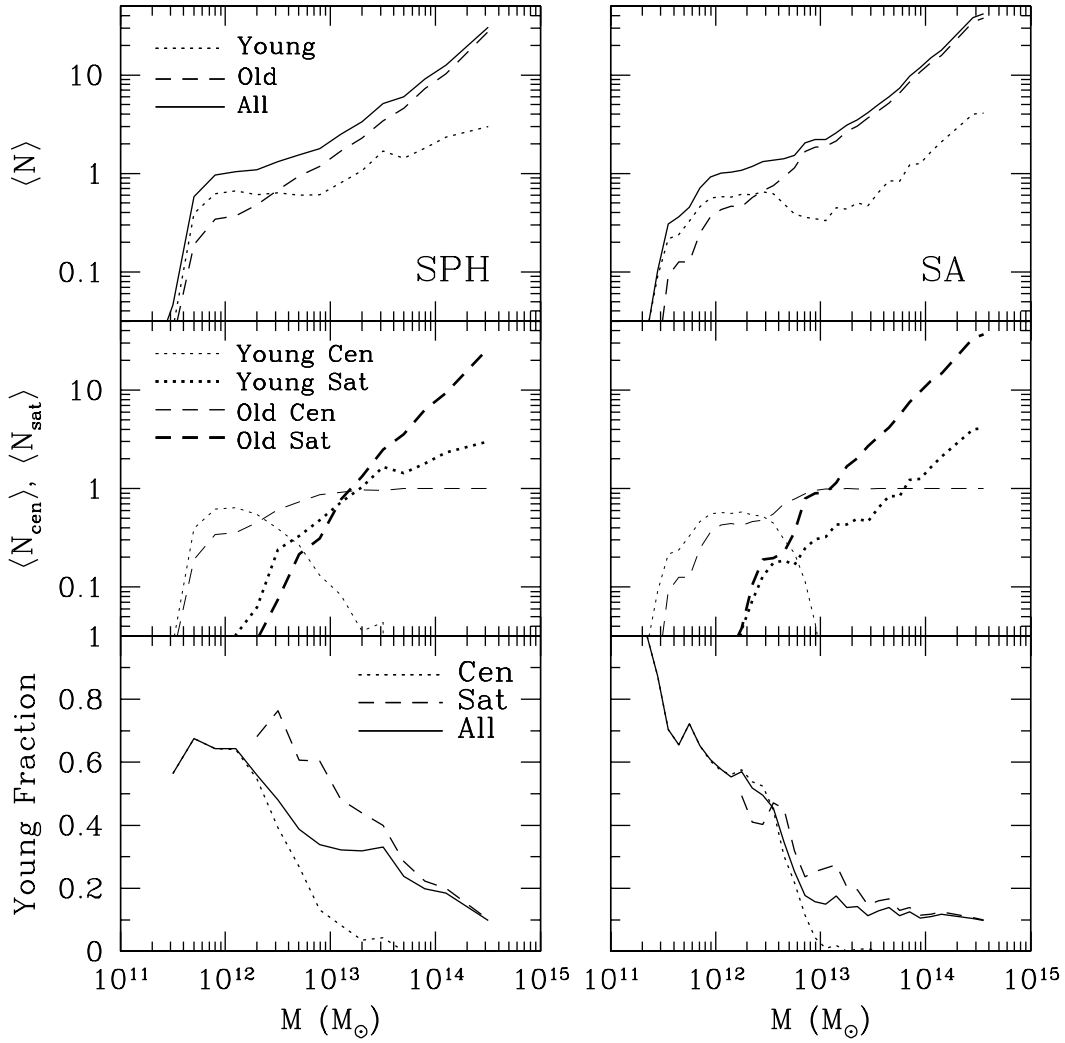

FIG. 4.—Age dependence of the HOD predicted by the SPH simulation (*left*) and by the SA model (*right*). For each model, the mean occupation functions of old and young galaxies are shown in the top panel, contributions from central and satellite galaxies to the mean occupation number are plotted in the middle panel, and the fraction of young galaxies (in central, satellite, and all galaxies) is plotted in the bottom panel.

and

$$\begin{aligned}\langle N(N-1)(N-2)\rangle = {} & \langle N_{\rm cen}(N_{\rm cen}-1)(N_{\rm cen}-2)\rangle \\ & + 3\langle N_{\rm cen}(N_{\rm cen}-1)N_{\rm sat}\rangle \\ & + 3\langle N_{\rm cen}N_{\rm sat}(N_{\rm sat}-1)\rangle \\ & + \langle N_{\rm sat}(N_{\rm sat}-1)(N_{\rm sat}-2)\rangle. \end{aligned} \quad (5)$$

Since $N_{\rm cen} = 0$ or 1, the first term on the right-hand side of equation (4) and the first two terms on the right-hand side of equation (5) vanish. The first surviving terms represent pairs or triplets involving the central galaxy, while the last terms represent combinations that only involve satellites. To deal with the cross-correlation between occupations of central and satellite galaxies, we simply need to note that if $N_{\rm cen} = 0$, then $N_{\rm sat} = 0$ by definition. Therefore, $\langle N_{\rm cen}N_{\rm sat}\rangle = \langle N_{\rm sat}\rangle$ and $\langle N_{\rm cen}N_{\rm sat}(N_{\rm sat}-1)\rangle = \langle N_{\rm sat}(N_{\rm sat}-1)\rangle$. Assuming a Poisson distribution for $N_{\rm sat}$, equations (4) and (5) reduce to

$$\langle N(N-1)\rangle = 2\langle N_{\rm sat}\rangle + \langle N_{\rm sat}\rangle^2 \quad (6)$$

and

$$\langle N(N-1)(N-2)\rangle = 3\langle N_{\rm sat}\rangle^2 + \langle N_{\rm sat}\rangle^3. \quad (7)$$

We plot predictions based on these two equations and the fits to $\langle N_{\rm cen}\rangle_M$ and $\langle N_{\rm sat}\rangle_M$ in the middle and bottom panels of Figure 3. Symbols in these panels are measurements from the SPH and SA calculations. We see that the three-parameter model overpredicts the number of galaxy pairs and triplets in low-mass halos, mainly because the number of satellites is overpredicted by ignoring the profile of the low-mass cutoff. For the five-parameter fit to $\langle N\rangle_M$, the predicted numbers of pairs and triplets agree remarkably well with the measured values, providing further evidence that the Poisson approximation for $P(N_{\rm sat}|\langle N_{\rm sat}\rangle_M)$ is adequate for practical calculations.

3.2. *HOD for Young and Old Galaxies*

The top panels of Figure 4 show mean occupation functions for the young and old halves of the $\bar{n}_g = 0.02\,h^3\,{\rm Mpc}^{-3}$ sample, in comparison to $\langle N\rangle_M$ of the full sample. Solid curves in the bottom panels show the fraction of young galaxies as a function of halo mass. These are basically the same as in Figure 13 of B03. While the mean occupation number of old galaxies rises continuously as halo mass increases, the $\langle N\rangle_M$ of young galaxies first rises, then declines to a local minimum at $M \sim 10^{13}\,M_\odot$, and then rises again. Sheth & Diaferio (2001) find similar nonmonotonic behavior for the mean occupation of blue galaxies in the SA models of Kauffmann et al. (1999). The fraction of young galaxies

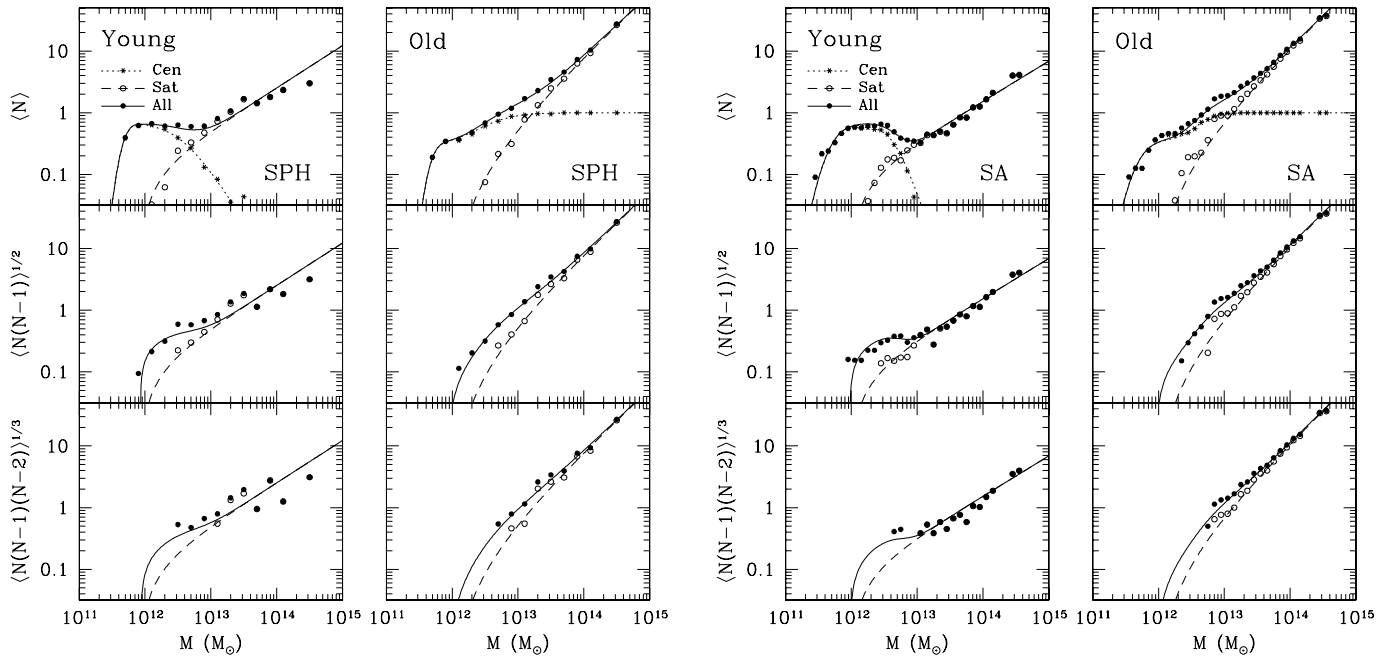

FIG. 5.—Same as Fig. 3, but for each model the left panels show the young galaxy sample and right panels the old galaxy sample. Symbols show the model results and curves show fits using the five-parameter model for all galaxies and the blue fraction parameterization described in the text.

(*bottom panels*) has a steep drop at low occupation number and decreases slowly toward higher occupation number.

The shapes of these mean occupation functions are easy to understand if we separate contributions from central and satellite galaxies, as shown in the middle panels of Figure 4. As shown by B03 (their Fig. 19), both the SPH simulation and the SA model predict that on average the (median/mean) stellar age of a halo's central galaxy is an increasing function of halo mass. As explained by B03, this correlation between stellar age of the central galaxy and the mass of the parent halo arises from two physical effects: higher mass halos begin to assemble earlier, allowing an earlier onset of star formation in the central galaxy, and gas accretion rates drop once a halo becomes massive enough to support a virial shock (Keres et al. 2005), choking off the formation of young stars at late times. In halos of mass near $M_{\rm min}$, therefore, central galaxies are usually young (i.e., below the sample's median age), while in high-mass halos they are all old. The minimum in $\langle N\rangle_M$ of young galaxies occurs for halos that are too massive to host young central galaxies but not massive enough to host satellite galaxies. The typical age of satellite galaxies also increases with halo mass, but the correlation is weaker (see B03, Fig. 19) because satellites experience most of their growth in lower mass halos that merge into their final host halo. In Figure 4, the young galaxy fraction for central galaxies drops rapidly with increasing halo mass and falls essentially to zero, while the young galaxy fraction for satellites drops more slowly and in the SA calculation reaches a nearly flat plateau at high mass. The different HODs of young and old galaxies reflect fundamental aspects of galaxy formation physics, and their origin is transparent once we separate central and satellite contributions.

As with the full sample HODs, we would like to find simple parameterized prescriptions that capture these age-dependent features, which we can do by describing the young galaxy central and satellite fractions as a function of halo mass. Based on the bottom panels of Figure 4, we adopt

$$f_{\rm young,cen} = f_c\left[1+\exp\left(\frac{\log M-\log M_c}{\sigma_{\log M_c}}\right)\right]^{-1}, \qquad (8)$$

where f_c sets the amplitude and M_c and $\sigma_{\log M_c}$ characterize the transition mass scale and speed. We approximate the fraction of young satellite galaxies by an exponential,

$$f_{\rm young,sat} = f_s\exp\left(-\frac{\log M-\log M_0}{\sigma_{\log M_s}}\right), \qquad (9)$$

where M_0 is the same satellite cutoff mass used in equation (3) and $\sigma_{\log M_s}$ characterizes the speed of the falloff. The top panels of Figure 5 show that the above functional forms for the $\bar{n}_g = 0.02\ h^3\ {\rm Mpc}^{-3}$ samples, together with our five-parameter model for the full galaxy $\langle N\rangle_M$, allow accurate fits to the SPH and SA mean occupation functions. The parameters we use for the SPH (SA) model are $f_c = 0.71$ (0.65), $\log M_c = 12.55$ (12.64), $\sigma_{\log M_c} = 0.26$ (0.14), $f_s = 0.99$ (0.80), and $\sigma_{\log M_s} = 1.50$ (1.10).

To predict numbers of galaxy pairs and triplets, we need additional assumptions. We cannot automatically simplify the central-satellite cross terms of equations (4) and (5) to obtain equations (6) and (7) because a halo may, for example, have an old central galaxy ($N_{\rm cen,\,young} = 0$) but nonetheless have young satellites. To proceed, we assume that the satellite population in a halo of a given mass does not depend on the age of the central galaxy and that the occupation numbers of young and old satellite galaxies follow independent Poisson distributions with respect to their individual means. Equations (6) and (7) are then modified by changing the first right-hand side terms to $2\langle N_{\rm cen}\rangle\langle N_{\rm sat}\rangle$ and $3\langle N_{\rm cen}\rangle\langle N_{\rm sat}\rangle^2$, respectively. The middle and bottom panels of Figure 5, which compare pair and triplet predictions from the mean occupation fits to the SPH and SA results, suggest that the above approximations are accurate enough for calculations of galaxy clustering.

Figure 6 shows HODs for samples defined by a bin of baryonic mass instead of a threshold; specifically, we take the less massive half of our $\bar{n}_g = 0.02\ h^3\ {\rm Mpc}^{-3}$ samples, and we again divide into young and old subsamples. The shapes of the mean occupation functions for all, young, and old galaxies look alike: a bump at lower halo masses resulting from central galaxies and an approximately power-law function toward higher halo

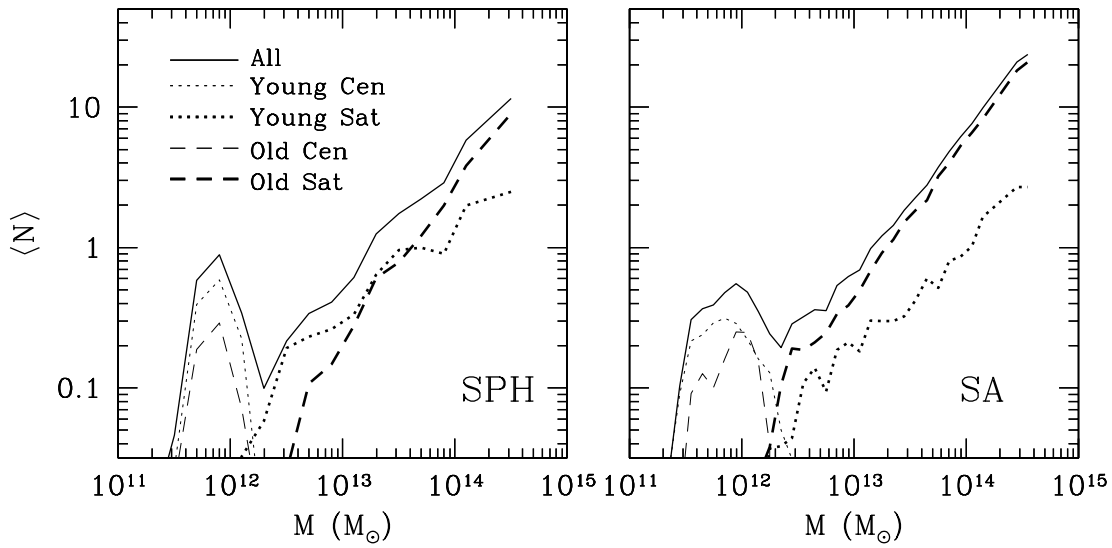

FIG. 6.—HOD predicted by the SPH simulation (*left*) and by the SA model (*right*) for a sample of galaxies in a bin of baryonic mass, containing the less massive half of the full galaxy sample. Dotted and dashed curves show young and old galaxy contributions, respectively.

masses representing the satellite contribution. Guzik & Seljak (2002) find a similar result for a luminosity bin sample using the Kauffmann et al. (1999) SA model. For our chosen mass bin, the majority of central galaxies are young, and (since the young and old samples are roughly equal) the majority of satellites are therefore old. Lower mass bins have a higher fraction of young central galaxies, while high-mass bins have predominantly old central galaxies. The three-parameter and five-parameter models described in § 3.1 are not appropriate for describing the HOD of a mass or luminosity bin because they do not allow an upper cutoff in $\langle N_{\rm cen}\rangle_M$. However, these models do provide good descriptions for samples defined by mass or luminosity thresholds, and the HOD of a bin sample is just the difference of the HODs for thresholds at the bin's upper and lower limits.

3.3. *HOD Parameters as a Function of Galaxy Mass/Luminosity*

As the threshold baryonic mass is increased, the predicted mean occupation function maintains the same general form but shifts horizontally toward higher masses in the $\log\langle N\rangle_M - \log M$ plane (B03, Fig. 9). Figure 7 examines the dependence of characteristic halo mass scales for central and satellite galaxies on the baryonic mass threshold. The lower solid curves represent $M_{\rm min}$, the characteristic minimum mass for hosting a central galaxy above the threshold. Since $\langle N_{\rm cen}\rangle_M$ is not a strict step function, we simply define $M_{\rm min}$ in this figure as the mass at which $\langle N\rangle_M = 0.5$; half of the halos with $M = M_{\rm min}$ contain a central galaxy above the threshold, and half do not. This halo mass threshold is linearly proportional to the baryonic mass threshold, $M_{\rm min} \propto M_{\rm gal}$, for $M_{\rm min} \lesssim 3\times10^{12}\,M_\odot$, implying that the baryonic mass of the central galaxy is proportional to the mass of the host halo in this regime. The dot-dashed curve in each panel marks $M_{\rm min} = (\Omega_b/\Omega_m)M_{\rm gal}$, expected if the ratio of central galaxy mass to halo virial mass is equal to the universal baryon fraction. The SPH results lie close to this limiting case for low-halo masses, while the SA central galaxies never accrete more than about 25% of the available baryons. As discussed by B03, this difference can be attributed to treatments of the gas core radius and stellar feedback in the SA model, which are adjusted to fit the observed galaxy LF. The SPH simulation predicts excessively massive galaxies, assuming a standard stellar initial mass function (IMF). Additional processes such as photoionization by the UV background or supernova gas blowout may suppress accretion in very low mass halos and produce an upturn in the $M_{\rm min}$-$M_{\rm gal}$ relation, but they are not evident in the mass range probed here.

In high-mass halos, some of the baryonic mass goes into satellite galaxies, and the efficiency of gas cooling drops so that a smaller fraction of baryons are accreted onto galaxies (B03, Fig. 5). As a result, $M_{\rm min}$ begins to grow faster than $M_{\rm gal}$ at masses $M_{\rm min} \gtrsim 3\times10^{12}\,M_\odot$, corresponding to group or cluster halos. The SA model predicts a much steeper rise than the SPH simulation, a consequence of resolution effects in the SPH simulation and of the different ways the two models treat cooling and feedback. The SPH simulation includes thermal feedback from supernovae, but the thermal energy is usually deposited in dense gas and radiated away before it can drive a galactic wind (Katz et al. 1996). This relatively efficient cooling produces overly luminous galaxies at the centers of groups and clusters (again assuming a standard stellar IMF), although the baryonic masses and hence luminosities are reduced somewhat when the numerical resolution is increased (see B03, Fig. 5). In the SA model, the formation of very bright galaxies is suppressed by adjusting the core radius in the gas density profile, which controls how much of the gas can cool, so that the model matches the bright end of the observed galaxy LF (Cole et al. 2000). The core radius mechanism is somewhat artificial, and the "solution" in the real universe probably involves some additional physics such as thermal conduction or AGN-driven superwinds (see Benson et al. 2003a) or reduced cooling efficiency in multiphase halo gas (Maller & Bullock 2004). However, the trend of $M_{\rm min}$ with $M_{\rm gal}$ is likely to resemble the SA curve shown here for any model that matches the observed galaxy LF, since the halo mass function (MF) is fixed by the cosmology and the LF by observations.

The upper solid curves in each panel of Figure 7 show the dependence of M_1 on galaxy baryonic mass, where M_1 is the mass of halos that on average contain one satellite galaxy (i.e., $\langle N\rangle_M = 2$) above the baryonic mass threshold. These curves look remarkably like scaled versions of the $M_{\rm min}$ curves. The dashed curves represent $14M_{\rm min}$ and $18M_{\rm min}$ for the SPH and SA models, respectively, and they match the M_1 curves almost perfectly except for very massive halos. The smaller scaling factor in the SPH simulations

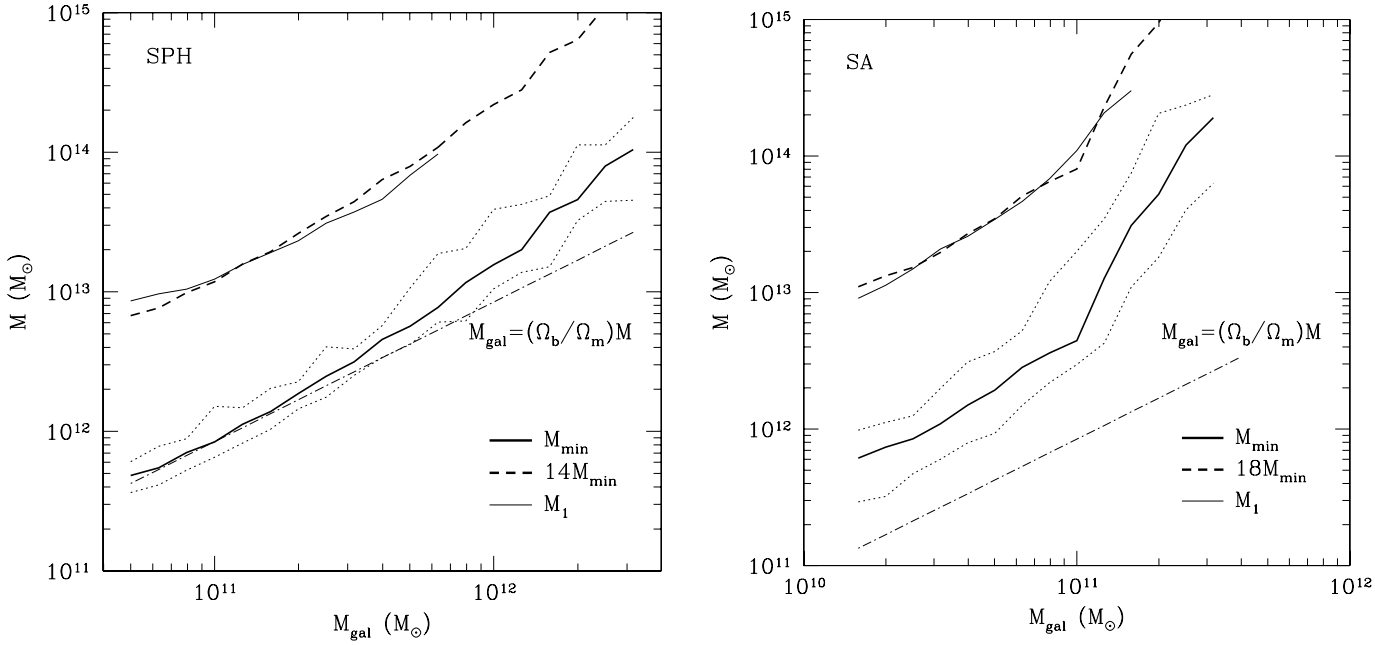

FIG. 7.—HOD parameters as a function of galaxy baryonic mass threshold (*left*: SPH simulation; *right*: SA model). In each panel, the thick solid curve is $M_{\rm min}$, the characteristic minimum mass of halos that can host galaxies above a given threshold, which is defined here as the mass of halos that have $\langle N_{\rm cen}\rangle_M = 0.5$. The two dotted curves denote masses of halos that have $\langle N_{\rm cen}\rangle_M = 0.1$ and 0.9, respectively. The thin solid curve is M_1, which is the mass of halos that can on average host one satellite galaxy above the given threshold. The dashed curve indicates $14M_{\rm min}$ (SPH) or $18M_{\rm min}$ (SA). For comparison, the dot-dashed curve shows the baryonic mass corresponding to the universal baryon fraction.

is at least partly a consequence of the more efficient cooling onto central galaxies, discussed above, which leads the simulation to produce galaxies that are increasingly overluminous as halo mass increases. The large gap between $M_{\rm min}$ and M_1, more than an order of magnitude in both calculations, produces the plateau in the mean occupation function between $\langle N\rangle_M \sim 1$ and $\langle N\rangle_M \sim 2$. As discussed by B03, a halo that is only a few times $M_{\rm min}$ usually "spends" its extra baryons building a larger central galaxy, instead of making two galaxies of comparable mass. A halo has to be $\sim$20 times more massive than $M_{\rm min}$ before it has a high chance of merging with a halo massive enough to contribute a satellite above the threshold.

For N-body subhalos analyzed in K04, the scaling factor between $M_{\rm min}$ and M_1 is about 25 at $z = 0$ (and becomes smaller at higher z, e.g., $\sim$10 at $z = 3$). The modest difference between our results and theirs may be caused partly by slight differences in defining $M_{\rm min}$ and partly by the difference in the cosmological models adopted in the calculation. K04 also find that the high-mass slope (α) of the satellite mean occupation is very close to 1, independent of the mass threshold. We find similar results through our fits to the satellite occupation functions for galaxy samples with different baryonic mass thresholds (see Table 1).

4. RELATIONS BETWEEN GALAXY MASS/LUMINOSITY AND HALO MASS

4.1. *Conditional Mass/Luminosity Function as a Function of Halo Mass*

The CLF (Yang et al. 2003) is the LF $\Phi(L|M)$ of galaxies that reside in halos of a given mass M. The CLF encodes the luminosity dependence of the mean occupation function, since a complete characterization of $\Phi(L|M)$ allows one to examine the M dependence for any specified range of L. This formalism has been used to model observations of the 2dFGRS and the DEEP2 redshift survey by simultaneously fitting measurements of the galaxy LF and luminosity-dependent clustering (Yang et al. 2003; van den Bosch et al. 2003a; Yan et al. 2003). A derived or assumed CLF can be used to construct mock catalogs, allowing more detailed tests of the CLF and cosmological model (e.g., Mo et al. 2004; Yan et al. 2004; Yang et al. 2004). The spirit of CLF modeling is similar to that of fitting parameterized HOD models like those of § 3 to observed clustering in different luminosity ranges, as undertaken for the SDSS by Zehavi et al. (2005; see also Seljak et al. [2005] for an analysis using galaxy-galaxy lensing measurements). In HOD fits, information about the galaxy LF enters through the number density constraints. The CLF approach attempts to solve for the entire luminosity dependence of the HOD at once, with the consequence that it requires a more detailed parameterized form for model fitting. The above-mentioned papers usually assume that the CLF follows a Schechter (1976) function at each halo mass, and they parameterize the halo mass dependence of the normalization, faint-end slope, and characteristic luminosity of this Schechter function with various functional forms. The central galaxy in each halo is chosen to be the most luminous one of the galaxies drawn from the CLF.

Since our galaxy sample is constructed based on a baryonic mass threshold, here we first examine the theoretical predictions for conditional galaxy baryonic MFs, which we refer to as the CMF (note that M here represents galaxy baryonic mass, not halo mass). The CMF should be similar to the CLF, although the systematic mass dependence and scatter of stellar mass-to-light ratios will stretch and smooth the CLF somewhat relative to the CMF. The qualitative results for the SPH simulation are similar to those for the SA model except for the overall shift in baryonic mass scale. We concentrate on the SA model because it has better statistics and because it better matches the observed galaxy LF. The SA model also predicts galaxy luminosities, and Benson et al. (2003b) have presented SA calculations of CLFs. We compare CLF and CMF results later in this section.

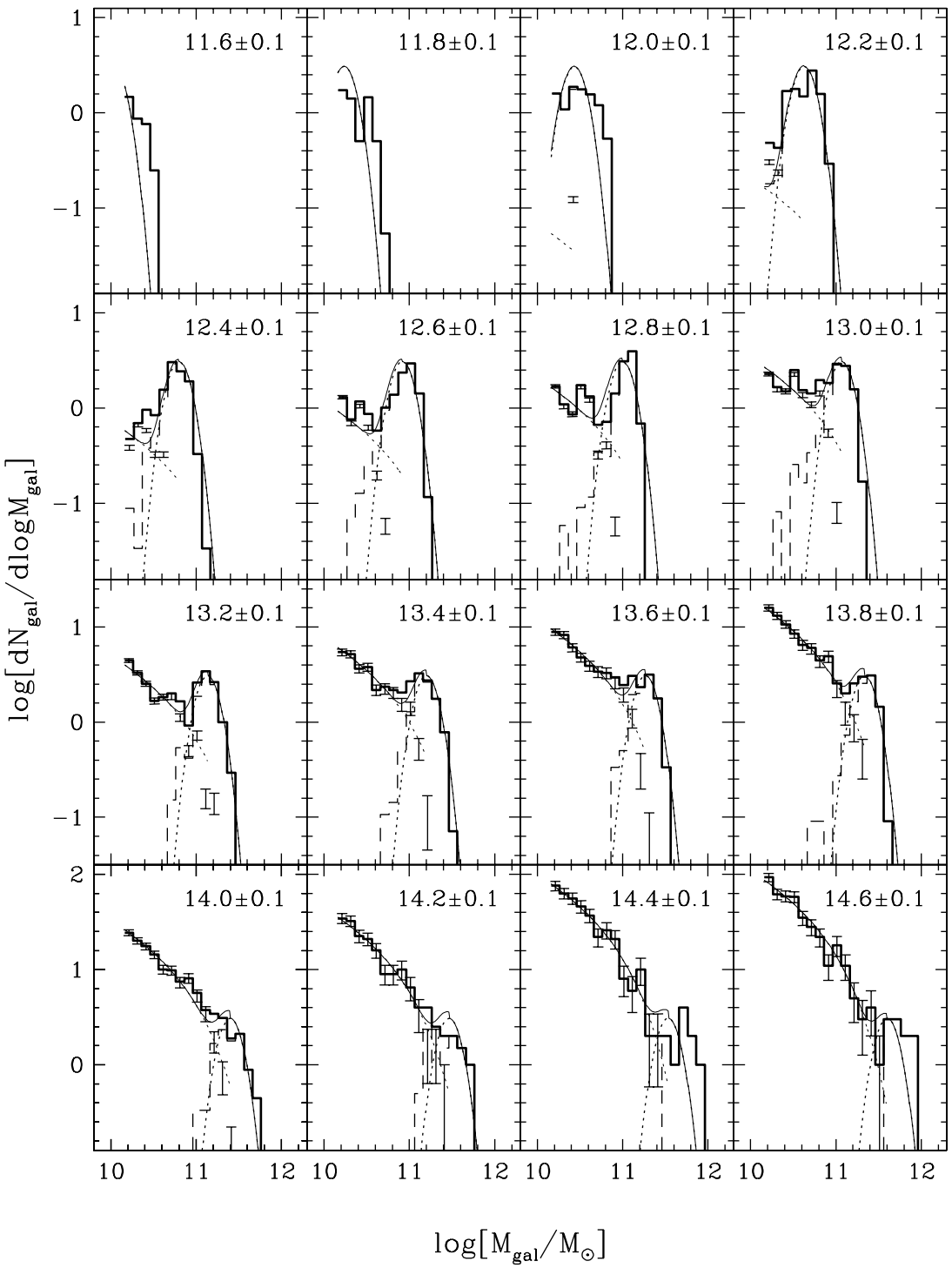

FIG. 8.—CMFs predicted by the SA model as a function of halo mass [the label in each panel is the range of log ($M/M_{\odot}$)]. In each panel, the CMF is normalized in such a way that the number of central galaxies in a halo is unity. The total CMF (*thick solid histogram*) is decomposed into contributions from central galaxies (*dashed histogram*) and satellites (*dots with Poisson error bars*). The thin solid curve is a sum of a truncated Schechter function (*dotted curve for satellite galaxies*) and a Gaussian function (*dotted curve for central galaxies*), using the parameters shown by the dashed curves in the middle and bottom panels of Fig. 9 and further described in the text.

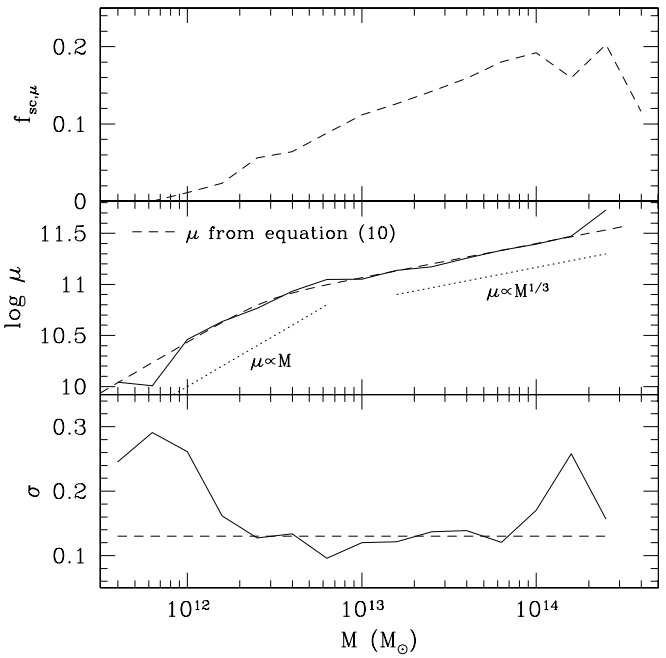

FIG. 9.—Parameters related to the CMF in Fig. 8. The middle and bottom panels show the best-fit values of the mean μ and the width σ of the Gaussian function representing the central galaxy CMF. For Gaussian fits shown in Fig. 8, an analytic function (*dashed curve in the middle panel*) is used to represent the mean μ, a constant σ is adopted (*dashed curve in the bottom panel*), and the normalization is set to have one central galaxy in a halo. The top panel shows the ratio of the amplitudes of the truncated Schechter function in Fig. 8 (CMF for satellite galaxies) and the Gaussian function (CMF for central galaxies) evaluated at the mean μ of the Gaussian function.

Figure 8 plots the CMF in bins of halo mass running from $\log M = 11.6 \pm 0.1$ to 14.6 ± 0.1. Solid histograms show the full CMF, dashed histograms the contribution of central galaxies only, and points with Poisson error bars the contributions of satellite galaxies only. Since we consider only galaxies with baryonic mass $M_{\rm gal} > 1.45 \times 10^{10}\ M_\odot$, the CMF in halos of mass $M \lesssim 10^{12}\ M_\odot$ is completely dominated by central galaxies; the two contributions become equal at halo mass $M \approx 10^{13}\ M_\odot$, where $\langle N_{\rm sat}\rangle_M = 1$ (Fig. 7). The central galaxy CMF is sharply peaked at a fixed halo mass, and it can be approximated by a Gaussian in $\log M_{\rm gal}$. Solid curves in Figure 9 show the parameters of such Gaussian fits as a function of M. The dispersion is $\sigma \approx 0.13$ over a wide range, and much of this width reflects the ± 0.1 size of our $\log M$ bins. The dispersion rises at low and high halo masses. At low M, the mean central galaxy mass is linearly proportional to halo mass, while the relation at high mass is much shallower, approximately $M_{\rm gal} \propto M^{1/3}$. The ratio of the integrated central galaxy CMF to the total CMF follows $\langle N_{\rm cen}\rangle_M/(\langle N_{\rm cen}\rangle_M + \langle N_{\rm sat}\rangle_M)$ once the baryonic mass scale of the central galaxies is well above the threshold baryonic mass, which happens in halos more massive than $\sim 2 \times 10^{12}\ M_\odot$.

The satellite galaxy CMF rises monotonically toward lower $M_{\rm gal}$ at any halo mass, and it can be approximated by a Schechter function that is truncated at the middle of the Gaussian representing the central galaxy CMF. Thin solid curves in Figure 8 show a model in which the satellite contribution is a sharply truncated Schechter function with $\alpha_{\rm S} = -1.5$ and $M^*_{\rm gal} = 10^{11}\ M_\odot$ at every halo mass[9] and the central galaxy contribution is a Gaussian in $\log M_{\rm gal}$ with constant width $\sigma = 0.13$. The mean of the Gaussian, shown by the dashed curve in the middle panel of Figure 9, is

$$\mu(M) = \mu_t \left[\frac{1}{1+c} \left(\frac{M}{M_t} \right)^{-\gamma} + \frac{c}{1+c} \left(\frac{M}{M_t} \right)^{-\gamma/3} \right]^{-1/\gamma}, \quad (10)$$

where $\mu_t = 5.9 \times 10^{10}\ M_\odot$, $M_t = 2.3 \times 10^{12}\ M_\odot$, $c = 0.49$, and $\gamma = 5.8$. The normalization of the truncated Schechter function is determined by matching the average number of satellite galaxies above the baryonic mass threshold in each halo mass bin, and the central galaxy CMF integrates to unity by definition (i.e., we assume that all halos contain a central galaxy of some mass, although it might be below our adopted threshold). The top panel of Figure 9 shows the relative amplitudes of central and satellite CMFs evaluated at the mean of the Gaussian component. Even though the relative normalization is the only parameter adjusted on a bin-by-bin basis, this global fit describes the SA model CMF quite well over the full range of halo masses plotted in Figure 8.

With the CMF/CLF, one can populate halos from simulations to study various properties of galaxy clustering. For a halo of a given mass, one chooses the baryonic mass or luminosity for the central galaxy according to the central galaxy CMF/CLF and does similarly for the satellites. This procedure implicitly assumes that there is no correlation between the masses/luminosities of central galaxies and satellites in halos of fixed mass. Figure 10 tests the validity of this assumption by comparing satellite CMFs at different values of central galaxy baryonic mass in three narrow bins of halo mass. The central and satellite CMFs are normalized so that there is one central galaxy per halo. For $10^{13}\ M_\odot$ halos, there is a clear trend for a lower amplitude satellite CMF in halos with a more massive central galaxy. The trend is present but weak in the more massive halos. This trend could be a consequence of galactic cannibalism, with more massive central galaxies growing by consuming more satellites. Note that in a small fraction of halos, the most massive satellite exceeds the mass of the central galaxy. While the SA model always places the more massive galaxy from a pair of merged halos at the center of the new halo, that galaxy may be more gas-rich than its partner, in which case feedback from star formation can leave it slightly lower mass in the end.

A clear consequence of the central galaxy contribution is that the CMF at a given halo mass *cannot* be well approximated by a Schechter function, especially at intermediate masses $10^{12.5}\ M_\odot \lesssim M \lesssim 10^{14}\ M_\odot$ where the central galaxy "bump" rises well above the extrapolation of the satellite CMF. At high M the fractional contribution of the central galaxy is small, and a Schechter function is a reasonable approximation for most purposes. At low M, where central galaxies dominate, the Gaussian CMF can be roughly approximated by a Schechter function with $\alpha_{\rm S} > 0$. Yang et al. (2003) indeed find $\alpha_{\rm S} > 0$ at low masses in their (Schechter based) fits to the 2dFGRS CLF (see their eq. [20] and Table 1). Note, however, that if we lowered our galaxy mass threshold, we would pick up lower mass (fainter) satellites in low-mass halos, and a Schechter form would again become a poor description.

Benson et al. (2003b) study the CLF in SA models and find a similar trend to that seen here in the CMF: in low-mass halos the CLF has a "hump" at the bright end and a roughly power-law shape at the faint end, and in massive halos the hump disappears and the CLF approaches the Schechter form (see their Fig. 1). Figure 11 presents similar results for the cosmology and version of GALFORM adopted here, in Sloan g and r bands. We truncate the histograms at $M_g = -18.75$ and $M_r = -19.25$, since scatter in the mass-luminosity relation would make our mass-limited

[9] Note that we use $\alpha_{\rm S}$ to distinguish the Schechter function slope from the slope α of $\langle N_{\rm sat}\rangle_M$.

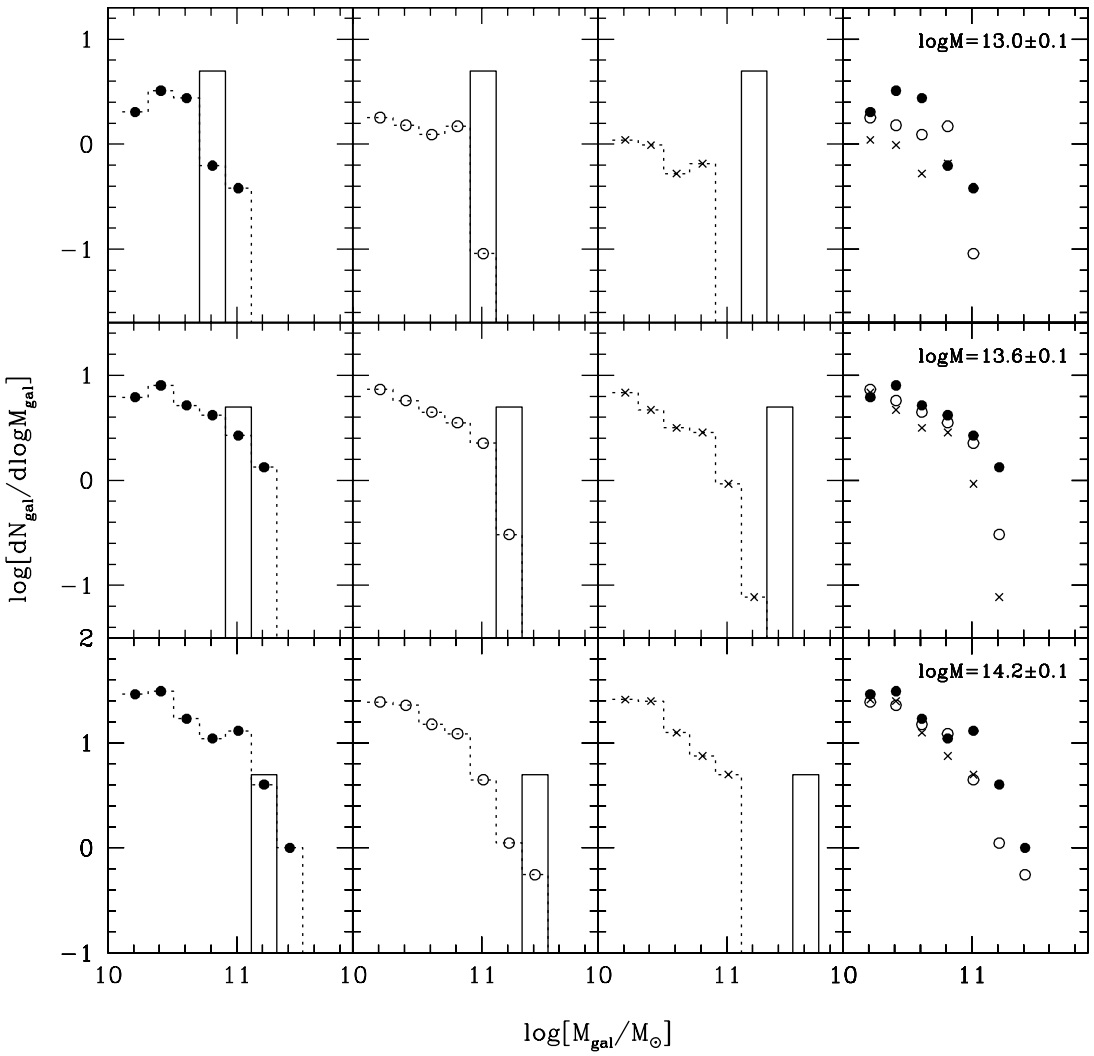

FIG. 10.—Satellite CMF as a function of central galaxy mass in three halo mass bins, $\log(M/M_{\odot}) = 13.0$ (*top*), 13.6 (*middle*), and 14.2 (*bottom*). For each halo mass, the first three panels show satellite CMFs (*dotted histograms with filled circles*) at different central galaxy masses (solid histograms are central galaxy CMFs); satellite CMFs in these three panels are plotted together in the right panel for direct comparison. CMFs are normalized such that there is one central galaxy in a halo.

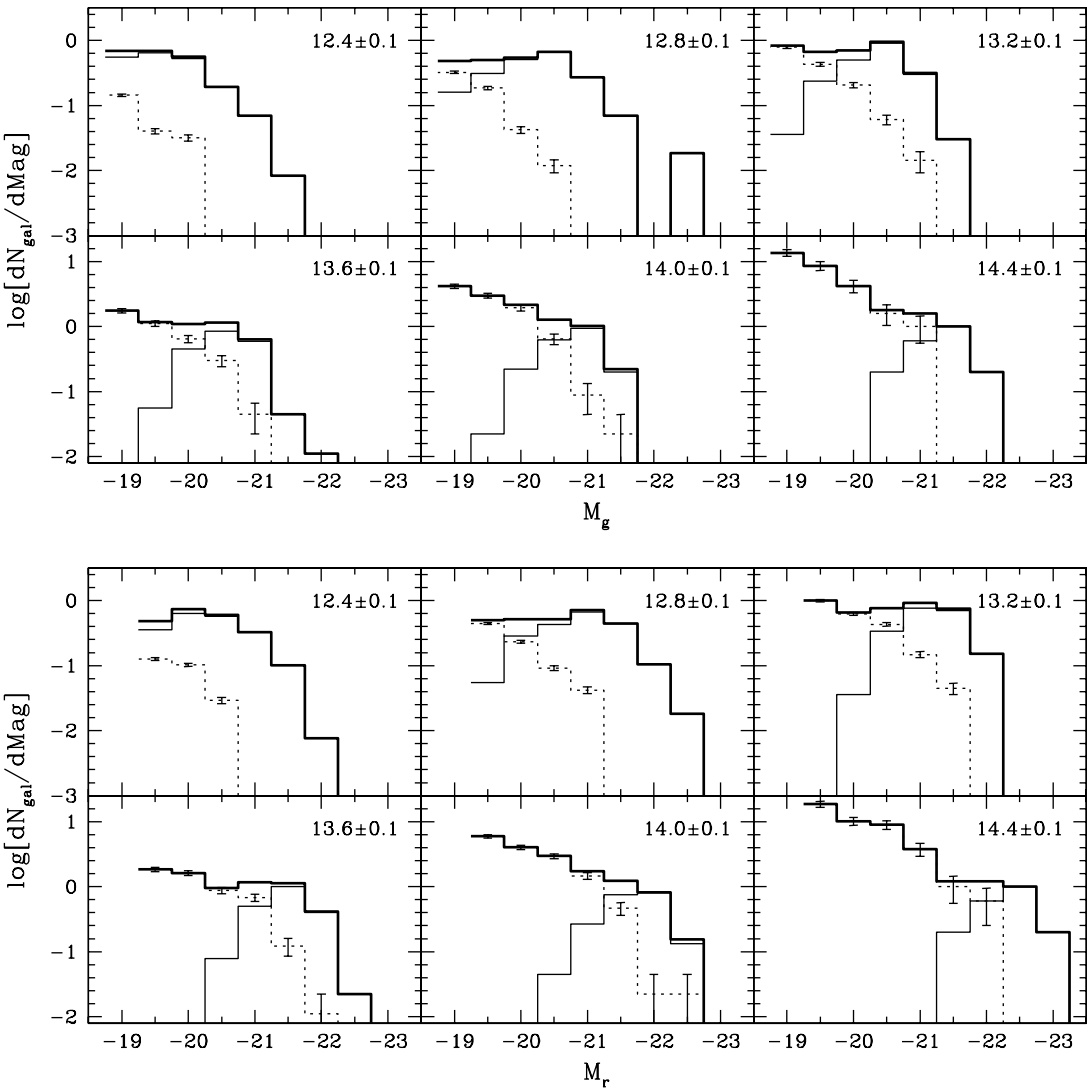

FIG. 11.—CLFs in *g* band (*top*) and *r* band (*bottom*) predicted by the SA model as a function of halo mass. Similar to Fig. 8, the label in each panel marks the range of $\log(M/M_{\odot})$. The dotted, thin solid, and thick solid histograms are CLFs for satellite, central, and all galaxies, respectively. The faintest bin of the CLF is set by the luminosity above which the completeness fraction is unity (see the text).

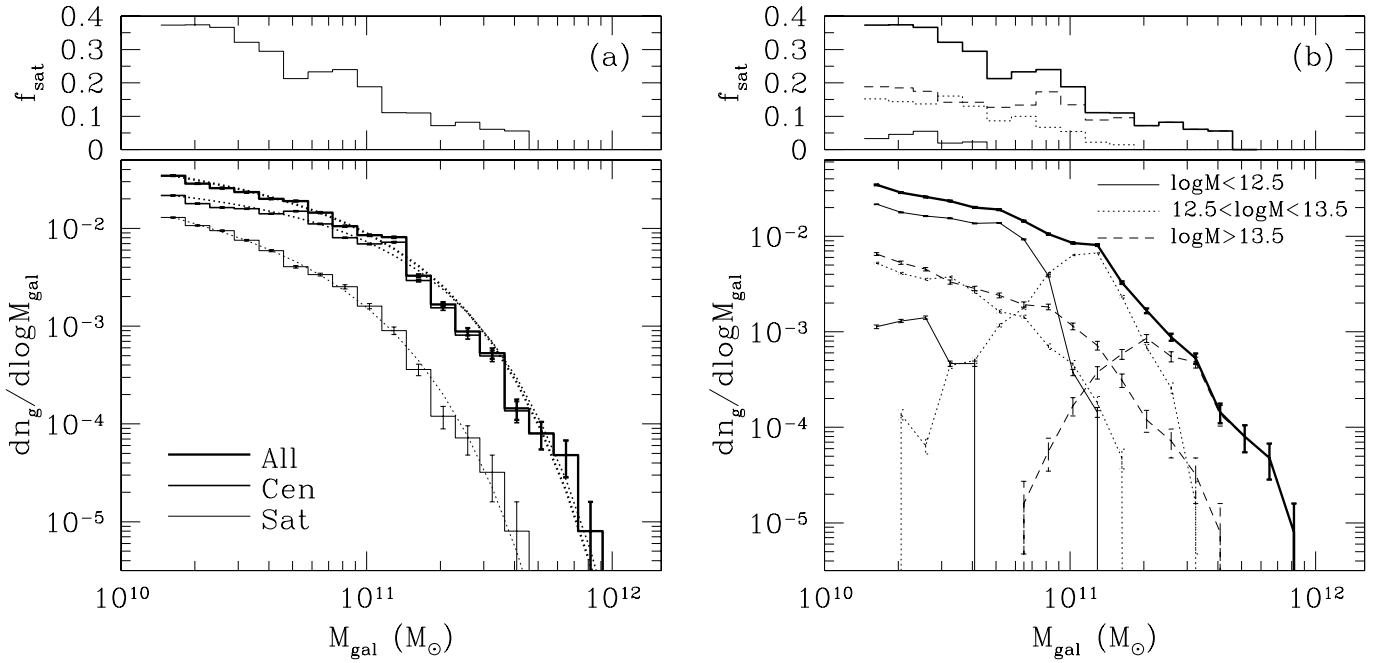

FIG. 12.—Contributions of central and satellite galaxies to the overall galaxy MF of the SA model. In the bottom left panel, the thick histogram shows the total MF (in units of h^3 Mpc^{-3} dex^{-1}), while thinner histograms show the central and satellite contributions. Dotted curves show Schechter function fits. The top left panel shows the fraction of satellite galaxies in each galaxy mass bin. The right panels show the separate contributions from different mass ranges, with curves to the right showing central galaxy contributions and curves to the left showing satellite contributions.

galaxy samples incomplete at fainter magnitudes. The central galaxy hump is much broader than in the CMF, and it therefore merges more continuously into the total CLF, although it is still visible in the low-mass histograms. While the total CLF roughly resembles a Schechter function with a faint-end slope that changes steadily with halo mass, our results suggest that a better model for empirical fitting would be a suitably generalized version of the one used in Figure 8, with parameterized forms of $\langle N_{\rm sat}\rangle_M/\langle N_{\rm cen}\rangle_M$ and $\mu(M)$, and freedom (perhaps mass dependent) added to the width σ of the central galaxy CLF and the Schechter parameters of the satellite CLF. Yang et al. (2003) considered a model of this form in their first paper on the CLF and showed that it yields a good fit to the 2dFGRS data, although they have mostly adopted the Schechter form in their later work.

How does separating central and satellite galaxies change our view of the overall galaxy MF and hence the LF? Figure 12*a* shows the galaxy baryonic MF of the SA model and the individual contributions of central and satellite galaxies. Both of these contributions, as well as the total MF, can be reasonably well fitted by Schechter functions, even though the CMF at fixed halo mass cannot. Central galaxies dominate the total MF at every mass; the satellite contribution is ~37% in our lowest mass bin (corresponding to luminosity $\sim 0.18L_*$) and $\lesssim$18% for $M_{\rm gal} \gtrsim 10^{11}\ M_\odot$ ($L \gtrsim 1.20L_*$). The dominance of central galaxies is a consequence of the large gap between the minimum halo mass $M_{\rm min}$ for hosting a central galaxy and the mass $M_1 \approx 18M_{\rm min}$ required to host a satellite of the same mass. The more massive halos are rarer, especially in the exponential cutoff region of the halo MF, so even though a massive halo can host multiple satellites, the central galaxies of abundant, lower mass halos dominate by number. Figure 12*b* amplifies this point by dividing the MF into contributions from different halo mass ranges. In observational terms, most galaxies are found in group environments, but these groups (at least if defined at an overdensity ~200) typically have a central galaxy that is substantially brighter than its neighbors. If one chooses a random galaxy of a given luminosity (especially with $L > L_*$), then it is more likely to be the dominant galaxy of its own group rather than the satellite of a more luminous system.

Our results on the decomposition of the MF are in agreement with those found for the LF by Benson et al. (2003b; see their Fig. 3). They also find that, except for very low mass halos (where the photoionization suppression of galaxy formation becomes important), the faint-end slope of the CLF is steeper than that of the overall LF, $\alpha_{\rm S} \sim -1.5$ versus $\alpha_{\rm S} \sim -1.2$. The shallower faint-end slope of the overall LF reflects the contribution of central galaxies. Our results for the galaxy MF are also similar to results found by other authors for the MF of dark matter subhalos in high-resolution *N*-body simulations. Vale & Ostriker (2004), based on the simulations of Weller et al. (2004), propose that the satellite subhalo MF in a given halo can be fitted by a Schechter function with low-mass slope $\alpha_{\rm S} \sim -1.91$, which is consistent with the power-law slope of roughly -2 found by De Lucia et al. (2004). Compared with their results, we find a shallower low-mass slope ($\alpha_{\rm S} \sim -1.5$), probably because of the difference between dark matter and baryonic mass. Figure 1 of Vale & Ostriker (2004) shows the overall MFs of (satellite) subhalos and parent halos, which conveys similar information to the bottom panel of our Figure 12*a*.

4.2. *Mass Distribution of Host Halos for Galaxies at Fixed Mass/Luminosity*

Orthogonal to the CMF/CLF is the conditional mass distribution of *halos* hosting galaxies at fixed baryonic mass or luminosity. This conditional distribution is just the product of the halo MF and the mean occupation function of such galaxies (see Fig. 6 for the mean occupation function for a sample of galaxies in a bin of baryonic mass). This quantity is relevant to the interpretation of galaxy-galaxy lensing measurements or to other methods of estimating the average mass distribution surrounding galaxies of specified properties.

Figure 13 shows the SA model predictions for the conditional mass distribution of host halos in narrow bins of galaxy baryonic

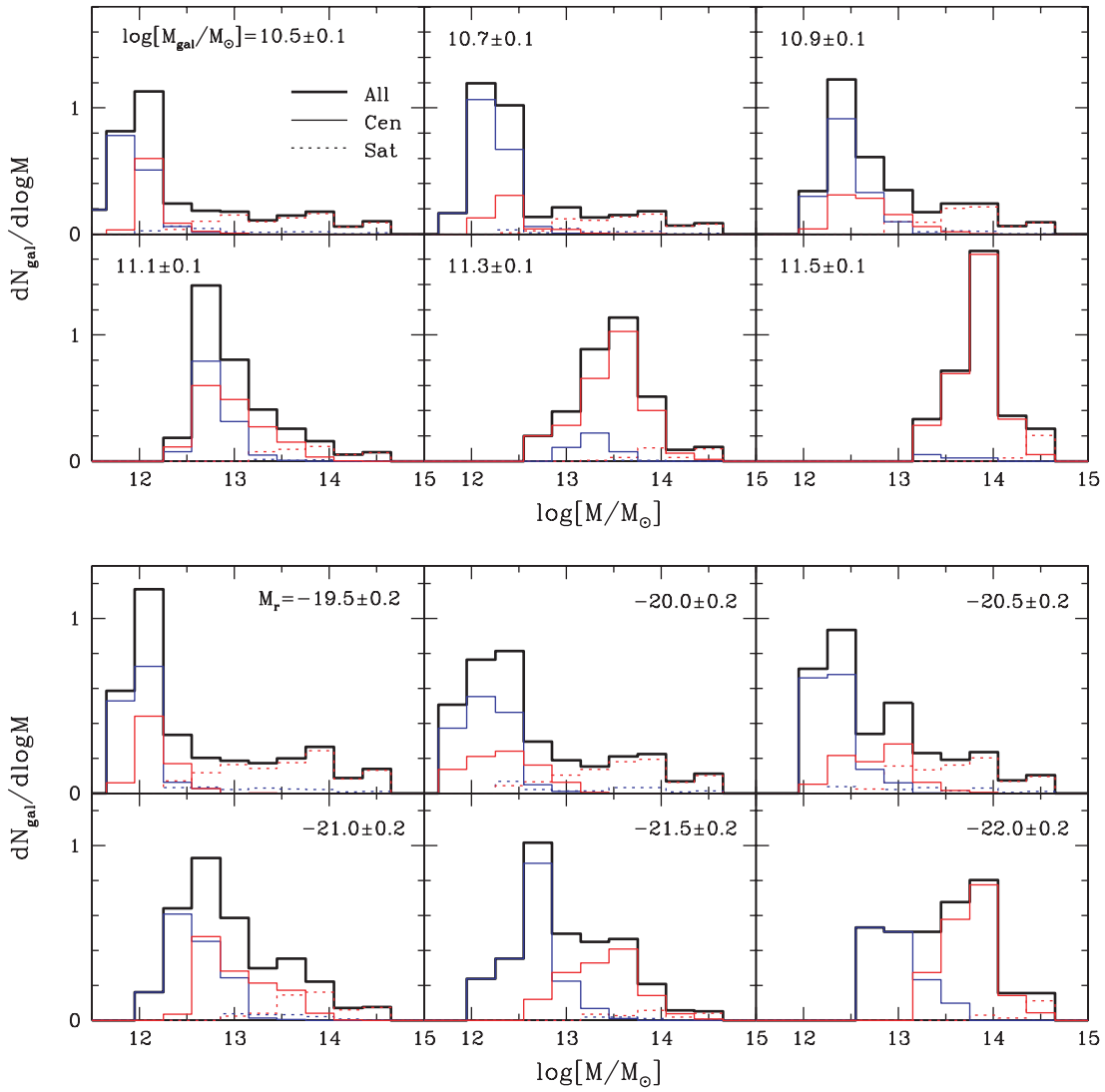

FIG. 13.—Mass distribution of host halos for galaxies at fixed baryonic mass (*top*) and at fixed *r*-band luminosity (*bottom*) predicted by the SA model. The label in each panel marks the range of log ($M_{\rm gal}/M_\odot$) or M_r. Contributions from central and satellite galaxies are represented by thin solid and dotted histograms, respectively. They are further divided into red and blue histograms for red and blue galaxies with a color cut at $g - r = 0.734$. Distributions are normalized such that the total area under the thick solid histogram (for all galaxies) is unity.

mass (*top panels*) and *r*-band luminosity (*bottom panels*). The distribution is separated into contributions from central and satellite galaxies, as well as blue and red galaxies. The division of blue and red galaxies is based on a color cut at $g - r = 0.734$, which results in roughly equal numbers of blue and red galaxies. For galaxies of fixed baryonic mass, the host halos span a wide range of masses, with a peak at low masses contributed by central galaxies and a fairly flat tail to high masses from satellites. The host halos of red galaxies, whether central or satellite, tend to be more massive than the host halos of blue galaxies. As the galaxy baryonic mass increases, the contributions to the conditional halo mass distribution from satellite galaxies and from blue galaxies decrease, as expected.

As a function of galaxy luminosity, the conditional mass distribution of host halos has a trend similar to that with baryonic mass. However, the scatter between galaxy luminosity and baryonic mass broadens the distribution, especially the central galaxy peak, and it leads to greater separation between the typical halo masses of blue and red central galaxies. This result is analogous to the finding of Mandelbaum et al. (2005), who assign luminosities to subhalos identified in high-resolution dissipationless simulations and show that scatter between subhalo circular velocity and galaxy luminosity makes the halo mass distribution at fixed luminosity wider. The width and asymmetry of the conditional mass distributions imply, first, that the mean halo mass estimated for a given type of galaxy can be very different from the typical mass of an *isolated* halo that hosts such a galaxy (i.e., from the location of the central galaxy peak in these histograms). Second, it suggests that interpretation of galaxy-galaxy lensing measurements will be simpler if one can classify galaxies by baryonic mass instead of luminosity and if one can separate isolated galaxies from those that are likely to be satellites in more massive halos.

5. SUMMARY AND DISCUSSION

A number of previous studies have investigated the halo occupation function $P(N|M)$ predicted by semianalytic models, hydrodynamic simulations, and high-resolution *N*-body simulations (Kauffmann et al. 1997; Benson et al. 2000; Seljak 2000; Scoccimarro et al. 2001; Sheth & Diaferio 2001; White et al. 2001; Yoshikawa et al. 2001; Cooray & Sheth 2002; Guzik & Seljak 2002; Scranton 2003; B03; K04), showing good agreement

in the qualitative features predicted by the different methods. Here we have followed the lead of Guzik & Seljak (2002) and K04 by separating the contributions of central and satellite galaxies to $P(N|M)$. When we consider galaxy samples defined by thresholds in baryonic mass, analogous to observational samples defined by thresholds in luminosity, our results for the Cole et al. (2000) SA model and the Davé et al. (2002) SPH simulation are similar to those found by K04 for subhalos in high-resolution N-body simulations. In particular, the separation of central and satellite galaxies naturally explains the general shape of the mean occupation function $\langle N\rangle_M$ and the transition from sub-Poisson fluctuations in $P(N|M)$ at low N to roughly Poisson fluctuations at high N. The correlation between halo mass and central galaxy mass is tight, so $\langle N\rangle_M$ for central galaxies rises sharply at a threshold mass $M_{\rm min}$ and can be approximated by a step function. A halo has either zero or one central galaxies, so the width of $P(N_{\rm cen}|M)$ is sub-Poisson by definition. The mean occupation of satellites has a roughly power-law form, $\langle N_{\rm sat}\rangle_M \approx (M/M_1)^\alpha$ with $\alpha \approx 1$, and the fluctuations of $P(N_{\rm sat}|M)$ about the mean are close to Poisson. However, a halo must be $\sim 10M_{\rm min}$–$20M_{\rm min}$ before it hosts on average one satellite galaxy above the baryonic threshold; in the mass decade above $M_{\rm min}$, a larger halo typically hosts a larger central galaxy instead of multiple galaxies above the threshold (B03). In the halo mass regime where central galaxies make a significant contribution to the total galaxy counts, the mean occupation rises slowly, and the width of the total $P(N|M)$ is substantially narrower than a Poisson distribution.

When a sample of galaxies above a baryonic mass threshold is divided in two on the basis of stellar population age, the HODs of young and old galaxies are markedly different. Halos near the cutoff mass $M_{\rm min}$ tend to host young central galaxies, while more massive halos host old central galaxies. The mean occupation function of young galaxies exhibits a local minimum at halo masses $M \sim 10M_{\rm min}$, where halos are too massive to have a young central galaxy but not massive enough to have satellites. The old galaxy population has a monotonically rising $\langle N\rangle_M$, and the number of old satellites in massive halos is larger than the number of young satellites. The statistics of old and young satellites are consistent with Poisson distributions about their respective means, uncorrelated with the age of the central galaxy. If one starts with a sample of galaxies in a bin of baryonic mass (instead of a mass-threshold sample), then the halos that host central galaxies occupy a relatively narrow mass range, and the ratio of old central galaxies to young central galaxies depends on the galaxy mass bin. For a low-mass (or low-luminosity) sample, most central galaxies are young, and the sample's older galaxies are therefore satellites in higher mass halos. With a higher mass (or luminosity) sample, more of the old galaxies are central objects. These differences in the HODs of young and old galaxies naturally explain much of the observed dependence of galaxy clustering on color, spectral type, and morphology, with red/early-type galaxies exhibiting stronger correlations and residing in higher density environments.

Two striking features of the theoretically predicted HODs are the near-constant ratio and large gap between the minimum host halo mass $M_{\rm min}$ and the mass M_1 at which an average halo hosts one satellite above the baryonic mass threshold. The ratio is $M_1/M_{\rm min} \approx 14$ for the SPH simulation and $M_1/M_{\rm min} \approx 18$ for the SA model, over a wide range in galaxy mass. The large value of $M_1/M_{\rm min}$ accounts for the extended plateau in $\langle N\rangle_M$ at low halo masses, and it has a number of important consequences. First, the CMF at fixed halo mass is not well described by a Schechter function. Central galaxies produce a "bump" in the CMF at high galaxy masses that rises well above the Schechter extrapolation of the satellite population. Scatter in the relation between baryonic mass and luminosity smears out but does not eliminate this bump in the CLF. Second, at any given galaxy mass or luminosity, the total MF/LF is dominated by central galaxies because the massive halos that can host multiple satellites are much less abundant than halos with $M \sim M_{\rm min}$. The satellite fraction is larger at low masses/luminosities, where halos with $M \sim M_1$ are not yet on the exponential tail of the halo MF.

The large value of $M_1/M_{\rm min}$ also plays a key role in shaping the galaxy correlation function $\xi(r)$. As first emphasized by Benson et al. (2000) and confirmed in many subsequent investigations (e.g., Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002), the sub-Poisson width of $P(N|M)$ at low masses is crucial to reproducing the observed, roughly power-law form of $\xi(r)$ because it enables low-mass halos to host large numbers of galaxies without hosting large numbers of small separation pairs. As shown here and in K04, this sub-Poisson width is a consequence of the nearest-integer statistics of central galaxy occupations, and it holds over an extended range of halo masses because $M_1/M_{\rm min}$ is large. The roughly constant value of $M_1/M_{\rm min}$ enables galaxies with a wide range of luminosity to exhibit power-law correlation functions. In detail, HOD models can also explain the deviations of $\xi(r)$ from a power law found in high-precision measurements, as shown by Zehavi et al. (2004).

The nearly constant value of $M_1/M_{\rm min}$ and the nearly Poisson form of $P(N_{\rm sat}|M)$ at fixed M are pleasingly simple results, and in qualitative terms they seem intuitively sensible. However, we do not have a quantitative explanation for either one of them. An important clue is that K04 find essentially the same results for samples of dark matter subhalos defined by circular velocity thresholds in purely gravitational simulations, which suggests that they emerge mainly from dark matter dynamics and that gas physics and star formation serve to light up the dark matter substructures in a fairly straightforward fashion. The $M_1/M_{\rm min}$ scaling and form of $P(N_{\rm sat}|M)$ should then be determined mainly by the statistics of halo merger histories, although they will also be affected by tidal disruption of subhalos and galaxies and by mergers induced by dynamical friction. Analytic methods have been developed to model these processes (Bond et al. 1991; Bower 1991; Lacey & Cole 1993; Bullock et al. 2000; Taylor & Babul 2004). A recent theoretical study of the formation and evolution of substructures using the methods of Zentner et al. (2005) shows that tidal stripping and disruption are the most important mechanisms shaping the HOD, with dynamical friction subdominant (A. Kravtsov 2004, private communication). Further investigation along these lines may lead to a more fundamental theory of the galaxy HOD and generic predictions for its dependence on redshift and on cosmological parameters.

Recent observational analyses show qualitative and to some degree quantitative agreement with many of our theoretical predictions. Zehavi et al. (2005) show that the projected correlation functions of SDSS galaxy samples with a wide range of luminosity thresholds can be well fitted by the three-parameter ($M_{\rm min}$, M_1, α) HOD models described in § 3.1 for "concordance" values of cosmological parameters. The fitted values of $M_{\rm min}$ and M_1 show the predicted linear scaling relation, with $M_1 \approx 23M_{\rm min}$, and the dependence of $M_{\rm min}$ on luminosity resembles the predicted dependence on galaxy baryonic mass (compare Fig. 18 of Zehavi et al. [2005] to our Fig. 7). The measured ratio $M_1/M_{\rm min} \approx 23$ is higher than the value of 14 predicted by the SPH simulation, probably for the same reasons that the SPH simulation fails to match the observed galaxy LF. The SA model is designed to match the LF fairly well, and its predicted ratio of

18 is closer to the measured value (the factor $\sim$25 in the dissipationless simulations of K04 is closer still). The remaining discrepancy could largely reflect the difference between a luminosity-threshold and mass-threshold galaxy sample, since scatter in stellar mass-to-light ratios allows some galaxies of the luminosity-threshold sample to occupy lower mass halos, reducing $M_{\rm min}$ and simultaneously raising M_1 to keep the number density fixed. The difference in cosmological models (we assume higher Ω_m and lower σ_8 than Zehavi et al. 2005) and errors in deriving $M_{\rm min}$ and M_1 from the clustering data with a simplified HOD model can also contribute to the discrepancy.

Zehavi et al. (2005) also find that the clustering of red and blue subsamples can be well described by HOD models like those described in § 3.2 for old and young galaxy populations. Fits to the blue fraction of central and satellite galaxies as a function of halo mass show good qualitative agreement with our predictions (compare Fig. 22 of Zehavi et al. [2005] to our Fig. 4). Van den Bosch et al. (2003b) carry out CLF fits to the 2dFGRS data, and their inferred dependence of late-type galaxy fraction on halo mass is qualitatively similar to our prediction for blue galaxy fraction. Except for the lowest luminosity bin, their inferred mean occupation function for luminosity bin galaxy samples (see their Fig. 10) does not show a clear bump like that seen in our Figure 6, probably because of their assumed Schechter function form of the CLF (see below).

Zehavi et al. (2005) use their HOD fits as a function of luminosity to infer the CLF in different ranges of halo mass. Given the good agreement with the predicted HOD results, it is not surprising that their inferred CLFs resemble our predicted CMFs (compare their Fig. 21 to our Fig. 8). In particular, the large $M_1/M_{\rm min}$ ratio leads to a central galaxy "bump" in the LF of intermediate-mass halos. Indeed, these bumps are more prominent than those seen in our CLF predictions (Fig. 11), perhaps indicating that the SA model used here produces too much scatter in the baryonic mass-luminosity relation for central galaxies. Hansen et al. (2005) have measured the CLF directly in an SDSS cluster catalog (Bahcall et al. 2003) and find that bumps caused by the brightest cluster galaxies emerge in the CLFs of low-richness clusters. They also find that, as a result of the significant contributions from the brightest cluster galaxies, a Schechter function is not a good fit for the CLFs in many of the cluster richness bins. Eke et al. (2004) look for such central galaxy bumps in the CLF of groups and clusters in the 2dFGRS Percolation-Inferred Galaxy Group (2PIGG) catalog and do not find them, but they demonstrate using SA mock catalogs that measurement errors in the group masses would wash them out. Using a halo-based group finder (Yang et al. 2005b), Yang et al. (2005a) identify groups from the 2dFGRS and measure the HOD as a function of halo mass. They find a tight correlation between the mean luminosity of central galaxies and the halo mass, qualitatively similar to that shown in the middle panel of Figure 9 of this paper. Although they obtain reasonably good fits to the CLF over the halo mass range $13 \lesssim \log [M/(h^{-1}\ M_\odot)] \lesssim 14.5$ with a Schechter function, the bright end of the CLF in halos of $\log [M/(h^{-1}\ M_\odot)] \sim 13$ starts to show a clear enhancement caused by the central galaxies. Lin et al. (2004) and Lin & Mohr (2004) find that the K-band LF of Two Micron All Sky Survey (2MASS) cluster galaxies can be fitted by a Schechter function if the brightest cluster galaxies are excluded. The brightest cluster galaxies seem to follow a different distribution and become less important in the total cluster light as the cluster mass increases. They also find that the distribution of satellite numbers at a fixed estimated mass is roughly Poisson. Miles et al. (2004) study a sample of 25 groups drawn from the Group Evolution Multi-Wavelength Study (GEMS), and they find prominent bumps at the bright end of the B-band and R-band LFs of the group galaxies, which are very likely caused by central galaxies. The isolated, luminous, X-ray–bright elliptical galaxies known as "fossil groups" (Vikhlinin et al. 1999; Jones et al. 2003) may represent an extreme example of the central galaxy bump in intermediate-mass halos. All of these results are in good qualitative agreement with the predictions here, although more careful replication of the observational selection would be needed for quantitative comparison.

In our decomposed CMFs and CLFs (Figs. 8 and 11), the satellite contributions are fairly well described by a truncated Schechter function with constant faint-end slope $\alpha_{\rm S}$, but the shape of the total CMF/CLF changes with halo mass because of the changing relative amplitude of central and satellite contributions. Yang et al. (2003) infer the CLF from 2dFGRS data *assuming* a Schechter form at each halo mass, and they find a steadily changing $\alpha_{\rm S}$ that becomes positive in low-mass halos, which might plausibly be a consequence of describing the roughly Gaussian LF of the central galaxies with a Schechter function. The Schechter+Gaussian model described in § 4 should be a better parameterized form for fitting observational data. It is not clear how forcing a Schechter function fit might influence the *cosmological* conclusions from the CLF method (Yang et al. 2003; van den Bosch et al. 2003a), but we would expect some impact. For example, a Schechter-based CLF fit that gives the correct halo occupations of L_* galaxies will give incorrect occupations, and hence incorrect bias factors, for $2L_*$ galaxies and thus alter the inferred amplitude of mass clustering. Yang et al. (2003) investigate an alternative parameterization for the CLF similar to that advocated here: they assume the central galaxy CLF to be a lognormal function for low-mass halos, with the width being a free parameter. With this alternative parameterization and a concordance cosmological model, they find that the inferred width in the central CLF is more or less consistent with the scatter in the Tully-Fisher relation, and their fits to the global LF and the luminosity dependence of the correlation length do not change much (see their Fig. 7). The resulting best-fit HOD has noticeable changes, especially at the low halo mass end. This test suggests that the assumed CLF form can have a significant impact on inferred halo galaxy populations but may have little effect on cosmological constraints derived from the galaxy LF and luminosity dependence of galaxy clustering strength.

The qualitative agreement between our predictions and the results of Zehavi et al. (2005) suggests that current theoretical models of galaxy formation are on largely the right track. There should be interesting progress in the near future as more clustering measurements are incorporated into the observational determinations of the HOD. For example, the group multiplicity function places fairly direct constraints on $P(N|M)$ at high halo masses (Peacock & Smith 2000; Marinoni & Hudson 2002; Berlind & Weinberg 2002; Kochanek et al. 2003), and void statistics and galaxy scaling relations are sensitive to behavior in the cutoff region near $M_{\rm min}$ (Berlind & Weinberg 2002), so combinations of these measures with the correlation function can constrain more flexible parameterizations of the HOD. Our five-parameter model provides a near-perfect description of the theoretical results and offers a good starting point for fitting observations, but ultimately one would like to test the ingredients of this model by allowing more general forms for $\langle N \rangle_M$, non-Poisson statistics for satellites, and so forth. Direct measurements of galaxy profiles in groups and clusters can refine the standard assumption (supported by the simulation results of B03) that satellite galaxies trace the underlying dark matter distribution

within halos. Detailed empirical determinations of the HOD for different classes of galaxies will test theoretical models of galaxy formation in greater detail than previously possible, and they may provide guidance to additional physics that should be incorporated in these models.

A still more ambitious goal is to constrain cosmological parameters and the HOD simultaneously using observed galaxy clustering, either on its own or in combination with other observables. Based on the CLF approach, van den Bosch et al. (2003a) have demonstrated that the galaxy LF, luminosity dependence of the galaxy correlation length, and mass-to-light ratios of galaxy clusters can impose interesting constraints on cosmological parameters, either on their own or in combination with cosmic microwave background (CMB) data. Abazajian et al. (2005) combine CMB data with the projected correlation function of SDSS galaxies brighter than $M_r = -21$, adopting parameterized HOD models like those in § 3.1. Cosmological constraints from this combination of measurements are as tight as those found by combining CMB data with the large-scale galaxy power spectrum (e.g., Percival et al. 2002; Spergel et al. 2003; Tegmark et al. 2004), where instead of an HOD model one assumes that galaxy bias is scale independent in the linear and near-linear regime. Tinker et al. (2005) show that the SDSS correlation function constraints and observational estimates of cluster mass-to-light ratios imply that the values of Ω_m and/or σ_8 are lower than the traditionally adopted "concordance" values of 0.3 and 0.9, in agreement with the van den Bosch et al. (2003a) conclusion based on CLF analysis of the 2dFGRS. The HOD approach will become more powerful as more clustering measurements are incorporated, especially observables like redshift space distortions and galaxy-galaxy lensing that are directly sensitive to mass. Even with a very flexible HOD parameterization, galaxy clustering measurements alone can pin down cosmological parameters like Ω_m and σ_8 with reasonable precision (Z. Zheng and D. Weinberg 2005, in preparation). In concert with CMB anisotropy, the Lyα forest, Type Ia supernovae, and other observables that probe different scales and redshifts, careful modeling of low-redshift galaxy clustering allows sharpened tests of the nature of dark energy, the masses of neutrinos, and the spectrum of density and gravity wave fluctuations that emerged from the early universe.

We thank Andrey Kravtsov and Idit Zehavi for helpful conversations. We thank Xiaohu Yang, Frank van den Bosch, and Houjun Mo for useful comments on an earlier draft of the paper. We also thank the referee, Adi Nusser, for constructive comments. We thank Chris Kochanek for suggesting the analysis that led to Figure 13. Z. Z. was partly supported by a Presidential Fellowship from the Graduate School of Ohio State University. Z. Z. also acknowledges the support of NASA through Hubble Fellowship grant HF-01181.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. This research was also supported by NSF grants AST 00-98584, AST 02-05969, and AST 04-07125 and NASA ATP grant NAGS-13308. C. G. L. acknowledges support from the PPARC rolling grant for extragalactic astronomy and cosmology at Durham. A. J. B. and C. M. B. are supported by Royal Society University Research Fellowships.

REFERENCES

Abazajian, K., et al. 2005, ApJ, 625, 613
Avila-Reese, V., Firmani, C., & Hernandez, X. 1998, ApJ, 505, 37
Bahcall, N. A., et al. 2003, ApJS, 148, 243
Benson, A. J., Bower, R. G., Frenk, C. S., Lacey, C. G., Baugh, C. M., & Cole, S. 2003a, ApJ, 599, 38
Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. 2000, MNRAS, 311, 793
Benson, A. J., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2003b, MNRAS, 343, 679
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Berlind, A. A., et al. 2003, ApJ, 593, 1 (B03)
Blanton, M. R., et al. 2003, ApJ, 592, 819
Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440
Bower, R. 1991, MNRAS, 248, 332
Bullock, J. S., Kravtsov, A. V., & Weinberg, D. H. 2000, ApJ, 539, 517
Cen, R., & Ostriker, J. P. 1992, ApJ, 399, L113
Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E. 1994, MNRAS, 271, 781
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168
Colless, M., et al. 2001, MNRAS, 328, 1039
Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
Davé, R., Katz, N., & Weinberg, D. H. 2002, ApJ, 579, 23
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
De Lucia, G., Kauffmann, G., Springel, V., White, S. D. M., Lanzoni, B., Stoehr, F., Tormen, G., & Yoshida, N. 2004, MNRAS, 348, 333
Eke, V. R., et al. 2004, MNRAS, 355, 769
Evrard, A. E., Summers, F. J., & Davis, M. 1994, ApJ, 422, 11
Guzik, J., & Seljak, U. 2002, MNRAS, 335, 311
Hansen, S., et al. 2005, ApJ, 633, 122
Jones, L. R., Ponman, T. J., Horton, A., Babul, A., Ebeling, H., & Burke, D. J. 2003, MNRAS, 343, 627
Katz, N., Hernquist, L., & Weinberg, D. H. 1992, ApJ, 399, L109
Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188
Kauffmann, G., Nusser, A., & Steinmetz, M. 1997, MNRAS, 286, 795
Kauffmann, G., White, S. D. M., & Guideroni, B. 1993, MNRAS, 264, 201
Kauffmann, G., et al. 2003, MNRAS, 341, 33
Keres, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, in press (astro-ph/0407095)
Kochanek, C. S., White, M., Huchra, J., Macri, L., Jarrett, T. H., Schneider, S. E., & Mader, J. 2003, ApJ, 585, 161
Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottloeber, S., Allgood, B., & Primack, J. R. 2004, ApJ, 609, 35 (K04)
Lacey, C. G., & Cole, S. 1993, MNRAS, 262, 627
Lin, Y., & Mohr, J. J. 2004, ApJ, 617, 879
Lin, Y., Mohr, J. J., & Stanford, S. A. 2004, ApJ, 610, 745
Ma, C., & Fry, J. N. 2000, ApJ, 543, 503
Madgwick, D. S., et al. 2003, MNRAS, 344, 847
Maller, A. H., & Bullock, J. S. 2004, MNRAS, 355, 694
Mandelbaum, R., Tasitsiomi, A., Seljak, U., Kravtsov, A. V., & Wechsler, R. H. 2005, MNRAS, 362, 1451
Marinoni, C., & Hudson, M. J. 2002, ApJ, 569, 101
Miles, T. A., Raychaudhury, S., Forbes, D. A., Goudfrooij, P., Ponman, T. J., & Kozhurina-Platais, V. 2004, MNRAS, 355, 785
Mo, H. J., Yang, X. H., van den Bosch, F. C., & Jing, Y. P. 2004, MNRAS, 349, 205
Murali, C., Katz, N., Hernquist, L., Weinberg, D. H., & Davé, R. 2002, ApJ, 571, 1
Norberg, P., et al. 2002, MNRAS, 332, 827
Peacock, J. A., & Smith, R. E. 2000, MNRAS, 318, 1144
Pearce, F. R., et al. 1999, ApJ, 521, L99
Percival, W., et al. 2002, MNRAS, 337, 1068
Schechter, P. 1976, ApJ, 203, 297
Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20
Scranton, R. 2003, MNRAS, 339, 410
Seljak, U. 2000, MNRAS, 318, 203
Seljak, U., et al. 2005, Phys. Rev. D, 71, 043511
Sheth, R. K., & Diaferio, A. 2001, MNRAS, 322, 901
Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087
Spergel, D. N., et al. 2003, ApJS, 148, 175
Tasitsiomi, A., Kravtsov, A. V., Wechsler, R. H., & Primack, J. R. 2004, ApJ, 614, 533
Taylor, J. E., & Babul, A. 2004, MNRAS, 348, 811
Tegmark, M., et al. 2004, Phys. Rev. D, 69, 103501
Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41

Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661
Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189
van den Bosch, F. C., Mo, H. J., & Yang, X. H. 2003a, MNRAS, 345, 923
———. 2003b, MNRAS, 340, 771
van Kampen, E., Jimenez, R., & Peacock, J. A. 1999, MNRAS, 310, 43
Vikhlinin, A., McNamara, B. R., Hornstrup, A., Quintana, H., Forman, W., Jones, C., & Way, M. 1999, ApJ, 520, L1
Weinberg, D. H., Davé, R., Katz, N., & Hernquist, L. 2004, ApJ, 601, 1
Weller, J., Ostriker, J. P., & Bode, P. 2004, MNRAS, in press (astro-ph/0405445)
White, M., Hernquist, L., & Springel, V. 2001, ApJ, 550, L129
White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52
Yan, R., Madgwick, D. S., & White, M. 2003, ApJ, 598, 848
Yan, R., White, M., & Coil, A. L. 2004, ApJ, 607, 739
Yang, X. H., Mo, H. J., Jing, Y. P., & van den Bosch, F. C. 2005a, MNRAS, 358, 217
Yang, X. H., Mo, H. J., Jing, Y. P., van den Bosch, F. C., & Chu, Y. 2004, MNRAS, 350, 1153
Yang, X. H., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057
Yang, X. H., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005b, MNRAS, 356, 1293
York, D. G., et al. 2000, AJ, 120, 1579
Yoshikawa, K., Taruya, A., Jing, Y. P., & Suto, Y. 2001, ApJ, 558, 520
Zehavi, I., et al. 2002, ApJ, 571, 172
———. 2004, ApJ, 608, 16
———. 2005, ApJ, 630, 1
Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H. 2005, ApJ, 624, 505