CONSTRAINTS ON THE RELATIONSHIP BETWEEN STELLAR MASS AND HALO MASS AT LOW AND HIGH REDSHIFT

BENJAMIN P. MOSTER\(^1\), RACHEL S. SOMERVILLE\(^1,2\), CHRISTIAN MAULBETSCHI\(^3\), FRANK C. VAN DEN BOSCH\(^1\), ANDREA V. MACC\O\(^1\), THORSTEN NAAB\(^4\), AND LUDWIG OSER\(^5\)

The Astrophysical Journal, submitted

ABSTRACT

We use a statistical approach to determine the relationship between the stellar masses of galaxies and the masses of the dark matter halos in which they reside. We obtain a parameterized stellar-to-halo mass (SHM) relation by populating halos and subhalos in an N-body simulation with galaxies and requiring that the observed stellar mass function be reproduced. We find good agreement with constraints from galaxy-galaxy lensing and predictions of semi-analytic models. Using this mapping, and the positions of the halos and subhalos obtained from the simulation, we find that our model predictions for the galaxy two-point correlation function (CF) as a function of stellar mass are in excellent agreement with the observed clustering properties in the SDSS at \(z = 0\).

We show that the clustering data do not provide additional strong constraints on the SHM function and conclude that our model can therefore predict clustering as a function of stellar mass. We compute the conditional mass function, which yields the average number of galaxies with stellar masses in the range \(m \pm dm/2\) that reside in a halo of mass \(M\). We study the redshift dependence of the SHM relation and show that, for low mass halos, the SHM ratio is lower at higher redshift. The derived SHM relation is used to predict the stellar mass dependent galaxy CF and bias at high redshift. Our model predicts that not only are massive galaxies more biased than low mass ones at all redshifts, but the bias increases more rapidly with increasing redshift for massive galaxies than for low mass ones. We present convenient fitting functions for the SHM relation as a function of redshift, the conditional mass function, and the bias as a function of stellar mass and redshift.

Subject headings: cosmology: theory — dark matter — galaxies: clusters: general — galaxies: evolution — galaxies: halos — galaxies: high-redshift — galaxies: statistics — galaxies: stellar content — large-scale structure of universe

1. INTRODUCTION

In the standard Cold Dark Matter (CDM) paradigm, the formation of galaxies is driven by the growth of the large-scale structure of the Universe and the formation of dark matter halos. Galaxies form by the cooling and condensation of gas in the centers of the potential wells of extended virialized dark matter halos (White & Rees 1978; Fall & Efstathiou 1980; Blumenthal et al. 1984). In this picture, galaxy properties, such as luminosity or stellar mass, are expected to be tightly coupled to the depth of the halo potential and thus to the halo mass.

There are various different approaches to link the properties of galaxies to those of their halos. A first method attempts to derive the halo properties from the properties of its galaxy population using e.g. galaxy kinematics (Erickson et al. 1987; Zuritsky et al. 1993; Carlberg et al. 1996; More et al. 2009a,b) or gravitational lensing (Mandelbaum et al. 2005, 2006; Cacciato et al. 2008), or X-ray studies (Lin et al. 2003; Lin & Mohr 2004).

A second approach is to attempt to model the physics that shapes galaxy formation \textit{ab initio} using either large numerical simulations including both gas and dark matter (Katz et al. 1996; Springel & Hernquist 2003) or semi-analytic models (SAMs) of galaxy formation (e.g. Kauffmann et al. 1993; Cole et al. 1994; Somerville & Primack 1999). In “hybrid” SAMs (e.g. Croton et al. 2006, Bower et al. 2006), dark matter “merger trees” are extracted from a dark matter only N-body simulation, and gas processes are treated with semi-analytic recipes. An advantage of this method is that high-resolution N-body simulations can track the evolution of individual subhalos (Klypin et al. 1999; Springel et al. 2001) and thus provide the precise positions and velocities of galaxies within a halo. However, many of the physical processes involved in galaxy formation (such as star formation and various kinds of feedback) are still not well understood, and in many cases simulations are not able to reproduce observed quantities with high accuracy.

With the accumulation of data from large galaxy surveys over the last decade, a third method has been developed, which links galaxies to halos using a statistical approach. The Halo Occupation Distribution (HOD) formalism specifies the probability distribution for a halo of mass \(M\) to harbour \(N\) galaxies with certain intrinsic properties, such as luminosity, color, or type (e.g. Peacock & Smith 2000; Seljak 2000; White 2001; Berlind & Weinberg 2002). More complex formulations of this kind of modelling, such as the conditional luminosity function (CLF) formalism (Yang et al. 2003; van den Bosch et al. 2003; Yang et al. 2004) have extended the HOD approach. These methods have the advantage that they do not rely on assumptions about the (poorly understood) physical processes that drive galaxy formation. In this way, it is possible to constrain the relationship between galaxy and halo properties (and thus, indirectly, the underlying physics), and to construct mock catalogs that reproduce in detail a desired observational quantity (such as the luminosity function).

One disadvantage of the classical HOD approach was that one had to make assumptions about the distribution of positions...
and velocities of galaxies within their host halos. In addition, the results of the HOD modelling can be difficult to interpret in terms of the underlying physics of galaxy formation.

In recent years, HOD models have been introduced that make use of information about the positions, velocities and masses of halos and subhalos extracted from a dissipation-less N-body simulation. The (sub)halo mass is then empirically linked to galaxy properties by requiring that a statistical observational quantity (e.g. galaxy luminosity function and/or galaxy two-point-correlation-function) is reproduced. This is either done by assuming parameterized functions to relate galaxy properties (such as luminosity) to halo mass or by assuming a non-parametric monotonic relation. It has been shown that these simple models reproduce galaxy clustering as a function of luminosity over a wide range in redshift (Vale & Ostriker 2004; Tassis & Madau 2004; Tinker et al. 2005; Vale & Ostriker 2008; Conroy et al. 2006; Shankar et al. 2006; Wang et al. 2006; Conroy & Wechsler 2008).

Observationally, it is well known that galaxy clustering is a function of spatial scale, galaxy properties (such as luminosity and type), and redshift. Luminous (massive) galaxies are more strongly clustered than less luminous (less massive) galaxies (Norberg et al. 2001; 2002; Zehavi et al. 2002; 2005; Li et al. 2006). One can split the galaxy two-point correlation function (2PCF) into two separate parts: the one-halo and the two-halo terms. The one-halo term, which dominates on small scales, depends strongly on the galaxy distribution within the halo as well as the details of the HOD. The two-halo term, which dominates on scales that are much larger than a typical halo, is proportional to the auto-correlation of the halo population. In general the two terms are not expected to combine to produce a featureless power-law, but generally show a break or dip at the scale where the transition from the one-halo to the two-halo term occurs (Zehavi et al. 2004).

The extensive multi-wavelength spectrophotometric information that is now available for large numbers of galaxies allows us to estimate physical parameters of galaxies, such as stellar masses, instead of relying on observational properties such as magnitudes (Bell & de Jong 2001; Kauffmann et al. 2003; Panter et al. 2004). These estimates can even be obtained — with a proper measure of caution — for high redshift galaxies. Stellar mass estimates have been presented in the literature for galaxies up to redshifts as high as $z \sim 6$ (Yan et al. 2006; Eyles et al. 2007), and stellar mass function estimates have been presented up to $z \sim 5$ (Drorv et al. 2005; Fontana et al. 2006; Elsner et al. 2008). The goal of our paper is to develop a "Conditional Stellar Mass Function" (CMF) formalism, which is the stellar mass analog of the CLF. The CMF yields the average number of galaxies with stellar masses in the range $M+z dm$ as a function of the host halo mass $M$ and can be regarded as the stellar mass function (SMF) for halos of mass $M$. We apply this formalism at low redshift and up to the highest redshifts where reliable observational stellar mass estimates are available ($0.1 \leq z \leq 4$). In this way, we derive a relationship between dark matter halo mass and galaxy mass as a function of redshift. We present convenient parameterizations of our results.

Using the CMF derived only from constraints from the observed SMF, we compute the predicted (projected) galaxy CF at $z \sim 0$ as a function of stellar mass, and find good agreement with the observational results of Li et al. (2006). Furthermore, we show that adding the clustering constraints does not tighten the constraints on our model parameters; i.e., any model that satisfies the mass function constraints will produce the correct clustering. Based on this result, we use our redshift-dependent CMF results to predict the clustering as a function of stellar mass and redshift. To date, observational measurements of clustering as a function of stellar mass at high redshift have not yet been published. However, very soon it will be possible to test these predictions with the results from deep wide-field surveys (e.g. MUSYC, UKIDDS, etc). We again present convenient fitting functions for the galaxy bias as a function of both stellar mass and redshift. In a companion paper we will employ our estimates of galaxy bias in order to compute the "cosmic variance", the uncertainty in observational estimates of the volume density of galaxies arising from the underlying large-scale density fluctuations.

This paper is organized as follows: in section 2 we describe the N-body simulation, the halo finding algorithm that was used to obtain a halo catalogue and the treatment of "orphaned" galaxies. Section 3 specifies our model: we motivate the form of the stellar-to-halo mass (SHM) relation and constrain it by requiring that the observed SMF is reproduced. The clustering properties of galaxies are then inferred from those of the halo population. We discuss the meaning of the parameters of the SHM relation and demonstrate that clustering puts only weak constraints on them. In section 5 we introduce the CMF, which describes how halos are occupied by galaxies, and compute the occupation numbers. Section 6 gives a comparison between our results and several other models and observations. In section 7 we apply our method to higher redshifts and determine the redshift dependence of the SHM relation. We make predictions of the stellar mass dependent galaxy CF at higher redshift which we use to compute the galaxy bias. Finally, we summarize our methods and conclusions in section 8.

Throughout this paper we assume a $\Lambda$CDM cosmology with $(\Omega_m, \Omega_\Lambda, h, \sigma_8, n) = (0.26, 0.74, 0.72, 0.77, 0.95)$. We employ a Kroupa (2001) initial mass function (IMF) and convert all stellar masses to this IMF. In order to simplify the notation we will use the capital $M$ to denote dark matter halo masses and the lower case $m$ to denote galaxy stellar masses.

2. THE SIMULATION AND HALO CATALOGS

High-resolution dissipation-less N-body simulations have shown that distinct halos contain subhalos which orbit within the potential of their host halo. These subhalos were distinct halos in the past, and entered the larger halo via merging during the process of hierarchical assembly. We will refer to the galaxy at the center of a distinct halo as a central galaxy, and the galaxies within subhalos as "satellites", and we will use the term ‘halo’ to refer to the distinct halo for central galaxies and to the subhalo in which the galaxy originally formed for satellite galaxies.

Ab initio models of galaxy formation predict that the stellar mass of a galaxy is tightly correlated with the depth of the potential well of the halo in which it formed. For distinct halos, the relevant mass is the virial mass at the time of observation. Subhalos, however, lose mass while orbiting in a larger system as their outer regions are tidally stripped. Stars are centrally concentrated and more tightly bound than the dark matter, however, and so the stellar mass of a galaxy which is accreted by a larger system probably changes only slightly until most of the dark matter has been stripped off. Therefore

4 We refer to virialized halos that are not subhalos of another halo as "distinct."
the subhalo mass at the time of observation is probably not a good tracer for the potential well that shaped the galaxy properties. A better tracer is the subhalo mass at the time that it was accreted by the host halo, or its maximum mass over its history.

The population of dark matter halos used in this work is drawn from an N-body simulation run with the simulation code GADGET-2 (Springel 2005) on a SGI AltixII at the University Observatory Munich. The cosmological parameters of the simulation are chosen to match results from WMAP-3 (Spergel et al. 2007) for a flat $\Lambda$CDM cosmological model: $\Omega_m = 0.26$, $\Omega_{\Lambda} = 0.74$, $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.72$, $\sigma_8 = 0.77$ and $n = 0.95$. The initial conditions were generated using the GRAFIC software package (Bertschinger 2001). The simulation was done in a periodic box with side length 100 Mpc, and contains 512$^3$ particles with a particle mass of $2.8 \times 10^5 \text{M}_\odot$ and a force softening of 3.5 kpc.

Dark matter halos are identified in the simulation using a friends-of-friends (FoF) halo finder. Substructures inside the FoF groups are then identified using the SUBFIND code described in Springel et al. (2001). For the most massive sub-group in a FoF group the virial radius and mass are determined with a spherical overdensity criterion: the density inside a sphere centered on the most bound particle is required to be greater than or equal to the value predicted by the spherical collapse model for a top hat perturbation in a $\Lambda$CDM cosmology (Bryan & Norman 1998). As discussed above, for subhalos we use the maximum mass over its past history, which is typically the mass when the halo was last a distinct halo and did not yet overlap with its later host. Merger trees were constructed out of the halo catalogs at 94 time-steps, equally spaced in expansion factor ($\Delta a = 0.01$), based on the particle overlap of halos at different time-steps.

Due to the finite mass resolution of the simulation ($M_{\text{min,halo}} \simeq 10^{10} \text{M}_\odot$), subhalos can no longer be identified when their mass has dropped below this limit due to tidal stripping. Since mass loss can be substantial (>90%) this is important even for fairly massive subhalos. A special treatment of these so-called “orphans” is necessary. We determine the orbital parameters at the last moment when a subhalo is identified in the simulation and use them in the dynamical friction recipe of Bovlan-Kolchin et al. (2008), which is applicable at radii $r < r_{\text{vir}}$. We also tried an alternate recipe in which we make no explicit use of the subhalo information, but apply the dynamical friction formula from the time when the satellites first enter the host halo. We obtained very similar subhalo mass functions and radial distributions with the alternate recipe, confirming the self-consistency of the approach.

For the halo positions in the determination of CFs, we use the coordinates of the most bound particle for distinct and subhalos. For orphans, by definition, the position is not known, so we follow the position of the most bound particle from the last time-step when a subhalo was identified. Since the dynamical friction force vanishes in the dark matter only simulation after a subhalo is dissolved, yet not in reality when a galaxy is present at the center of the subhalo, the distance to the center of the host halo might be slightly overestimated with this prescription.

---

5 In an idealized situation, halo mass should increase monotonically with time until the halo becomes a subhalo, at which point the mass begins to decrease due to tidal stripping.

3. CONNECTING GALAXIES AND HALOS

In this section we describe how we derive the relationship connecting the stellar mass of a galaxy to the mass of its dark matter halo. In the standard picture of galaxy formation, gas can only cool and form stars if it is in a virialized gravitationally bound dark matter halo (White & Rees 1978). In this model the gas cooling rate, the star formation rate and thus the properties of the galaxy depend mainly on the virial mass of the host halo. Thus we expect the stellar mass of a central galaxy to be strongly correlated with the virial mass of the halo in which the galaxy formed. As we discussed in the last section, this corresponds to the virial mass for central galaxies, and to the maximum mass over the halo’s history for satellite galaxies. In the rest of this work, unless noted otherwise, the halo mass $M$ will represent:

\[ M = \begin{cases} M_{\text{vir}} & \text{for host halos} \\ M_{\text{max}} & \text{for subhalos} \end{cases} \tag{1} \]

Note that we have also experimented with instead using the present mass for subhalos, and found that we were not able to reproduce the galaxy clustering properties (see also Conroy et al. 2006).

3.1. The stellar-to-halo mass relation

In order to link the stellar mass of a galaxy $m$ to the mass of its dark matter halo $M$ we need to specify the SHM ratio. A direct comparison of the halo mass function $\phi (M)$ and the galaxy mass function $\phi (m)$ helps to constrain the stellar-to-halo mass function. If we assume that every host (sub) halo contains exactly one central (satellite) galaxy and that each system has exactly the same SHM ratio $m/M$, the galaxy stellar mass function can be derived trivially from the halo mass function and has the same features. The galaxy mass function derived for $m/M = 0.05$ is compared to the observed SDSS galaxy mass function in Figure 1. The observed galaxy mass function is steeper for high masses and shallower for low masses than the one derived from the halo mass function. Thus, for a constant SHM ratio there will inevitably be too many galaxies at the low and high mass end.

This implies that the actual SHM ratio $m/M$ is not constant, but increases with increasing mass, reaches a maximum around $m^*$ and then decreases again. Hence we adopt the following parameterization, similar to the one used in Yang et al. (2003):

\[ \frac{m(M)}{M} = 2 \left( \frac{m}{M} \right)_0 \left[ \left( \frac{M}{M_1} \right)^{-\beta} + \left( \frac{M}{M_1} \right)^{\gamma} \right]^{-1} \tag{2} \]

It has four free parameters: the normalization of the stellar-to-halo mass ratio $(m/M)_0$, a characteristic mass $M_1$, where the SHM ratio is equal to $(m/M)_0$, and two slopes $\beta$ and $\gamma$ which indicate the behavior of $m/M$ at the low and high mass ends respectively. We use the same parameters for the central and satellite populations, since – unlike luminosity – the stellar mass of satellites changes only slightly after they are accreted by the host halo.

Note that though both $\beta$ and $\gamma$ are expected to be positive, they are not restricted to be so. The SHM relation is therefore not necessarily monotonic.

3.2. Constraining the free parameters

Having set up the model we now need to constrain the four free parameters $M_1$, $(m/M)_0$, $\beta$ and $\gamma$. To do this, we populate the halos in the simulation with galaxies. The stellar
masses of the galaxies depend on the mass of the halo and are derived according to our prescription (equation [2]). The positions of the galaxies are given by the halo positions in the $N$-body simulation.

Once the simulation box is filled with galaxies, it is straightforward to compute the mass function $\Phi_{\text{mod}}(m)$. As we want to fit this model mass function to the observed mass function $\Phi_{\text{obs}}(m)$ by [Panter et al. 2007], we choose the same stellar mass range ($10^{10.5} - 10^{11.85} M_\odot$) and the same binsize. The observed SMF was derived using spectra from the Sloan Digital Sky Survey Data Release 3 (SDSS DR3); see [Panter et al. 2004] for a description of the method.

Furthermore it is possible to determine the stellar mass dependent clustering of galaxies. For this we compute projected galaxy CFs $w_p(r_p, m_i)$ in several stellar mass bins which we choose to be the same as in the observed projected galaxy CFs of [Li et al. 2006]. These were derived using a sample of galaxies from the SDSS DR2 with stellar masses estimated from spectra by [Kauffmann et al. 2003].

We first calculate the real space CF $\xi(r)$. In a simulation this can be done by simply counting pairs in distance bins:

$$\xi(r_i) = \frac{dd(r_i)}{N_p(r_i)} - 1$$  \hspace{1cm} (3)

where $dd(r_i)$ is the number of pairs counted in a distance bin and $N_p(r_i) = 4\pi N^2 r_i^2 \Delta r_i / L_{\text{box}}^3$ where $N$ is the total number of galaxies in the box. The projected CF $w_p(r_p)$ can be derived by integrating the real space correlation function $\xi(r)$ along the line of sight:

$$w_p(r_p) = 2 \int_0^\infty \frac{dr}{\sqrt{r^2 + r_p^2}} = 2 \int_{r_p}^\infty \frac{r \xi(r)}{\sqrt{r^2 - r_p^2}}$$  \hspace{1cm} (4)

where the comoving distance ($r$) has been decomposed into components parallel ($r_i$) and perpendicular ($r_p$) to the line of sight. The integration is truncated at 45 Mpc. Due to the finite size of the simulation box ($L_{\text{box}} = 100$ Mpc) the model correlation function is not reliable beyond scales of $r > 0.1 L_{\text{box}} \sim 10$ Mpc.

In order to fit the model to the observations we use Powell’s directions set method in multidimensions (e.g. Press et al. 1992) to find the values of $M_1$, $(m/M)_0$, $\beta$ and $\gamma$ that minimize either

$$\chi^2_r = \frac{\chi^2(\Phi)}{N_p}$$  \hspace{1cm} (mass function fit)

or

$$\chi^2_i = \frac{\chi^2(\Phi)}{N_p} + \frac{\chi^2(w_p)}{N_r N_m}$$  \hspace{1cm} (mass function and projected CF fit)

with $\chi^2(idm)$ and $\chi^2(idf)$, the errors for the SMF and projected CFs, respectively.

3.3. Estimation of parameter errors

In order to obtain estimates of the errors on the parameters, we need their probability distribution $\text{prob}(A|I)$, where $A$ is the parameter under consideration and $I$ is the given background information. The most likely value of $A$ is then given by: $A_{\text{best}} = \max(\text{prob}(A|I))$.

As we have to assume that all our parameters are coupled, we can only compute the probability for a given set of parameters. This probability is given by:

$$\text{prob}(M_1, (m/M)_0, \beta, \gamma|I) \propto \exp(-\chi^2)$$

In a system with four free parameters $A, B, C$ and $D$ one can calculate the probability distribution of one parameter (e.g. $A$) if the probability distribution for the set of parameters is known, using marginalization:

$$\text{prob}(A|I) = \int_{-\infty}^{\infty} \text{prob}(A, B|I) dB$$

Once the probability distribution for a parameter is determined, one can assign errors based on the confidence intervals. This is the shortest interval that encloses a certain percentage $X$ of the area under the posterior probability distribution. For the 1-sigma error $X = 68\%$ while for the 2-sigma error $X = 95\%$. Assuming that the probability distribution has been normalized to have unit area we seek $A_1$ and $A_2$ such that

$$\int_0^{A_1} \text{prob}(A|I) dA = \int_{A_2}^{\infty} \text{prob}(A|I) dA = \frac{1 - X}{2}.$$}

Finally the parameter $A$ is given as $A = A_{\text{best}} + \sigma_A$ with $\sigma_+ = A_2 - A_{\text{best}}$ and $\sigma_- = A_{\text{best}} - A_1$.

4. FITTING RESULTS

Here we present the results we obtain by fitting to the stellar mass function only, and for the combined fit to the SMF and the projected CF.
First we fit to the SDSS SMF and use the derived best-fit parameters to calculate the model projected correlation functions. Note that for now, we do not take into account any possible scatter in the $m(M)$ relation. We will consider scatter in §4.5.

We see in Figure[1] that our fit produces excellent agreement with the observed SMF. Using the approach described above we also compute the errors on the parameters. The results are summarized in Table[1].

Having derived the best-fit parameters, we can predict the projected CFs. We present the results both including and not including orphan galaxies, where we have fitted to the SMF for each case.

Figure[2] shows a comparison between our model and the SDSS projected correlation functions in five stellar mass bins ranging from $\log m/M_\odot = 9.0$ to $\log m/M_\odot = 11.5$ with a bin-size of 0.5 dex. The correlation function that has been derived without orphans is too low at small scales and can be regarded as a lower limit. Neglecting these galaxies results in an under-prediction of satellite galaxy clustering. As on small scales the projected CF depends mainly on the one-halo term this results in the underprediction of $w_p(r_p)$. This effect weakens for the clustering of more massive galaxies as they are more likely to be central galaxies and thus not effected by tidal stripping at all.

The agreement with the observationally derived $w_p(r_p)$ for the catalogue including orphaned galaxies is very good, which is also reflected in the low value of $\chi^2(w_p) = 3.83$. Note that this value has been calculated with the parameters from the mass function fit given above and does not correspond to a fit to the projected CFs.

As a test we also used the present mass instead of the maximum mass for subhalos. We then found that the projected CF was underpredicted particularly on small scales. This effect is due to tidal stripping of subhalos and is thus strongest at small scales where the subhalo contribution dominates.

### 4.1. The stellar mass function fit

We now investigate whether we can improve the agreement between the model and the observed projected CFs by performing a combined fit as described above. We obtain the

| $\log M_1$ ($m/M_\odot$) | $\beta$ | $\gamma$ | $\chi^2$ |
|-------------------------|--------|---------|--------|
| best fit               | 11.884 | 0.02820 | 1.057  | 0.5560  |
| $\sigma^+$             | 0.030  | 0.00061 | 0.054  | 0.0102  |
| $\sigma^-$             | 0.023  | 0.00053 | 0.046  | 0.0038  |

**NOTE.** No scatter included. All masses are in units of $M_\odot$.
same parameters as those we derived from the fit to the SMF alone. This seems surprising, but on further inspection we find that this is due to \( \chi^2(m) \) being a lot more sensitive to changes of the parameters than \( \chi^2(w_p) \). This means that if one changes the parameters a little in order to improve the fit to the projected correlation functions, one can get a slightly better agreement between the model and the observed projected CFs only at the cost of a large disagreement between the model and the observed stellar mass functions. In other words: \( \chi^2(w_p) \) is much flatter around its minimum than \( \chi^2(m) \), as shown in Figure 3.

We conclude that the SMF provides a much stronger constraint on the SHM function than the clustering data. Therefore, we can use our model to predict clustering as a function of stellar mass.

### 4.3. The resulting stellar-to-halo mass relation

The upper panel of Figure 4 shows the derived stellar mass as a function of halo mass. The light shaded area gives the 68% confidence interval while the dark and light shaded areas together give the 95% confidence interval. These have been derived using a set of different models computed on a mesh, as described in §3.3.

For the SHM ratio we apply the same procedure. The result is shown in the lower panel of Figure 4. We see that the SHM ratio has the form we expected: it increases with increasing halo mass, reaches its maximum value around \( M_1 \), and then decreases again.

### 4.4. Meaning of parameters and correlations

We now explore the effects of changing each parameter in order to understand how they affect the SMF. If we keep \( M_1 \) and \( \gamma \) fixed and only vary \( (m/M)_0 \), this corresponds to changing the stellar mass of the galaxy that lives inside each halo by a constant factor. This has no impact on the form of the SMF. Its shape stays the same, while only the position on the stellar mass axis changes. Due to the monotonic form of the SMF this directly determines the value of the normalization \( \phi^* \). For a larger value of \( (m/M)_0 \) we get a larger value of \( \phi^* \).

Varying only \( M_1 \) we find that the shape of the SMF changes drastically. For a higher \( M_1 \) than our best fit value, we get too many massive galaxies and too few low mass galaxies, while for a lower value of \( M_1 \) we get too few massive galaxies and too many low mass galaxies. This is because \( M_1 \) is the characteristic mass corresponding to the highest SHM ratio. In the SMF, this corresponds to the knee and we get a SMF which has its knee at the stellar mass corresponding to \( M_1 \). For a larger \( M_1 \) the knee is shifted to a higher stellar mass. Together, \( M_1 \) and the maximum stellar-to-halo mass ratio \( (m/M)_0 \) determine the normalization of the stellar mass function \( \phi \) and the characteristic mass \( m^* \).

Changing \( \beta \) affects mainly the low mass slope of the stellar mass function. For larger values of \( \beta \) the slope becomes shallower. As \( \beta \) influences mainly the slope of the low mass end of the SMF, it is strongly related to the parameter \( \alpha \) of the Schechter function. A small value of \( \beta \) corresponds to a high value of \( \alpha \).

If we change \( \gamma \), this mainly impacts the slope of the massive end of the SMF. For larger values of \( \gamma \) than for its best-fit value the slope of the massive end becomes steeper. As \( \gamma \) affects mainly the slope of the massive end of the SMF it is not coupled to a parameter of the Schechter function though it is related to the high-mass cutoff, assumed to be exponential in a Schechter function.

Figure 5 shows the contours of the two-dimensional probability distributions for the parameters pairs. We see a correlation between the parameters \([M_1, \gamma]\) and \([(m/M)_0, \gamma]\) and an anti-correlation between \([\beta, \gamma]\), \([\beta, M_1]\) and \([(m/M)_0, M_1]\). There does not seem to be a correlation between \([\beta, (m/M)_0]\).
4.5. Introducing scatter

Up until now we have assumed that there is a one-to-one, deterministic relationship between halo mass and stellar mass. However, in nature, we expect that two halos of the same mass \( M \) may harbor galaxies with different stellar masses, since they can have different halo concentrations, spin parameters and merger histories.

For each halo of mass \( M \), we now assign a stellar mass \( m \) drawn from a log-normal distribution with a mean value given by our previous expression for \( m(M) \) (Equation 2), with a variance of \( \sigma_m^2 \). We assume that the variance is a constant for all halo masses, which means that the percent deviation from \( m \) is the same for every galaxy. This is consistent with other halo occupation models, semi-analytic models and satellite kinematics (Cooray 2006; van den Bosch et al. 2007; More et al. 2009b).

Assuming a value of \( \sigma_m = 0.15 \) dex and fitting the stellar mass function only, we find the values given in Table 2. These values lie within the \((2\sigma)\) error bars of the best-fit values that we obtained with no scatter. The largest change is on the value of \( \gamma \), which controls the slope of the SHM relation at large halo masses.

In Figure 6, we compare our model without scatter with the model including scatter. We have also included the relation between halo mass and the average stellar mass for the model with scatter in Figure 6.

5. THE CONDITIONAL MASS FUNCTION

In the previous section we derived a model that specifies the stellar mass of a central galaxy as a function of the virial mass of its host halo and the stellar mass of a satellite galaxy as a function of the maximum mass of the subhalo in which it lives. It has become common to represent the population of host halos by the Halo Occupation Distribution (HOD). This includes the halo occupation function \( P(N|M) \) which is the probability distribution that a halo of mass \( M \) contains \( N \) galaxies (of a specific type). A close relative of the HOD is the “conditional luminosity function” (CLF; e.g. Yang et al. 2003; van den Bosch et al. 2007; Yang et al. 2004). It extends the halo occupation function \( P(N|M) \) (which gives only information about the total number of galaxies per halo in a given luminosity range) and yields the average number of galaxies with luminosities in the range \( L \pm dL/2 \) as a function of the virial mass \( M \) of their host halo.

We define its analog, the “conditional mass function” (CMF), or the average number of galaxies with stellar masses in the range \( m \pm dm/2 \) as a function of the virial mass \( M \) of
Fig. 7.— The conditional mass function (CMF) predicted by our model at $z = 0$. We plot the derived SMFs ($d\tilde{n}_g/d\log m$) in a subsample of halo mass bins. The left panels show the CMF for a model without scatter while the right panels show the CMF with scatter of $\sigma_m = 0.15$. The label in each panel is the range of host halo mass $\log M / M_\odot$. The stellar mass functions are normalized such that a host halo contains exactly one central galaxy. The total CMF consists of a central galaxy part (crosses) and a satellite part (diamonds). The central part is described by a lognormal distribution (solid line) and the satellite part is described by a truncated Schechter function (dashed line) using the parameters that were derived by a fit to the CMF. The dotted line shows the completeness limit used in the fit to the satellite contribution.
their host halo. This provides a direct link between the SMF \( \Phi(m) \) and the host halo mass function \( dn(M)/dM \): 

\[
\Phi(m) = \int_0^\infty \Phi(m|\Delta M)d\Delta M 
\tag{5}
\]

A host halo of mass \( M \) can contain a whole population of galaxies with different stellar masses \( m \). If we count the number of galaxies living in host halos with a virial mass in the range \( M \in [M_1, M_2] \) we can compute the SMF of the halo bin \([M_1, M_2]\):

\[
\tilde{\Phi}(m) = \int_{M_1}^{M_2} \Phi(m|\Delta M)d\Delta M \approx \Phi(m|\Delta M)\Delta n
\tag{6}
\]

The tilde over a function represents the fact that it is computed in a halo mass bin. We have replaced the integral by a “tophat” with a width of \( \Delta n \) (number of host halos in the bin) and a height of \( \Phi(m|M_0) \), where \( M = (M_1 + M_2)/2 \).

This equation allows us to put constraints on \( \Phi(m|\Delta M) \) by calculating \( \tilde{\Phi}(m)/\Delta n \). We can then choose an adequate parameterization of \( \Phi(m|\Delta M) \) and fit these parameters to \( \tilde{\Phi}(m)/\Delta n \) in every halo mass bin. Finally we can investigate the halo mass dependence of the parameters.

### 5.1. Parameterization

In order to specify the CMF \( \Phi(m|M) \) we divide the galaxy population into a central and a satellite part, as in the updated CLF formalism (Cooray 2006; Yang et al. 2008; Cacciato et al. 2008). The central part is \( \Phi_c(m|M) \) and the satellite part is \( \Phi_s(m|M) \). Then the total CMF is the sum of both parts:

\[
\Phi(m|M) = \Phi_c(m|M) + \Phi_s(m|M)
\tag{7}
\]

Note that both \( \Phi_c(m|M) \) and \( \Phi_s(m|M) \) are statistical functions and should not be regarded as the mass functions of galaxies living in a given individual halo.

For the central population we expect the CMF to have a peak around the stellar mass \( m_c \) that corresponds to the host halo’s virial mass \( M \) in the SHM relation (equation 8). Due to the halo mass bin size this distribution gets smeared out, because halos in the interval \([M_1, M_2]\) contain central galaxies of stellar masses \( m \in [m_1(M), m_2(M)] \). Thus \( \Phi(m)/\Delta n \) will be finite inside the interval \([m_1(M), m_2(M)]\) and zero elsewhere with a normalization such that the number of central galaxies per halo equals one. This can be regarded as scatter \( \sigma_{\text{bin}} \) due to the binning. If we add intrinsic scatter \( \sigma_m \) to relation (2), we expect \( \Phi_c(m|M) \) to be a lognormal with a maximum around \( m_c(M) \) and a variance of \( \sigma_m^2 \). This scatter the binning scatter \( \sigma_{\text{bin}} \) adds in quadrature (assuming that \( \sigma_{\text{bin}} \) and \( \sigma_m \) are uncorrelated), resulting in a total scatter of \( \sigma^2 = \sigma_{\text{bin}}^2 + \sigma_m^2 \). For both cases (\( \sigma_m = 0 \) and \( \sigma_m \neq 0 \)) we use a lognormal distribution:

\[
\Phi_c(m|M) = \frac{1}{\sqrt{2\pi}\ln10m_\sigma} \exp \left[ -\frac{\log^2(m/m_\sigma)}{2\sigma_c^2} \right],
\tag{8}
\]

where the mean \( m_\sigma(M) \) and width \( \sigma_c^2(M) \) are parameterized functions of the halo mass \( M \).

For the satellite population we adopt a Schechter function with a steeper slope for the massive end. This is done by squaring the argument of the exponential function in the Schechter function:

\[
\Phi_s(m|M) = \frac{\Phi_s^*}{m_s} \left( \frac{m}{m_s} \right)^{\alpha_s} \exp \left[ -\left( \frac{m}{m_s} \right)^2 \right].
\tag{9}
\]

Also here the parameters \( \Phi_s^*(M), m_s(M) \) and \( \alpha_s(M) \) are functions of the host halo mass \( M \). They are the normalization, the characteristic mass and the low mass slope of the satellite population of host halos of mass \( M \).

### 5.2. Constraining the conditional mass function

We populate the halos and subhalos in our simulation with central and satellite galaxies according to the prescription in section 3. Then we choose halo mass bins between \( \log M/M_\odot = 10.2 \) and \( \log M/M_\odot = 15.0 \) with a bin size of \( \Delta M = 0.4 \) dex. In every halo mass bin we seek all galaxies which live in a host halo with a mass in that bin, which we divide between central and satellite galaxies. For these populations we then compute two separate SMFs which we normalize such that the number of central galaxies per host halo equals one. This procedure then yields for every halo mass bin a central and a satellite distribution \( (d\tilde{n}_g/d\log M)\Delta n_h \).

Using equation (6) we can now relate the stellar mass function in a halo mass bin to the CMF:

\[
\frac{d\tilde{n}_g(m)}{d\log M} \frac{1}{\Delta n_h} = \frac{\ln 10 \ 1}{M} \frac{d\tilde{n}_g(m)}{dM} = \frac{\ln 10 M \ 1}{\Delta n_h} \approx \ln 10 M \ 1 \ 1.
\tag{10}
\]

Now we can fit the five parameters \( m_c(M), \sigma_c(M), m_s(M), \Phi_s^*(M) \) and \( \alpha_s(M) \) to the SMFs in each halo bin. We compute and fit the central and the satellite parts separately.

The left panels of Figure 7 show the CMF in a subsample of halo mass bins running from \( \log M/M_\odot = 10.2 \pm 0.2 \) to \( 15.0 \pm 0.2 \), where we have not included intrinsic scatter in the SHM relation. For the satellite part, only galaxies with a mass above the completeness limits for each halo mass bin (as indicated in Figure 7) have been used in the fit.

In low-mass halos \( (\log M/M_\odot < 11.0) \) the contribution from satellite galaxies is very small and the central contribution dominates until \( \log M/M_\odot = 12.0 \). For massive halos \( (\log M/M_\odot > 13.0) \) the satellite contribution dominates by number. The mean of the lognormal fit to the central contribution also increases with halo mass as stipulated by the model derived in Section 3. The characteristic mass scale of the satellite contribution also increases with halo mass meaning that the most massive satellite galaxies have a mass which is comparable to the mass of the central galaxy.

The scatter of the central contribution \( \sigma_x(M) \) decreases with halo mass. As we did not include any scatter in the model, this scatter reflects the width (0.4 dex) of the halo mass bins \( \sigma_{\text{bin}} \). The halo mass dependence of \( \sigma_x(M) \) arises because a fixed halo mass bin is mapped to a smaller galaxy mass bin for larger halo mass due to the shape of the SHM relation. Another feature of the CMF is the slope for low mass satellite galaxies \( \alpha_x(M) \) which becomes shallower with increasing halo mass.

### 5.3. The parameters of the conditional mass function

In this section we investigate the halo mass dependence of the five parameters of the CMF: \( m_c(M), \sigma_c(M), m_s(M), \Phi_s^*(M) \) and \( \alpha_s(M) \). They have been fixed by fitting to the stellar mass functions in each halo mass bin. We introduce a parameterization in order to describe the dependence on halo mass and constrain these by a fit to each parameter. The results are presented in Table 3. This provides a complete description of the CMF.
As we have already determined the mean relation between the stellar mass of a galaxy and the mass of its halo, the form of \( m_c(M) \) has to be the same and can thus be described by equation (2):

\[
m_c(M) = 2 M \left( \frac{m_c}{M} \right)_0 \left[ \left( \frac{M}{M_{1c}} \right)^{-\beta_c} + \left( \frac{M}{M_{1c}} \right)^{\gamma_c} \right]^{-1}
\]

This yields four parameters \( \left( m_c/M_0 \right)_0, M_{1c}, \beta_c \) and \( \gamma_c \).

In the upper left panel of Figure 8, \( m_c(M) \) is plotted as a function of halo mass. Note that by construction, it has the same form as the SHM relation.

The scatter of the central galaxy contribution is high for low halo masses and decreases for more massive halos. The middle left panel of Figure 8 shows \( \sigma_c(M) \) as a function of halo mass. As one can see, \( \sigma_c(M) \) goes to a constant value both for low and high halo masses while it decreases with halo mass. We therefore choose the following parameterization:

\[
\sigma_c(M) = \sigma_\infty + \sigma_1 \left[ 1 - \frac{2}{\pi} \arctan \left( \frac{\xi \log \frac{M}{M_2}}{2} \right) \right]
\]

This yields four more parameters \( \sigma_\infty, \sigma_1, \xi \) and \( M_2 \). Here, \( \sigma_\infty \) sets the high mass limit of \( \sigma_c(M) \) while \( \sigma_1 \) sets the difference between the low and high mass limits of \( \sigma_c(M) \). The parameter \( M_2 \) determines the mass scale at which the transition occurs and \( \xi \) sets the strength. For a large (small) value of \( \xi \) the transition occurs in a small (large) interval around \( M_2 \).

The specific shape of \( \sigma_c(M) \) can be explained by the form of the SHM relation (equation 2). As we have not included any scatter in this relation (\( \sigma_m = 0 \)), the width of the lognormal function of the central galaxy distribution arises from the width of the halo mass bin (\( \sigma_c = \sigma_{\text{bin}} \)). A halo mass interval \([M_1, M_2]\) contains only central galaxies with stellar masses of \( m \in [m_1(M_1), m_2(M_2)] \). The lower left panel of Figure 8 illustrates this by showing how halo mass bins affect the bin size of the stellar mass. If we choose the same bin size for low and high mass halos, we get different bin sizes for low and high...
mass galaxies, due to the changing slope of \(m(M)\). Therefore the transition occurs where the slope of \(m(M)\) changes which is around \(M_1\), so the value of \(M_2\) is very close to that value.

As Figure 7 shows that the satellite contribution falls off around the mean mass of the central galaxy, we expect the characteristic mass of the modified Schechter function \(m_{\ast}(M)\) to follow \(m_c(M)\). We describe \(m_c(M)\) with the same function we used for the parameterisation of \(m_c(M)\):

\[
m_c(M) = 2M \left[ \frac{m_c}{M} \right]^\alpha \left[ \left( \frac{M}{M_{1s}} \right)^{-\beta} + \left( \frac{M}{M_{1s}} \right)^\gamma \right]^{-1}
\]

This function yields four parameters \((m_c/M_0, M_{1s}, \beta, \gamma)\). The upper right panel of Figure 8 plots \(m_c(M)\) as a function of halo mass. We see that the shape is similar to that of \(m_s(M)\). Note that \(m_c(M)\) is always lower than \(m_s(M)\), while the deviation increases with increasing halo mass. This implies that for high halo masses the satellite contribution to the CMF falls off before the mean mass of the central galaxy.

The normalization of the modified Schechter function is small for low halo masses and increases with the mass of the host halo. The middle right panel of Figure 8 shows \(\Phi_s(M)\) as a function of halo mass. We see that \(\Phi_s(M)\) can be described by a power law and choose the following parametrisation:

\[
\Phi_s(M) = \Phi_0 \left( \frac{M}{M_0} \right)^\lambda
\]

We get two more parameters, \(\Phi_0\) and \(\lambda\). The normalization of \(\Phi_s(M)\) is given by \(\Phi_0\) and the slope by \(\lambda\). The shape of \(\Phi_s(M)\) implies that the probability for a host halo to harbor satellite galaxies (in a given stellar mass range) increases with increasing halo mass.

The slope of the modified Schechter function for the satellite contribution becomes shallower for more massive halos. The lower right panel of Figure 8 shows \(\alpha_s(M)\) as a function of halo mass and shows that \(\alpha_s(M)\) goes to a constant value for both low and high halo masses. Similar to \(\sigma_c(M)\), we choose the parameterization:

\[
\alpha_s(M) = \alpha_\infty + \alpha_1 \left[ 1 - 2 \arctan \left( \frac{\zeta}{\log \left( \frac{M}{M_1} \right)} \right) \right]
\]

This yields four more parameters \(\alpha_\infty, \alpha_1, \zeta \) and \(M_1\). Here, \(\alpha_\infty\) sets the high mass limit of \(\alpha_s(M)\) while \(\alpha_1\) sets the difference between the low and high mass limits of \(\alpha_s(M)\). The mass scale at which this transition occurs is determined by \(M_1\) and \(\zeta\) sets its strength. The transition occurs in a small (large) interval around \(M_1\) for a small (large) value of \(\zeta\).

5.4. The occupation numbers

In order to compare our results to other HOD models it is useful to compute the average number of galaxies per halo \(\langle N \rangle\), as this is the main prediction of the HOD approach. To compute \(\langle N \rangle\) from the CMF we simply integrate \(\Phi_m|M|\) over the desired stellar mass range:

\[
\langle N \rangle = \int m_1 \Phi(m|M) dm
\]

As we have divided \(\Phi(m|M)\) into a central galaxy contribution \(\Phi_c(m|M)\) and a satellite galaxy contribution \(\Phi_s(m|M)\), we can compute separate occupation numbers for central and satellite galaxies:

\[
\langle N \rangle = \int m_1 \Phi_c(m|M) dm + \int m_1 \Phi_s(m|M) dm = \langle N_c \rangle + \langle N_s \rangle
\]

Table 3

| \(\sigma_m = 0.0\) | \(\sigma_m = 0.15\) |
|------------------|------------------|
| \(\log M_{1s}\) | 11.9805 ± 0.0939 | 11.8298 ± 0.0278 |
| \(\sigma_\infty\) | 0.0450 ± 0.0087 | 0.1590 ± 0.0031 |
| \(\alpha_1\) | 0.1308 ± 0.0279 | 0.0462 ± 0.0030 |
| \(\xi\) | 3.0420 ± 0.5415 | 4.1392 ± 0.9732 |

The average number of central galaxies per halo \(\langle N_c \rangle\) is given by

\[
\langle N_c \rangle = \frac{1}{2} \left[ \text{erf}(\eta_2 - \text{erf}(\eta_1)) \right],
\]

with the error-function \(\text{erf}(x)\) and the integration boundaries

\[
\eta_1 = \frac{\log (m_1/m_\ast)}{\sqrt{2}\sigma_c} \quad \text{and} \quad \eta_2 = \frac{\log (m_2/m_\ast)}{\sqrt{2}\sigma_c}.
\]

The average number of satellite galaxies per halo \(\langle N_s \rangle\) is

\[
\langle N_s \rangle = \frac{\Phi_s}{\Phi} \left( \frac{\alpha_s}{\alpha_1} + \frac{1}{\sigma_\infty} \right) \left( \Gamma \left( \frac{\alpha_s}{\alpha_1} + \frac{1}{\sigma_\infty} \right) - \Gamma \left( \frac{\alpha_s}{\alpha_1} + \frac{1}{\sigma_\infty} \right) \right),
\]

with the upper incomplete gamma function \(\Gamma(a, x)\) and the integration boundaries

\[
\kappa_1 = (m_1/m_\ast)^2 \quad \text{and} \quad \kappa_2 = (m_2/m_\ast)^2.
\]

Figure 9 shows the resulting occupation numbers for the values of the CMF parameters that were derived in section 5.3 (without using scatter). The five lines in each panel correspond to the same stellar mass bins that have also been used for the clustering calculation.

The left panel shows the average number of central galaxies per halo \(\langle N_c \rangle\) as a function of halo mass. In the middle panel, the average number of satellite galaxies per halo \(\langle N_s \rangle\) as a function of halo mass is shown. The right panel plots the average number of all galaxies per halo \(\langle N_{\text{tot}} \rangle\) as a function of halo mass. A galaxy of low stellar mass can thus either be a central galaxy of a low mass halo, or a satellite galaxy of a massive halo. It is not likely to live in a halo of intermediate mass.

As it is common in the literature to plot occupation numbers for stellar mass intervals, but for galaxy samples with a mass above a given threshold, we need to adjust equations (14) and (15). The stellar mass threshold is then given by \(m_1\) while \(m_2 \to \infty\). This yields for the average number of central galaxies

\[
\langle N_c \rangle(M, m_1) = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{\log (m_1/m_\ast)}{\sqrt{2}\sigma_c} \right) \right],
\]
since $\text{erf}(x \to \infty) \to 1$, and for the average number of satellite galaxies

$$\langle N_s \rangle (M, m_1) = \frac{\Phi_s}{2} \Gamma \left[ \frac{\alpha_s}{2} + \frac{1}{2} \right] \left( \frac{m_1}{m_s} \right)^2$$

(20)

since $\Gamma(a, x \to \infty) \to 0$.

Figure 10 shows occupation numbers for different stellar mass thresholds. The left panel shows the average number of central galaxies per halo $\langle N_c \rangle (M)$ as a function of halo mass. The middle panel plots the average number of satellite galaxies per halo $\langle N_s \rangle (M)$ as a function of halo mass. It is similar to the middle panel of Figure 9 while it is larger at a given halo mass. In the right panel the average number of all galaxies per halo $\langle N_{tot} \rangle (M)$ as a function of halo mass is shown.

5.5. The impact of scatter

Until now, we have used the SHM relation (2) without any intrinsic scatter. In this section we investigate how the CMF and the parameters change if we include a scatter $\sigma_m$ as described in section 4.5. This scatter is again assumed to be constant with host halo mass.

The right panels of Figure 7 show the resulting CMF in a subsample of halo mass bins for an intrinsic scatter of $\sigma_m = 0.15$. The central part is now no longer near-constant in the interval $[m(M - \Delta M/2), m(M + \Delta M/2)]$ as in the left panels of Figure 7 (where $\sigma_m = 0$) but has the form of a lognormal with a broader distribution for bigger $\sigma_m$. As the scatter has been taken from a lognormal distribution, the central galaxy contribution to the CMF is distributed in the same way. Hence, $\sigma_c(M)$ changes with respect to the model that does not include artificial scatter. We notice that at the massive end the binning scatter $\sigma_{bin}$ and the intrinsic scatter $\sigma_m$ add to the total scatter $\sigma_{tot}^2$. At the low mass end, however, the total scatter is less than what has been obtained by using no intrinsic scatter. This shows that the two forms of scatter do not add in quadrature and indicates that they are correlated.

We compare $m_c(M), \sigma_c(M), m_s(M), \Phi_s^c(M)$ and $\alpha_s(M)$ for $\sigma_m = 0$ and $\sigma_m = 0.15$ and show the resulting parameters in Table 3 (columns four and five) and in Figure 8. The mean mass of the central galaxy $m_c(M)$ does not change much if artificial scatter is introduced. The most likely stellar mass of a central galaxy is still given by the SHM relation, so the mean of the gaussian in logarithmic space stays the same. Also the parameters of the satellite population $[m_s(M), \Phi_s^c(M)$ and $\alpha_s(M)]$ do not change significantly.

6. COMPARISON

6.1. Other HOD models

Numerous variations on halo occupation models have been presented in the literature. In this section we describe some of the most popular ones and compare them to our model. As many authors use different initial mass functions and definitions of halo masses, we convert all results to the conventions that we have used in this work (Kroupa IMF and virial overdensity).

In the Non-Parametric model (Vale & Ostriker 2006; Conroy et al. 2006; Shankar et al. 2006), galaxy properties, such as luminosity and stellar mass, are monotonically related to the mass of dark matter halos. Using the observed galaxy SMF, the most massive halo is matched to the most massive galaxy:

$$n_g(>m_l) = n_h(>M_l)$$

(21)

In this way, the observed SMF is automatically reproduced. Applying this procedure and fitting the parameters of the SHM relation to the result, we have derived the values given in Table 4. These are in good agreement with the parameters of our model, except for $\beta$. We find that this is due to the shape of the SHM ratio for low masses. For the Non-Parametric model, $m(M < M_l)$ can not be perfectly described by a single power law, as is assumed in our model.

Adding an additional parameter and assuming a fitting function with five free parameters, we are able to fit the SHM relation predicted by the non-parametric model quite precisely. The fifth parameter accounts for the deviation from the power-law at high and low masses. Using the parameteri-
As we discussed in the introduction, semi-analytic models (SAMs) of galaxy formation attempt to predict the relationship between dark halo mass and stellar mass by a priori modelling of physical processes, such as the growth of structure, cooling, star formation, and stellar and AGN feedback. We compare our results with predictions from the latest version of the semi-analytic models of Somerville & Primack (1999); see Somerville et al. (2008). For this we compute the mean stellar mass of central galaxies as a function of the mass of the host halo in halo mass bins. The results are shown in Figure 11 and are in good agreement with our model.

### 6.2. Gravitational lensing

The relation between stellar mass and halo mass can be constrained observationally using galaxy-galaxy lensing. Gravitational lensing induces shear distortions of background objects around foreground galaxies, allowing the mass of the dark matter halo to be estimated. Mandelbaum et al. (2005, 2006) have used SDSS data to calibrate the predicted signal from a halo model which has been derived from a dissipation-less simulation. They have extracted the mean halo mass as a function of stellar mass. The lensing data for combined early and late-type galaxies (Mandelbaum, private communication) are shown in Figure 11 and are in excellent agreement with our model.

### 6.3. Semi-analytic models

As the mass function is steep at high masses and shallow for low masses, a change in the scatter will influence the number of massive galaxies strongly, while it will have only a small effect on the low mass end. As the other three parameters $M_1$, $(m/M)_0$ and $\beta$ are coupled to the Schechter function parameters, there are two parameters to constrain the slope of the massive end of the SMF. This degeneracy can cause the difference in $\gamma$ between the two models. The fact that in the Millennium simulation the cosmology is different to that of our simulation also affects the value of the parameters.

#### Table 5

**FIT PARAMETERS FOR EQUATION (22)**

| log$m_0$ | log$M_1$ | $\gamma_1$ | $\gamma_2$ | $\beta$ |
|---------|---------|------------|------------|--------|
| 10.864  | 10.456  | 7.17       | 0.201      | 0.557  |
| ± 0.043 | 0.211   | 1.16       | 0.018      | 0.031  |

**NOTE.** — All masses are in units of $M_\odot$.  

In Wang et al. (2006) a model similar to ours is used to constrain the SHM ratio. The halo catalogue is taken from the Millennium simulation (Springel et al. 2005); halos are identified using a friends-of-friends group finder while substructure is found using the SUBFIND algorithm of Springel et al. (2001). As observational constraints, the authors use a SMF which they compute from the SDSS DR2 data using the mass estimates of Kauffmann et al. (2003) and the projected CFs of Li et al. (2000).

The parameterization they use is similar to ours, with four free parameters that can easily be converted to $M_1$, $(m/M)_0$, $\beta$ and $\gamma$ and an unconstrained scatter. These are fixed by generating a grid of models and the best-fit model is defined as the one for which $\chi^2 = \chi^2(\Phi) + \chi^2(w_p)$ is minimal. They find that their fit improves if they take a different set of parameters for central and satellite galaxies. In Table 4 we compare our best-fit parameters with their central galaxy best-fit parameters which have been updated in Wang et al. (2007). We show these results in Figure 11.

The values of $M_1$ and $(m/M)_0$ are in very good agreement with our values, but the slopes are both higher, resulting in fewer massive and fewer low mass galaxies. The reason for the difference in the low mass end is the different simulation used. As the resolution of the simulation in our model is higher, the low mass end can be constrained more tightly. For the massive end the difference in $\gamma$ can be explained by the additional unconstrained scatter that is used in Wang et al. (2006).
Fig. 11.— Comparison of the stellar-to-halo mass relation \( m(M) \) between our model (solid line), models from other authors and galaxy-galaxy lensing (symbols). The blue areas are the 1-\( \sigma \) and 2-\( \sigma \) levels and the error-bars on the symbols are the 2-\( \sigma \) levels of the halo mass.

(described above) to obtain the parameter estimates from the SAM. We summarize these results in Table [4] and show them in Figure [11].

The two slopes are in very good agreement with our results. However, the normalization in the Croton et al. (2006) SAM is \( \sim 25\% \) higher and the characteristic mass is \( \sim 25\% \) lower than what we found and what Yang et al. (2006) find for their model. This is because the SAM of Croton et al. (2006) does not produce a perfect fit to the observed SMF.

6.4. SDSS group catalogue

Another direct way of studying galaxy properties as a function of halo mass is using the SDSS group catalogue presented in Yang et al. (2007). In this approach, galaxies are first linked together into “groups” using a friends-of-friends algorithm. Each group is then assigned a total halo mass by matching to the theoretical dark matter halo mass function. Yang et al. (2003) present the relation between the mean stellar mass of the central galaxy and the host halo mass. We fit the parameters of equation [2] to their relation and present the results in Table [4].

We note that the characteristic mass and the normalization derived from the group catalogue are both higher than our model parameters. The high mass slope of the SHM relation in the group catalogue is shallower than that of our model. The low mass slope is also shallower, however, the constraints on the low mass slope in the group catalogue are weak, since the lowest halo masses are \( \log(\frac{M}{M_\odot}) \sim 11.7 \). This can also be seen in Figure 11 where we show the SHM relation of the group catalogue for comparison.

7. HIGH REDSHIFT

The discussion in the previous sections has focussed solely on the present day universe. In this section we extend our analysis to higher redshifts and derive the redshift dependence of the stellar-to-halo mass relation. Having chosen a particular observed stellar mass function at a given redshift, we can investigate how the parameters of the SHM ratio change with time. This allows us to learn about the evolution of galaxies. Also, with this information, we can populate the N-body simulation snapshots with galaxies at different redshifts using the appropriate redshift dependent SHM relation, and then use the spatial information from the simulation to compute the stellar mass dependent correlation functions.

Since at the present time there are no high redshift clustering data as a function of stellar mass available, we fit the four parameters of equation [2] to the observed SMFs at a given redshift. We argued in section [4.2] that the SMFs provide much stronger constraints on the SHM ratio than the clustering data. Thus we should be able to use our model to predict clustering as a function of stellar mass at any redshift.

7.1. Which survey for which redshift

In order to constrain the SHM relation we have to first select observational stellar mass functions at the redshifts we want to investigate. Because of the trade-off between surveying large areas and obtaining deep samples, measurements of the SMF at high redshift tend to suffer from limited dynamic range. Therefore it is important to think about how the constraints on our four SHM function parameters arise from the observations.

The characteristic mass \( M_1 \) and the maximum SHM ratio \( (m/M)_0 \) mostly depend on galaxies and halos of intermediate mass. The high mass slope \( \gamma \) is fixed by the number of massive galaxies since these live in the massive halos. On the other hand, the low mass slope \( \beta \) is set by the number of low mass galaxies since these live in the low mass halos.

For a survey with a fixed area on the sky, the observed volume is smaller for low redshifts \( (z \lesssim 1) \) than for high redshifts. In order to compute the SMF at high galaxy masses, the observed volume has to be relatively large, as massive galaxies are rare. Thus for low redshifts one has to choose a wide survey (large area) to determine the SMF for massive galaxies and properly constrain \( \gamma \). Constraining the SMF at the low mass end requires a high level of completeness for low mass galaxies, which are very faint objects. Hence we have to choose a deep survey that can detect faint galaxies in order to constrain \( \beta \).

Taking these considerations into account, we choose the stellar mass functions presented in Drory et al. (2004) to constrain the parameters \( M_1, (m/M)_0 \) and \( \gamma \) at low redshifts. The authors derive the SMFs using MUNICS which is a wide area, medium-deep survey selected in the \( K \) band. The detection limit is \( K \approx 19.5 \) and the subsample the authors use covers 0.28 deg\(^2\). We apply our method using these mass functions and take the three parameters from that analysis.

However, the MUNICS survey is not deep enough to detect galaxies that are fainter than the characteristic mass of the SMF (the knee) and thus is not sufficient to constrain the parameter \( \beta \). To constrain \( \beta \) we choose the SMFs derived in Fontana et al. (2006). This work is based on the GOODS-MUSIC sample, a multicolor catalogue extracted from the survey conducted over the Chandra Deep Field South. The catalogue is selected in the \( z_{850} \) and \( K \) bands, covers an area of 143.2 arcmin\(^2\), and is complete to a typical magnitude of \( K \approx 23.5 \). We apply our method using the SMFs computed with the \( z_{850} \) band selected sample and take the parameter \( \beta \) from that analysis.

For high redshift \( (z \gtrsim 1) \) we use the SMFs presented in Fontana et al. (2006) to constrain all four parameters.
Fig. 12.— Comparison between the model and the observed stellar mass functions for different redshifts. The observed stellar mass functions are taken from Drory et al. (2004) (for \(z \leq 0.9\)) and from Fontana et al. (2006) (for \(z \geq 1.1\)) and are represented by the symbols. The model stellar mass functions have been fitted to the observations and are represented by the solid lines. The dashed lines are the theoretical mass function we obtain from the redshift-dependent parameterization. The redshift is indicated at the top of each panel.

For high redshifts, the volume of a redshift bin becomes large enough to sample massive galaxies, and therefore the GOODS-MUSIC sample is sufficient to constrain \(\gamma\).

We convert all SMFs which use a Salpeter initial mass function to the Kroupa/Chabrier initial mass function.

### 7.2. Evolution of the parameters

Having selected the observational SMFs for a set of different redshifts, we fit the four free parameters \(M_1\), \((m/M)_0\), \(\beta\) and \(\gamma\) to the observations. The errors on the parameters are derived in a similar way as explained in section 3.3, but instead of using confidence intervals we have fitted a Gaussian to the probability distributions of \(M_1\), \((m/M)_0\) and \(\gamma\) and a lognormal to the probability distribution of \(\beta\).

Figure[12] shows the observed and the model stellar mass functions for different redshifts (indicated at the top of each panel). The values of the resulting four parameters for the different redshifts are shown in Table[6] and the redshift evolution is plotted in Figure[13]. The characteristic mass \(M_1\) grows with increasing redshift, while the normalization of the SHM ratio \((m/M)_0\) becomes smaller with increasing redshift. This means that there is less stellar content in a halo of a given mass at a higher redshift.

| \(z\) | \(\log(M_1)\) | \(\pm (m/M)_0\) | \(\pm \beta\) | \(-\) | \(+\) | \(\gamma\) |
|-------|----------------|----------------|-------------|---|---|-----|
| 0.0   | 11.88 | 0.02 | 0.0282 | 0.0005 | 1.06 | 0.05 | 0.05 | 0.56 | 0.00 |
| 0.5   | 11.95 | 0.24 | 0.0254 | 0.0047 | 1.37 | 0.22 | 0.27 | 0.55 | 0.17 |
| 0.7   | 11.93 | 0.23 | 0.0215 | 0.0048 | 1.18 | 0.23 | 0.28 | 0.48 | 0.16 |
| 0.9   | 11.98 | 0.24 | 0.0142 | 0.0034 | 0.91 | 0.16 | 0.19 | 0.43 | 0.12 |
| 1.1   | 12.05 | 0.18 | 0.0175 | 0.0060 | 1.66 | 0.26 | 0.31 | 0.52 | 0.40 |
| 1.5   | 12.15 | 0.30 | 0.0110 | 0.0044 | 1.29 | 0.25 | 0.32 | 0.41 | 0.41 |
| 1.8   | 12.28 | 0.27 | 0.0116 | 0.0051 | 1.53 | 0.33 | 0.41 | 0.41 | 0.41 |
| 2.5   | 12.22 | 0.38 | 0.0130 | 0.0037 | 0.90 | 0.20 | 0.24 | 0.30 | 0.30 |
| 3.5   | 12.21 | 0.19 | 0.0101 | 0.0020 | 0.82 | 0.72 | 1.16 | 0.46 | 0.21 |

**Note:** For \(M_1\), \((m/M)_0\) and \(\gamma\) the errors are drawn from a Gaussian and thus are symmetric (indicated by the symbol \(\pm\)). For \(\beta\) the errors are drawn from a lognormal distribution and thus there is a lower error (indicated by the symbol \(-\)) and an upper error (indicated by the symbol \(+\)). All quoted masses are in units of \(M_\odot\).
The high mass slope $\gamma$ can be constrained only weakly. This is due to the limitation of the available galaxy surveys. As the area of the survey is small, the volume in which galaxies are detected is limited, and thus massive galaxies are very rare. This results in large error bars for the SMF of massive galaxies which propagate into the error bars of $\gamma$. The situation improves slightly for higher redshifts as the volume of higher redshift bins is larger and thus more massive galaxies can be observed. The value of $\gamma$ decreases with increasing redshift. For higher redshifts ($z > 1$) the error bars on $\gamma$ become very large because of the limited area covered by the available deep surveys (in this case, GOODS). We leave it up to the reader to assess the reliability of our results at $z > 1$ based on our quoted error bars.

The low mass slope $\beta$ seems to increase with redshift until $z \approx 2$ and then drops to a low value. However, as the redshift increases it becomes more and more difficult to observe low mass galaxies which are very faint. Thus the high redshift values for $\beta$ are not very well constrained and perhaps not to be fully trusted. We therefore assume that $\beta$ grows with increasing redshift.

As we explained in Section 4.4, $\beta$ is strongly related to the parameter $\alpha$ of the Schechter function. A small value of $\beta$ corresponds to a large absolute value of $\alpha$ while a large value of $\beta$ results in a low absolute value of $\alpha$. This would mean that for higher redshifts the stellar mass function would become shallower, in contradiction with observations (e.g. Fontana et al. 2006 show that the absolute value of $\alpha$ increases with redshift). However, one has to remember that the halo mass function also changes with redshift and becomes steeper. Thus the halo mass function steepens more than the SMF, so $\beta$ has to increase in order to compensate.

With the derived parameter values it becomes possible to interpolate and find the SHM ratio at any redshift. This is done by choosing a redshift-parameterization for each of the parameters. As $M_1$ and $(m/M)_0$ do not change much above a redshift of $z > 1.5$ we choose power laws for the redshift dependence:

$$\log M_1(z) = \log M_0 \cdot (z+1)^\mu.$$  \hfill (23)

and

$$(m/M)_0(z) = (m/M)_{z=0} \cdot (z+1)^\nu.$$  \hfill (24)

with the normalizations $M_0$ and $(m/M)_{z=0}$ and the slopes $\mu$ and $\nu$.

To parameterize $\gamma$ over redshift, a linear dependence would lead to a negative $\gamma$ at a certain redshift. Though this is not forbidden, it leads to a SHM ratio which would be increasing monotonically with halo mass which is inconsistent with feedback processes at the massive end. Hence we also choose a power-law parameterization for $\gamma$:

$$\gamma(z) = \gamma_0 \cdot (z+1)^\gamma_1.$$  \hfill (25)

with the normalization $\gamma_0$ and the slope $\gamma_1$. From Figure 13 we are not able to infer whether $\beta$ converges to a constant value. Thus we adopt a simple linear parameterization:

$$\beta(z) = \beta_1 \cdot (z+\beta_0).$$  \hfill (26)

Note that we have also tried other parameterizations (constant $\beta$, decreasing $\beta$) but could not reproduce the observed stellar mass functions. Using the linear parameterization for $\beta$ and the power laws for the other parameters we were able to compute stellar mass functions that are in good agreement with the observed ones.

A fit to the derived values presented in Table 6 yields the parameters given in Table 7. As we do not fully trust the derived values of $\beta$ for $z \gtrsim 2$ we neglect these two values and fit a line to the remaining values of $\beta$. 

![Figure 13](https://example.com/figure13.png)

**Fig. 13.** — Evolution of the stellar-to-halo mass relation parameters with redshift. The symbols correspond to the derived values while the solid line is a fit to the data. For $M_1$, $(m/M)_0$ and $\gamma$ this is a power-law, while for $\beta$ it is a straight line.
TABLE 7
PARAMETERS FOR REDSHIFT DEPENDENT STELLAR-TO-HALO MASS RELATION

| $M_0$  | $\mu$ | $(m/M_*)$ | $\nu$ | $\gamma_0$ | $\gamma_1$ | $\beta_0$ | $\beta_1$ |
|-------|-------|-----------|-------|-----------|-----------|---------|---------|
| 11.88 | 0.019 | 0.0282    | -0.72 | 0.556     | -0.26     | 1.06    | 0.17    |
| ±     | 0.002 | 0.0003    | 0.06  | 0.001     | 0.05      | 0.06    | 0.12    |

NOTE. — All quoted masses are in units of $M_\odot$.

Fig. 14.— Stellar mass as a function of halo mass for different redshifts. The solid lines show different redshifts, which are indicated at the top of the panels.

7.3. The stellar-to-halo mass relation for different redshifts

Having developed a redshift dependent model of the stellar-to-halo mass relation we now test this model by computing interpolated stellar mass functions for different redshifts. For this we use the method described in section 4. However, now we do not use the parameters that have been derived at each redshift by fitting the model to the observations but we use the eight parameters of the redshift dependent SHM relation that have been derived in the previous section.

The resulting interpolated SMFs are compared to the observations (and the fitted mass functions) in Figure 12. For $z \lesssim 2$ we see excellent overall agreement, the interpolated mass functions mostly overlap with the error bars of the observations.

The SMFs for the high redshifts $z \gtrsim 2$ are too low. The deviations are largest at the low mass end. However, if we look at Figure 12 we see that $\beta$ is higher than the derived value for the two highest redshifts which results in a low mass slope that is too shallow.

To compare the relation at different redshifts, we use the redshift dependent SHM relation with the eight parameters that have been derived in the previous section. Figure 14 plots stellar mass versus halo mass for different redshifts. The plot shows that at a fixed low halo mass (e.g., $M = 10^{11} M_\odot$), galaxies that live in such halos are more massive at low redshift ($m \sim 10^9 M_\odot$ for $z = 0$) than galaxies that live in a halo of the same mass at a higher redshift ($m \sim 10^8 M_\odot$ for $z = 2$). In contrast, massive halos contain more massive galaxies at high redshift, while at low redshifts the galaxies in massive halos have less mass. However, as halos also become more massive over time, one cannot identify a halo of a certain mass at high redshifts with a halo of the same mass at low redshifts. Thus the fact that at a given (high) halo mass the mass of the central galaxy is lower at present than at an earlier epoch does not imply that individual galaxies lose mass during their evolution. This only means that large halos accrete dark matter faster than large galaxies grow in stellar mass, while the growth of low mass halos is slower than that of the central galaxies they harbor (see also Conroy & Wechsler 2008). Because of its statistical nature, our model is not suitable for following the evolution of an individual galaxy through cosmic time.

7.4. Clustering at higher redshift

Having determined the SHM relation as a function of redshift we are now able to populate halos with galaxies at any redshift. We choose a set of redshifts and populate the halos with galaxies, deriving the stellar masses from the redshift dependent SHM relation. We divide these galaxies into six samples of different stellar mass between $\log m/M_\odot = 8.5$ and 11.5. For each of these samples we compute the real space CF $\xi(r)$ by counting pairs in distance bins (equation 4). This leads to six CFs for every selected redshift.

Figure 15 shows the CFs for six different redshifts as a function of stellar mass. We also plot the correlation function of dark matter at the respective redshifts for comparison. For all redshifts we see that massive galaxies are clustered more strongly than low mass galaxies. The higher the redshift, the more the CFs for different stellar masses differ. For high redshift, there are very few massive galaxies in our limited volume simulation box, and so the error bars become larger.

7.5. The galaxy bias

The bias of any object may be defined as the square root of the ratio between the CF of the object $\xi_o(r)$ and the CF of dark matter particles $\xi_{dm}(r)$:

$$b(r) = \sqrt{\frac{\xi_o(r)}{\xi_{dm}(r)}}. \quad (27)$$

Here we focus on the galaxy two-point CF $\xi_{gg}(r, m, z)$, which in addition to the distance between the galaxies also depends on the redshift and the stellar mass of the galaxies:

$$b(r, m, z) = \sqrt{\frac{\xi_{gg}(r, m, z)}{\xi_{dm}(r, z)}}. \quad (28)$$

From our predicted galaxy CFs, we compute the bias at a scale of $r \approx 6$ Mpc for every redshift and stellar mass by averaging the values of $b(r, m, z)$ between $r = 2$ Mpc and 10 Mpc. Figure 16 shows the redshift dependence of the bias. The bias is computed at $r \approx 6$ Mpc as a function of redshift for the different stellar masses (as one can see from Figure 15 the scale dependence of the bias is quite weak). The symbols represent the averaged value of the bias while the solid lines correspond to a fit to the symbols. For this we have used a power law form:

$$b(z) = b_0(z+1)^{b_1} + b_2 \quad (29)$$

where the parameters $b_0$, $b_1$, and $b_2$ are functions of stellar mass.
Correlation functions as a function of stellar mass at high redshift. The different panels correspond to different redshifts, which are given at the bottom of each panel. The different lines are correlation functions for six stellar mass bins, which are given in the upper left panel. The error-bars on the most massive sample are from Poisson statistics. The correlation function of dark matter particles (thick solid line) at the respective redshifts is also shown for comparison. At high redshift the correlation function of the massive samples is only shown on large scales, since there is no relevant one-halo term.

The galaxy bias at a fixed scale (≈6 Mpc) as a function of redshift for different stellar masses. The symbols have been derived by averaging the bias over a distance interval while the lines are fits to the symbols.

This shows that the bias at a fixed stellar mass increases with increasing redshift. Massive galaxies are always biased more strongly than galaxies of lower mass as the value of $b(r \approx 6 \text{ Mpc}, m, z)$ for massive galaxies exceeds that of low mass galaxies at any redshift.

The fit parameters are given in Table 8.

8. CONCLUSIONS

The goal of this paper is to characterize the relationship between the stellar masses of galaxies and the masses of the dark matter halos in which they live at low and high redshift, and to make predictions of stellar mass dependent galaxy clustering at high redshift.

We used a high-resolution N-body simulation and identified halos and subhalos. Halos and subhalos were populated with central and satellite galaxies using a parameterized SHM relation. For host halos the mass was given by the virial mass $M_{\text{vir}}$ while for subhalos we used the maximum mass of the halo over its history $M_{\text{max}}$ since we expect the stellar mass of

| log $m_G$ | $b_0$ ± uncertainty | $b_1$ ± uncertainty | $b_2$ ± uncertainty |
|-----------|----------------------|----------------------|----------------------|
| 8.5 - 9.0 | 0.062 ± 0.017        | 2.59 ± 0.18          | 1.025 ± 0.062        |
| 9.0 - 9.5 | 0.074 ± 0.008        | 2.58 ± 0.26          | 1.039 ± 0.028        |
| 9.5 - 10.0| 0.042 ± 0.003        | 3.17 ± 0.05          | 1.147 ± 0.021        |
| 10.0 - 10.5| 0.053 ± 0.014       | 3.07 ± 0.17          | 1.225 ± 0.077        |
| 10.5 - 11.0| 0.069 ± 0.014       | 3.19 ± 0.13          | 1.269 ± 0.087        |
| 11.0 - 11.5| 0.173 ± 0.035       | 2.89 ± 0.20          | 1.438 ± 0.061        |

NOTE. — All quoted masses are in units of $M_\odot$.
the satellite galaxy to be more tightly linked to this quantity. We described the ratio between stellar and halo mass by a function with four free parameters, a low-mass slope $\beta$, a characteristic mass $M_1$, a high-mass slope $\gamma$, and a normalization $(m/M_0)$. We fit for the values of these parameters by requiring that the observed galaxy SMF is reproduced. We find that the SHM function has a characteristic peak at $M_1 \sim 10^{12} M_\odot$, and declines steeply towards both smaller mass ($\beta \sim 1$) and less steeply towards larger mass halos ($\gamma \sim 0.6$). The physical interpretation of this behavior is the interplay between the various feedback processes that impact the star formation efficiency. Supernova feedback is more effective at reheating and expelling gas in low mass halos, while AGN feedback is more effective in high mass halos (e.g. Shankar et al. 2006, Croton et al. 2006, Bower et al. 2006, Somerville et al. 2008). In this picture, the characteristic mass $M_1$ is the halo mass where the efficiency of these two processes crosses.

We have thoroughly discussed the meaning of the parameters. We have also investigated the effects on the SHM relation that arise from introducing scatter to the relation. To do this we have added scatter drawn from a lognormal distribution with a typical variance of $\sigma_m = 0.15$ to the SHM function. We showed that the impact of such a scatter on the three of the four parameters is negligible, with a small but significant impact on the high-mass slope $\gamma$.

We showed that adding constraints from stellar mass dependent galaxy clustering did not change the values of our best-fit parameters. Put another way, the likelihood (here $\chi^2$) function for the clustering constraint is much “flatter” than that for the mass constraint, so adding the clustering constraint does not significantly change the distribution for the most likely (best-fit) parameter values. Fitting to the SMF only, we found that the observed projected CFs of galaxies for five samples of different stellar mass were reproduced well. This means that the clustering properties of galaxies are predominantly driven by the clustering of the halos and subhalos in which they reside. From this we concluded that our model can predict clustering as a function of stellar mass at any redshift.

In order to describe how galaxies of different masses populate host halos, we introduced the conditional mass function $\Phi(m|M)$, which yields the average number of galaxies with stellar masses in the range $m \pm \Delta m/2$ that live in a distinct halo of mass $M$. It is described by five parameters which are functions of halo mass. We divided the conditional mass function into a contribution from central galaxies (described by a lognormal distribution) and a contribution from satellite galaxies (described by a modified Schechter function). We computed the SMF in different halo mass bins and fitted the five parameters in each bin. Introducing halo mass dependent functions for every parameter and fitting these to the derived values in the halo mass bins, we determined the halo mass dependence of the five parameters and thus fully described the conditional mass function. We also computed the occupation numbers of halos which give the average number of galaxies of a given stellar mass that live inside a halo of mass $M$.

We compared the results for our SHM function with those that have been derived using other approaches. These include other halo occupation type models, gravitational lensing and semi-analytic models. We showed that all methods yield consistent SHM relations.

Using SMFs at higher redshifts, we applied our model at earlier epochs of the universe. We thus constrained the SHM relation at a given set of redshifts between $z = 0$ and $z \sim 4$. This allowed us to study how the four parameters of the SMF function depend on redshift. For each parameter we introduced a redshift dependent function. We found that the characteristic mass increases with redshift while the normalization decreases with redshift. This indicates that there is less stellar content in halos at higher redshifts. As the halo mass function steepens more with redshift than the stellar mass function, the low mass slope increases with redshift. We present an eight parameter fitting function describing the redshift dependent SHM relation.

Using the SHM relation that we derived in this way, along with spatial information for halos from the N-body simulation, we predicted the high-redshift real space CFs for five stellar mass intervals. We find that for all redshifts, massive galaxies are more clustered than galaxies of lower mass. Using the real space CF of dark matter we calculated the galaxy bias as a function of distance, redshift and stellar mass. Averaging over spatial scale in an interval around $r \approx 6$ Mpc, we demonstrated that the galaxy bias increases with redshift, and presented fitting formulae for the galaxy bias as a function of stellar mass and redshift. In a companion paper (Moster et al. 2009) we will use these bias results to present predictions for the cosmic variance $\sigma$ for galaxies of different stellar mass.

We thank Benjamin Panter, Cheng Li, Adriano Fontana, Niv Drory and Rachel Mandelbaum for providing their data in electronic form. We also thank Eric Bell, Ramin Skibba, Risa Wechsler and Charlie Conroy for helpful discussions. This research was supported in part by the DFG cluster of excellence ‘Origin and Structure of the Universe’.

REFERENCES

Bell, E. F. & de Jong, R. S. 2001, ApJ, 550, 212
Berlind, A. A. & Weinberg, D. H. 2002, ApJ, 575, 587
Bertschinger, E. 2001, ApJS, 137, 1
Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517
Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2006, MNRAS, 370, 645
Boylan-Kolchin, M., Ma, C.-P., & Quataert, E. 2008, MNRAS, 383, 93
Bryan, G. L., & Norman, M. L. 1999, ApJ, 495, 50
Cacciato, M., van den Bosch, F. C., More, S., Li, R., Mo, H. J., & Yang, X. 2008, ArXiv e-prints
Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abraham, R., Gravel, P., Morris, S., & Pritchet, C. J. 1996, ApJ, 462, 32
Cooray, A., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E. 1994, MNRAS, 271, 781
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Conroy, C., & Wechsler, R. H. 2008, ArXiv e-prints
Croton, D. J., et al. 2006, MNRAS, 365, 11
Drory, N., Bender, R., Feulner, G., Hopp, U., Maraston, C., Snigula, J., & Hil, G. J. 2004, ApJ, 608, 742
Drory, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., & Pannella, M. 2005, ApJ, 619, L13
Elsner, F., Feulner, G., & Hopp, U. 2008, A&A, 477, 503
Erickson, L. K., Gottesman, S. T., & Hunter, Jr., J. H. 1987, Nature, 325, 779
Eyles, I. P., Bunker, A. J., Ellis, R. S., Lacey, M., Stanford, E. R., Stark, D. P., & Chiu, K. 2007, MNRAS, 374, 910
Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189
Fontana, A., et al. 2006, A&A, 459, 745
Fry, E., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Kauffmann, G., et al. 2003, MNRAS, 341, 33
Katz, N. Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Klypin, A., Gottlüber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999, ApJ, 516, 350
Kroupa, P. 2001, MNRAS, 322, 231
Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715
Shankar, F., Eke, V., Sefusatti, E., & Hill, G. J. 2004, MNRAS, 368, 715
Mandelbaum, R., Tasitsiomi, A., Seljak, U., Kravtsov, A. V., & Wechsler, R. H. 2005, MNRAS, 362, 1451
Mare, S., van den Bosch, F. C., & Cacciato, M. 2009a, MNRAS, 392, 917
More, S., van den Bosch, F. C., Cacciato, M., Mo, H. J., Yang, X., & Li, R. 2009b, MNRAS, 392, 801
Moster, B. P., Somerville, R. S., & Newman, J. A. 2009, in prep
Li, C., Kauffmann, G., Jing, Y. P., White, S. D. M., Börner, G., & Cheng, F. Z. 2006, MNRAS, 368, 21
Lin, Y.-T., Mohr, J. J., & Stanford, S. A. 2003, ApJ, 591, 749
Lin, Y.-T., & Mohr, J. J. 2004, ApJ, 617, 879
Norberg, P., et al. 2001, MNRAS, 328, 64
Norberg, P. et al. 2002, MNRAS, 332, 827
Panter, B., Heavens, A. F., & Jimenez, R. 2004, MNRAS, 355, 764
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, MNRAS, 378, 1550
Peacock, J. A. & Smith, R. E. 2000, MNRAS, 318, 1144
Seljak, U. 2000, MNRAS, 318, 203
Shankar, F., Lapi, A., Salucci, P., De Zotti, G., & Danese, L. 2006, ApJ, 643, 14
Somerville, R. S. & Primack, J. R. 1999, MNRAS, 310, 1087
Springel, V., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Spergel, D. N., et al. 2007, ApJS, 170, 377
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289
Springel, V. 2005, MNRAS, 364, 1105

Springel, V., et al. 2005, Nature, 435, 629
Tasitsiomi, A., Kravtsov, A. V., Wechsler, R. H., & Primack, J. R. 2004, ApJ, 614, 533
Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41
Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189
Vale, A., & Ostriker, J. P. 2006, MNRAS, 371, 1173
van den Bosch, F. C., Mo, H. J. 2003, MNRAS, 340, 771
van den Bosch, F. C., et al. 2007, MNRAS, 376, 841
Wang, L., Li, C., Kauffmann, G., & De Lucia, G. 2006, MNRAS, 371, 537
Wang, L., Li, C., Kauffmann, G., & De Lucia, G. 2007, MNRAS, 377, 1419
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
White, M. 2001, MNRAS, 321, 1
Yan, H., Dickinson, M., Giavalisco, M., Stern, D., Eisenhardt, P. R. M., & Ferguson, H. C. 2006, ApJ, 651, 24
Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057
Yang, X., Mo, H. J., Jing, Y. P., van den Bosch, F. C., & Chu, Y. 2004, MNRAS, 350, 1153
Yang, X., Mo, H. J., van den Bosch, F. C., Pasquali, A., Li, C., & Barden, M. 2007, ApJ, 671, 153
Zehavi, I., et al. 2002, ApJ, 571, 172
Zehavi, I., et al. 2004, ApJ, 608, 16
Zehavi, I., et al. 2005, ApJ, 630, 1