



Quantifying Scatter in Galaxy Formation at the Lowest Masses

Ferah Munshi^{1,2,3}, Alyson M. Brooks², Elaad Applebaum², Charlotte R. Christensen⁴, T. Quinn⁵, and Serena Sligh¹

¹Homer L. Dodge Department of Physics & Astronomy, University of Oklahoma, 440 W. Brooks Street, Norman, OK 73019, USA; fdm@ou.edu

²Department of Physics & Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Road, Piscataway, NJ 08854, USA

³Department of Physics and Astronomy, Vanderbilt University, PMB 401807, Nashville, TN 37206, USA

⁴Department of Physics & Astronomy, Grinnell College, Grinnell, IA 50112, USA

⁵Astronomy Department, University of Washington, Box 351580, Seattle, WA, 98195–1580, USA

Received 2021 January 14; revised 2021 June 8; accepted 2021 June 21; published 2021 December 8

Abstract

We predict the stellar mass–halo mass (SMHM) relationship for dwarf galaxies, using simulated galaxies with peak halo masses of $M_{\text{peak}} = 10^{11} M_{\odot}$ down into the ultra-faint dwarf range to $M_{\text{peak}} = 10^7 M_{\odot}$. Our simulated dwarfs have stellar masses of $M_{\text{star}} = 790 M_{\odot}$ to $8.2 \times 10^8 M_{\odot}$, with corresponding *V*-band magnitudes from -2 to -18.5 . For $M_{\text{peak}} > 10^{10} M_{\odot}$, the simulated SMHM relationship agrees with literature determinations, including exhibiting a small scatter of 0.3 dex. However, the scatter in the SMHM relation increases for lower-mass halos. We first present results for well-resolved halos that contain a simulated stellar population, but recognize that whether a halo hosts a galaxy is inherently mass resolution dependent. We thus adopt a probabilistic model to populate “dark” halos below our resolution limit to predict an “intrinsic” slope and scatter for the SMHM relation. We fit linearly growing log-normal scatter in stellar mass, which grows to more than 1 dex at $M_{\text{peak}} = 10^8 M_{\odot}$. At the faintest end of the SMHM relation probed by our simulations, a galaxy cannot be assigned a unique halo mass based solely on its luminosity. Instead, we provide a formula to stochastically populate low-mass halos following our results. Finally, we show that our growing log-normal scatter steepens the faint-end slope of the predicted stellar mass function.

Unified Astronomy Thesaurus concepts: Dwarf galaxies (416)

1. Introduction

In the Λ Cold Dark Matter (Λ CDM) paradigm of cosmological structure formation, dwarf galaxies are predicted to be the smallest, most abundant, yet least luminous galaxies in the universe. Attempts to link dwarf galaxies to their parent dark matter halos via abundance matching have led to discrepancies between theory and observations (e.g., Ferrero et al. 2012; Garrison-Kimmel et al. 2014a; Brook & Di Cintio 2015; Papastergis et al. 2015). Abundance matching matches a stellar mass or luminosity at a given abundance to dark matter halos with the same abundance, derived from a dark matter–only simulation. A monotonic relationship is generally assumed (Guo et al. 2010; Behroozi et al. 2013; Moster et al. 2013). Critically, abundance matching also assumes that every dark matter halo is occupied by a galaxy. Abundance matching studies generally yield fair agreement for the stellar mass–halo mass (SMHM) relation for halos of masses $\gtrsim 10^{11} M_{\odot}$. Additionally, for halos of roughly Milky Way mass and greater, abundance matching also reproduces clustering statistics (e.g., Conroy & Wechsler 2009; Wechsler & Tinker 2018).

However, derivations of the SMHM relation at lower masses have yielded discrepancies (e.g., Behroozi et al. 2013; Moster et al. 2013; Brook et al. 2014; Garrison-Kimmel et al. 2014b; Read et al. 2017; Jethwa et al. 2018). If the SMHM relation has the form $M_{\text{star}} \propto M_{\text{halo}}^{\alpha}$, the range of derived α varies from 1.4 to 3.1 for galaxies smaller than $M_{\text{halo}} < 10^{11.5} M_{\odot}$. However, Moster et al. (2013) and Behroozi et al. (2013) did not have data in order to derive the SMHM relation below stellar masses of a few $\times 10^7 M_{\odot}$. While their results are unconstrained at lower masses, the slopes at their lowest measured mass were quite different, $\alpha = 1.4$ (Behroozi et al. 2013) versus $\alpha = 2.4$ (Moster et al. 2013). Read et al. (2017) find a fairly shallow

relation between $10^7 < M_{\text{star}}/M_{\odot} < 10^9$, with $\alpha \approx 1.2$, for field dwarfs in the Sloan Digital Sky Survey (SDSS). They also derive the halo masses of isolated, local dwarfs via rotation curve fitting (Read et al. 2016b) and find that the SMHM relation for the individual galaxies is well described by their derived SMHM relation from abundance matching, including when extrapolated to lower stellar masses. They attribute this match to the fact that they use an isolated galaxy sample, arguing that including galaxies processed in a group environment leads to a steeper SMHM relation. Brook et al. (2014) and Garrison-Kimmel et al. (2014b), however, used Local Group galaxy data to determine the SMHM relation at $10^6 < M_{\text{star}}/M_{\odot} < 10^8$. Again, they came to quite different conclusions about the value of α , 3.1 in Brook et al. (2014) and 1.9 in Garrison-Kimmel et al. (2014b).⁶

Similarly, both Jethwa et al. (2018) and Nadler et al. (2020) examine the SMHM using Local Group data, specifically using Milky Way satellites. However, unlike Brook et al. (2014) and Garrison-Kimmel et al. (2014b), they do not require that every dark matter halo contains a galaxy. Jethwa et al. (2018) explore a range of assumptions, from the standard abundance matching, to models that allow unoccupied halos and varying forms of the stellar mass function (SMF). Nadler et al. (2020) use a more constrained model, building on earlier work to determine which subhalos are likely to contain luminous satellites, and the known selection function of current surveys to determine which satellites should be detected. Both Jethwa et al. (2018) and Nadler et al. (2020) found that most halos must be occupied down to $M_{\text{peak}} \sim 10^8 M_{\odot}$. It is common in abundance matching to assume that all halos continue to host galaxies.

⁶ Though the slope is dependent on the normalization at higher masses, and the values come into better agreement when a consistent normalization is adopted; see Garrison-Kimmel et al. (2017).

The idea that all halos down to $M_{\text{peak}} \sim 10^8 M_{\odot}$ are occupied is contrary to traditional wisdom, which holds that reionization may stifle star formation entirely in such low-mass halos (e.g., Gnedin 2000; Governato et al. 2007; Okamoto et al. 2008; Brooks et al. 2013). Simulators generally find that reionization causes low-mass halos to be entirely devoid of stars, or to contain so few as to be beyond detection (e.g., O’Shea et al. 2015; Ocvirk et al. 2016; Wheeler et al. 2019; Katz et al. 2020; Ocvirk et al. 2020). However, the mass scale at which this occurs remains an open question. Sawala et al. (2016) used simulations representative of the Local Group to show that the fraction of halos that are not populated (not luminous) increases drastically for $z = 0$ halo masses less than $10^9 M_{\odot}$, suggesting a plummeting galaxy formation efficiency at these mass scales. Fitts et al. (2018) find a similar drop in galaxy formation below $M_{\text{halo}} = 10^9 M_{\odot}$. Ocvirk et al. (2020) also find a drop below this halo mass, though at $z = 6$. However, other simulators have found that galaxies populate lower-mass halos, closer to $M_{\text{halo}} \sim 10^8 M_{\odot}$ (Revaz & Jablonka 2018; Wheeler et al. 2019; Katz et al. 2020). Simulations of the low-mass halos at high z have shown that stars can form in even lower-mass halos at $z > 6$ (e.g., Wise et al. 2012; Côté et al. 2018; Latif & Khochfar 2019; Nakatani et al. 2020; Skinner & Wise 2020).

In addition to the slope of the SMHM relation and the occupation fraction of low-mass halos, there is also uncertainty in the (intrinsic) scatter in the SMHM relation. The scatter at the high-mass end of the SMHM relation is consistently measured to be relatively small, ~ 0.2 dex (Behroozi et al. 2013; Reddick et al. 2013; Matthee et al. 2017; Kravtsov et al. 2018), though see Taylor et al. (2020). However, the scatter at the low-mass end may be much larger. Observations indicate that dwarf galaxies over a range of stellar masses may all occupy dark matter halos with a narrow range of masses (e.g., see Figure 1 of Klypin et al. 2015; Ferrero et al. 2012; Strigari et al. 2008). Fitts et al. (2017) simulated 15 galaxies all with roughly the same halo mass at $z = 0$ ($\approx 10^{10} M_{\odot}$), yet the stellar content varied by two orders of magnitude ($10^{5-7} M_{\odot}$). The star formation history at dwarf galaxy scales is likely to depend on the mass accretion history of the halo (Brooks & Zolotov 2014; Weisz et al. 2015; Sawala et al. 2016). Rey et al. (2019) showed that the stellar content of a *single* dark matter halo (with the same final halo mass at $z = 0$) could vary based on its growth rate. If the halo grew rapidly and was more massive at early times, it contained a higher stellar mass at $z = 0$, by as much as an order of magnitude. The results for the single halo in Rey et al. (2019) represent a lower limit on how much scatter in stellar mass might be expected at a fixed halo mass.

Garrison-Kimmel et al. (2017) explored the implications of scatter within M_{star} at a given M_{halo} using Local Group data down to $M_{\text{star}} \sim 10^5 M_{\odot}$. They demonstrated that there is a degeneracy between the slope and the scatter of the SMHM relation when using the SMHM relation to derive the SMF. Small halos are more likely to scatter to large stellar masses due to the rapidly rising mass function in a Λ CDM cosmology. Hence, large scatter requires a steeper SMHM slope in order to reproduce the observed SMF. A steeper SMF may, e.g., alleviate the mismatch between the number of predicted classical dwarfs in CDM with what is observed (Garrison-Kimmel et al. 2017). That said, Nadler et al. (2020) find that the scatter must be small in order to not overpopulate the observed satellite luminosity function (LF) of the Milky Way. The inherent steepness of the LF at dwarf scales also has

implications for how reionization proceeded. In order for galaxies to reionize the universe, the UVLF likely has to maintain a steep slope down to UV magnitudes as faint as about -10 (e.g., Kuhlen & Faucher-Giguère 2012; Robertson et al. 2013; Weisz & Boylan-Kolchin 2017). Measurements of lensed dwarfs at $1 < z < 3$ show no break in the UVLF down to a UV magnitude of roughly -14 (Alavi et al. 2016). Results from lower-resolution simulations using the code adopted in this work show that a LF with no break down to -15 can reionize the universe at a rate consistent with observations (Anderson et al. 2017). In this work we quantify the slope and scatter of the SMHM relation inherent in our simulations down to the ultra-faint dwarf (UFD) range, and explore the impact on the slope of the SMF.

A number of simulators have used baryonic simulations to show that they match the derived SMHM relation at halo masses above $10^{10} M_{\odot}$ (e.g., Brook et al. 2012; Aumer et al. 2013; Munshi et al. 2013; Hopkins et al. 2014), but only a few have examined the SMHM of simulated dwarf galaxies below this mass (Munshi et al. 2013; Shen et al. 2014; Oñorbe et al. 2015; Sawala et al. 2015). Most works do not examine a large enough sample of dwarfs to be able to define an SMHM relation, but can instead only compare to extrapolated abundance matching results. Sawala et al. (2015) studied a larger population of dwarfs, tracing halos down to $\sim 10^8 M_{\odot}$. They showed that an abundance matching that used only *populated* halos leads to placing higher stellar mass galaxies into halos than traditional abundance matching (see also Jethwa et al. 2018). Considering only occupied halos and the fact that baryons alter the halo mass function (Munshi et al. 2013; Sawala et al. 2013; Benson 2020) leads to an SMHM relation that indicates the typical halo mass at a given stellar mass, unlike traditional abundance matching that indicates the typical stellar mass at a given halo mass. Thus, taking into account baryonic effects leads to an SMHM relation that is much shallower at the low-mass end than that found by traditional abundance matching.

In this paper we address the question, how do galaxies populate low-mass dark matter halos? We do this using a suite of zoomed-in simulations that contain over 200 dwarfs, the Marvel-ous Dwarfs and Justice League simulations. These simulations were purposely designed to yield a large sample of dwarf galaxies, from LMC-mass at the most massive end, down into the UFD galaxy range for the first time at the low-mass end. We examine both field dwarfs and their satellites, and satellites around Milky Way-mass galaxies. However, we make a radical departure from earlier works that assumed that their simulated “dark” halos were those impacted by reionization. Many simulators treat these dark halos as a prediction of the simulation. Instead, we assume that our dark halos are impacted by our resolution limit, and explore the intrinsic slope and scatter of the SMHM relation including these unresolved galaxies.

We present our new suites of high-resolution simulations in Section 2. In Section 3 we show that the scatter in the SMHM relation grows as halo mass decreases. We quantify the slope and scatter of the SMHM relation, first for only the galaxies that are well resolved in our simulation, and then extrapolated to include unresolved galaxies. We demonstrate the impact of our predicted scatter on the resulting SMF that can be tested by, e.g., the Vera Rubin Observatory’s Legacy Survey of Space and Time (LSST). In Section 4 we discuss other factors that we

have not necessarily explored in this paper but may also impact the resulting SMHM relation. We summarize our results in Section 5.

2. Simulations

For the first time, we present the full sample of 211 dwarf galaxies including both the “Marvel-ous Dwarfs” zoom simulations, along with the “DC Justice League” Milky Way–mass zoom simulations. The “Marvel-ous Dwarfs” (hereafter Marvel) are slightly different than traditional zoom simulations, which generally select one halo of interest and place the highest resolution particles on that halo out to a few virial radii. For the Marvel runs, we instead selected regions of the universe that contain dozens of dwarf galaxies and ran the entire region as a zoom-in simulation, with the goal of generating one of the largest samples of simulated dwarf galaxies at incredibly high resolution. We ran four such simulations (named CptMarvel, Elektra, Rogue, and Storm) within a WMAP3 cosmology (Spergel et al. 2007), each containing a dozen to a few dozen dwarfs. The most massive galaxies in the Marvel runs are $\sim$ LMC-mass ($\sim 10^{11} M_{\odot}$ in halo mass), but the high resolution of the simulations (60 pc force resolution, gas, initial star, and dark matter masses of $1410 M_{\odot}$, $420 M_{\odot}$, and $6650 M_{\odot}$, respectively) allows galaxies as low as $M_{\text{star}} \sim 3000 M_{\odot}$ UFDs to be resolved. A total of 68 Marvel dwarfs are used in this work, 11 of which are satellites (within the virial radius) of dwarf galaxies.

The regions selected for the Marvel runs are roughly 1.5 to 7 Mpc away from a Milky Way–mass galaxy and can be considered representative of the Local Volume. To complement these regions, we also use four zoom simulations of Milky Way–mass galaxies and their surrounding environments (out to ~ 1 Mpc), the DC Justice League simulations (named after the first four women who have served on the US Supreme Court; Sandra, Ruth, Sonia, and Elena). Two of the DC Justice League simulations (Ruth and Sonia) are run at “Near Mint” (NM) resolution, which is slightly lower resolution than the Marvel dwarfs (170 pc force resolution, initial gas, initial star, and dark matter masses of $2.7 \times 10^4 M_{\odot}$, $8000 M_{\odot}$, and $4.2 \times 10^4 M_{\odot}$, respectively) within a Planck cosmology (Planck Collaboration et al. 2016). However, we include dwarfs from the “Mint” resolution DC Justice League (Sandra and Elena) presented in Applebaum et al. (2021), which have a resolution within a factor of two of Marvel (87 pc force resolution and gas, dark, and initial star particle masses of 3310, 17,900, and $994 M_{\odot}$, respectively). This combined “Mint” and “Near Mint” set of four simulations yields 143 dwarfs: 64 field dwarfs (47 of which are backplash dwarfs of the Milky Way–mass hosts as defined in Applebaum et al. 2021) and 79 satellites.⁷

Both suites of simulations were run with the N-Body + SPH code CHANGA (Menon et al. 2015). CHANGA adopts the hydrodynamic modules of GASOLINE2 (Wadsley et al. 2004, 2017) but uses a faster gravity solver, as well as the CHARM++ (Kalé & Krishnan 1993) runtime system for dynamic load balancing and communication. This allows CHANGA to scale up to thousands of cores. It is the excellent scalability of CHANGA that allowed the Marvel simulation suite and the “Mint” DC Justice League simulations to be run.

Both sets of simulations utilize the gas cooling and star formation scheme introduced in Christensen et al. (2012). Metal line cooling and the diffusion of metals is included (Shen et al. 2010), and the nonequilibrium formation and destruction of molecular hydrogen, H_2 , is followed. We apply a uniform, time-dependent UV field from Haardt & Madau (2012) in order to model photoionization and heating, and the Lyman–Werner radiation from young stars is tracked. Star formation is restricted to occur only in the presence of H_2 .

Star formation occurs stochastically when gas particles become cold ($T < 1000$ K) and dense ($n > 0.1 m_H \text{ cm}^{-3}$). Although the density threshold is low, in practice the requirement that H_2 be present restricts stars to forming only in gas that reaches a density threshold $n > 100 m_H \text{ cm}^{-3}$. The probability, p , of spawning a star particle in a time Δt is a function of the local dynamical time t_{form} :

$$p = \frac{m_{\text{gas}}}{m_{\text{star}}} (1 - e^{-c_0^* X_{H_2} \Delta t / t_{\text{form}}}) \quad (1)$$

where m_{gas} is the mass of the gas particle and m_{star} is the initial mass of the potential star particle. A star formation efficiency parameter, $c_0^* = 0.1$, multiplied by the fraction of nonionized hydrogen in H_2 , X_{H_2} , gives the correct normalization of the Kennicutt–Schmidt relation (Christensen et al. 2014).

We adopt the “blastwave” supernova feedback approach (Stinson et al. 2006), in which mass, thermal energy, and metals are deposited into nearby gas when massive stars evolve into supernovae. The thermal energy deposited among those nearby gas neighbors is 1.5×10^{51} erg per supernova event. Subsequently, gas cooling is turned off until the end of the momentum-conserving phase of the supernova blastwave. The coupling of the supernova thermal energy into the interstellar medium, combined with the turning off of cooling in the affected gas particles, is designed to mimic the effect of energy deposited in the local interstellar medium (ISM) by *all* processes related to young stars, including UV radiation from massive stars (see Wise et al. 2012; Agertz et al. 2013).

It is the recent success of simulations in matching dwarf galaxy properties (e.g., Governato et al. 2010; Brooks & Zolotov 2014; Shen et al. 2014) that allow us to undertake this work. At the dwarf galaxy scale, the different slopes between the observed galaxy SMF and the Λ CDM predicted halo mass function require that galaxies at halo masses below $10^{11} M_{\odot}$ have gas cooling and star formation efficiencies much lower than those of Milky Way–sized galaxies. While this trend was historically difficult to produce in cosmological simulations, recent high-resolution cosmological simulations that resolve scales on the order of giant molecular clouds can include more realistic models for star formation and feedback, resulting in simulations that can successfully reproduce the observed trends in star formation efficiency (Brook et al. 2012; Aumer et al. 2013; Munshi et al. 2013; Simpson et al. 2013; Stinson et al. 2013; Hopkins et al. 2014; Governato et al. 2015; Wheeler et al. 2015; Christensen et al. 2016; Fitts et al. 2017). The success of models in reproducing reliable and accurate dwarf galaxies lies in the ability to be able to resolve the impact of baryonic processes on the interstellar medium and star formation (Christensen et al. 2014; Munshi et al. 2014). When this happens, the simulations also simultaneously reproduce additional observed trends in dwarf galaxies, such as cored dark matter density profiles (Governato et al. 2012; Di Cintio et al. 2014; Pontzen & Governato 2014; Maxwell et al. 2015;

⁷ Due to different criteria for inclusion and different halo definitions, our sample, while overlapping with Applebaum et al. (2021), includes fewer galaxies from Sandra and Elena.

Oñorbe et al. 2015; Read et al. 2016a; Dutton et al. 2019) and bulgeless disks (Brook et al. 2011; Brooks & Christensen 2016).

In the following Sections we show that the stochasticity of star formation and mass loss of satellites after infall add to the scatter in the relationship between stellar mass and halo mass. We do so with simulations that, a priori, require no further tuning to successfully match observed properties, including published mass–metallicity relationships (Brooks et al. 2007; Christensen et al. 2018), cold gas fractions (Munshi et al. 2013; Brooks et al. 2017), dark matter profile shapes (Governato et al. 2012), and a host of observed scaling relations for Local Group dwarfs (Applebaum et al. 2021). A future work (Munshi et al., in prep.) will present the Marvel dwarf properties in full detail.

Individual halos are identified using AMIGA’S HALO FINDER⁸ (AHF, Gill et al. 2004; Knollmann & Knebe 2009). Throughout this work, the virial radius of a halo is defined to be the radius for which the average halo density is 200 times the critical density of the universe at a given redshift, $200 \rho_{\text{crit}}(z)$. For subhalos, AHF identifies the virial radius as the point where the lowest density is reached before the density profile increases again due to the contribution from the parent halo. For all halos in this work, we trace back the main progenitor to find the peak halo mass that the halo attained, defined as M_{peak} . At each snapshot, the main progenitor is defined to be the halo in the previous step that contains the majority of the particles in the current halo.

3. Results

In Figure 1, we show the stellar mass of the simulated galaxies as a function of $z=0$ halo mass (left panel) and as a function of peak halo mass (right panel). Data points for the simulated galaxies are colored based on the galaxy V-band magnitude⁹ at $z=0$. The stellar masses for dwarfs brighter than a V-band magnitude of -8 are calculated using each simulated galaxy’s photometric color, as described in Munshi et al. (2013). This photometric correction reduces the total simulated stellar masses by $\sim 40\%$, primarily because of an aperture correction when assuming SDSS photometry, which many of the abundance matching techniques have used to derive the SMHM relation at higher masses.¹⁰ For the fainter galaxies, we adopt the simulation stellar mass because the stellar mass of observed UFDs is generally based on resolved star counts rather than photometric color. Circles represent central (isolated) galaxies, squares represent backplash galaxies, and stars indicate galaxies that are satellites at $z=0$. Smaller data points are galaxies from the two Near Mint DC Justice League simulations, while the larger data points are from the high-resolution Marvel simulations and two Mint DC Justice League simulations. It can be seen that the brightest galaxies in this sample are roughly LMC-mass ($M_{\text{halo}} \sim 10^{11} M_{\odot}$; Kallivayalil et al. 2013; Besla 2015; Peñarrubia et al. 2016; Dooley et al.

2017), while the faintest galaxies are UFDs (e.g., $M_{\text{star}} < 10^5 M_{\odot}$, $V \gtrsim -8$; Simon 2019).

The scatter is greatest if using $z=0$ halo masses (left panel), due to the inclusion of satellite galaxies (shown by stars) that can be stripped of their dark matter after infall to their parent halo. In some cases subhalos lose more than an order of magnitude in halo mass after infall. Stellar masses are more robust, and only those halos with significant halo stripping have lost about a factor of two in stellar mass. This is consistent with earlier findings that $\sim 90\%$ of the dark matter mass can be stripped before stars are stripped (Muñoz et al. 2008; Peñarrubia et al. 2008; Libeskind et al. 2011; Chang et al. 2013; Brooks & Zolotov 2014). Considering the peak halo mass (right panel) for the galaxies reduces the scatter, though the lowest-mass halos still host almost 2 orders of magnitude in stellar mass. The peak halo mass is typically used in all previous derivations of SMHM relations, including those shown in Figure 1.

The galaxies in Figure 1 were all chosen to have resolved star formation histories, which we define to be star formation timescales that span more than 100 Myr. This choice ensures that, despite multiple supernovae having occurred, the star formation in these galaxies is robust to the feedback. This choice will exclude galaxies from our simulated sample that undergo a single burst of fast star formation and then quench, but it is not clear if such an event is reliable since it is sensitive to the resolution of the star particles and the feedback prescription. In practice, our choice of resolved star formation histories leads to a lower limit of 14 star particles in the Marvel and Mint DC Justice League galaxies, and 3 star particles in the Near Mint DC Justice League simulations. Despite the potentially low number of particles in some Near Mint Justice League galaxies, the SMHM relation of the faintest DC Justice League galaxies blends smoothly into the relation of the higher-resolution dwarfs (3 stars is $\sim 1.5 \times 10^4 M_{\odot}$), indicating that there are no obvious resolution effects impacting the SMHM results shown here. We examine the effect of resolution more carefully in Appendix A.

At the faintest end of the SMHM relation presented in Figure 1, a galaxy cannot be assigned a unique halo mass based solely on its luminosity. As stellar masses decline, the range of halo masses that host a given galaxy increases. Likewise, as halo masses decline, the scatter in the stellar masses of the simulated galaxies increases.

3.1. Quantifying the Scatter

In this Section we quantify the scatter in our SMHM relation. We use the M_{peak} results for halo masses, as is commonly adopted for abundance matching or halo occupation studies. We use the results from these fits to make predictions presented later in the paper.

3.1.1. Well-resolved Halos

Figure 1 shows that the scatter in the SMHM relation increases with decreasing halo mass. To quantify the scatter, we assume that stellar masses can be described by log-normal scatter about a mean SMHM relation.

At a given halo mass, the scatter in stellar mass is of order 0.3 dex for halos more massive than $\sim 10^{10} M_{\odot}$. The scatter in stellar mass increases to 1.2 dex at the smallest halo masses for these well-resolved halos. With this scatter, there is no longer a

⁸ AHF is available for download at <http://popia.ft.uam.es/AHF/Download.html>.

⁹ The V-band magnitude is calculated in *pynbody* (Pontzen et al. 2013), which utilizes the Padova simple stellar population models (Marigo et al. 2008; Girardi et al. 2010) found at <http://stev.oapd.inaf.it/cgi-bin/cmd>.

¹⁰ Munshi et al. (2013) found that aperture size was the primary factor in determining photometric stellar masses, and that mass-to-light ratios played a subdominant role.

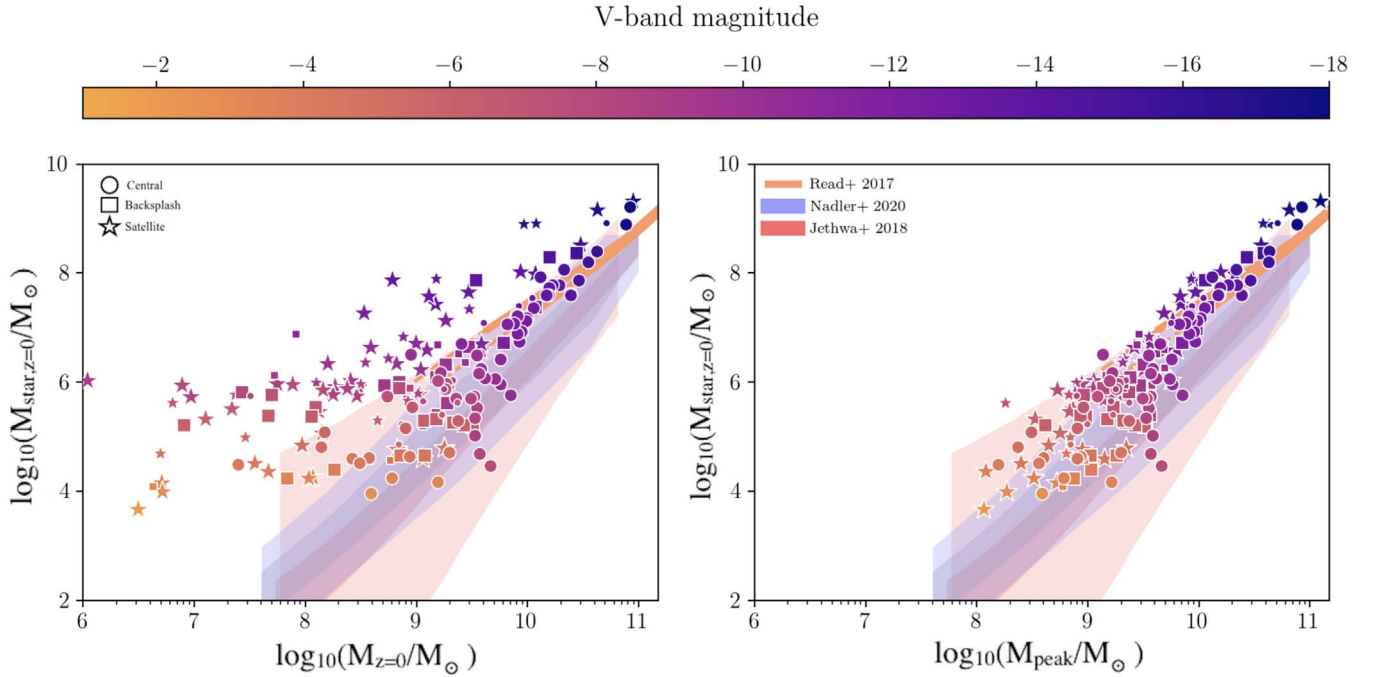


Figure 1. *Stellar Mass versus Halo Mass.* Simulated galaxies are color-coded by their V-band magnitude at $z = 0$. **Left:** Stellar masses at $z = 0$ of the galaxies in the Marvel and Justice League simulations versus their $z = 0$ halo mass. **Right:** Stellar masses at $z = 0$ versus their peak halo mass. In both panels we display results derived in previous works (Read et al. 2017; Jethwa et al. 2018; Nadler et al. 2020). The red contours correspond to $P(M_{\text{star}}|M_{\text{vir}})$ from Jethwa et al. (2018) which overlap the blue contours from Nadler et al. (2020). Darker colors for each author represent the 68% confidence interval, while the lighter colors represent the 95% confidence interval. The stellar masses of classical dwarfs (with M_V brighter than -8) are calculated based on photometric colors (see Munshi et al. 2013), while fainter dwarfs use stellar masses directly from the simulations to mimic resolved star counts in UFDs. Galaxies represented by circles are isolated galaxies at $z = 0$, while galaxies represented by stars are satellites, and squares are backplash galaxies of the Milky Way-mass halos. Small points are galaxies from the Near Mint DC Justice League simulations, and larger points are from the Marvel suite and Mint DC Justice League runs. While the scatter is decreased by considering M_{peak} , the scatter in the relation increases with decreasing halo mass.

single stellar to halo mass relationship for the faintest dwarf galaxies.

Garrison-Kimmel et al. (2017) demonstrated the impact of scatter in the SMHM relation on the resulting SMF. They explored both a model in which the scatter is constant as a function of halo mass, and a model in which the scatter increases as halo mass declines. Clearly, our results favor a model in which the scatter increases toward low halo masses. Garrison-Kimmel et al. (2017) quantify the increasing scatter as follows:

$$\sigma = \sigma_0 + \gamma(\log_{10} M_{\text{halo}} - \log_{10} M_1) \quad (2)$$

where σ is the scatter, γ is the rate as which the scatter grows, and M_1 is a characteristic mass above which the scatter remains constant, at a value σ_0 . We fit our relationship with a broken power law that breaks at $M_1 = 10^{10} M_{\odot}$. Our high-mass slope is 1.9 above the knee, and the low-mass slope is 2.0 below the knee. Above M_1 , we assume a constant scatter of $\sigma_0 = 0.3$ dex, consistent with scatter studies at higher masses (Behroozi et al. 2013), though at the highest masses, the scatter may actually be much smaller (Bradshaw et al. 2020). Using Equation (2) with the well-resolved simulated galaxies (and photometry based stellar masses for M_V brighter than -8) yields $\gamma = -0.43$.

We have verified that the fit remains unchanged if we instead use stellar masses directly from the simulations. Because the primary effect of the photometric correction to the stellar masses is to shift the masses of the brighter galaxies down uniformly by $\sim 40\%$, the overall slope and scatter in the SMHM remains largely unchanged.

The knee in our SMHM at M_1 corresponds visibly to a knee in our resolved halos (see also Figures 2 and 5). Our knee corresponds to a halo mass above which all halos in our simulations host a galaxy, but not below. We discuss the origin of the knee further in Section 4.2. Using only the halos that are well resolved from a simulation is inherently resolution dependent. We address this in the Section below. However, the results of using only the well-resolved halos confirm that the scatter is in fact growing, and can be described by log-normal growing scatter.

3.1.2. All Halos

Next we extend the discussion to include dark halos: because of limited resolution, some or all of our dark halos may in fact host a galaxy with a stellar mass below our mass resolution. As such, we seek to estimate the underlying SMHM relation in a way that is fully independent of our resolution. To address this, we take the limiting assumption that our unresolved halos would be populated if we were not limited by our mass resolution and model (subgrid) assumptions. This limiting assumption is supported by earlier work (Jethwa et al. 2018; Nadler et al. 2020) that shows that all surviving halos down to $M_{\text{peak}} = 3 \times 10^8 M_{\odot}$ must be occupied in order to match the completeness-corrected abundance of Milky Way satellites. Assuming fully populated halos steepens the best-fit SMHM relation at low masses (below the knee, where halos start to become unoccupied) due to the inclusion of less massive galaxies at a given halo mass.

As in the previous section, we model the SMHM as a broken power law. In Appendix B we show that a broken power law

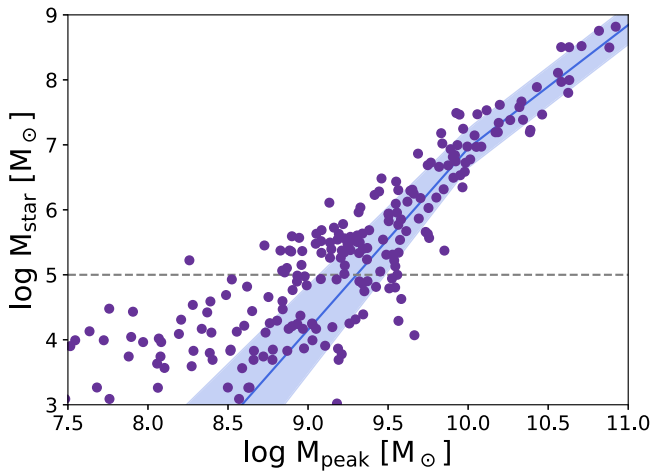


Figure 2. Stellar mass vs. halo mass with fitted relation. We show the $z=0$ stellar masses and peak halo masses of all simulated galaxies (including galaxies that did not meet the “well resolved” criteria for Figure 1). The solid line shows the best-fit mean relation following the fitting procedure that accounts for the presence of “dark” halos; see Section 3.1.2 for details. The solid band shows the best-fit log-normal scatter in stellar mass (NB: the band does not represent uncertainty in the best-fit SMHM relation). The best-fit relation is a broken power law separated at $\log(M_{\text{peak}}) = 10$. The high-mass end is described by a slope $\alpha = 1.9$ and constant scatter $\sigma = 0.3$ dex, and the low-mass end by $\alpha = 2.81$ and σ given by Equation (2) with $\gamma = -0.39$, $\sigma_0 = 0.3$, and $\log M_1 = 10$. Accounting for dark halos steepens the underlying SMHM relation. We include the dashed line to indicate the stellar mass above which we are complete, and above which we fit our model.

and growing scatter are both statistically preferred over a single power law or constant scatter. For $\log(M_{\text{peak}}) \geq 10$, for which all halos in our simulations host galaxies (see Figure 4), we fit one slope. For $\log(M_{\text{peak}}) < 10$, we fit another slope, with growing scatter according to Equation (2). We enforce continuity at the break in both the mean stellar mass, and the scatter, so that we apply Equation (2) with $\log M_1 = 10$, $\sigma_0 = 0.3$.

We fit the SMHM relation in a manner that does not assign individual (unresolved) stellar masses to dark halos. Rather, we assume that stellar masses are log-normally distributed along the entire relation, *including for halos that form no stars in our simulations*. “Dark” halos are treated as populated galaxies that are merely below our stellar mass resolution. In doing so, we are able to avoid making assumptions about the halo occupation fraction as a function of mass, which in simulations is a function of underlying physics as well as resolution. We discuss this further below.

For the SMHM relation below $\log(M_{\text{peak}}) = 10$, we find our parameters θ (i.e., the slope α and scatter growth rate γ) with the following procedure. We divide the SMHM space into $N_k = N_i \times N_j$ bins, with bins $i \in \{1, \dots, N_i\}$ in $\log(M_{\text{peak}})$ and bins $j \in \{1, \dots, N_j\}$ in $\log(M_{\text{star}})$. We count the number of galaxies in each bin $k = (i, j)$. The likelihood of finding the set of simulated galaxies, $\mathbf{n}$, given our SMHM model parameters, θ , is

$$P(\mathbf{n}|\theta) = \prod_{k=1}^{N_k} P(n_k|\lambda_k), \quad (3)$$

where n_k is the number of simulated galaxies in bin k , and λ_k is the mean number of galaxies in bin k predicted by the model at a given θ . The likelihood assumes that n_k is complete in the stellar mass range of the bin, with no systematic undercounts

due to resolution issues. The counts in each bin can be modeled by a Poisson distribution, so that

$$P(n_k|\lambda_k) = \frac{\lambda_k^{n_k} e^{-\lambda_k}}{n_k!}. \quad (4)$$

Let $\xi(m|\theta)$ give the mean log stellar mass at a given log halo mass, m . Since the range of possible stellar masses is normally distributed about $\xi(m)$, the probability that a halo with mass m , in bin i , will host a galaxy in stellar mass bin j (with bin upper and lower limits of u and ℓ , respectively), will be

$$p(j|m, \theta) = \int_u^\ell \mathcal{N}(\xi(m), \sigma) = F(\ell) - F(u), \quad (5)$$

where $F(x)$ is the cumulative density function of the normal distribution with mean μ and standard deviation σ ,

$$F(x|\mu, \sigma) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{x - \mu}{\sigma\sqrt{2}} \right) \right]. \quad (6)$$

The expected number of galaxies in bin $k = (i, j)$ according to our model is therefore¹¹

$$\lambda_k = n_i p(j), \quad (7)$$

where n_i is the total number of halos in $\log(M_{\text{peak}})$ bin i , including dark and occupied halos.

To summarize, our likelihood is given by Equations (3) and (4), with the model galaxy counts given by Equation (7). We note that the likelihood compares for each bin the expected number of galaxies, λ_k , to the number of simulated galaxies in the bin, n_k . Therefore, for bins in which we calculate the likelihood, we must ensure that we are not systematically undercounting the number of simulated galaxies due to resolution limitations. In other words, we have to restrict our bins to a range of stellar mass in which we can be reasonably sure we are complete. We expect the fit to be less dependent on resolution limitations affecting our identification of dark matter halos because both λ_k and n_k are proportional to the number of halos in a given bin. We thus fit to galaxy counts in the range $7.5 \leq \log(M_{\text{peak}}) \leq 11$ and $5 \leq \log(M_{\text{star}}) \leq 9$. We found that binning above $\log(M_{\text{star}}) = 5$ led to converged results, while including fainter galaxies in the fit led to a steeper slope. We also found varying the lower bounds on $\log(M_{\text{peak}})$ had little effect. We interpret this to mean that we have resolved all galaxies above $\log(M_{\text{star}}) = 5$, and below this there is an extended (i.e., non-step function) transition until we no longer resolve any galaxies.

We find the best-fitting parameters using the affine-invariant Markov Chain Monte Carlo sampler EMCEE (Foreman-Mackey et al. 2013), assuming flat priors in the region $0 < \alpha < 5$ and $-2 < \gamma < 0$. We run 10^4 steps using 32 walkers, and discard a burn-in period of 500 steps, or ~ 15 times the autocorrelation length.

Using the above procedure, the best-fit SMHM below $\log(M_{\text{peak}}) = 10$ has a slope $\alpha = 2.81^{+0.12}_{-0.11}$, with a growing scatter given by Equation (2), with $\gamma = -0.39^{+0.05}_{-0.06}$, $\log M_1 = 10$, and $\sigma_0 = 0.3$ dex (recall, the latter two values were fixed as boundary conditions). The best-fit relation is shown in Figure 2. As expected, this relation is steeper than

¹¹ Formally, we would have to calculate $p(j)$ separately for every m in bin i . In practice, it is sufficient to calculate it once for each bin, using the mean m of the bin. Our results are insensitive to bin size, indicating this approximation is robust.

that found by just fitting to occupied halos, which bias the underlying relation to higher, resolved masses. Readers may further explore how resolution and scatter can bias the inferred slope of the SMHM in an interactive applet¹², which allows users to adjust parameters of a basic SMHM toy model.

3.1.3. Comparison to Previous Work

Our fit to all halos has growing log-normal scatter toward lower halo masses, and reaches $\sigma > 1.0$ dex at $\log(M_{\text{peak}}) = 8$. We are not aware of any other simulations that have allowed scatter to be quantified in the SMHM relation at these low masses, but we can compare our results to simulations of isolated dwarfs at varying masses. The large scatter in our simulations encompasses the range seen by Rey et al. (2019), who found that a single UFD galaxy could vary by an order of magnitude in stellar mass at $M_{\text{peak}} = 1.5 \times 10^9 M_{\odot}$. Likewise, the simulated dwarf galaxies of Revaz & Jablonka (2018) and the NIHAO galaxies in Buck et al. (2019) show a comparable amount of scatter in M_{star} at a $z = 0$ halo mass of $10^9 M_{\odot}$ as in our Figure 1. Our galaxies extend to slightly lower stellar masses than those in Revaz & Jablonka (2018), as might be expected due to our higher stellar mass resolution. The spread in Buck et al. (2019) is comparable to our Figure 1, despite their lower resolution, because they plot galaxies with as few as one star particle, while our Figure 1 shows only well-resolved galaxies.

However, Fitts et al. (2017) simulated 15 field dwarfs with halo masses $\sim 10^{10} M_{\odot}$ and a range of halo concentrations, and found a range of $M_{\text{star}} = 10^{5-7} M_{\odot}$, as well as one dark halo. This appears inconsistent with our fixed scatter of 0.3 dex at a similar halo mass. The FIRE scheme used by Fitts et al. (2017) is stronger and burstier in dwarf galaxies than in the Marvel and Justice League simulations (Iyer et al. 2020), resulting in a steeper SMHM relation in FIRE than is presented here, as was also noted by Revaz & Jablonka (2018). The distribution of stellar masses in Fitts et al. (2017) is also not well described by a log-normal distribution, but rather with a distribution that peaks at higher M_{star} and has a tail to low M_{star} . This again may be due to the stronger feedback in FIRE, but may also be influenced by their chosen range of halo concentration.

By matching Local Group galaxies to the ELVIS catalog (Garrison-Kimmel et al. 2014a), Garrison-Kimmel et al. (2014b) found that a large log-normal scatter of 2 dex, constant across M_{peak} , alleviated abundance matching discrepancies. If there is no scatter in the SMHM relation, then the number of predicted Local Group and Local Field galaxies in the range $5 < \log(M_{\text{star}}/M_{\odot}) < 6$ is larger than currently observed. Increasing the scatter to 2 dex while steepening the SMHM relation reduces the predicted number, because the halo mass function is steep in CDM, causing more low-mass halos to scatter up than high-mass halos to scatter down. Therefore, a given stellar mass is hosted by lower-mass halos than might otherwise be expected. Likewise, Jethwa et al. (2018) find that scatter in the SMHM relation leads to the inference of a shallow SMHM relation if we measure halo mass at a given stellar mass, $P(M_{\text{vir}}|M_{\text{star}})$. They find that scatter can help to explain the fact that faint field dwarfs (Ural et al. 2015; Read et al. 2016b) are sitting in lower-mass halos than might otherwise be expected.

At face value, our results appear to be in tension with Nadler et al. (2020), who found that the scatter in stellar mass (or luminosity, which is the quantity they used instead) must be small, $\sigma < 0.2$ dex, at all halo masses. There are a couple of differences in our analysis that can bring our results closer together, though not necessarily reconcile the difference. First, they anchor their growing scatter model at a larger M_1 value of $10^{11} M_{\odot}$, while we anchor at a lower mass of $10^{10} M_{\odot}$. If we consider that γ is the rate of change of scatter, including the higher-mass halos where scatter is essentially constant should decrease γ as the scatter does not grow between $10^{11} M_{\odot}$ to $10^{10} M_{\odot}$. Fitting all of our resolved halos with a single slope and growing scatter model anchored at $10^{11} M_{\odot}$ (instead of growing scatter only below the knee, anchored at $10^{10} M_{\odot}$), our value for γ decreases from -0.39 to -0.21 .

Also, Nadler et al. (2020) examine scatter in luminosity as a function of v_{peak} , while we are tracing based on M_{peak} . Previous works have shown that M_{star} correlates more strongly with v_{peak} , due to the effects of halo assembly bias (e.g., Reddick et al. 2013; Chaves-Montero et al. 2016). This correlation should lead to smaller scatter in the $M_{\text{star}}-v_{\text{peak}}$ relation than in the SMHM relation. We have verified that our scatter is reduced when using v_{peak} , e.g., decreasing our scatter above the knee from 0.3 dex to 0.17 dex. However, our scatter still increases below the knee, blowing up to over 1 dex at the lowest v_{peak} values we trace, as with M_{peak} . Thus, while the change in variables can explain some difference in the results, it does not explain it all.

Finally, Nadler et al. (2020) found that the scatter must be small to reproduce the observed LF of Pan-STARRS1 and the Dark Energy Survey (DES, Bechtol et al. 2015; Drlica-Wagner et al. 2015); with large scatter, low-mass halos host satellite galaxies that scatter to observable luminosities too often, and they were unable to reproduce the satellite LF of the Milky Way. The Milky Way may (Carlsten et al. 2021; Mao et al. 2021) or may not (Wang et al. 2021) have a typical satellite LF for a galaxy of its luminosity, but the four Justice League simulations used here have been shown to match the range of satellite LFs of observed Milky Way–mass galaxies (Akens et al. 2021; Applebaum et al. 2021). Nadler et al. (2020) derived their constraints using only two Milky Way realizations, which could impact their scatter results. A full accounting of this discrepancy requires further exploration.

3.2. Scatter and the Stellar Mass Function

In this subsection, we use our predicted slopes and scatter in the SMHM relation in order to calculate a predicted SMF. To do this, we draw 1000 random halos with $10^{7.8} \leq M_{\text{peak}} \leq 10^{11.5}$ from the ELVIS catalogs (Garrison-Kimmel et al. 2014b), and populate them with stellar masses according to our predicted SMHM slopes and scatter. The resulting SMFs are shown in Figure 3 for three cases: (i) no scatter in the SMHM relation (“No Scatter”), (ii) constant log-normal scatter of 0.6 dex in M_{star} at a given M_{peak} (“Constant Scatter”, see Appendix B), and (iii) adopting our results for growing scatter from Section 3.1.2 (“Growing Scatter”). For the Growing Scatter case, recall that we adopt a slope of the SMHM $\alpha = 1.9$ for halos from $10 < \log(M_{\text{peak}}) < 11.5$. We adopt a steeper slope of $\alpha = 2.8$ at lower halo masses, utilizing the slope derived from probabilistically populating halos below our resolution limit. The scatter is constant at 0.3 dex above $M_1 = 10^{10} M_{\odot}$, and linearly grows below this halo mass, i.e.,

¹² Available at <https://github.com/emapple/smhbm-toy-model>.

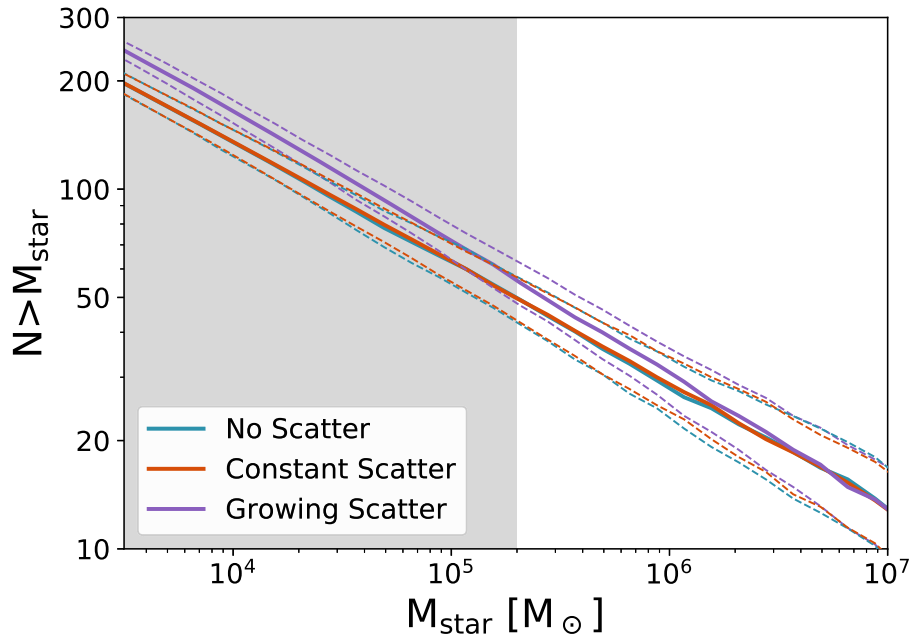


Figure 3. The effect of SMHM scatter on the predicted stellar mass function. Constant Scatter is very similar to No Scatter, both in slope and normalization. Growing Scatter, however, steepens the SMF. We generate each SMF 1000 times, and show in the solid lines the median relation and in the dashed lines the inner 68% range. We have additionally scaled each SMF so that they have 12 galaxies above $10^7 M_\odot$, consistent with the number within 1 Mpc of the Milky Way. The shaded region represents the approximate discovery space of UFD galaxies.

$\gamma = -0.38$. We generate each SMF 1000 times, and show in the solid lines the median relation with the inner 68% range indicated by the dashed lines. We have scaled each SMF so that they have 12 galaxies above $10^7 M_\odot$, consistent with the number within 1 Mpc of the Milky Way.

Garrison-Kimmel et al. (2017) demonstrated that large scatter in the SMHM relation impacts the observed SMF. Due to the rapidly rising halo mass function, there are more small halos available to scatter to larger stellar masses than large halos to scatter to lower stellar masses, an effect that is increasingly noticeable as scatter increases. The change from the No Scatter case to the Constant Scatter case amounts to a uniform shift upwards in the number of galaxies of a given stellar mass in log space in the SMF. This shift upwards disappears after we rescale the SMF to $N = 12$ for $M_{\text{star}} > 10^7 M_\odot$, which is why the No Scatter and Constant Scatter mean SMF lie on top of each other in Figure 3 (note that this will occur no matter the magnitude of the scatter chosen for the Constant Scatter case). If scatter is not constant, however, and grows with decreasing halo mass, then the shift in the SMF is not uniform. Hence, the Growing Scatter SMF steepens in Figure 3.

Our scatter grows in such a way that it only reaches relatively large scatter (>1 dex) at very low halo masses that host primarily UFD galaxies. This is where the scatter begins to have a noticeable impact on the SMF, compared to the Constant Scatter and No Scatter cases (though we caution that this is also the range below which our simulations are considered complete, as discussed in Section 3.1.2). The mass range where there is greatest difference between SMFs corresponds to luminosities where current observations are incomplete, but are being or will be probed by surveys like DES, HSC-SSP, or with the Vera Rubin Observatory (Tollerud et al. 2008; Walsh et al. 2009). In Figure 3, we shade the mass region of the new dwarf satellites that have been found in the first two years of the DES (Bechtol et al. 2015; Drlica-Wagner

et al. 2015). Tollerud et al. (2008) estimates that such faint dwarf galaxies should be observed out to ~ 1 Mpc by the Vera Rubin Observatory’s LSST after the full co-added data are collected. Thus, the slope of the SMF, when complete, can possibly constrain the magnitude of scatter in the SMHM, and possibly the smallest halo that hosts a galaxy.

4. Discussion

In this Section we discuss various factors that might influence our predicted SMHM relation.

4.1. Occupation Fraction

The occupation fraction in the real universe—the fraction of halos at a given mass that host a galaxy—is dependent on many physical processes (discussed further below), including the physics of star formation, gas cooling, self-shielding, and the strength and timing of reionization. Quantifying this “intrinsic” occupation fraction in a cosmological simulation would require the ability to resolve large numbers of both the smallest halos that can host galaxies as well as the smallest stellar mass that can constitute a galaxy, which is still observationally unconstrained. Given that galaxies have been observed with masses as low as $\sim 10^2 M_\odot$ (e.g., Drlica-Wagner et al. 2015; Homma et al. 2018; Longeard et al. 2018), cosmological simulations to date do not have the ability to reliably resolve the intrinsic occupation fraction. Rather, whether a halo is “occupied” or “dark” in the simulations is subject to resolution.

We show this simplistically in Figure 4 by choosing various threshold stellar masses below which all dark matter halos are devoid of stars (i.e., do not host a galaxy). This, in effect, is varying the mass resolution for galaxies in our simulations. We also indicate (black dashed line) the occupation fraction we would infer from our simulations if we did not recognize that it

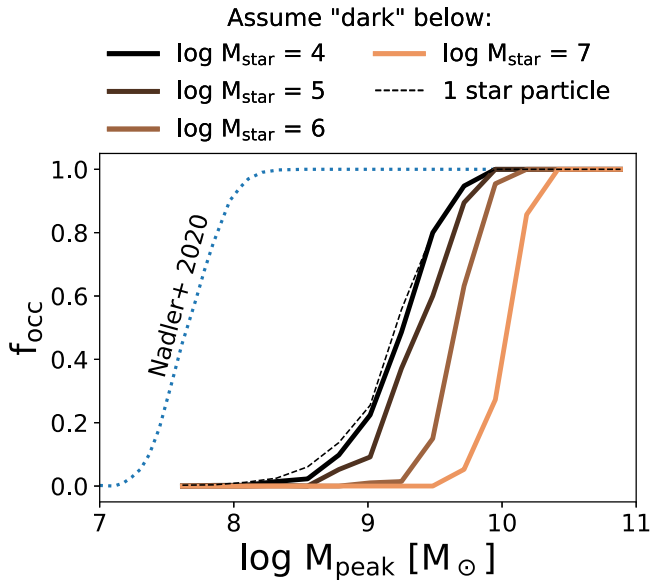


Figure 4. The dependence of occupation fraction on simulation resolution. As simulations have finite mass resolution, the minimum mass of an occupied halo is inherently tied to resolution. We demonstrate that the simulated occupation fraction changes significantly when the minimum resolvable stellar mass (and in effect, the mass resolution) is varied. Both the peak halo mass where galaxies go “dark” and how sharply the curve declines to zero are affected. The sharpness of the decline increases as resolution decreases as a result of the decreasing scatter in the SMHM relation at higher masses. In the dotted line we show the occupation fraction that we would infer from the simulations if we assumed that halos that host at least one star particle contain galaxies. For comparison, we show the occupation fraction inferred by Nadler et al. (2020), which is constrained by Milky Way observations.

is resolution dependent, assuming all halos with at least one star particle host a galaxy.¹³ As we increase the threshold stellar mass, not only does the halo mass where galaxies go “dark” increase, the sharpness of the simulated occupation fraction becomes closer to a step function. The sharper decline comes from the smaller scatter in the SMHM relation at higher stellar masses.

Both Jethwa et al. (2018) and Nadler et al. (2020), using observations combined with a model for the galaxy–halo connection, find that essentially all halos with peak mass above $\sim 3 \times 10^8 M_\odot$ need to host a galaxy in order to be consistent with the completeness-corrected observations of dwarf galaxies. Jethwa et al. followed galaxies down to $M_V = -1.5$, while Nadler et al. modeled galaxies as faint as $M_V = 0$ ($M_{\text{star}} \sim 100 M_\odot$). Our simulations are unable to resolve galaxies this faint. The inferred occupation fraction derived by Nadler et al. (2020) is shown by the blue dotted line in Figure 4. While our higher-resolution simulations have $z=0$ star particle masses as low as $174 M_\odot$, the occupation fraction inferred from the simulations drops to zero at substantially higher M_{peak} than in Nadler et al. This is because we start to be unable to resolve galaxies that have low stellar mass that would reside in halos with $M_{\text{peak}} < 10^9 M_\odot$. This motivates the approach we have taken in this work, in which we do not impose an occupation fraction a priori, nor do we impose the occupation fraction explicitly defined by our simulations, as we know this to be resolution dependent. Instead, we assume that

stellar masses are log-normally distributed along the entire SMHM relationship, as our resolved halos show a log-normal growing scatter. In doing so, we treat dark halos as populated, but populated below our resolution limit. This approach then allows us to make predictions for the SMF to highlight the effect of scatter in the SMHM relation, removing the effect of a resolution-dependent occupation fraction.

We note that this approach of treating the occupation fraction as resolution dependent is very different to what has been assumed in the past. It has long been thought that reionization should remove baryons from low-mass halos, leaving a population of low-mass halos that are completely “dark” and without a galaxy. Early works suggested that the halo mass cutoff for galaxy formation was fairly massive (e.g., Benson et al. 2002; Somerville et al. 2004). The idea that reionization should imprint itself at a halo mass that is somewhere on the order of 10^8 to $10^{10} M_\odot$ has persisted, leaving simulators (e.g., Sawala et al. 2015) to treat their halo occupation fractions as predictions rather than recognizing the resolution dependence inherent in the results. It is only the recent reconciliation of observed UFD galaxy counts with halo counts that have forced a re-examination of this assumption (see also Kim et al. 2018; Newton et al. 2018; Read & Erkal 2019).

However, if some halos with a peak mass greater than $10^8 M_\odot$ are truly dark, then this will have a strong impact on the resulting SMF. For example, Munshi et al. (2019) showed that the star formation prescription we use in this work, in which H_2 must be present for stars to form, inhibits the formation of UFD galaxies in halos with $\log M_{\text{halo}} > 8.5$. They also showed that a different star formation prescription, which required high densities and low temperatures but not necessarily the presence of H_2 , allowed galaxies to form in halos at least an order of magnitude lower in mass. It is not clear that either of these models is truly reflective of star formation in UFDs, but the result demonstrates that there may be an influence beyond reionization that impacts the halo occupation distribution. A comparison of our Figure 3 with Figure 6 of Munshi et al. (2019) shows that occupation fraction can have a much stronger impact on the resulting SMF than the large scatter in the low-mass SMHM relation we have found in this work.

4.2. Reionization

Our SMHM relation has a bend at $\sim 10^{10} M_\odot$ below which the slope of the SMHM steepens. This is true independent of cosmology or resolution, as can be seen in Appendix A, where the knee is readily apparent. This knee is also reflected in the full sample in Figure 1, though the large number of dwarf galaxies and large scatter at low masses make it less obvious visually. This knee in the relationship, dividing the SMHM into two slopes, is an expected result of reionization. There will be some halo mass (the knee) below which star formation is suppressed, either due to gas loss, gas suppression, and/or gas ionization. Gnedin (2012) finds that the contribution of ionized gas becomes significant in galaxies with $v_{\text{max}} \sim 40\text{--}50 \text{ km s}^{-1}$. Our knee at $10^{10} M_\odot$ corresponds to $v_{\text{max}} \sim 40 \text{ km s}^{-1}$. The onset of reionization more strongly impacts the earlier forming (i.e., lower-mass) halos (Bose et al. 2018; Benitez-Llambay & Frenk 2020; Bose et al. 2020). The impact of reionization at a given halo mass may depend on local density and timing of the onset of reionization (discussed below), but it follows that there should be two different slopes in the SMHM relation, with a

¹³ We use all halos with one star particle despite the two different resolutions used in this work. The lower-resolution simulations will reach zero occupation at a slightly higher M_{peak} than the higher-resolution simulations, but the number of halos in that mass range is so large that it has little impact on f_{occ} .

knee below which reionization is suppressing star formation in halos (see also Garrison-Kimmel et al. 2019).

As discussed in Munshi et al. (2019), many of the low-mass halos in our H_2 star formation model are not able to form stars before reionization prevents them from doing so. This implies that our results are sensitive to our selected reionization model. We have adopted the same model in all simulation volumes, following Haardt & Madau (2012). However, Haardt & Madau (2012) has been shown to heat the intergalactic medium earlier ($z \sim 15$) than it should (Oñorbe et al. 2015), potentially making the impact of reionization particularly strong on our results.

Arguably the more important limitation of our reionization model is the fact that it is uniform throughout the simulation volume. This is common for cosmological galaxy simulations, as the radiative transfer required to explicitly follow patchy reionization is computationally expensive. Simulations that self-consistently model reionization find that the baryon fraction of low-mass halos is highly dependent on the timing of reionization due to local densities, with suppression due to reionization occurring earlier in higher density environments (Wu et al. 2019; Katz et al. 2020; Ocvirk et al. 2020). The Marvelous Dwarfs are $\sim 1.5\text{--}7$ Mpc away from a Milky Way-mass galaxy, meaning that, in a more realistic reionization scenario, they are in a lower density region that may not ionize as early as the higher density regions surrounding the Justice League Milky Ways. The fact that our full sample blurs out the knee in the SMHM relation more than near individual massive galaxies (compare Figure 1 to Appendix A) likely points to an environmental dependence on reionization at a fixed halo mass even with our spatially uniform UV background. However, a more realistic reionization model may allow some of our dark halos in our lower density regions to form stars. Benítez-Llambay & Frenk (2020) explicitly look at the effect of reionization on the $z=0$ occupation fraction, and find that it varies with how early or late reionization begins. Understanding the impact of reionization will require further simulations and study, but will be essential to explore the uncertainties that will impact the interpretation of LSST observations.

4.3. Star Formation and Feedback Model

As discussed in Section 4.1, Munshi et al. (2019) finds that there is a significant reduction in overall efficiency of star formation in simulated UFD galaxies when adopting a nonequilibrium H_2 -based star formation prescription relative to a prescription that adopts a temperature-density threshold. The reduction in star formation in the H_2 model is due the long formation times of H_2 at low metallicities in low-mass halos. This yields a significant difference in the number of predicted galaxies at low masses and results in a different predicted SMF between the two star formation recipes. In this paper, we adopt the same H_2 -based star formation model as in Munshi et al. (2019) across all eight volumes, and use our large sample of dwarf galaxies to test the effect of the slope and scatter of the SMHM on the predicted SMF. Figure 3 shows that there is no appreciable change in the faint-end slope of the SMF or number of predicted UFDs when considering No Scatter versus the Constant Scatter case. However, in the Growing Scatter case, the SMF is steepened toward lower masses. A comparison with Figure 6 of Munshi et al. (2019) shows that the occupation fraction imposed by the star formation prescription can have a much larger effect on the number of

predicted UFDs: the large scatter in the SMHM relation derived in this paper impacts the number of UFDs by $\sim 25\%$. The star formation prescription contributes a comparable change *if* the slope of the SMHM relation is unchanged. If the star formation prescription also changes the slope of the SMHM relation, then the number of UFDs can be a factor of 2–3 different.

Alternatively, Byrne et al. (2019) demonstrate that tying star formation to shielded gas rather than the presence of H_2 allows stars to form in lower density gas. This may change the resulting SMHM in the regime where dwarf galaxies do not self-regulate. Latif & Khochfar (2019) find H_2 self-shielding is critical: in addition to the delay in star formation, the collapse and evolution of halos is tied to the strength of the UV background. Furthermore, our H_2 model does not let stars form (at our current resolution) in halos below $10^8 M_\odot$ at $z=6$. Skinner & Wise (2020) and Latif & Khochfar (2019) show halos below this mass should be forming Population III stars, which are not included in our model. Both emphasize the balance between UV flux and self-shielding that sets the halo mass that can form stars, and thus the $z=0$ occupation fraction. These results emphasize that the number of newly discovered UFDs by LSST will place constraints on both the process of star formation and the UV background at high redshift.

In addition to the star formation recipe, feedback strength and implementation varies between simulations. Galaxy stellar masses are sensitive to specific feedback implementations. For example, Agertz et al. (2020) find that varying feedback models changes the stellar mass of their test halo by over 1 dex, but still within the scatter of our SMHM relation. In simulations run with CHANGA and GASOLINE, superbubble feedback (Keller et al. 2014) leads to a factor of ~ 2 reduction in stellar mass in Milky Way-like halos (Keller et al. 2015). Mina et al. (2021) simulate dwarf galaxies using superbubble feedback, and they find that it has a varying effect on the SMHM relation. A larger sample is needed to assess any potential systematic effect of superbubble feedback on dwarf stellar masses. Specific results may also be sensitive to feedback details like the inclusion of radiative feedback. For example, Smith et al. (2021) show that ionizing radiation reduces supernova clustering, leading to suppressed supernova-driven outflows, while Smith et al. (2019) stress that the strength of supernova feedback in dwarfs may also be highly dependent on ISM turbulence, runaway massive stars, and early stellar feedback. To constrain these varying physical processes, it may be that a more detailed accounting of the observed stellar mass fraction as a function of halo mass (e.g., Read & Erkal 2019; Romeo et al. 2020) is required.

Finally, coupled with nonsupernova feedback, the effects of IMF sampling on star formation are more difficult to predict (Smith 2021). Applebaum et al. (2020) shows that implementing a stochastically populated initial mass function (IMF) in CHANGA can systematically lower stellar masses in UFDs.

In summary, how galaxies populate dark matter halos in simulations is not only resolution dependent, but also tied to both the star formation model and reionization. Since our choice of star formation model and our reionization model can affect the slope and scatter of halos we predict, we are unable to say that the number of UFDs predicted in this paper is exactly what LSST should expect to find. However we *can* constrain the effect of scatter in the SMHM relation on predictions compared to a no scatter case as long as our underlying models and assumptions remain constant.

5. Summary

In this paper, we use a large sample of extremely high resolution simulated dwarf galaxies (LMC-mass to UFDs) in a range of environments in order to predict the SMHM relation at low halo masses. This is the first prediction of the SMHM relation for dwarf galaxies this low in mass at $z=0$, and the first time that the scatter at the low-mass end has been robustly quantified. In doing so, we demonstrate that (1) derived SMHM relations cannot be simply extrapolated from higher masses to lower masses, and thus (2) the halo mass of a faint dwarf galaxy cannot be inferred a priori.

Our SMHM relation is best fit by including a break in the slope of the SMHM relation. This break corresponds to a visible knee in the relation, above which all halos are occupied by simulated galaxies, and below which we start encountering dark halos. This knee is the natural consequence of reionization. The details of the reionization model will matter for the exact location of the knee. In this work, in which we adopt a uniform UV background based on Haardt & Madau (2012), we find a break at a halo mass of $10^{10} M_{\odot}$. The slope above the knee is $\alpha = 2.0$, and below the knee it steepens to $\alpha = 2.8$.

The scatter in stellar mass at a given (peak) halo mass is constant above the knee at 0.3 dex. Below the knee, some halos may truly be dark due to reionization, but it is also likely that these low-mass halos can host galaxies that are below our mass resolution limit. Assuming that all halos are occupied, and that the log-normal scatter that grows linearly about our best-fit slope of $\alpha = 2.8$ below the knee, the rate of increase in scatter is quantified by γ , where $\gamma = -0.38$ (see Equation (2)).

When using $z=0$ halo masses, the scatter in the relation is much larger, due to the inclusion of satellite and backsplash galaxies in the SMHM relation. The satellites of both our dwarf and Milky Way galaxies as well as backsplash galaxies can lose substantial mass after infall, including satellites that lose an order of magnitude or more in halo mass. The results presented above instead use peak halo mass, similar to earlier abundance matching studies, which significantly tightens the SMHM relation.

Our best-fit SMHM relation slopes and scatters can be used to stochastically populate theoretical models at low masses instead of relying on abundance matching. We demonstrate this by populating a halo mass function and showing the predicted SMF (see Figure 3), and exploring the effect of scatter on the SMF. The resulting SMF is essentially unchanged when considering the Constant Scatter and No Scatter cases, due to the fact that the SMF is normalized to have 12 galaxies above $M_{\text{star}} = 10^7 M_{\odot}$ within 1 Mpc, comparable to the Local Group. However, the faint-end slope of the SMF is steepened in the case of Growing Scatter. More UFD galaxies are predicted if there is Growing Scatter in the SMHM relation. The luminosity/mass range where this can be tested is currently being probed by surveys like DES, which discovered 16 new UFDs in its first two years (Bechtol et al. 2015; Drlica-Wagner et al. 2015), and the HSC-SSP, which has discovered three dwarfs (Homma et al. 2016, 2018, 2019). Additional dwarfs should be discovered by the Rubin Observatory’s LSST when it comes online, allowing the SMF in this range to be probed observationally.

This work is part of a series that has explored uncertainties in modeling UFD galaxies. Applebaum et al. (2020) showed that adopting a stochastic IMF can alter the stellar masses of UFD galaxies. Munshi et al. (2019) showed that two commonly

adopted simulation star formation prescriptions can yield different low-mass SMHM relations and occupation fractions. In this work we have discussed the role of the occupation fraction, the reionization model, and the star formation model on our resulting SMHM relation. Because of these uncertainties, we are unable to say that our predicted SMF is what LSST is going to find. However, by systematically exploring these effects we can quantify the uncertainties in the predicted SMF given known unknowns, and begin to search for ways to break degeneracies.

Despite the caveats, the large intrinsic scatter in our simulated SMHM demonstrates that the commonly adopted assumption of a monotonic relationship between stellar mass and halo mass that is adopted in abundance matching breaks down at low masses. Reionization should impose a break in the relation at low masses. At the faintest end of the SMHM probed by our simulations, a galaxy cannot be assigned a unique halo mass based solely on its stellar mass or luminosity. The lack of a monotonic relation between stellar mass and halo mass has implications for interpreting observations of dwarf galaxies.

FDM acknowledges support from the Vanderbilt Initiative for Data-Intensive Astrophysics (VIDA) through a VIDA Postdoctoral Fellowship, and from the University of Oklahoma. Long before the University of Oklahoma was established, the land on which the University now resides was the traditional home of the “Hasinai” Caddo Nation and Kikirikis Wichita & Affiliated Tribes. FDM acknowledges this territory once also served as a hunting ground, trade exchange point, and migration route for the Apache, Comanche, Kiowa and Osage nations. Today, 39 tribal nations dwell in the state of Oklahoma as a result of settler and colonial policies that were designed to assimilate Native people. FDM and AMB acknowledge support from HST AR-13925 provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. FDM and SS acknowledge support from NSF grant PHY-2013909. AMB acknowledges support from HST AR-14281. EA and AMB acknowledge support from NSF grant AST-1813871. EA acknowledges support from the National Science Foundation (NSF) Blue Waters Graduate Fellowship. CRC acknowledges support from the NSF under CAREER grant No. AST-1848107. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. This research also made use of the NSF supported Frontera project operated by the Texas Advanced Computing Center (TACC) at The University of Texas at Austin. This research was supported in part by the National Science Foundation under grant No. NSF PHY-1748958.

The authors thank Yao-Yuan Mao, Ethan Nadler, Martin Rey, Priya Natarajan, and Jordan Van Nest for assistance and helpful comments that improved this manuscript.

Appendix A

The Effect of Cosmology and Resolution on Scatter

As mentioned in Section 2, the Marvel dwarfs are run within a WMAP Year 3 cosmology (Spergel et al. 2007), while the Justice League simulations are run with a Planck cosmology (Planck Collaboration et al. 2016). In the top panel of Figure 5 we examine the effect of different cosmologies on the resulting

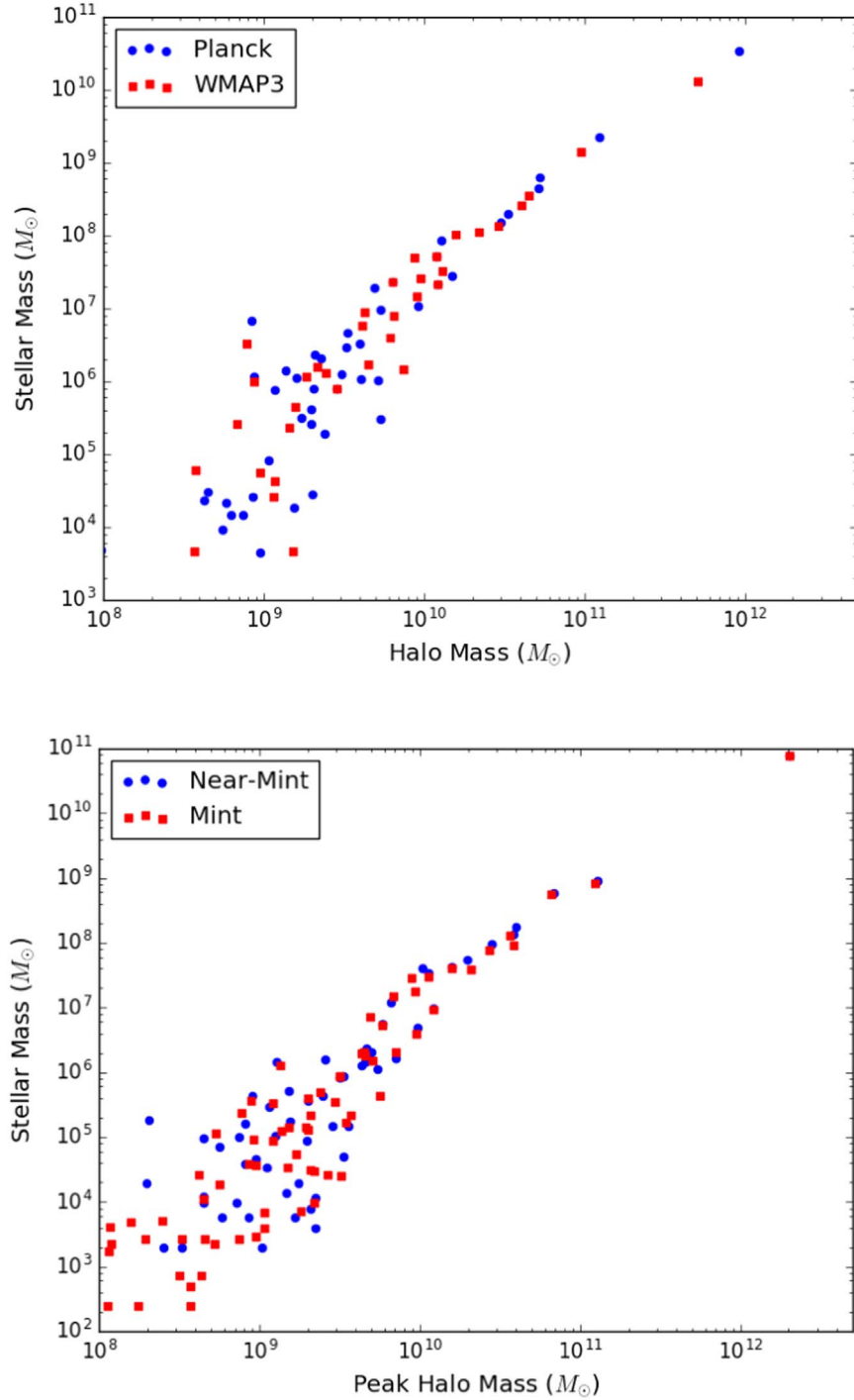


Figure 5. SMHM relationship for various versions of Sandra. Top panel: we compare $z = 1.4$ results from both the WMAP3 cosmology and the Planck cosmology at NM resolution to show convergence between different cosmologies. Bottom panel: we show $z = 0$ results using a Planck cosmology between both the NM run and Mint run. Both resolutions converge to a similar SMHM. Both panels use simulation stellar masses, rather than photometric.

SMHM relation. The top panel of Figure 5 compares the $z = 1.4$ results for Sandra’s (one of the Justice League Milky Ways) SMHM in both WMAP3 cosmology (red points) and in Planck cosmology (blue points), at the same NM resolution. We find that the SMHM relation is consistent across cosmologies. To quantify, we fit a single slope and scatter across the entire range of masses. That is, rather than fit a “well-resolved” sample as in Section 3.1.1, or all halos (including dark) as in Section 3.1.2, we fit all halos that contain at least one star particle, i.e., any “occupied” halos. For

reference, if we fit a single slope and scatter for our occupied sample at $z = 0$ instead of the two-component fit, the slope is $\alpha = 2.0$ with a scatter of $\gamma = -0.21$. The slope and scatter of the $z = 1.4$ SMHM relations with WMAP3 and Planck cosmologies are consistent with each other and with the occupied sample to within one sigma ($\alpha = 1.97/\gamma = -0.2$ and $\alpha = 2.0/\gamma = -0.2$, respectively).

In the bottom panel of Figure 5 we compare the $z = 0$ NM and Mint Sandra simulations in a Planck cosmology to test the effect of resolution. The stellar masses converge between the

two simulations (see also Applebaum et al. 2021), although with the increased Mint resolution we resolve more UFDs. Here, too, we find similar slopes and scatters between the two resolutions, within one sigma of each other ($\alpha = 2.02/\gamma = -0.2$ and $\alpha = 2.05/\gamma = -0.2$ for NM and Mint, respectively). Since the cosmological comparison is at a higher redshift and both comparisons (resolution and cosmology) only include one simulation, we compare slope and scatter in the above substantially simplified form.

Appendix B Model Details

In Section 4.2 we have discussed that a broken power law is physically motivated. Here, we demonstrate that a broken power law joined at $\log M_{\text{peak}} = 10$ is also statistically preferred over a single power-law relation. To show this, we compare the Akaike Information Criterion (AIC; Akaike 1974) of the two models, using the likelihood described in Section 3.1.2. In both cases we use Equation (2) for the scatter, with $\sigma_0 = 0.3$ dex above $M_1 = 10^{10} M_{\odot}$ and growing scatter below, such that the two-part power law only has one additional fitted parameter (the second slope). We find the two-part power law is preferred, with $\text{AIC}_1 - \text{AIC}_2 = 19.8$. Even if we were to include the (fixed) location of the knee and the value of σ_0 as additional degrees of freedom, the two-part power law would still be preferred. We find that the best-fit slope with a single power law is $\alpha = 2.38$, with growing scatter given by $\gamma = -0.26$ (note that this fit is to all halos, including dark halos). In other words, the high-mass slope is steeper and the low-mass slope is shallower compared to the two-part power law. While neither regime is dramatically changed, both are poorer fits to the data, resulting in the preference of the two-part power law.

To test whether the growing scatter is statistically preferred over a constant scatter, we once again compare AIC values. As above, we compare the model from Section 3.1.2 to two otherwise identical models: instead of using Equation (2) for σ , we either (1) leave a constant value of σ_0 as a free parameter, or (2) use $\sigma = 0.3$ dex for $\log M_{\text{peak}} \geq 10$ and a constant σ_0 as a free parameter for lower masses only. We find the former case yields a best-fit scatter of 0.55 dex and $\Delta\text{AIC} = 24.3$ (which is also inconsistent with the observed scatter at the high-mass end), while the latter yields a scatter (below M_1) of 0.62 dex and $\Delta\text{AIC} = 9.5$. Both of the constant scatter models are disfavored, then, compared to the growing scatter model.

ORCID iDs

Ferah Munshi  <https://orcid.org/0000-0002-9581-0297>
 Alyson M. Brooks  <https://orcid.org/0000-0002-0372-3736>
 Elaad Applebaum  <https://orcid.org/0000-0001-8301-6152>
 Charlotte R. Christensen  <https://orcid.org/0000-0001-6779-3429>

References

- Agertz, O., Kravtsov, A. V., Leitner, S. N., & Gnedin, N. Y. 2013, *ApJ*, **770**, 25
- Agertz, O., Pontzen, A., Read, J. I., et al. 2020, *MNRAS*, **491**, 1656
- Akaike, H. 1974, *ITAC*, **19**, 716
- Alkins, H. B., Christensen, C. R., Brooks, A. M., et al. 2021, *ApJ*, **909**, 139
- Alavi, A., Siana, B., Richard, J., et al. 2016, *ApJ*, **832**, 56
- Anderson, L., Governato, F., Karcher, M., Quinn, T., & Wadsley, J. 2017, *MNRAS*, **468**, 4077
- Applebaum, E., Brooks, A. M., Christensen, C. R., et al. 2021, *ApJ*, **906**, 96
- Applebaum, E., Brooks, A. M., Quinn, T. R., & Christensen, C. R. 2020, *MNRAS*, **492**, 8
- Aumer, M., White, S. D. M., Naab, T., & Scannapieco, C. 2013, *MNRAS*, **434**, 3142
- Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015, *ApJ*, **807**, 50
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, **770**, 57
- Benítez-Llambay, A., & Frenk, C. 2020, *MNRAS*, **498**, 4887
- Benson, A. J. 2020, *MNRAS*, **493**, 1268
- Benson, A. J., Frenk, C. S., Lacey, C. G., Baugh, C. M., & Cole, S. 2002, *MNRAS*, **333**, 177
- Besla, G. 2015, arXiv:1511.03346
- Bose, S., Deason, A. J., Belokurov, V., & Frenk, C. S. 2020, *MNRAS*, **495**, 743
- Bose, S., Deason, A. J., & Frenk, C. S. 2018, *ApJ*, **863**, 123
- Bradshaw, C., Leauthaud, A., Hearin, A., Huang, S., & Behroozi, P. 2020, *MNRAS*, **493**, 337
- Brook, C. B., & Di Cintio, A. 2015, *MNRAS*, **450**, 3920
- Brook, C. B., Di Cintio, A., Knebe, A., et al. 2014, *ApJL*, **784**, L14
- Brook, C. B., Governato, F., Roškar, R., et al. 2011, *MNRAS*, **415**, 1051
- Brook, C. B., Stinson, G., Gibson, B. K., Wadsley, J., & Quinn, T. 2012, *MNRAS*, **424**, 1275
- Brooks, A., & Christensen, C. 2016, *Galactic Bulges*, **418**, 317
- Brooks, A. M., Governato, F., Booth, C. M., et al. 2007, *ApJL*, **655**, L17
- Brooks, A. M., Kuhlen, M., Zolotov, A., & Hooper, D. 2013, *ApJ*, **765**, 22
- Brooks, A. M., Papastergis, E., Christensen, C. R., et al. 2017, *ApJ*, **850**, 97
- Brooks, A. M., & Zolotov, A. 2014, *ApJ*, **786**, 87
- Buck, T., Macciò, A. V., Dutton, A. A., Obreja, A., & Frings, J. 2019, *MNRAS*, **483**, 1314
- Byrne, L., Christensen, C., Tsekitsidis, M., Brooks, A., & Quinn, T. 2019, *ApJ*, **871**, 213
- Carlsten, S. G., Greene, J. E., Peter, A. H. G., Beaton, R. L., & Greco, J. P. 2021, *ApJ*, **908**, 109
- Chang, J., Macciò, A. V., & Kang, X. 2013, *MNRAS*, **431**, 3533
- Chaves-Montero, J., Angulo, R. E., Schaye, J., et al. 2016, *MNRAS*, **460**, 3100
- Christensen, C., Quinn, T., Governato, F., et al. 2012, *MNRAS*, **425**, 3058
- Christensen, C. R., Davé, R., Brooks, A., Quinn, T., & Shen, S. 2018, *ApJ*, **867**, 142
- Christensen, C. R., Davé, R., Governato, F., et al. 2016, *ApJ*, **824**, 57
- Christensen, C. R., Governato, F., Quinn, T., et al. 2014, *MNRAS*, **440**, 2843
- Conroy, C., & Wechsler, R. H. 2009, *ApJ*, **696**, 620
- Côté, B., Silvia, D. W., O'Shea, B. W., Smith, B., & Wise, J. H. 2018, *ApJ*, **859**, 67
- Di Cintio, A., Brook, C. B., Macciò, A. V., et al. 2014, *MNRAS*, **437**, 415
- Dooley, G. A., Peter, A. H. G., Carlin, J. L., et al. 2017, *MNRAS*, **472**, 1060
- Drlica-Wagner, A., Bechtol, K., Rykoff, E. S., et al. 2015, *ApJ*, **813**, 109
- Dutton, A. A., Macciò, A. V., Buck, T., et al. 2019, *MNRAS*, **486**, 655
- Ferrero, I., Abadi, M. G., Navarro, J. F., Sales, L. V., & Gurovich, S. 2012, *MNRAS*, **425**, 2817
- Fitts, A., Boylan-Kolchin, M., Bullock, J. S., et al. 2018, *MNRAS*, **479**, 319
- Fitts, A., Boylan-Kolchin, M., Elbert, O. D., et al. 2017, *MNRAS*, **471**, 3547
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Garrison-Kimmel, S., Boylan-Kolchin, M., Bullock, J. S., & Kirby, E. N. 2014a, *MNRAS*, **444**, 222
- Garrison-Kimmel, S., Boylan-Kolchin, M., Bullock, J. S., & Lee, K. 2014b, *MNRAS*, **438**, 2578
- Garrison-Kimmel, S., Bullock, J. S., Boylan-Kolchin, M., & Bardwell, E. 2017, *MNRAS*, **464**, 3108
- Garrison-Kimmel, S., Hopkins, P. F., Wetzell, A., et al. 2019, *MNRAS*, **487**, 1380
- Gill, S. P. D., Knebe, A., & Gibson, B. K. 2004, *MNRAS*, **351**, 399
- Girardi, L., Williams, B. F., Gilbert, K. M., et al. 2010, *ApJ*, **724**, 1030
- Gnedin, N. Y. 2000, *ApJ*, **535**, 530
- Gnedin, N. Y. 2012, *ApJ*, **754**, 113
- Governato, F., Brook, C., Mayer, L., et al. 2010, *Natur*, **463**, 203
- Governato, F., Weisz, D., Pontzen, A., et al. 2015, *MNRAS*, **448**, 792
- Governato, F., Willman, B., Mayer, L., et al. 2007, *MNRAS*, **374**, 1479
- Governato, F., Zolotov, A., Pontzen, A., et al. 2012, *MNRAS*, **422**, 1231
- Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, *MNRAS*, **404**, 1111
- Haardt, F., & Madau, P. 2012, *ApJ*, **746**, 125
- Homma, D., Chiba, M., Komiyama, Y., et al. 2019, *PASJ*, **71**, 94
- Homma, D., Chiba, M., Okamoto, S., et al. 2016, *ApJ*, **832**, 21
- Homma, D., Chiba, M., Okamoto, S., et al. 2018, *PASJ*, **70**, S18
- Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, *MNRAS*, **445**, 581
- Iyer, K. G., Tacchella, S., Genel, S., et al. 2020, *MNRAS*, **498**, 430

- Jethwa, P., Erkal, D., & Belokurov, V. 2018, *MNRAS*, **473**, 2060
- Kalé, L., & Krishnan, S. 1993, in Proceedings of OOPSLA'93, ed. A. Paepcke (New York: ACM Press), 91
- Kallivayalil, N., van der Marel, R. P., Besla, G., Anderson, J., & Alcock, C. 2013, *ApJ*, **764**, 161
- Katz, H., Ramsay, M., Rosdahl, J., et al. 2020, *MNRAS*, **494**, 2200
- Keller, B. W., Wadsley, J., Benincasa, S. M., & Couchman, H. M. P. 2014, *MNRAS*, **442**, 3013
- Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2015, *MNRAS*, **453**, 3499
- Kim, S. Y., Peter, A. H. G., & Hargis, J. R. 2018, *PhRvL*, **121**, 211302
- Klypin, A., Karachentsev, I., Makarov, D., & Nasonova, O. 2015, *MNRAS*, **454**, 1798
- Knollmann, S. R., & Knebe, A. 2009, *ApJS*, **182**, 608
- Kravtsov, A. V., Vikhlinin, A. A., & Meshcheryakov, A. V. 2018, *AstL*, **44**, 8
- Kuhlen, M., & Faucher-Giguère, C.-A. 2012, *MNRAS*, **423**, 862
- Latif, M. A., & Khochfar, S. 2019, *MNRAS*, **490**, 2706
- Libeskind, N. I., Knebe, A., Hoffman, Y., Gottlöber, S., & Yepes, G. 2011, *MNRAS*, **418**, 336
- Longard, N., Martin, N., Starkenburg, E., et al. 2018, *MNRAS*, **480**, 2609
- Mao, Y.-Y., Geha, M., Wechsler, R. H., et al. 2021, *ApJ*, **907**, 85
- Marigo, P., Girardi, L., Bressan, A., et al. 2008, *A&A*, **482**, 883
- Matthee, J., Schaye, J., Crain, R. A., et al. 2017, *MNRAS*, **465**, 2381
- Maxwell, A. J., Wadsley, J., & Couchman, H. M. P. 2015, *ApJ*, **806**, 229
- Menon, H., Wesolowski, L., Zheng, G., et al. 2015, *ComAC*, **2**, 1
- Mina, M., Shen, S., Keller, B. W., et al. 2021, *A&A*, **655**, A22
- Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, **428**, 3121
- Muñoz, R. R., Majewski, S. R., & Johnston, K. V. 2008, *ApJ*, **679**, 346
- Munshi, F., Brooks, A. M., Christensen, C., et al. 2019, *ApJ*, **874**, 40
- Munshi, F., Christensen, C., Quinn, T. R., et al. 2014, *ApJL*, **781**, L14
- Munshi, F., Governato, F., Brooks, A. M., et al. 2013, *ApJ*, **766**, 56
- Nadler, E. O., Wechsler, R. H., Bechtol, K., et al. 2020, *ApJ*, **893**, 48
- Nakatani, R., Fialkov, A., & Yoshida, N. 2020, *ApJ*, **905**, 151
- Newton, O., Cautun, M., Jenkins, A., Frenk, C. S., & Helly, J. C. 2018, *MNRAS*, **479**, 2853
- Ocvirk, P., Aubert, D., Sorce, J. G., et al. 2020, *MNRAS*, **496**, 4087
- Ocvirk, P., Gillet, N., Shapiro, P. R., et al. 2016, *MNRAS*, **463**, 1462
- Okamoto, T., Gao, L., & Theuns, T. 2008, *MNRAS*, **390**, 920
- Oñorbe, J., Boylan-Kolchin, M., Bullock, J. S., et al. 2015, *MNRAS*, **454**, 2092
- O'Shea, B. W., Wise, J. H., Xu, H., & Norman, M. L. 2015, *ApJL*, **807**, L12
- Papastergis, E., Giovanelli, R., Haynes, M. P., & Shankar, F. 2015, *A&A*, **574**, A113
- Peñarrubia, J., Gómez, F. A., Besla, G., Erkal, D., & Ma, Y.-Z. 2016, *MNRAS*, **456**, L54
- Peñarrubia, J., McConnachie, A. W., & Navarro, J. F. 2008, *ApJ*, **672**, 904
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Pontzen, A., & Governato, F. 2014, *Natur*, **506**, 171
- Pontzen, A., Roškar, R., Stinson, G., & Woods, R. 2013, pynbody: N-Body/SPH analysis for python, Astrophysics Source Code Library, ascl:1305.002
- Read, J. I., Agertz, O., & Collins, M. L. M. 2016a, *MNRAS*, **459**, 2573
- Read, J. I., & Erkal, D. 2019, *MNRAS*, **487**, 5799
- Read, J. I., Iorio, G., Agertz, O., & Fraternali, F. 2016b, *MNRAS*, **462**, 3628
- Read, J. I., Iorio, G., Agertz, O., & Fraternali, F. 2017, *MNRAS*, **467**, 2019
- Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, *ApJ*, **771**, 30
- Revaz, Y., & Jablonka, P. 2018, *A&A*, **616**, A96
- Rey, M. P., Pontzen, A., Agertz, O., et al. 2019, *ApJL*, **886**, L3
- Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, *ApJ*, **768**, 71
- Romeo, A. B., Agertz, O., & Renaud, F. 2020, *MNRAS*, **499**, 5656
- Sawala, T., Frenk, C. S., Crain, R. A., et al. 2013, *MNRAS*, **431**, 1366
- Sawala, T., Frenk, C. S., Fattahi, A., et al. 2015, *MNRAS*, **448**, 2941
- Sawala, T., Frenk, C. S., Fattahi, A., et al. 2016, *MNRAS*, **456**, 85
- Shen, S., Madau, P., Conroy, C., Governato, F., & Mayer, L. 2014, *ApJ*, **792**, 99
- Shen, S., Wadsley, J., & Stinson, G. 2010, *MNRAS*, **407**, 1581
- Simon, J. D. 2019, *ARA&A*, **57**, 375
- Simpson, C. M., Bryan, G. L., Johnston, K. V., et al. 2013, *MNRAS*, **432**, 1989
- Skinner, D., & Wise, J. H. 2020, *MNRAS*, **492**, 4386
- Smith, M. C. 2021, *MNRAS*, **502**, 5417
- Smith, M. C., Bryan, G. L., Somerville, R. S., et al. 2021, *MNRAS*, **506**, 3882
- Smith, M. C., Sijacki, D., & Shen, S. 2019, *MNRAS*, **485**, 3317
- Somerville, R. S., Lee, K., Ferguson, H. C., et al. 2004, *ApJL*, **600**, L171
- Spergel, D. N., Bean, R., Doré, O., et al. 2007, *ApJS*, **170**, 377
- Stinson, G., Seth, A., Katz, N., et al. 2006, *MNRAS*, **373**, 1074
- Stinson, G. S., Brook, C., Macciò, A. V., et al. 2013, *MNRAS*, **428**, 129
- Strigari, L. E., Bullock, J. S., Kaplinghat, M., et al. 2008, *Natur*, **454**, 1096
- Taylor, E. N., Cluver, M. E., Duffy, A., et al. 2020, *MNRAS*, **499**, 2896
- Tollerud, E. J., Bullock, J. S., Strigari, L. E., & Willman, B. 2008, *ApJ*, **688**, 277
- Ural, U., Wilkinson, M. I., Read, J. I., & Walker, M. G. 2015, *NatCo*, **6**, 7599
- Wadsley, J. W., Keller, B. W., & Quinn, T. R. 2017, *MNRAS*, **471**, 2357
- Wadsley, J. W., Stadel, J., & Quinn, T. 2004, *NewA*, **9**, 137
- Walsh, S. M., Willman, B., & Jerjen, H. 2009, *AJ*, **137**, 450
- Wang, W., Takada, M., Li, X., et al. 2021, *MNRAS*, **500**, 3776
- Wechsler, R. H., & Tinker, J. L. 2018, *ARA&A*, **56**, 435
- Weisz, D. R., & Boylan-Kolchin, M. 2017, *MNRAS*, **469**, L83
- Weisz, D. R., Dolphin, A. E., Skillman, E. D., et al. 2015, *ApJ*, **804**, 136
- Wheeler, C., Hopkins, P. F., Pace, A. B., et al. 2019, *MNRAS*, **490**, 4447
- Wheeler, C., Oñorbe, J., Bullock, J. S., et al. 2015, *MNRAS*, **453**, 1305
- Wise, J. H., Abel, T., Turk, M. J., Norman, M. L., & Smith, B. D. 2012, *MNRAS*, **427**, 311
- Wu, X., Kannan, R., Marinacci, F., Vogelsberger, M., & Hernquist, L. 2019, *MNRAS*, **488**, 419